Chanter avec les mains : interfaces chironomiques pour les instruments de musique numériques

Olivier Perrotin

Université Paris-Sud

Ecole Doctorale 427
Informatique Paris-Sud

Laboratoire : LIMSI-CNRS

Thèse de Doctorat

Spécialité Informatique

Présentée par :

Olivier Perrotin

Chanter avec les mains : Interfaces chironomiques pour les instruments de musique numériques

Soutenue le Mercredi 23 septembre 2015 devant le jury composé de :

Myriam Desainte-Catherine	Professeur ENSEIRB / LaBRI	Rapporteur
Marcelo M. Wanderley	Professeur MacGill University / CIRMMT	Rapporteur
Michel Beaudouin-Lafon	Professeur Université Paris-Sud / LRI	Examinateur
Boris Doval	Maître de conférences UPMC / d'Alembert	Examinateur
Thierry Dutoit	Professeur Université de Mons / TCTS Lab	Examinateur
Christophe d'Alessandro	Directeur de Recherche CNRS / LIMSI	Directeur de thèse

Groupe Audio et Acoustique
LIMSI-CNRS
Campus universitaire Bâtiment 508
Rue John von Neumann
91405 Orsay Cedex, France

EDIPS
Université Paris-Sud - ED 427
UFR Sciences Orsay - Bat 650 PCRI - aile nord
Rue Noetzlin
91405 Orsay Cedex, France

Résumé

Chanter avec les mains : Interfaces chironomiques pour les instruments de musique numériques

Le travail de cette thèse porte sur l'étude du contrôle en temps réel de synthèse de voix chantée par une tablette graphique dans le cadre de l'instrument de musique numérique *Cantor Digitalis*. La pertinence de l'utilisation d'une telle interface pour le contrôle de l'intonation vocale a été traitée en premier lieu, démontrant que la tablette permet un contrôle de la hauteur mélodique plus précis que la voix réelle en situation expérimentale. Pour étendre la justesse du jeu à toutes situations, une méthode de correction dynamique de l'intonation a été développée, permettant de jouer en dessous du seuil de perception de justesse et préservant en même temps l'expressivité du musicien. Des évaluations objective et perceptive ont permis de valider l'efficacité de cette méthode. L'utilisation de nouvelles interfaces pour la musique pose la question des modalités impliquées dans le jeu de l'instrument. Une troisième étude révèle une prépondérance de la perception visuelle sur la perception auditive pour le contrôle de l'intonation, due à l'introduction d'indices visuels sur la surface de la tablette. Néanmoins, celle-ci est compensée par l'important pouvoir expressif de l'interface. En effet, la maîtrise de l'écriture ou du dessin dès l'enfance permet l'acquisition rapide d'un contrôle expert de l'instrument. Pour formaliser ce contrôle, nous proposons une suite de gestes adaptés à différents effets musicaux rencontrés dans la musique vocale. Enfin, une pratique intensive de l'instrument est réalisée au sein de l'ensemble *Chorus Digitalis* à des fins de test et de diffusion. Un travail de recherche artistique est conduit tant dans la mise en scène que dans le choix du répertoire musical à associer à l'instrument. De plus, un retour visuel dédié au public a été développé, afin d'aider à la compréhension du maniement de l'instrument.

Mots-clefs : Instrument de musique numérique, interface gestuelle, tablette graphique, synthèse vocale, voix chantée.

Abstract

Singing with hands: Chironomic interfaces for digital musical instruments

This thesis deals with the real-time control of singing voice synthesis by a graphic tablet, based on the digital musical instrument *Cantor Digitalis*. The relevance of the graphic tablet for the intonation control is first considered, showing that the tablet provides a more precise pitch control than real voice in experimental conditions. To extend the accuracy of control to any situation, a dynamic pitch warping method for intonation correction is developed. It enables to play under the pitch perception limens preserving at the same time the musician's expressivity. Objective and perceptive evaluations validate the method efficiency. The use of new interfaces for musical expression raises the question of the modalities implied in the playing of the instrument. A third study reveals a preponderance of the visual modality over the auditive perception for the intonation control, due to the introduction of visual clues on the tablet surface. Nevertheless, this is compensated by the expressivity allowed by the interface. The writing or drawing ability acquired since early childhood enables a quick acquisition of an expert control of the instrument. An ensemble of gestures dedicated to the control of different vocal effects is suggested. Finally, an intensive practice of the instrument is made through the *Chorus Digitalis* ensemble, to test and promote our work. An artistic research has been conducted for the choice of the *Cantor Digitalis*' musical repertoire. Moreover, a visual feedback dedicated to the audience has been developed, extending the perception of the players' pitch and articulation.

Keywords: Digital musical instrument, gestural interface, graphic tablet, voice synthesis, singing voice.

Remerciements

Je tiens à remercier en premier lieu Christophe d'Alessandro pour ces trois années d'encadrement. Sa curiosité, sa culture et son imagination débordante m'ont ouvert de nombreuses pistes d'exploration qui ont rendu ces travaux très riches autant sur le plan scientifique, musical, qu'humain. Sa confiance dans mon travail m'a permis d'acquérir une certaine autonomie dans mes projets de recherche. J'ai beaucoup apprécié la dimension musicale de cette thèse qui m'a permis de découvrir des mondes que je connaissais peu. Merci pour ces moments musicaux, des répétitions au LIMSI "pour le travail" aux sessions d'improvisation tardives dans une salle de Georgia Tech. Merci pour tous les autres moments insolites qui rendent cette thèse inoubliable, du choix aléatoire d'un menu en Coréen, au Happy Hour Tacos/Margarita d'Atlanta, en passant par le taillage de moustache en Lorraine.

Je remercie ensuite les rapporteurs de cette thèse, Myriam Desainte-Catherine qui n'a malheureusement pu se déplacer pour la soutenance et Marcelo Wanderley, et les examinateurs Michel Beaudouin-Lafon, Boris Doval et Thierry Dutoit, pour avoir accepté de faire partie du jury et pour leurs remarques, commentaires, questions très pertinentes et encourageantes.

Mes pensées vont ensuite à Lionel Feugère qui en développant sur les bases posées par Sylvain Le Beux a savamment construit le Cantor Digitalis qui a servi de support pour ces travaux. Merci de m'avoir accueilli dans l'équipe, fait découvrir le monde merveilleux de Max/MSP et fait de moi un démêleur de spaghettis aguerri, fait découvrir les joies de l'open source, échangé les versions 1.18b426 du Cantor. Mais aussi en dehors du LIMSI, compagnon de voyage à travers 3 continents, de la Corée aux States en passant par le Clavistan qu'il maîtrise du bout de ses baguettes avec Ô-liostère.

Merci à Boris Doval, un des plus grands chanteurs par tablette graphique au monde, qui m'a fait découvrir la musique indienne, et surtout pour ses discussions toujours intéressantes et porteuses de nouvelles idées sur le geste, sur la manipulation de l'instrument. Mes travaux n'auraient certainement pas été significatifs sans l'aide d'Albert Rilliard, qui a su m'expliquer que tout n'est pas qu'une question de p-value. Merci à Samuel Delalez, arrivé pendant ma dernière année, qui a su mettre le Cantor Digitalis en transe et qui continuera à développer brillamment de nouveaux instruments chanteurs.

Le double aspect scientifique et artistique de ces travaux n'aurait pu avoir lieu sans les répétitions intenses du Chorus Digitalis dont je remercie vivement les membres pour leur patience, leurs idées, leur bonne humeur et leurs talents musicaux venus d'horizons diverses et variées : Annelies, Samuel, Boris, Christophe, Lionel, Simon, Hélène, et tonton Yéyé le special guest pour ses imitations remarquables du Cantor.

Que seraient les journées de travail sans la joie et la bonne humeur permanente des collègues du labo ? Je remercie d'abord les ex-futurs docteurs qui m'ont montré la voie : Gaëtan, Tifanie, à peine arrivé déjà partis, Matthieu plus présent au 334 qu'au LIMSI, David

D. le dealer de cuivre, Paul la force tranquille du B18, Lionel open-tablas, David P.Q. qui met des chansons dans la tête dès 9h du matin, Trang qui aime parler français, Marc double papa en rédaction ; les futurs ex-doctorants : Bart l'historien, Justin philly-ch'ti, Samuel psychotranse ; les 6 post-docs vietnamo-hispano-franco-gréco-britanico-canadiens : Thahn, David G., Laurent, Areti, Peter, Caroline ; ceux de plus court passage : Frédéric, Renaud, Prune, Simon, Maxime, Thomas, Varvara, Daniel, Julie ; et les permanents : Christophe, Albert, Brian, Nathalie, Laurent. Au-delà de AA, merci à tous les collègues du LIMSI, endroit où il fait bon vivre.

Une pensée au 502 bis qui nous a brusquement quitté et dont sa fraîcheur printanière / automnale nous manquera.

Une thèse ça reste dans la tête jour et nuit, et je remercie tous les gens qui m'ont permis de m'évader quelques instants, notamment par la musique, en commençant par les musiciens de l'Atelier Manouche pour trois ans de Gipsy jazz : Manila, Marylou, Etienne, Raphaël, Jérémy, Pierre, Ossama, Alexander, Thibault, Samuel, Augustin, Lucile, Théo, Yvan, Matthieu ; les musiciens de l'Afreubo pour 6 mois de "biture" et qui sont bien trop nombreux à citer ; ceux de l'ensemble Itinér'air pour une escapade en Sologne ; Dustin et Violaine pour ce trio improvisé.

Enfin je remercie tous mes proches pour leurs nombreux encouragements malgré la distance (« En fait tu passes tes journées à jouer de la tablette ? », « Et les consonnes, c'est pour quand ? »). Merci aux amis de plus loin qui m'ont permis de voyager aux quatre coins de l'Europe : Pauline et Clément famille nombreuse, Claire constructrice de chalets Suisses, Camille le routard/couch surfer. Merci à mes parents, mes deux frères pour m'avoir permis d'en arriver là. Enfin, un immense merci à Manny Murphy pour son soutien sans faille, toujours de bon conseil et à l'écoute de mes questionnements, pour la relecture du manuscrit, et surtout pour sa présence qui me permet chaque jour d'avancer.

Table des matières

Notations et expressions

Notation ou expression	Signification
A	Amplitude du geste pour atteindre une cible
ANOVA	Analyse de variance
A_i avec $i \in \{1, 2, 3, 4, 5\}$	Amplitude du filtre formantique i
AV_c	Vitesse moyenne critique (*critical Average Velocity*)
α_m	Coefficient d'asymétrie de l'ODGD
B	Boutons du stylet
B.p.m.	Battements par minute
B_i avec $i \in \{1, 2, 3, 4, 5\}$	Bande-passante du filtre formantique i
CALM	Modèle linéaire causal anticausal (*Causal Anticausal Linear Model*)
Cents	Centièmes de demi-ton
C	Coût d'un mouvement
CS	Pas de correction (*Correction Step*)
CV	Transition Consonne-Voyelle
DPW	Déformation de hauteur dynamique (*Dynamic Pitch Warping*)
DTW	Déformation temporelle dynamique (*Dynamic Time Warping*)
F_i avec $i \in \{1, 2, 3, 4, 5\}$	Fréquence centrale du filtre formantique i
F_0	Fréquence fondamentale de vibration des plis vocaux

Notation ou expression	**Signification**
FOF	Fonctions d'Ondes Formantiques
FSR	Capteur de force résistif (*Force Sensing Resistor*)
I	Intervalle de détection
ID	Indice de difficulté
IHM	Interface Homme-Machine
IP	Indice de performance
Jitter	Perturbation de la fréquence de vibration des cordes vocales
LPC	Codage prédictif linéaire (*Linear Predictive Coding*)
Mapping	Mise en correspondance de paramètres de sortie d'un système vers les paramètres d'entrée d'un deuxième
MIDI	Protocole de communication pour instruments de musique numériques (*Musical Instrument Digital Interface*)
MOS	Note d'opinion moyenne (*Mean Opinion Score*)
MT	Temps moyen
ODG	Onde de Débit Glottique
ODGD	Onde de Débit Glottique Dérivée
O_q	Quotient ouvert de l'ODGD
Phase d'articulation	Position des articulateurs entre deux phonèmes
P	Pression du stylet sur la tablette graphique
Rapport C-D	Rapport Contrôle-Affichage (*Control-Display ratio*)
RT-CALM	Implémentation temps-réel du CALM (*Real-Time Causal Anticausal Linear Model*)
Shimmer	Perturbation de l'amplitude de vibration des cordes vocales

Notation ou expression	**Signification**
ST	Demi-ton (*Semi-tone*)
T	Temps nécessaire pour atteindre une cible
T_c	Temps critique
T_t	Temps de transition
TS	Pas temporel (*Time Step*)
TTS	Transformation de texte en parole (*Text-to-Speech*)
VC	Transition Voyelle-Consonne
VE	Effort vocal
W	Largeur d'une cible (*Width*)
W_e	Largeur d'une cible effective
X	Position horizontale du stylet ou des doigts sur la tablette graphique
Y	Position verticale du stylet ou des doigts sur la tablette graphique

Liste des fichiers audio-visuels

Les fichiers audios et vidéos accompagnant ce manuscrit sont disponibles à l'adresse : https://perso.limsi.fr/operrotin/these.fr.php.

L'instrument *Cantor Digitalis* et sa pratique au sein de l'ensemble *Chorus Digitalis* sont présentés sur un site dédié : https://cantordigitalis.limsi.fr

Le détail des liens de téléchargement pour chaque fichier est donné ci-dessous.

Chapitre 2 : Le *Cantor Digitalis*

- Le logiciel est téléchargeable sur le site https://cantordigitalis.limsi.fr/download_fr.php
- Le manuel utilisateur ainsi que la documentation technique se trouvent à l'adresse https://perso.limsi.fr/operrotin/these.fr.php#Publications

Chapitre 4 : Méthodes de correction dynamique de la justesse mélodique

- Des exemples audios et vidéos des corrections présentées en figure 4.4 se trouvent à l'adresse https://perso.limsi.fr/operrotin/these.fr.php#Annexes

Chapitre 7 : Evaluation de l'instrument par sa pratique au sein du *Chorus Digitalis*

- Les vidéos de concerts réalisés par le *Chorus Digitalis* sont disponibles à l'adresse https://cantordigitalis.limsi.fr/chorusdigitalis_fr.php
- Des exemples de voix extrêmes sont donnés à l'adresse https://perso.limsi.fr/operrotin/these.fr.php#Annexes

Introduction

Chanter avec ses mains, contexte et problématique

Alors que des voix de synthèse de qualité nous parlent au quotidien, depuis nos modes de transports (p. ex. gares SNCF, GPS), jusque dans nos téléphones ou nos ordinateurs (p. ex. *Siri* d'Apple, voix du groupe Acapela [1]), le contrôle en temps réel de la synthèse vocale reste un enjeu majeur aujourd'hui. En effet, la voix est d'abord un outil d'interaction, de communication, et il semble donc naturel que la synthèse utilisée comme une substitution à la voix réelle soit capable de produire des phrases synthétisées en temps réel, en dialogue avec un interlocuteur. Cela constituerait une application médicale prometteuse par exemple. Les premiers contrôleurs de voix de synthèse apparus dès le début du 20^{e} avec le *Vodeur* permettaient l'articulation de phrases simples au prix de plusieurs mois d'entrainement intensif de l'opérateur [DRW39]. Un siècle plus tard, peu d'avancées ont été effectuées pour la production de parole intelligible en temps réel. La complexité et la dextérité des mouvements de nos articulateurs rendent difficile le transfert de l'articulation vers un paradigme de contrôle gestuel.

La voix est aussi un moyen d'expression. Dans le cas de la voix parlée, ce sont essentiellement les informations prosodiques (intonation, qualité vocale, rythme, accent, ...) qui sont porteuses d'expressivité. Dans le cas de la voix chantée, l'intelligibilité du signal importe moins que ses qualités expressives. L'expression se manifeste alors dans les variations subtiles d'intonation, de nuance, de rythme, propres à l'interprétation du musicien. De nombreux travaux s'intéressent à la création de règles automatiques pour la synthèse expressive. Cela a permis l'émergence de chanteurs virtuels tels que les avatars de *Vocaloid* de Yamaha [2], très célèbres dans le monde de la "J-pop". Il s'agit d'un logiciel de musique assistée par ordinateur (MAO) permettant de créer et d'intégrer en temps différé des parties vocales synthétiques dans des pièces musicales. Néanmoins, bien que l'aspect robotique des voix soit compensé par la création d'un univers visuel autour des chanteurs virtuels, l'expression musicale de ces chanteurs est encore loin d'égaler l'expression transmise par l'homme. Par ailleurs, il a été montré au fil des siècles à travers l'art chorégraphique ou l'art pictural que le geste est l'un des principaux convoyeurs d'expressivité du corps humain. La question de l'utilisation du geste pour le contrôle de synthèse de voix chantée semble alors tout à fait pertinente.

La conjonction de l'étude de gestes de contrôle d'une part et de moteurs de synthèse d'autre part pour la production de chant artificiel nous entraîne dans le domaine de la conception d'instruments de musique numériques, à cheval entre interaction homme-machine et traitement du signal. L'apparition d'interfaces de plus en plus performantes fait émerger de nouvelles perspectives dans ce domaine, à travers des conférences internationales dédiées, *New Interfaces for Musical Expression (NIME)*, *International Computer Music Confe-*

1. http://www.acapela-group.com/?lang=fr (vérifié le 22 octobre 2015)
2. http://vocaloid.fr (vérifié le 22 octobre 2015)

rence (ICMC) ainsi que de nombreux évènements art-sciences. Le découplage entre gestes de contrôle et moteur de synthèse permet de plus grandes possibilités d'exploration sonore que les instruments de musique acoustiques, où les combinaisons geste/son sont contraintes par des lois physiques. Néanmoins, l'étude du geste musical et l'expérience acquise dans la conception de nouveaux instruments numériques au cours des dernières décennies montrent qu'une prise de liberté aveugle dans l'association des gestes aux tâches musicales peut conduire à des instruments trop simples ou trop complexes, joués uniquement par leurs concepteurs et devenant rapidement obsolètes. Interfaces et moteur de synthèse doivent donc être choisis minutieusement et un travail de recherche approfondi est nécessaire pour valider l'adéquation de l'interface à la synthèse, afin de produire un nouvel instrument riche en possibilités musicales, associant identité sonore, finesse de contrôle, expressivité et un important potentiel d'exploration.

Ces considérations nous amènent donc à réfléchir sur le type de geste à associer au contrôle de la synthèse de voix chantée. La chironomie, jeu expressif des mains ou des doigts dans l'accompagnement de la parole, peut en particulier désigner l'accompagnement d'une ligne mélodique par le geste manuel. On définit alors une interface chironomique comme une interface captant les mouvements de la main de l'utilisateur pour un contrôle musical. Différentes équipes de recherche ont proposé l'usage de la tablette graphique pour le contrôle vocal au cours des deux dernières décennies et ces différents travaux ont soulevé le potentiel de l'utilisation du geste chironomique pour le contrôle de la voix chantée.

Le *Cantor Digitalis* est un exemple d'instrument de musique numérique de voix chantée contrôlé en temps réel par une tablette graphique. Développé depuis une dizaine d'années au LIMSI-CNRS, il est équipé d'un moteur de synthèse par formants permettant un calcul peu coûteux du signal vocal et une grande flexibilité du modèle utilisé. Le choix de la tablette graphique comme contrôleur de voix chantée s'est fait par expérience, notamment par les tests de plusieurs interfaces telles que l'ensemble clavier et pédale, ou le gant intelligent. Néanmoins, peu de travaux ont caractérisé l'adéquation d'une telle interface pour le contrôle de la voix chantée.

Cette thèse a pour but d'étudier la pertinence du contrôle du *Cantor Digitalis* par une tablette graphique, ou plus généralement, l'analogie entre gestes d'écriture ou de dessin et production de voix chantée. En allant plus loin que le test d'interface par sa pratique, cette thèse pose des bases scientifiques sur le choix de la tablette graphique comme outil musical, notamment en termes de quantification de performances et d'étude cognitive sur les modalités perceptives engagées dans le jeu de l'instrument. En quoi la tablette graphique est-elle un bon candidat pour le contrôle de la voix chantée ? Ses propriétés permettent-elle de jouer une mélodie avec la précision permise par la voix réelle ? Les fines fluctuations d'intonation, de nuance, de rythme propres à l'expression vocale sont-elles reproductibles à la tablette graphique ? Une telle interface propose-t-elle un apprentissage de l'instrument similaire aux instruments acoustiques ? L'instrument obtenu est-il praticable en conditions de concert ?

Les différentes études présentées dans ce manuscrit apportent des réponses à ces questions. Plus particulièrement nous mettons en évidence dans ce travail l'adéquation de cette interface pour le contrôle mélodique de la voix, en augmentant cette dernière d'outils logiciels pour la correction automatique de justesse. Par ailleurs, la question de l'importance de la vision dans le maniement de l'instrument est soulevée. Enfin un travail de recherche sur le jeu de l'instrument est conduit, justifiant de manière artistique l'utilisation de la tablette graphique pour le contrôle vocal.

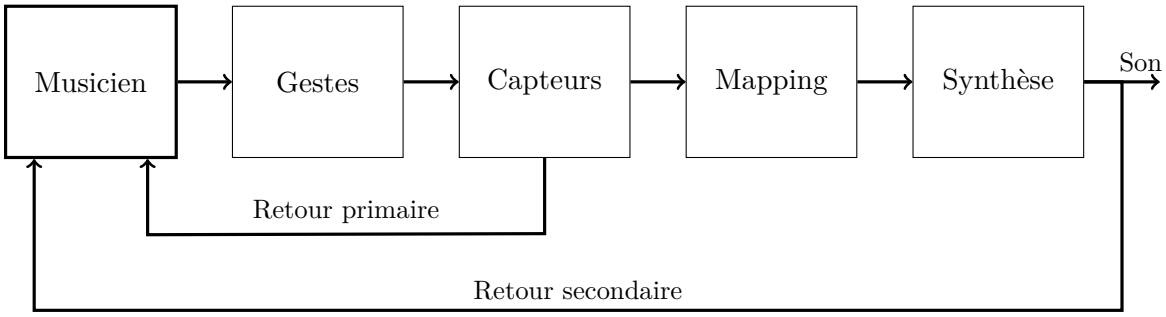

Figure 1 – *Schéma d'un instrument de musique numérique, d'après [WD04].*

Plan du manuscrit

Le schéma présenté en figure 1 résume l'organisation d'un instrument de musique numérique, selon la description de Wanderley *et al.* [WD04]. Le musicien interagit avec l'instrument par un ou plusieurs gestes. Ces gestes sont captés par l'interface de contrôle, faisant le lien entre monde physique et monde numérique. Les paramètres des gestes sont ensuite transformés en paramètres sonores par une étape de *mapping*. Ces paramètres sonores servent alors à la synthèse du son de l'instrument. Le musicien reçoit différent types de retours : un retour primaire lié à la manipulation de l'interface, et un retour secondaire, auditif, et relatif au son produit par l'instrument. L'instrument de musique numérique se décompose donc en trois blocs : une interface équipée de capteurs, un moteur de synthèse, et une phase faisant le lien entre les deux appelée *mapping*. Dans le cas d'un instrument acoustique, ces trois blocs sont confondus. Pour la conception d'un nouvel instrument de qualité, il est essentiel qu'aucun de ces blocs ne soit négligé. C'est pourquoi ce schéma est utilisé comme guide dans l'évaluation de l'interface chironomique pour le contrôle de la voix chantée. Chaque élément du schéma fait l'objet d'un chapitre dans cette thèse.

Le premier chapitre présente l'état de l'art nécessaire à la mise en contexte de ce travail. L'appareil vocal que l'on cherche à imiter est décrit en détail, ainsi que les différentes méthodes de synthèse utilisées aujourd'hui. Chaque bloc caractérisant un instrument de musique numérique est ensuite décrit en deuxième partie de ce chapitre, présentant les contraintes associées à la conception d'instruments numériques. Une revue brève de quelques instruments de synthèse de voix chantée termine se chapitre.

Le *Cantor Digitalis*, utilisé comme support dans cette thèse, est détaillé au chapitre 2. La première partie fait état des travaux effectués précédemment au LIMSI-CNRS dans la conception de l'instrument, de l'évolution du moteur de synthèse au choix de l'interface de contrôle. La deuxième partie présente les étapes de diffusion du *Cantor Digitalis* sous forme de logiciel libre réalisées durant cette thèse. L'interface graphique permettant une utilisation simple du logiciel et ses fonctionnalités proposées au grand public sont en particulier détaillées dans cette partie.

Le chapitre 3 est une évaluation du *capteur* utilisé. Il s'agit plus particulièrement de l'évaluation de la tablette graphique pour le contrôle de l'intonation en comparaison avec la voix réelle. Une expérience d'imitation d'intervalles mélodiques est conduite et analysée en terme de justesse et de précision de jeu.

Afin de faciliter le contrôle de l'intonation, différents *mappings* entre paramètres de la tablette et hauteur mélodique sont explorés. Le chapitre 4 propose le développement d'une méthode adaptative de correction dynamique de la hauteur mélodique. Celle-ci permet de corriger la justesse et la précision de la mélodie contrôlée tout en préservant les subtiles modulations mélodiques porteuses de l'expressivité du musicien. Ce mapping est évalué en termes d'expressivité préservée et d'apport de justesse de manière objective et perceptive.

Dans le chapitre 5, la question des *retours* primaire et secondaire est abordée dans le cas du *Cantor Digitalis*. La présence d'indices placés sur la surface de la tablette introduit délibérément une modalité visuelle de retour primaire ignorée chez les instruments de musique acoustiques. Une expérience est conduite pour quantifier l'impact de la modalité visuelle du retour primaire sur la modalité auditive du retour secondaire.

Toute interface impose un type de *gestes* pour le contrôle d'un instrument. Le chapitre 6 s'intéresse à la compatibilité des gestes proposés par la tablette graphique avec le contrôle de la voix chantée. Une suite de gestes pour le contrôle d'effets musicaux vocaux est proposée et confrontée aux lois décrivant les mouvements biologiques du corps humain.

Enfin, le chapitre 7 propose une évaluation de l'instrument dans sa globalité par sa pratique au sein de l'ensemble *Chorus Digitalis*. Il s'agit d'une approche artistique, à la recherche d'un répertoire musical propre à l'identité de l'instrument. La présentation scénique de l'ensemble est aussi discutée, notamment avec l'introduction d'un retour visuel permettant une meilleure compréhension de la manipulation de l'instrument en situation de concert.

Les résultats apportés par ces études ainsi que les perspectives de travaux futurs sont résumés en conclusion de cette thèse.

Un ensemble d'annexes complète ce manuscrit. D'abord, les calculs théoriques effectués dans l'étude des coûts du mouvement lors du chapitre 6 sont donnés en annexe A. Ensuite, la liste des communications scientifiques ou artistiques associées à cette thèse sont données en annexe B.

Chapitre 1

Contrôle temps réel de la synthèse vocale : cadre d'étude et état de l'art

Sommaire

1.1 Introduction

L'appareil vocal est un système très complexe utilisé par de nombreuses fonctions de notre organisme. Les poumons permettent de respirer, la langue, les dents sont nécessaires à l'alimentation, et une coordination de tous ces éléments permet la production de voix, essentielle à la communication. Par ailleurs, les possibilités de choisir les hauteurs du son avec précision et de contrôler leur rythme de production nous permettent de chanter. L'appareil vocal est un des premiers instruments de musique utilisé par l'homme, est un des plus étudiés aujourd'hui. C'est la volonté de compréhension de l'appareil vocal qui a fait émerger les premiers synthétiseurs, comme l'explique Sundberg à ce sujet [Sun06] :

"If you want to describe what characterizes a singer's voice, you simply synthesize it. As soon as your synthesis contains all the timbral characteristics of the original, you know that from a perceptual point of view your synthesis is exhaustive."

Alors que la connaissance de l'appareil vocal s'est développée, les synthétiseurs vocaux se sont perfectionnés pour être aujourd'hui utilisés dans un but de reproduction vocale, pour la parole (voix de systèmes GPS, métro, gares, assistance, etc.) ou pour le chant (instruments de musique numériques).

Parallèlement au développement des synthétiseurs, qu'ils soient vocaux ou non, l'apparition de nouvelles interfaces de contrôle a favorisé l'émergence des instruments de musique numériques, créant une nouvelle branche dans le domaine de l'interaction homme-machine, spécialisée dans le contrôle musical. De nombreuses recherches ont ainsi été effectuées sur les notions de geste, de capteurs, de relation entre gestes et tâches musicales, afin de proposer des contrôleurs parfois insolites mais surtout adaptés au contrôle d'un instrument de musique.

Cette partie a pour but de présenter ces différentes recherches, en commençant par une description de l'appareil vocal, puis en décrivant les méthodes de synthèses vocales les plus utilisées, et enfin en présentant les enjeux du contrôle d'instruments de musique numériques. Quelques exemples d'instruments de synthèse vocale sont ensuite exposés, résultants de combinaisons entre méthodes de synthèses et interfaces diverses.

1.2 La production de parole

1.2.1 L'appareil vocal

La parole et le chant sont les fruits de la perturbation dans le larynx du flux d'air sortant de nos poumons, et sa résonance dans l'ensemble du conduit vocal, c'est-à-dire cavités buccale et nasale. On distingue alors trois étapes dans la production de la parole.

La création d'un flux d'air

Afin d'exciter les différentes membranes responsables de la création d'un son, un flux d'air doit circuler dans notre appareil de production vocal. Le système de soufflerie utilisé est simplement notre appareil respiratoire, autrement dit nos poumons. De l'air est expiré (ou inspiré dans une moindre mesure dans le cas de voix ingressives) par la trachée et vers le larynx situé à son sommet.

La perturbation du flux

Le larynx, schématisé en figure 1.1, est une cavité située au sommet de la trachée et source de la production sonore. Il repose sur le cartilage cricoïde, en forme d'anneau qui fait la jonction entre trachée et larynx. La face antérieure du larynx est fermée par le cartilage thyroïde en forme de livre ouvert, dont la partie saillante forme la pomme d'Adam. Sur la face postérieure se situent deux cartilages aryténoïdes de formes pyramidales, attachés au cricoïde et mobiles. En position supérieure se situe le cartilage épiglottique, fermant le passage vers la trachée lors de la déglutition. Deux ligaments recouverts d'une muqueuse appelés cordes vocales ou plis vocaux sont tendus de manière symétrique et horizontalement entre le cartilage thyroïde et chacun des cartilages aryténoïdes. L'articulation des aryténoïdes permet de rapprocher ou d'éloigner les cordes vocales. On appelle *glotte* l'espace entre celles-ci. Lorsque les cordes vocales se touchent, elles obstruent le passage de l'air de la trachée. A l'inverse, lorsque celles-ci sont en position ouverte, le passage de l'air est possible.

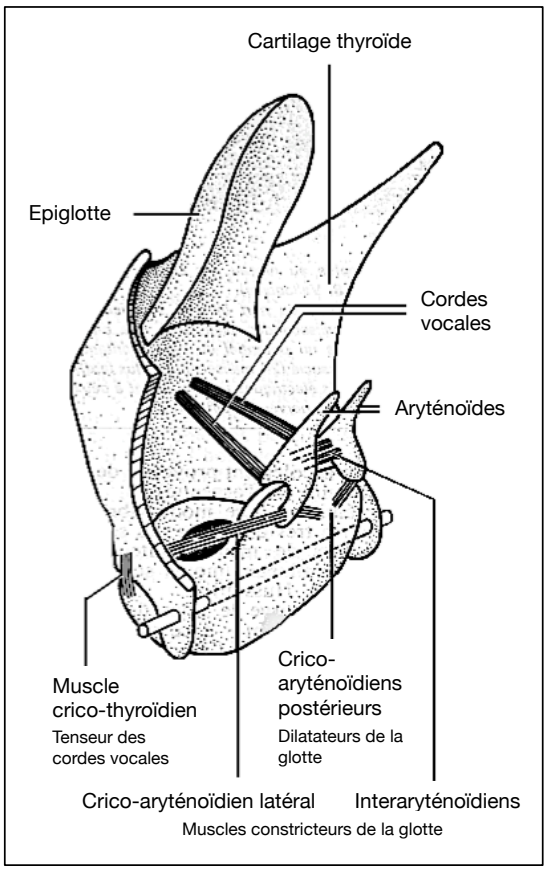

FIGURE 1.1 – *Schéma simplifié du larynx selon une vue postérieure gauche, d'après H. Lullies.*

La position ouverte est la plus courante puisque nécessaire à la respiration. De plus, lorsque le flux d'air est suffisamment rapide, des turbulences se créent dans le larynx et un son bruité est alors produit, caractéristique du chuchotement par exemple. On appelle ce mode de parole *non-voisé*.

Le basculement du cartilage thyroïde permet de tendre les cordes vocales longitudinalement. La mise en concurrence de la tension des cordes vocales et de la pression de l'air expiré induit une vibration des muqueuses, alternant ouverture et fermeture de la glotte : la

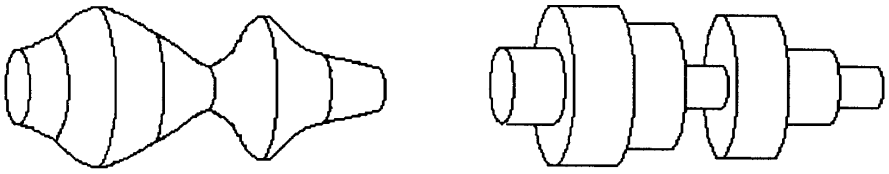

Figure 1.2 – *Schéma du tube acoustique représentant le conduit vocal à gauche, et sa version discrétisée à droite, d'après [Coo90].*

tension des cordes vocales ferme la trachée et obstrue le passage de l'air. Toutefois, lorsque la pression de l'air sous-glottique est suffisamment importante, celle-ci écarte doucement les cordes vocales et provoque l'ouverture de la glotte. L'air passe, entraînant une dépression. Par élasticité les cordes vocales se referment alors brusquement et obstruent à nouveau la trachée. La répétition de se phénomène produit une vibration des cordes vocales, produisant un son dit voisé, c'est la *phonation*. Un son voisé est caractérisé par la fréquence de vibration des cordes vocales exprimée en cycles par secondes ou hertz (Hz), directement liée à la perception de la hauteur du son.

Résonance

L'air perturbé traverse ensuite le conduit vocal. Ce dernier peut être considéré comme un tube dont le diamètre varie en fonction des obstacles rencontrés [KL62], notamment le plat et la pointe de la langue, mais aussi en fonction de ses parois mobiles, comme les lèvres et la mâchoire inférieure, ou le voile du palais qui ouvre le passage vers le conduit nasal. On appelle ses obstacles et parois mobiles les *articulateurs*. Une modélisation approchant le tube en une suite de cylindres permet un calcul simple de la propagation de l'onde acoustique dans le conduit vocal. La figure 1.2 donne un exemple de modélisation.

L'interaction entre le flux d'air perturbé et chaque articulateur amplifie une certaine bande de fréquence dans le spectre du signal glottique. Cette "bosse" fréquentielle ou résonance est appelée formant. Chaque articulateur est responsable d'un formant, caractérisé par sa fréquence centrale, son amplitude et sa largeur. Les articulateurs étant mobiles, les caractéristiques de chaque formant varient en fonction de la position des articulateurs, produisant différents timbres sonores, et perçus comme les différentes voyelles. Seuls les trois premiers formants (classés par ordre de fréquence centrale croissante) suffisent à l'intelligibilité des voyelles du français et sont liés respectivement à l'ouverture de la mâchoire, la position avant-arrière de la langue et l'ouverture des lèvres.

On peut alors représenter les positions des voyelles dans le plan des deux premiers formants, soit (ouverture de la mâchoire $\times$ position de la langue), comme présenté en figure 1.3. L'organisation des voyelles sur ce plan donne à cette représentation la dénomination de *triangle vocalique*. Celle-ci nous montre que l'ouverture de la mâchoire fait la différence entre un /a/ (mâchoire ouverte) ou un /u/ comme dans "poule" (mâchoire fermée), la position de la langue différencie un /o/ (arrière) d'un /e/ comme dans "été" (avant), et l'écartement des lèvre différencie un /u/ (serrées) d'un /i/ (écartées). La combinaison de ces trois articulateurs permet d'obtenir toutes les voyelles buccales du français. Les voyelles nasales sont produites par la combinaison de positions articulatoires buccales et l'ouverture du voile du palais.

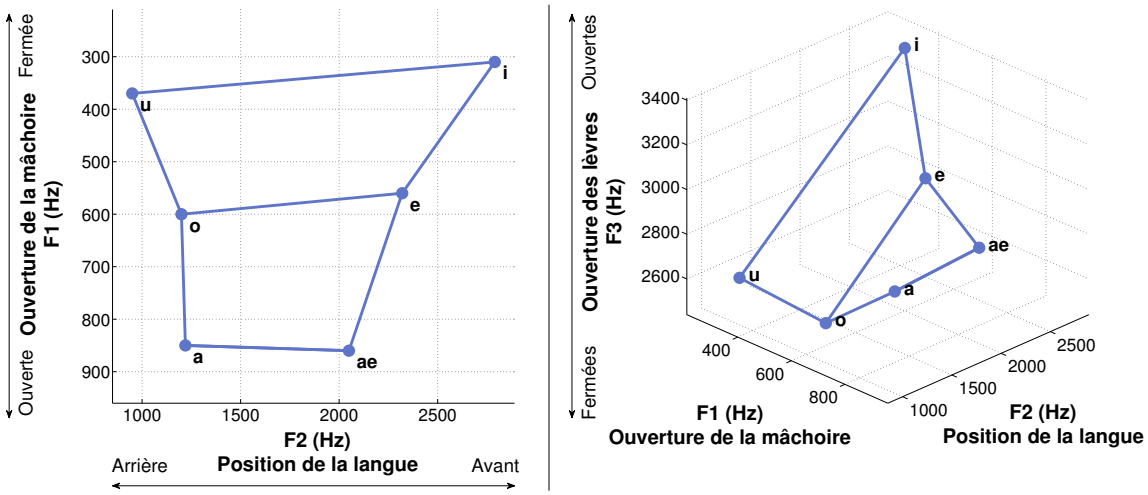

FIGURE 1.3 – *Représentation des voyelles extrêmes du français dans le plan des deux premiers formants (gauche) et dans l'espace des trois premiers formants (droite).*

1.2.2 Le modèle source-filtre

La décomposition par étapes successives de la production de parole amène à introduire le modèle "source-filtre". Ce dernier stipule que la perturbation du flux d'air dans le larynx (partie "source") et sa résonance dans le conduit vocal (partie "filtre") sont indépendantes et peuvent être chacune modélisée par une combinaison de filtres linéaires [Fan70]. Le modèle source-filtre se décompose comme suit :

1. La partie "source" modélise la production sonore. Elle est vue comme un filtre linéaire G (pour glottique) transformant soit un train d'impulsion de période T_0 caractérisant la fréquence fondamentale du signal dans le cas "voisé", soit un bruit dans le cas "non-voisé", en une Onde de Débit Glottique (ODG). L'ODG modélise donc le flux d'air à la sortie du larynx.
2. La partie "filtre" et composée de deux sous-parties. D'abord elle englobe toutes les résonances induites par le conduit vocal représentées par un filtre linéaire V. Celui-ci prend en entrée l'ODG. Lors de la sortie du conduit vocal, l'onde acoustique est rayonnée par les lèvres, se traduisant en une dérivation du signal de sortie de V. Un filtre dérivateur L (pour lèvres) modélise donc le rayonnement.

Un schéma du modèle source-filtre est présenté en figure 1.4. Du point de vue du traitement du signal, on peut alors exprimer le signal de sortie s comme la convolution successive des réponses impulsionnelles du modèle :

$$s(t) = e(t) * g(t) * v(t) * l(t) \tag{1.1}$$

Où $e(t)$ est l'entrée du système (train d'impulsion ou bruit). Dans le domaine fréquentiel, il s'agit de la multiplication du spectre d'entrée par les réponses en fréquences des filtres du modèle :

$$S(f) = E(f) \times G(f) \times V(f) \times L(f) \tag{1.2}$$

Cette décomposition permet théoriquement d'isoler le signal de source et les effets du conduit vocal. Une méthode courante de séparation source-filtre développée dans les années 1970 est la méthode LPC (*Linear Predictive Coding*) [AH71]. Il s'agit d'exprimer chaque

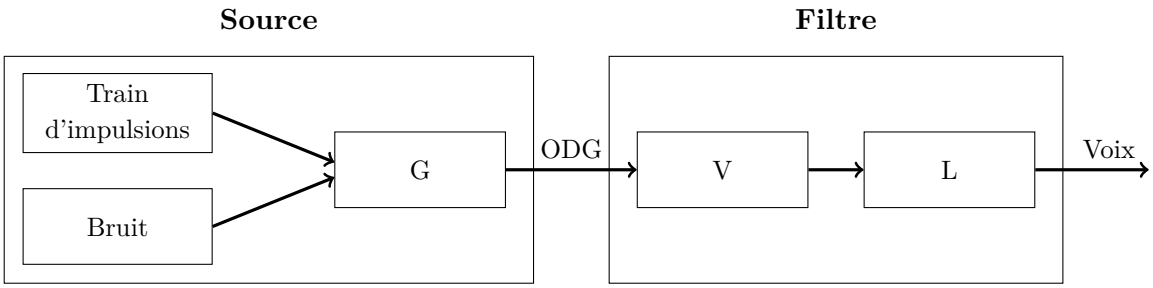

Figure 1.4 – *Modèle source-filtre.*

échantillon du signal de sortie s comme une combinaison linéaire des n échantillons précédents, n étant l'ordre de la LPC. L'erreur de prédiction obtenue s'exprime alors comme :

$$e(n) = s(n) - \sum_{k=1}^{n} a_k s(n-k) \tag{1.3}$$

Ce qui se traduit dans le domaine fréquentiel par :

$$\frac{S(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^{n} a_k z^{-k}} \tag{1.4}$$

Le membre de droite correspond à un filtre tout-pôle, caractéristique des résonances dans le conduit vocal. En supposant le modèle parfait, on en déduit que l'erreur de prédiction $e(n)$ représente le signal de source, et les coefficients de LPC représentent les pôles du filtre du conduit vocal. En connaissant le signal de parole s et en déconvoluant ce dernier par la réponse impulsionnelle du filtre LPC, on peut alors obtenir le signal de source, comme montré en figure 1.5. On observe sur la rangée du haut les signaux de paroles enregistrés par une voix d'homme pour les voyelles /a/, /i/ et /o/. Une analyse LPC d'ordre 64 permet d'extraire les enveloppes spectrales affichées en bleues, et induites par le conduit vocal. Les pics marqués en vert sur l'enveloppe spectrale correspondent aux premiers formants. On voit bien que le 1er formant du /a/ est plus avancé que pour le /i/ et /o/. A l'inverse, le 2^{e} formant du /i/ est beaucoup plus élevé que pour les autres voyelles. Les signaux de source obtenus par déconvolution sont affichés sur la rangée du bas. On observe alors une grande ressemblance entre ces signaux, prononcés par le même locuteur avec une qualité de voix similaire. Les différences spectrales caractérisant les voyelles sont bien propres au conduit vocal et n'apparaissent pas au niveau de la source.

Bien que couramment utilisé, le modèle source-filtre est construit sur l'indépendance de la source et du filtre qui n'est pas vérifiée en pratique.

1.2.3 Les caractéristiques acoustiques de la voix

La génération du son par la source est un phénomène complexe, dû aux multiples paramètres régissant la régularisation du flux d'air, ainsi que l'ouverture et la tension des cordes vocales. Les combinaisons infinies de tous ces paramètres permettent une production de parole très riche en sonorités dont les principaux effets perceptifs ont été recensés par d'Alessandro [d'A06]. On dit que ces différents effets décrivent la *qualité de voix*. La perception de chacune de ses qualités de voix est possible grâce au comportement spécifique d'un ensemble de paramètres spectraux pour chacune d'elles [IRDC14].

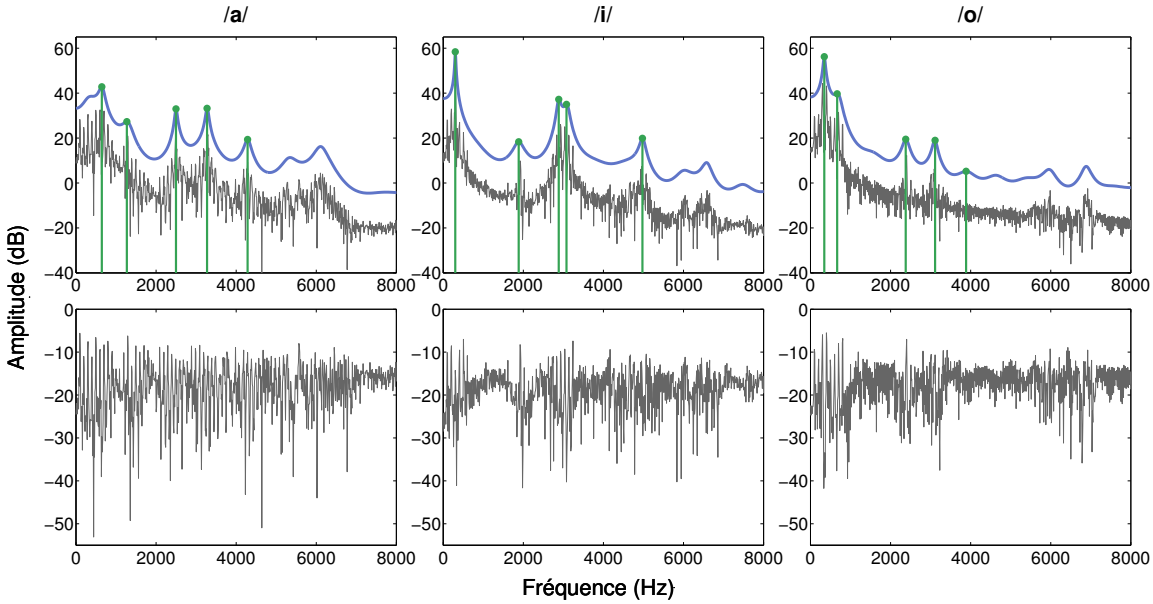

FIGURE 1.5 – *Décomposition source-filtre par analyse LPC d'ordre 64. Le signal de parole est représenté en haut. L'enveloppe spectrale induite par le conduit vocal et extraite par LPC est indiquée en bleue. Les 5 premiers maxima correspondant aux premiers formants sont relevés en vert. Le signal de source obtenu par déconvolution du signal de parole est affiché en bas. De gauche à droite, les voyelles /a/, /i/ et /o/.*

Hauteur

En mode phonatoire, lorsque les cordes vocales vibrent, la fréquence de vibration exprimée en cycles par seconde est étroitement corrélée à la perception de hauteur. Par exemple, une vibration de 440 cycles par seconde (440 Hz) produit un La4 (le la du diapason). La hauteur vocale peut être contrôlée de manière extrêmement précise, nous permettant de chanter.

Registre ou mécanisme laryngé

Pour parcourir une gamme de hauteur importante, les cordes vocales adoptent quatre modes de vibrations. On parle de registre vocal ou mécanisme laryngé. Le mécanisme M0, dit *voix craquée*, correspond à une vibration très lente (10-20 Hz) des cordes vocales, très détendues et produisant une voix très grave. En mécanisme M1, dit *voix de poitrine*, les cordes vocales vibrent sur toutes leurs longueurs. C'est le mécanisme couramment utilisé pour la parole par les hommes. En mécanisme M2, dit *voix de tête*, les cordes vocales sont plus tendues et ne vibrent que sur leurs parties antérieures. Les hauteurs produites sont plus aiguës. Les mécanismes M1 et M2 sont alternativement utilisés en parole par les voix de femme. Enfin, le mécanisme M3 dit *voix de sifflet* correspond à une voix très aiguë non utilisée en parole. Le schéma 1.6 résume la position des différents mécanismes sur l'axe des sections croissantes de la glotte, et le mode de production de parole associée. Le changement de registre se fait souvent de manière abrupte, et accompagné d'un saut de hauteur. La hauteur maximum possible en mécanisme M1 étant limitée pour les chanteurs hommes par exemple, il est courant de passer en voix de tête pour étendre la tessiture. Un entraînement est nécessaire pour changer de registre de manière fluide et inaudible.

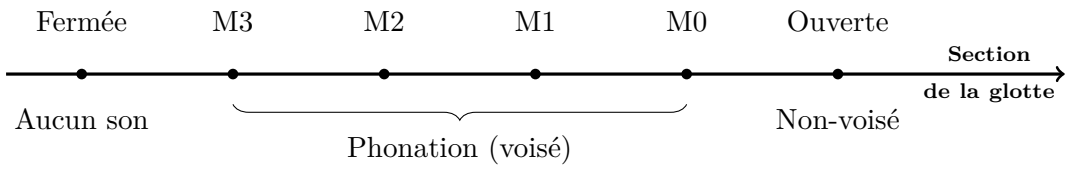

FIGURE 1.6 – *Evolution des mécanismes et des modes de production de parole correspondants en fonction de la section de la glotte.*

Apériodicités

Un bruit résulte d'une perturbation non périodique du flux d'air. Deux cas de figure coexistent dans la production de parole. D'abord, les turbulences causées par la forme du larynx sur le flux d'air génèrent un bruit. Celui-ci peut être exclusif dans le cas de chuchotements où la glotte est grande ouverte. Il en résulte de la parole non-voisée. Le bruit peut être aussi additif. Pendant la phonation chez certaines personnes, il arrive que les cordes vocales ne se referment jamais complètement. De l'air passe en permanence et produit donc un bruit additif au son voisé. On parle alors de *voix soufflée*.

La deuxième forme d'apériodicité rencontrée concerne la vibration des cordes vocales. En général, dans une voix non travaillée ces dernières subissent des perturbations aléatoires produisant une voix instable. Ces perturbations concernent soit la fréquence de vibration des cordes vocales (*jitter*) soit l'amplitude des vibrations (*shimmer*). Plus les perturbations sont élevées, plus la voix perçue est rauque.

Tension

Le rapprochement des cartilages aryténoïdes permet de tendre ou de relâcher les cordes vocales. Lorsque les aryténoïdes sont rapprochés, les cordes vocales sont tendues et une voix pressée est produite. A l'inverse, l'écartement des aryténoïdes détend les cordes vocales et une voix relâchée est obtenue.

Effort vocal

Lorsque l'on cherche à varier l'amplitude de notre voix, plusieurs facteurs entrent en jeu : la pression sous-glottique ainsi que la tension des cordes vocales sont augmentées pour produire une voix plus puissante. La combinaison de ces paramètres entraîne la perception d'un *effort vocal*. C'est par cet effort que nous sommes capable de différencier une voix murmurée d'une voix criée, quelque soit le volume de la voix perçue.

1.3 La synthèse vocale

La voix est un appareil de production sonore extrêmement riche tant au niveau des sonorités permises par les différentes qualités vocales, qu'à la malléabilité de la caisse de résonance introduite par la mobilité des articulateurs. Cette richesse a poussé des équipes de recherche à s'intéresser à son fonctionnement, et à tenter de la reproduire par différentes méthodes de synthèse. Celles-ci sont nombreuses, et peuvent être partagées en différentes catégories.

La catégorie de méthodes la plus ancienne est la création d'onde sonores, où la voix produite est entièrement synthétisée, et calculée à partir de la connaissance du fonctionnement de l'appareil vocal [Coo96] [Coo98]. Cette catégorie se divise à nouveau en deux groupes : les méthodes par modèle physique, et les méthodes par modèle spectral. Les premières reposent sur l'étude et la modélisation de la production sonore. Les mouvements et les interactions des organes de l'appareil vocal sont simulés, proposant une approche intuitive de la production de parole. Cependant, le nombre de paramètres nécessaires pour le contrôle de chaque organe (source, articulateurs) peut être très important et nuit à l'aspect intuitif introduit par ce modèle.

Les méthodes spectrales quant à elles reposent sur la perception de la parole. En effet, la cochlée, organe responsable de transformer les sons perçus par le tympan en signaux électriques, est une bande d'environ 30 mm enroulée sur elle-même dont le diamètre décroît de 10 à 4 mm. Cette décroissance de diamètre rend les différentes parties de la cochlée sensibles à des fréquences particulières. Ainsi, les vibrations perçues par le tympan sont décomposées spatialement sur la cochlée, fournissant une analyse spectrale du son. Les modèles spectraux consistent donc à créer de la parole dans le domaine fréquentiel en synthétisant des spectres, c'est-à-dire en choisissant l'amplitude de chaque fréquence constituant le son. Bien que moins intuitifs que les modèles physiques, ceux-ci sont fortement utilisés en raison de leur plus grande simplicité en terme de contrôle et de coût de calcul.

En opposition aux systèmes de création d'onde sonore, il existe des systèmes de synthèse à partir de parole réelle dite par corpus [Rod02]. Cette deuxième catégorie utilise des bases de données de parole enregistrée dont des informations sont extraites pour la synthèse. Ces informations peuvent être des tronçons de paroles découpés dans la base et mis bout à bout dans un ordre différent pour la synthèse. Il s'agit alors de la synthèse par concaténation. A mi-chemin entre synthèse par concaténation et synthèse de forme d'onde, on voit apparaître des systèmes paramétriques par modèles statistiques où des paramètres descriptifs de la parole tels que la fréquence fondamentale ou des descripteurs spectraux sont extraits de bases de données de parole réelle, et la synthèse est réalisée à partir de ces descripteurs.

Rodet met en évidence la nécessité d'établir un système de règles permettant le contrôle des synthétiseurs [Rod02]. Il distingue deux niveaux de règles : un bas niveau permettant la synthèse la plus naturelle possible de la voix, et un haut niveau décrivant l'aspect expressif. Un des systèmes de règles les plus élaborés est le système RULSYS [Ber95] créé à KTH en collaboration avec des musiciens pour trouver des règles d'expressivité.

Finalement, les applications de la synthèse vocale sont diverses. Elles peuvent avoir un but de production de parole. De nombreux exemple de système TTS (*Text-To-Speech*) de la fin du 20^{e} siècle sont donnés par d'Alessandro [d'A01]. La synthèse a aussi une portée académique, comme outil à l'étude de la voix. Sundberg donne différents exemples de travaux réalisés à KTH tels que l'étude du formant du chanteur, l'identification des caractéristiques vocales différenciant un bon d'un mauvais chanteur, ou l'étude de l'intonation [Sun06]. Enfin la synthèse vocale a suscité l'intérêt de nombreux musiciens par la synthèse du chant. Le tableau 1.1 classifie les différentes méthodes de synthèse utilisées aujourd'hui. Chacune est ensuite présentée plus en détail dans la suite de cette partie.

Synthèse par génération d'onde sonore	**Synthèse à partir de parole réelle**
Modèles physiques	*Synthèse par concaténation*
Modèles spectraux	*Synthèse paramétrique par modèle statistique*
- Synthèse FM	*Synthèse par transformation de voix parlée*
- Synthèse par modélisation sinusoïdale	*Synthèse par mélange de voix*
- Synthèse par formants	

Table 1.1 – *Résumé des différentes méthodes de synthèse utilisées aujourd'hui.*

1.3.1 Synthèse par modèle physique

La synthèse par modèle physique étudie les mouvements et interactions des organes de l'appareil vocal et calcule l'expression de l'onde acoustique traversant ce dernier. On distingue quatre étapes dans un modèle [KB08] : la modélisation des mouvements du conduit vocal (contrôle) ; la modélisation de la production sonore (source) ; la modélisation de la géométrie du conduit vocal (conduit vocal) ; la génération des signaux acoustiques depuis les informations d'articulation.

- Les paramètres de contrôle du système sont les descriptions phonologiques des sons à jouer ainsi que les indications prosodiques. Dans le cas du chant, des paramètres musicaux sont introduits (hauteur, durée, nuance des notes).
- La source est modélisée soit par une fonction paramétrique, soit calculée par des équations mécaniques. Dans le cas voisé, il s'agit soit de l'introduction d'une fonction d'onde glottique paramétrique, soit d'un système d'oscillation masse-ressort. Dans le cas non-voisé, le bruit peut être paramétrique ou calculé par les équations aéro-acoustiques.
- Le conduit vocal est caractérisé par exemple par les positions de la mâchoire, des lèvres, de la pointe de la langue, du corps de la langue, du vélum et du larynx. Ces positions peuvent être relevées de manière statistique sur un corpus de mouvements mesurés par IRM ou rayon X, soit calculées en représentant les articulateurs par la méthode des éléments finis ou par modèles géométriques.
- L'onde acoustique résultante est le débit d'air traversant le conduit vocal et rayonnant aux lèvres. Celle-ci peut-être obtenue selon les coefficients de réflexion/transmission du conduit vocal (impédance des jonctions des cylindres dans le modèle à tubes, figure 1.2), par le calcul de la fonction de transfert acoustique du conduit vocal, ou par la fonction de propagation d'onde acoustique.

Le SPASM (*Singing Physical Articulatory Synthesis Model*) créé par Perry Cook [Coo90], [Coo93] est un exemple de synthèse par modèle physique ou articulatoire. Il modélise le conduit vocal par un tube formé de cylindres de sections différentes (figure 1.2), calculé à partir de paramètres de formes du conduit vocal et composé d'un conduit buccal et nasal. Il se comporte comme un guide d'onde. Plusieurs qualités de sources sont données par des tables et des turbulences sont ajoutées dans le conduit vocal pour réaliser les consonnes. La forme du conduit vocal est affichée visuellement et peut être modifiée par l'utilisateur par le biais d'une interface dédiée.

1.3.2 Synthèse par modèle spectral

Les méthodes par modèle spectral consistent à reproduire le spectre de signaux de paroles dans le domaine fréquentiel, perceptif.

Synthèse FM

Bien que non utilisée pour la voix, la méthode FM a eu un succès fulgurant dans la synthèse d'instruments de musique. La modulation en fréquence (FM) consiste à faire varier la fréquence d'une sinusoïde appelée fréquence porteuse par un signal dit modulant. Ce principe est d'abord utilisé pour la transmission d'ondes radios où le signal audio (modulant) est décalé dans le domaine des très hautes fréquences par la porteuse pour sa transmission. Chowning a utilisé ce concept pour la création de spectres audios [Cho73]. Dans ce cas, fréquence porteuse et signal modulant sont tous deux dans le domaine audible. L'ajout d'une modulation étend la bande de fréquence de la porteuse, et le choix de l'amplitude et de la fréquence du signal modulant ainsi que de la fréquence porteuse permettent d'obtenir une très grande variété de spectres. La combinaison de la modulation et d'une enveloppe d'amplitude adaptée permet la synthèse de nombreux instruments. Brevetée conjointement par l'université de Standford et Yamaha, cette méthode a eu un succès fulgurant par le biais du synthétiseur DX-7 [1].

Synthèse par modélisation sinusoïdale

En partant du principe que tout signal périodique peut se décomposer en une somme de sinusoïdes appelées partiels, une méthode d'extraction de ces partiels a été mise en place [SSI90]. Construite en deux étapes, elle permet de modéliser séparément la partie harmonique du signal de la partie inharmonique. La première étape consiste à extraire les trajectoires des harmoniques en fonction du temps en découpant le signal en courtes trames et par extraction des pics sur le spectre de chaque trame. L'observation du spectre par trame est réalisée par une STFT (*Short Time Fourier Transform)*. Par soustraction du signal original et du signal harmonique modélisé, on obtient le signal bruité. L'enveloppe spectrale de ce dernier est alors lissée. Il est ensuite possible de resynthétiser de la parole par synthèse additive suivant les trajectoires des sinusoïdes et en filtrant un bruit blanc par l'enveloppe spectrale de la partie inharmonique. Ce type de modèle permet une modification facile des modèles extraits pour créer des nouveaux sons.

Dans un but de contrôle de la synthèse additive pour la composition, la méthode de Synthèse Additive Structurée (SAS) permet d'exprimer la synthèse additive selon quatre paramètres uniquement, proches de la perception musicale : l'amplitude globale du signal ; la fréquence fondamentale du signal ; la couleur, ou l'enveloppe spectrale ; l'inharmonicité exprimée comme une fonction de déformation de la position des partiels en référence à un son harmonique (équidistants) [DCM99a], [DCM99b]. La formulation de la synthèse par ces quatre paramètres uniquement permet un contrôle aisé du son produit par un musicien et facilite la communication entre composition musicale et création sonore.

Synthèse par formants

La synthèse par formants modélise les résonances du conduit vocal de manière individuelle et les applique sur la source. Ces résonances sont modélisées fréquentiellement par des filtres résonants du second ordre décrits par leurs fréquences centrales, bandes passantes et amplitudes et placés en cascade ou en parallèle [Kla80], [Hol83]. Le modèle en cascade est plus fidèle à la production physique de parole où chaque résonance est associée à une position spatiale du conduit vocal. En revanche, le modèle en parallèle est plus adapté à la perception humaine, permettant d'agir indépendamment sur chaque bande de fréquence du signal obtenu. Holmes [Hol83] conclu que bien que le modèle parallèle nécessite plus de contrôle (un

1. https://www.youtube.com/watch?v=F3rrjQtQe5A (vérifié le 22 octobre 2015)

gain par formant), ce dernier permet de modéliser l'effort vocal et simplifie sa manipulation. De plus un modèle parallèle est indispensable pour le contrôle des consonnes [Kla80]. Les valeurs des formant sont obtenues en général par une analyse LPC de signaux de paroles.

Un des synthétiseurs par formants les plus aboutis est le synthétiseur MUSSE (*MUsic and Singing Synthesis Equipment*) [Lar77], [ZGS84] réalisé à KTH et implémenté sous forme numérique MUSSE DIG en 1989. Il est joué par un clavier et des potentiomètres permettent de contrôler la fréquence du vibrato, la quantité de bruit glottique, le jitter et le shimmer. Le système de règles de contrôle RULSYS pour la production de chant expressive a par la suite été développé pour commander le MUSSE DIG [Ber95].

Une deuxième approche de la synthèse à formant est la modélisation des résonances dans le domaine temporel par des Fonctions d'Ondes Formantiques (FOF) [RPB84]. Bien qu'exprimées dans le domaine temporel, celles-ci sont contrôlées par la fréquence centrale, la bande passante et l'amplitude du formant correspondant. Les FOF sont implémentées en parallèle dans le programme CHANT [2][RD85].

1.3.3 Synthèse à partir de parole réelle

La deuxième catégorie de méthodes de synthèse ne considère plus les mécanismes de production de la parole. Elle utilise uniquement une base de donnée de parole réelle. On parle de synthèse par corpus.

Synthèse par concaténation

La synthèse par concaténation d'unités consiste à mettre les uns derrière les autres des morceaux de parole enregistrés, afin de reconstituer un extrait de parole. Une base de donnée importante doit alors être enregistrée pour disposer de tous les sons nécessaires à la synthèse. Les unités extraites des enregistrements sont généralement des diphones, c'est-à-dire un enchaînement de deux voyelles V ou consonnes C (VV, VC, CC, CV). Cela permet d'utiliser la transition naturelle entre les sons réalisée par le locuteur. Les diphones sont concaténés au niveau de leurs parties stables (voyelle ou consonne tenue). Afin de limiter la taille des enregistrements, les unités sont parfois modifiées, notamment pour changer leurs hauteurs. La synthèse se déroule en deux étapes. Les unités sont d'abord sélectionnées dans la base selon deux coûts : le coût de concaténation (proximité entre deux unités à accoler) et coût de cible (ressemblance entre l'unité et le son cible). Ces coûts peuvent être appris par apprentissage statistique. Ensuite, les unités sont concaténées en ajustant la phase entre les deux [BL03], en lissant la transition entre leurs spectres, ou en s'aidant de modèles paramétriques [BCL+01].

On peut citer comme applications le MBROLA (*Multi-Band Resynthesis OverLap Add*), conçu par le TCTS Lab de la Faculté Polytechnique de Mons [3] [DPP+96] qui prend en entrée la liste des phonèmes et les informations prosodiques à synthétiser. Le logiciel commercial *Vocaloid* [4] [KO07] de Yamaha est probablement le système le plus connu du grand public, permettant d'obtenir une prestation chantée par un des avatars célèbres du logiciel à partir seulement de la partition et des paroles d'une chanson. D'autres systèmes prennent en entrée les paroles, mais les informations prosodiques sont fournis par l'extraction de paramètres vocaux [5] [JBB06]. Enfin, un système de synthèse par concaténation du chant en français est actuellement développé sous le projet ANR ChaNTeR [Ard13].

2. http://recherche.ircam.fr/anasyn/peeters/PSOLA/AUDIO/reine.aiff (vérifié le 22 octobre 2015)
3. http://tcts.fpms.ac.be/synthesis/mbrola.html (vérifié le 22 octobre 2015)
4. http://vocaloid.fr (vérifié le 22 octobre 2015)
5. http://www.dtic.upf.edu/~jjaner/dafx06/ (vérifié le 22 octobre 2015)

Synthèse paramétrique par modèle statistique

Toujours d'après le modèle source-filtre, il est possible d'extraire d'un signal de parole un ensemble de paramètres permettant de décrire ce dernier. Ces paramètres sont la fréquence fondamentale ainsi que le voisement pour la source, et les coefficients mel-cepstraux pour la partie filtre, ainsi que les variations dynamiques de ces paramètres. Ce principe s'appelle Vocodeur (*Voice Coder*).

La reconstitution d'un signal de parole à partir de ces paramètres de vocodeur est couramment utilisée pour les systèmes de conversion de texte en parole (*Text To Speech - TTS*). L'enregistrement d'un grand nombre de signaux permet de couvrir des contextes vocaux relativement larges. Pour chaque signal enregistré, deux extractions sont faites : une extraction des paramètres liés au texte au niveau de la phrase, du mot, de la syllabe et du phonème, et une extraction des paramètres du signal vocal (vocodeur). La manipulation de ces paramètres se fait par approche statistique. Le modèle le plus utilisé jusqu'à ce jour est le modèle de Markov caché (*Hidden Markov Model - HMM*) [YTK+99], [OMNT12], [TNT+13]. Préliminaire à la synthèse, une étape d'apprentissage établit les liens entre les paramètres du texte et les paramètres de parole. Ensuite, une étape de synthèse sélectionne les paramètres de parole adaptés à un nouveau texte fourni en entrée selon les règles apprises précédemment et reconstruit le signal par Vocodeur. Cette méthode offre les avantages de pouvoir interpoler les paramètres de différents locuteurs, d'être robuste au changement de langue et d'être très peu coûteux en termes de mémoire. Il est aussi possible d'inclure des paramètres expressifs pour enrichir la qualité de synthèse, ou musicaux pour la synthèse de voix chantée. Par exemple, des travaux ont mis en évidence les liens entre paramètres acoustiques du signal et émotions dans la voix chantée [MDCGPS13].

Une comparaison de la synthèse par concaténation et synthèse statistique montre qu'à ce jour, la synthèse par concaténation offre le meilleur rendu sonore. En revanche, la synthèse par modèle statistique offre beaucoup plus de souplesse et nécessite bien moins de mémoire que la synthèse par concaténation. Des modèles hybrides ont d'ailleurs été proposés pour tirer partie des avantages des deux méthodes [ZTB09].

1.3.4 Synthèse par transformation de voix

Une dernière catégorie de synthèse consiste à modifier les caractéristiques d'un segment de voix enregistrée pour en produire un nouveau. Les applications présentées ici concernent essentiellement la synthèse du chant.

Synthèse du chant depuis la voix parlée

Une méthode de transformation de parole en chant a été développée par Saitou [STUA04]. Celle-ci combine l'extraction des paramètres vocaux par Vocodeur, et leur modification. La parole est ensuite resynthétisée depuis les paramètres modifiés. Parmi les modifications, un modèle de contrôle de hauteur pour le chant a été développé, et les paramètres spectraux spécifiques à la voix chantée sont ajoutés, notamment la présence d'un vibrato, et d'une résonance appelée formant glottique.

Calliphony est un système de modification en temps réel de l'intonation [LBRd07], [LB09] et du rythme [LBdRD10] de parole enregistrée. Intonation et rythme sont modifiées par l'algorithme TD-PSOLA selon les positions verticales et horizontales du stylet sur une tablette graphique. De la parole enregistrée peut donc être chantée en modifiant à souhait rythme et intonation. Il est aussi possible de modifier l'intonation et le rythme de synthèse à partir de texte, ajoutant un contrôle expressif à la synthèse.

Synthèse par mélange de voix

Lors de la réalisation du film *Farinelli* de Gérard Corbiau, il était nécessaire d'obtenir une voix de castrat. Un tel type de voix n'existant plus de nos jours, Depalle *et al.* ont créé une voix de castrat virtuelle en mélangeant une voix de contre-ténor et une voix de soprano colorature [DGR94]. Bien que très fastidieuse car non automatique, cette méthode a permis d'obtenir 39 min de matériel audio de qualité remarquable [6].

1.4 Contrôle et chironomie

Qu'ils mettent en jeu les doigts, le bras (archets, baguettes, coulisses), le pied (pédales), la bouche (embouchures) ou l'appareil vocal en entier (voix), tout instrument de musique acoustique est exclusivement contrôlé par le geste. De plus, si ces gestes sont nécessaires à la production du son, le corps tout entier est en mouvement lors d'une production musicale [WD04]. Lors de la conception d'instruments de musique numériques, le découplage entre production sonore et contrôle nous entraîne dans une branche du domaine de l'interaction homme-machine (IHM) [Cad94], [OSW01]. Néanmoins, dans les débuts de l'IHM, d'autres canaux de communication ont été favorisés. En effet, afin de se rapprocher de la communication entre humains, les recherches se sont principalement concentrées sur l'interaction visuelle, auditive et par la parole, réduisant la communication gestuelle à des tâches simples telles que la manipulation d'une souris ou d'un clavier [Cad94]. Lors de l'avènement des instruments de musique numériques, une attention particulière a été portée à la richesse du canal gestuel, et à son exploitation par le biais de systèmes d'acquisition performants et pertinents afin de contrôler de manière la plus expressive possible le moteur de synthèse.

1.4.1 Typologie des gestes musicaux

En comparant le canal gestuel aux autres canaux d'interaction du corps humain, Cadoz [Cad94] affirme que :

"*Une particularité fondamentale du geste, qui le distingue d'emblée des autres canaux, et qu'il est deux fois double : tout d'abord il est moyen d'action sur le monde physique et moyen de communication informationnelle, ensuite, dans la seconde fonction il est à double sens, c'est-à-dire moyen d'émission et de réception d'information.*"

Il formalise alors ces différents rôles selon les trois dénominations suivantes :

- Le geste *épistémique* permet de prendre conscience de l'environnement qui nous entoure. Il met en jeu trois niveaux de perception : le toucher, sensible à l'état de la surface des objets ; la perception kinesthésique introduisant la notion de mobilité, de parcours sur un objet à identifier ; la proprioception nous renseignant sur la position de nos membres dans l'espace. On peut alors parler de perception *tactilo-proprio-kinesthésique*.
- Le geste *sémiotique*, à l'inverse permet de transmettre des informations. Il peut être purement informatif et structuré (langage des signes, direction du chef d'orchestre, écriture), ou esthétique (chorégraphie, dessin).
- Le geste *ergotique* communique de l'énergie au monde physique qui l'entoure et agit donc sur ce dernier sous la forme de déplacements, déformations.

6. https://www.youtube.com/watch?v=t9h7oB0TpLY (vérifié le 22 octobre 2015)

De ces trois rôles, le geste instrumental est essentiellement ergotique puisque celui-ci consiste à manipuler l'instrument, à agir sur ce dernier. Mais il est aussi épistémique, car le musicien cherche en permanence des informations sur l'instrument, sur la position de ses membres. Enfin, il a aussi un caractère sémiotique puisque le jeu d'un instrument est créateur d'information auditive qu'est la musique produite. Il s'agit donc d'un geste extrêmement riche de par ces différents rôles mais aussi de par ces variétés. En effet, en ne considérant que le rôle ergotique du geste instrumental, il existe de multiples façons d'agir sur un instrument.

Cadoz identifie trois types d'actions [Cad88] :

- Le geste *d'excitation* transmet de l'énergie à l'instrument pour la production sonore. Il s'agit par exemple du mouvement de l'archet sur un violon ou de la vibration des lèvres dans l'embouchure d'un cuivre.
- Le geste de *modification* change la nature de l'objet vibrant. Cette modification peut être quantitative ou continue, comme le déplacement d'un doigt sur une corde. On parle de *modification paramétrique*. Elle peut être aussi qualitative ou discrète comme l'utilisation d'une pédale de piano. On parle alors de *modification structurelle*.
- Le geste de *sélection* choisit ou met en relation les différents objets à faire vibrer. C'est le cas de l'appui sur une touche de piano ou du changement de corde par l'archet sur un violon.

Quelque soit le type d'action effectuée, les gestes ont des niveaux de complexité différents qui se traduisent par le nombre de variables transmises à l'instrument, ou degrés de libertés. Cadoz définit trois notions hiérarchisant les gestes [Cad88].

- Un *canal gestuel* transmet un ensemble de variables dépendantes et associées à un dispositif. Par exemple le mouvement d'un archet sur une corde est caractérisé par sa vitesse, sa pression, son inclinaison, la position du contact avec la corde.
- Une *voie gestuelle* est une variable isolée d'un canal gestuel.
- Une *unité gestuelle* est un ensemble de canaux gestuels. Un clavier de piano est une unité gestuelle composée d'un ensemble de touches, chacune étant un canal gestuel composées d'une seule voie gestuelle (vélocité d'enfoncement).

1.4.2 L'acquisition des gestes

Les gestes instrumentaux, aussi complexes qu'ils soient doivent être transformés d'énergie physique en énergie électrique dans le monde numérique. L'interface entre monde réel et signal électrique se fait par le biais de capteurs. A cause de l'extrême richesse et diversité des gestes instrumentaux, il n'existe pas de capteur optimal pour l'acquisition du geste. Il est donc nécessaire de réfléchir dans un premier temps au mode d'acquisition à adopter, puis dans ce mode, choisir le capteur le plus pertinent pour la tâche musicale auquel il sera associé.

Les différents modes d'acquisition

Wanderley *et al.* ont catégorisé dans différents articles les modes d'acquisition des gestes [Wan97], [WD99], [WD04]. D'abord, ils différencient les *acquisitions directes* et *indirectes* du son. L'acquisition indirecte consiste à retrouver le geste du musicien par analyse du son mesuré à l'aide d'un microphone avec des techniques de traitement du signal et par une connaissance des mécanismes de l'instrument [TDW03]. Par exemple, l'extraction de la fréquence fondamentale du signal permet de retrouver le doigté effectué par le musicien. Cette technique est

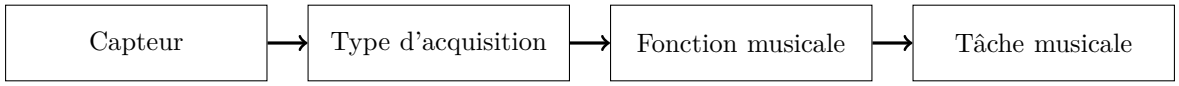

Figure 1.7 – *Des capteurs à la tâche musicale.*

pertinente dans le cas où des instruments acoustiques sont utilisés comme contrôleurs d'instruments numériques. Dans la suite, on ne s'intéressera qu'aux cas d'acquisitions directes, où les gestes sont mesurés directement.

Cadoz distingue trois modes d'acquisition des gestes [Cad94]. Des dispositifs tels que des caméras ou des sonars permettent d'effectuer des acquisitions *sans contact*. Bien que limitées en degrés de libertés, celles-ci ont l'avantage de ne pas interférer avec les actions du musicien. A l'opposé, on cherche parfois à capter le geste *par contact*, avec des *capteurs* effectuant des mesures sur le geste en immersion (gants, combinaisons, exosquelettes), ou en vis-à-vis (mesure de l'interaction entre un geste et un objet seulement). Cela peut aussi se faire par des *effecteurs*, mesurant un transfert d'énergie. Un retour tactile ou d'effort est alors nécessaire.

De ces différents modes d'acquisition, il est possible de combiner différents capteurs ou effecteurs pour construire un contrôleur permettant d'agir sur le moteur de synthèse. Wanderley *et al.* tracent trois catégories de contrôleurs [WD99], [WD04].

- Les contrôleurs *imitatifs* sont identiques aux instruments acoustiques originaux. Cela permet d'exploiter pleinement les techniques de jeu déjà apprises sur les instruments acoustiques. Les plus beaux exemples sont les claviers de piano, intensément utilisés pour le contrôle des premiers synthétiseurs.
- Les contrôleurs *analogues* ressemblent aux instruments acoustiques pour exploiter les techniques de jeu des musiciens mais proposent des ouvertures de jeu.
- Les contrôleurs *alternatifs* sont inspirés de contrôleurs non conçus pour la musique. Bien que nécessitant un apprentissage nouveau, ils permettent une plus grande exploration des gestes musicaux possibles. C'est le cas de la tablette graphique par exemple.

Du capteur à la tâche musicale

Une fois le mode d'acquisition et la catégorie de contrôleur choisis, il est primordial de sélectionner les capteurs de manière pertinente. Pour cela, une réflexion sur le lien entre capteur et tâche musicale associée est à conduire. La figure 1.7 schématise ce lien. Il convient en premier lieu d'identifier les tâches musicales proposées par l'instrument. Orio *et al.* proposent une liste des tâches musicales les plus rencontrées [OSW01] que l'on peut hiérarchiser de la sorte :

- Production de tons isolés (choix de la note et de l'amplitude)
- Gestes musicaux basiques (glissandos, trilles)
- Modulation continue de caractéristiques sonores (vibrato pour la hauteur, timbre)
- Gammes et arpèges simples
- Rythmes simples
- Phrases musicales
- Synchronisation avec d'autres musiciens

Tâche musicale	**Fonction musicale**
Tons isolés	Dynamique absolue
Gestes musicaux basiques (glissandos, trilles)	Dynamique absolue
Modulation continue de caractéristiques (hauteur, timbre)	Dynamique relative

TABLE 1.2 – *Tâches musicales et fonctions musicales, d'après [VUK96] et [OSW01].*

On remarque que les trois premières sont des tâches "unitaires" à partir desquelles sont construites les autres tâches. Ce sont sur ces tâches unitaires qu'il faut associer des gestes en priorité. La combinaison et l'enchaînement de ces gestes permettront alors de construire les tâches plus complexes que sont les gammes, les rythmes ou les phrases musicales.

Selon la variation des paramètres associés à chaque tâche musicale, Vertegaal *et al.* classifient ces dernières en trois fonctions musicales [VUK96]. Une tâche peut d'abord être statique, nécessitant des paramètres stationnaires comme l'accordage d'un instrument. A l'inverse, une tâche dynamique nécessite des paramètres évoluant au cours du temps. Il s'agit alors soit de paramètres *dynamiques absolus* tels que le choix d'une note, ou de paramètres *dynamiques relatifs* tels que la modulation d'une caractéristique sonore. Le tableau 1.2 met en relation les tâches musicales relevées par Orio et les fonctions musicales de Vertegaal.

La variété de capteurs existants de nos jours est immense et permet un très grand nombre de possibilités pour la conception du contrôleur. Marshall *et al.* ont recensé les principaux capteurs utilisés par les instruments de musique numériques présentés aux huit premières conférences *New Interfaces for Musical Expression (NIME)*. Parmi les plus utilisés, on retrouve les capteurs de force résistifs FSR (*Force-Sensing Resistors*), les accéléromètres, la caméra vidéo pour une acquisition sans contact, des boutons ou interrupteurs, des potentiomètres, des capteurs de position, etc. Chacun de ces capteurs possède des propriétés différentes selon les mesures réalisées. Les travaux de Card [CMR91], Vertegaal [VUK96] et Wanderley [WD99] ont abouti une classification des capteurs selon trois critères :

- Les *degrés de libertés* mesurés par chaque capteur, parmi les trois translations et trois rotations possibles.
- La *grandeur physique* mesurée (position, force, angle, couple)
- La résolution du capteur (binaire, discret, continu)

De cette classification est né un mode de représentation des capteurs montré en figure 1.8. Chaque colonne représente un degré de liberté. Chaque ligne représente une grandeur physique mesurée (de haut en bas : position, vitesse, force, variation de force). Dans chaque case, on représente de gauche à droite la résolution (binaire, discret, continu). Les différents capteurs listés dans [VUK96] et [WD99] sont représentés par des cercles blancs. Chaque cercle correspond à une voie gestuelle. Lorsque le capteur possède plusieurs voies, celles-ci sont reliées par un segment.

Enfin, Vertegaal établit des liens entre types d'acquisition et fonctions musicales [VUK96], montrées en table 1.3. Pour chaque fonction musicale sont indiqués les types d'acquisition les plus adaptés, par ordre décroissant. Les capteurs de position sont préférés pour les fonctions musicales statiques et dynamique absolues alors que les capteurs de forces semblent plus adaptés aux fonctions musicales dynamiques relatives.

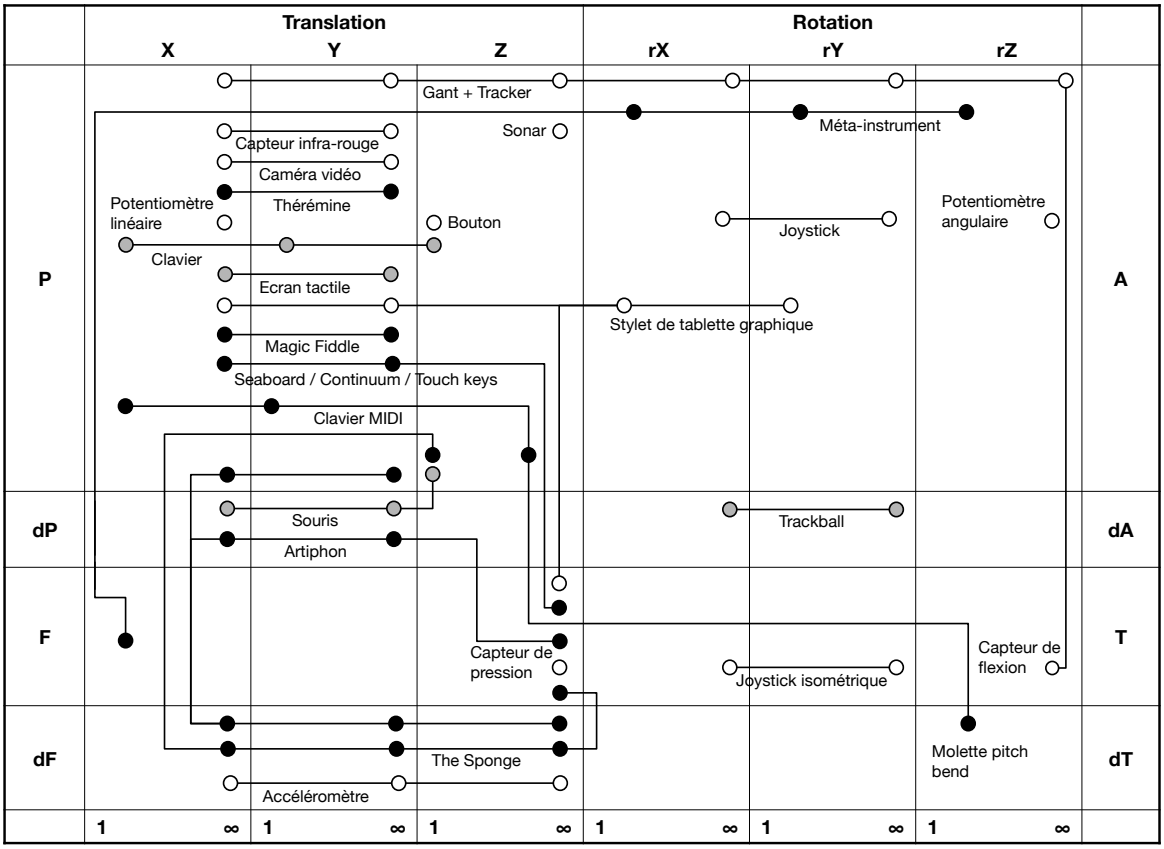

Figure 1.8 – *Capteurs classés selon leur mode d'acquisition [VUK96], [OSW01], d'après la représentation de [CMR91]. Les colonnes sont les différents degrés de liberté, les lignes les grandeurs mesurées, et la résolution des capteurs dépend de leurs positions horizontales dans chaque case. En blanc, les voies gestuelles des capteurs individuels ; en gris, les voies gestuelles d'interfaces pour l'ordinateur ; en noir, les voies gestuelles d'instruments de musique numériques. Les voies appartenant aux mêmes canaux gestuels sont reliées par des segments.*

Fonction musicale	**Type d'acquisition**
Statique	Position linéaire Position angulaire Force isotonique angulaire
Dynamique absolue	Position linéaire Force isométrique linéaire Force isométrique angulaire
Dynamique relative	Force isométrique Force isotonique Vitesse Position

Table 1.3 – *Fonctions musicales et types d'acquisition préférés dans l'ordre décroissant, d'après [VUK96].*

Pour conclure, les trois classifications des tables 1.2, 1.3 et de la figure 1.8 permettent à partir d'une tâche musicale donnée de trouver le ou les capteurs qui semblent les plus pertinents pour son contrôle. Pour aller plus loin, des méthodes statistiques peuvent être

envisagées dans l'association des gestes aux tâches musicales. Par exemple, l'étude de corrélations entre geste de l'auditeur et propriétés sonores ont mis en évidence des associations récurrentes entre geste et son [CBS10]. Une étude similaire sur les gestes de l'instrumentiste permettrait de renforcer la classification de Vertegaal.

Exemples d'applications et évaluations

A titre d'exemples, nous avons classé huit instruments de musique numériques ou électroniques commercialisés ou au point de l'être selon leurs modes d'acquisition en figure 1.8. Parmi ces instruments on note quatre instruments de type analogue. Le *Continuum*[7] [HTW98], le *Seaboard*[8] [LR11] et le *TouchKeys* [MGS13] sont trois claviers augmentés proposant une acquisition continue des doigts sur les dimensions X et Y du clavier. *L'Artiphon*[9] [BJ14] est un manche de guitare augmenté de capteurs de pression et d'un accéléromètre. Les quatre autres instruments sont alternatifs, utilisant des interfaces initialement non conçues pour la musique. Le plus ancien est le *Thérémine*[10] dont la proximité des mains avec deux antennes permet de contrôler la hauteur et le volume du son produit. Le *Méta-Instrument*[11] [dLG06] est un exosquelette captant les mouvements de rotation des articulations des membres supérieurs. The *Sponge*[12] [Mar10] est un bloc de mousse équipé de capteurs de pressions, de boutons et d'accéléromètres. Le *Magic Fiddle* [WOL11] est une application pour tablette numérique utilisant les capteurs de cette dernière (écran tactile).

On remarque que la grande majorité de ces instruments utilisent des mesures de translation plutôt que de rotation. De plus, tous les instruments sauf un ont des mesures de position continues, reliées au contrôle de la hauteur. L'exception est *The Sponge* qui utilise des combinaisons de boutons selon le système binaire. Les mesures d'accéléromètre sont pour la plupart reliées au contrôle du timbre, alors que les mesures de pression contrôlent l'amplitude du son produit. Finalement, par l'analyse succincte de quelques instruments numériques commercialisés, on remarque globalement que les mêmes types de capteurs sont utilisés pour le contrôle de tâches musicales similaires.

Wanderley *et al.* proposent une mise en application des classifications présentées ci-dessus dans l'association d'une tablette graphique augmentée de capteurs de pression au contrôle de la voix chantée (synthèse par FOF). Afin de confronter le choix du capteur issu de la classification aux préférences des utilisateurs, les auteurs comparent un contrôle du vibrato (dynamique relatif) par un capteur de pression (force isométrique), un capteur de position linéaire (position linéaire) et l'inclinaison du stylet (position angulaire). D'après le tableau 1.3 le capteur de pression semble le plus adapté. Les résultats montrent effectivement que les sujets préfèrent largement le capteur de pression, l'angle su stylet étant le moins apprécié.

Marshall *et al.* ont conduit deux expériences similaires. Pour la première, différents capteurs (Potentiomètres linéaires, angulaires, capteurs de pression, linéaire et de flexion) sont testés pour les tâches de sélection de note (dynamique absolue), de modulation de note (dynamique relative) et les deux en même temps [MW06]. Il s'avère que le capteur linéaire est plus apprécié pour la sélection de note, comme le prédit la classification (tableau 1.3), et que le capteur de pression et le capteur linéaire sont également préférés pour le vibrato. En

7. http://www.hakenaudio.com/Continuum/ (vérifié le 22 octobre 2015)
8. https://www.roli.com/products/seaboard-grand (vérifié le 22 octobre 2015)
9. http://www.artiphon.com (vérifié le 22 octobre 2015)
10. https://www.youtube.com/watch?v=w5qf9O6c20o (vérifié le 22 octobre 2015)
11. http://www.pucemuse.com/recherche-developpement/meta-instrument/ (vérifié le 22 octobre 2015)
12. http://www.martinmarier.com/wp/?page_id=12 (vérifié le 22 octobre 2015)

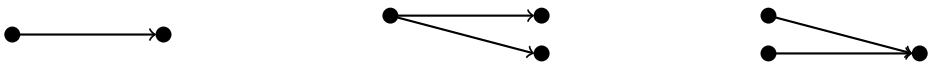

Figure 1.9 – *Mappings unitaire, divergent et convergent.*

revanche, le capteur linéaire semble plus adapté pour la combinaison des deux tâches.

La deuxième expérience compare le jeu d'un vibrato chez des pianistes et violonistes avec des mouvements de glissement ou de roulement sur un capteur linéaire ou un enfoncement sur un capteur de pression [MHWL09]. Les musiciens globalement ont favorisé le capteur de pression. Le transfert de compétence attendu chez les violonistes (roulement du doigt) ne s'est pas manifesté.

Il en ressort de ces trois études que la classification de Vertegaal 1.3 semble refléter raisonnablement les attentes des utilisateurs sur l'association des capteurs et des tâches musicales. Toutefois, ce n'est que par le test intensif de l'instrument que l'on peut choisir au mieux les capteurs associés à chaque tâche musicale. Nous pouvons donc conclure cette partie en proposant l'usage de la classification pour extraire un petit ensemble de capteurs les plus pertinents à la tâche musicale étudiée. Ensuite, c'est par le test de chacun de ces capteurs que le choix final doit s'effectuer.

1.4.3 Des capteurs au synthétiseur : le mapping

Les capteurs fournissent à l'instrument des paramètres relatifs aux gestes (positions, vitesses, etc.) qui doivent contrôler le synthétiseur. Il est donc nécessaire d'établir une correspondance entre les paramètres de sortie du contrôleur et les paramètres d'entrées du synthétiseur. On appelle communément cette correspondance le "mapping", pouvant présenter des niveaux de complexité très variés.

Complexité

Wanderley distingue trois manières de faire correspondre un ensemble de paramètres d'entrée à un ensemble de paramètres de sortie [Wan97], [WD99], [WD04]. La plus simple consiste à faire correspondre un paramètre de sortie à un paramètre d'entrée par une certaine transformation. Il s'agit du mapping *unitaire* (ou *one-to-one*). Il est ensuite possible de contrôler plusieurs paramètres de sortie avec un seul paramètre d'entrée. On parle alors de mapping *divergent* ou *one-to-many*. A l'inverse, on peut combiner plusieurs paramètres d'entrées pour contrôler un paramètre de sortie. Il s'agit alors d'un mapping *convergent* ou *many-to-one*. La figure 1.9 illustre ces trois types de mapping.

Couches

La création de plusieurs étapes de transformations est généralement souhaitée dans la conception d'instruments de musique numériques, menant à un contrôle complexe de l'instrument qui élargit les possibilités de jeu et d'expressivité. Une identification des différents états des paramètres et des transformations associées a été réalisée en parallèle par Arfib [ACKV02] et Hunt [HW02], et est illustrée en figure 1.10.

La première transformation fait le lien entre les *paramètres de contrôle* issus du contrôleur, et des paramètres plus abstraits liés au gestes du musiciens qu'on appellera *paramètres gestuels*. Ces derniers sont principalement des normalisations, des extrapolations ou des dérivées des paramètres de contrôle reflétant les mouvements physiques du musicien. La dernière

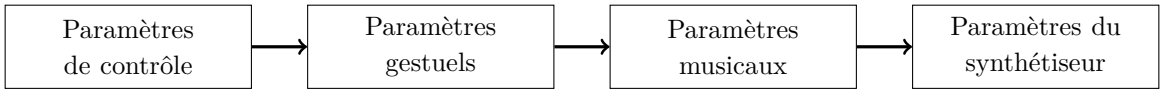

FIGURE 1.10 – *Différentes couches de mapping, d'après [ACKV02] et [HW02].*

transformation fait la liaison entre les *paramètres sonores* ou *musicaux* associés aux caractéristiques acoustiques du son (hauteur, timbre), et les *paramètres de synthèse* injectés à l'entrée du synthétiseur. Hunt *et al.* [HW02] proposent une chaîne de transformation à trois étapes, où les paramètres gestuels et sonores coïncident. Arfib *et al.* [ACKV02] vont plus loin et proposent une chaîne d'évolution à quatre étapes où une transformation intermédiaire fait le lien entre paramètres gestuels et paramètres musicaux.

Dynamique

Afin d'enrichir d'avantage le mapping entre les paramètres du contrôleur et ceux du synthétiseur, il est possible de rendre les fonctions de transformations des paramètres dynamiques, c'est-à-dire qui varient en fonction du temps. Parmi les exemples trouvés dans la littérature, on peut relever des mappings évoluant par eux-mêmes selon des règles prédéfinies, introduisant une production sonore non causée directement par le musicien, comme les Modèles Intermédiaires Dynamiques [GGGD11]. On peut aussi relever des mappings qui s'adaptent au jeu du musicien afin de faciliter le jeu de l'instrument. Le chapitre 4 traite plus en détail des mappings adaptatifs.

La conception du mapping est souvent négligée dans la conception d'instruments de musique. Pourtant, c'est celui-ci qui définit la richesse d'un instrument [HWP03]. Une correspondance trop simple entre les paramètres rend l'instrument peu intéressant car très limité en termes de contrôle. A l'inverse, un mapping trop complexe peut rendre incompréhensible le fonctionnement de l'instrument. Un juste milieu doit donc être trouvé, suffisamment complexe pour proposer une richesse de contrôle et un potentiel d'exploration suscitant un intérêt pour l'instrument, tout en restant intuitif pour le musicien.

1.4.4 Le retour

Enfin, comme indiqué en figure 1, le musicien reçoit des retours sur ses actions par les différents blocs de la chaîne de l'instrument de musique numérique [WD99]. On distingue un retour primaire et un retour secondaire. Le retour primaire concerne les informations apportées suite à la manipulation de l'instrument. Il s'agit donc d'informations tactilo-kinesthésiques (dimension épistémique du geste instrumental) et visuelles. Le retour secondaire est le son produit par l'instrument perçu par le canal auditif (dimension sémiotique du geste instrumental). Il existe aussi un retour primaire auditif concernant les bruits découlant des mécanismes de l'instrument, mais ce dernier est souvent masqué par le retour secondaire.

Finalement, toutes les études présentées dans cette section fournissent un appui scientifique dans l'élaboration de chaque bloc de la figure 1 constituant le contrôleur d'un instrument de musique numérique. Parallèlement à celles-ci, les travaux de Cook proposent une liste non exhaustive de principes pragmatiques pour la conception d'instruments, énoncés selon l'expérience de luthiers numériques [Coo01], [Coo09].

1.5 Synthétiseurs de voix chantée contrôlés en temps réel

Le choix important des méthodes de synthèses vocales existantes et leurs évolutions constantes ont amené de nombreux chercheurs ou ingénieurs à développer des instruments numériques vocaux de toutes sortes. Quelques exemples sont donnés dans cette partie.

Voder

Le VODER (*Voice Operation DEmonstrator*) est un des premiers synthétiseurs de parole par formant construit en 1939 par Dudley, aux laboratoires Bell [DRW39]. Le synthétiseur est constitué de deux sources (vocale et bruitée) et de dix filtres résonants constituant les formants. Le processus de contrôle est le suivant : le poignet gauche appuie sur une barre permettant de passer du mode voisé à non voisé. L'intonation est contrôlée continûment par l'enfoncement d'une pédale. 10 capteurs de pressions contrôlés indépendamment par les 10 doigts permettent d'atténuer l'amplitude de chacun des 10 formants. Le contrôle des voyelles se fait donc par combinaison des doigts. Trois boutons additionnels permettent de déclencher les consonnes plosives non-voisées (/t/,/p/,/k/) ou voisées (/d/,/b/,/g/) selon la position du poignet (figure 1.11). Il en résulte un contrôle relativement complexe nécessitant plusieurs mois d'apprentissage intensif au bout desquels des discours très simples sont produits avec une intelligibilité moyenne [13].

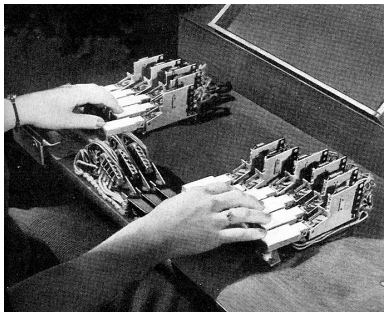

Figure 1.11 – *Le contrôleur du Voder, d'après [DRW39].*

SPASM

Cook a développé une interface de contrôle temps-réel pour le synthétiseur SPASM [14], [15] [Coo93]. Les paramètres de contrôle performatif sont la hauteur mélodique, la fréquence et la quantité de vibrato, ainsi que la quantité de variations aléatoires du vibrato. La section du conduit vocal, du vélum, et le contrôle du bruit sont aussi modifiables pour changer le son en temps réel. Des modules tels que l'interpolateur de forme de conduit, l'interpolateur d'espace vocal ou l'interpolateur d'effort vocal permettent de contrôler continûment et respectivement la forme du conduit dans un espace 2D, le choix des voyelles dans un espace 2D et l'effort vocal dans un espace 1D (figure 1.12). Des contrôleurs MIDI peuvent être utilisés pour ces contrôles. Cook a ensuite développé des interfaces spécifiques pour la voix chantée [Coo01], [Coo05], testant plusieurs types de synthèse (formants, FOF, modèle physique).

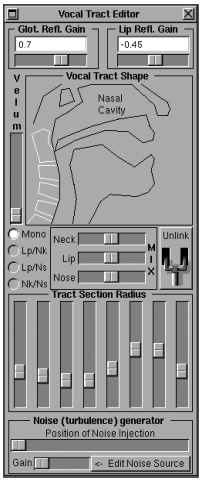

Figure 1.12 – *L'interface de contrôle du conduit vocal du SPASM, d'après [Coo01].*

13. https://www.youtube.com/watch?v=5hyI_dM5cGo (vérifié le 22 octobre 2015)
14. http://www.cs.princeton.edu/~prc/DaisyD2.mp3 (vérifié le 22 octobre 2015)
15. http://www.cs.princeton.edu/~prc/HappyBirthdayMAXFEST07.qt (vérifié le 22 octobre 2015)

La famille des SqueezeVoxen sont des accordéons augmentés de capteurs FSR linéaires et de flexion (figure 1.13). Le clavier de l'accordéon contrôle la hauteur mélodique, avec des modulations fines et continues possibles par les capteurs FSR. Le souffle est contrôlé par le soufflet de l'accordéon. Les boutons de la main gauche sélectionnent des diphones / mots / phrases présélectionnés. Quatre capteurs de flexion correspondent aux quatre positions d'articulation (ouverture de la mâchoire, position de la pointe de la langue, position du plat de la langue, ouverture du vélum). D'autres contrôleurs tels que le COWE (*Controller, One With Everything*) mélangeant capteur de souffle, FSR, boutons et accéléromètres ou le VOMID (*Voice-Oriented Melodica Interface Device*) basé sur le principe du mélodica ont aussi été développés.

FIGURE 1.13 – *Les SqueezeVoxen Lisa et Bart, d'après [Coo01].*

Glove-Talk

Bien que le *Glove-Talk* ne soit pas un instrument de voix chantée, il est l'un des contrôleurs de synthèse de parole permettant d'articuler en temps réel. Une première version construite en 1992 par Fels [FH92] est constituée d'un synthétiseur TTS, d'un gant intelligent possédant deux capteurs de flexion par doigt, ainsi qu'un accéléromètre et d'un gyromètre. Un réseau de neurones est implémenté pour apprendre chaque geste que l'utilisateur choisit d'associer aux mots du dictionnaire.

FIGURE 1.14 – *Gant de la main gauche du Glove-Talk II, d'après* [16].

Dans sa deuxième version construite en 1998 [FH98], le *Glove-Talk II* ne sélectionne plus des mots mais les paramètres du synthétiseur tels que les formants. Cela permet d'offrir un vocabulaire illimité à l'utilisateur. Un gant similaire ainsi qu'accéléromètre et gyromètre sont toujours placés sur la main droite. Une position ouverte de la main déclenche des voyelles contrôlées par la position de celle-ci dans le plan X-Y. La fermeture des doigts déclenche des consonnes. La hauteur de la main contrôle l'intonation et une pédale contrôle l'amplitude. A nouveau, un réseau de neurones permet d'apprendre les gestes de l'utilisateur. Fels déclare que 100 heures d'entraînement permettent de créer un discours intelligible [16]. Le contrôle du Glove-Talk a ensuite été utilisé par le GRASSP dans une application pour la voix chantée [17] [PF06].

FOF

La carte *Vocalise* est la première implémentation en temps réel du synthétiseur CHANT de Rodet [RPB84] sur le microprocesseur TMS 320, réalisée en 1984 [DdR84]. La synthèse est contrôlée par un ensemble de potentiomètres, et l'instrument fût présenté une dizaine d'années à la Cité des Sciences à Paris.

16. https://www.youtube.com/watch?v=hJpGkroFP3o (vérifié le 22 octobre 2015)
17. https://vimeo.com/8983689 (vérifié le 22 octobre 2015)

Wanderley *et al.* proposent l'utilisation d'une tablette graphique pour le contrôle de CHANT [WVIR00]. Un capteur de position et de pression est utilisé pour contrôler la fréquence fondamentale (position) et le vibrato (pression). La position du stylet sur la tablette permet de sélectionner les voyelles dans un espace vocalique décrit par la fréquence du 1er formant (axe Y) et du 2e formant (axe X). Les fréquences des 3 formants suivants ainsi que les amplitudes et les bandes passantes sont fixées.

Figure 1.15 – *Contrôleur de FOF proposé par Hunt, d'après [HHW00].*

Un contrôle plus exotique des FOF est proposé par Hunt *et al.* [HHW00]. Deux capteurs de pression FSR sont placés sur une balle de tennis pour contrôler la hauteur et l'amplitude de la voix. Trois autres capteurs sont placés en triangle sur une surface spongieuse, recouverts d'une surface solide. En pressant sur cette dernière l'utilisateur peut répartir la pression sur chacun des trois capteurs chacun relié à l'amplitude des trois premiers formants.

Voicer

Le *Voicer* de Kessous est l'un des premiers instruments de synthèse de voix chantée contrôlée en temps réel par une tablette graphique [Kes04a], [Kes04b]. Trois moteurs de synthèse par formants ont été testés : un signal en dent de scie traversant trois filtres du second ordre en cascade ; un signal en dent de scie traversant 5 FOFs ; un signal modélisant l'onde de débit glottique traversant 5 filtres résonants en cascade. Le contrôle de la hauteur mélodique se réalise avec la main habile de manière cyclique avec une octave par tour, correspondant à une représentation hélicoïdale de la hauteur. Une tablette graphique ainsi qu'un gant sur un écran ont été testés pour ce contrôle. L'autre main sélectionne les voyelles de manière continue dans un plan en deux dimensions avec soit un joystick soit une deuxième tablette. De plus un retour visuel est présenté à l'utilisateur. Une démonstration de l'instrument peut être écoutée en suivant le lien en bas de page [18].

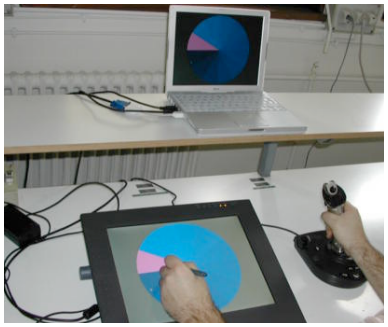

Figure 1.16 – *Le Voicer, d'après [Kes04b].*

Handsketch

Handsketch [19] [dD09b], [dD09a] est un des instruments de synthèse vocale performative qui avec le *Cantor Digitalis* ont fait suite au séminaire eNTERFACE 2008 [ddLB+08].

Tout comme le *Cantor Digitalis*, il utilise le synthétiseur de source RT-CALM [ddLBD06b], [dWF+07] décrit au chapitre 2. Trois contrôles sont proposés : la hauteur mélodique est choisie continûment par la position angulaire du stylet sur la tablette. La position radiale du stylet contrôle la qualité

Figure 1.17 – *Handsketch, d'après [dD09b].*

18. http://www.jmc.blueyeti.fr/Videos/Voicer.mpg (vérifié le 22 octobre 2015)
19. https://vimeo.com/20461414 (vérifié le 22 octobre 2015)

de voix. L'intensité est quant à elle liée à la pression du stylet sur la tablette. 8 FSR sont rajoutés sur la tablette et manipulés par la main non-habile (figure 1.17). Selon les combinaisons choisies, il est possible de choisir la hauteur de manière discrète, d'ajuster celle-ci en l'attirant vers des notes justes, et d'articuler des phonèmes.

Transvoice Table

La *Transvoice Table*[20] est un système permettant la modification en temps réel de signaux de parole [dDD08]. Trois modifications sont proposées : la modification de l'intonation pour passer de voix d'homme (plus grave) à femme (plus aigüe) ou personne âgée (moins stable) ; la création de voix soufflées par extraction de l'enveloppe spectrale du signal et son application à un bruit rose ; le contrôle d'une voix de synthèse par le système *Handsketch*. Deux autres modifications hors-temps sont aussi proposées : la modification de la pente spectrale du signal associée à l'effort vocal ; la séparation des valeurs médianes et des fluctuations fines de la hauteur mélodique pour amplifier ou diminuer ces dernières, liées à l'expressivité.

MAGE

MAGE est une plate-forme de contrôle temps réel de synthèse paramétrique par une tablette graphique [AdD12]. Le système de synthèse HTS [ZTB09] décrit plus haut est utilisé pour la phase d'apprentissage. Pour la phase de synthèse, celui-ci est modifié pour permettre une synthèse en temps réel, trame par trame. L'interface *Handsketch* est par la suite utilisée pour contrôler la hauteur, la vitesse, l'intensité et la longueur du conduit vocal associés au signal de parole[21].

1.6 Conclusion

Ce chapitre a montré à la fois les enjeux de la synthèse vocale, ainsi que les différentes techniques mises en œuvres pour aboutir à des sons vocaux artificiels de qualité. Parallèlement, une formalisation du geste et du contrôle instrumental est proposée, fondant les bases de la lutherie numérique. L'association entre geste et tâche musicale nécessite d'être soigneusement pensée, par la sélection de capteurs les plus appropriés pour une tâche donnée. Divers synthétiseurs de voix chantée ont émergé de ces études, explorant les méthodes de synthèse spectrales, physiques ou par corpus, et utilisant des interfaces allant des plus simples (potentiomètres) aux plus sophistiquées (tablette graphique). Cette dernière semble être la plus appréciée pour le contrôle temps réel de la voix chantée et son utilisation est l'objet de cette thèse, à travers un autre instrument, le *Cantor Digitalis*, présenté dans le chapitre suivant.

20. http://www.dailymotion.com/video/x809dn_numediart-transvoice-table-video1_tech?start=3 (vérifié le 22 octobre 2015)
21. https://www.youtube.com/watch?v=W7OwfU0A_HM (vérifié le 22 octobre 2015)

Chapitre 2

Le Cantor Digitalis

Sommaire

2.1 Introduction

Le *Cantor Digitalis* est un synthétiseur de voix chantée contrôlé en temps réel par une tablette graphique et développé au LIMSI-CNRS depuis une dizaine d'années. Il est actuellement implémenté sur la plateforme Max/MSP développée par Cycling [1]. L'idée a émergé lors des workshops eNTERFACE 2005 [ddLB+05] et 2006 [dDLB+06] où les projets portaient sur le contrôle gestuel en temps réel de la voix chantée expressive. Différents types de synthèses ont été essayés, notamment le moteur MBROLA [DPP+96] pour de la synthèse de parole à partir du texte, ou des synthèses par formants basés sur divers modèles de source (LF [FLL85] et CALM [DdH03]). De même, plusieurs contrôleurs ont été expérimentés, comme des gants intelligents, des claviers MIDI équipés de pédales, ou des joysticks. Il est apparu que le synthétiseur par formants basé sur la source CALM et le gant de donnée permettaient la plus grande expressivité. De plus, une étude de d'Alessandro *et al.* montre qu'après comparaison du modèle temporel LF et du modèle spectral CALM de la source, le modèle spectral est moins coûteux en calculs et plus facile à contrôler dû à l'aspect perceptif de la représentation spectrale. De plus, le lien avec le conduit vocal est plus immédiat [ddLBD06a].

De ces explorations, Sylvain Le Beux a mis au point pendant sa thèse de doctorat au LIMSI-CNRS [LB09] le premier moteur de synthèse du *Cantor Digitalis* basé sur une implémentation temps réel du modèle de source CALM (RT-CALM [ddLBD06b], [dWF+07]). Plusieurs types de contrôle ont été étudiés, comme le gant de données P5, la tablette graphique Wacom, et le Méta-instrument de PuceMuse [dLG06]. Lionel Feugère a par la suite perfectionné le moteur de synthèse en introduisant un modèle de conduit vocal de qualité, basé sur un système par règles de formants en parallèle, lors de sa thèse de doctorat au LIMSI-CNRS [Feu13]. Il fixe aussi le contrôle de la synthèse à la tablette graphique, définissant ainsi l'instrument tel qu'il est utilisé aujourd'hui.

La description technique du *Cantor Digitalis* est donnée en première partie de ce chapitre, résumant les travaux effectués précédemment [2]. Nous décrivons le modèle de source RT-CALM du *Cantor Digitalis* mis en place pendant la thèse de Le Beux, ainsi que le modèle de filtre conçu par Feugère. Nous présenterons dans un deuxième temps l'interface de contrôle finale choisie par ce dernier. Le travail effectué sur l'instrument pendant cette thèse est présenté dans la section 2.3. Il s'agit de la diffusion du *Cantor Digitalis* sous forme de logiciel [3]. Cette section présente le travail de présentation du code et de l'interface effectué pour un usage simple de l'instrument, la rédaction de la documentation associée, ainsi que les méthodes de diffusion du logiciel au grand public.

2.2 Description technique

2.2.1 Le synthétiseur

Le *Cantor Digitalis* est un synthétiseur par formants. Il est donc construit sur le modèle source-filtre, où la source vocale est synthétisée dans un premier temps, puis filtrée par le conduit vocal constitué de formants en parallèle. Afin de perfectionner le modèle, des interactions entre la source et le conduit vocal ont été introduites.

1. https://cycling74.com (vérifié le 22 octobre 2015)
2. Ces travaux ont été soumis aux Transactions in Audio, Language and Signal Processing (TASLP) [FdDP]
3. https://cantordigitalis.limsi.fr (vérifié le 22 octobre 2015)

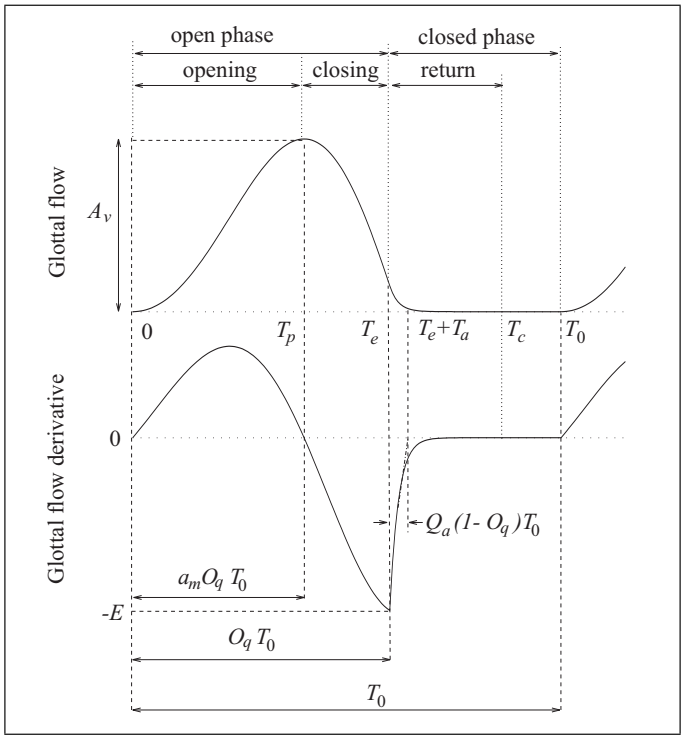

FIGURE 2.1 – *Paramètres décrivant l'ODG (haut) et sa dérivée (bas), d'après [DdH06].*

Source

Selon la description du principe de phonation donné en section 1.2.1, un cycle de vibration des cordes vocales se décompose en quatre phases : une phase d'ouverture puis de fermeture des cordes vocales pendant lesquelles la glotte est ouverte, et une phase de retour puis une phase statique pendant lesquelles la glotte est fermée. Le flux d'air traversant la glotte appelé Onde de Débit Glottique (ODG) croit donc à l'ouverture des cordes vocales, puis décroit à leur fermeture. Il est nul pendant la phase fermée. On appelle cette bosse l'impulsion glottique. La figure 2.1 montre un exemple d'ODG sur un cycle de phonation (haut). Le rayonnement de l'ODG aux lèvres pouvant être approché par la dérivée de cette dernière, on représente en bas de la figure l'ODG dérivée ou ODGD.

Parmi les différents modèles de source les plus utilisés (LF [FLL85], KLGLOT88 [KK90], R++ [Vel98]), Henrich et Doval *et al.* montrent qu'il est à chaque fois possible de résumer la description de la courbe d'ODG à cinq paramètres [Hen01], [DdH06].

- Fréquence fondamentale F_0 : nombre de cycles par seconde.
- Excitation maximale E : variation maximale du déplacement des cordes vocales, c'est-à-dire le maximum de la dérivée de l'onde de débit glottique, situé à l'instant de fermeture.
- Quotient ouvert O_q : rapport de la durée pendant laquelle la glotte est ouverte sur une période de cycle.
- Coefficient d'asymétrie α_m : rapport entre le temps d'ouverture de la glotte sur le temps pendant lequel elle est ouverte.
- Coefficient de retour Q_m : rapport entre le temps de retour et le temps pendant lequel la glotte est fermée. Décrit la fermeture des cordes vocales (abrupte / lisse).

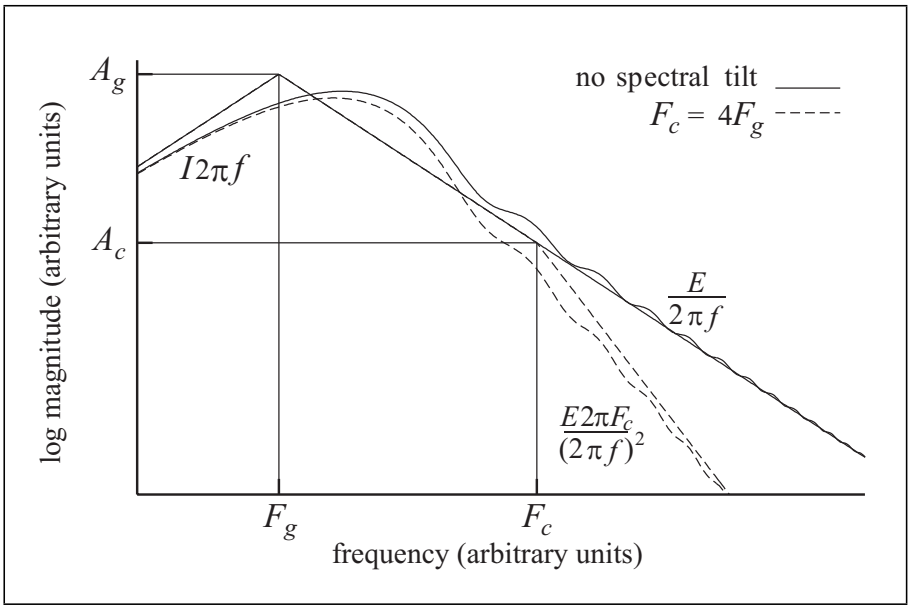

Figure 2.2 – *Représentation spectrale de la source glottique, d'après [DdH06].*

La fréquence fondamentale F_0 indique le nombre de cycles se déroulant par seconde. Perceptivement, cela se traduit par la hauteur du son perçu. L'excitation maximale E dépend de l'amplitude d'écartement des cordes vocales. Plus celles-ci sont écartées, plus elles se refermeront rapidement par élasticité. E est donc relié à l'amplitude du signal.

Les paramètres O_q et α_m définissent la phase d'ouverture. Le quotient ouvert varie entre 0 et 1. Plus il est faible, plus l'impulsion glottique est étroite et cela correspond à une voix tendue. A l'inverse, un temps d'ouverture plus long s'associe à une voix plus relâchée. Le coefficient α_m quantifie l'asymétrie de l'impulsion glottique. L'ouverture de la glotte étant toujours plus longue que la fermeture, α_m est compris entre 0.5 et 1. L'impulsion glottique est presque sinusoïdale avec une asymétrie faible, propre d'une voix relâchée, et proche d'une impulsion avec une forte asymétrie, caractéristique d'une voix tendue.

Le coefficient de retour Q_m se rapporte à la phase de fermeture. Si celui-ci est nul, cela correspond à une fermeture brutale des cordes vocales propre aux voix criées. A l'inverse, un coefficient de retour proche de 1 indique une fermeture lisse, caractéristique d'une voix douce.

Cette décomposition entre phase ouverte et phase fermée de la glotte permet de modéliser le spectre de l'ODGD de manière simple (figure 2.2). La phase ouverte de la glotte se traduit par l'apparition d'une bosse dans le spectre appelée formant glottique. Il se modélise simplement par un filtre passe-bande du second ordre. La position du formant glottique dépend essentiellement du quotient ouvert. Lorsque ce dernier diminue, la voix se tend et le formant est décalé vers les fréquences plus élevées. La largeur du formant glottique dépend quant à elle du coefficient d'asymétrie. Une asymétrie faible entraîne un signal presque sinusoïdal et donc un formant glottique étroit. A l'inverse, une forte asymétrie créé un signal plus riche spectralement et donc un formant glottique plus large.

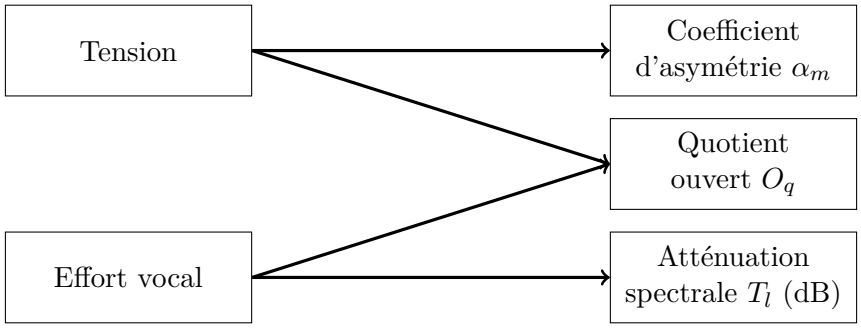

FIGURE 2.3 – *Relation entre paramètres de qualité de voix et paramètres de source, d'après [dWF+07].*

La phase fermée de la glotte définit la pente du spectre dans les hautes fréquences. Celle-ci se modélise par un filtre passe-bas du premier ou deuxième ordre à une fréquence de coupure F_c. Une fermeture abrupte des cordes vocales entraîne une richesse spectrale et donc une pente relativement faible. A l'inverse, une fermeture lisse des cordes vocales est moins riche spectralement et se traduit par une pente plus importante réduisant les hautes fréquences de la voix. Enfin, le maximum d'excitation E agit principalement comme un gain général sur le spectre de l'ODG.

Le modèle linéaire causal-anticausal *CALM - Causal-Anticausal Linear Model* modélise spectralement la source glottique [DdH03]. En analysant l'ODGD selon les phases ouverte et fermée, les auteurs constatent que la phase ouverte peut être considérée comme la réponse à un filtre causal divergent, ou de manière équivalente à un filtre anticausal convergent. Ils modélisent ainsi le formant glottique par un filtre passe-bande anticausal du deuxième ordre, et la pente spectrale par un filtre passe-bas causal du premier ordre.

Une implémentation temps-réel du CALM appelée RT-CALM est réalisée par d'Alessandro *et al.* [ddLBD06b], [dWF+07]. Une première implémentation calcule la partie anticausale par retournement temporel de la réponse à un filtre causal. Pour réduire les coûts de calcul et éviter des discontinuités liées à ce retournement, la deuxième solution calcule directement la réponse à un filtre divergent. Le signal est alors coupé à l'instant de fermeture glottique. Les expressions analytiques des relations entre les paramètres de qualité de voix (tension, effort vocal) et les paramètres de source (quotient d'ouverture, coefficient d'asymétrie, et atténuation de la pente spectrale) sont données dans [dWF+07], selon le schéma 2.3. On a donc deux mappings divergents pour la tension et l'effort vocal, deux mappings unitaires pour le coefficient d'asymétrie et l'atténuation spectrale, et un mapping convergent pour le quotient ouvert. Le RT-CALM est programmé en C et exporté comme objet Max.

La dernière étape de synthèse de la source est l'ajout d'un filtrage supplémentaire à la sortie du bloc RT-CALM par Feugère [Feu13] pour renforcer la dynamique de variation de la pente spectrale. Il s'agit donc d'un filtrage passe-bas du premier ordre, atténuant la pente de 0 à 25 dB selon l'effort vocal.

Parallèlement à la source voisée, une source de bruit blanc filtré passe-bas est ajoutée, et modulée par l'ODGD. Son gain est choisi selon la volonté d'obtenir une voix soufflée ou non. Il est ensuite ajouté au signal voisé.

	Fréquences F_i (Hz)					**Bande-passantes B_i (Hz)**					**Amplitudes A_i (dB)**				
/i/	215	1900	2630	3170	3600	10	18	20	30	40	-13	-23	-2	1	-30
/e/	460	1550	2570	2980	3600	10	15	20	30	40	-1	-3	-2	-2	-5
/a/	700	1200	2500	2800	3600	13	13	40	60	40	0	0	-5	-7	-24
/o/	440	880	2160	2860	3600	10	12	20	30	40	-6	-1	-18	-10	-28
/u/	290	750	2300	3080	3600	10	10	20	30	40	-12	-9	-14	-11	-11
/y/	250	750	2160	3060	3600	10	10	20	30	40	-12	-9	-14	-11	-11

Table 2.1 – *Valeurs des formants du Cantor Digitalis pour une voix de ténor.*

Filtre

La modélisation du conduit vocal a été réalisée finement par Lionel Feugère [Feu13]. Il ramène le conduit vocal à six formants pour le conduit buccal. Ceux-ci sont modélisés par des filtres résonants, passe-bandes du deuxième ordre. Les formants sont décrits par leurs fréquences centrales F_i, leurs bandes passantes B_i ainsi que leurs amplitudes A_i. Ces grandeurs sont choisies pour une voix de ténor pour les six voyelles /i/, /e/, /a/, /o/, /u/ et /y/ bordant le triangle vocalique. Le tableau 2.1 reporte les grandeurs choisies pour le *Cantor Digitalis*. Lorsque le contrôle des voyelles est réalisé dans un plan en deux dimensions, seules les voyelles /a/,/o/,/u/ et /y/ sont utilisées. Dans le cas d'un contrôle en une dimension, les voyelles de référence /i/, /e/, /a/, /o/ et /u/ sont utilisées. Les voyelles intermédiaires sont obtenues en interpolant les valeurs de référence.

Les sinus piriformes sont deux cavités de part et d'autre du larynx introduisant une antirésonance dans le spectre du conduit vocal. Cette dernière est modélisée par un filtre coupe-bande de fréquence centrale 4500 Hz.

Enfin, une modélisation de la taille du conduit vocal est réalisée, permettant de simuler un déplacement vertical du larynx se traduisant perceptivement par l'allongement ou le rétrécissement de la taille du conduit vocal du chanteur ténor de référence. Un conduit vocal plus long résulte en des formants plus éloignés en fréquence et inversement. Par conséquent, un facteur est introduit permettant d'écarter ou de rapprocher les formants. En choisissant une tessiture appropriée, on peut alors transformer la voix de ténor en voix de soprano, d'enfant en diminuant la taille du conduit vocal, ou en voix de basse ou de gros animaux en augmentant la taille du conduit vocal. La fréquence centrale de l'antirésonance des sinus piriformes est aussi modifiée en conséquence. On peut noter que quelle que soit la taille du conduit vocal choisi, la perception des voyelles est possible puisque celle-ci réside dans les rapports entre les positions des formants et non dans leurs positions absolues.

Interactions source-filtre

Une des principales limites du modèle source-filtre est qu'il suppose l'indépendance de la source et du conduit vocal, ce qui n'est pas vérifié en pratique. Afin de combler cette lacune, Feugère a introduit trois interactions entre source et conduit vocal [Feu13].

D'abord, afin d'éviter l'amplification non-naturelle des harmoniques ou de la fréquence fondamentale par les formants, leurs amplitudes sont diminuées lorsqu'ils coïncident avec les harmoniques. Ensuite, la position du premier formant varie avec l'effort vocal. Un effort vocal élevé entraîne un premier formant plus aigu. Enfin, pour optimiser la puissance de leurs voix, les chanteurs adaptent la position des premiers formants par la forme de leurs conduits vocaux à la hauteur de leurs voix. Pour imiter cette technique, les positions des deux premiers formants suivent la fréquence fondamentale dès que celle-ci dépasse un certain seuil.

Contrôleurs	Type d'acquisition	Fonction musicale	Tâche musicale
Position stylet X	Position linéaire	Dynamique absolue	Hauteur
Position stylet X	Position linéaire	Dynamique relative	Vibrato
Position stylet Y	Position linéaire	Dynamique relative	Timbre
Pression stylet	Force isométrique	Dynamique absolue	Effort vocal
Position doigt X	Position linéaire	Dynamique absolue	Articulation
Position doigt Y	Position linéaire	Dynamique absolue	Articulation

TABLE 2.2 – *Les contrôleurs du Cantor Digitalis, et leurs tâches musicales associées.*

2.2.2 L'interface

Parmi toutes les interfaces testées pour le contrôle expressif et temps réel de la voix chantée, c'est la tablette graphique qui a été choisie. Il s'agit de tablettes graphiques Wacom Intuos [4], initialement conçues pour le dessin sur ordinateur est utilisées ici comme contrôleur alternatif pour la voix de synthèse.

Tablette graphique

De nombreuses utilisations de la tablette graphique émergent de la littérature, aussi bien pour des applications musicales non vocales [WWF97], [Cou02], [ZWMC07] que vocales [WVIR00], [Kes04a], [Kes04b], [dD09b], [dD09a], [ddLB+08], [AdD12]. Cet attrait pour cet outil peut être justifié de trois manières. D'abord, du point de vue technologique, la tablette est un outil de qualité fournissant de multiples paramètres à hautes résolutions spatiales et temporelles (figure 1.8). Comparée à un écran tactile où le contrôle se fait au doigt, une étude montre que l'utilisation du stylet permet de plus petits mouvements, et est plus précise dans le dessin de formes complexes [TRZ12]. La tablette fournit aussi la dimension supplémentaire de la pression du stylet. A titre d'exemple, la tablette Wacom Intuos 5 utilisée pour le *Cantor Digitalis* possède une surface active de 233 × 146 mm de résolution spatiale 0.005 mm. 1024 niveaux de pressions sont détectés par le stylet, et la résolution temporelle des mesures est de 5 ms (200 Hz).

De plus, le geste requis pour l'utilisation de cet outil est le maniement d'un stylet. Cette tâche est relativement complexe puisqu'elle consiste dans l'écriture occidentale par exemple, à tracer des boucles, des traits, de manière très précise et coordonnée et pourtant maîtrisée par la plupart de la population. La tablette permet donc l'exploitation d'un geste expert naturel chez l'utilisateur moyen.

Enfin, bien qu'étant un contrôleur alternatif, c'est-à-dire n'étant pas initialement conçue pour la musique, celle-ci fait néanmoins appel à la sensibilité de l'utilisateur par le geste esthétique qu'est celui du dessin. En effet, l'espace gestuel du dessin est beaucoup plus ouvert et moins contraint par les formes imposées par l'écriture. Entre écriture fonctionnelle et dessin libre, on peut associer les gestes instrumentaux de la tablette à la calligraphie, où l'utilisateur doit tout de même suivre certaines règles de tracé, pour commander le synthétiseur, tout en laissant libre cours à son expressivité. Le *Melodic Brush* [HTL+12] est un exemple d'instrument de musique à métaphore calligraphique.

Les tablettes Wacom utilisées captent aussi la position des doigts, ce qui permet de combiner manipulations au stylet et aux doigts. Le tableau 2.2 résume les contrôleurs de la tablette utilisés dans le contrôle du *Cantor Digitalis*, ainsi que les tâches musicales associées, selon la classification de Vertegaal *et al.* [VUK96].

4. http://www.wacom.com/fr-fr/products/pen-tablets/intuos-pro-medium (vérifié le 22 octobre 2015)

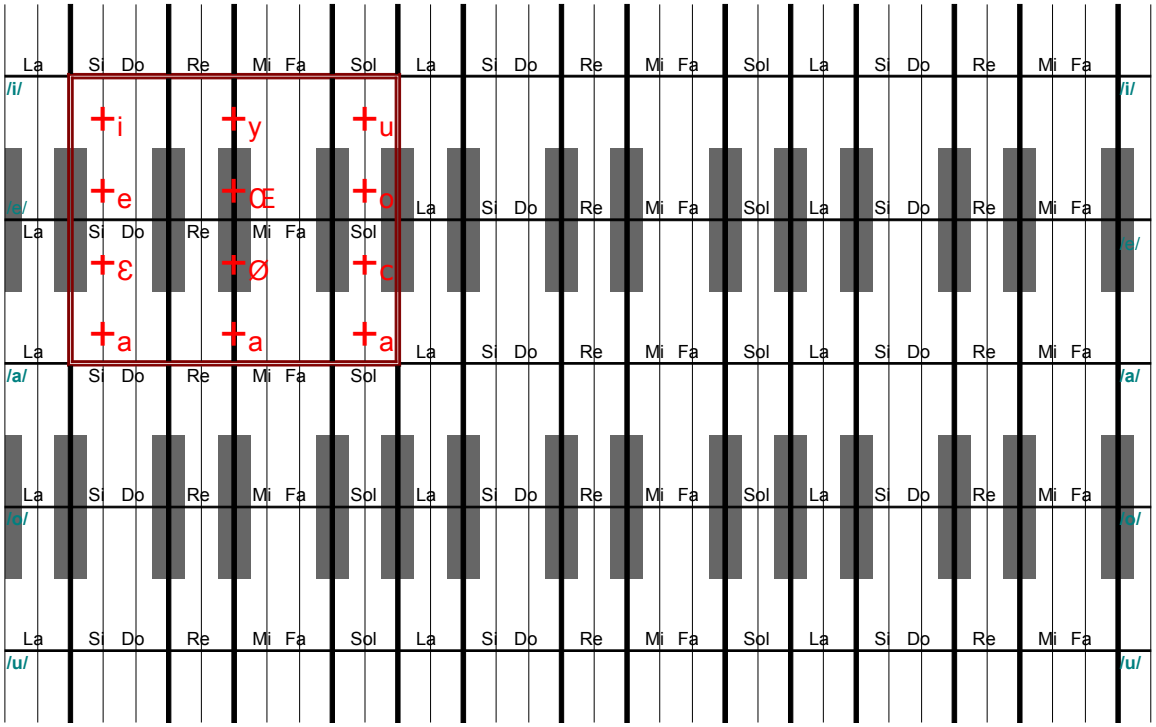

Figure 2.4 – *Calque appliqué à la surface de la tablette pour le jeu de musique occidentale tempérée.*

Les tâches de contrôle de la hauteur et de l'articulation étant dynamiques absolues, celles-ci sont reliées au capteurs de positions linéaires. Le stylet est attribué à la hauteur car celle-ci demande une résolution bien plus importante que l'articulation. Cette dernière se fait donc avec le doigt dans un plan en deux dimensions représentant le triangle vocalique. Afin d'aider ce contrôle, le calque montré en figure 2.4 est imprimé et appliqué sur la surface de la tablette. Sur l'axe horizontal sont représentées des touches de piano, favorisant le repérage rapide des positions des notes, cette représentation étant la plus commune. Il est important de préciser que le contrôle de la hauteur est continu, et que les "touches" dessinées ne sont qu'un indicateur de la note exacte la plus proche dans la gamme tempérée. Celles-ci sont jouées sur les lignes verticales et toute position intermédiaire du stylet entraîne le jeu d'une note intermédiaire. Trois octaves sont accessibles sur la tablette. L'articulation est contrôlée dans le cadre rouge supérieur gauche. Chaque croix représente l'emplacement des voyelles cibles. De même, les voyelles sont contrôlées continûment. Ce calque est parfaitement adapté pour la musique tempérée occidentale. Pour jouer d'autres types de musique, différents calques ont été réalisés.

La figure 2.5 montre un exemple de calque utilisée pour le jeu du Raga Yaman d'Inde du Nord. Sur cette représentation, seules les notes de la gamme utilisée sont inscrites, selon les termes indiens, soit *Sa*, *Re*, *Ga*, *Ma'*, *Pa*, *Dha* et *Ni*. Le mode utilisé est le mode *Kaylan* qui se rapproche du mode *Lydien* en musique occidentale (mode de Fa). De plus, il s'agit d'une gamme non tempérée. La tierce *Sa*-*Ga* est pure, c'est-à-dire qu'elle possède un rapport de fréquence de 5/4, soit un intervalle de 386 centièmes de demi-tons (cents), inférieur à un intervalle de tierce majeure tempérée (400 cents). Cela se traduit par un *Ga* plus proche du *Re* que du *Ma'* (Ma#) sur le calque.

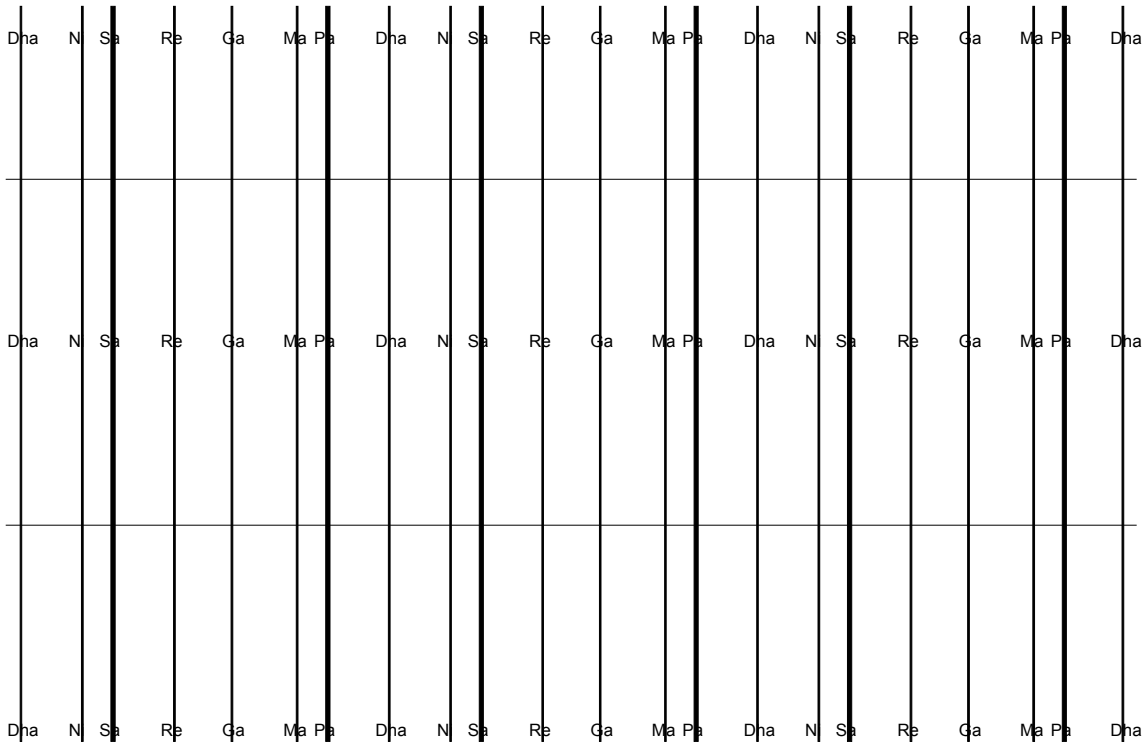

FIGURE 2.5 – *Calque appliqué à la surface de la tablette pour le jeu du Raga Yaman, réalisé par Boris Doval.*

Le choix du contrôle du vibrato, du timbre et de l'effort vocal peut sembler paradoxal puisque les acquisitions des positions linéaires sont associées aux fonctions dynamiques relatives, et l'acquisition de force isométrique à la fonction dynamique absolue. Toutefois, il a été montré par Wanderley *et al.* [MW06] que lorsque la hauteur et le vibrato étaient tous deux contrôlés, il était plus aisé d'utiliser le même capteur. De plus, le contrôle du vibrato par la position du stylet permet un contrôle total de ses paramètres (fréquence, amplitude). L'effort vocal a pendant un temps été contrôlé par la position verticale du stylet sur la tablette. Il s'avère que ce contrôle, bien que respectant la classification de Vertegaal, est moins intuitif que l'utilisation de la pression, elle-même étant un effort. Enfin, il est probable que le contrôle du timbre soit mieux adapté à une acquisition de force comme c'est le cas chez beaucoup d'instruments utilisant des accéléromètres par exemple. Un tel contrôle sur le *Cantor Digitalis* nécessiterait l'ajout de capteurs supplémentaires à la tablette, ce qui rendrait plus difficile la diffusion de l'instrument. C'est pourquoi la position verticale du stylet a été choisie. Plusieurs contrôles du timbre peuvent être sélectionnés, comme la taille du conduit vocal (plus grand vers le bas et plus petit vers le haut), ou l'ajout de tension/souffle (tension vers le bas et souffle vers le haut).

Interface visuelle

Le timbre ne peut être modifié que finement par la position du stylet. De plus, le choix de la tessiture, du taux de rugosité (jitter, shimmer) n'est pas possible sur la tablette. Le logiciel propose donc une interface visuelle sur l'ordinateur, sur laquelle le musicien peut paramétrer son instrument avant de jouer.

Cinq paramètres de qualité de voix sont modifiables : la tessiture (cinq gammes de trois octaves) ; la taille du conduit vocal, par sélection un facteur variant de 0.5 à 2.2 ; la tension ; la quantité de bruit de souffle ; le taux de rugosité (jitter+shimmer). Tous les paramètres sauf la tessiture sont continus. Il est ensuite possible de sélectionner le paramètre de timbre qui sera contrôlé plus finement sur la tablette. Pour faciliter le choix de voix prédéfinies telles que soprano, alto, ténor, basse, des réglages par défaut sont proposés, associés à ces cinq paramètres. Plus de détails sur l'interface visuelle sont donnés en section 2.3.

2.2.3 Mapping

Les paramètres de contrôle proposés par la tablette graphique et l'interface visuelle sont mis en relation avec les paramètres du synthétiseur suivant plusieurs couches de mapping. La figure 2.6 résume l'évolution des paramètres dans l'instrument.

La mise en relation des paramètres s'effectue par un mapping à trois couches, tel que proposé par [HW02]. Il s'agit de la représentation de la figure 1.10 où les paramètres gestuels et musicaux sont confondus. L'encadré supérieur de la figure 2.6 (en bleu) contient les différents modes de contrôle de l'instrument, que sont la tablette graphique et l'interface visuelle. Ceux-ci envoient alors les paramètres de contrôle vers le deuxième encadré en vert.

Ce dernier correspond à la première mise en relation des paramètres, transformant les paramètres de contrôle en paramètres musicaux aussi appelés paramètres de haut niveau. On remarque que les paramètres de source (à gauche) et les paramètres d'articulation (à droite) sont traités séparément. Plus de détails sur les mappings intervenant à cette étape sont donnés dans la documentation technique du *Cantor Digitalis*[5].

Les paramètres musicaux ou de haut niveau sont ensuite transformés en paramètres bas niveau pour le synthétiseur par le troisième encadré, en orange. Encore une fois, les paramètres de source sont traités d'un côté, calculant le quotient d'ouverture O_q, le coefficient d'asymétrie α_m ou la pente spectrale. Parallèlement, les fréquences F_i, amplitudes A_i et bandes passantes B_i des 5 formants ($i \in [1, 5]$) sont calculées. Ensuite, ces valeurs sont transformées par le module d'interaction avec la source, prenant aussi en entrée la hauteur et l'effort vocal.

Enfin, ces paramètres bas niveau sont utilisés pour synthétiser le signal vocal, comme montré dans l'encadré inférieur (en rouge). La synthèse commence par le calcul du signal de source par le module RT-CALM, suivi de l'ajout d'une pente spectrale, d'un seuil de phonation et d'une modulation par le bruit de souffle. Le signal de source est ensuite envoyé vers un banc de filtres résonants en parallèle, commandés par les grandeurs F_i, A_i et B_i. Finalement, le signal résultant est filtré coupe-bande pour la modélisation des sinus piriformes.

5. https://cantordigitalis.limsi.fr/download_en.php (vérifié le 22 octobre 2015)

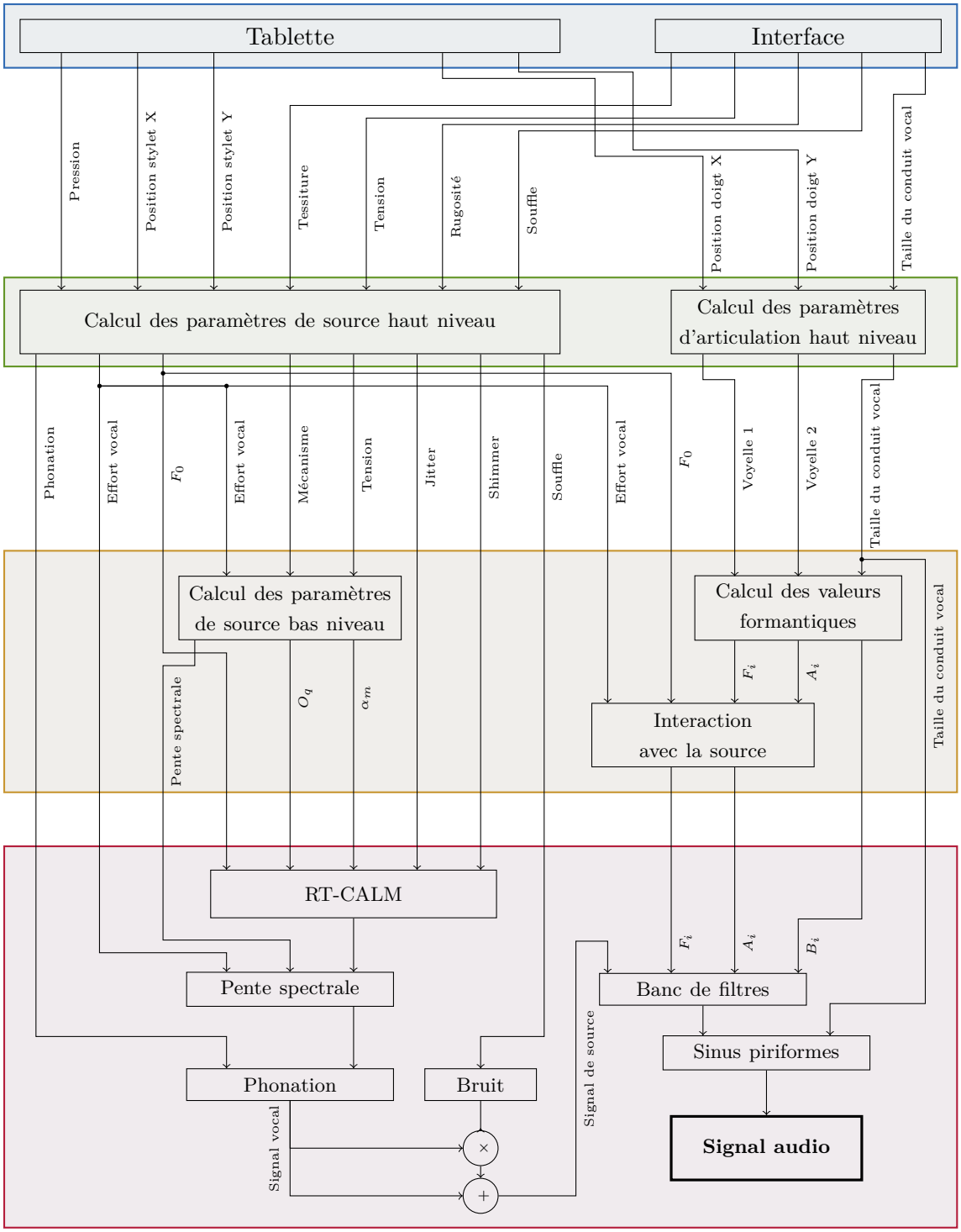

FIGURE 2.6 – *Schéma d'évolution des paramètres du Cantor Digitalis.*

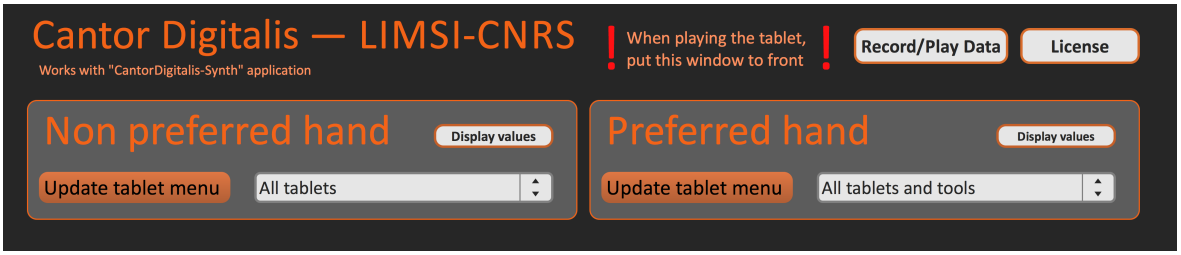

Figure 2.7 – *Interface de l'application CantorDigitalis_Tab.*

2.3 Présentation et diffusion du logiciel

L'usage intensif de notre instrument lors des répétitions du *Chorus Digitalis* nous a permis de corriger de nombreuses erreurs d'implémentation et d'aboutir à une version relativement stable du logiciel. La participation au concours international du Logiciel Musical (Lomus) organisé par l'Association Francophone d'Informatique Musicale (AFIM) en Mai 2014 a été l'occasion de produire la première version propre de notre logiciel, destinée à être partagée en dehors de notre équipe de recherche. La création d'une version finie du logiciel s'est faite en trois étapes : la mise en place d'un code source propre et accessible, l'écriture d'une documentation, et le choix des moyens de diffusion.

2.3.1 Mise en forme du logiciel

Pour des soucis de coûts de calcul, le *Cantor Digitalis* se présente sous la forme de deux logiciels, le *CantorDigitalis_Tab* et le *CantorDigitalis_Synth*. Le premier est chargé de récupérer les données de la tablette graphique et le deuxième contient le moteur de synthèse et la mise en correspondance des paramètres de tablette et de synthèse.

Récupération des données de la tablette - *CantorDigitalis_Tab*

L'application *CantorDigitalis_Tab* récupère les données de la tablette graphique de manière formatée à l'aide des objets *s2m.wacom* et *s2m.wacomtouch* développés au LMA [6]. L'interface (figure 2.7) montre qu'il est possible de choisir la ou les tablettes graphiques à sélectionner pour le contrôle de la main préférée (celle qui tient le stylet) ou pour l'autre main. Une sous-fenêtre donne accès à un programme d'enregistrement des données de la tablette. Il est aussi possible de lire des flux de tablettes déjà enregistrés. Les données formatées de la tablette sont envoyées à l'application *CantorDigitalis_Synth* contenant le moteur de synthèse par protocole UDP sur le réseau local de l'ordinateur.

Moteur de synthèse - *CantorDigitalis_Synth*

L'application *CantorDigitalis_Synth* inclut les différents étages de mise en correspondance des paramètres et le moteur de synthèse. Son interface principale est présentée en figure 2.8. Celle-ci est segmentée en quatre zones numérotées permettant de guider l'utilisateur de manière efficace dans le réglage de son instrument. La première zone permet de choisir le type de contrôle pour l'instrument. Le mode par défaut rectangle vocalique (*Vocalic Rectangle*) nécessite l'utilisation d'une tablette graphique *Touch* pour le contrôle des voyelles au doigt de la main non habile dans le cadre rouge en haut à gauche de la tablette (figure 2.4). Dans le

6. http://www.maxobjects.com/?v=libraries&id_library=163 (vérifié le 22 octobre 2015)

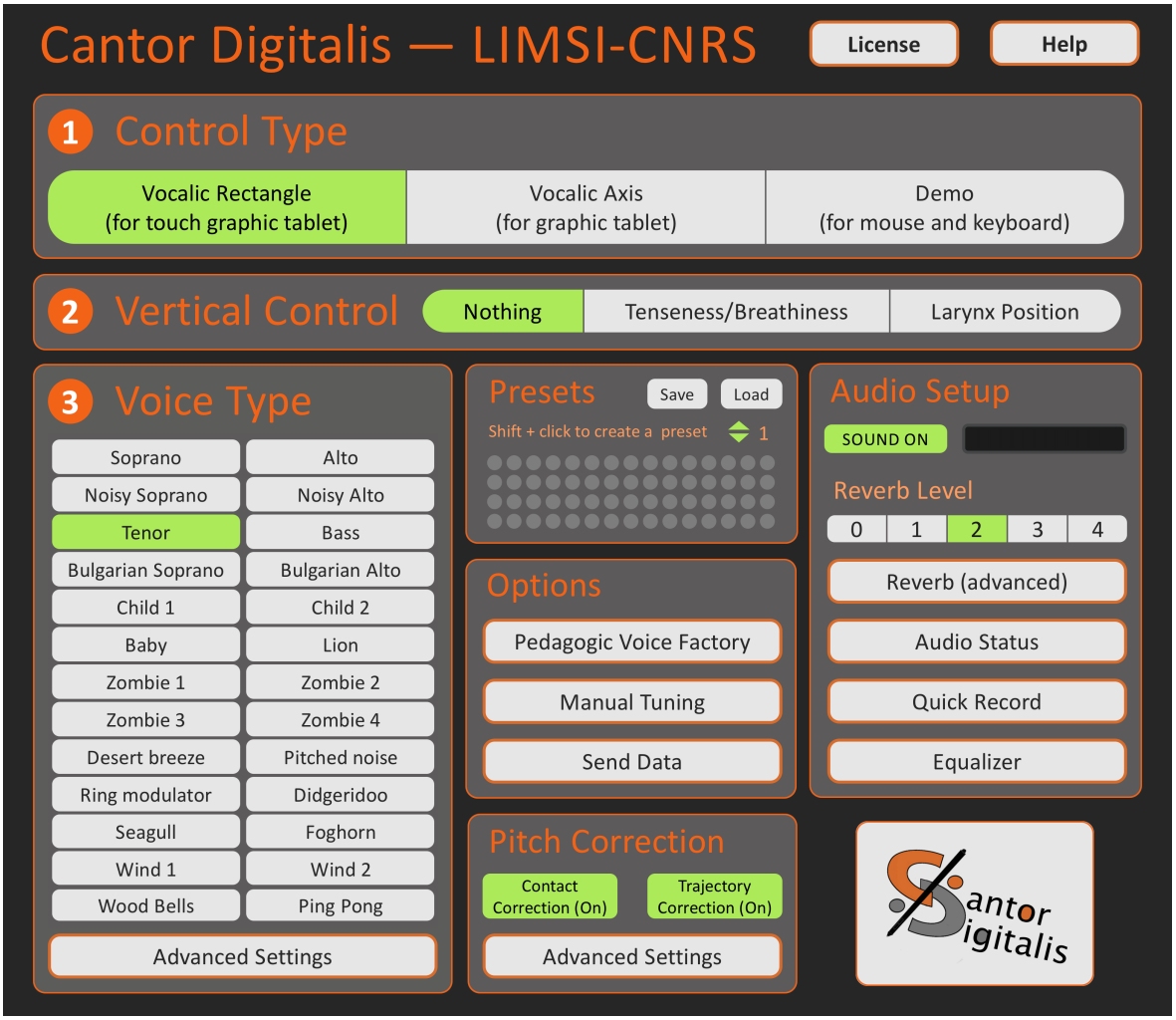

FIGURE 2.8 – *Interface de l'application CantorDigitalis_Synth.*

cas où l'utilisateur ne disposerait pas d'une tablette graphique tactile, le mode axe vocalique (*Vocalic axis*) est proposé et permet le contrôle des voyelles suivant la position verticale du stylet sur la tablette. Les lignes horizontales sur le calque indiquent les positions de chaque voyelle. Leurs noms respectifs sont écrits en vert sur les bords gauche et droit de la tablette (figure 2.4). Enfin, à titre de démonstration, le dernier mode propose un contrôle à la souris, permettant à des utilisateurs non détenteurs d'une tablette d'avoir un aperçu de l'instrument. Néanmoins, le contrôle grossier proposé par la souris comparé aux fines modulations de position possibles par un stylet entraîne un rendu sonore nettement inférieur.

La deuxième zone permet d'étendre le contrôle de l'instrument aux paramètres de source, sous réserve que le mode *Vocalic rectangle* ait été choisi précédemment. L'utilisateur peut alors effectuer des variations fines de la tension et de la quantité de souffle, ou de la position du larynx avec la position verticale du stylet sur la tablette.

La troisième zone permet ensuite de sélectionner la voix de l'instrument. 26 pré-réglages sont proposés, allant de voix réelles aux sons non-vocaux explorés pour le concours Guthman d'instruments de musique. Une sous-fenêtre montrée en figure 2.9 permet à l'utilisateur de configurer lui-même sa voix en choisissant finement les paramètres de source.

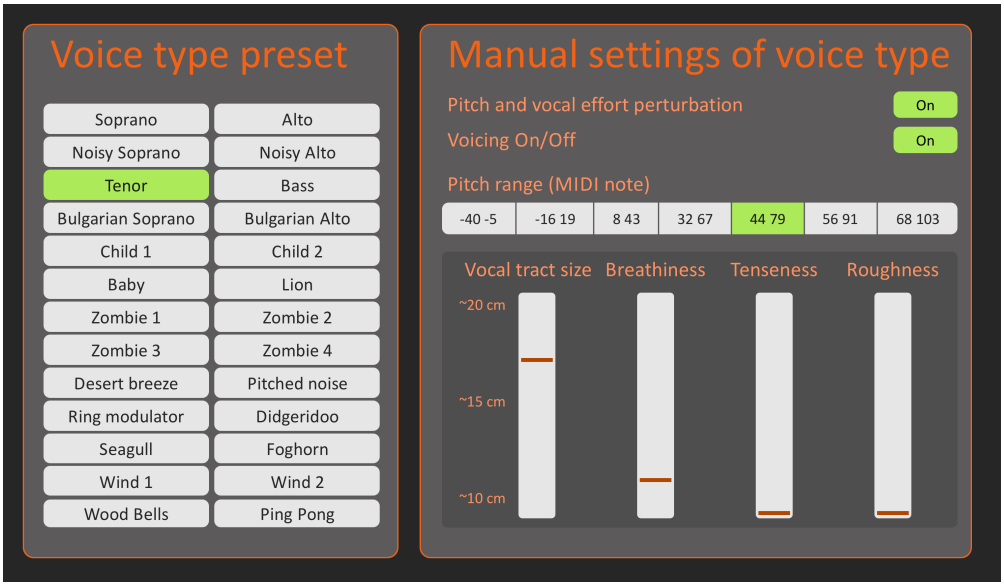

Figure 2.9 – *Panneau de contrôle pour le choix des paramètres de source.*

Les encarts restants constituent la quatrième zone et permettent le réglage d'options plus avancées dans le contrôle de l'instrument. L'encart *Audio Setup* permet de choisir le taux de réverbération de la sortie audio, ainsi que les canaux de sortie à utiliser.

L'encart *Options* propose trois fonctions. D'abord l'ouverture du programme de déconstruction de l'appareil vocal (*Pedagogic Voice Factory*) propose une décomposition du modèle source-filtre, permettant l'écoute et la visualisation du signal de source, en sélectionnant la quantité de formants à introduire petit à petit [Fd13]. La fonction *Manual Tuning* permet de choisir précisément les notes aux extrémités de la tablette (par défaut, 3 octaves) ainsi que d'accorder la note de référence (par défaut, la4 = 440 Hz). Enfin, la fonction *Send Data* permet l'envoi des données du synthétiseur vers le programme de visualisation des données présenté en section 7.2.

L'encart *Pitch Correction* propose l'activation de la correction automatique de justesse présentée au chapitre 4. Une sous-fenêtre permet de régler finement les paramètres de correction (figure 2.10).

Enfin, l'encart *Presets* permet de sauvegarder différents réglages effectués sur la fenêtre principale. Celui-ci est utilisé à chaque concert, lorsque pour chaque morceau le musicien change de voix et de réglages associés.

A l'ouverture de l'application, des réglages par défaut sont proposés, indiqués en vert sur la figure 2.8, et permettent de jouer immédiatement de l'instrument.

La présentation du code

Le *Cantor Digitalis* est implémenté sous Max/MSP [7] version 6. Afin de faciliter l'exploration du code et sa manipulation, ce dernier a été segmenté en sous-programmes montrés en figure 2.11. Cette segmentation repose sur les différentes étapes de la chaîne d'évolution des paramètres montrées en figure 2.6. En haut de la hiérarchie se trouvent quatre modules

7. https://cycling74.com (vérifié le 22 octobre 2015)

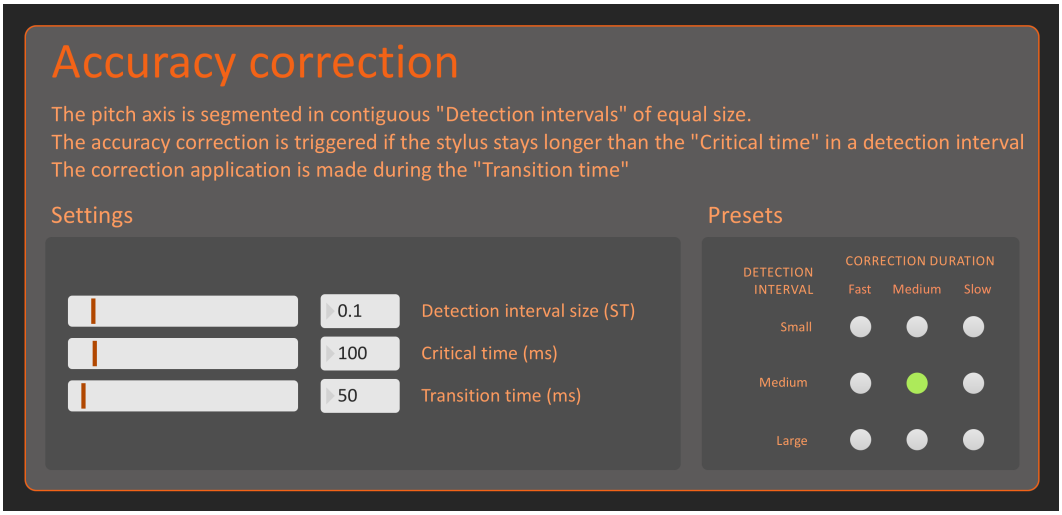

FIGURE 2.10 – *Panneau de contrôle pour le choix des paramètres de correction de justesse.*

constituant les différentes possibilités de contrôle de l'instrument. Le programme principal en noir (*Main patcher*) regroupe les éléments de l'interface de l'instrument (figure 2.8). Le programme de contrôle (*Control*) récupère les données de la tablette envoyées par *CantorDigitalis_Tab* en temps-réel. Le programme d'individualisation des voix *Voice Individualization* permet le choix des paramètres de qualité de voix (figure 2.9). Le programme *Voice Factory* propose de décomposer le modèle vocal.

A un deuxième niveau, les programmes *Glottis Mapping HL* et *Vowel Mapping* transforment les paramètres de contrôle en paramètres gestuels/musicaux de la source et du conduit vocal respectivement. Les programmes *Glottis LL*, *Vowel Rules* et *Source Filter Dependencies* transforment ensuite les paramètres gestuels/musicaux en paramètres pour le synthétiseur.

Enfin, les programmes *Glottis* et *Vocal Tract* synthétisent le signal de source et final respectivement. Le programme *Audio* s'occupe de la diffusion du signal synthétisé.

Cette segmentation du code s'avère très utile pour sa manipulation par des utilisateurs extérieurs. Par exemple, deux adaptations de l'instrument ont été effectuées pour un contrôle par le *Continuum Fingerboard* et le *Soundplane*[8]. Les développeurs ont simplement remplacé le module *Control* chargé de la récupération des données tablette par un module récupérant les données de leurs contrôleurs, en respectant le format des données de sortie explicitement fourni dans la documentation technique. Les deux contrôleurs sont des surfaces continues sensibles à la pression. Le jeu se fait aux doigts comme sur un piano. Les vidéos de démonstration faites par ces développeurs montrent des attaques plus percussives, se rapprochant de consonnes bien que le moteur de synthèse ne produise que des voyelles. On peut expliquer ce phénomène par deux caractéristiques de ces interfaces. D'abord, celles-ci sont étroites en largeur. Par conséquent, lorsque le déplacement du larynx est contrôlé suivant la position verticale du doigt, un petit déplacement entraîne une grande variation des formants. Combiné avec différentes sensibilités de pression, par leurs capteurs mais aussi par leurs matériaux (le *Continuum* est en mousse et le *Soundplane* est en bois), on obtient des attaques percussives. Ces appropriations du moteur de synthèse démontrent que les possibilités sonores de l'instrument ne sont pas propres au moteur de synthèse, mais dépendent aussi de l'interface.

8. https://cantordigitalis.limsi.fr/use_fr.php (vérifié le 22 octobre 2015)

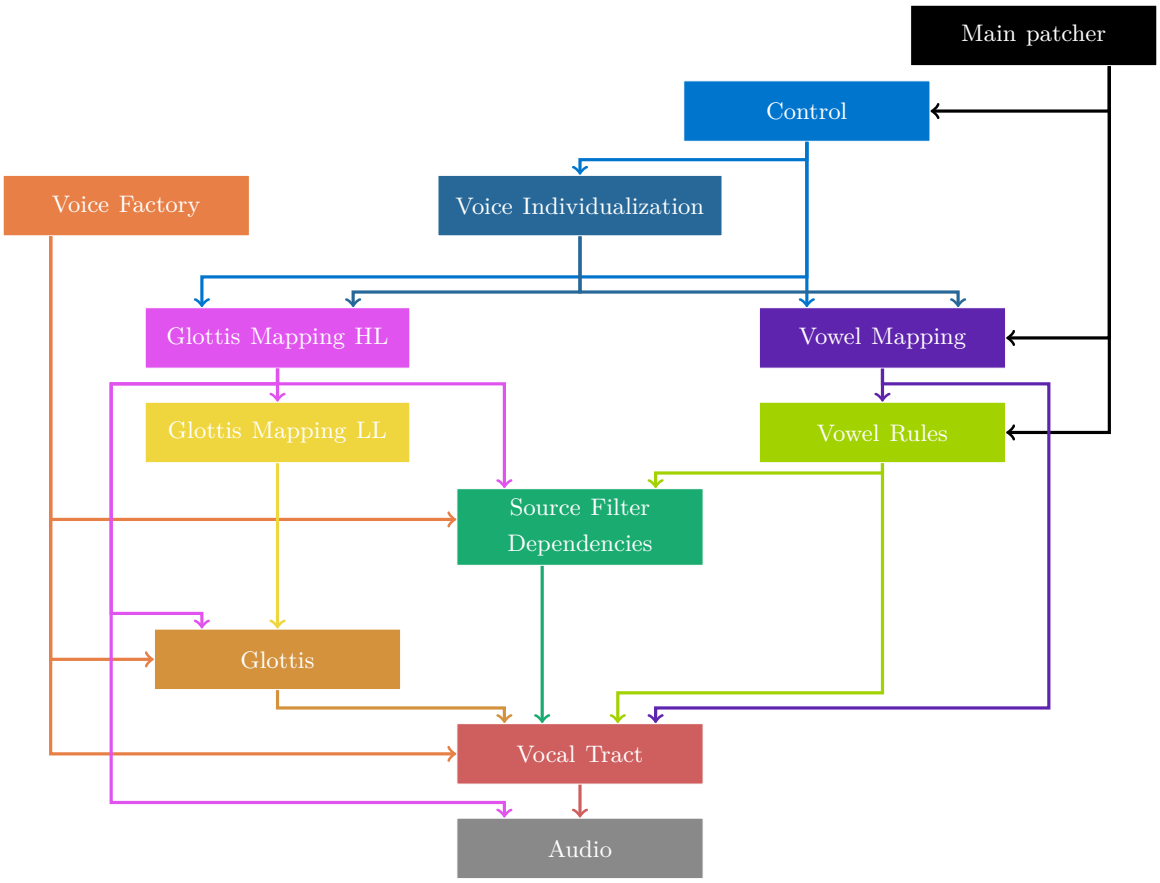

Figure 2.11 – *Schéma de la hiérarchie du code.*

2.3.2 La documentation

Toute diffusion de produits nécessite une documentation associée décrivant son fonctionnement aux nouveaux utilisateurs. Deux documentations ont été rédigées. La première, technique, cible un public de développeurs désireux d'explorer le code source. La deuxième, sous forme de manuel, explique les possibilités de manipulation de l'instrument [9].

Documentation technique

La documentation technique a pour but de présenter la structure du programme (figure 2.11), ainsi que l'échange et l'évolution des données dans le programme. Sous Max/MSP, deux modes de communication existent pour l'échange de données entre objets. La méthode la plus simple est de relier ces deux objets par un câble. Pour éviter la surcharge visuelle, ou lorsque les objets sont éloignés dans le programme, il est possible d'utiliser un couple d'objet *send/receive* qui permet de communiquer à distance. Bien que pratique, une utilisation importante de cette solution rend difficile la visualisation des communications entre divers modules du programme. C'est pourquoi toutes les entrées et les sorties de chaque sous-programme sont explicitées, en indiquant les provenances et les directions de chacune. Ensuite, l'évolution des données au sein de chaque module est présentée sous forme de schéma bloc.

9. https://perso.limsi.fr/operrotin/these.fr.php#Publications (vérifié le 22 octobre 2015)

Manuel utilisateur

Pour aider à la prise en main de l'instrument, un manuel a été rédigé en français et en anglais, décrivant toutes les actions à accomplir, de l'installation du logiciel à la préparation d'un concert pour le jeu du *Cantor Digitalis*.

2.3.3 Diffusion

Choix de la licence

Une étape importante dans la diffusion du logiciel a été de discuter des droits que l'on souhaitait accorder à l'utilisateur. Dans une optique de recherche plutôt que commerciale, nous avons opté pour une diffusion libre de l'instrument, sous la licence CeCILL élaborée conjointement par le CEA, le CNRS et l'INRIA, se rapprochant de la licence américaine GPL. Celle-ci permet l'utilisation, la modification et la redistribution du programme par les utilisateurs.

Le paquetage

Le logiciel est téléchargeable depuis la page https://cantordigitalis.limsi.fr/download_en.php sous forme d'un paquet. Ce dernier contient :

- Les applications *CantorDigitalis_Synth* et *CantorDigitalis_Tab*. Celles-ci sont compatibles uniquement sur Mac OS X version 10.6 ou ultérieure.
- Les sources. On y trouve les codes Max/MSP de l'instrument, ainsi que les objets *s2m.wacom* et *rtcalm* écrits en C, permettant respectivement de récupérer les données de la tablette et de calculer le modèle de source.
- La documentation technique et les manuels utilisateurs français et anglais.
- Une version PDF du calque pour une gamme tempérée (figure 2.4) proposée en format M et L des tablettes Intuos. Le fichier source ODS (Open Office) est aussi fourni.
- L'ensemble des publications associées aux travaux de recherches sur le *Cantor Digitalis*.

Les outils de communication

La diffusion du logiciel se fait principalement depuis le site web https://cantordigitalis.limsi.fr/ que nous avons développé. Ce site contient une description de la recherche effectuée sur l'instrument ainsi que la liste des publications associées, une page consacrée au téléchargement du logiciel, une présentation du *Chorus Digitalis*, ensemble de *Cantor Digitalis* (voir chapitre 7) ainsi que des vidéos associées, une présentation des divers projets entrepris par des développeurs extérieurs à notre équipe, et un recensement des différents articles de presse parlant de l'instrument.

Parallèlement, nous avons créé deux listes de diffusion sur lesquelles les utilisateurs sont invités à s'inscrire. La première, *cantordigitalis.news@limsi.fr* permet de recevoir les informations d'actualité sur l'instrument, telles que les nouvelles mises à jour ou les concerts du *Chorus Digitalis*. La deuxième, *cantordigitalis.forum@limsi.fr* permet aux utilisateurs de discuter autour de l'instrument. 41 utilisateurs sont inscrits sur la première et 20 sur la deuxième. En pratique, très peu de messages ont été envoyés sur la liste forum.

2.3.4 Les versions

A ce jour, deux versions ont été diffusées, et une troisième est actuellement en cours de préparation. La première (v1.0) a été réalisée pour le concours Lomus en Avril 2014. Peu de communication a été effectuée autour de celle-ci, s'agissant plus d'une version beta que d'une version officielle. A l'occasion de l'inscription au concours Guthman d'instruments de musique, en novembre 2014, une deuxième version (v1.1) a été diffusée et une notification a été envoyée à de nombreuses listes de diffusion de la communauté. Un travail conséquent d'amélioration de l'interface graphique a été réalisé, ainsi que la correction d'un grand nombre de bogues restants. Le mode de contrôle par la souris a aussi été rajouté. Enfin, pour la participation au concours ainsi qu'au concert suivant à Metz, nous avons préparé une troisième version (v1.2) comportant entre autres de nouvelles voix, une amélioration de la source par un ajout de hautes fréquences, et toujours une correction de bogues restants. Les interfaces présentées par les figures de cette annexe proviennent de la version (v1.2), mais cette dernière n'a pas encore été diffusée au grand public.

2.4 Conclusion

Le *Cantor Digitalis* présenté ici est l'aboutissement d'une dizaine d'années de recherche. Le logiciel associé est disponible sous licence libre depuis avril 2014. La plupart des efforts de conception se sont concentrés sur l'élaboration du moteur de synthèse. En revanche, peu de travaux ont été effectués sur l'évaluation de l'interface associée pour son contrôle. C'est dans ce cadre que s'inscrivent les chapitres suivants.

Chapitre 3

Justesse et précision de l'intonation vocale et chironomique

Sommaire

3.1 Introduction

Le contrôle de la hauteur vocale dépend à la fois des capacités du musicien et de l'interface qui lui est proposée. L'importante résolution spatiale de la tablette graphique introduite en section 2.2.2 permet d'atteindre une résolution de hauteur de 0.004 centièmes de demi-tons (cents), valeur très inférieure au seuil de perception de hauteur d'environ 4 cents [Moo73]. L'interface n'impose donc pas de limites théoriques à la justesse de jeu. La question posée dans ce chapitre est donc celle des capacités d'un musicien à utiliser la tablette graphique pour jouer une mélodie, ou en d'autres termes, l'adéquation de l'interface pour une telle tâche.

Il a été montré précédemment qu'il était possible d'imiter très fidèlement les contours intonatifs de phrases données dans un contexte d'imitation de parole [dRLB11]. Toutefois, ce dernier est moins restrictif que l'imitation du chant. En effet, la parole ne demande qu'une reproduction de hauteur relative, et permet une déviation de justesse jusqu'à 1 demi-ton, sans altérer la qualité de l'imitation, contre 5 à 10 cents pour le chant sans vibrato. De plus, une contrainte temporelle est imposée au chant par un tempo. Ces exigences musicales rendent la tâche d'imitation du chant plus difficile et une nouvelle étude est par conséquent nécessaire.

De nombreuses études ont été réalisées sur les performances d'imitations mélodiques vocales en mesurant la justesse d'intervalle, c'est-à-dire l'erreur moyenne entre les intervalles chantés et les intervalles cibles sur un extrait musical. Il a été montré que la majorité de la population est capable de chanter juste des mélodies célèbres, en particulier lorsqu'un tempo lent est imposé, permettant d'atteindre des justesses similaires à des chanteurs professionnels (moyenne : 30 cents ; écart-type : 0 cent) [DBGP07], à l'exception d'une minorité de personnes qualifiées de *poor-pitch singers* en anglais (10-16% de la population), dont une mauvaise correspondance entre perception et production moteur entraînerait des transpositions et des compressions d'intervalles systématiques [PB07]. Des mesures de justesse similaires ont été reportées dans différentes études s'intéressant à l'influence de la technique vocale [LMM12] (moyenne : 50.83 cents ; écart-type : 3.29 cents), [LMMM13] et [LMMM14] (moyenne : 28.7 cents ; écart-type : 3.36 cents). De plus il a été montré que ces mesures de justesse d'intervalles sont fortement corrélées avec la perception de juges experts [LMLS+13].

Néanmoins, la mesure de justesse d'intervalles prend en compte uniquement des moyennes. Elle donne une tendance globale de justesse mais ne rend pas compte de la constance du chanteur dans ses erreurs. Une mesure d'écart-type [TS88] appelée précision permet de combler cette lacune. Un travail de Pfordresher *et al.* comparant les différents types de mesures de justesse et précision [PBM+10] montre qu'en dépit des bonnes justesses des chanteurs, ceux-ci sont peu précis. Autrement dit, les erreurs moyennes observées sont faibles, mais la variabilité des erreurs est grande. De plus, une relation de causalité est mise en évidence entre justesse et précision : un chanteur non juste est très probablement imprécis. Enfin, au regard des faibles opinions que la population se fait d'elle même sur sa capacité à chanter juste [PB07], on peut émettre l'hypothèse que cette auto-évaluation est corrélée aux mesures de précision.

Afin d'évaluer les performances de justesse et précision du contrôle chironomique de la synthèse vocale, une expérience d'imitation de mélodies a été conduite précédemment au sein du laboratoire mais les données n'ont pas été exploitées. Ce chapitre présente l'analyse des résultats obtenus. Des mesures de justesse et précision sont réalisées conjointement sur des imitations vocales et chironomiques afin de les comparer [1]. Le protocole réalisé précédemment est décrit en section 3.2. Les résultats sont présentés en section 3.3 et discutés en section 3.4.

1. L'ensemble de ces travaux est paru dans [dFLB+14].

3.2 Expérience

Le but de cette étude est de comparer justesse et précision des intonations vocale et chironomique par expérience d'imitation. Des extraits musicaux courts sont présentés aux sujets. Il leur est demandé de les reproduire soit à la voix (imitation vocale), soit à l'aide de la tablette graphique (imitation chironomique). Pour aller plus loin, afin de tester l'influence de la présence d'une aide visuelle sur la tablette, une imitation chironomique sans retour audio est aussi requise. Il en résulte trois conditions d'imitation : une imitation *Vocale*, une imitation *Chironomique*, et une imitation *Chironomique muette*.

3.2.1 Matériel

L'expérience s'est déroulée dans une cabine traitée acoustiquement et insonorisée. Les extraits musicaux à imiter sont joués par un synthétiseur MIDI (Instrument *choir Aah 2* du logiciel *SimpleSynth*[2]) produisant un son de chœur synthétique et émis par le biais d'une carte son RME Fireface 400 et d'un casque DTX900 Beyerdynamic. Une partition comportant le nom des notes est affichée sur un écran placé devant les sujets. Les imitations vocales sont enregistrées par un microphone DPA 4006-TL. Les imitations chironomiques sont réalisées à l'aide du *Cantor Digitalis*. Les contrôles possibles pour cette expérience sont cependant restreints à celui de la hauteur suivant l'axe horizontal du stylet, et au contrôle de l'effort vocal par la pression du stylet. Bien que ce dernier ne soit pas exploité dans l'expérience, il ajoute une dimension plus réaliste à la voix de synthèse. L'axe vertical du stylet est laissé libre pour contraindre au minimum les gestes des sujets. La tablette utilisée est une tablette graphique Wacom Intuos 3 Large de surface active 228.6 × 304.8 mm, présentant 1024 niveaux de pression du stylet, une résolution temporelle de 200 Hz et une résolution spatiale de 0.25 mm. Contrairement à la configuration originale du *Cantor Digitalis*, le calque contenant les indices visuels ne comporte qu'une octave (figure 3.1). Seuls les noms des notes nécessaires à l'expérience sont indiqués (Do, Ré, Mi, Fa, Sol, La Si et Do), et chaque demi-ton est séparé de 1.4 cm. La résolution de hauteur résultante est de 1.7 cents, légèrement inférieure au seuil de perception. Deux tessitures sont proposées : de 125 à 250 Hz pour les sujets hommes et de 250 à 500 Hz pour les sujets femmes. Le son produit par les imitations chironomiques provient du moteur de synthèse du *Cantor Digitalis*, fixé sur une voyelle /a/. Un métronome indique le tempo de manière visuelle (sur l'écran) et auditive.

3.2.2 Stimuli

L'expérience est découpée en trois blocs, chacun présentant un motif musical différent. Le tableau 3.1 résume les modalités de chaque bloc.

Bloc 1 : Intervalles

Les motifs de ce premier bloc sont des intervalles de deux notes, ascendants et descendants, extraits d'une gamme de Do majeur. Les intervalles de septièmes sont retirés car plus difficiles à jouer. Ces motifs sont présentés en haut de la figure 3.2, chaque mesure représentant un stimulus. Le tempo imposé aux sujets pour ce bloc est de 120 battements par minute (b.p.m.). Les motifs étant très courts et facile à mémoriser, les sujets ne peuvent les écouter qu'une fois. Chaque motif est imité trois fois dans chacune des trois conditions (*vocale*, *chironomique*,

2. http://simplesynth.sourceforge.net (vérifié le 22 octobre 2015)

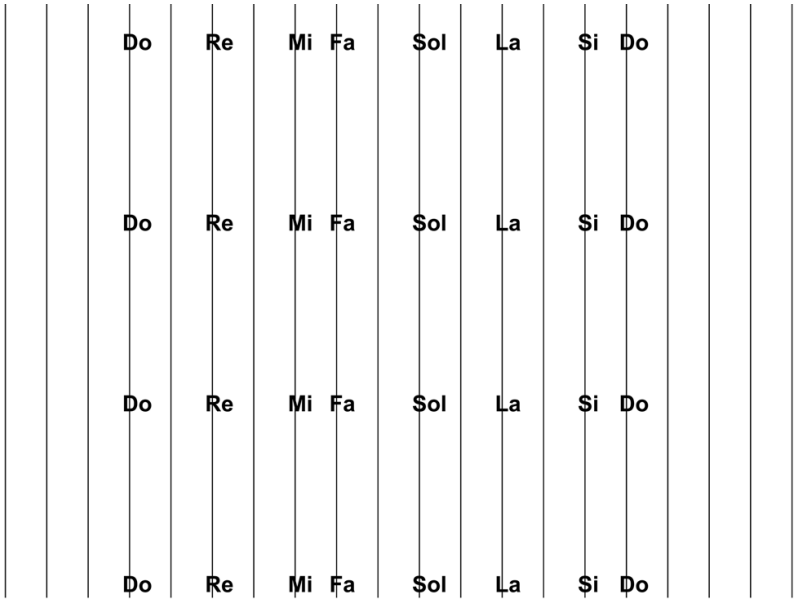

Figure 3.1 – *Calque appliqué sur la tablette pour les modalités chironomiques.*

	Bloc 1	**Bloc 2**	**Bloc 3**
Motif	Intervalles (2 notes)	Mélodies (6-7 notes)	Doubles Intervalles (3 notes)
Nombre de motifs	12	5	12
Conditions	Chironomique Chironomique muette Vocale	Chironomique Chironomique muette Vocale	Chironomique
Tempo	120 b.p.m.	120 b.p.m.	120 b.p.m. 180 b.p.m. 240 b.p.m.
Nombre d'écoutes	1	≥ 1	1
Nombre d'essais	3	≥ 3	3

Table 3.1 – *Résumé des modalités de chaque bloc de l'expérience justesse et précision de l'intonation.*

chironomique muette). Seul le meilleur des trois essais est conservé pour chaque condition et chaque motif. Le bloc 1 est donc constitué de 12 motifs joués 3 fois sous 3 conditions.

Bloc 2 : Mélodies

Le bloc 2 est similaire au bloc 1 mais avec des motifs mélodiques plus longs : 5 mélodies de 6 à 7 notes composées d'intervalles présentés dans le bloc 1 (milieu de la figure 3.2). Les mélodies étant plus difficiles à mémoriser, le nombre d'écoutes n'est pas limité. De plus, comme il est parfois difficile de chanter des mélodies inconnues, un nombre illimité d'essais peut être enregistré pour chaque mélodie, avec un minimum de trois. Le meilleur des essais est conservé. Chaque mélodie est imitée dans les trois conditions (*vocale*, *chironomique*, *chironomique muette*) avec un tempo de 120 b.p.m. Le bloc 2 est donc constitué de 5 motifs joués au moins 3 fois sous 3 conditions.

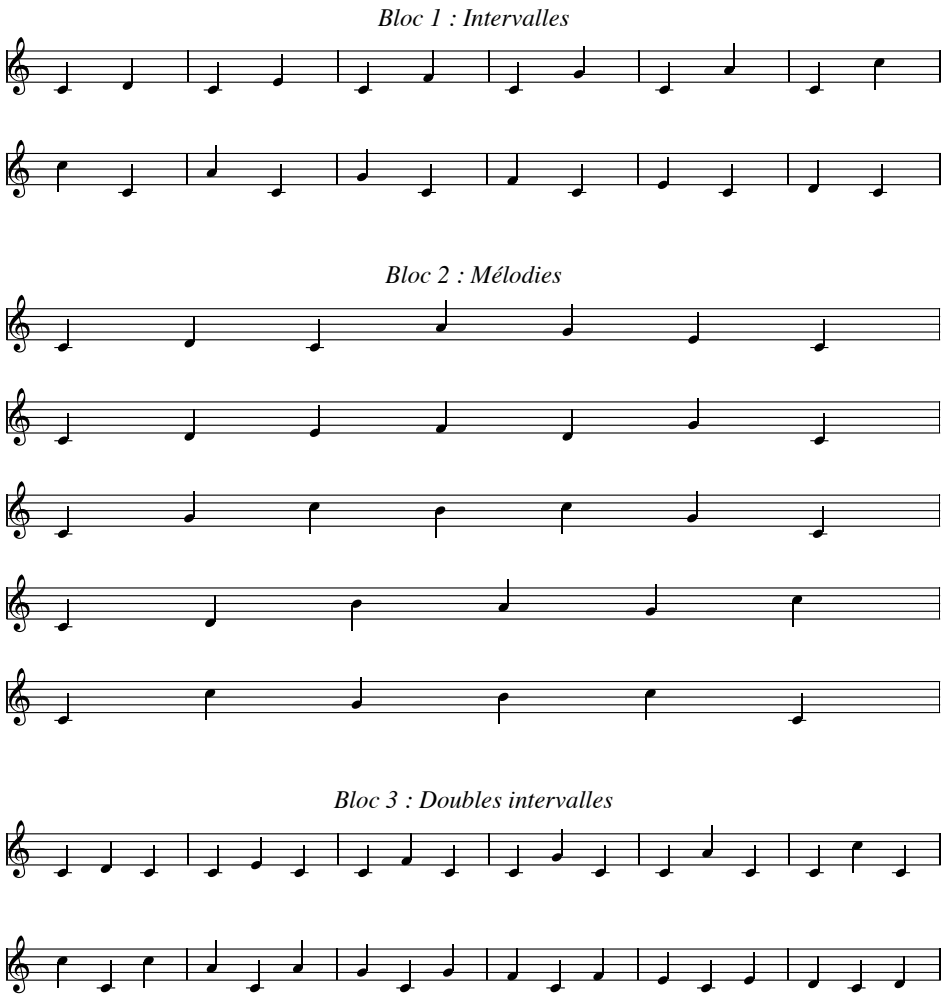

FIGURE 3.2 – *Motifs utilisés pour chaque bloc de l'expérience. Chaque mesure correspond à un motif.*

Bloc 3 : Tempo

Le but du dernier bloc est d'étudier l'influence du tempo sur l'imitation *chironomique*. Les 12 motifs présentés en bas de la figure 3.2 sont des enchaînements de 3 notes montant/descendant ou descendant/montant dont les premières et dernières notes sont identiques. Les motifs étant très courts et faciles à mémoriser, les sujets ne peuvent les écouter qu'une fois. Chaque motif est imité trois fois dans la seule condition *chironomique*, sous trois tempi différents : 120, 180 et 240 b.p.m. Seul le meilleur des trois essais est conservé pour chaque condition et chaque tempo. Le bloc 3 est donc constitué de 12 motifs joués 3 fois sous 3 tempi.

3.2.3 Tâche

Chaque bloc de stimuli est introduit séparément aux sujets. Les stimuli sont présentés dans un ordre aléatoire au sein d'un bloc. Une interface dédiée a été implémentée sur Max/MSP

par Sylvain Le Beux pour contrôler le déroulement de l'expérience. Le sujet commence par presser la barre d'espace pour initier une tâche d'imitation. Le métronome et l'extrait musical sont joués et affichés à l'écran. Pour l'imitation *vocale*, le sujet doit maintenir le stylet en contact avec la tablette. Pour les imitations *chironomique* et *chironomique muette*, il est aussi demandé aux sujets de ne pas rompre le contact entre stylet et tablette pendant l'imitation d'un motif. Pour chaque condition, l'essai suivant est présenté automatiquement dès que le stylet est relevé de la tablette. Une session d'entraînement est proposée avant l'expérience afin de se familiariser avec l'interface. Afin de limiter l'effet d'apprentissage, un protocole similaire mais des motifs différents sont utilisés dans ce cas.

3.2.4 Participants

Un groupe de 20 sujets a réalisé les blocs 1 et 2 (moyenne d'âge 31 ans ; 6 femmes, 14 hommes). Dans ce groupe, 14 ont reçu un entraînement musical et/ou pratiquent régulièrement la musique. La durée moyenne de pratique musicale est de 18 ans. Aucun sujet n'a reporté de problèmes auditifs, mais 12 sujets ont évalué leurs capacités vocales comme faibles en terme de justesse ou précision. 16 sujets étaient droitiers et 4 gauchers.

Un groupe de 28 sujets a réalisé le bloc 3 (moyenne d'âge 29 ans ; 11 femmes, 17 hommes). Dans ce groupe, 18 ont reçu un entraînement musical et/ou pratiquent régulièrement la musique. La durée moyenne de pratique musicale est de 16 ans. Aucun sujet n'a reporté de problèmes auditifs, mais 15 sujets ont évalué leurs capacités vocales comme faibles en terme de justesse ou précision. 23 sujets étaient droitiers et 5 gauchers.

Tous les sujets étaient membres du laboratoire et ont participé à l'expérience sur la base du volontariat sans contrepartie financière.

3.2.5 Outils d'analyse

Extraction des données

L'extraction des notes jouées par les sujets se fait en trois étapes : obtention de la hauteur brute ; identification des notes dans la trajectoire ; extraction des valeurs de hauteur jouées sur les parties stables de la trajectoire.

La hauteur jouée est directement fournie par les données de la tablette dans les conditions *chironomie* et *chironomie muette*. Dans le cas *vocal*, la hauteur est extraite par le programme de détection de hauteur STRAIGHT [KMKdC99]. La courbe intonative est convertie en demi-tons (ST) suivant la norme MIDI : $ST = 12\log_2(Hz/440) + 69$. Des exemples de traces intonatives sont présentées en figure 3.3. Les courbes pleines représentent les traces intonatives chironomique (en haut) et vocale (en bas), et les segments en pointillés indiquent les cibles à atteindre.

Une fois la courbe intonative extraite, les transitions entre chaque note sont repérées en regardant les maxima locaux de la dérivée de la trajectoire. Le signal est alors segmenté à chaque transition afin d'isoler chaque note jouée.

L'extraction de la valeur de la note sur chaque segment de signal est réalisée à l'aide d'une procédure de stylisation semi-automatique. L'axe temporel est d'abord subdivisé en intervalles très courts de tailles égales (10 ms) et le signal est moyenné sur ces intervalles. Deux intervalles consécutifs dont la différence des valeurs de hauteur est inférieure à un certain seuil sont regroupés. Il en résulte un intervalle dont la valeur de hauteur est la moyenne des valeurs des deux intervalles. Les seuils utilisés par défaut sont 50 cents pour les imitations *vocales* et 10 cents pour les imitations *chironomique* et *chironomique muette*. Le regroupement

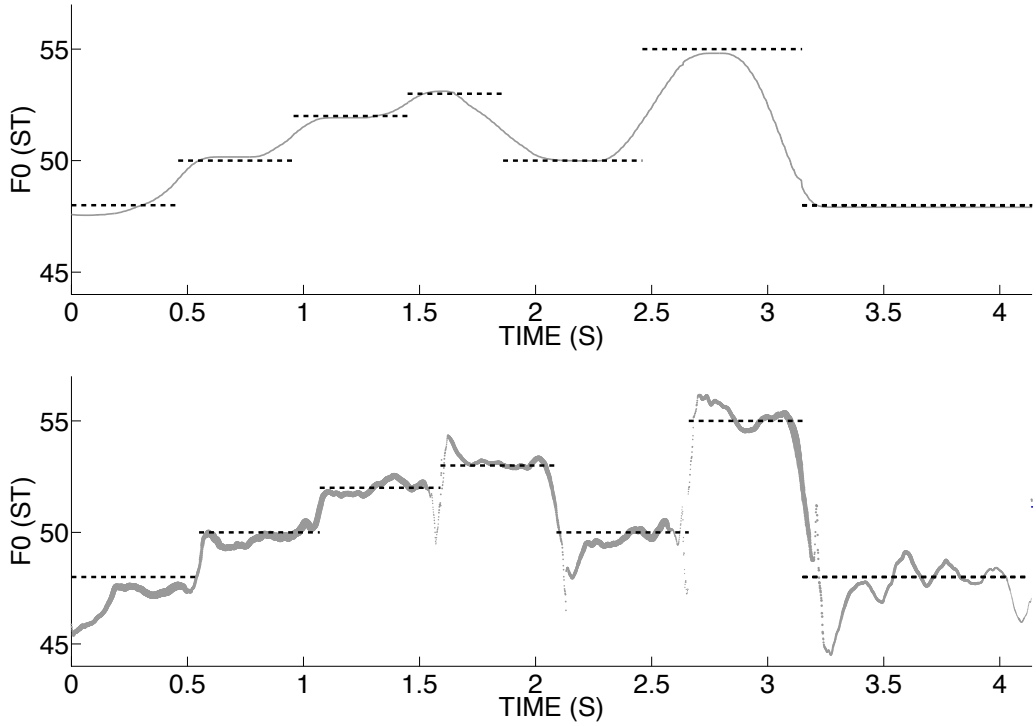

FIGURE 3.3 – *Exemples de courbes intonatives extraites d'imitations chironomique (haut) et vocale (bas). Les traits en pointillés représentent les notes cibles.*

d'intervalles est réalisé jusqu'à ce que les valeurs de tous les intervalles diffèrent d'au moins un seuil avec leurs voisins. La courbe obtenue est constituée de paliers de tailles variables. Plus un palier est long, plus la courbe intonative initiale est stable. La valeur de la note jouée sur le segment intonatif est alors la valeur du palier le plus long du segment. La figure 3.4 montre un exemple d'extraction de notes par stylisation. La trajectoire intonative vocale est représentée en gris. La courbe stylisée y est superposée en trait plein noir. Les segments pointillés indiquent les notes cibles, et les $\times$ indiquent les notes extraites. Chaque extraction a été vérifiée et les paramètres parfois ajustés manuellement en cas d'extraction erronée.

L'extraction fournit donc un ensemble de notes par stimulus à comparer avec les notes cibles. Les imitations dont le nombre de notes ne correspond pas au motif sont écartées.

Mesures

Les mesures de justesse et précision utilisées par Pfordresher *et al.* [PBM+10] sont utilisées. Sur un ensemble de N notes, en notant S_i une note jouée et T_i la note cible correspondante, on définit la justesse de notes NA comme la moyenne des erreurs observées entre notes jouées et notes cibles :

$$NA = \frac{1}{N} \sum_{i=1}^{N} (S_i - T_i) \tag{3.1}$$

Cette mesure de justesse définit la tendance d'un sujet à jouer trop haut (justesse positive) ou trop bas (justesse négative). Une valeur nulle indique que le sujet atteint en moyenne correctement les notes cibles.

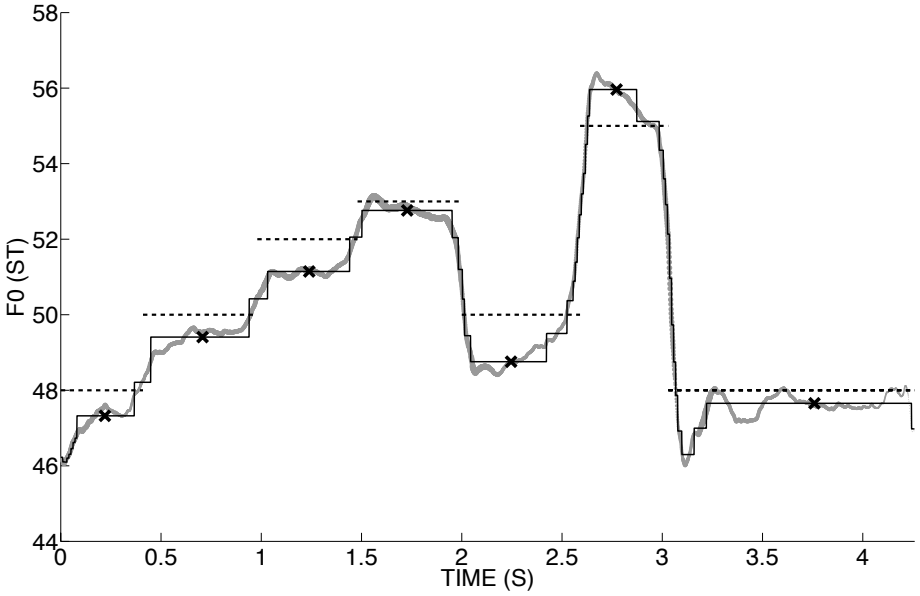

Figure 3.4 – *Estimation des notes jouées par stylisation. La courbe stylisée (noir) est superposée à la courbe intonative (gris). Les notes extraites sont marquées par des* $\times$.

On définit ensuite la justesse d'intervalles IA comme la moyenne des erreurs observées entre les intervalles joués et les intervalles cibles :

$$IA = \frac{1}{N-1} \sum_{i=2}^{N} (|S_i - S_{i-1}| - |T_i - T_{i-1}|) \tag{3.2}$$

La justesse d'intervalle définit la tendance d'un sujet à jouer des intervalles trop grands (justesse positive) ou trop petits (justesse négative). Une justesse d'intervalle nulle indique que le sujet joue en moyenne des intervalles de tailles attendues.

Soit M_{nc} la moyenne des N_{nc} notes jouées ayant la même cible. On définit la précision autour d'une note cible NP_{nc} comme l'écart-type des erreurs observées entre les notes jouées et la note cible :

$$NP_{nc} = \sqrt{\frac{1}{N_{nc}} \sum_{i=1}^{Nnc} (S_i - M_{nc})^2} \tag{3.3}$$

La précision de notes de l'ensemble NP est définie comme la moyenne des précisions autour de chaque note cible. En posant NC le nombre de notes cibles différentes dans l'ensemble on a :

$$NP = \frac{1}{NC} \sum_{nc=1}^{NC} NP_{nc} \tag{3.4}$$

La précision de note définit la tendance d'un sujet à réaliser des erreurs constantes (valeur faible ou nulle) ou des erreurs variables (valeur élevée) autour des notes cibles.

Blocs	Groupe	Facteurs	Niveaux
1 et 2	Sujets	Sujets × Conditions	20 × 3
1 et 2	Motifs	Motifs × Conditions	17 × 3
1 et 2	Intervalles	Intervalles × Conditions	16 × 3
3	Tempi	Tempi × Sujets × Motifs	3 × 28 × 3

TABLE 3.2 – *Résumé des groupes d'étude pour l'expérience justesse précision de l'intonation.*

De la même manière on définit la moyenne M_{ic} des N_{ic} intervalles joués ayant la même cible. La précision autour d'un intervalle cible IP_{ic} se définit comme l'écart-type des erreurs observées entres les intervalles joués et l'intervalle cible.

$$IP_{ic} = \sqrt{\frac{1}{N_{ic}} \sum_{i=1}^{Nic} (S_i - M_{ic})^2} \tag{3.5}$$

Enfin, la précision d'intervalles de l'ensemble IP est définie comme la moyenne des précisions autour de chaque intervalle cible. En posant IC le nombre d'intervalles cibles différents dans l'ensemble on a :

$$IP = \frac{1}{IC} \sum_{ic=1}^{IC} IP_{ic} \tag{3.6}$$

Cette mesure définit la tendance d'un sujet à effectuer des erreurs constantes dans la taille des intervalles (valeur faible ou nulle), ou des erreurs variables (valeur élevée).

Parmi l'ensemble des essais obtenus pour l'imitation d'un stimulus, le meilleur essai est défini comme celui ayant la somme des justesses de note et d'intervalle la plus faible.

3.3 Résultats

3.3.1 Regroupement des données

Les mesures de justesse et de précision sont calculées sur des ensembles de notes. Afin de mettre en évidence les effets de différents facteurs, quatre groupes d'étude sont considérés : les groupes *Sujets*, *Motifs* et *Intervalles* pour les blocs 1 et 2, et le groupe *Tempi* pour le bloc 3. Dans le groupe *Sujets*, chaque ensemble est constitué de toutes les notes jouées par un sujet dans une condition. L'interaction des facteurs "sujet" (20 niveaux) et "condition" (3 niveaux) résulte en 60 séries de mesures de justesse et précision. Dans le groupe *Motifs*, chaque ensemble contient toutes les notes des imitations d'un même motif dans une condition. L'interaction des facteurs "motif" (17 niveaux) et "condition" (3 niveaux) entraîne 51 séries de mesures de justesse et précision. Dans le groupe *Intervalles*, chaque ensemble comprend toutes les notes précédées du même intervalle dans une condition. La première note de chaque motif est classifiée comme précédée d'un intervalle unisson. L'interaction des facteurs "intervalle" (16 niveaux) et "condition" (3 niveaux) conduit à 48 séries de mesures de justesse et précision. Enfin, dans le groupe *Tempi*, chaque ensemble est constitué des notes jouées par un sujet pour un certain motif et un certain tempo. L'interaction des facteurs "sujets" (28 niveaux), "motifs" (12 niveaux) et "tempi" (3 niveaux) résulte en 1008 séries de mesures pour ce groupe. Le tableau 3.2 résume les différents groupes étudiés. Les justesses et précisions des différents groupes sont comparées à l'aide d'un test de Wilcoxon par paires sur la plateforme de traitement statistique R [tea13].

3.3.2 Observations générales

Les justesses et précisions calculées pour les groupes *Sujets*, *Motifs* et *Intervalles* pour les trois conditions sont présentées en figure 3.5. Chaque boîte contient les 2e et 3e quartiles des valeurs, et la ligne épaisse représente la médiane.

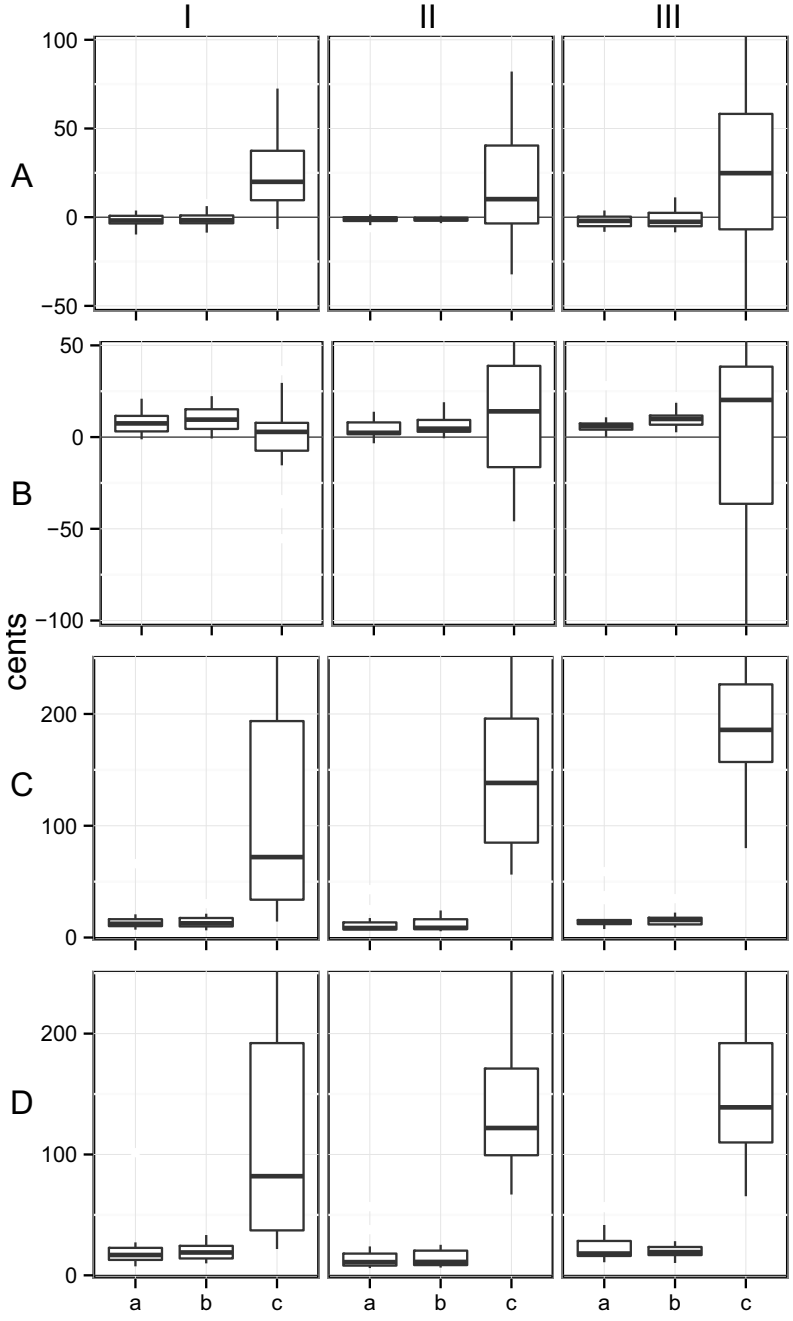

Figure 3.5 – *Observations générales – Justesses de notes (A), d'intervalles (B) et précisions de notes (C) et d'intervalles (D) pour le groupe des Sujets (I), des Motifs (II) et des Intervalles (III) pour les conditions "chironomique" (a), "chironomique muette" (b) et "vocale" (c).*

Nous observons principalement des valeurs de justesses et précisions systématiquement faibles pour les imitations à la tablette et une dispersion plus grande pour les imitations vocales. Les valeurs de précisions de notes et intervalles obtenues pour la tablette sont significativement inférieures à celles des imitations vocales ($W \leq 36, p < 0.05$). La modalité *vocale* paraît donc plus difficile que les modalités *chironomique* et *chironomique muette*.

De plus les justesses de notes sont proches de 0 pour les deux modalités tablette. Elles sont significativement plus faibles que celles de la modalité *vocale* pour le groupe des *Sujets* (W = 49, p < 0.05 entre *chironomie* et *voix*, et W = 50, p < 0.05 entre *chironomie muette* et *voix*) et le groupe *Intervalles* (W = 75, p < 0.05 entre *chironomie* et *voix*).

Les justesses d'intervalles des modalités tablette sont toujours inférieures à 25 cents, mais jamais significativement plus faibles que celles de la modalité *vocale*. A l'inverse, les valeurs de justesses d'intervalles de la modalité *vocale* sont significativement inférieures à celles des justesses *chironomiques muettes* pour le groupe des *Sujets* (W = 286, p < 0.05). Néanmoins, la différence des médianes est négligeable du point de vue du seuil de perception.

Enfin, aucune différence significative n'est observée entre les modalités *chironomique* et *chironomique muette*, quelque soit la mesure considérée. Les justesses sont comprises entre -3 et 10 cents, proche du seuil de perception, et les précisions entre 8 et 19 cents. 7 cents correspondent à un millimètre sur la tablette, soit approximativement la largeur de la pointe du stylet. En absence d'apprentissage, la largeur de la pointe du stylet devient une limite à la haute résolution de la tablette. Les excellents résultats obtenus pour la modalité *chironomie muette* rendent difficile l'observation de l'influence du retour audio.

3.3.3 Effet de la voix et de l'entraînement musical

La figure 3.6 montre les précisions d'intervalles, de notes et justesses d'intervalles et de notes de chaque sujet classées par précision d'intervalles décroissantes. On choisit une mesure d'intervalle car celles-ci sont plus représentatives d'une tâche musicale que les mesures sur les notes. De plus, une mesure de précision renseigne sur la consistance d'un sujet, et un sujet précis est en général juste [PBM+10]. La figure 3.7 montre les années de pratique musicale des sujets, ordonnées selon le même critère.

Les meilleurs sujets en terme de précision d'intervalles ont des résultats comparables dans les 3 modalités, avec parfois des meilleurs résultats pour la voix. De plus, 9 sujets ont des meilleures justesses d'intervalles pour la voix, contre seulement 3 sujets en regardant les justesses de notes. Cela laisse supposer que les aides visuelles proposées par la tablette favorisent l'atteinte de cibles en position absolue, donc la justesse de notes, alors que dans le cas de la voix, les sujets se concentrent sur la reproduction d'intervalles. La figure 3.7 montre que les meilleurs sujets ont globalement plus d'expérience musicale. L'entraînement musical a donc une influence non négligeable sur les performances vocales des sujets. De manière globale, les meilleures performances sont réalisées à la tablette par l'ensemble des sujets.

La figure 3.8 représente les moyennes des justesses et précisions de notes et d'intervalles de chaque sujet sur des plans justesse×précision. Seules les valeurs de précision inférieures à 100 cents et les valeurs de justesse comprises entre -50 et 50 cents sont présentées. On observe à nouveau une dispersion des valeurs de la modalité *vocale* supérieure à celles des modalités tablettes, notamment pour les valeurs de précision.

Pfordresher *et al.* proposent différents seuils de justesse et précision au-delà desquels un sujet est considéré comme non-juste ou imprécis [PBM+10] (50 cents, 100 cents, 250 cents). Le plus petit intervalle possible en musique tonale occidentale étant le demi-ton, un seuil

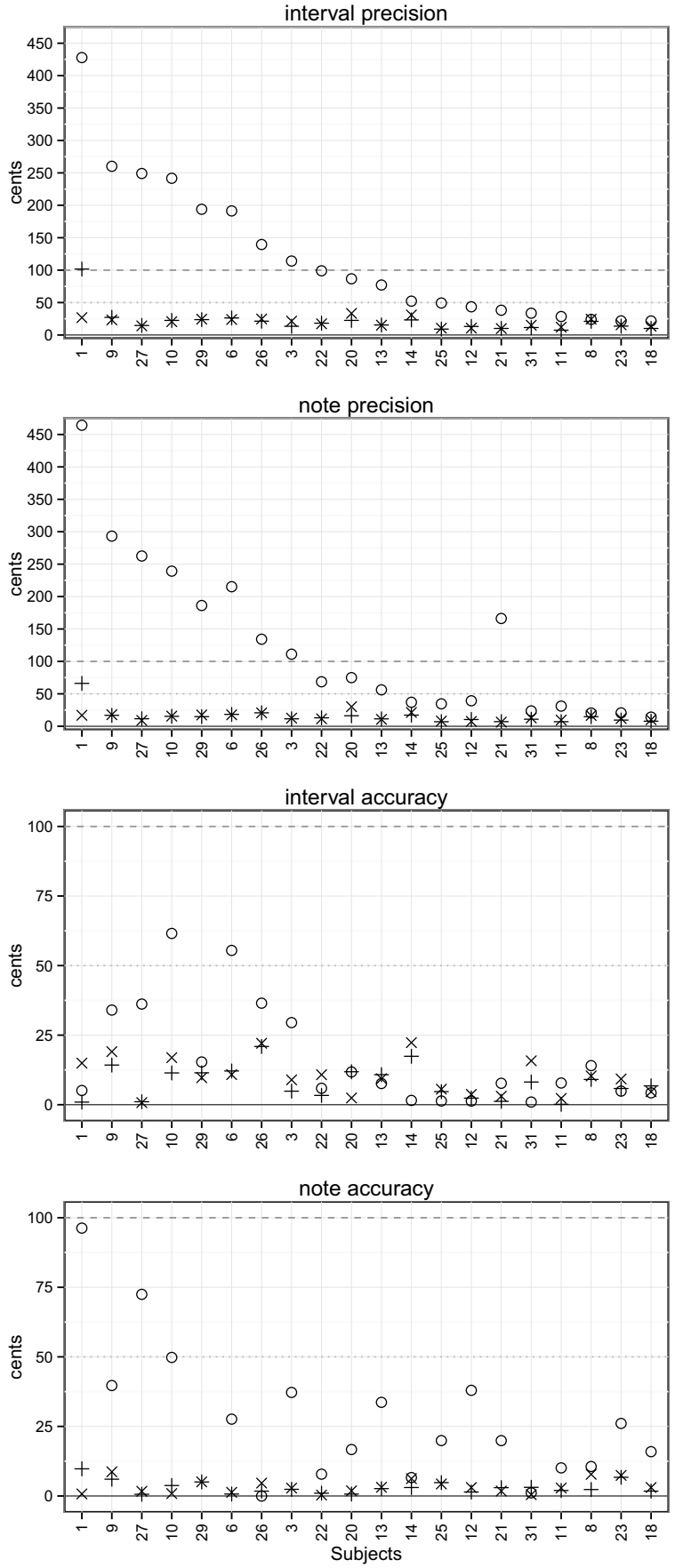

Figure 3.6 – *Effet de l'entraînement musical – Moyenne des valeurs de précisions d'intervalles, de notes, et justesses d'intervalles et de notes pour chaque sujet, ordonnées par valeurs de précisions d'intervalles décroissantes. Les modalités sont représentées par des + pour la chironomie, des × pour la chironomie muette, et des ∘ pour la voix.*

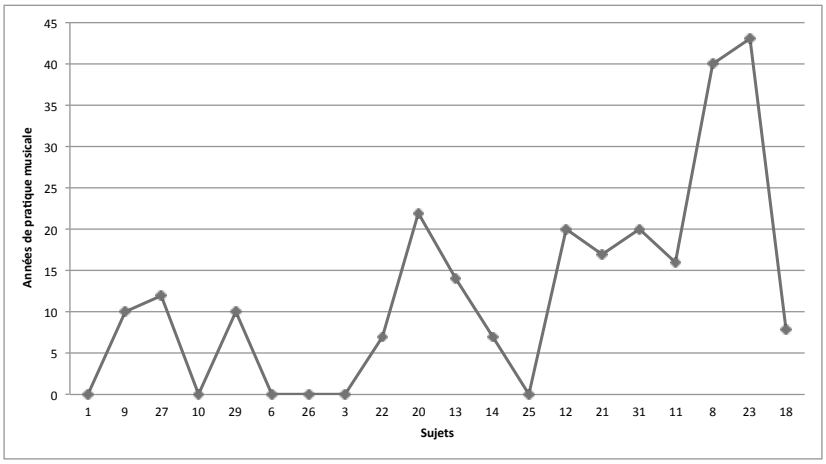

FIGURE 3.7 – *Années de pratique musicale par sujet, ordonnées par valeurs de précisions d'intervalles décroissantes (voir figure 3.6).*

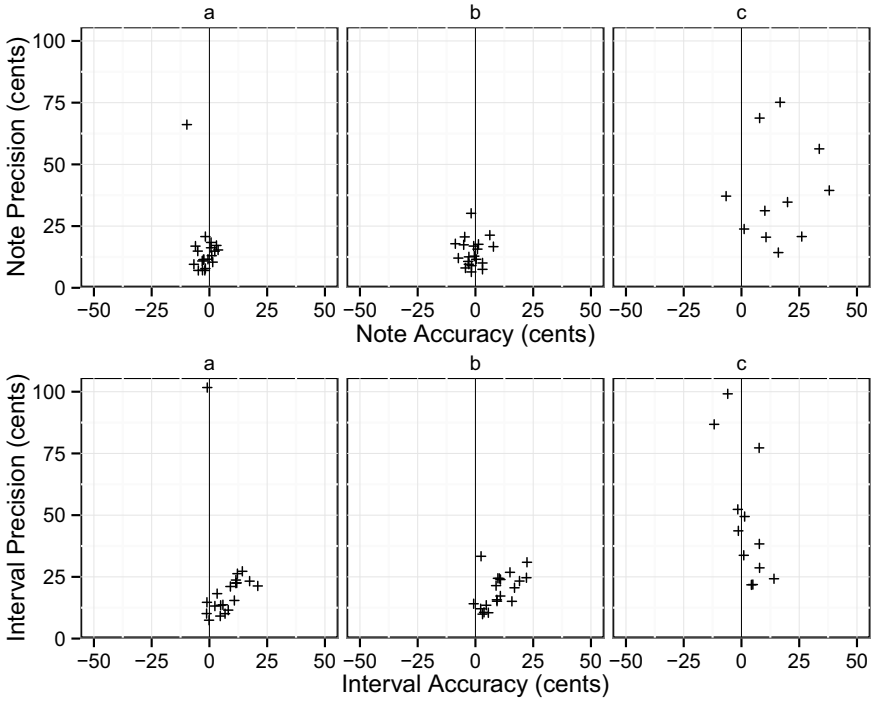

FIGURE 3.8 – *Moyenne des précisions en fonction des moyennes des justesses de chaque sujet pour les notes (haut) et les intervalles (bas) pour les modalités "chironomie" (a), "chironomie muette" (b) et "vocale" (c).*

de 50 cents permet de faire la différence entre des notes correctes et incorrectes. Le tableau 3.3 montre le nombre de sujets justes et/ou précis en termes de justesses et précisions de notes pour les trois modalités considérant un seuil de 50 cents. Le tableau 3.4 reporte les pourcentages de sujets justes et précis de l'étude de Pfordresher calculés avec le même seuil.

Voix	Précis (%)	Imprécis (%)	
Juste (%)	40	45	85
Non-juste (%)	0	15	15
	40	60	
Chironomie	Précis (%)	Imprécis (%)	
Juste (%)	100	0	100
Non-juste (%)	0	0	0
	100	0	
Chironomie muette	Précis (%)	Imprécis (%)	
Juste (%)	95	5	100
Non-juste (%)	0	0	0
	95	5	

Table 3.3 – *Pourcentage des sujets justes et précis (justesse et précision de notes) pour chaque condition selon un seuil de 50 cents parmi les 20 sujets des blocs 1 et 2.*

Voix	Précis (%)	Imprécis (%)	
Juste (%)	24	45	69
Non-juste (%)	3	28	31
	27	73	

Table 3.4 – *Pourcentage de sujets justes et précis (justesse et précision de notes) reportées par Pfordresher et al. [PBM+ 10] selon un seuil de 50 cents.*

On note 85 % de sujets justes et 40 % de sujets précis pour la modalité *vocale* dans notre étude, contre 69 % de sujets justes et 27 % de sujets précis dans celle de Pfordresher. De plus, 40 % de nos sujets sont à la fois justes et précis, contre 25 % pour la précédente étude. Cela peut s'expliquer par un entraînement musical moyen important chez les sujets de notre étude (18 ans), alors que Pfordresher *et al.* ont sélectionné des sujets sans expérience musicale.

Par ailleurs, tous nos sujets sauf un sont justes et précis pour les deux modalités tablettes, bien qu'aucun n'ait d'expérience dans l'usage d'une tablette graphique. Cela montre que contrairement à la modalité *vocale* où un entraînement musical est nécessaire à une bonne performance, les modalités *chironomiques* et *chironomiques muettes* permettent de jouer juste et précis quelque soit l'expérience musicale du sujet.

3.3.4 Effet des motifs

Les blocs 1 et 2 de l'expérience présentent des motifs de tailles variables, et la longueur d'un motif est un facteur de difficulté lors d'une tâche d'imitation. La figure 3.9 reporte les justesses et précisions de notes (I et II) et d'intervalles (III et IV) du groupe des *Motifs*, en séparant les intervalles du bloc 1 (I et III) et les mélodies du bloc 2 (II et IV).

Aucune différence entre les deux blocs ne se manifeste pour les justesses de notes (I et II). En revanche, la justesse d'intervalles de la modalité *vocale* est meilleure pour les mélodies du bloc 2 que pour les intervalles du bloc 1 ($W = 38$, $p = 0.43$). A l'inverse, la justesse d'intervalles des modalités tablette est meilleure pour les intervalles du bloc 1 que pour les mélodies du bloc 2 (*chironomie* : $W = 2$, $p < 0.01$; *chironomie muette* : $W = 2$, $p < 0.01$). Cela s'explique par la nature très musicale de la tâche mélodique, plus proche de l'expérience vocale des sujets que de leur expérience de manipulation d'un stylet.

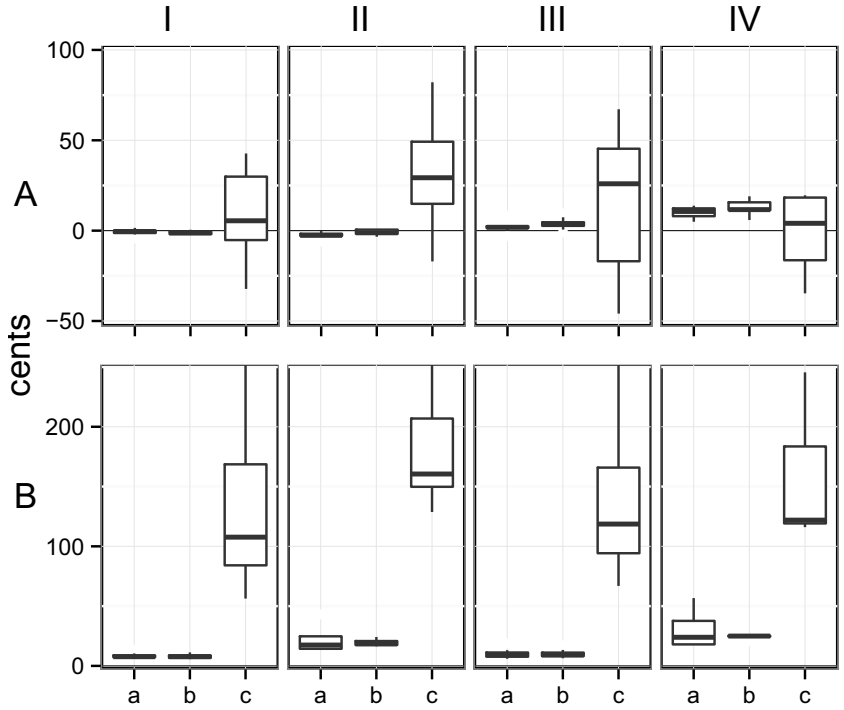

FIGURE 3.9 – *Effet des motifs – Justesses (A) et précisions (B) de notes (I et II) et d'intervalles (III et IV) pour le groupe des Motifs. Les blocs 1 (I et III) et 2 (II et IV) sont représentés distinctement pour les trois modalités "chironomie" (a), "chironomie muette" (b) et "vocale" (c).*

3.3.5 Effet des tailles d'intervalles

Pour les tâches chironomiques, la taille d'un intervalle à jouer est corrélée à la distance du déplacement du stylet sur la tablette. Par conséquent, des intervalles plus larges entraînent des mouvements plus amples et plus difficiles à réaliser. La figure 3.10 montre les justesses (A) et précisions (B) de notes (I, III et IV) et d'intervalles (II et V) pour le groupe des *Intervalles* en distinguant les intervalles descendants (I et II), unisson (III) et montants (IV et V).

Les justesses de notes des modalités tablettes sont systématiquement négatives pour les mouvements descendants (I) et positives pour les mouvements ascendants (IV) indiquant un effet de dépassement des cibles. Les différences entre mouvements descendants et ascendants sont significatives pour les deux modalités : 5 cents de différence pour la modalité *chironomie* ($W = 9$, $p < 0.05$) et 9 cents de différence pour la modalité *chironomie muette* ($W = 0$, $p < 0.05$). La différence étant plus faible avec un retour audio, on en déduit que ce dernier réduit l'effet de dépassement. Ce résultat se manifeste aussi par une justesse d'intervalle plus faible pour la modalité *chironomique* que pour la modalité *chironomique muette* ($W = 56$, $p < 0.05$).

Des résultats inexpliqués sont les justesses de la modalité *vocale* systématiquement positives, quelque soit la direction de l'intervalle, alors que les justesses des modalités tablettes sont équitablement réparties dans chaque direction. L'algorithme d'extraction de notes a été testé sur des exemples de synthèse mais aucun biais introduisant des décalages systématiques n'a été relevé.

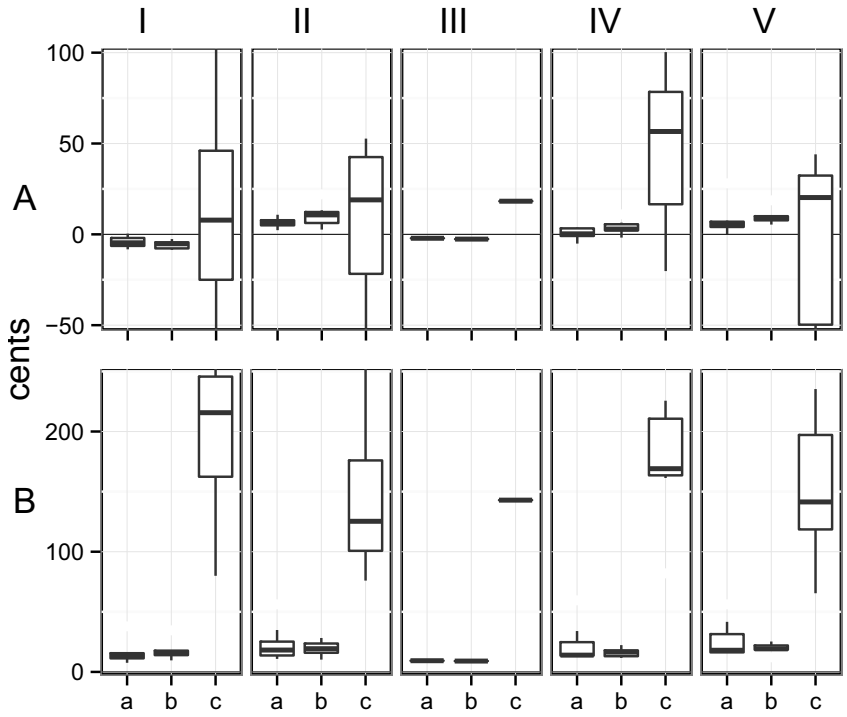

Figure 3.10 – *Effet des intervalles – Justesses (A) et précisions (B) de notes (I, III et IV) et d'intervalles (II et V) pour le groupe des Intervalles. Les intervalles descendants (I et II), unisson (III) et montants (IV et V) sont représentés distinctement pour les trois modalités "chironomie" (a), "chironomie muette" (b) et "vocale" (c).*

3.3.6 Effet du tempo

L'effet du tempo est étudié sur les imitations du bloc 3. La figure 3.11 montre les justesses (A) et précisions (B) de notes (I) et d'intervalles (II) pour les trois tempi proposés. Bien qu'on pourrait s'attendre à une détérioration des performances avec une augmentation du tempo, aucune différence significative n'apparaît entre les différents tempi. Les sujets sont justes et précis dans l'imitation des doubles intervalles au trois tempi. Ces résultats indiquent qu'un tempo critique au-delà duquel les performances se dégradent n'a pas été atteint.

3.4 Discussion et conclusion

3.4.1 Résumé des résultats

Deux résultats principaux se dégagent de cette expérience. Tout d'abord, les imitations chironomiques sont globalement plus justes et plus précises que les imitations vocales. L'exception à ces observations est la meilleure performance vocale des imitations de mélodies, comparées aux imitations d'intervalles en terme de justesse d'intervalle. L'étude de l'influence de l'entraînement musical des sujets a mis en évidence la nécessité d'une connaissance musicale pour réaliser des imitations justes et précises vocalement, alors que celle-ci n'est pas nécessaire pour les imitations chironomiques. Il a été montré cependant que les imitations chironomiques subissaient les effets moteurs du mouvement de la main caractérisés par des

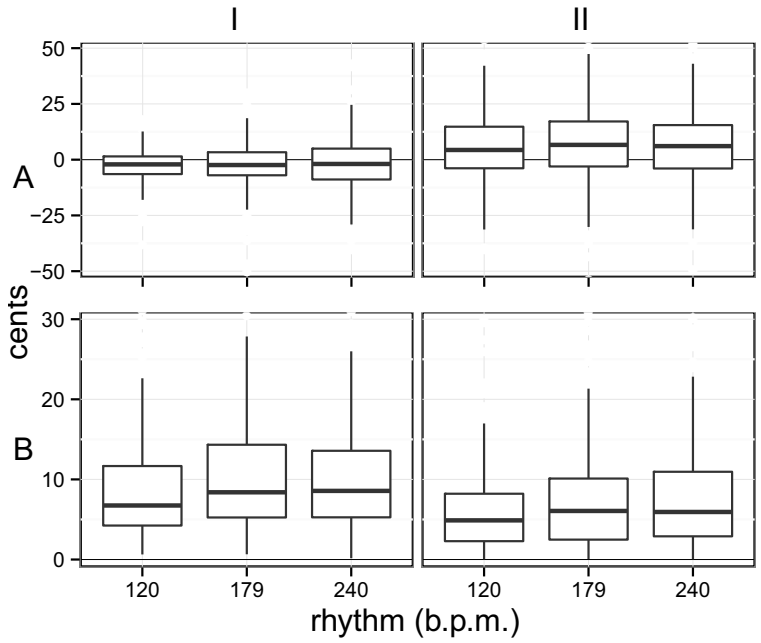

FIGURE 3.11 – *Effet du tempo – Justesses (A) et précisions (B) de notes (I) et d'intervalles (II) pour les trois tempi proposés.*

dépassements systématiques des cibles. L'étude de l'effet du tempo n'a pas permis d'établir un tempo critique au delà duquel les performances d'imitation de doubles intervalles seraient dégradées. Ensuite, très peu de différences sont apparues entre les modalités *chironomiques* et *chironomiques muettes*. La nécessité d'un retour audio n'a donc pas été démontrée ici.

3.4.2 Du chant à l'écriture

Alors que les qualités des performances vocales des sujets sont dispersées, toutes les performances chironomiques sont justes et précises. Seuls les sujets ayant le plus d'entraînement musical sont parvenus à obtenir des résultats comparables à la tablette et à la voix. Les sujets non-entraînés quant à eux présentent des bons résultats par chironomie uniquement. L'aptitude chironomique semble donc être beaucoup plus répandue que l'aptitude vocale.

Cela s'explique par une expérience d'écriture ou de dessin très importante chez tous les sujets, et plus généralement chez une majeure partie de la population. Comme l'écriture est acquise depuis l'enfance, chaque sujet, quelque soit son expérience musicale, montre des prédispositions au contrôle chironomique. De plus, les indicateurs visuels placés sur la tablette ont permis au sujets non-musiciens de localiser les notes cibles très facilement, transposant ainsi la tâche de chant en une tâche de dessin ou d'écriture.

Par ailleurs, la présence d'indicateurs visuels favorise la recherche de cible en terme de position absolue. La justesse de notes est donc plus adaptée à la mesure de tâches chironomiques. A l'inverse, l'appareil vocal ne fournit comme référence que la note jouée précédemment. La recherche de cible est donc relative à la précédente, et la justesse d'intervalles est donc privilégiée. Des justesses d'intervalles plus faibles ont effectivement été observées pour la voix dans un contexte mélodique, alors que les modalités chironomiques présentent de meilleures justesses de notes pour la même tâche.

Concernant la virtuosité des tâches demandées, la limite de dégradation des performances n'a pas été atteinte avec un tempo maximal de 240 b.p.m. Cependant, ce résultat est à tempérer du fait que les motifs du bloc 3 sont relativement simples à jouer : doubles intervalles symétriques. Il serait par la suite pertinent d'aller plus loin dans l'investigation de ce tempo critique, en imitant à la fois des motifs plus représentatifs d'une virtuosité mélodique, et tout en augmentant le tempo.

Enfin, il a été prouvé que la nature du timbre du stimulus à écouter influence la justesse d'imitation vocale [GIKGK13]. La qualité de la voix synthétique présentée dans l'expérience étant relativement mauvaise, des meilleures justesses et précisons vocales pourraient être atteintes après présentation de stimuli vocaux, réduisant ainsi la différence entre performances vocales et chironomiques.

3.4.3 Importance du retour audio

Le peu de différences observées entre les modalités *chironomiques* et *chironomiques muettes* tendent à montrer que le retour audio est inutile quand des indices visuels sont présents sur la tablette. A l'inverse, la suppression du retour audio dans le chant entraîne une forte dégradation des performances vocales, le retour kinesthésique seul ne permettant pas de contrôler l'intonation précisément [MPHS02].

Afin de tester l'influence du retour audio sur la justesse et la précision des sujets, il pourrait être envisagé de réaliser une condition *chironomie aveugle* ou les sujets ne disposeraient pas de retour visuel. Cela poserait cependant un important problème d'apprentissage. Le sujet se retrouverait dans une situation similaire à un violoniste débutant, confronté à apprendre à contrôler la hauteur du son aveuglément sur une dimension spatiale continue. Le violoniste parvient à maîtriser son instrument en développant une technique de placement de la main suivant différentes positions sur le manche seulement après de nombreuses années d'apprentissage. Nous pouvons émettre l'hypothèse qu'un apprentissage similaire serait nécessaire avec la tablette, tout comme il l'a été au thérémine. Un tel apprentissage n'étant pas concevable pour notre expérience, l'introduction de la condition *chironomie aveugle* aurait probablement produit des résultats aberrants. Un protocole expérimental différent étudiant l'influence du retour audio comparé au retour visuel doit être considéré, et est abordé au chapitre 5.

Finalement, bien que le retour visuel semble prépondérant pour la justesse et la précision chironomique, il n'est évidemment pas envisageable d'interpréter une pièce musicale sans retour audio, ce dernier étant essentiel à la fois pour l'expressivité du musicien, et pour le jeu d'ensemble.

3.4.4 Conclusion

La comparaison des performances vocales et chironomiques a prouvé que l'interface proposée pour le contrôle de la synthèse vocale permettait d'égaler, et même de dépasser les justesses et précisions obtenues vocalement. Par ailleurs, l'expérience nécessaire pour un contrôle chironomique de l'intonation juste et précis est acquise par la majorité de la population lors de l'enfance par l'apprentissage de l'écriture. Cela fournit alors une grande accessibilité à notre instrument et prouve que la tablette graphique est un bon candidat pour le contrôle de l'intonation de la synthèse vocale.

Chapitre 4

Méthodes de correction dynamique de la justesse mélodique

Sommaire

4.1 Introduction

La plupart des styles de musiques sont construits sur des gammes proposant un ensemble fini et discret de notes. Il est indispensable pour le musicien de pouvoir atteindre ces notes précisément car la moindre erreur détériore grandement le rendu de la prestation, quelle que soit la qualité sonore de l'instrument. Par ailleurs, il est d'usage d'agrémenter la trajectoire mélodique de modulations reflétant une des formes d'expression possibles du musicien [JL03], qu'on appellera par la suite expressivité mélodique ou expressivité. Trois modulations de trajectoires communément utilisées sont le *portamento*, transition continue et courte entre deux notes, le *glissando*, transition continue et longue entre deux notes, et le *vibrato*, oscillation autour d'une hauteur stable. Savoir combiner justesse et expressivité mélodiques est une étape importante dans l'apprentissage d'un instrument.

Cette dualité justesse vs. expressivité est corrélée à l'espace de contrôle de la hauteur mélodique. Les instruments de musique acoustiques mélodiques sont catégorisés en trois espaces de contrôle. Certains possèdent un espace entièrement discret (p. ex. claviers). Seules les notes d'une échelle donnée sont proposées, induisant une justesse parfaite mais sans modulation possible. A l'inverse, d'autres instruments présentent un espace de contrôle mélodique continu (p. ex. cordes frottées, voix). Atteindre une note juste est plus difficile car la moindre déviation est perçue comme une fausse note. En revanche, toute modulation de hauteur est possible, favorisant l'expressivité mélodique du musicien. La troisième catégorie possède à la fois une échelle discrète proposée par un certain doigté, et un contrôle continu local réalisé à l'embouchure (p. ex. instruments à clés ou à pistons). L'identité d'un instrument est en partie définie par son espace de contrôle mélodique, par conséquent le choix de ce dernier est inévitable lors de la construction d'un nouvel instrument de musique numérique mélodique. Il est d'abord possible de concevoir un espace de contrôle discret par l'usage de contrôleurs de type clavier, tout en permettant des variations de hauteur expressives autour des notes cibles (par exemple molette "Pitch bend" sur claviers Minimoog ou Yamaha DX7), à l'image de la troisième catégorie d'instruments. Avec l'avènement de nombreuses interfaces continues telles que les écrans tactiles, les claviers continus, les tables tangibles ou les tablettes graphiques, il semble pertinent de se pencher sur le problème inverse : la conception d'un espace de contrôle continu augmenté d'un ajustement de hauteur mélodique facilitant la justesse.

En effet, le *Cantor Digitalis* propose un espace de contrôle continu, en imitation de la voix. L'étude précédente (chapitre 3) a montré que la tablette graphique permettait un contrôle de la synthèse vocale juste et précis, quelque soit l'entraînement musical des sujets. Toutefois, les mesures ont été effectuées dans un contexte de laboratoire, sur des motifs mélodiques sortis de tout contexte musical. Dans des conditions de jeu réelles, en concert ou en ensemble polyphonique, la tâche est plus difficile. C'est pourquoi nous proposons une méthode d'ajustement de justesse afin de faciliter le jeu du *Cantor Digitalis*, et plus généralement d'instruments numériques à espace de contrôle continu [1]. Après une présentation des méthodes existantes dans la suite de l'introduction, notre méthode est détaillée en section 4.2 et évaluée en terme d'amélioration de la justesse mélodique et de préservation de l'expressivité de l'instrumentiste en section 4.3.

1. Méthode d'ajustement et évaluation ont été publiées dans [Pd13], [Pd15], et [Pd].

4.1.1 Méthodes de correction dans le domaine visuel

Les méthodes de correction de justesse mélodique actuelles sont inspirées de méthodes de correction visuelles, couramment utilisée pour l'aide au maniement du curseur à l'écran. En effet, les instruments de musique numériques, tout comme les interfaces graphiques, effectuent une transformation depuis le domaine moteur vers des domaines perceptifs : auditif pour les instruments, exprimé en fréquences, et visuel pour les interfaces graphiques, exprimé en pixels. Ainsi, les actions consistant à jouer une mélodie en atteignant des fréquences spécifiques dans le domaine auditif ou à placer le curseur de la souris sur un icône en atteignant une position spécifique à l'écran dans le domaine visuel, correspondent toutes deux à une tâche spatiale de pointage dans le domaine moteur. Corriger la justesse mélodique ou la position d'un curseur à l'écran sont donc deux ajustements de même nature, effectués dans le domaine perceptif en réponse à un manque de précision d'une tâche moteur de pointage, mais ayant chacun des contraintes adaptées à leurs domaines respectifs (auditif ou visuel).

Aide au pointage sur les interfaces graphiques (GUI)

Deux principes d'aide au contrôle d'un curseur dans le contexte d'une interface souris ressortent de la littérature : élargir la taille de la cible visée, ou réduire l'amplitude du mouvement à réaliser [Bal04]. En effet, étendre visuellement une cible lorsque le curseur est suffisamment proche selon la méthode *Expanding target* [MB02], [MB05], y ajouter une bulle autour [CF03] ou étendre la zone de sélection du curseur [CLP09] permet de faciliter l'acquisition de cibles. Toutefois, ces méthodes introduisent des distorsions à l'affichage visuel, pouvant potentiellement altérer l'acquisition des cibles voisines. De plus, les utilisateurs ont tendance à anticiper l'expansion des cibles et deviennent moins stricts sur leurs performances, visant la zone autour de la cible et non la cible elle-même.

Une solution alternative permettant l'expansion d'une cible sans modifier l'affichage visuel est d'étendre celle-ci dans le domaine moteur. Le rapport contrôle-affichage C-D (pour *Control-Display ratio*) est un gain transformant le mouvement de l'utilisateur dans le domaine moteur, mesuré en mètres, en déplacement dans le domaine visuel, mesuré en pixels [MR94]. Diminuer localement la valeur du rapport C-D autour de la cible rend cette dernière plus large dans le domaine moteur, et par conséquent plus facile à atteindre [BGBL04]. C'est le principe de la méthode *Sticky Icons* [WWBH97]. A l'inverse, l'augmentation du rapport C-D pendant un déplacement du curseur vers la cible réduit l'amplitude du mouvement dans le domaine moteur.

La comparaison entre les méthodes *Expanding targets* et *Sticky Icons* montre des performances similaires dans l'aide à l'acquisition de cibles [CF03]. Néanmoins, la méthode *Sticky Icons* est préférée des utilisateurs car elle est considérée moins intrusive, et parce qu'elle n'introduit pas de paresse lors de l'atteinte de cible, car l'expansion est moteur et non visuelle et donc plus difficile à anticiper.

Aide au pointage distant

L'acquisition de cibles est plus difficile qu'avec la souris dans le cas de pointages distants [PKS+12], [PTIK13], que ce soit avec une télécommande (e.g. Nintendo Wii) avec les mains (e.g. Kinect de Microsoft) ou par des techniques hybrides (e.g. stylet pour viser des cibles lointaines sur une tablette). Des méthodes d'aide au pointage ont été adaptées pour le pointage distant au stylet [PMNI05] ou à la main [MHT14]. Finalement, la méthode *Sticky Icons* fournit de meilleures performances que la méthode *Expanding targets* pour ces applications.

4.1.2 Méthodes de correction dans le domaine auditif

Les méthodes de correction auditives sont apparues il y a une vingtaine d'année pour la correction de la hauteur mélodique de chanteurs. Auto-tune de la société Antares [2] [Hil99] est un des systèmes les plus célèbres dans ce domaine. Pour la correction du chant, la hauteur doit d'abord être extraite du signal original et ce dernier est modifié. Cette étape de transformation du signal sonore associée à la discrétisation de la hauteur entraîne des distorsions audibles, souvent recherchées par certains styles de musique.

Dans le cas d'instruments de musique numériques, nous cherchons à minimiser tout artéfact lié à la correction. Cela est facilité par l'accès à la trajectoire de hauteur avant la phase de production sonore. De nombreux constructeurs d'instruments de musique numériques à espace de contrôle continu ont intégré des ajustements de justesse pour améliorer le confort du musicien. C'est le cas du *Continuum Fingerboard* de Haken [HTW98], le *Seaboard* de Roli [LR11], le *LinnStrument* de Linn [3], les applications iOS *Morphwiz* [4] et *Garageband* [5], ou le *TouchKeys* de McPherson [MGS13].

L'ajustement de justesse consiste à découpler la *hauteur d'entrée* jouée par le musicien sur l'interface de la *hauteur de sortie* liée au synthétiseur et perçue par le musicien en y introduisant une étape de *mapping*. Bien que les implémentations des méthodes d'ajustement de certains produits commerciaux ne soient pas explicitées, deux catégories d'ajustement sont identifiées. La première englobe les méthodes de convergence de hauteur. La trajectoire de hauteur de sortie est continuellement ajustée vers la note la plus proche d'une échelle prédéfinie. La deuxième catégorie est composée de méthodes de déformation de hauteur (*pitch warping*). La hauteur de sortie est alors calculée selon une fonction de la hauteur d'entrée.

Méthodes de convergence de hauteur

Les méthodes de convergence consistent à systématiquement attirer la trajectoire de hauteur de sortie vers la note la plus proche d'une gamme prédéfinie. Auto-tune [Hil99] est un exemple de méthode de convergence à un degré de liberté, c'est-à-dire dont la convergence est régie par un seul paramètre. Celui-ci est appelé *retune speed* et définit la vitesse de convergence de la trajectoire vers la cible. Une valeur de 0 indique un ajustement immédiat de la hauteur sur la cible, soit une discrétisation complète de la hauteur de sortie. Des valeurs plus larges induisent un compromis entre justesse et expressivité, contourné par deux options : le choix d'une seconde valeur pour les notes tenues, et l'addition d'un vibrato artificiel.

La correction Haken [Hak09] implémentée sur le *Continuum Fingerboard* [HTW98] est une autre méthode de convergence de hauteur, mais à deux degrés de liberté. La convergence est définie par deux paramètres : un pas de correction CS (*Correction Step*) en demi-tons et un pas temporel TS (*Time Step*) en secondes. A chaque intervalle de temps TS, la trajectoire d'entrée est augmentée ou diminuée d'un pas de correction fixe $\pm CS$ de telle sorte que la valeur moyenne locale de la trajectoire se rapproche de la note cible, sans toucher aux variations relatives à l'expressivité. Les degrés de libertés associés aux paramètres sont la fréquence de correction indiquée par le pas temporel TS, et la préservation de l'expressivité donnée par le pas de correction CS. Chaque modulation de hauteur dont la vitesse est inférieure à CS/TS sera distordue. Par conséquent, ce rapport doit être plus faible que la vitesse de la modulation la plus lente à préserver.

2. http://www.antarestech.com/products/detail.php?product=Auto-Tune_8_66 (vérifié le 22 octobre 2015)
3. http://www.rogerlinndesign.com/linnstrument.html (vérifié le 22 octobre 2015)
4. http://www.wizdommusic.com/products/morphwiz.html (vérifié le 22 octobre 2015)
5. http://www.apple.com/fr/ios/garageband/ (vérifié le 22 octobre 2015)

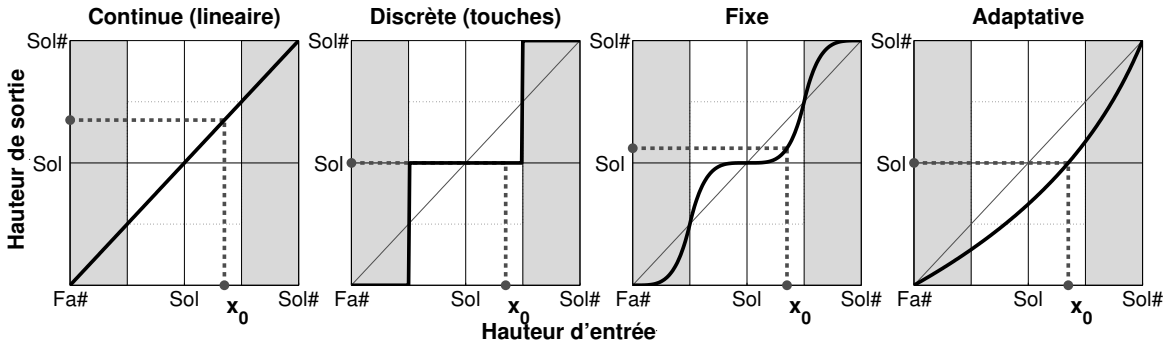

FIGURE 4.1 – *Exemples de fonctions de déformation. De gauche à droite : fonction linéaire (continu) ; fonction en escalier (discret) ; fonction intermédiaire (fixe) ; fonction adaptative.*

Méthodes de déformation de hauteur

Les déformations de hauteur les plus simples consistent en l'application de relations statiques entre hauteurs d'entrée et de sortie. Une revue de ces relations est fournie par Goudard *et al.* [GGF14] et des exemples sont illustrés en figure 4.1. La fonction de déformation neutre est la fonction linéaire (gauche) qui décrit la fonction d'un instrument à contrôle continu sans correction. A l'opposé, la fonction en escalier (deuxième graphe) arrondit ou discrétise la hauteur d'entrée en hauteurs de sorties correspondant aux notes de la gamme choisie. Sa forme ne permet pas d'effectuer des modulations de hauteur. Une fonction intermédiaire fréquemment utilisée est la fonction du troisième panneau. Par analogies aux corrections visuelles cette fonction réduit le rapport C-D autour des notes cibles. Enfin, la dernière fonction est dite adaptative car calculée en fonction de la hauteur d'entrée.

Appliquer ces fonctions de manière statique, c'est-à-dire systématiquement quelle que soit la hauteur d'entrée, impose des contraintes importantes, nuisibles à l'expressivité. Des transformations dynamiques à l'inverse permettent d'adapter les relations entrée/sortie en fonction de la dynamique de la hauteur et sont préférées.

Méthodes de déformation dynamiques

Le *TouchKeys* de McPherson [MGS13] implémente une méthode de déformation dynamique à deux degrés de liberté. L'instrument permet de réaliser des inflexions de hauteur (*pitch bend*) par des glissements verticaux des doigts sur les touches. La correction ajuste dynamiquement la hauteur à la fin du glissement pour que la note d'arrivée sonne juste. Les deux paramètres mis-en-jeu sont la taille d'intervalles autour des cibles appelés *snap zones*, et un seuil de vitesse. La correction s'applique si la hauteur entre dans une *snap zone* à une vitesse inférieure au seuil donné. La fonction de déformation ainsi que la valeur du seuil de vitesse ne sont pas explicitement donnés dans le papier.

Le compromis entre justesse et expressivité introduit par les deux méthodes de convergence de hauteur découle d'un ajustement permanent de la hauteur propre aux méthodes statiques. En introduisant un seuil de vitesse, McPherson distingue les notes stables à corriger des modulations de hauteur à laisser libre. Comme ces dernières présentent des variations plus grandes que les notes ciblées, la vitesse de hauteur est un bon indicateur de distinction entre les deux. Si la hauteur d'entrée est inférieure à une vitesse critique, celle-ci est corrigée. Dans le cas contraire, aucune correction n'est appliquée. Pour aller plus loin que la correc-

Degrés de liberté	Convergence de hauteur Statique	Déformation de hauteur Statique	Déformation de hauteur Dynamique
0		Mapping fixe [GGF14]	
1	Auto-tune [Hil99] *Retune speed*		
2	Haken [Hak09] *Correction step* *Time step*		McPherson [MGS13] *Snap zone size* *Speed threshold*
3			DPW *Intervalle de détection* *Temps critique* *Temps de transition*

Table 4.1 – *Résumé des types de correction selon leurs degrés de libertés donnés en italique.*

tion d'inflexions de hauteur, nous proposons une méthode à trois degrés de liberté appelée déformation de hauteur dynamique ou DPW (*Dynamique Pitch Warping*) permettant de corriger n'importe quelle trajectoire. Le tableau 4.1 résume l'ensemble des types de corrections présentées ainsi que notre correction DPW.

4.2 Méthodes d'ajustement

La méthode de déformation de hauteur dynamique est décrite en deux étapes. Il est d'abord nécessaire de définir le type de déformation entre hauteurs d'entrée et de sortie. Il faut ensuite décrire le mode de déclenchement de l'ajustement, en fonction de la dynamique de la trajectoire mélodique. Ces deux étapes sont présentées successivement après la présentation du principe de la correction.

4.2.1 Principe de la correction dynamique

Un exemple de trajectoire mélodique (*Fa-Sol-Fa*) est donné en figure 4.2. Les hauteurs d'entrée et de sortie sont les courbes épaisses respectivement en pointillés et continue. Quatre zones sont numérotées et correspondent aux quatre étapes du processus de correction :

1. La hauteur d'entrée entre dans un intervalle de détection I (lignes horizontales pointillées) et y reste plus longtemps que le temps critique T_c. Cela indique que la trajectoire d'entrée est suffisamment stable pour correspondre à une note. La correction est déclenchée.
2. La correction est appliquée graduellement et continûment pendant le temps de transition T_t, attirant la hauteur de sortie vers une des notes cibles représentées par les lignes pleines horizontales.
3. La hauteur de sortie évolue avec la fonction de déformation non-linéaire et converge vers la hauteur d'entrée lorsque la note exacte suivante est atteinte (Fa#).
4. Une correspondance linéaire est appliquée entre hauteurs d'entrée et de sortie de telle sorte que les courbes soient superposées jusqu'à l'application d'une nouvelle correction.

Finalement, lorsque la correction est déclenchée, la hauteur est corrigée vers la note la plus proche de la gamme prédéfinie, car la fonction de correspondance est déclenchée dynamiquement selon la hauteur d'entrée. Dans le cas de variations de la hauteur d'entrée après application de la correction, la hauteur de sortie change selon la fonction de correspondance, permettant des modulations de hauteur.

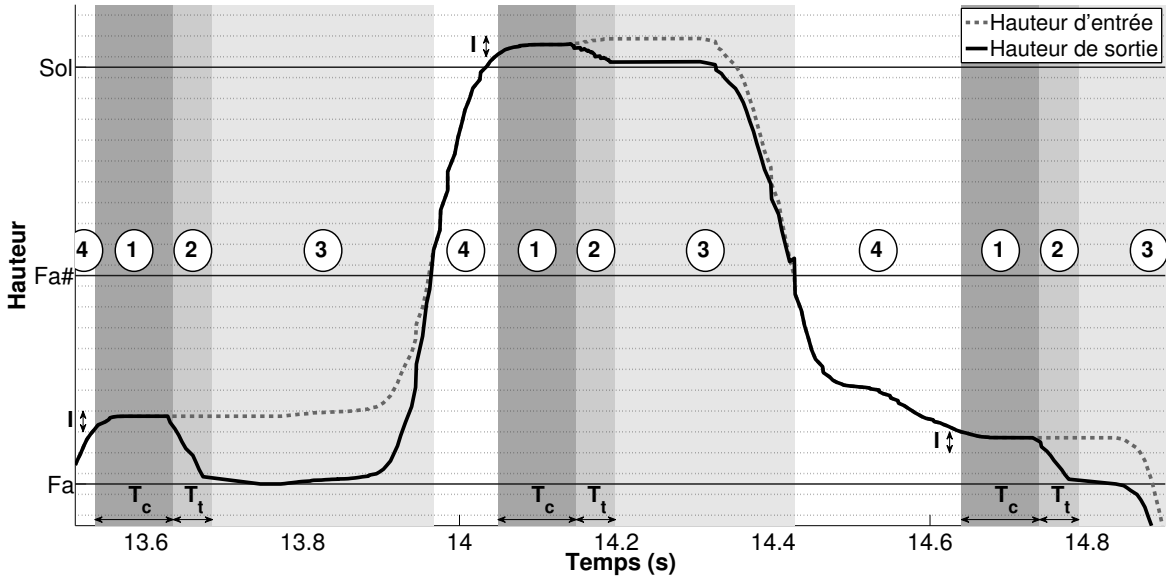

FIGURE 4.2 – *Correction note élastique. Les courbes épaisses en pointillés et continue sont respectivement les trajectoires de hauteur d'entrée et de sortie. Les lignes continues horizontales marquent les notes exactes à atteindre et les lignes horizontales pointillées sont les intervalles de détection* $I = 0.1$ *demi-ton. Quatre zones sont mises en évidence : (1) La hauteur d'entrée reste dans un intervalle de détection* I *plus longtemps que le temps critique* $T_c = 100$ *ms ; (2) La correction est appliquée pendant le temps de transition* $T_t = 50$ *ms ; (3) La hauteur évolue avec une fonction de correspondance non-linéaire ; (4) Une fonction de correspondance linéaire est appliquée après que le demi-ton suivant soit atteint (Fa#).*

4.2.2 Fonctions de correspondance statiques : fixe vs. adaptative

La relation entrée/sortie définie par la fonction de déformation peut-être caractérisée par deux familles de fonctions : les fonctions fixes et adaptatives. Les fonctions *fixes* sont calculées en fonction des notes cibles à atteindre et une fois définies lors du réglage de la correction, leurs expressions ou formes sont invariantes dans le temps. Elles sont couramment utilisées pour des ajustements statiques [GGF14] et sont équivalentes aux fonctions des applications visuelles telles que les *Sticky Icons* [WWBH97]. Un exemple de forme de fonction *fixe* récurrent est la présence d'un aplatissement de la courbe autour des notes cibles afin d'élargir les zones de justesse (figure 4.1 - troisième courbe). On utilisera par la suite cette forme que l'on appellera fonction *notes étendues*.

Les fonctions dites *adaptatives* sont uniques à chaque application d'une correction puisque celles-ci sont calculées par rapport à la hauteur d'entrée déclenchant la correction. Ainsi, deux hauteurs d'entrées différentes se verront appliquer deux fonctions adaptatives différentes. Celles-ci sont par exemple utilisées par le piano augmenté de McPherson [MGS13]. Elles déplacent la position de la note cible sur la position d'entrée afin que toute hauteur d'entrée donne une note juste à l'activation de l'ajustement (figure 4.1 - droite). Son fonctionnement est proche de la méthode *Expanding targets*, où quelque soit la position du curseur dans une zone autour de la cible, cette dernière est sélectionnée puisqu'agrandie. On utilisera par la suite une fonction adaptative en arche que l'on appellera fonction *élastique*.

Ces fonctions sont deux alternatives de la méthode DPW et sont présentées et comparées

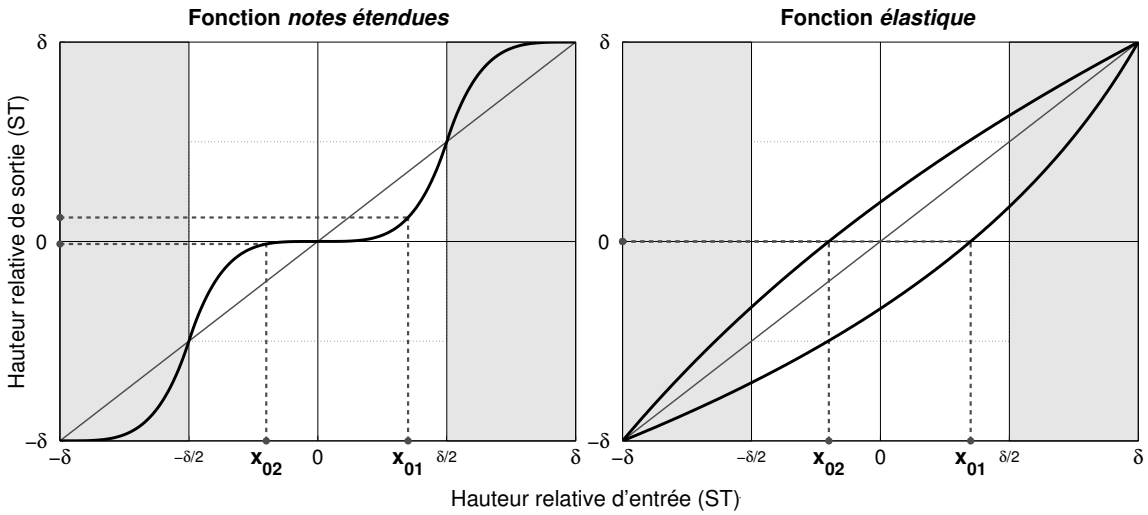

FIGURE 4.3 – *Fonctions de relation entre hauteurs d'entrée et de sortie exprimées en demi-tons (ST) relativement à la note cible (point (0,0)). Gauche : ajustement fixe "notes étendues" ; Droite : ajustements adaptatifs "élastiques", une fonction est calculée pour chaque position d'entrée.*

dans ce chapitre. Elles sont définies depuis l'espace des hauteurs d'entrée vers l'espace des hauteurs de sortie. Considérons ici une échelle générale d'intervalles de taille δ, chaque note exacte étant un multiple de δ. Les fonctions sont exprimées relativement à la note cible courante, où le point $(0, 0)$ correspond à la hauteur d'entrée jouant la cible exactement. $(-\delta, -\delta)$ et (δ, δ) désignent respectivement les notes exactes précédente et suivante. Les fonctions sont tracées en figure 4.3.

Fonction fixe : *notes étendues*

Tout comme le rapport C-D est augmenté autour de la cible par la méthode *Sticky Icons* [WWBH97], le rapport *hauteur d'entrée/hauteur de sortie* est ici augmenté autour de la note cible. Il en résulte un plateau autour cette dernière. Deux cas de correction émergent d'une telle fonction (figure 4.3 - gauche) : une hauteur d'entrée proche de la cible x_{02} entraîne une hauteur de sortie parfaitement juste. Une hauteur d'entrée plus éloignée x_{01} est améliorée mais imparfaitement corrigée. De plus, il est impossible de jouer un vibrato autour de la cible, car la hauteur de sortie est constante quelque soit la hauteur d'entrée. Pour contourner ce problème, il est possible d'introduire une pente autour de la cible mais cela entraîne toujours un compromis entre justesse (faible pente) et expressivité (large pente).

La fonction *notes étendues* vérifie 4 conditions :

- Passe par le point $(0, 0)$ correspondant à la cible à atteindre
- Est symétrique par rapport à l'origine
- Est continûment dérivable aux points $(-\delta/2, -\delta/2)$ et $(\delta/2, \delta/2)$ pour assurer une transition lisse entre les notes
- Choix du degré de distorsion de la fonction

Pour remplir ces conditions, la fonction est définie comme la somme de deux monômes de degrés respectifs z ($z > 1, z \in \mathbb{R}$) et 1. Le premier monôme est une fonction impaire dont

le degré contrôle la largeur du plateau. Le deuxième, de coefficient b, définit la tangente à l'origine, soit la pente du plateau. En respectant les conditions aux limites, la hauteur de sortie y_{NE} définie sur $[-\delta/2, \delta/2]$ s'exprime en fonction de la hauteur d'entrée x sous la forme :

$$y_{NE}(x) = (1-b)\left(\frac{2}{\delta}\right)^{z-1} \times signe(x) \times |x|^z + bx \tag{4.1}$$

Par la suite, on choisira $b = 0$ et $\delta = 1$ amenant à l'équation simplifiée :

$$y_{NE}(x) = 2^{z-1} \times signe(x) \times |x|^z \tag{4.2}$$

Obtention de l'expression de la fonction notes étendues. La lecture de la démonstration est optionnelle et ne conditionne pas la compréhension de la suite.

Expression générale : Soit la fonction élastique y_{NE} définie comme la somme de deux monômes de degrés respectifs z ($z > 1, z \in \mathbb{R}$) et 1. On pose donc :

$$y_{NE}(x) = ax^z + bx \tag{4.3}$$

Afin que la fonction soit symétrique par rapport à l'origine, c'est-à-dire impaire, on transforme le monôme ax^z en produit de sa valeur absolue et de son signe.

$$y_{NE}(x) = a \times signe(x) \times |x|^z + bx \tag{4.4}$$

Calcul de a : Les conditions aux limites de l'expression sont :

$$\begin{cases} y_{NE}\left(-\frac{\delta}{2}\right) = a \times signe\left(-\frac{\delta}{2}\right) \times \left|-\frac{\delta}{2}\right|^z - b\frac{\delta}{2} = -\frac{\delta}{2} & (1) \\ y_{NE}\left(\frac{\delta}{2}\right) = a \times signe\left(\frac{\delta}{2}\right) \times \left|\frac{\delta}{2}\right|^z + b\frac{\delta}{2} = \frac{\delta}{2} & (2) \end{cases} \tag{4.5}$$

Soit en simplifiant l'équation (2) du système :

$$a\left(\frac{\delta}{2}\right)^z + b\frac{\delta}{2} = \frac{\delta}{2} \tag{4.6}$$

$$a = (1-b)\left(\frac{2}{\delta}\right)^{z-1} \tag{4.7}$$

Expression finale : En remplaçant a dans l'équation 4.4 on obtient finalement :

$$y_{NE}(x) = (1-b)\left(\frac{2}{\delta}\right)^{z-1} \times signe(x) \times |x|^z + bx \tag{4.8}$$

□

Fonction adaptative : *élastique*

Le but de la fonction *élastique* est de fournir une hauteur parfaitement juste quelque soit la hauteur d'entrée à l'instant de correction, tout en permettant des modulations de hauteur a posteriori de l'application. Ainsi celle-ci doit remplir les deux contraintes suivantes :

- Passe par le point $(x_0, 0)$ donnant une note juste pour la position initiale du stylet, tout en demeurant continue sur l'intervalle [-δ,δ] comme illustré en figure 4.3 - droite.
- Pour éviter un décalage persistant entre hauteurs d'entrée et de sortie, la fonction doit converger vers une relation linéaire lorsque les notes précédentes et suivantes sont atteintes. La fonction doit donc passer par les points $(-\delta, -\delta)$ et (δ, δ).

Pour remplir ces conditions, la fonction est définie comme un arc de courbure γ, dépendant de la hauteur initiale x_0. En respectant les conditions aux limites, la hauteur de sortie y_E définie sur $[-\delta, \delta]$ s'exprime en fonction de la hauteur d'entrée x sous la forme :

$$\begin{cases} y_E(x) &= \frac{1}{\gamma}\left[\log\left[(e^{2\gamma\delta}-1)(\frac{x}{\delta}+1)\frac{1}{2}+1\right]\right]-\delta & \text{si } \gamma \neq 0 \\ y_E(x) &= x & \text{si } \gamma = 0 \end{cases} \tag{4.9}$$

Pour chaque nouvel ajustement, la condition pour avoir une note juste à la hauteur initiale x_0 est $y_E(x_0, \gamma_0) = 0$, où γ_0 est la courbure à l'instant d'ajustement. Cela conduit à l'expression :

$$\gamma_0 = \frac{1}{\delta}\log\left(\frac{\delta - x_0}{\delta + x_0}\right) \tag{4.10}$$

Par la suite on choisira $\delta = 1$ amenant aux équations simplifiées :

$$\begin{cases} y_E(x) &= \frac{1}{\gamma}\left[\log\left[(e^{2\gamma}-1)^{\frac{(x+1)}{2}}+1\right]\right]-1 & \text{si } \gamma \neq 0 \\ y_E(x) &= x & \text{si } \gamma = 0 \end{cases} \tag{4.11}$$

$$\gamma_0 = \log\left(\frac{1 - x_0}{1 + x_0}\right) \tag{4.12}$$

Obtention de l'expression de la fonction élastique. La lecture de la démonstration est optionnelle et ne conditionne pas la compréhension de la suite.

Expression générale : Soit la fonction g définie comme un arc de courbure γ s'écrivant sous la forme :

$$g(y) = Ae^{\gamma(y+B)} + C \tag{4.13}$$

avec pour conditions aux limites :

$$\begin{cases} f(-\delta) &= Ae^{\gamma(-\delta+B)}+C &= -\delta & (1) \\ f(\delta) &= Ae^{\gamma(\delta+B)}+C &= \delta & (2) \end{cases} \tag{4.14}$$

Calcul de C : On soustrait et additionne les équations du système 4.14 :

$$\begin{cases} Ae^{\gamma B}\left[e^{\gamma\delta}+e^{-\gamma\delta}\right]+2C &= 0 & (1)+(2) \\ Ae^{\gamma B}\left[e^{\gamma\delta}-e^{-\gamma\delta}\right] &= 2\delta & (1)-(2) \end{cases} \tag{4.15}$$

On remplace :

$$\begin{aligned} C &= -\delta\tfrac{e^{\gamma\delta}+e^{-\gamma\delta}}{e^{\gamma\delta}-e^{-\gamma\delta}} \\ C &= -\delta\left[1+\tfrac{2e^{-\gamma\delta}}{e^{\gamma\delta}-e^{-\gamma\delta}}\right] \\ C &= -\delta\left[1+\frac{2}{e^{2\gamma\delta}-1}\right] \end{aligned} \tag{4.16}$$

Calcul de A : En remplaçant C dans (1) du système 4.14 on obtient :

$$A = 2\delta \frac{e^{\gamma(\delta - B)}}{e^{2\gamma\delta} - 1} \tag{4.17}$$

Expression finale : En remplaçant A et C dans l'équation 4.13

$$g(y) = 2\delta \frac{e^{\gamma(\delta - B)}}{e^{2\gamma\delta} - 1} e^{\gamma(y+B)} - \delta \left[1 + \frac{2}{e^{2\gamma\delta} - 1}\right]$$

$$\begin{cases} g(y) &= \delta \left[2\frac{e^{\gamma(y+\delta)} - 1}{e^{2\gamma\delta} - 1} - 1\right] & \text{si } \gamma \neq 0 \\ g(y) &= y & \text{si } \gamma = 0 \end{cases} \tag{4.18}$$

La fonction g obtenue est bijective sur $[-\delta, \delta]$. Pour simplifier les calculs d'obtention de la courbure γ, la fonction de déformation élastique y_E est définie comme l'inverse de g. On a alors :

$$\begin{cases} y_E(x) &= \frac{1}{\gamma}\left[\log\left[(e^{2\gamma\delta} - 1)(\frac{x}{\delta} + 1)\frac{1}{2} + 1\right]\right] - \delta & \text{si } \gamma \neq 0 \\ y_E(x) &= x & \text{si } \gamma = 0 \end{cases} \tag{4.19}$$

Calcul de la courbure pour une position donnée :

On cherche γ_0 tel que $y_E(x_0, \gamma_0) = 0$:

$$\frac{1}{\gamma_0}\left[\log\left[(e^{2\gamma_0\delta} - 1)(\frac{x_0}{\delta} + 1)\frac{1}{2} + 1\right]\right] - \delta = 0$$

En prenant la réciproque :

$$x_0 = \delta\left[2\frac{e^{\gamma_0\delta} - 1}{e^{2\gamma_0\delta} - 1} - 1\right]$$

On pose $u = e^{\gamma_0\delta}$:

$$x_0 = \delta\left[2\frac{u - 1}{u^2 - 1} - 1\right]$$

On obtient un polynôme du second degré en u.

$$\left(\frac{x_0}{\delta} + 1\right)u^2 - 2u + (1 - \frac{x_0}{\delta}) = 0$$

Le discriminant $\Delta = 4\frac{x_0^2}{\delta^2} > 0$ donc le polynôme possède deux racines réelles :

$$u = \frac{\delta \pm x_0}{\delta + x_0}$$

On cherche $\gamma_0 \neq 0$ donc $u \neq 1$. Finalement $u = \frac{\delta - x_0}{\delta + x_0}$ et :

$$\gamma_0 = \frac{1}{\delta}\log\left(\frac{\delta - x_0}{\delta + x_0}\right) \tag{4.20}$$

□

4.2.3 Dynamique de l'ajustement

Si les fonctions de déformation présentées ci-dessus étaient systématiquement appliquées, de sérieuses distorsions de hauteur apparaîtraient, particulièrement avec la fonction adaptative *élastique* qui fournirait une sortie discrète. Ainsi, différents degrés de liberté sont introduits pour appliquer la correction de hauteur en fonction de la dynamique de la hauteur d'entrée.

On identifie d'abord deux techniques de jeu pour lesquelles les techniques d'application de la correction sont différentes : un musicien peut soit jouer *staccato*, en relâchant le contact du stylet entre chaque note, soit jouer *legato*, en glissant le stylet sur la tablette sans relâcher le contact, afin d'enchainer les notes selon une trajectoire de hauteur continue. Dans le cas d'un jeu *staccato*, la correction est appliquée immédiatement et systématiquement à chaque contact. Cela permet de jouer juste immédiatement lors de l'attaque d'une note. On appellera par conséquent *correction d'attaque* cet ajustement. Il faut noter que ce type de correction ne permet pas de réaliser des attaques en faisant glisser la note par le bas ou par le haut. La fonction de déformation la plus adaptée pour une correction d'attaque est la fonction adaptative *élastique* qui propose une note juste quelle que soit la position d'attaque. On ne s'intéressera pas ici à une correction d'attaque de fonction *notes étendues*.

Dans le cas d'un jeu *legato*, on utilisera une correction dite *de contour*, qui doit ajuster la trajectoire de hauteur de manière lisse. Pour ce faire, la correction de contour est appliquée en fonction de la dynamique de la hauteur d'entrée en suivant les contraintes suivantes : elle doit être appliquée temporairement, seulement sur les notes cibles ; la correction doit être appliquée graduellement et de manière lisse pour éviter les sauts de hauteur ; la fin de l'application doit être synchronisée précisément avec le retour à une relation linéaire entre hauteurs d'entrée et de sortie. Les paramètres régissant le déclenchement, l'application et le retrait de la correction sont détaillés dans cette section.

Déclenchement

Le cas de déclenchement de la correction d'attaque est simple : à la détection d'un nouveau contact entre stylet et tablette, la correction est déclenchée immédiatement.

Dans le cas *legato*, la détection de notes à corriger est plus subtile. L'expressivité mélodique se traduit par des variations de hauteur. A l'inverse, les notes à corriger sont supposées stables. L'ajustement est donc appliqué uniquement pour des variations faibles de hauteur, et désactivé en cas contraire. La vitesse de la hauteur d'entrée permet donc de distinguer les notes à corriger (variations faibles non-intentionnelles) des variations expressives (variations intentionnelles). Prendre en compte la vitesse instantanée entraverait le jeu des vibratos, car tout changement de direction de la hauteur conduit à une vitesse nulle et déclencherait la correction. On définit alors la vitesse moyenne critique AV_c de la hauteur (*critical Average Velocity*) comme seuil au delà duquel l'ajustement ne sera pas appliqué. Celle-ci s'exprime comme le temps critique T_c nécessaire pour parcourir un intervalle I appelé intervalle de détection : $AV_c = I/T_c$. Ainsi, si la hauteur d'entrée reste dans un intervalle de détection I plus longtemps que T_c, la vitesse de la hauteur mélodique est inférieure à AV_c et l'ajustement est appliqué. Inversement, si la hauteur mélodique quitte l'intervalle I avant T_c, la vitesse de la hauteur est supérieure à AV_c et l'ajustement n'est pas réalisé. L'intervalle de détection I et le temps critique T_c sont donc des paramètres pour le réglage de la dynamique d'application de l'ajustement.

L'intervalle de détection I est obtenu en divisant l'axe des hauteurs d'entrées en un nombre entier d'intervalles par demi-ton. La taille de l'intervalle de détection est directement liée à la fréquence d'ajustement. La hauteur d'entrée est susceptible de changer d'intervalle de détection plus souvent lorsque ceux-ci sont petits (p. ex. 0.1 demi-ton) que lorsque ceux-ci sont grands (0.5 demi-ton). Un nombre plus important d'ajustements aura lieu avec de petits intervalles, favorisant la justesse. A l'inverse, de grands intervalles permettent plus de libertés de variations expressives. Cependant les variations non-intentionnelles au sein des intervalles ne sont pas corrigées.

Une fois l'intervalle de détection choisi, le temps critique permet d'ajuster le seuil de vitesse AV_c. Un faible vibrato d'amplitude ± 0.1 demi-ton et de fréquence 5 Hz observé en chant choral [Sun94] donne une vitesse moyenne $AV = 2$ demi-ton/s. Pour préserver les vibratos, la vitesse moyenne critique AV_c doit donc être inférieure. On peut imaginer des glissandos plus lents (p. ex. 1 demi-ton/s). Il faut alors diminuer la vitesse moyenne critique en augmentant le temps critique pour préserver les glissandos. Plus le temps critique est long plus les variations expressives lentes sont préservées, mais moins les ajustements sont réactifs. Le processus de déclenchement de la correction est représenté par les zones (1) de la figure 4.2.

Application

Lors d'un jeu *staccato*, la correction est appliquée immédiatement pour fournir une attaque juste au moment même du contact avec la tablette.

En revanche, dans le cas *legato*, l'ajustement doit s'appliquer graduellement pour ne pas introduire de sauts brusques dans la hauteur. La fonction de correspondance est donc modifiée continûment d'une relation linéaire (sans ajustement) vers la fonction voulue en un temps de transition prédéfini T_t. Cela se fait en modifiant continûment le degré de distorsion z de la fonction *notes étendues* ou la courbure γ de la fonction *élastique*. Si la hauteur d'entrée évolue pendant l'application de la correction, la hauteur de sortie est calculée selon la fonction de relation transitoire. Dans le cas *élastique*, si une nouvelle correction est déclenchée pendant l'application de la correction, le paramètre γ_0 est mis à jour immédiatement et le γ courant évolue alors vers ce nouveau paramètre.

Le temps de transition entre deux notes pour des chanteurs entraînés et non-entraînés semble presque constant quelque soit l'intervalle chanté [Sun73], [XS00], d'une valeur de 140 ms. Le temps de réponse d'un chœur entraîné à une variation de hauteur est aussi proche de 140 ms, en ne prenant pas en compte le temps de réaction des chanteurs [GST+09]. Par conséquent, 140 ms servent de référence au temps de transition. Un temps de transition plus long (p. ex. 250 ms) entraînera une application plus lisse, avec des transitions presque inaudibles. A l'inverse, un temps de transition court (p. ex. 50 ms), bien qu'audible permet une correction plus réactive. L'application de la correction est représentée par les zones (2) de la figure 4.2.

Retrait

Le maintien de l'ajustement après son application ne dépend pas de la technique de jeu (*staccato* ou *legato*) mais diffère selon la fonction de déformation choisie. L'aplatissement de la fonction *notes étendues* autour de la cible rend les variations de hauteur fines telles que le vibrato difficiles à produire. En effet, toute oscillation de la hauteur d'entrée dans la zone aplatie résulte en une hauteur de sortie constante. C'est pourquoi l'ajustement *notes étendues*

Technique	Staccato	Legato	
Correction	d'attaques	de contours	
Fonction	Elastique		Notes étendues
Déclenchement	Immédiat	Choix de la fréquence d'application avec I Choix de la réactivité avec T_c	
Application	Immédiat	Choix de la rapidité avec T_t	
Retrait	Quitte l'intervalle $[-\delta, \delta]$		Quitte l'intervalle I

Table 4.2 – *Résumé des dynamiques de la correction DPW en fonction des techniques de jeu et des fonctions de déformation.*

ne doit pas être maintenu durant un vibrato et est donc supprimé à la moindre variation de hauteur, c'est-à-dire lorsque celle-ci quitte l'intervalle de détection ayant déclenché l'ajustement. A l'inverse, la fonction *élastique* distord peu les modulations de hauteur. Elle peut donc être maintenue durant toute variation expressive. Elle est alors retirée uniquement en cas de changement de note, i.e. lorsque la hauteur quitte l'intervalle $[-\delta, \delta]$ autour de la note cible. Le retrait de la correction est représenté par les zones (4) de la figure 4.2.

Résumé et algorithmes

Pour conclure, la correction de contour de hauteur est appliquée dynamiquement selon trois degrés de liberté dirigés par trois paramètres : la fréquence de l'application de la correction donnée régie par l'intervalle de détection I ; la réactivité de la correction donnée par le temps critique T_c ; la rapidité de la correction commandée par T_t. La correction d'attaque est quant à elle appliquée immédiatement sans degrés de libertés. Les différentes dynamiques d'application sont résumées en tableau 4.2.

L'application de la correction DPW est résumée par les algorithmes 1 (déclenchement de la correction), 2 et 3 (calcul de la fonction de correspondance) et 4 (correction).

4.2.4 Préservation de l'expressivité

Le choix des paramètres de déclenchement I et T_c est primordial pour la préservation de l'expressivité du musicien. Pour illustrer le rôle de ces paramètres, différents exemples de valeurs de paramètres sont donnés et l'expressivité est analysée par observation de trajectoires contenant les trois principaux effets de modulation de hauteur : *portamento*, *glissando* et *vibrato*. Idéalement, l'ajustement DPW doit corriger la trajectoire tout en préservant les modulations expressives du musicien. Les rôles de la vitesse critique AV_c et de la vitesse de convergence CS/TS de la correction Haken sont similaires car ils décrivent le même degré de liberté : la préservation de l'expressivité. Cette dernière est conservée uniquement si ces valeurs sont inférieures à la vitesse moyenne des modulations à conserver. Par conséquent, la méthode DPW sera confrontée à la méthode Haken par les trois réglages d'ajustement présentés en table 4.3.

Les réglages A à C sont utilisés pour la correction DPW. Le réglage A est un ajustement "rapide" caractérisé par un petit intervalle de détection et un faible temps critique. A l'inverse le réglage B est un ajustement "lent". Son intervalle de détection plus grand et son long temps critique permettent plus de liberté de variations expressives. Le réglage C possède un petit intervalle de détection et un long temps critique. Un temps de transition $T_t = 50$ ms est utilisé pour les 3 réglages. Les réglages D à F sont utilisés pour la correction Haken. Le réglage D est donné en exemple dans [Hak09]. Ensuite, la vitesse de convergence CS/TS est réduite en augmentant TS (réglage E) ou en diminuant CS (réglage F).

Algorithme 1: Déclenchement de la correction DPW

Entrées : Hauteur d'entrée, I, T_c
Output : Signal de déclenchement

```
// Déclenche la correction (zone 1)
```
tant que *la correction DPW est activée* **faire**
  si *un nouveau contact est détecté* **et** *la correction d'attaques est active* **alors**
    Envoie le signal de déclenchement
  fin
  tant que *le stylet reste en contact avec la tablette* **et que** *la correction de contours est active* **faire**
    Initialise le chronomètre T;
    tant que *la hauteur d'entrée reste dans un intervalle de détection* I **faire**
      Incrémente un chronomètre T;
      si $T > T_c$ **alors**
        Envoie le signal de déclenchement;
      fin
    fin
  fin
fin

Algorithme 2: Calcule la fonction de correspondance dans le cas *notes étendues*

Entrées : Hauteur d'entrée, I, T_t, Signal de déclenchement
Output : Cible courante, Fonction de déformation

```
// Applique la correction (zone 2)
```
si *le signal de déclenchement est reçu* **alors**
  Définit la fonction de déformation courante y_{NE} avec l'équation 4.2 en changeant continûment la valeur courante z_T en z pendant la durée T_t;
fin

```
// Retrait de la correction (zone 4)
```
si *la hauteur d'entrée quitte l'intervalle de détection* I *ayant déclenché la correction* **alors**
  Applique immédiatement une relation linéaire entre les hauteurs d'entrée et de sortie ($z = 1$);
fin

Algorithme 3: Calcule la fonction de correspondance dans le cas *élastique*

Entrées : Hauteur d'entrée, T_t, Signal de déclenchement
Output : Cible courante, Fonction de déformation

`// Applique la correction (zone 2)`
si *le signal de déclenchement est reçu* **alors**
  Empêche toute modification de la note cible;
  Utilise la hauteur d'entrée initiale x_0 pour calculer la courbure γ_0 avec l'équation 4.12 pour obtenir une sortie juste;
  si *le déclenchement est dû à un nouveau contact* **alors**
    Calcule la nouvelle fonction de déformation y_E avec l'équation 4.11 en posant $\gamma = \gamma_0$
  sinon
    Calcule la nouvelle fonction de déformation y_E avec l'équation 4.11 en changeant continûment la valeur courante γ en γ_0 pendant la durée T_t;
  fin
fin

`// Retrait de la correction (zone 4)`
si *la hauteur d'entrée quitte l'intervalle* $[-\delta, \delta]$ *autour de la note cible* **alors**
  Définit la cible comme la note exacte la plus proche de la hauteur d'entrée;
  Applique immédiatement une relation linéaire entre les hauteurs d'entrée et de sortie ($\gamma = 0$);
fin

Algorithme 4: Correction de hauteur

Entrées : Hauteur d'entrée, Cible courante, Fonction de correspondance
Output : Hauteur de sortie

Calcule la hauteur d'entrée relativement à la note cible;
Calcule la hauteur de sortie depuis la hauteur d'entrée relative avec la fonction de correspondance actuelle;

DPW Correction			
	Intervalle de de détection I	**Temps critique T_c**	**Vitesse moyenne critique AV_c**
A	0.1 ST	0.1 s	1 ST/s
B	0.5 ST	0.25 s	2 ST/s
C	0.1 ST	0.25 s	0.4 ST/s
Haken Correction			
	Pas de correction CS	**Pas temporel TS**	**Vitesse de convergence CS/TS**
D	0.01 ST	0.005 s	2 ST/s
E	0.01 ST	0.025 s	0.4 ST/s
F	0.001 ST	0.005 s	0.2 ST/s

TABLE 4.3 – *Paramètres de déclenchement utilisés pour l'étude de l'expressivité.*

Ces ajustements sont appliqués à une même mélodie reflétant les trois modulations détaillées plus haut (portamento, glissando et vibrato). Les étapes de la mélodie sont : début sur un Do ; deux portamentos successifs vers les notes Do# et Ré ; vibrato ; glissando descendant vers la note Do. Les valeurs d'entrée et de sortie sont représentées en figure 4.4, en courbes pointillée et pleine respectivement. Les lignes pleines horizontales sont les notes cibles et les lignes pointillées horizontales sont les limites des intervalles de détection. Les mêmes couleurs d'arrière-plan que la figure 4.2 sont utilisées pour la représentation des zones 1 à 4 de la correction DPW. Les lettres A à F correspondent aux réglages présentés en tableau 4.3. Des exemples audios et vidéos de ces différentes corrections sont fournis à l'url en bas de page [6].

Déclenchement dynamique - DPW correction

Une première observation des réglages A, B et C montre que le portamento est bien ajusté pour chaque réglage, et chaque méthode d'ajustement. Seul le temps de réaction diffère, directement lié au temps critique (T_c = 100 ms pour le réglage A, et T_c = 250 ms pour les réglages B et C). En revanche, le vibrato est centré autour de la note cible uniquement par l'ajustement *élastique*. Cela est lié aux conditions de suppression de l'ajustement pour les deux méthodes (section 4.2.3). Dans le cas de la fonction *notes étendues*, l'ajustement est supprimé dès que la hauteur quitte l'intervalle de détection ayant déclenché l'ajustement. C'est pourquoi dans le cas de vibratos d'amplitudes supérieures à l'intervalle de détection, l'ajustement n'est pas conservé.

En haut figurent les résultats de l'ajustement rapide (réglage A). Les faibles valeurs d'intervalle de détection et de temps critique induisent un déclenchement fréquent de l'ajustement, principalement pendant les glissandos. A chaque changement d'intervalle de détection, la hauteur est ajustée en introduisant des distorsions. Dans le cas de la fonction *notes étendues* l'ajustement est successivement appliqué, puis retiré, et ainsi de suite, créant des pics dans la trajectoire. Dans le cas de la fonction *élastique*, l'ajustement est maintenu toute la durée du glissando, conduisant à l'apparition de paliers. Concernant le vibrato, l'ajustement est bien supprimé pour la fonction *notes étendues*, à 5.4 s. De plus, on observe de très légères distorsions au début et à la fin des vibratos pour chaque méthode. Cela ce produit lorsque l'amplitude du vibratos devient proche de la taille de l'intervalle de détection. Comme le temps critique T_c = 100 ms est inférieur à une période de vibrato (150-200 ms), la hauteur d'entrée reste suffisamment longtemps dans un intervalle de détection pour déclencher

6. https://perso.limsi.fr/operrotin/these.fr.php#Annexes (vérifié le 22 octobre 2015)

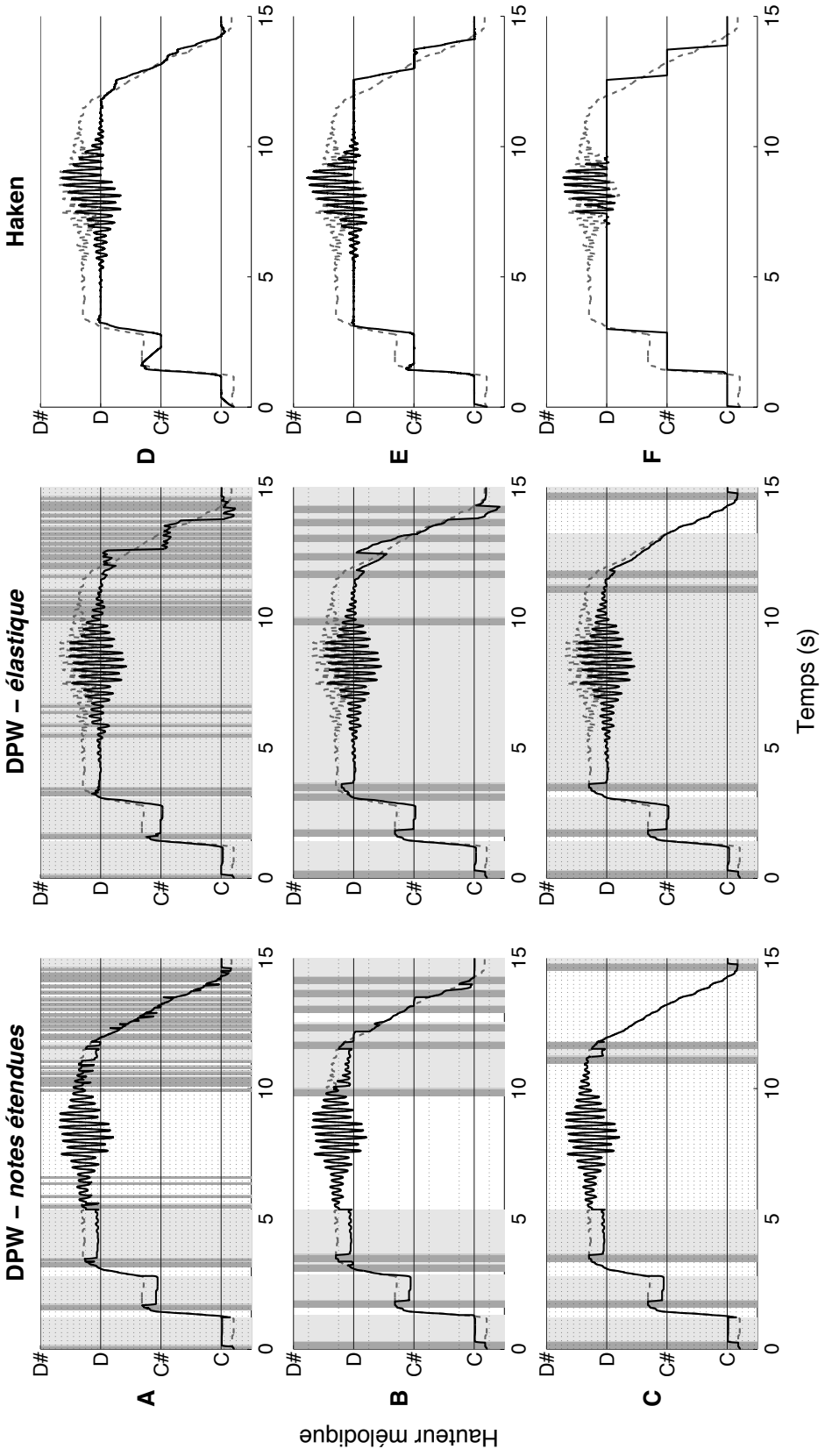

Figure 4.4 – *Exemples d'ajustements (gauche : DPW - notes étendues ; milieu : DPW - élastique) ; droite : Haken avec différents réglages (voir table 4.3). Les courbes en pointillés sont les trajectoires de la hauteur d'entrée. Les courbes pleines sont les trajectoires des hauteurs de sortie. Les lignes pleines sont les notes cibles. Les lignes pointillées représentent les intervalles de détection I (A et C : 0.1 ST, B : 0.5 ST).*

l'ajustement au sein même d'un cycle. Ces réglages sont donc à la limite de préservation des vibratos. Un temps critique plus court entraînerait des plus grandes distorsions. Finalement, si les ajustements sont trop rapides, ils deviennent intrusifs et empêchent certaines formes d'expressivité.

Les résultats de l'ajustement lent (réglage B) sont montrés au milieu de la figure 4.4. De même qu'avec le réglage A, l'ajustement *notes étendues* est à nouveau supprimé à 5.4 s pour le vibrato. Il est cependant intéressant de noter qu'après 10 s, le vibrato est inclus dans un intervalle de détection, et est donc ajusté. Des intervalles de détection plus larges sont donc nécessaires pour ajuster les vibratos avec la fonction *notes étendues*, aux dépens de distorsions introduites par la fonction de correspondance. A l'inverse, le vibrato est correctement ajusté sans introduction de distorsions avec la fonction *élastique*. Enfin, les paramètres d'ajustement induisent une vitesse moyenne critique de 2 ST/s, supérieure à la vitesse du glissando. Cela explique donc les distorsions observées.

Le réglage C est défini par le petit intervalle de détection du réglage A ($I = 0.1$ ST) et le long temps critique du réglage B ($T_c = 250$ ms). Le bas de la figure 4.4 montre les résultats des deux méthodes avec ce réglage. Aucune distorsion n'apparaît pour chacun des effets avec les deux méthodes. Comme précédemment, le vibrato n'est pas ajusté avec la méthode *notes étendues*, mais ce dernier n'est pas distordu.

La comparaison entre les réglages B et C indique que dans le cas de l'ajustement *élastique*, un petit intervalle de détection entraîne une meilleure détection des zones stables de la trajectoire mélodique. Cela permet à la fois un ajustement fréquent de la trajectoire et une préservation de toutes les formes expressives. A l'inverse, dans le cas de l'ajustement *notes étendues*, un grand intervalle de détection est nécessaire pour l'ajustement du vibrato, aux dépens de la fréquence d'ajustement.

La comparaison entre les réglages A et C montre qu'un temps critique long permet de mieux préserver les modulations expressives. Néanmoins, l'enchaînement des notes est lent. Pour des tempi plus rapides, un temps critique plus court est nécessaire, au détriment des modulations expressives (réglages A).

Déclenchement statique - Haken correction

Avec la correction Haken, chaque variation de hauteur d'entrée ayant une vitesse moyenne inférieure à la vitesse de convergence CS/TS est distordue. Il en résulte l'analyse suivante : le réglage D introduit une vitesse de convergence élevée ($CS/TS = 2$ demi-tons/s). Par conséquent, le glissando est transformé en marches d'escalier plus prononcées qu'avec le réglage A. Le vibrato est aussi distordu. Avec le réglage E, la vitesse de convergence $CS/TS = 0.4$ demi-tons/s est réduite pour correspondre à la vitesse moyenne critique du réglage C. Bien que les marches d'escalier n'apparaissent plus pour le glissando, une faible distorsion est toujours présente sur le vibrato. La vitesse de convergence CS/TS est à nouveau réduite à 0.2 demi-tons/s avec le réglage F. Les distorsions du vibrato sont moindres par rapport au réglage précédent, mais la vitesse de convergence devient trop lente pour corriger les portamentos aussi précisément que les réglages B ou C.

L'observation des réglages D, E et F met en évidence le compromis entre justesse (vitesse de convergence rapide) et expressivité (vitesse de convergence lente) introduit par la correction Haken. La comparaison entre les réglages C, E et F montre que la correction DPW évite ce compromis en choisissant quand appliquer la correction.

Conclusion

Pour conclure, notre méthode dynamique de déformation de hauteur est plus appropriée que la méthode Haken de convergence de hauteur en terme de préservation de l'expressivité. En effet, par son application dynamique, la méthode DPW introduit un minimum de distorsion sur les trajectoires mélodiques tout en corrigeant la justesse efficacement, sous réserve que des paramètres appropriés soient choisis (réglage C). Cependant, la fonction *notes étendues* introduit un compromis entre fréquence d'ajustement et correction et distorsion des vibratos. Elle et donc moins robuste que la fonction *élastique* pour la préservation de l'expressivité.

4.3 Evaluation de la correction

Les ajustements de justesse mélodique sont développés et testés dans le contexte du *Cantor Digitalis*. Dans cette étude, le contrôle de l'instrument est limité à la position du stylet sur la tablette ainsi qu'à la pression du stylet permettant d'ajouter du réalisme à la voix de synthèse en contrôlant l'effort vocal. Seul le contrôle de la hauteur mélodique est étudié ici. Cette dernière est modifiée continûment par la position horizontale du stylet sur la tablette. L'axe vertical est laissé libre pour contraindre au minimum le geste du musicien. Pour aider le joueur à viser les notes précisément, le support visuel du *Cantor Digitalis* montré en figure 2.4 est placé sur la surface de la tablette. On rappelle que bien que ce dernier soit inspiré d'un clavier de piano, les "touches" sont uniquement des indicateurs visuels de la note la plus proche. L'instrumentiste doit viser les lignes pour jouer juste. Une hauteur intermédiaire sera jouée dans le cas contraire.

4.3.1 Réduction de la difficulté

La loi de Fitts est un outil d'évaluation de la justesse et vitesse d'acquisition de cibles [Fit54]. L'indice de difficulté I_D d'une tâche de pointage est calculé comme le rapport logarithmique entre l'amplitude du mouvement A, soit la distance à parcourir jusqu'à la cible, et la largeur de la cible W (*width*), la tolérance autorisée autour du centre de la cible [Mac92] :

$$I_D = log_2\left(\frac{A}{W} + 1\right) \tag{4.21}$$

Ainsi, plus la cible et loin, plus la tâche est difficile, et plus la cible est large, plus la tâche est aisée. Cet indice est exprimé en bits et représente la quantité d'information portée par la tâche de pointage. Il a été largement utilisé dans la quantification de performances d'interfaces de pointage [CMR91].

Dans le cas de voyelles synthétiques, le seuil de perception de hauteurs différentes est d'environ 7 centièmes de demi-ton (environ 0.5 Hz à 120 Hz [FS58]). Il s'agit d'une largeur de cible auditive très faible. En musique occidentale où le plus petit intervalle sur une échelle musicale est le demi-ton, le plus petit mouvement possible entre deux hauteurs d'un demi-ton donne un indice de difficulté de $I_D = 3.9$ bits. Un saut d'octave a un indice de difficulté de $I_D = 7.4$ bits.

La fonction *notes étendues* élargit la cible par sa forme aplatie. En prenant par exemple une largeur de plateau de $W = 0.5$ demi-tons, on obtient alors un indice de difficulté $I_D = 1.6$ bit pour un déplacement d'un demi-ton et $I_D = 4.6$ pour un saut d'octave.

La fonction *élastique* étend quant à elle la largeur de la note cible à un demi-ton, car toute note est automatiquement corrigée à la note la plus proche au moment de la correction. En

prenant $W = 1$ on obtient alors un indice de difficulté $I_D = 1$ bit pour un déplacement d'un demi-ton et $I_D = 3.7$ pour un saut d'octave.

En guise de comparaison, un mouvement de curseur de 800 pixels ciblant un icône de 48 pixels de large a un indice de difficulté de $I_D = 4.1$ bits. Viser un point de 5 pixels de large dans un fichier texte depuis une même distance donne un indice de difficulté de $I_D = 7.3$ bits. Cela correspond à jouer une octave sans correction. Un saut d'octave avec la fonction *élastique* est équivalent à atteindre un icône de 64 pixels depuis une distance de 800 pixels.

La méthode DPW réduit donc bien la difficulté d'atteinte de hauteur précise. La fonction *élastique* permet une visée plus aisée car elle élargit la cible plus amplement que la fonction *notes étendues*.

4.3.2 Apport de justesse et précision en jeu staccato - correction d'attaques

Une première expérience est menée dans le but de quantifier l'apport de justesse et précision apportée par la correction DPW d'attaque, c'est-à-dire à chaque nouveau contact entre le stylet et la tablette. Seule la fonction *élastique* est considérée ici. Il s'agit d'une expérience d'imitation d'extraits musicaux dont le protocole est fortement inspiré de l'étude sur la justesse et la précision chironomique (chapitre 3). Les sujets doivent reproduire à la tablette les mélodies proposées avec et sans l'application de la correction, ainsi qu'avec et sans retour auditif. Il en résulte quatre conditions : une imitation *Chironomique corrigée* (CC) ; une imitation *Chironomique non corrigée* (C) ; une imitation *Chironomique muette corrigée* (CMC) et une imitation *Chironomique muette non corrigée* (CM).

Protocole

Les extraits musicaux présentés sont joués par un synthétiseur MIDI (Instrument *piano 1* du logiciel *SimpleSynth*[7]) produisant un son de piano synthétique et émis par le biais d'un casque DTX900 Beyerdynamic. Une partition comportant le nom des notes est affichée sur un écran placé devant les sujets. Les imitations chironomiques sont réalisées à l'aide du *Cantor Digitalis* équipé d'une tablette Wacom 4M de surface active 233×146 mm, présentant 1024 niveaux de pression du stylet, une résolution temporelle de 200 Hz et une résolution spatiale de 0.005 mm. Contrairement à la configuration actuelle du *Cantor Digitalis*, chaque demi-ton est séparé de 1.25 cm. Cela fournit donc une résolution de hauteur de 0.04 centièmes de demi-tons, ce qui est grandement inférieur au seuil de perception de hauteur différentes pour des voyelles synthétiques (environ 0.5 Hz à 120 Hz, soit 7 cents [FS58]). La position horizontale du stylet contrôlant la hauteur est enregistrée, ainsi que le son produit par les imitations chironomiques provenant du moteur de synthèse du *Cantor Digitalis*, fixé sur une voyelle /a/. Un métronome indique le tempo de manière visuelle (sur l'écran) et auditive.

Les mélodies utilisées pour cette expérience sont inspirées des basses d'Alberti. Habituellement présentes dans les accompagnements de piano de l'époque classique, ces motifs enchaînent les notes d'une triade majeure dans l'ordre suivant : basse, aigu, intermédiaire, aigu et ainsi de suite. Construites sur Do majeur (Do, Sol, Mi, Sol, Do), les notes intermédiaires et aiguës sont transposées pour obtenir un intervalle de taille variable entre la 1^{e} et la 2^{e} note et des tierces fixes entre les 2, 3 et 4^{e} notes. Tous ces motifs ont été transposés une quinte au-dessus pour doubler le nombre de motifs présentés en figure 4.5.

7. http://simplesynth.sourceforge.net (vérifié le 22 octobre 2015)

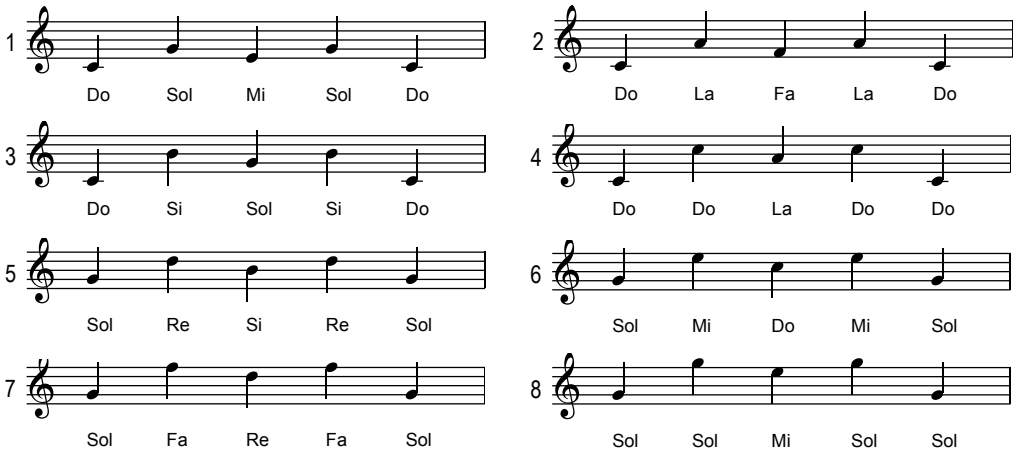

Figure 4.5 – *Motifs à imiter pour l'évaluation de la justesse et précision de la correction DPW d'attaque.*

Le tempo imposé aux sujets et de 120 b.p.m. Chaque stimulus est joué 5 fois par chaque sujet dans chaque condition (C, CC, CM, CMC). Au total, le bloc de stimuli est constitué de 4 motifs joués 5 fois sous 4 conditions.

Les stimuli sont présentés aux sujets aléatoirement et peuvent être écoutés autant de fois que nécessaire. L'interface utilisée dans l'étude précédente (chapitre 3) a été mise à jour et réutilisée ici. Pour chaque tâche le sujet commence par écouter l'extrait mélodique, la partition est affichée à l'écran et le tempo est donné visuellement et auditivement. La condition de correction appliquée ne lui est pas indiquée. Le sujet doit alors reproduire la mélodie à la tablette en effectuant un nouveau contact pour chaque note. La fin de l'essai s'effectue en appuyant sur la touche espace. Chaque sujet a été informé au préalable des modalités de l'expérience et a réalisé une session d'entraînement, présentant le même protocole que l'expérience mais des motifs différents.

7 sujets (28 ans et 11 années d'expérience musicale en moyenne) ont participé à l'expérience. Aucun des sujets ne présente de déficience auditive et tous sont droitiers. Chaque sujet a été informé des modalités de l'expérience et a réalisé une session d'entraînement au préalable, présentant les mêmes stimuli et protocole que l'expérience.

Analyse

Pour chaque essai, les notes jouées par les sujets sont identifiées à la fois sur les hauteurs d'entrée et de sortie comme les trajectoires de pression non-nulles entre les différents contacts. Selon la méthode présentée en section 3.2.5, deux valeurs sont extraites pour chaque note :

- La valeur au contact : extraite de la hauteur de sortie immédiatement corrigée après l'application de l'ajustement *staccato*.
- La valeur stable : extraite de la hauteur de sortie après stabilisation de la hauteur d'entrée (la note entendue par le sujet).

L'algorithme ajuste la hauteur correctement seulement si celle-ci est comprise dans un intervalle de $[-0.5, 0.5]$ autour de la cible. Dans le cas contraire, la hauteur sera ajustée vers

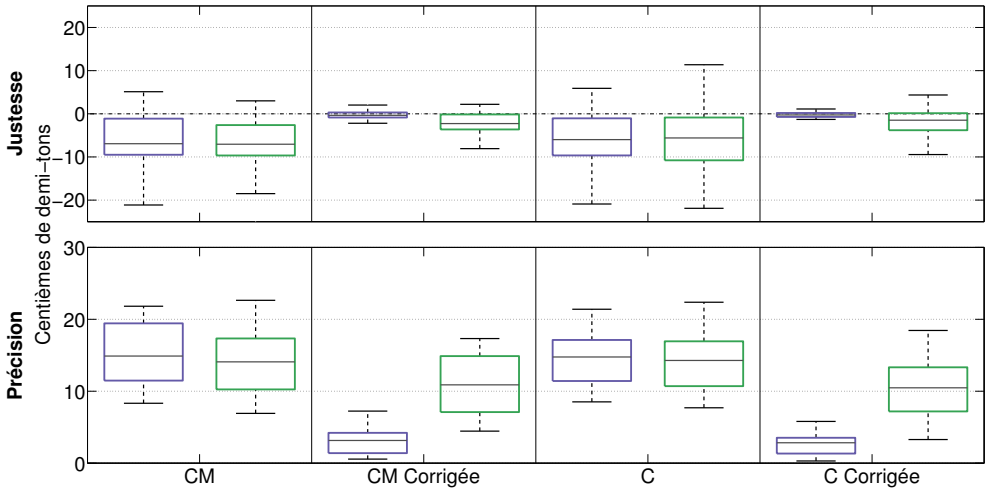

FIGURE 4.6 – *Justesse (haut) et précision (bas) des sujets en fonction de la condition d'ajustement : Chironomie Muette, Chironomie Muette Corrigée, Chironomie, et Chironomie Corrigée. Pour chaque condition la boîte de gauche en violet (resp. droite en vert) contient les valeurs de contact (resp. stable) de chaque réalisation.*

une note adjacente qui n'est pas celle attendue. Pour étudier l'efficacité de l'ajustement dans le seul cas où celui-ci se fait vers la note cible, les erreurs supérieures à un demi-ton sont écartées. L'analyse de la performance des ajustements repose sur les notions de justesse et précision dont les expressions sont fournies en section 3.2.5. Seules les justesses et précisions de notes sont étudiées ici.

Résultats

La figure 4.6 montre la justesse (haut) et la précision (bas) exprimées en centièmes de demi-tons (cents) des sujets dans les 4 conditions d'imitation. Pour chaque condition, la boîte de gauche contient les justesses (resp. précisions) des valeurs de contact de chaque réalisation de chaque sujet, et la boîte de droite contient les justesses (resp. précisions) des valeurs stables de chaque réalisation de chaque sujet. Les boîtes contiennent 50% des valeurs et les lignes représentent les médianes. Les différences statistiques entre justesse et précision parmi les différents ajustements sont étudiées avec un test de Wilcoxon par paires depuis l'environnement R [tea13].

On note d'abord que les valeurs de justesse et précision sont d'un ordre de grandeur semblable à celles observées au chapitre 3 (voir figure 3.5 encarts I-A et I-C). De plus, les résultats présentent une dispersion globale relativement faible, reflétant à nouveau l'habilité des sujets à manier un stylet acquise dès le plus jeune âge (voir discussion en section 3.4.2). D'ailleurs, aucune différence significative n'est notable entre les conditions avec et sans retour auditif. Par la suite, on comparera uniquement les résultats avec retour auditif.

Les boîtes de gauche de chaque panneau montrent les justesses et précisions des sujets à l'instant du contact. Bien que théoriquement la justesse attendue au contact après correction doit être de valeur nulle, l'implémentation en temps-réel du système introduit des artefacts. Malgré cela, l'amélioration de la justesse est spectaculaire, passant d'environ -7 cents pour la condition non corrigée à moins d'un cent avec la correction ($W = 273$, $p < 0.01$). Il en

est de même pour les précisions, améliorées significativement par la correction ($W = 1224$, $p < 0.01$). Par ailleurs, les valeurs de justesse et précision obtenues à l'instant de contact sont toutes inférieures au seuil de perception des hauteurs (7 cents pour les voyelles synthétiques). La correction est donc bien efficace à cet instant.

Les boîtes de droite de chaque panneau montrent les justesses et précisions des sujets après le contact, lorsque la hauteur est stabilisée sur la tablette. C'est la hauteur perçue. Celles-ci diffèrent des hauteurs d'entrée car le stylet glisse légèrement sur la tablette après du contact. Les valeurs de justesses des notes stables et corrigées sont légèrement mais significativement plus élevées que les justesses au contact ($W = 871$, $p < 0.01$). Elles restent cependant significativement inférieures aux justesses non corrigées ($W = 344$, $p < 0.01$). Des comportements similaires sont observés pour la précision, dégradée après le contact ($W = 45$, $p < 0.01$) mais de valeurs plus faibles que sans correction ($W = 915$, $p < 0.01$).

Finalement, cette expérience montre l'efficacité remarquable de l'ajustement d'attaque à l'instant de contact, produisant une attaque de son parfaitement juste. Néanmoins, le stylet dévie parfois de sa position initiale et produit une note légèrement fausse après stabilisation. C'est là qu'intervient la correction de contour, capable de rectifier la trajectoire du stylet après le contact.

4.3.3 Apport de justesse et précision en jeu legato - correction de contours

La justesse apportée par les ajustements est quantifiée dans un deuxième temps sur une technique de jeu *legato*, où les notes sont sélectionnées par le tracé d'un contour continu sur la tablette. Il s'agit à nouveau d'une expérience d'imitation mais les modalités de retour audio ne sont pas observées ici. Les deux fonctions *élastique* et *notes étendues* de la correction DPW sont proposées selon deux réglages : un ajustement rapide ($I = 0.1$ ST ; $T_c = 0.1$ s ; $T_t = 0.05$ s) et un ajustement lent ($I = 0.5$ ST ; $T_c = 0.25$ s ; $T_t = 0.05$ s). Ils correspondent aux réglages A et B de la table 4.3. Ces quatre conditions ainsi qu'une condition "sans ajustement" sont proposées pour chaque motif mélodique.

Protocole

Un protocole similaire à l'expérience précédente sur la correction d'attaque (section 4.3.2) est suivi. Toutefois, la tablette utilisée est une Wacom Intuos 5M, équipée du calque décrit en section 2.2.2. Un demi-ton correspond à 6 mm sur la tablette, et 35 demi-tons sont accessibles, de sol# à sol, fournissant donc une résolution de hauteur de 0.08 centièmes de demi-tons, ce qui est toujours inférieur au seuil de perception de hauteur différentes. La position du stylet est enregistrée, ainsi que le son produit par les imitations chironomiques provenant du moteur de synthèse du Cantor Digitalis. Un métronome indique le tempo de manière visuelle (sur l'écran) et auditive.

Les mélodies proposées dans cette expérience sont 4 motifs de 9 notes inspirés d'exercices vocaux et sont présentées en figure 4.7. Deux tempi différents sont utilisés : 120 et 240 battements par minute (b.p.m.) et chaque stimulus est joué 3 fois par chaque sujet pour chaque condition. Au total, chaque sujet doit jouer 3 fois 4 mélodies sous 5 conditions d'ajustement pour 2 tempi.

Les stimuli sont présentés aux sujets aléatoirement et sont écoutés autant de fois que nécessaire. La même interface que précédemment est utilisée. Pour chaque tâche le sujet

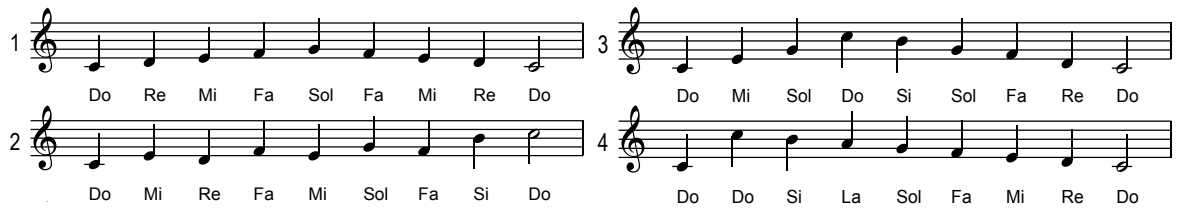

FIGURE 4.7 – *Motifs à imiter pour l'évaluation de la justesse et précision de la correction DPW - legato.*

commence par écouter l'extrait mélodique, la partition est affichée à l'écran et le tempo est donné visuellement et auditivement. La condition de correction appliquée ne lui est pas indiquée. Le sujet doit alors reproduire la mélodie à la tablette sans rompre le contact entre stylet et surface de la tablette pendant la durée d'un essai. La fin de l'essai s'effectue en appuyant sur la touche espace. Chaque sujet a été informé au préalable des modalités de l'expérience et a réalisé une session d'entraînement, présentant le même protocole et les mêmes motifs que pour l'expérience.

10 sujets (29 ans et 12 années d'expérience musicale en moyenne) ont participé à l'expérience. On distingue trois groupes de sujets : les "Non-musiciens", trois sujets n'ayant aucune expérience musicale ; les "Musiciens", quatre sujets avec plusieurs années de pratique instrumentale ; les "joueurs de *Cantor Digitalis*", trois sujets avec plus de 10 ans de pratique musicale et jouant régulièrement du *Cantor Digitalis* (environ 50h de pratique). Aucun des sujets n'a reporté de déficience auditive et tous sont droitiers. Chacun a été informé des modalités de l'expérience et a réalisé une session d'entraînement au préalable, présentant les mêmes stimuli et protocole que l'expérience.

Analyse

Pour chaque essai, les notes jouées par les sujets ont été identifiées à la fois sur les hauteurs d'entrée et de sortie comme les parties stables entre les transitions inter-notes (pics dans la dérivée), selon la méthode présentée en section 3.2.5. Pour chaque note, deux valeurs sont extraites :

- La valeur d'entrée : extraite de la hauteur d'entrée avant l'application de l'ajustement (la note jouée lorsque le stylet est stable).
- La valeur de sortie : extraite de la hauteur de sortie après une éventuelle application de l'ajustement (la note entendue par le sujet).

Dans le cas où l'ajustement est attendu mais pas réalisé, les notes concernées sont retirées et le pourcentage des notes restantes est discuté plus bas.

L'analyse de la performance des ajustements repose sur les notions de justesse et précision dont les expressions sont fournies en section 3.2.5. Seules les justesses et précisions de notes seront étudiées ici. Chaque sujet a joué l'ensemble des stimuli 3 fois. Par conséquent, chaque réalisation de l'ensemble des stimuli sera considérée séparément pour donner un couple de valeurs justesse/précision par sujet et par réalisation.

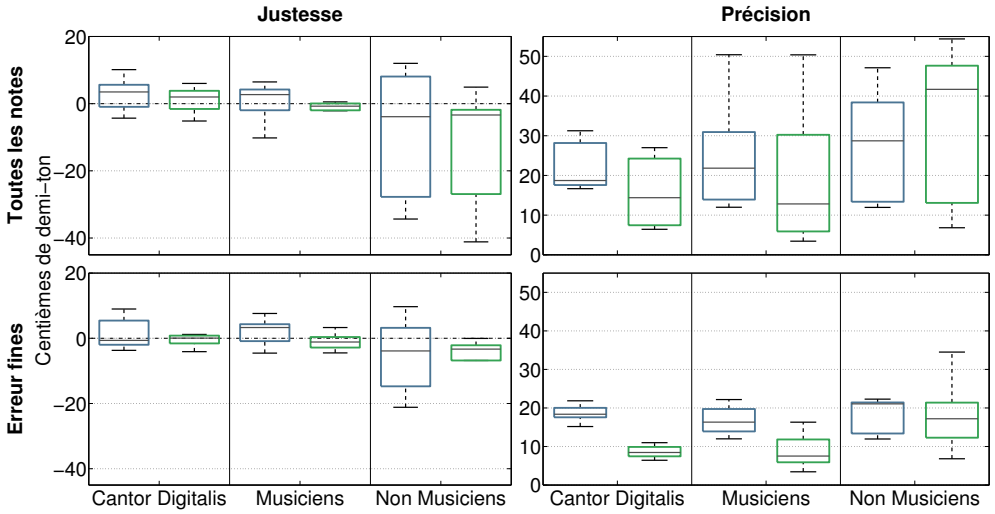

Figure 4.8 – *Justesse (gauche) et précision (droite) exprimées en centièmes de demi-tons pour chaque groupe de sujets en considérant toutes les notes (haut) ou les notes jouées avec une erreur ≤ 0.5 demi-tons (bas). Pour chaque condition, la boîte de gauche en bleu (resp. droite en vert) contient les valeurs d'entrée (resp. de sortie).*

Effet des sujets et de l'entraînement

La correction ajuste la hauteur d'entrée vers la note exacte la plus proche. Par conséquent, l'ajustement n'est efficace seulement si la hauteur d'entrée est plus proche de la note cible que des notes adjacentes, autrement dit si l'erreur est inférieure à 0.5 demi-ton dans notre cas. Lorsque l'erreur dépasse ce seuil, la hauteur est ajustée vers une mauvaise note et amplifie alors l'erreur initiale. On appellera *erreurs fines* les erreurs inférieures à 0.5 demi-tons, et *erreurs grossières* les erreurs supérieures à 0.5 demi-tons et entraînant une mauvaise correction.

La figure 4.8 montre la justesse (gauche) et la précision (droite) des sujets dans deux cas différents : lorsque toutes les notes jouées par les sujets sont considérées (haut) ; lorsque seulement les notes jouées avec une erreur fine sont prises en compte (bas). Les trois groupes de sujets sont représentés par chaque panneau. Chaque panneau contient deux boîtes contenant les valeurs d'entrée (gauche) ou les valeurs de sortie (droite). Les différences statistiques entre les répartitions des valeurs de justesse et précisions parmi les groupes sont étudiées par un test de Wilcoxon par paires depuis l'environnement R [tea13].

On observe des valeurs de justesse et de précision d'entrée légèrement supérieures à celle obtenues dans l'étude sur la justesse et précision du jeu chironomique (figure 3.5, encarts I-A et I-C). Il faut prendre en compte la difficulté plus importante de la tâche demandée ici : les mélodies sont plus longues (9 notes contre 2, 6 ou 7 notes précédemment), et les tempi plus rapides (120 et 240 b.p.m. ici contre seulement 120 b.p.m. précédemment). Aucune amélioration significative n'apparaît entre valeurs d'entrée et de sortie en prenant en compte toutes les notes (haut). De plus, moins les sujets ont d'expérience musicale, moins ils sont précis. Une dégradation marginalement significative des précisions de sortie est observée entre les joueurs de *Cantor Digitalis* et les Non-musiciens ($W = 62, p = 0.06$). Inversement, dans le cas des notes ayant une erreur fine (bas), la correction améliore significativement les précisions de sortie pour les joueurs de *Cantor Digitalis* ($W = 81, p < 0.01$) et les Musiciens

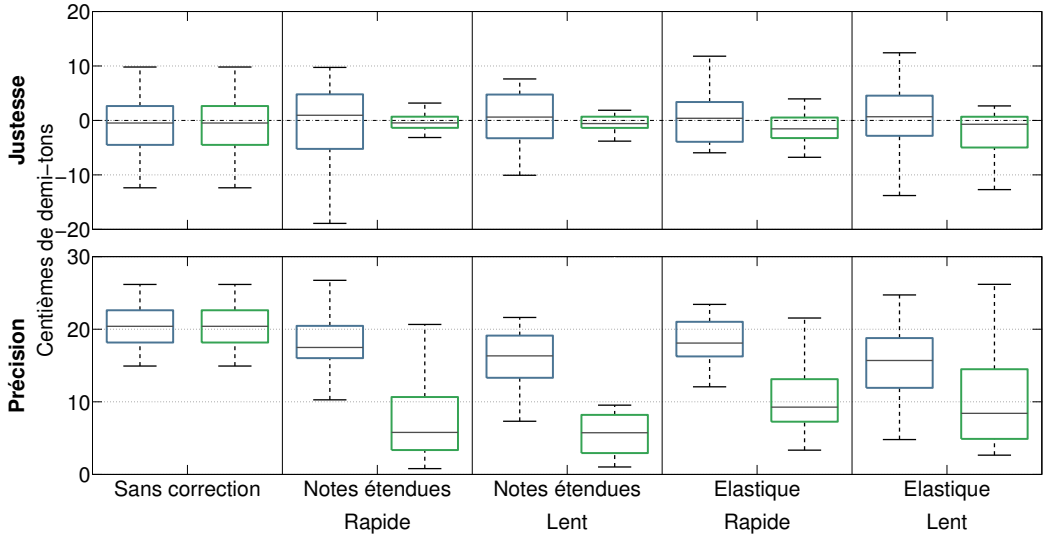

FIGURE 4.9 – *Justesse (haut) et précision (bas) des sujets en fonction de la condition d'ajustement. Pour chaque condition la boîte de gauche en bleu (resp. droite en vert) contient les valeurs d'entrée (resp. sortie) de chaque réalisation.*

($W = 135, p < 0.01$). Les valeurs de précisions de sortie ne sont pas améliorées de manière significative pour les Non-Musiciens. Ces observations mènent à deux conclusions :

- Les Musiciens et les joueurs de *Cantor Digitalis* sont plus enclins à atteindre le minimum de précision requis pour déclencher une correction efficace (erreur inférieure à 0.5 demi-ton dans notre cas).
- Seuls les Musiciens et les joueurs de *Cantor Digitalis* obtiennent une amélioration significative de leur précision après correction sur les notes bien ciblées. En effet, la correction nécessite une certaine stabilité de la hauteur durant son application pour être efficace, condition qui pourrait ne pas être remplie par les Non-musiciens.

Pour l'étude de l'effet de la correction et du tempo, seuls les notes avec une erreur fine sont conservées afin de n'observer que les effets désirables de la correction.

Effet des fonctions et réglages

On étudie l'effet des fonctions de déformation et des réglages par le calcul de justesses et précisions sur l'ensemble de notes *Correction* contenant toutes les notes jouées pour chaque condition de correction. On considère alors 3 facteurs : le facteur "sujet" (10 niveaux) ; le facteur "essai" (3 niveaux) et le facteur "correction" (5 niveaux : sans correction, correction *notes étendues* rapide, correction *notes étendues* lente, correction *élastique* rapide, correction *élastique* lente) conduisant à 150 mesures de justesse et précision.

La figure 4.9 montre la justesse (haut) et la précision (bas) exprimées en centièmes de demi-tons (cents) des sujets dans les 5 conditions d'ajustement. Pour chaque condition, la boîte de gauche contient les justesses (resp. précisions) des valeurs d'entrée de chaque réalisation de chaque sujet, et la boîte de droite contient les justesses (resp. précisions) des valeurs de sortie de chaque réalisation de chaque sujet. Les différences statistiques entre justesse et précision parmi les différents ajustements sont étudiées avec un test de Wilcoxon par paires.

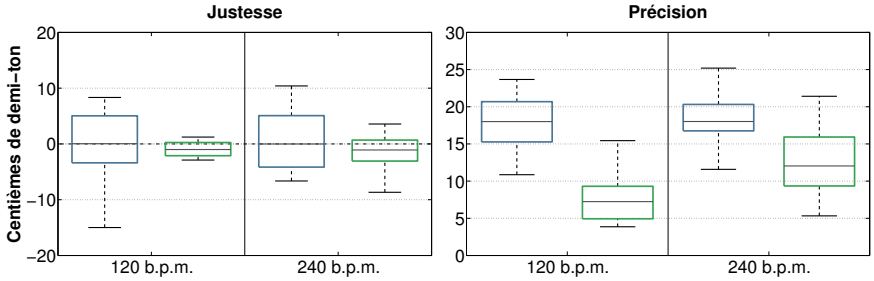

Figure 4.10 – *Justesse (gauche) et précision (droite) des sujets en fonction du tempo. Pour chaque condition la boîte de gauche en bleu (resp. droite en vert) contient les valeurs d'entrée (resp. sortie) de chaque réalisation.*

Les justesses obtenues pour chaque condition ont toutes une médiane proche de 0 et une dispersion inférieure à 5 cents. Les valeurs d'entrées des joueurs étant déjà très justes, la justesse n'est pas significativement améliorée pour aucun des ajustements. Les médianes des précisions des valeurs d'entrée sont situées entre 15 et 20 cents. Cela correspond à une distance d'environ 1 mm sur la tablette, de l'ordre de grandeur de la largeur de la pointe du stylet. On a donc une précision proche de la limite imposée par la tâche, déjà observée dans l'étude précédente (section 3.3.2). Toutes les précisions des valeurs de sortie sont en dessous de 15 cents et sont significativement plus faibles pour chacun des ajustements : *notes étendues* rapide ($W = 845, p < 0.001$) ; *notes étendues* lent ($W = 761, p < 0.001$) ; *élastique* rapide ($W = 804, p < 0.001$) ; *élastique* lent ($W = 664, p < 0.01$).

On observe ensuite un effet de l'ajustement sur les précisions des valeurs de sortie. La valeur de précision fournie par l'ajustement *notes étendues* rapide (resp. lent) est significativement plus faible que la valeur de précision fournie par l'ajustement *élastique* rapide ($W = 242, p < 0.01$) (resp. lent ($W = 290, p < 0.05$)). L'ajustement *notes étendues* définit une zone autour de la note cible dans laquelle toutes les positions d'entrées produisent cette cible. L'ajustement *élastique* déplace la cible vers la position d'entrée actuelle, mais autorise de faibles déviations autour de la position corrigée. Par conséquent ce dernier est plus sensible aux faibles mouvements et conduit à une correction moins stable.

Effet du tempo

Les justesses et précisions sont calculées ici sur l'ensemble *Tempo* constitué de trois facteurs. L'interaction des facteurs "sujet" (10 niveaux), "essais" (3 niveaux) et "tempi" (2 niveaux : 120 et 240 b.p.m.) entraîne 60 mesures de justesse et précision. De plus, l'ensemble *Tempo* ne contient que les stimuli avec correction. La figure 4.10 montre l'effet du tempo sur l'efficacité de la correction. Bien que l'effet ne soit pas visible sur les justesses, il est significatif sur les précisions : la médiane des valeurs de sorties et 5 cents plus élevée lorsque le tempo est doublé ($W = 256, p < 0.01$). Cela est dû à une trajectoire moins stable du stylet à des tempi élevés, mais qui n'empêche pas les sujets de viser les notes précisément.

Il est intéressant de regarder quand la correction est vraiment appliquée. Le nombre de notes corrigées est en effet plus faible que le nombre de notes jouées, toutes conditions confondues, comme indiqué en table 4.4. Seulement la moitié des notes sont corrigées au total. La correction rapide a un temps critique de $T_c = 100$ ms et un temps de transition $T_t = 50$ ms, ce qui rend la correction efficace après 150 ms. La correction lente est effective après 250

	Correction rapide	**Correction lente**	**Total**
120 b.p.m.	87	66	76
240 b.p.m.	42	11	26
Total	64	38	51

TABLE 4.4 – *Pourcentage des notes corrigées selon la vitesse de correction et du tempo.*

ms. Un tempo de 240 b.p.m. donne une pulsation toutes les 250 ms. Cela laisse donc peu de temps pour des ajustements après stabilisation de la note. Par conséquent, les réglages définissent un tempo limite au-delà duquel la correction est susceptible de ne pas s'appliquer. De plus, certains sujets n'ont jamais joué de notes suffisamment stables pour permettre à la correction d'être déclenchée.

Pour conclure, les deux méthodes d'ajustement améliorent significativement la précision des joueurs, avec de meilleures performances pour l'ajustement *notes étendues*. Néanmoins, la précision des valeurs de sortie est proche de la limite de perception de la hauteur dans tous les cas. Malgré ces performances, l'ajustement DPW a deux limites : un minimum de précision et de stabilité est requis pour que l'ajustement soit appliqué efficacement. Le joueur doit viser chaque note avec une erreur inférieure à 0.5 demi-tons, ce qui demande un minimum d'expérience avec l'interface. De plus, la stabilité de la hauteur pendant la phase de correction ne peut être obtenue seulement si le joueur a des intentions musicales (et n'accomplit pas uniquement une tâche de pointage), et si le tempo est suffisamment faible pour laisser du temps à la correction de s'appliquer.

4.3.4 Etude perceptive

L'évaluation objective de la justesse et précision apportée par la correction a montré des performances très satisfaisantes. Bien qu'il ait été montré que des mesures objectives de justesse de hauteur sont fortement corrélées à la perception [LMLS+13], il semble tout de même important d'étudier la pertinence de la correction d'un point de vue auditif. Des tests perceptifs sont conduits afin de voir si la contribution de la correction est perceptible.

Protocole

Un paradigme MOS (*Mean Opinion Score)* est utilisé pour l'évaluation perceptive de la correction. L'expérience consiste à écouter les enregistrements de stimuli joués dans différentes conditions, et de noter leur justesse sur échelle MOS de 1 (pauvre) à 5 (excellente). Il est demandé aux sujets de ne considérer que la justesse de jeu, et non les autres aspects du son entendu. Le matériel audio utilisé pour cette expérience est constitué des stimuli enregistrés lors de l'évaluation objective de la justesse. Un sujet a été écarté car la faible musicalité dont il a fait preuve aurait probablement interféré avec la perception de justesse. Afin de réduire le nombre de stimuli, seuls ceux utilisant la fonction *élastique* ont été utilisés, cette dernière corrigeant légèrement moins efficacement que la fonction *notes étendues*. Seuls les 1e essais de chaque stimulus ont été conservés, conduisant à 216 stimuli (9 sujets, 2 tempi, 4 mélodies et 3 corrections : sans correction, *élastique* rapide et *élastique* lent). Les stimuli sont présentés aléatoirement à travers deux enceintes amplifiées Genelec dans une salle acoustique traitée et insonorisée.

9 sujets (moyenne 33 ans) ont participé à l'expérience. Tous ont un parcours musical (moyenne 21 ans) et aucun n'a rapporté de déficience auditive. Trois ont aussi participé à

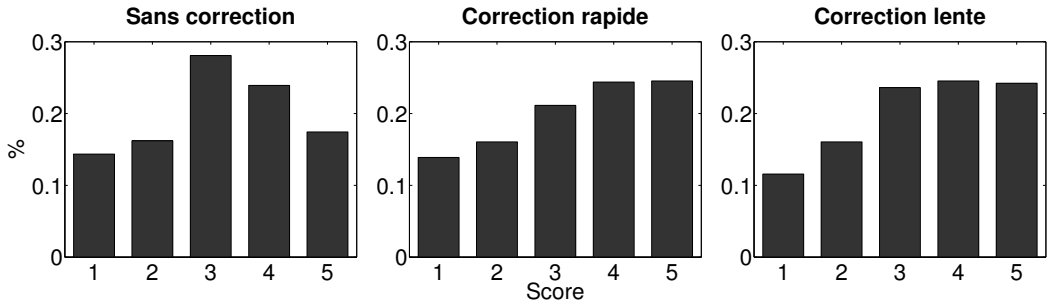

Figure 4.11 – *Résultats des MOS des auditeurs en fonction des trois conditions de correction (sans correction, élastique rapide, élastique lent).*

l'évaluation objective. Pour distinguer les sujets des expériences objectives et perceptives, les premiers sont appelés *joueurs* et les deuxième *auditeurs*. Tous les auditeurs ont été informé de la tâche demandée et ont réalisé une session d'entraînement présentant 20 des stimuli sélectionnés représentatifs de la gamme des différentes justesses rencontrées.

Résultats

Les MOS des auditeurs sont représentés en figure 4.11 en fonction des trois types de correction. Les panneaux de gauche, milieu et droite représentent respectivement la répartition des scores pour les conditions sans correction, correction *élastique* rapide et correction *élastique* lente. Des différences significatives apparaissent entre les distributions sans correction et correction rapide ($W = 35686, p < 0.01$) ainsi qu'entre les distributions sans correction et correction lente ($W = 35600.5, p < 0.01$). Aucune différence significative n'est observée entre les deux corrections.

Ces mesures montrent que les auditeurs tendent à attribuer de meilleurs scores aux stimuli avec correction qu'aux stimuli sans. Ce résultat s'accorde donc avec les mesures objectives de justesse et précision réalisées en section 4.3.3, démontrant donc que la correction DPW améliore perceptivement les justesse et précision de hauteur.

4.4 Discussion et conclusion

4.4.1 Justesse et expressivité

Jouer d'un instrument de musique numérique est plus simple avec l'aide d'une correction automatique de hauteur. Une analyse des méthodes de correction existantes a mis en évidence deux catégories d'algorithmes : les corrections par convergence de hauteur où cette dernière est constamment ajustée vers la note cible la plus proche, et les corrections par déformation de hauteur où celle-ci est calculée par une fonction de la hauteur d'entrée. Il est alors possible d'appliquer ces correction de manière statique, c'est-à-dire avec de manière similaire quelque soit l'instant, ou de manière dynamique, en fonction de la trajectoire de la hauteur.

La correction *Dynamic Pitch Warping* ou DPW est une méthode de déformation de hauteur dynamique, qui améliore la justesse et la précision sans altérer les modulations expressives du musicien. Celle-ci a été proposée avec deux types de fonctions de déformations : une fonction fixe appelée *notes étendues* et une fonction adaptative calculée pour chaque hauteur

d'entrée appelée *élastique*. Enfin, celle-ci a été appliquée dans deux conditions différentes : dans un jeu *staccato*, où un contact est effectué pour chaque nouvelle note ; dans un jeu *legato* où le changement de note est à identifier dans un contour continu de hauteur.

La correction DPW d'attaques est pleinement efficace pour le jeu *staccato* car elle ajuste immédiatement la hauteur au contact fournissant une attaque juste, et est supprimée au changement de note suivant. L'application de la correction DPW de contours pour un jeu *legato* est plus subtile. La correction doit être déclenchée sur les notes stables et retirée pendant les modulations expressives entre les notes. Cela suppose que des notes stables sont jouées. De plus les notes doivent être atteintes avec une précision minimale pour ne pas déclencher de fausses corrections. De par ces conditions, la correction s'adresse à des joueurs ayant une intention musicale mélodique. Il a été montré que chez les musiciens, la correction permet d'atteindre des justesses et précisions proches ou en dessous du seuil de perception, et inférieures à la limite imposée par la largeur de la pointe du stylet. Des résultats similaires ont été obtenues pour les fonctions *élastique* et *notes étendues*.

Quant aux modulations expressives, le choix des paramètres s'avère primordial. Pour un ajustement réactif, nous suggérons l'utilisation d'un intervalle de détection $I = 0.1$ ST. Le temps critique doit être ajusté en fonction du tempo de la pièce. La section 4.2.4 montre que $T_c = 250$ ms entraîne le moins de distorsions pour un tempo lent. Un temps critique plus petit doit être choisi pour des tempi plus importants, comme il a été montré en section 4.3.3. La limitation de l'expressivité introduite par un temps critique court est compensée par le fait qu'un musicien montre naturellement moins d'expressivité mélodique lors de successions de notes rapides tout en conservant d'autres formes d'expressivité.

L'effet de la correction a été confirmé par une expérience perceptive. En regardant de plus près les scores attribués à chaque joueur, il apparaît que l'effet de la correction est plus important chez les joueurs ayant une expérience musicale. A l'inverse celle-ci a peu d'effet chez les non-musiciens. La correction est donc prouvée efficace objectivement et perceptivement uniquement chez les joueurs ayant des intentions musicales. De plus, un effet plus faible de la correction a été ressenti chez les joueurs de *Cantor Digitalis* que chez les musiciens. Cela illustre le rôle de l'apprentissage de l'instrument. Bien qu'utile, la correction devient moins essentielle lorsque le joueur gagne en expérience dans le maniement de l'instrument.

La correction est un outil précieux dans l'apprentissage de l'instrument. Comme sur un violon, la précision de hauteur est d'abord limitée par la taille de l'objet (doigt sur la touche ou pointe du stylet sur la tablette) contrôlant la hauteur. Ce n'est qu'après des années d'entraînement que le violoniste apprend à intégrer la largeur de ses doigts pour jouer précisément. On peut supposer qu'une tendance d'apprentissage similaire quoique moins longue existe pour les instruments de musique numériques à hauteur continue. Puisque la correction entraîne une précision inférieure à celle proposée par le stylet, celle-ci sera utile pendant la période d'apprentissage de l'instrument. De plus, comme aucune différence n'a été relevée entre les hauteurs d'entrées corrigées et non corrigées, on en déduit que la correction n'introduit pas de paresse incitant les joueurs à viser seulement dans une zone autour de la note, s'appuyant fortement sur la correction. Par conséquent la correction n'altère pas l'apprentissage.

4.4.2 Corrections visuelles et auditives

Dans le contexte du *Cantor Digitalis*, la correction DPW *élastique* est comparable à la méthode *Expanding target*. Les indices visuels sur la tablette représentent les touches d'un clavier de piano, et sont par conséquent des cibles étendues visuellement. Lorsque la correction

DPW *élastique* est active et le stylet est placé sur une cible étendue (touche de piano sur le visuel), alors la correction produit une hauteur juste à l'instant de correction. De même sur un écran, la cible est correctement sélectionnée si le curseur se situe dans la zone étendue par la méthode *Expanding targets*. Toutefois, dans le cas de tâches visuelles la méthode *Expanding targets* et moins attractive que d'autres méthodes car elle introduit des distractions visuelles [MB05]. En effet, en étendant la cible, les cibles voisines sont compressées ou masquées. Dans le cas du *Cantor Digitalis*, les cibles sont suffisamment espacées pour qu'elles ne se superposent pas. Il est donc possible de fournir un visuel étendu sans distorsions.

La correction DPW *notes étendues* quant à elle est très proche de la méthode *Sticky Icons* par la forme de la fonction. Cependant, dans le cas visuel, les modulations subtiles autour de la cible sont considérées comme du bruit et sont masquées par la méthode. En revanche, dans un contexte musical, ces modulations sont essentielles.

Finalement, dans un contexte visuel les contraintes associées font de la méthode *Sticky Icons* la plus appréciée des sujets car elle corrige précisément les cibles sans introduire de distraction visuelle. A l'inverse, dans un contexte auditif, la fonction fixe *notes étendues* est trop stricte, ne prenant jamais en compte la hauteur d'entrée. La fonction adaptative *élastique* est alors préférée, permettant une correction efficace, tout en préservant l'expressivité et sans introduire de distractions visuelles.

4.4.3 Conclusion

Basée sur de simples fonctions de transformation, la correction DPW présentée ici permet de corriger en temps réel la hauteur jouée sur une interface de contrôle continue comme une tablette graphique. Les expériences ont montré que la méthode peut ajuster automatiquement la hauteur avec une justesse et une précision remarquable. Cela a été vérifié de manière perceptive. Concernant l'expressivité, seule la fonction *élastique* permet d'éviter toute distorsion pour les trois types de modulations les plus courantes (vibrato, glissando et portamento). On en conclut donc que la fonction de correspondance *élastique* entre hauteurs d'entrée et de sortie est plus favorable à l'expressivité mélodique.

La correction est implémentée sur le *Cantor Digitalis* et a été utilisée avec succès lors de représentations publiques (voir chapitre 7). Dans ce contexte, les musiciens ont pu apprécier le confort apporté par la correction DPW, particulièrement pour les performances rapides.

La méthode a été développée dans le contexte d'un contrôle musical par une tablette graphique, mais il est possible d'envisager d'autres applications avec d'autres interfaces. Il est d'autant plus intéressant que la tablette graphique offre déjà une précision de pointage important. On peut alors s'attendre à de meilleures améliorations de correction avec d'autres appareils, telles que les interfaces tactiles. De plus la correction DPW est capable de corriger n'importe quelle grandeur physique continue proposant des cibles discrètes à intervalles réguliers. Par exemple, dans le contexte de la synthèse vocale, une méthode de correction dynamique par déformation de voyelles (DVW) pourrait être implémentée pour aider la sélection de voyelles dans un espace continu d'articulation.

Après avoir mis en évidence l'adéquation de la tablette pour le contrôle chironomique de la hauteur vocale (chapitre 3), l'augmentation logicielle de l'interface par une correction automatique permet un contrôle de la hauteur remarquable. Cela exploite à la fois l'habilité du joueur à manier un stylet, acquise dès le plus jeune âge par la majorité de la population et fournissant une justesse et précision déjà supérieure à la voix, et l'expérience musicale du joueur tirant parti des propriétés de la correction et ramenant les erreurs de notes en dessous du seuil de perception de hauteur.

Chapitre 5

Multi-modalité de la pratique de l'instrument

Sommaire

5.1 Introduction

Une des étapes cruciales de la conception d'instruments de musique numériques est le choix de l'interface. Ses caractéristiques influencent directement les possibilités musicales et l'expressivité du joueur. Les nouvelles interfaces actuelles créées pour l'expression musicale sont parfois des systèmes déjà existants et conçus initialement dans des buts sans liens directs avec la musique (tablettes graphiques [ZWMC07], kinects, tablettes numériques [WOL11], objets de la vie courante [ACDCH12] ...). Dans d'autres situations, de nouveaux contrôleurs sont spécialement développés pour un nouvel instrument [JGAK07], [Mar10]. Dans tous les cas, les procédés cognitifs associés à la manipulation de ces nouveaux contrôleurs pour la production musicale sont peu connus et ont retenu notre attention en termes de modalités sensori-motrices impliquées dans le jeu de l'instrument. Cette étude se concentre sur les différentes modalités impliquées dans le jeu du *Cantor Digitalis*.

Parmi les contrôles proposés par la tablette graphique, seul celui de l'intonation est étudié ici. La hauteur est modifiée par la position horizontale du stylet sur la tablette. Pour aider les joueurs à viser des hauteurs précisément, le calque montré en figure 2.4 est appliqué sur la tablette. Jouer du *Cantor Digitalis* implique des actions liées au corps dites proximales, comme le mouvement du stylet sur la tablette, ainsi que des actions extracorporelles ou action distales comme la superposition de la pointe du stylet avec les indices visuels sur la tablette, ou la variation auditive de hauteur mélodique. Ces actions correspondent chacune à un canal perceptif : les actions proximales sont perçues à travers le retour kinesthésique et les actions distales à travers les retours visuels ou auditifs. Les retours kinesthésiques et visuels sont dits primaires car directement liés à la manipulation de l'instrument. Le retour auditif est secondaire. Le tableau 5.1 résume les actions et perceptions impliquées dans le contrôle mélodique du *Cantor Digitalis*.

	Action	**Perception**
Proximale	Mouvement du stylet	Kinesthésique
Distale	Correspondance entre pointe du stylet et indices visuels	Visuelle
	Variations de hauteur mélodique	Auditive

Table 5.1 – *Résumé des actions et perceptions distales et proximales impliquées dans le contrôle mélodique du Cantor Digitalis.*

L'influence des modalités mises en jeu dans le contrôle du *Cantor Digitalis* a été partiellement explorée au chapitre 3 lors de la comparaison des justesses vocales et chironomiques. Trois conditions ont été testées : l'imitation de motifs mélodiques à la voix, à la tablette avec retour auditif du synthétiseur, et à la tablette sans retour auditif. Le calque présentant des indices visuels était présent sur la tablette. Les deux conditions chironomiques ont montré des meilleures justesses mélodiques que la voix, et présentaient des résultats comparables, indépendamment de la présence du retour auditif. Par conséquent, il a été relevé que les indices visuels influencent les joueurs à se concentrer essentiellement sur le retour visuel, tandis que la présence de retour auditif n'a pas d'impact significatif sur la justesse mélodique. A l'inverse, la voix ou la manipulation de tout instrument de musique acoustique est apprise à l'aide des retours auditifs et kinesthésiques seulement [MPHS02].

Le but de cette étude est de déterminer le degré d'impact de chaque modalité (visuelle, auditive, kinesthésique) sur le contrôle mélodique du *Cantor Digitalis*. Une revue des études sur les interférences sensori-motrices est présentée dans la suite de l'introduction. L'expérience est décrite en section 5.2 et les résultats sont discutés en section 5.3.

5.1.1 Action, perception et théorie de *l'event coding*

Le principe idéomoteur affirme que chaque action est planifiée pour obtenir le résultat dont on anticipe la perception [Gre70]. Autrement dit, par expérience résultant d'un apprentissage, le résultat et les effets produits par une action peuvent être prédits, et c'est par la volonté d'obtenir ces résultats que l'action est réalisée. Une action effectuée pour la première fois est donc plus difficile à produire car les effets attendus ne sont pas connus. A l'inverse, une forte expérience rend les actions plus faciles à réaliser.

Dans le cas du *Cantor Digitalis*, le tableau 5.1 montre que les effets attendus sont multiples et de différentes natures (proximale / kinesthésique et distale / auditive et visuelle). Nous pouvons donc nous demander le(s)quel(s) parmi ces effets influence(nt) l'action de changement de hauteur sur l'instrument.

Pour aller plus loin, la théorie de *l'event coding* affirme que action et perception partagent le même domaine de représentation cognitif [HMAP01]. Ceux-ci sont représentés sous la forme de codes caractéristiques (*feature codes*) décrivant l'évènement perçu ou à accomplir. Par exemple, percevoir une balle de tennis activera parmi d'autres les codes *jaune*, *petit*, *rond*. Percevoir une balle de basket activera les codes *orange*, *gros* (de manière relative) et *rond*. Deux étapes interviennent dans l'utilisation de ces codes. D'abord, tous les codes relatifs aux évènements (action ou perception) planifiés sont activés, c'est-à-dire sont amenés au premier plan du processus cognitif. Si on doit chercher sur une image les balles de tennis et de basket on activera les codes *jaune*, *orange*, *petit*, *gros* et *rond*. Dans un deuxième temps, lorsqu'un évènement est réalisé, les codes relatifs à celui-ci sont intégrés, c'est-à-dire liés entre eux et associés à l'évènement en cours. Ils ne sont alors plus disponibles pour d'autres évènements simultanés. Dans l'exemple précédent, si on décide de se concentrer sur les balles de tennis, on intègrera *jaune*, *petit* et *rond*. Le code *rond* n'est donc plus disponible pour la recherche simultanée des balles de basket. Lors de la recherche des balles de tennis, on ne pourra identifier les balles de basket que par leur couleur *orange* ou leur taille *gros* mais plus par leur forme. Une fois l'évènement accompli, les codes retournent en état d'activation ou sont désactivés si aucun évènement futur et planifié ne les sollicite. Ces étapes permettent l'enchaînement rapide d'actions ou perceptions partageant les mêmes caractéristiques car les codes sont alors activés en permanence. En revanche, si deux évènements simultanés partagent les mêmes codes, alors ces derniers sont intégrés par un seul évènement et le deuxième se verra détérioré en termes de performance. Quand une action est réalisée, tout évènement partageant les mêmes codes caractéristiques ne pourra être perçu au même moment. Dans l'exemple précédent, on peut soit chercher les balles de tennis, soit les balles de basket efficacement. Enfin, l'intégration des codes caractéristiques relatifs à un évènement se fait en pondérant chaque code activé pour sélectionner les plus pertinents.

En utilisant ce principe de *l'event coding*, il a été montré que la perception de la hauteur mélodique se fait selon une représentation mentale spatiale [RKG+06]. Rusconi *et al.* ont demandé à des participants non-musiciens de comparer deux hauteurs de sons et à des sujets musiciens et non-musiciens de comparer deux timbres de sons de hauteurs différentes. A chaque fois, les réponses étaient données par des touches situées à gauche et à droite ou en haut et en bas d'un clavier et les temps de réponses étaient mesurés. Les sons de hauteurs plus aiguës (resp. graves) étaient identifiés plus rapidement en appuyant sur les touches de droite ou du haut (resp. de gauche ou du bas), quelle que soit la tâche (comparaison de la hauteur ou du timbre). Cela suggère donc une représentation cognitive (codes caractéristiques) commune entre hauteur aiguë et position haute ou droite, et entre hauteur grave et position basse ou

gauche. Cet effet appelé SMARC (*Spatial-Musical Association of Response Code*), propose une représentation mentale linéaire de la hauteur dans les directions de l'espace verticale ou horizontale. Ce résultat est confirmé par [KBLH08], qui lors de tâches de comparaison de hauteurs entre différents sons montre que le temps de réponse décroît linéairement avec la taille de l'intervalle en demi-tons, à l'image d'autres grandeurs physiques.

On en déduit que la planification du contrôle de la hauteur sur la tablette est directement liée à la distance du mouvement à accomplir et active alors trois codes caractéristiques, chacun relatif aux représentations kinesthésique, visuelle et auditive. Lors de l'accomplissement du mouvement, une pondération de ces codes permet de choisir le(s)quel(s) de ces codes est (sont) le(s) plus pertinent(s) à être intégrés. Afin de quantifier cette pondération, nous cherchons par la suite à introduire des interférences entre retours visuels, auditifs et kinesthésiques produisant des distances perçues différentes pour chaque retour. L'étude de l'action résultante de la perception de ces retours discordants permet alors d'identifier le ou les codes intégrés lors de l'accomplissement du mouvement.

5.1.2 Influence des évènements visuels sur les évènements proximaux

De nombreuses preuves de la dominance de la modalité visuelle sur la modalité kinesthésique émergent de la littérature [SSRM13]. Rieger *et al.* ont étudié l'adaptation motrice à des changements dans l'environnement visuel [RKP05]. Dans leur étude, les sujets ont pour tâche d'atteindre de manière répétée deux cibles alignées verticalement sur un écran avec un curseur. Ce dernier est contrôlé par la position verticale du stylet sur une tablette graphique cachée de la vue des sujets. Après six mouvements par défaut, un gain est introduit entre les amplitudes du stylet et du curseur pendant les six mouvements suivants. Ces deux séquences sont répétées en appliquant différents gains, changeant seulement l'amplitude du curseur (perturbation distale), ou seulement l'amplitude du stylet (perturbation proximale). Une troisième condition est testée, n'introduisant pas de gains mais changeant la position de la cible inférieure à l'écran. L'adaptation des sujets à chaque nouvelle condition est observée. La compensation suite aux changements de positions de cibles est plus rapide que les compensations dues aux changements de gains. De plus, les compensations de la perturbation distale sont plus rapides que les compensations des perturbations proximales. Rieger *et al.* concluent donc que l'adaptation à l'environnement se base principalement sur la perception distale, c'est-à-dire sur le retour visuel.

D'autres preuves de la dominance de la modalité visuelle sur la modalité kinesthésique apparaissent sur la perception de formes dessinées par des mouvements de la main [MS09], [WSM+12]. Dans la première étude, il est demandé aux participants de dessiner des cercles à l'écran, en contrôlant le curseur à l'aide d'une tablette graphique masquée. Des gains sont appliqués successivement sur les axes horizontal et vertical de la tablette faisant dessiner des ellipses aux sujets. Il apparait que la plupart des participants sont peu conscients des mouvements de leurs mains car des gains relativement importants sont nécessaires pour que les sujets perçoivent des ellipses et non des cercles. Dans la deuxième étude, les participants doivent déplacer un bras haptique placé hors de leur vue sur une trajectoire triangulaire prédéfinie, tout en regardant une trajectoire visuelle conflictuelle se dessiner à l'écran. L'angle supérieur des triangles est différent entre la trajectoire haptique et visuelle (aigu ou obtus). Une condition sans retour visuel est aussi testée. De même, les sujets se révèlent très incertains de la trajectoire de leur main, et leur perception est faussement influencée par les visuels conflictuels.

Tandis que les deux études précédentes sont perceptives, une troisième série de travaux a pour but de mesurer et de quantifier les répercussions observées dans des tâches de réplication de mouvements manuels soumises à des interférences entre mouvements moteurs et retour visuel [LSM12], [LSM13], [WSS14]. Commun aux trois études, le principal protocole expérimental demande aux sujets d'atteindre une cible visuelle affichée à l'écran en contrôlant un curseur avec une tablette graphique masquée. Le retour visuel est ensuite supprimé et les sujets doivent reproduire un mouvement de stylet inverse pour revenir à leurs positions initiales. Différents gains sont appliqués entre les amplitudes du stylet et du curseur soit en changeant seulement l'amplitude du curseur (perturbation distale) soit en variant seulement l'amplitude du stylet (perturbation proximale). Des dépassements (resp. mouvements trop courts) du stylet sont observés lorsque l'amplitude du curseur est supérieure (resp. inférieure) au mouvement du stylet. Ces résultats sont plus prononcés lorsque les directions des mouvement du stylet et du curseur sont identiques [LSM12] et quand la forme de leurs trajectoires sont les mêmes (ligne droite) [WSS14]. De plus, Ladwig *et al.* [LSM13] ont montré que la modalité visuelle influe sur la modalité motrice alors que cette dernière n'a pas d'effets sur le visuel.

En examinant les interférences entre modalités visuelles et motrices, ces études attestent l'hypothèse émise au chapitre 3 : le retour visuel a une forte influence sur le contrôle moteur.

5.1.3 Influence des évènements auditifs sur les évènements proximaux

L'aspect visuel d'un geste est facilement identifiable étant donnée sa dimension spatiale. En revanche, l'association d'un son à un geste est moins immédiate. On peut différencier dans ce cas deux catégories de sons : les sons naturels résultant du geste et de son interaction avec l'environnement, et des sons sans liens apparents avec le geste mais dont on cherche des analogies ou métaphores (sonification).

Dans le cas de sons naturels liés au geste comme le frottement d'un stylo sur une feuille de papier, il a été montré que certains gestes peuvent être identifiés uniquement à partir du son qu'ils produisent [TAKM+14]. Cette étude prouve qu'il est possible de synthétiser des sons de frottements représentatifs de mouvements biologiques basés sur la loi de puissance 2/3 [LTV83], et que les sujets sont capables de reconnaître plusieurs formes dessinées seulement en écoutant le son de frottements, réel ou synthétisé, simulant le tracé. Il existe donc un fort lien cognitif entre geste et son naturel associé.

Pour aller plus loin, les mêmes auteurs ont étudié l'influence des retours visuels et auditifs sur le tracé de formes [TAB+14a], [TAB+14b]. Il est demandé aux participants de tracer à l'aveugle sur une tablette graphique les formes perçues visuellement (données sur un écran) et/ou auditivement parmi des cercles ou des ellipses. Trois distracteurs sont introduits : le tracé de la trajectoire à l'écran (ellipse ou cercle), le parcours de cette trajectoire par un point de cinématique variable (ellipse ou cercle) et l'écoute d'un son de synthèse simulant le tracé d'une trajectoire elliptique ou circulaire. La combinaison des trois entraîne des retours concordants ou discordants suivant les cas. Cette étude conclut que dans le cas visuel, la cinématique biologique du mouvement présentée à l'écran influence grandement le mouvement moteur des sujets. Dans le cas auditif, un retour discordant affecte énormément la perception de la trajectoire affichée (aplatissement des cercles et ellipses plus rondes), d'autant plus quand la cinématique visuelle est aussi discordante. Lorsqu'il s'agit de sons naturels associés au geste, ces études montrent une grande dépendance auditive et kinesthésique.

La deuxième approche est de considérer des sons non liés au geste mais présentant des analogies par leurs caractéristiques. C'est l'approche adoptée par la conception d'instruments de musique numériques puisqu'il s'agit d'associer gestes de contrôle à des sons musicaux. Des corrélations entre cinématique gestuelle (positions, vitesses, accélérations) et certaines caractéristiques du son (amplitude, clarté) ont été mises en évidence par Caramiaux *et al.* [CBS10].

Beaucoup de travaux ont étudié l'effet de l'apport d'un retour auditif sur l'aide à la visée de cible, mais les retours auditifs proposés sont dits discrets et associés à un évènement particulier (curseur au-dessus de la cible [AMH95], [CB05], [CJDS10], curseur hors du trajet proposé [SRC10]). A l'inverse, peu de travaux ont exploré l'influence d'un retour auditif continu dont les propriétés correspondent au mouvement de l'utilisateur. Un exemple est celui d'Andersen *et al.* [AZ10] qui se sont intéressés à la sonification de la trajectoire du stylet sur la tablette pour le dessin de motifs simples. Celle-ci est réalisée en synthétisant une somme de sons purs et associant position verticale du stylet à la hauteur mélodique, position horizontale avec richesse spectrale, apériodicité et inharmonicité, et vitesse du stylet avec amplitude. Les sujets doivent dessiner différents motifs sans regarder la tablette dans quatre conditions : sans retour, avec retour auditif, avec retour visuel, avec retours visuel et auditif. Alors que le retour visuel permet d'améliorer certains aspects de la reproduction (taille, vitesse, fermeture), le retour auditif a peu d'impact (fermeture uniquement). Ces observations sont appuyées par les théories du contrôle à boucle ouverte, où la perception est trop lente pour aider le geste moteur. Néanmoins, une deuxième expérience montre que l'apport d'un retour auditif permet de rendre plus attractives les tâches de dessin, suscitant l'amusement des sujets.

Bien que cette étude ne montre pas d'impact important du retour auditif sur la performance des sujets, le retour est relativement complexe. Ici on s'intéresse uniquement à la relation entre distance sur une dimension de la tablette et hauteur mélodique. Ce travail a donc pour but d'explorer l'influence de l'audition dans le contrôle de la hauteur du *Cantor Digitalis*. Pour comparer l'impact des modalités visuelles et auditives, des mesures quantitatives sont effectuées en créant des interférences entre modalité visuelle, auditive et kinesthésique par une adaptation du protocole de Ladwig *et al.* [LSM12].

5.2 Expérience sur l'influence des perceptions auditive et visuelle sur le geste moteur

Afin de simuler le contrôle mélodique du *Cantor Digitalis*, le mouvement proximal étudié ici est un trait horizontal dessiné à l'aveugle sur une tablette graphique. Contrairement à l'usage régulier de l'instrument, la tablette est placée hors de la vue des sujets et un retour visuel est affiché sur un écran pour découpler spatialement les évènements moteurs et visuels. La hauteur est toujours contrôlée linéairement suivant la position horizontale du stylet.

Nous avons mis au point une tâche de réplication similaire à [LSM12], divisée en deux phases : le sujet commence par déplacer le stylet horizontalement pour atteindre une cible distale donnée par le retour visuel et/ou auditif. Une fois la cible atteinte, la phase 2 commence, consistant à retourner à la position initiale du stylet le plus précisément possible en effectuant le mouvement inverse de la phase 1 sans l'aide des retours visuel et/ou auditif. La distance entre les positions initiale et finale appelée erreur de reproduction est mesurée pour quantifier l'influence des retours distaux sur la perception du mouvement.

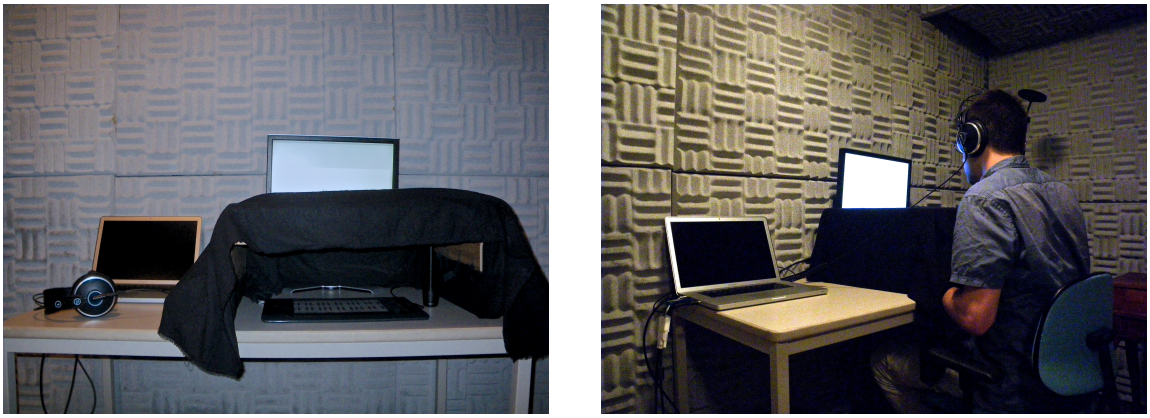

FIGURE 5.1 – *Disposition de l'expérience. Gauche : la tablette est placée sous un support opaque et un écran derrière le support affiche le retour visuel. Droite : Position adoptée par un sujet pour l'expérience.*

L'objectif de l'expérience est triple : celle-ci cherche à étudier 1) l'influence du retour visuel (perception distale) sur le contrôle manuel (action proximale) ; 2) l'impact du retour auditif sur le contrôle manuel ; 3) l'interférence entre retours visuel et auditif et leurs impacts sur le contrôle manuel. Trois conditions expérimentales sont proposées pour répondre à ces objectifs : une condition visuelle V présentant uniquement un retour visuel et servant de référence pour comparer nos résultats aux études précédentes ; une condition auditive A proposant uniquement un retour auditif ; une condition audiovisuelle AV présentant en même temps les retours visuel et auditif. Trois hypothèses découlent de ces conditions. H1 : les interférences visuelles et auditives s'annulent et aucune répercussion n'est observée sur les mouvement du stylet. H2 : la modalité visuelle domine la modalité auditive et le mouvement du stylet est influencé par les interférences visuelles. H3 : la modalité auditive est prépondérante et le mouvement du stylet est impacté par les interférences auditives.

5.2.1 Matériel, tâche et stimuli

L'expérience s'est déroulée dans une cabine traitée acoustiquement et insonorisée, et est implémentée sur l'environnement Max/MSP sur un ordinateur Apple Macintosh. Une tablette Wacom Intuos 5M est utilisée comme interface. Un couvercle en carton est placé autour de la tablette graphique afin que les sujets ne puissent pas voir leurs mouvements, comme montré sur la figure 5.1. Le retour visuel est affiché sur un écran DELL 2007FP de résolution (1600 × 1200) pixels et le retour audio est joué à travers un casque fermé AKG-K271.

Pour la condition V, deux barres rectangulaires de taille 0.8×0.2 cm ainsi qu'un curseur circulaire de diamètre 0.4 cm sont affichés à l'écran (figure 5.2). La phase 1 démarre lorsque le curseur est placé sur la barre de départ. La phase 2 commence lorsque le curseur atteint la deuxième barre. Les barres et le curseur ne sont plus affichés en phase 2. La relation entre les amplitudes du stylet et du curseur est perturbée par 9 gains différents : trois amplitudes de stylet (6, 12, 18 cm) combinées à trois amplitudes de curseur (6, 12, 18 cm). Des valeurs identiques que celles proposées dans [LSM12] sont choisies à des fins de comparaison. Les combinaisons (stylet = 12 cm × curseur = (6, 12, 18 cm)) sont appelées *curseur perturbé* et les combinaisons (stylet = (6, 12, 18 cm) × curseur = 12 cm) *stylet perturbé* selon la terminologie employée dans [LSM12]. Les gains de la condition V sont résumés dans le tableau 5.2.

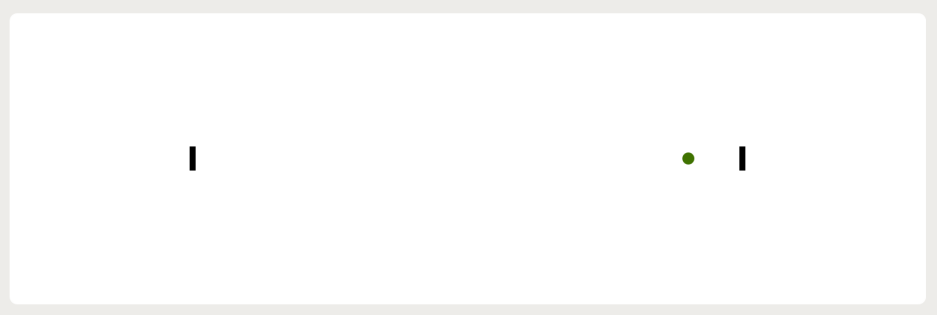

Figure 5.2 – *Capture d'écran du retour visuel affiché en phase 1 de la condition V. Les positions initiales et finales du mouvement sont indiquées par les barres verticales noires.*

		Curseur		
		Petit (6 cm)	Moyen (12 cm)	Grand (18 cm)
Stylet	Petit (6 cm)	Non perturbée	Stylet perturbé	
	Moyen (12 cm)	Curseur perturbé	Non perturbée	Curseur Perturbé
	Grand (18 cm)		Stylet perturbé	Non perturbée

Table 5.2 – *Résumé des 9 gains proposées pour la condition V. Les cases vides découlent seulement d'une absence de terminologie pour le gain associé.*

		Hauteur		
		Petite (8 demi-tons)	Moyenne (16 demi-tons)	Grande (24 demi-tons)
Stylet	Petit (6 cm)	Non perturbée	Stylet perturbé	
	Moyen (12 cm)	Hauteur perturbée	Non perturbée	Hauteur Perturbée
	Grand (18 cm)		Stylet perturbé	Non perturbée

Table 5.3 – *Résumé des 9 gains proposés pour la condition A. Les cases vides découlent seulement d'une absence de terminologie pour le gain associé.*

Pour la condition A, le retour audio est produit par le synthétiseur du *Cantor Digitalis* calibré sur une voyelle /a/. La hauteur est modifiée continûment lorsque le stylet se déplace horizontalement sur la tablette. La phase 1 démarre avec une hauteur initiale associée au stylet jouée en même temps qu'une note de référence aussi produite par le synthétiseur. La phase 2 commence lorsque la hauteur contrôlée atteint la note de référence. Aucun son n'est joué durant la phase 2. Dans la condition V, la cible montrée à l'écran a une certaine largeur qui introduit une tolérance dans la visée. Pour simuler cette largeur dans le domaine auditif, la correction de justesse DPW *élastique* introduite au chapitre 4 est activée. Celle-ci ajuste la hauteur au demi-ton près lorsque la variation de hauteur est faible, soit quand le stylet est proche de la cible. Les mêmes trois amplitudes de stylet (6, 12, 18 cm) sont combinées à trois amplitudes de hauteur (8, 16, 24 demi-tons), conduisant à 9 gains pour cette condition. Les trois amplitudes de hauteur sont choisies comme étant équidistantes, sans introduire d'intervalles de quinte car très consonantes et pouvant amener à des erreurs d'atteinte de cible, et en limitant l'intervalle le plus large à deux octaves, ce qui correspond approximativement à la dynamique de hauteur d'un chanteur. Ces intervalles induisent un mapping de 1.33 demi-tons / cm sur la tablette graphique. Les combinaisons (stylet = 12 cm × hauteur = (8, 16, 24 demi-tons) sont appelées *hauteur perturbée* et les combinaisons (stylet = (6, 12, 18 cm) × hauteur = 16 demi-tons) *stylet perturbé*. Les gains de la condition A sont reportés dans le tableau 5.3.

Petite amplitude de stylet		**Curseur**		
		Petit (6 cm)	Moyen (12 cm)	Grand (18 cm)
Hauteur	Petite (8 demi-tons)	Non perturbée	Stylet perturbé	
	Moyenne (16 demi-tons)			
	Grande (24 demi-tons)			

Moyenne amplitude de stylet		**Curseur**		
		Petit (6 cm)	Moyen (12 cm)	Grand (18 cm)
Hauteur	Petite (8 demi-tons)		Hauteur perturbée	Interférences
	Moyenne (16 demi-tons)	Curseur perturbé	Non perturbée	Curseur Perturbé
	Grande (24 demi-tons)	Interférences	Hauteur perturbée	

Grande amplitude de stylet		**Curseur**		
		Petit (6 cm)	Moyen (12 cm)	Grand (18 cm)
Hauteur	Petite (8 demi-tons)		Stylet perturbé	
	Moyenne (16 demi-tons)			
	Grande (24 demi-tons)			Non perturbée

TABLE 5.4 – *Résumé des 27 gains proposées pour la condition AV. Les cases vides découlent seulement d'une absence de terminologie pour le gain associé.*

Les deux références visuelle et auditive sont fournies dans la condition AV. Les trois amplitudes de stylet, de curseur et de hauteur sont testées ici, conduisant à 27 gains pour cette condition. Les combinaisons *stylet perturbé* sont (stylet = (6, 12, 18 cm) × curseur = 12 cm × hauteur = 16 demi-tons), les combinaisons *curseur perturbé* sont (stylet = 12 cm × curseur = (6, 12, 18 cm) × hauteur = 16 demi-tons) et les combinaisons *hauteur perturbée* sont (stylet = 12 cm × curseur = 12 cm × hauteur = (8, 16, 24 demi-tons)). Les deux conditions d'interférences sont (stylet = 12 cm × curseur = 6 cm × hauteur = 24 demi-tons) et (stylet = 12 cm × curseur = 18 cm × hauteur = 8 demi-tons) où le curseur a une amplitude plus large et la hauteur une amplitude plus courte que l'amplitude du stylet, et inversement. Les gains de la condition AV sont résumés dans le tableau 5.4.

Pour chaque condition, les deux directions sont observées : phase 1 avec un mouvement vers la gauche et hauteur plus aiguë et inversement. Finalement, 3 amplitudes de stylet × 3 amplitudes de curseur × 2 directions = 18 stimuli sont proposées pour la condition V, 3 amplitudes de stylet × 3 amplitudes de hauteur × 2 directions = 18 stimuli sont proposées pour la condition A et 3 amplitudes de stylet × 3 amplitudes de curseur × 3 amplitudes de hauteur × 2 directions = 54 stimuli sont proposées pour la condition AV.

5.2.2 Déroulement

Les stimuli des trois conditions sont mélangés aléatoirement dans un bloc de 90 stimuli. Chaque participant est soumis à 5 blocs, précédés d'une session d'entraînement de 54 stimuli. L'expérience a duré en moyenne 120 min par sujet, incluant les temps de repos entre les blocs.

Pour toutes les conditions, chaque stimulus commence par une phase d'initialisation, où une barre est affichée à l'écran avec un curseur rouge afin d'indiquer la position initiale du

stylet aux sujets. Une fois la barre atteinte, le curseur devient orange et le sujet presse un des boutons de la tablette pour passer en phase 1. Les deux barres sont affichées et le curseur devient vert pour les conditions V et AV, ou la barre de départ et le curseur sont supprimés pour la condition A. La hauteur liée au stylet et la référence sont jouées pour les conditions A et AV uniquement. Lorsque le curseur atteint la barre cible ou la hauteur de référence, les participants appuient une deuxième fois sur le bouton de la tablette pour débuter la phase 2. Ces derniers doivent alors répliquer leurs mouvements en sens inverse et presser une troisième fois le bouton de la tablette lorsqu'ils pensent avoir atteint leur position initiale. Les sujets doivent finalement presser une dernière fois le bouton pour passer au stimulus suivant. Il est demandé aux participants de prêter attention à leur mouvement lors de la phase 1 et d'atteindre la cible et leur position initiale le plus précisément possible, sans contrainte de temps, sans changer de direction, ni stopper le mouvement pendant chaque phase. Les participants pouvaient se reposer entre chaque stimulus.

Pour chaque condition, la déviation du stylet entre les phases 1 et 2 est analysée. Les essais des sujets sont écartés lorsque la phase 2 est initiée avant d'avoir atteint la cible, lorsque la cible est dépassée, et quand la direction du mouvement change ou le mouvement est stoppé au sein de chaque phase.

5.2.3 Participants

15 sujets ont pris part à l'expérience (5 femmes), recrutés au sein du laboratoire et de l'université. Agés de 18 à 42 ans (moyenne 23 ans), aucun d'eux n'a reporté de problèmes auditifs et moteurs ou de vision non corrigée. 12 ont reçu une éducation musicale (11.7 années en moyenne), 6 ont une pratique régulière ou occasionnelle de la tablette graphique et tous sont familiers avec le contrôle d'un curseur à l'écran, 11 utilisant un trackpad et 4 une souris. Tous les participants étaient naïfs vis-à-vis de l'expérience, et ont participé en échange d'une rémunération de 20 euros.

5.3 Résultats et discussion

5.3.1 Résultats

Les déviations d'amplitude du mouvement entre les phases 1 et 2 sont extraites des trajectoires sans erreur (taux d'erreur de 2.89% pour la condition V, 30.1% pour la condition A et 2.10% pour la condition AV) et analysées séparément sur chaque condition utilisant un modèle d'effets mixtes. Les facteurs fixes pour la condition V sont les amplitudes du stylet (stylet : 3 niveaux) et du curseur (curseur : 3 niveaux). Les facteurs fixes pour la condition A sont les amplitudes du stylet (stylet : 3 niveaux) et de la hauteur (hauteur : 3 niveaux). Les facteurs fixes pour la condition AV sont les amplitudes du stylet (stylet : 3 niveaux), du curseur (curseur : 3 niveaux) et de la hauteur (hauteur : 3 niveaux). Pour les trois conditions, les influences des sujets et des répétitions à travers les 5 blocs sont modélisées par des facteurs aléatoires. Une procédure de simplification de chaque modèle est réalisée en supprimant progressivement les effets non significatifs des facteurs, tout en vérifiant que le modèle simplifié ne diffère pas de manière significative de l'original [Cra13]. Une analyse de variance (ANOVA) donne les déviances expliquées par chaque facteur. Les librairies *lme4* et *car* de l'environnement R [tea13] sont utilisées.

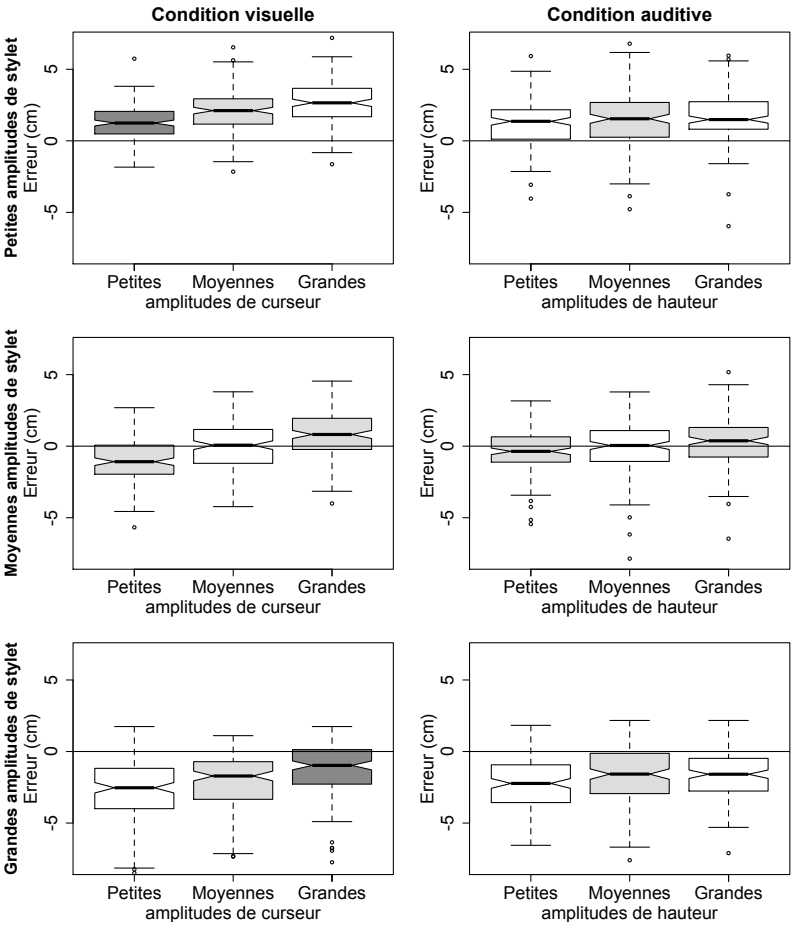

FIGURE 5.3 – *Déviation (cm) entre les mouvements guidés et répliqués sous les conditions V (gauche) et A (droite) en fonction des amplitudes du stylet (de haut en bas) et du curseur ou de la hauteur (sur chaque panneau). Les boîtes gris clair représentent les conditions de stylet et de curseur perturbés. Les boîtes gris foncé représentent les conditions non perturbées.*

Condition	**Facteur**	χ^2	**d.f.**	**p**
V	stylet	1490	2	$< 10^{-16}$
	curseur	224	2	$< 10^{-16}$
A	stylet	777	2	$< 10^{-16}$
	hauteur	32.1	2	$< 10^{-7}$

TABLE 5.5 – *Déviances expliquées par chaque facteur pour les conditions V et A. La significativité est testée selon une distribution χ^2.*

La figure 5.3 montre les déviations de chaque sujet et de chaque essai pour les conditions V (gauche) et A (droite) en fonction des amplitudes du stylet (haut : petite ; milieu : moyenne ; bas : grande) et du curseur (sur chaque panneau, gauche : petite ; centre : moyenne ; droite : grande). Chaque boîte contient 50% des valeurs et les lignes noires sont les médianes. Les boîtes gris clair représentent les conditions *stylet perturbé* et *curseur perturbé* ou *hauteur perturbée* comme décrites par [LSM12], et les boîtes gris foncé représentent les conditions non-perturbées (même amplitude de stylet et de curseur). Le tableau 5.5 reporte les résultats de l'analyse statistique.

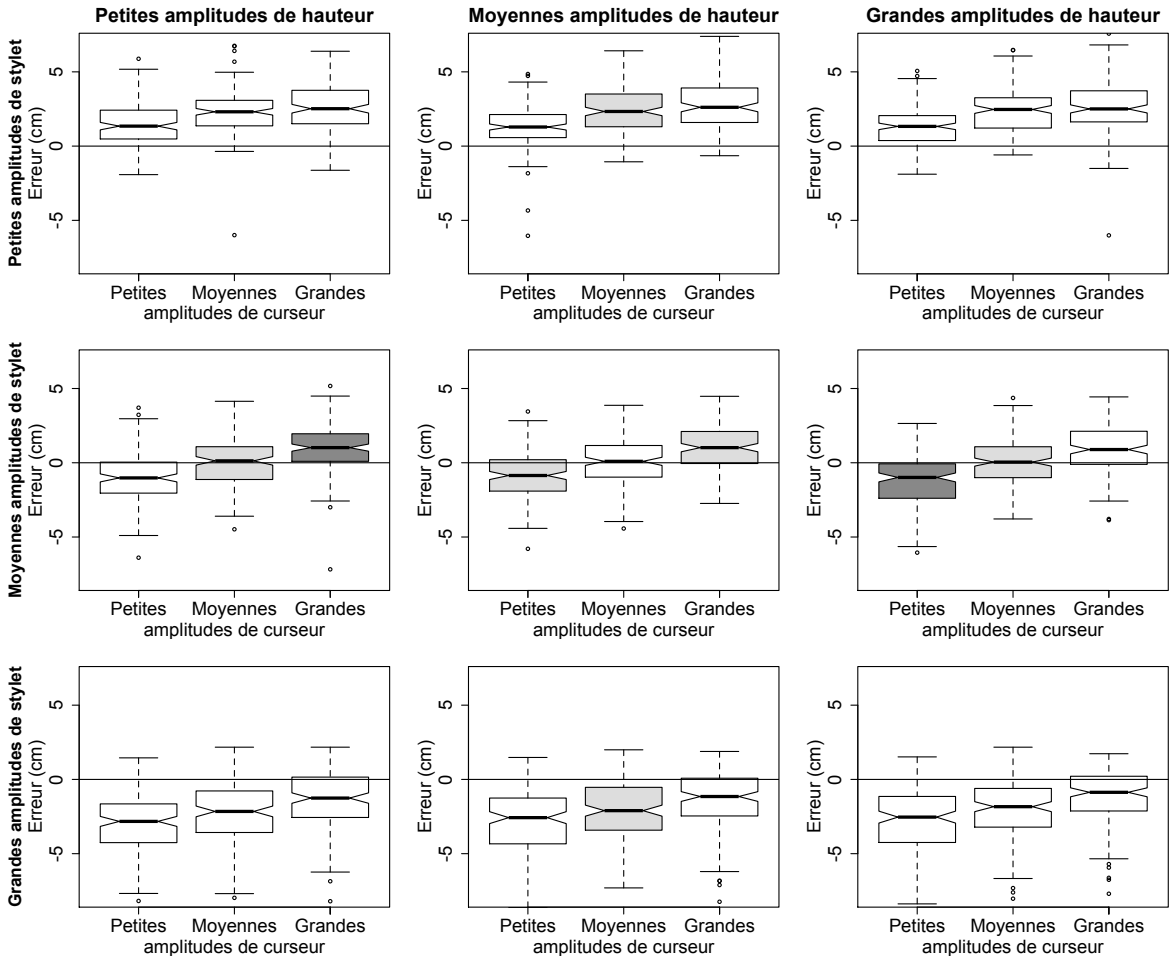

FIGURE 5.4 – *Déviation (cm) entre les mouvements guidés et répliqués sous la condition AV en fonction des amplitudes du stylet (de haut en bas), de la hauteur (de gauche à droite) et du curseur (sur chaque panneau). Les boîtes gris clair représentent les conditions de stylet, de curseur et de hauteur perturbés et les boîtes gris foncé les conditions d'interférences.*

Condition	**Facteur**	χ^2	**d.f.**	**p**
AV	stylet	4970	2	$< 10^{-16}$
	curseur	746	2	$< 10^{-16}$
	stylet×curseur	30.3	2	$< 10^{-6}$

TABLE 5.6 – *Déviances expliquées par chaque facteur pour la condition AV. La significativité est testée selon une distribution χ^2.*

Pour la condition V, des répercussions significatives apparaissent, dues aux amplitudes du stylet et du curseur. En comparaison avec l'amplitude du stylet intermédiaire, la petite amplitude induit des mouvements de retour plus large (+1.94 cm) tandis que la grande amplitude du stylet mène à des mouvements de retour plus courts (-1.78 cm). A l'inverse, les mouvements de retour sont significativement plus court (-0.65 cm) avec une petite amplitude de curseur et plus longs (+0.74 cm) avec une grande amplitude de curseur. Aucune interaction entre stylet et curseur n'est observée.

Des effets similaires se manifestent avec la condition A. L'effet principal est l'influence de l'amplitude du stylet avec laquelle les mouvements de retour sont 1.42 cm plus longs pour des petites amplitudes et -1.77 cm plus courts pour des grandes amplitudes en comparaison avec une amplitude intermédiaire. L'amplitude de la hauteur a quant à elle un effet plutôt faible, la médiane des mouvements de retour étant augmentée de 0.2 cm pour des amplitudes de hauteur large et réduite de -0.41 cm pour des amplitudes de hauteur faibles.

La figure 5.4 illustre les erreurs de réplication pour la condition AV. L'effet de l'amplitude du stylet est montré suivant les lignes, l'effet de la hauteur est affiché suivant les colonnes et l'effet du curseur est présenté sur chaque panneau. Les boîtes gris clair représentent les conditions stylet, curseur et hauteur perturbés, et les boîtes gris foncé montrent les conditions d'interférence. Le tableau 5.6 reporte les résultats de l'analyse statistique.

Comme pour les conditions V et A, l'amplitude du stylet a l'effet le plus prononcé, avec des mouvements de retour plus larges (+1.96 cm) pour des petites amplitudes de stylet et plus courts (-1.95 cm) pour des grandes amplitudes de stylet, en comparaison avec les amplitudes intermédiaires. L'amplitude du curseur a de nouveau l'effet inverse avec des mouvements de retour plus petits (-0.83 cm) pour des amplitudes faibles et plus larges (0.63 cm) pour des grandes amplitudes de curseur. Néanmoins, aucun effet significatif de la hauteur n'est observé. Par contre, une interaction significative entre les amplitudes du stylet et du curseur émerge. Celle-ci ne révèle aucune différence entre les amplitudes de curseur moyennes et grandes pour des petites amplitudes de stylet, alors qu'une extension du mouvement de retour est observée pour les autres amplitudes de stylet.

5.3.2 Discussion

Cette expérience propose des perceptions visuelles et/ou auditives discordantes de la perception kinesthésique dans le tracé de lignes horizontales sur la tablette. L'analyse de la longueur de ces tracés permet d'identifier quelle perception a guidé l'action proximale.

Effet de la modalité visuelle - comparaison avec les études précédentes

La condition V est un test de référence pour comparer cette étude aux précédentes. Les résultats suggèrent que les effets obtenus dans cette condition sont analogues et du même ordre de grandeur que ceux des études antérieures [LSM12]. Une grande amplitude de curseur présentée aux sujets tend à leur faire surestimer la taille de leur mouvements et inversement. Une interférence entre la perception distale et l'action proximale est par conséquent démontrée.

Un effet majeur observé ici et non mesuré par [LSM12] est l'influence de l'amplitude du stylet. Les boîtes gris foncé de la figure 5.3 représentent les conditions non-perturbées, lorsque les amplitudes du curseur et du stylet sont égales. Alors que les moyennes amplitudes du stylet et du curseur donnent une erreur centrée autour de 0, les mouvements de retour dépassent systématiquement la position initiale (resp. sont trop courts) pour des petites (resp. larges) amplitudes de curseur et stylet. Un geste de 6 cm sur la tablette (resp. 18 cm) est toujours surestimé (resp. sous-estimé) quelque soit l'amplitude du curseur. En d'autres termes, lorsque l'amplitude du stylet varie, un biais constant propre au mouvement du stylet et indépendant du retour visuel est ajouté aux effets dus à l'amplitude du curseur. Nous pouvons alors suggérer qu'une amplitude de stylet intermédiaire (12 cm) résulte en un mouvement confortable sur la tablette et facile à mémoriser pour les sujets. Lorsque ceux-ci sont confrontés à des amplitudes plus larges ou plus courtes, les sujets tendent à reproduire un mouvement

plus proche de l'amplitude intermédiaire, en allongeant les amplitudes courtes et rétrécissant les amplitudes longues.

L'effet plus prononcé de l'amplitude du stylet par rapport à l'amplitude du curseur fournit une explication à l'asymétrie des effets observés entre les conditions curseur perturbé et stylet perturbé montrée par les boîtes gris clair sur la figure 5.3 et identifiée lors des expériences précédentes [LSM12], [WSS14]. En effet, la dispersion des médianes des erreurs est plus grande pour les conditions stylet perturbé que curseur perturbé. Nos résultats laissent penser que l'asymétrie des effets découle plus d'une caractéristique exclusive à l'action motrice qu'à l'interférence entre les effets distaux et proximaux.

Effet de la modalité auditive

Bien que la condition A présente des tendances similaires à la condition V, les interférences liées à la hauteur mélodique ont beaucoup moins d'impact que les interférences visuelles sur les mouvements des sujets. Les effets sont deux fois moins prononcés que pour la condition visuelle. Plusieurs causes sont responsables de cette différence d'amplitude entre effets visuels et auditifs. D'abord, la nécessité de passer d'une représentation spatiale à une représentation musicale de l'amplitude rend la tâche plus complexe. Alors que le retour visuel introduit un effet distal de même dimension et de même ordre de grandeur que le mouvement proximal, l'audio présente une amplitude exprimée en demi-tons dont la relation avec le mouvement spatial n'est pas triviale. Deuxièmement, la perception d'amplitudes de hauteur est une tâche exigeante, peu commune dans la vie de tous les jours comparée à l'appréciation de distances spatiales. Bien que la plupart des sujets ont une expérience musicale, tous ont trouvé la condition A plus difficile que les autres, ce qui se reflète dans les taux d'erreurs observés pour chaque condition (30% pour la condition A pour moins de 3% pour les autres conditions). Néanmoins, une petite amplitude de hauteur influence les sujets à exécuter des mouvements légèrement plus courts alors qu'une grande amplitude de hauteur les incite à dessiner des mouvements plus larges que les mouvements initiaux. Bien que les effets soient plus faibles, la perception auditive interfère bien avec les mouvements proximaux. Une asymétrie entre les conditions hauteur perturbée (amplitude de stylet moyenne et amplitudes de hauteur variables) et stylet perturbé (amplitude de hauteur moyenne et amplitudes de stylet variables) apparaît, car un biais est à nouveau introduit par les différentes amplitudes de stylet.

Interaction des modalités distales

La condition AV confronte les deux modalités. L'influence plus prononcée du retour visuel inhibe l'impact du retour auditif qui n'induit pas d'effet significatif sur la réplication des mouvements. Comme le retour visuel est plus aisé à suivre que le retour auditif, les sujets semblent s'appuyer essentiellement sur le mouvement du curseur plutôt que sur les variations de hauteur pour achever la tâche de pointage, et sont par conséquent influencés seulement par le retour visuel. Les boîtes gris foncé de la figure 5.4 montrent les conditions d'interférences, où soit l'amplitude du curseur est plus grande que l'amplitude du stylet et l'amplitude de hauteur plus petite, soit l'inverse. Les retours visuel et auditif influencent alors le mouvement dans des directions opposées. Toutefois, comme les erreurs obtenues ne dépendent pas de l'amplitude de la hauteur, cela illustre la dominance de la modalité visuelle sur la modalité auditive. En conséquence des différents degrés d'influence du stylet (forte), du curseur (modérée) et de la hauteur (inexistante), une asymétrie se dessine entre les conditions curseur perturbé, hauteur perturbée et stylet perturbé, montrées par les boîtes gris clair de la figure 5.4.

Finalement, ces observations valident l'hypothèse H2 : la modalité visuelle domine la modalité auditive, et le mouvement du stylet est influencé par les interférences visuelles. Malgré cela, en absence de retour visuel, une influence significative de la modalité auditive se manifeste, comme il a été observé par Andersen *et al.* sur la réplication de gestes sur la tablette en combinant les retour visuel et auditif [AZ10]. Bien qu'un effet du retour auditif soit noté dans l'amélioration des trajectoires, l'effet du retour visuel est prépondérant.

5.4 Discussion générale et conclusion

L'impact du retour visuel sur le contrôle moteur a été démontré dans des usages visuels de la tablette graphique [RKP05], [MS09], [SMB11], [LSM12], [LSM13], [SSRM13], [WSS14]. Toutefois, dans le contexte d'un instrument de musique numérique tel que le *Cantor Digitalis*, la tâche de jouer d'un instrument est supposée être principalement auditive. L'expérience conduite dans cette étude apporte des éclaircissements sur l'impact de la modalité auditive sur le contrôle moteur et son poids comparé à l'impact de la modalité visuelle.

5.4.1 Le calque visuel, prépondérant dans le jeu du *Cantor Digitalis*

Deux principales conclusions peuvent être tirées de cette expérience. D'abord, la modalité auditive a bien un impact sur le contrôle moteur. Des amplitudes de hauteur discordantes des amplitudes de stylet mènent bien à des erreurs dans la réplication du mouvement. Ces discordances sont néanmoins à nuancer, car elles ne peuvent exister que s'il existe une référence. Les trois amplitudes de hauteur (8, 16, 24 demi-tons) sont associées aux amplitudes du stylet (6, 12, 18 cm) selon une correspondance de 1.33 ST/cm, proche du gain de 1.54 ST/cm implémenté par défaut sur l'instrument et amplement utilisé lors de concerts. Bien qu'il n'a pas été prouvé que cette correspondance entre distances spatiales et fréquentielles soit naturelle pour les sujets, nous pensons que l'effet du retour audio réside plus dans les différences relatives importantes entre les trois amplitudes d'intervalles que dans leurs valeurs en absolu. Le choix d'amplitudes plus resserrées dans une investigation future permettrait de définir jusqu'à quel point la modalité auditive à un impact sur la perception du mouvement.

Ensuite, lorsque les deux modalités sont présentées en même temps, la modalité visuelle inhibe la modalité auditive. En d'autres termes, le retour auditif perd toute influence sur le mouvement moteur en présence d'une modalité visuelle. Par conséquent, nous supposons que les joueurs débutants de *Cantor Digitalis* se reposent essentiellement sur les indices visuels placés sur la tablette plutôt que sur la hauteur produite pour atteindre les notes précisément, c'est-à-dire plus sur le retour primaire lié à la manipulation de l'instrument que sur le retour secondaire lié à la production sonore. L'influence plus forte du retour visuel permet un contrôle plus simple de l'instrument. Par conséquent, la courbe d'apprentissage du *Cantor Digitalis* dépend des modalités proposées, le contrôle de l'instrument avec retour audio seul nécessitant plus d'entraînement qu'un jeu avec les indices visuels.

Comme le retour auditif a tout de même une influence sur le mouvement lorsqu'il est présenté seul, ces conclusions soulignent le potentiel du *Cantor Digitalis* a être joué avec seulement les modalités auditive et proprioceptive comme un instrument de musique acoustique. Par ailleurs, il a été remarqué chez les joueurs de *Cantor Digitalis* pratiquant régulièrement au sein du *Chorus Digitalis* que la dépendance visuelle diminue avec la progression du musicien.

5.4.2 Spatial vs. temporel, des retours visuels et auditifs complémentaires

Une des raisons de la prépondérance de la modalité visuelle provient d'une représentation de l'espace visuel quasi-similaire à l'espace du mouvement, et propose une résolution de perception spatiale bien supérieure à la modalité auditive. A l'inverse, il a été prouvé que la modalité auditive domine la vision dans la perception des durées, autrement dit sur des aspects temporels [OGMGS14]. En effet, la modalité auditive propose une résolution temporelle supérieure à la résolution spatiale et est prépondérante indépendamment de sa pondération par rapport au visuel ou à l'attention des sujets. De par cette propriété, il a été montré qu'un retour auditif permet de se substituer à un retour kinesthésique dans la production de mouvements répétés chez des personnes privées de proprioception, soulignant l'analogie entre un modèle hiérarchique de la décomposition du mouvement et de la décomposition temporelle des stimuli audio proposés [GRDC00]. De plus, la réponse musculaire à des stimuli auditifs est plus rapide (80 ms) qu'avec des stimuli visuels (125 ms) [LSG67]. La nécessité d'une précision rythmique exemplaire dans la musique occidentale souligne donc l'importance d'un retour auditif, permettant une meilleure synchronisation du joueur avec un tempo et d'autres musiciens.

Finalement, la présence d'une dimension spatiale (intervalles de hauteur) et temporelle (intervalles de temps) dans la musique suggère une complémentarité des retours visuels et auditifs sur le *Cantor Digitalis*. Toutefois, bien que la modalité visuelle facilite grandement le jeu de débutants, nous conseillons d'apprendre à jouer de l'instrument sans retour visuel, bien que cette tâche demande plus d'entrainement. En s'affranchissant de la vue, cela permet d'apporter plus d'attention à la modalité auditive nécessaire pour le jeu en rythme à plusieurs, mais aussi pour faciliter le jeu expressif du musicien.

Chapitre 6

Les gestes pour le contrôle de la synthèse vocale

Sommaire

6.1 Introduction

Plus importante que le choix de l'interface, c'est la conception de l'interaction entre utilisateur et système qui prime sur le développement d'instruments de musique numériques [BL04]. Dans le cas du *Cantor Digitalis*, celle-ci se traduit par un lien entre geste manuel et geste vocal. En effet, par le biais de la tablette graphique, les caractéristiques vocales sont contrôlées par le tracé du stylet sur la tablette. Bien qu'utilisée fréquemment lors des performances du *Chorus Digitalis*, chœur de *Cantor Digitalis*, la question se pose de la pertinence de ce type de contrôle, du lien entre geste chironomique et geste vocal.

Le geste manuel est à la fois guidé et contraint par l'ensemble des muscles et articulations associés au bras et à la main, ainsi que par le moteur de contrôle qu'est le cerveau. On peut donc stipuler que ces contraintes induisent un certain nombre de règles ou de lois que le geste doit respecter, appelées lois biologiques du mouvement. Trois lois empiriques ressortent de la littérature [GKP04] : une loi régissant les aspects temporels du mouvement appelée *Loi de Fitts* [Fit54] ; une loi portant sur les aspects cinématiques du mouvement appelée *Loi de puissance 2/3* [VM83] ; et une loi définissant le coût optimal d'un mouvement appelée *Loi du coût de secousse* [Hog84].

Parallèlement, le geste vocal est aussi soumis à certaines règles. Celles-ci sont d'abord des contraintes physiologiques, limitant la voix chantée aux contraintes de l'appareil vocal. Les règles sont ensuite musicales. En effet, l'esthétique du son produit est une dimension essentielle dans le jeu d'un instrument de musique, et est bâtie à la fois sur des techniques musicales à respecter, et sur la subjectivité du musicien. Parmi ces techniques figurent le *portamento*, *le glissando*, ou le *vibrato* déjà mentionnées au chapitre 4.

La question de la compatibilité entre les contraintes du geste manuel et les contraintes du geste vocal se pose alors. En d'autres termes : existe-t-il des analogies entre gestes manuel et vocal ? Le but de ce chapitre est d'étudier la compatibilité des contraintes manuelles et vocales pour proposer un ensemble de gestes pertinents pour le contrôle expressif de la voix chantée. Des critères objectifs sont présentés et utilisés telles que les lois biologiques régissant le mouvement. Des critères subjectifs sont aussi proposés, telle que l'expérience acquise dans le jeu intensif de l'instrument, selon la sensibilité musicale des musiciens. Le chapitre est divisé en deux études. La première se concentre sur la temporalité des transitions de notes (section 6.2). La deuxième étudie différentes formes expressives et propose un geste adapté à chacune d'elles (section 6.3).

6.2 Temporalité du geste musical

La temporalité du geste musical consiste à étudier les temps nécessaires pour effectuer des transitions de notes. On fait alors apparaître l'analogie entre atteinte de note et visée de cible déjà mentionnée dans le chapitre 4. Il a été montré de nombreuses fois qu'il existe des lois empiriques décrivant la temporalité du mouvement de la main telle que la loi de Fitts. Par ailleurs, quelques études ont montré que la temporalité des transitions de notes dans l'intonation vocale suivait aussi une certaine loi. Le but de cette étude est donc de situer le geste chironomique de contrôle du *Cantor Digitalis* par rapport à ces deux théories. On détaillera successivement les lois chironomiques puis les lois intonatives. La visualisation des données chironomiques du *Cantor Digitalis* est effectuée en dernière partie.

6.2.1 Temporalité du geste

Loi logarithmique (Fitts)

La loi de Fitts [Fit54] a été énoncée et démontrée empiriquement dans le cas de gestes chironomiques. Elle stipule que pour une tâche de pointage dont la consigne est d'atteindre la cible le plus rapidement et le plus précisément possible, le temps moyen MT pour réaliser la tâche est proportionnel à un indice de difficulté ID défini comme suit. Soit A l'amplitude du mouvement, et W la largeur de la cible, c'est-à-dire l'erreur autorisée. L'indice de difficulté s'exprime comme :

$$ID = log_2\left(\frac{2A}{W}\right) \tag{6.1}$$

Il est exprimé en bits. Il croit avec l'amplitude, et décroit pour une plus grande largeur de cible. Une démonstration de cette formulation est obtenue en considérant le *modèle déterministe de correction itérative du mouvement*. Ce dernier suggère qu'un mouvement est conduit selon une série de sous-mouvements guidés par un retour visuel ou kinesthésique.

Une autre expression de l'indice de difficulté proposée par MacKenzie [Mac92] et plus largement utilisée est :

$$ID = log_2\left(\frac{A}{W} + 1\right) \tag{6.2}$$

Elle permet d'éviter d'avoir un indice de difficulté nul pour de petites amplitudes (ou grandes largeurs), et se rapproche plus de la formulation de Shannon dans ses travaux sur la capacité de transmission d'un canal de télécommunication. Il est aussi précisé que pour que la loi soit la plus juste possible, le taux d'erreur dans la réalisation de la tâche doit être de 4%. La largeur de la cible W est donc à modifier *a posteriori* dans le cas où le taux d'erreur serait différent. La relation entre temps moyen et indice de difficulté est donnée par l'équation suivante, où les coefficients a et b sont obtenus empiriquement :

$$MT = a + b.ID = a + b.log_2\left(\frac{A}{W} + 1\right) \tag{6.3}$$

Le coefficient a est l'ordonnée à l'origine, soit le temps moyen pour une tâche de difficulté nulle. Il correspond à l'ensemble des gestes qui ne participent pas à la réalisation de la tâche, mais qui prennent un certain temps (temps de réaction par exemple).

Le coefficient b est la pente de la droite obtenue. Il est exprimé en secondes par bit et son inverse est appelé indice de performance $IP = 1/b$ en $bit.s^{-1}$. Ce dernier représente la quantité d'information apportée pour un temps donné. Celui-ci étant théoriquement constant pour tout indice de difficulté, il permet donc de quantifier l'efficacité du geste réalisant la tâche. Un indice de performance élevé implique une pente faible et entraîne des temps courts pour réaliser tout type de difficulté. Un indice de performance faible entraîne des temps longs dès que la difficulté augmente.

La comparaison des indices de performance pour des tâches similaires avec des gestes différents est largement utilisée pour comparer l'efficacité des gestes. Card [CMR91] a comparé des indices de performance pour diverses interfaces (souris, headmouse, doigt) dans la sélection de texte. McGuffin [MB02] a comparé les indices de performance d'une tâche de sélection avec et sans aide à la sélection (*expanding targets*). Cockburn [CF03] a comparé les indices de performance pour différentes corrections (*sticky icons*, *expanding targets*, *goal crossing*).

Cette formulation est valable pour une tâche de pointage à une dimension. Celle-ci a été par la suite étendue pour être appliquée dans des conditions variées. On peut citer une extension pour une tâche à deux dimensions [MB92], la redéfinition de la loi pour une tâche de suivi de trajectoire [AZ97], ou une autre redéfinition pour une tâche de traversée de cible (*goal crossing*) [AGZ10].

Enfin, la loi de Fitts a été largement vérifiée dans de nombreux contextes [SM04], qu'ils soient dans le cadre d'une expérience de laboratoire ou dans des tâches courantes [CBBL07], [GRH12]. Sutter *et al.* ont par ailleurs démontré que dans le cas où un découplage sensori-moteur a lieu, où la perception est distale (visuelle) et l'action proximale, la loi de Fitts s'applique sur l'environnement distal du geste [SMB11].

Loi linéaire (Schmidt)

Bien que la loi de Fitts soit la plus couramment utilisée, celle-ci requiert la réalisation d'un mouvement le plus rapide et le plus précis possible. De plus, elle stipule l'existence du *modèle déterministe de correction itérative du mouvement* où un mouvement est conduit selon une série de sous-mouvements guidés par un retour visuel ou kinesthésique. La trajectoire progresse donc vers la cible itérativement. Dans le cas où ces conditions ne sont pas respectées, d'autres lois sont proposées.

Schmidt *et al.* affirment que lorsque la durée du mouvement est trop courte, (inférieure à 200 ms), le guidage du mouvement par les retours visuel ou kinesthésique n'a pas le temps d'avoir lieu [SZH+79]. C'est la théorie de *programmation moteur* qui est alors suivie, stipulant que tous les paramètres du mouvement sont établis *a priori* et que le mouvement n'est pas rectifié pendant sa course. Ce modèle met en jeu la variabilité de l'impulsion générant le mouvement et conduit à la formulation suivante :

$$W_e = k\frac{A}{MT} \tag{6.4}$$

W_e est la taille effective de la cible, A l'amplitude du mouvement à parcourir, MT le temps moyen observé pour la réalisation de plusieurs mouvements, et k une constante de proportionnalité. Il s'agit là d'une relation linéaire entre temps moyen et indice de difficulté de la tâche. Cette formulation a été affinée par Meyer *et al.*, notamment par la proposition d'un modèle d'impulsion du mouvement plus précis [MKSW82].

Finalement, Wright *et al.* ont conduit une expérience distinguant les modalités d'apparition de la loi logarithmique de Fitts et la loi linéaire [WM83]. Ils ont montré que la loi de Fitts obéit à une contrainte *spatiale*, c'est-à-dire l'atteinte d'une cible le plus rapidement possible. A l'inverse, la loi linéaire obéit à une contrainte *temporelle*, c'est-à-dire l'atteinte d'une cible en un temps donné.

Loi racine carrée (Meyer)

Une autre formulation de la loi temporelle est proposée par Meyer *et al.*, reposant sur un modèle stochastique du mouvement [MKA+88]. La loi de Fitts ne prend pas en compte les variabilités des sous-mouvements mis en jeu. Meyer décompose ici le mouvement en deux sous-mouvements, dont l'écart-type de leur position d'arrivée est proportionnel à la vitesse du sous-mouvement. Optimiser le mouvement total consiste à trouver un compromis entre vitesse du premier sous-mouvement et correction à effectuer ensuite. Une vitesse élevée au

départ entraîne une probabilité d'erreur plus grande nécessitant un deuxième sous-mouvement important rallongeant le temps d'exécution. A l'inverse, une vitesse faible au départ minimise l'erreur d'atteinte de cible et l'utilisation d'un deuxième sous-mouvement mais allonge le temps d'exécution. Un tel compromis est modélisé de manière simplifiée par :

$$MT = a + b\sqrt{\frac{A}{W}} \tag{6.5}$$

Les courbes racine carrée et logarithme étant de formes similaires, Meyer remarque que cette loi stochastique et la loi déterministe de Fitts donnent des résultats proches pour des rapports A/W compris entre 4 et 64.

6.2.2 Temporalité de l'intonation

Les temps de transitions moyens pour chanter des intervalles isolés ont été étudiés d'abord par Sundberg [Sun73] puis par Xu [XS00] et enfin Mori [MOKH04]. Bien qu'étudiant la même grandeur, les trois études diffèrent par le type de sujets, de stimuli et de mesures réalisées. D'abord, les groupes de sujets recrutés ne présentent pas les mêmes qualités. Sundberg compare des sujets entraînés et non-entraînés, ainsi que hommes et femmes. Xu compare des sujets de langues maternelles anglaise et chinoise. Mori fait des mesures sur un seul chanteur professionnel.

Ensuite, le choix des stimuli varie aussi d'une étude à l'autre. Sundberg et Xu utilisent des stimuli similaires : des oscillations de hauteur entre deux notes constantes. Les intervalles étudiés entre les deux notes sont 4, 7 et 12 demi-tons. Sundberg demande un minimum de 8 oscillations en partant de la note la plus basse. Xu demande 5 oscillations en partant soit de la note la plus basse, soit de la plus haute. Mori quant à lui demande des stimuli de 3 notes montant/descendant ou descendant/montant où les notes de départ et d'arrivée sont fixes, et la note du milieu définit l'intervalle. Les intervalles de 1 à 12 demi-tons sont étudiés. Il est demandé de chanter legato dans les trois études. En revanche le tempo diffère. Sundberg demande d'effectuer les transitions le plus vite possible, tout en imposant un tempo au métronome. Xu demande des oscillations de 4 ou 6 Hz, tout en demandant d'imiter un stimulus. Mori ne précise pas de tempo, seulement une interprétation naturelle.

Enfin les mesures effectuées diffèrent entre les études. Le temps de réponse T_r est défini par Sundberg par le temps nécessaire pour parcourir les 3/4 de l'excursion maximale, d'1/8 à partir du minimum jusqu'à 1/8 avant le maximum (ou vice-versa). Xu reprend cette définition et calcule aussi le temps T_e nécessaire pour parcourir l'excursion totale. Mori utilise aussi le temps d'excursion totale T_e. Il est démontré empiriquement que le temps total T_e est deux fois plus élevé que le temps de réponse T_r donné par Sundberg. Ces temps de transition sont représentés en figure 6.1. Xu et Mori calculent par ailleurs la taille de l'excursion et Xu calcule la vitesse d'excursion.

La figure 6.2 affiche les temps de transition moyen, sujets confondus, pour chaque intervalle, de Sundberg (rouge), Xu (bleu) et Mori (noir). Pour homogénéiser les résultats entre les temps de réponse T_r mesurés par Sundberg et les temps d'excursions T_e mesurés par Xu et Mori, les temps de réponse ont été multipliés par 2. Par ailleurs, les intervalles représentés sont les intervalles effectifs et non théoriques. Dans le cas de l'étude de Xu, bien que les intervalles théoriques sont de 4, 7 et 12 demi-tons comme Sundberg, les intervalles effectifs chantés par les sujets sont beaucoup plus faibles. On peut aussi noter que les données de Mori, ainsi que celles de Sundberg représentées par des carrés sont des données de chanteurs entraînées. Les autres sont des données de chanteurs non-entraînés.

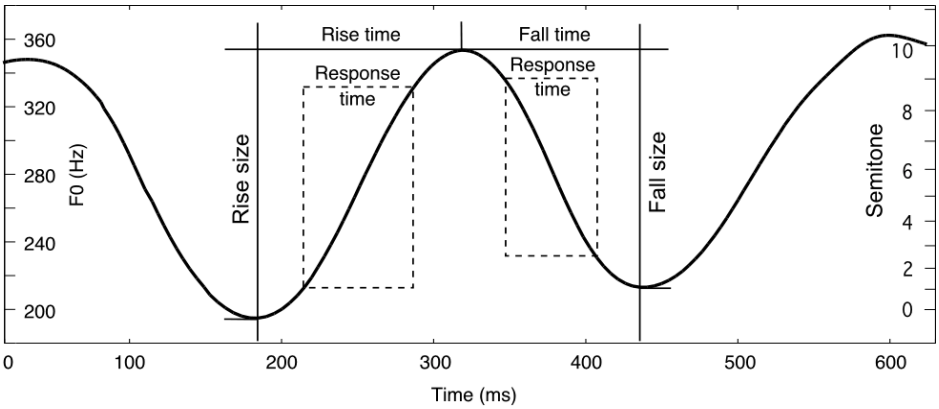

Figure 6.1 – *Définition du temps de réponse (Response time) et du temps d'excursion (Rise/Fall time), d'après [XS00].*

Les trois études montrent que le temps moyen d'exécution est dépendant de la taille de l'intervalle. On remarque aussi une asymétrie entre les intervalles montants et les intervalles descendants. En général, les temps moyens présentent beaucoup moins de variations pour des intervalles descendants. Ceci est vérifié chez les sujets non entraînés de Sundberg et le sujet de Mori. Cette asymétrie est expliquée dans les 3 études par la plus grande complexité à tendre les cordes vocales pour des notes de plus en plus aiguës, qu'à les détendre. Cependant, l'asymétrie est beaucoup moins nette pour les sujets entraînés de Sundberg et les sujets de Xu. Sundberg suppose que les chanteurs professionnels s'entraînent pour atténuer cette asymétrie et uniformiser leur technique. Cependant ceci est en contradiction avec le sujet de Mori : l'asymétrie présente dans ses données se rapproche plus des sujets non-entraînés de Sundberg. Il est difficile d'expliquer ce phénomène, du fait du nombre très réduit de sujets.

Par ailleurs, chaque étude propose des résultats spécifiques. Sundberg montre que les femmes ont des temps moyens de transitions plus faible que les hommes. Cependant ceci n'est pas vérifié par Xu. Ce dernier montre que les intervalles chantés par les anglophones sont plus grands. Il en résulte une plus grande vitesse de transition. Il montre aussi que les temps de transitions sont décorrelés de la vitesse de transition. En effet ses résultats montrent un temps moyen quasi-constant en fonction de l'intervalle. La vitesse augmenterait donc avec l'intervalle.

Mori étudie les dépassements en fin de transition et observe aussi une asymétrie entre intervalles montants et descendants. Les dépassements sont constants et faibles pour les intervalles montants, et dépendent de l'intervalle pour les intervalles descendants. Il conclu que la transition obéit à un système du deuxième ordre, amorti pour les intervalles montants, et sous-amorti pour les intervalles descendants.

Ces trois études mettent en évidence l'asymétrie des temps moyens de transitions en fonction du sens de l'intervalle. L'observation des résultats incite à suggérer la loi suivante décrivant l'évolution du temps de transition moyen MT en fonction de l'intervalle A (pour amplitude) :

$$\begin{cases} MT = a + b.A & si\ A > 0 \\ MT = c & si\ A < 0 \end{cases} \tag{6.6}$$

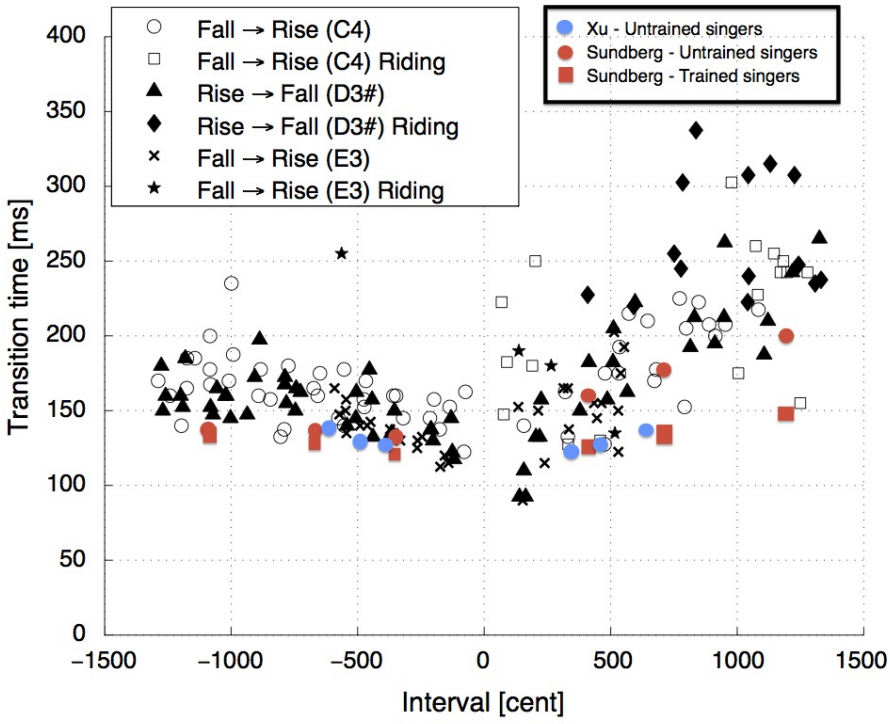

FIGURE 6.2 – *Temps d'excursion totale en fonction de la taille de l'intervalle, d'après [Sun73], [XS00] et [MOKH04].*

Cependant ces résultats sont incomplets : Sundberg et Xu n'ont travaillé que sur un nombre réduits d'intervalles, ne permettant pas d'extrapoler une tendance générale. A l'inverse Mori a exploré une gamme d'intervalles plus importants, mais un seul sujet ne suffit pas pour tirer des conclusions solides. Il serait donc nécessaire de compléter ces études pour pouvoir proposer une loi empirique.

6.2.3 Observation des gestes chironomiques

Les deux études précédentes, sur le geste et la voix ont émis différentes lois empiriques sur des tâches techniques similaires (atteinte de cible).

$$\begin{cases} & Chironomie & & \\ MT & = & a + b.log_2\left(\frac{|A|}{W} + 1\right) & si\ contrainte\ spatiale \\ MT & = & a + b.|A| & si\ contrainte\ temporelle \\ & Voix & & \\ MT & = & a + b.A & si\ A > 0 \\ MT & = & c & si\ A < 0 \end{cases} \tag{6.7}$$

Cela soulève donc la question de savoir s'il s'agit de deux mécanismes différents, décrits par des lois différentes, ou s'il s'agit d'un mécanisme commun où les lois exprimeraient de manière différente un comportement commun. Afin d'avoir un premier aperçu, une étude est réalisée sur différents ensembles de données récoltées lors des expériences précédentes.

Choix des données

On sélectionne d'abord les données d'imitation de mélodies récoltées pour le chapitre 3. Seuls les 9 meilleurs sujets en termes de justesse et précision sont conservés sur les modalités voix et chironomie avec retour audio. On sélectionne ensuite les données d'imitation de mélodies récoltées pour le chapitre 4 pour le jeu *legato*. Enfin, on utilise des enregistrements de répétitions du *Chorus Digitalis*, ensemble de *Cantor Digitalis*. Comme pour les expériences, les données de la tablette telles que les coordonnées spatiales du stylet, la pression et le temps, sont enregistrés pendant le jeu de morceaux en condition musicale réelle. Six sujets sont enregistrés à travers les différentes répétitions. Les données de chaque enregistrement sont traitées suivant les trois étapes suivantes.

Segmentation des données : Toutes les données sont rassemblées en une structure regroupant méta-données décrivant le contexte de l'enregistrement et données brutes (positions X, Y et pression P du stylet pour la tablette, et fichier audio pour la voix). La hauteur contrôlée est calculée à partir de la position X du stylet sur la tablette, et extraite avec l'algorithme STRAIGHT [KMKdC99] sur les données voix.

Alignement des données : Afin de pouvoir extraire les notes une par une, une première segmentation est réalisée au niveau de chaque transition. On crée un signal cible u constant par morceaux. Les notes de ce signal (paliers) sont parfaitement justes, et les transitions parfaitement identifiées (instantanées). Ce signal est généré à partir d'un fichier MIDI correspondant à la partition jouée de chaque extrait. Il est aligné sur le signal de hauteur par la méthode de *Dynamic Time Warping (DTW)*. Aucune condition de lissage n'est utilisée car u est constant par morceaux. Les transitions de u sont alors alignées avec celles de s, et il est donc possible d'isoler chaque note.

Extraction des notes par histogramme : Dans un contexte réel de chant, beaucoup d'ornements et d'irrégularités s'ajoutent à l'intonation pure, notamment du vibrato. Celui-ci rend donc la détection de note plus difficile. La valeur de la hauteur évoluant sans cesse au cours du temps, on peut supposer que la valeur de la hauteur cible est celle atteinte le plus grand nombre de fois. On utilisera donc une méthode par histogramme. Pour chaque portion de signal compris entre deux transitions, la valeur et la position de la note jouée sont extraites en prenant la valeur maximum de l'histogramme du signal.

Extraction des temps de transition par régression : Entre chaque note identifiée précédemment, la transition est modélisée par une fonction sigmoïde. Les caractéristiques de la transition sont ensuite extraites du modèle.

On considère la sigmoïde d'équation :

$$\sigma_j(x) = A\left(\frac{e^{\lambda x}-1}{e^{\lambda x}+1}\right) + D \tag{6.8}$$

Où A est l'amplitude de la sigmoïde, D la distance du centre de la sigmoïde à l'origine et λ un coefficient définissant la pente de la courbe à l'origine. En appelant s_j la portion de signal joué entre les notes identifiées n_j et n_{j+1}, A et D sont calculées selon :

$$\begin{cases} A &= \frac{n_{j+1}-n_j}{2} \\ D &= \frac{n_{j+1}+n_j}{2} \end{cases} \tag{6.9}$$

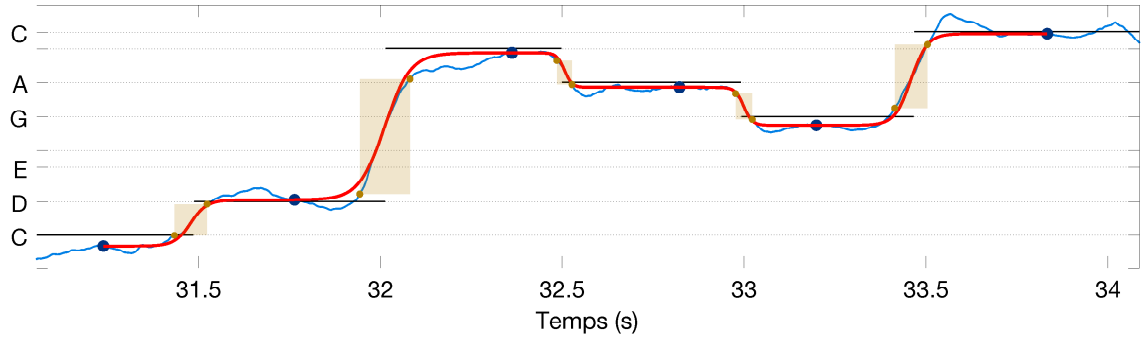

FIGURE 6.3 – *Extraction des caractéristiques d'une mélodie de voix chantée. La hauteur extraite du signal est en bleu et les disques foncés sont les notes identifiées. La modélisation par des sigmoïdes est en rouge et les rectangles orange indiquent les périodes de transition.*

Une régression non-linéaire est ensuite effectuée pour trouver le coefficient λ qui approche le mieux la sigmoïde du signal analysé.

En supposant la sigmoïde centrée sur l'instant t_0, on définit alors le temps de réponse t_{r_τ} de la sigmoïde comme le temps nécessaire pour aller de $\sigma\left(t_0 - \frac{t_{r_\tau}}{2}\right) = -\tau A + D$ à $\sigma\left(t_0 + \frac{t_{r_\tau}}{2}\right) = \tau A + D$ avec τ compris entre 0 et 1. On a alors :

$$t_{r_\tau} = \frac{2}{\lambda} ln\left(\frac{1+\tau}{1-\tau}\right) \tag{6.10}$$

En reprenant la définition de Sundberg [Sun73], le temps de réponse est le temps nécessaire pour effectuer 75% de l'excursion maximale. En prenant $\tau = 0.75$ on a alors $t_{r_{75}} = \frac{1.95}{\lambda}$.

La figure 6.3 montre l'extraction du temps de transition. La courbe rouge représente la modélisation en sigmoïdes. Les rectangles oranges montrent les périodes de transition à 75%.

Nous disposons alors pour chaque transition de son amplitude théorique, son amplitude réelle, et sa durée. L'ensemble de ces transitions sont partagées en trois sous-ensemble : le sous-ensemble des transitions chantées à la voix, extraites de l'expérience du chapitre 3 ; le sous-ensemble des transitions jouées à la tablette dans un contexte expérimental, extraites de la condition tablette et audio du chapitre 3 et du chapitre 4 ; le sous-ensemble des transitions jouées à la tablette dans un contexte musical, extraites des répétitions du *Chorus Digitalis*. Pour chaque sous-ensemble, on peut calculer la largeur effective W_e de la cible à atteindre en faisant en sorte que seules 4% des notes soient en dehors de la cible [Mac92].

Observations générales

Les temps de ces transitions pour chaque sous ensemble sont regroupés en fonction de l'amplitude théorique de l'intervalle à jouer dans des boîtes sur la figure 6.4. Chaque boîte contient 50% des valeurs et le temps médian pour chaque amplitude est indiqué par un disque noir. Les temps des transitions vocales sont affichés en haut, les transitions chironomiques en conditions expérimentales au milieu et les transitions chironomiques musicales en bas.

Deux caractéristiques émergent de cette figure. D'abord, les temps de transition de la voix chantée sont deux fois plus faibles que ceux des tracés chironomiques. On en déduit

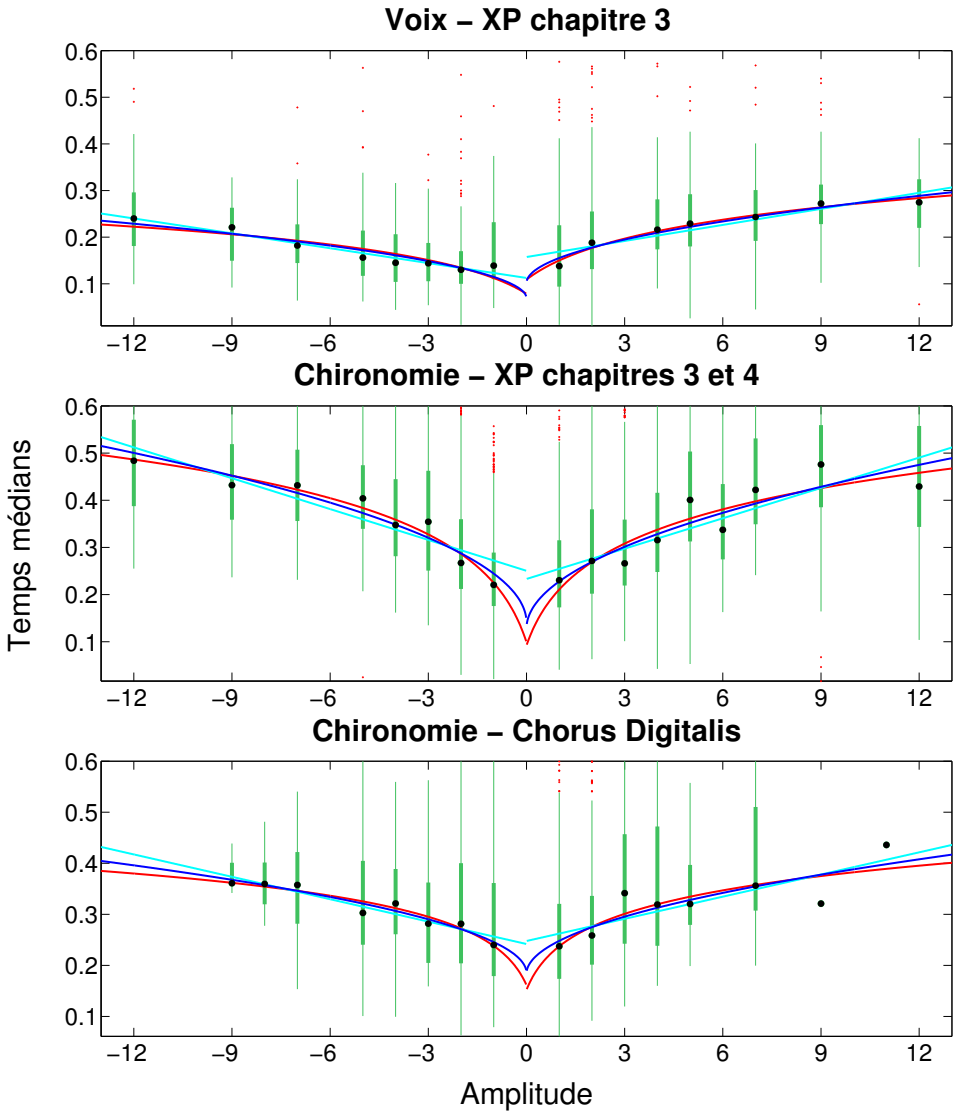

Figure 6.4 – *Temps médians en fonction de l'amplitude pour les trois conditions.*

que l'intonation se contrôle plus rapidement par les cordes vocales que par le geste manuel. Une explication est la différence de déplacement des organes responsables du contrôle de l'intonation. Dans le cas du chant, ce sont les positions des aryténoïdes et de la thyroïde qui contrôlent la hauteur, nécessitant un déplacement très faible (quelques millimètres). A l'inverse, le mouvement du bras contrôlant le stylet est de l'ordre du centimètre. En supposant une vitesse de déplacement proche pour chaque organe, il en résulte un temps de transition plus important pour le geste manuel.

Ensuite, la pente de l'évolution des temps médians vocaux en fonction des amplitudes positives est plus importante que pour les amplitudes négatives. On retrouve l'asymétrie entre intervalles montants et descendants observée sur les sujets chanteurs de Sundberg et Mori. Celle-ci s'explique par la plus grande facilité à détendre les muscles pour descendre un intervalle, que de les tendre pour monter. A l'inverse, pour le geste manuel, rien n'indique qu'un mouvement vers la droite est plus difficile qu'un mouvement vers la gauche. C'est pourquoi les courbes chironomiques sont symétriques.

	Voix	**Chironomie Expériences**	**Chironomie Chorus Digitalis**
Loi linéaire	$R^2_- = \mathbf{0.94}$ $R^2_+ = 0.86$	$R^2_- = 0.84$ $R^2_+ = 0.77$	$R^2_- = 0.91$ $R^2_+ = 0.70$
Loi racine carrée	$R^2_- = 0.87$ $R^2_+ = 0.94$	$R^2_- = 0.93$ $R^2_+ = 0.83$	$R^2_- = 0.93$ $R^2_+ = 0.72$
Loi logarithmique	$R^2_- = 0.82$ $R^2_+ = \mathbf{0.97}$	$R^2_- = \mathbf{0.96}$ $R^2_+ = \mathbf{0.83}$	$R^2_- = 0.93$ $R^2_+ = 0.72$

TABLE 6.1 – *Coefficients des régressions linéaire, racine carrée et logarithmique effectuées sur les données voix, chironomie expérimentale et chironomie musicale. R^2_- (resp. R^2_+) sont les coefficients des régressions sur les amplitudes négatives (resp. positives).*

Régressions

Pour chaque sous-ensemble, trois régressions sur les temps médians ont été effectuées : une régression linéaire décrivant la loi chironomique de Schmidt [SZH+79] ou la loi observée par Sundberg [Sun73] (cyan), une régression racine carrée décrivant la loi de Meyer [MKA+88] (bleu) et une régression logarithmique décrivant la loi de Fitts [Mac92] (rouge). Le tableau 6.1 reporte les coefficients de détermination R^2 obtenus pour chacune des régressions. Il faut noter que pour chaque sous-ensemble, les régressions ont été appliquées indépendamment sur les données d'amplitudes négatives (coefficients R^2_-) et positives (coefficients R^2_+) afin de détecter d'éventuelles asymétries.

On obtient des coefficients R^2 très proches pour les régressions de chaque sous-ensemble, ce qui se manifeste par des courbes relativement similaires. Celles-ci se différencient principalement à des amplitudes inférieures à 2 demi-tons ou supérieures à 9 demi-tons, là où nos données sont les plus éparses. Pour chaque sous-ensemble, nous pouvons malgré tout relever des tendances, qui sont plus hypothétiques que révélatrices de lois empiriques. Les coefficients de régression les plus élevés pour chaque sous-ensemble sont notés en gras dans le tableau.

Pour le sous-ensemble des transitions vocales, la loi linéaire décrit mieux l'évolution des temps de transitions en fonction des amplitudes d'intervalles négatives. Nous retrouvons alors une tendance similaire aux données de Sundberg [Sun73], Xu [XS00] et Mori [MOKH04]. L'évolution des temps de transition en fonction des amplitudes positives est mieux représentée par une loi logarithmique ou racine carrée qui sont très proches. Cette tendance est due au faible temps de transition observé pour l'amplitude de 1 demi-ton.

Le sous-ensemble chironomique en condition expérimentale est quant à lui modélisé par une loi logarithmique ou racine carrée, pour les amplitudes négatives et positives. L'imitation de mélodies demandée aux sujets se traduit par une tâche de pointage sur la tablette. Ces données semblent donc suivre une loi de Fitts [Mac92]. On relève des indices de performance (inverse de la pente de la loi de Fitts tracée en fonction de l'indice de difficulté) de $IP_- = -11.4$ bits/s et $IP_+ = 12.1$ bits/s. A titre de comparaison, les indices de performance d'une souris et d'un doigt sur un écran sont respectivement de $IP_{souris} = 10$ bits/s et $IP_{doigt} = 40$ bits/s [CMR91]. L'indice obtenu pour la chironomie est donc proche de celui de la souris, ce qui paraît être raisonnable quant à la maniabilité de chacune des interfaces.

Enfin, à cause d'une plus grande dispersion des données et un manque d'information pour les grandes amplitudes, les trois lois modélisent aussi bien les données chironomiques

dans un contexte musical. Plusieurs hypothèses peuvent être faites concernant les différentes lois. D'abord, on peut supposer que s'agissant d'une tâche de pointage chironomique, une loi logarithmique (Fitts) devrait apparaître. Néanmoins, dans un contexte musical, la contrainte temporelle est essentielle. Par conséquent, les observations de Wright *et al.* suggèrent qu'une loi linéaire serait plus appropriée [WM83]. De plus, si l'intention des musiciens est d'imiter l'intonation vocale dans leur jeu, il est possible que la loi linéaire observée pour la voix transparaisse dans le jeu chironomique musical. Enfin, une loi linéaire est synonyme d'une plus grande uniformité dans le jeu des intervalles, ce qui est recherché par les chanteurs professionnels [Sun73].

6.2.4 Discussion et conclusions

Trois lois empiriques ont été proposées et observées dans la description de l'évolution des temps de transition d'intervalles en fonction de leurs amplitudes. Dans le cas chironomique sur des tâches de pointage, la loi de Fitts (logarithmique) apparaît lorsque la tâche est conditionnée par la réalisation d'un geste rapide et précis [Fit54], [Mac92]. Lorsque la consigne est l'atteinte d'une cible en un temps donné, une loi linéaire décrit mieux les données [SZH+79], [MKSW82], [WM83]. Dans le cas vocal sur des tâches de chant d'intervalles réguliers, une loi linéaire apparaît avec une asymétrie entre intervalles montants et descendants [Sun73], [XS00], [MOKH04].

L'observation de données vocales et chironomiques a permis de retrouver certaines de ces lois, dans des conditions vocales (linéaire et asymétrique), et chironomiques expérimentales, soit des tâches de pointage (logarithmique, Fitts). En revanche, pour la tâche intermédiaire qu'est l'imitation de la voix chantée en condition musicale par un geste chironomique, il n'a pas été possible d'identifier une tendance. Ceci s'explique par la grande variabilité de ces données qui dépendent essentiellement de la sensibilité musicale des musiciens.

Finalement, l'observation de données récoltées dans plusieurs conditions ne nous permet pas de conclure sur la modélisation des temps de transitions dans le chant chironomique par une loi empirique. En revanche, cette étude nous a permis d'observer des tendances validant certaines hypothèses et incitant à poursuivre l'investigation dans cette direction. Une étape cruciale pour la poursuite de cette étude est l'homogénéisation de la récolte des résultats, notamment dans les enregistrements en condition musicale, en imposant des contraintes similaires à tous les sujets. La principale hypothèse extraite de cette étude est l'évolution linéaire des temps de transitions pour le jeu chironomique musical. Cela s'expliquerait par la contrainte temporelle imposée par la tâche musicale, et le désir d'uniformiser les temps de transition pour un jeu plus homogène, à l'image des chanteurs professionnels. Une piste de recherche serait donc l'étude de l'évolution de cette loi, d'une tâche chironomique de pointage (logarithmique) à une tâche chironomique musicale (linéaire).

6.3 Proposition de gestes musicaux et caractérisation

De nombreuses techniques de jeu permettent une expressivité mélodique dans la pratique d'instruments de musique. On peut nommer le *portamento*, le *glissando*, le *vibrato*, le *staccato* ou le jeu virtuose. Des méthodes gestuelles consensuelles proposées par différentes "écoles" permettent de réaliser chacune de ces techniques de manières confortable et efficace sur chaque instrument. Dans le cas du violon par exemple, le glissando s'effectue en déplaçant un doigt de manière continue sur la corde. Un vibrato se réalise en oscillant la main gauche, faisant rouler le doigt sur la corde.

Dans l'exploration du *Cantor Digitalis*, il paraît important de proposer une méthode de jeu adaptée aux contraintes chironomiques imposées par l'instrument. Cette partie propose donc une suite de gestes qui semblent les plus adaptés aux techniques expressives mentionnées précédemment. Deux approches sont possibles : les gestes peuvent être soit sélectionnés selon des mesures objectives, soit par expérience résultant d'une pratique intensive de l'instrument. Nous présentons d'abord les outils théoriques disponibles pour le choix des gestes que sont les coûts du mouvement. Une liste des techniques vocales les plus utilisées est ensuite donnée, détaillant les caractéristiques du son associées. Enfin, des gestes sont proposés pour chaque technique, satisfaisant les contraintes musicales associées et caractérisés en terme de coûts.

6.3.1 Lois du mouvement

Loi cinématique

La loi communément appelée "puissance 2/3", énoncée et étudiée par Viviani [VT82],[VM83], montre que pour un mouvement continu de trajectoire en deux dimensions on a :

$$V(t) = k.R(t)^{\beta} \tag{6.11}$$

avec V la vitesse tangentielle et R le rayon de courbure de la trajectoire. Il a été montré empiriquement que $\beta = 1/3$ [LTV83]. k est un gain constant sur des morceaux de trajectoires appelés unités. Il dépend du périmètre total de la trajectoire, et du couplage entre périmètre total et périmètre de l'unité de trajectoire. Il est aussi fonction du tempo d'exécution de la trajectoire [VC85].

Tout comme la loi de Fitts, cette relation empirique est vérifiée dans de multiples applications telles que l'écriture, mais aussi l'articulation de la parole [TW04], [PF08]. Il a été par ailleurs montré que nous sommes capable d'entendre la loi de puissance 2/3 par assimilation de la dynamique de trajectoire d'un stylo sur une surface et du son produit par ce dernier [TAKM+14].

Coûts du mouvement

Tout mouvement nécessite un effort et a par conséquent un coût. Afin d'éviter des efforts inutiles, chaque mouvement est planifié pour minimiser son coût, relativement à l'objectif du mouvement. Pour chaque objectif, il existe différents critères d'optimisation [Nel83]. La notion d'optimum dépend de la grandeur à minimiser. La figure 6.5 représente les différents profils de vitesse d'un mouvement sur une dimension en fonction du temps, minimisant différentes grandeurs détaillées par la suite. L'aire sous chaque courbe représente la distance à parcourir D et est la même pour chaque courbe. La pente de chaque courbe représente l'accélération. On notera par la suite x_0 la position initiale du mouvement au temps t_0, et T le temps total du mouvement.

Minimisation de la vitesse : La courbe pleine marquée V présente le profil de vitesse minimisant la vitesse maximale. Idéalement la courbe serait un créneau de valeur égale à la vitesse moyenne $V_{moy} = D/T$. En pratique, elle est plus proche d'un trapèze, où le début et la fin du mouvement sont réalisés avec une accélération et décélération maximales. Un tel profil permet un geste régulier. Il est par exemple pratiqué chez les violonistes avec l'archet sur la corde lors de notes tenues. En supposant la vitesse constante sur toute la durée du mouvement, on définit la trajectoire minimisant la vitesse comme une rampe, soit un polynôme d'ordre 1.

$$x_{i_{opt}}(t) = x_0 + \frac{D}{T}t \tag{6.12}$$

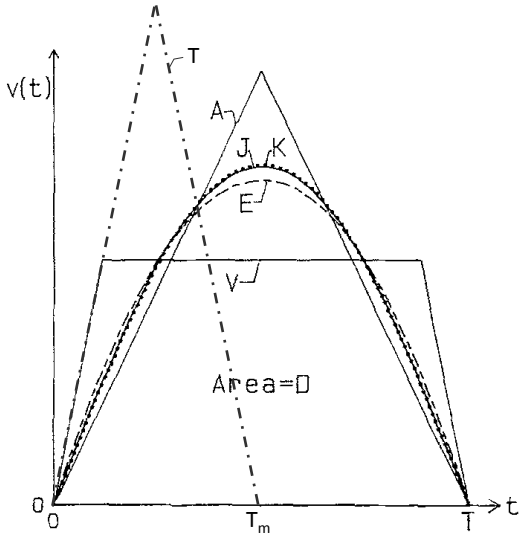

Figure 6.5 – *Profils de vitesse minimisant le coût de : T - Temps ; A - Accélération ; V - Vitesse ; E - Energie ; J - Jerk ; K - Profil de vitesse d'un système masse-ressort, d'après [Nel83].*

Le coût du mouvement à minimiser s'appelle le coût d'impulsion défini comme le maximum de la vitesse de la trajectoire :

$$C_{i_{opt}} = max|\dot{x}_{i_{opt}}(t)| = \frac{D}{T} \tag{6.13}$$

Minimisation de l'accélération : La courbe minimisant l'accélération pour parcourir une distance D en un temps T (indiqué en abscisses) est montrée par la courbe pleine marquée A. Il s'agit d'un profil triangulaire dont la pente (l'accélération) est minimum. Le deuxième principe de la mécanique, où l'accélération d'un objet est proportionnelle à la somme des forces qui s'appliquent sur ce dernier, nous indique qu'un mouvement d'accélération minimum est un mouvement nécessitant une force minimum. Le passage brutal entre accélération et décélération se montre néanmoins contraignant. Le profil de vitesse étant une rampe croissante (resp. décroissante) sur la première (resp. deuxième) moitié du mouvement, la trajectoire résultante est une concaténation de deux paraboles, soit un polynôme d'ordre 2. En prenant aux conditions aux limites des vitesses initiales et finales nulles on obtient l'expression :

$$\begin{cases} x_{f_{opt}}(t) &= x_0 + \frac{2D}{T^2}(t-t_0)^2 & \text{si } t \leq t_0 + \frac{T}{2} \\ x_{f_{opt}}(t) &= x_0 + \frac{2D}{T^2}\left[-\frac{T^2}{2} + 2T(t-t_0) - (t-t_0)^2\right] & \text{si } t > t_0 + \frac{T}{2} \end{cases} \tag{6.14}$$

La minimisation de l'accélération étant équivalente à la minimisation de la force à appliquer au mouvement, le coût à minimiser s'appelle le coût de force et est défini comme le maximum de l'accélération :

$$C_{f_{opt}} = max|\ddot{x}_{f_{opt}}(t)| = 4\frac{D}{T^2} \tag{6.15}$$

A l'inverse, il est possible de maximiser la force appliquée au mouvement afin d'obtenir des meilleures performances telles que maximiser la taille de l'intervalle D ou minimiser le

temps de parcours T. La courbe en tirets-pointillés marquée T montre le profil nécessaire pour parcourir la distance D en un temps minimal T_m. On voit que l'accélération est constante et maximale, sur la première partie du mouvement. La décélération est ensuite aussi constante et maximale sur la deuxième partie du mouvement. Cette accélération maximale est définie par les limites physiologiques associées à l'anatomie de la personne. Réaliser le mouvement le plus rapide possible demande beaucoup d'efforts pour accélérer et décélérer le mouvement.

Minimisation de la secousse : Les courbes en tirets marquée E et pleine marquée J sont proches et représentent respectivement la minimisation de l'énergie et de la secousse (*jerk*) du mouvement. La première permet le mouvement le moins fatiguant physiquement. Il s'agit de profils adoptés pour réduire la difficulté physique de la tâche, dans des tâches sportives par exemple. Le deuxième profil indique la fluidité du mouvement. La secousse étant la dérivée de l'accélération, minimiser la secousse revient à minimiser les changements brutaux d'accélération et de produire un mouvement lisse. Dans un contexte artistique, où les mouvements doivent être fluides, ce type de coût paraît approprié. Hogan *et al.* montrent que la trajectoire minimisant le coût de secousse est un polynôme d'ordre 5. En prenant pour conditions aux limites des vitesses et accélérations initiales et finales nulles on obtient l'expression :

$$x_{s_{opt}}(t) = x_0 + D\left[10\left(\frac{t-t_0}{T}\right)^3 - 15\left(\frac{t-t_0}{T}\right)^4 + 6\left(\frac{t-t_0}{T}\right)^5\right] \tag{6.16}$$

Le coût de secousse est définit comme la puissance de la secousse. Le coût minimal de secousse est alors :

$$C_{s_{opt}} = \frac{1}{2}\int_{t_0}^{T} \dddot{x}^2_{s_{opt}}(t)dt = 360\frac{D}{T^5} \tag{6.17}$$

Le coût de secousse a été particulièrement étudié car la trajectoire minimisant ce dernier apparaît empiriquement dans de nombreux mouvements chez le primate ou l'homme [Hog84], comme celui de l'écriture par exemple [EF87]. Le modèle a ensuite été perfectionné afin de prendre en compte le passage par des points intermédiaires [FH85], [HF87] ou d'obtenir des profils asymétriques de vitesse [Nag89]. Il a par ailleurs été montré qu'une trajectoire optimale au sens de la minimisation du coût de secousse vérifiait la loi de puissance 2/3 [VF95]. Il y a donc une compatibilité des lois du mouvement.

Analogie avec le système masse-ressort : La courbe en pointillés marquée K représente le profil de vitesse d'un système masse-ressort non-amorti. Certaines théories montrent qu'un mouvement est programmé par le changement *a priori* de la position neutre et de la raideur du muscle. Celui se comporte alors comme un système masse-ressort et se déplace vers sa nouvelle position de repos, créant le mouvement voulu. La position neutre et la raideur du muscle sont calculées à partir de l'amplitude et de la durée du mouvement à produire de telle sorte que le muscle soit en position finale à la première annulation de sa vitesse, correspondant au premier pic de la trajectoire non-amortie d'une réponse à un système du second-ordre [Nel83]. Le tracé du profil en figure 6.5 montre une forte analogie entre la théorie du système masse-ressort et la contrainte de minimisation de la secousse.

Bilan des coûts : Le calcul de chaque trajectoire et de chaque coût est donné en détail en annexe A et les trajectoires ainsi que leurs dynamiques sont tracées en figure 6.6. Le tableau 6.2 résume les coûts d'impulsion, de force et de secousse calculés sur chacune des trajectoires.

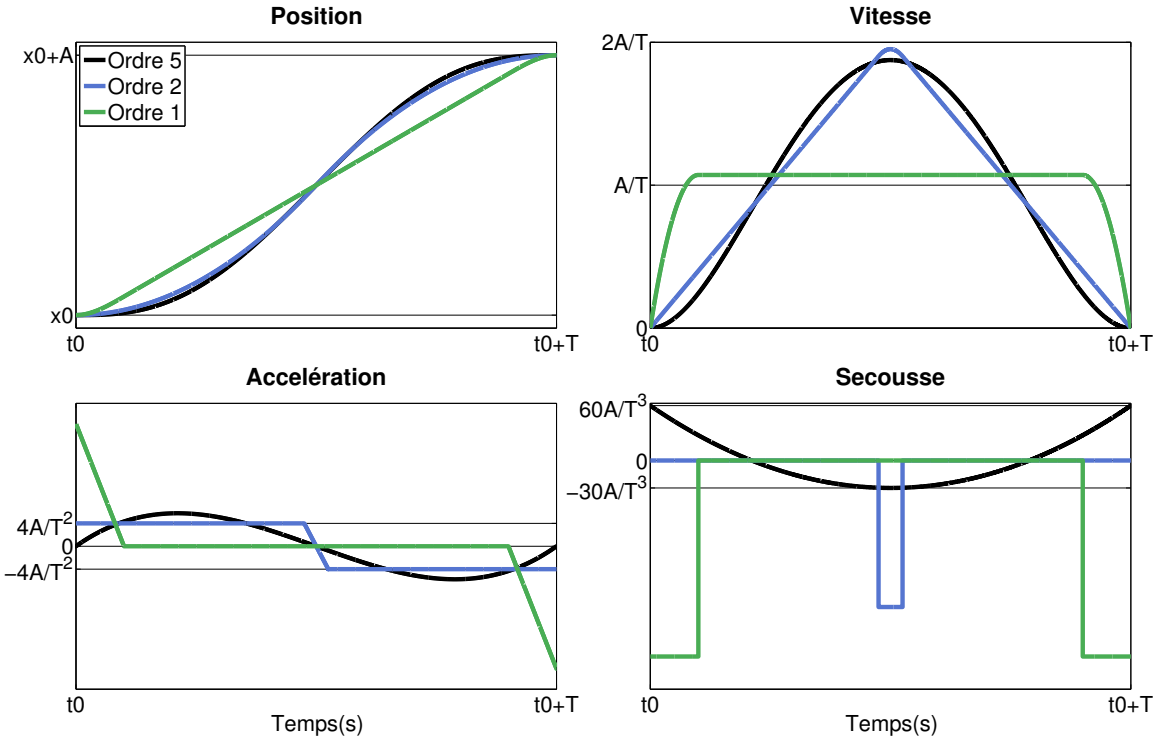

Figure 6.6 – *Trajectoires théoriques d'une dimension minimisant les coûts de secousse (ordre 5), de force (ordre 2) et d'impulsion (ordre 1) et leurs profils de vitesse, d'accélération et de secousse respectifs.*

Polynôme	Impulsion	Force	Secousse
Ordre 5	$\frac{15}{8}\frac{A}{T}$	$\frac{10}{\sqrt{3}}\frac{A}{T^2}$	$360\frac{A^2}{T^5}$
Ordre 2	$\frac{2A}{T}$	$\frac{4A}{T^2}$	∞
Ordre 1	$\frac{A}{T}$	∞	∞

Table 6.2 – *Bilan des coûts d'impulsion, de force et de secousse calculés pour les trois trajectoires polynomiales.*

6.3.2 Techniques musicales

Si l'enchaînement de notes sur un instrument de musique était une succession de simples tâches de pointage, le résultat obtenu serait d'une uniformité incompatible avec l'expression musicale. C'est pourquoi chaque musicien orne ses mélodies d'effets permettant d'ajouter une dimension subjective à la pièce jouée qu'on appelle l'interprétation. Ces effets sont en général associés à une technique de jeu propre à chaque instrument de musique. Ici, nous en recensons cinq couramment utilisés dans le chant.

Une transition entre deux notes peut s'effectuer de deux manières en chant. Lorsque le flux d'air est interrompu entre les notes par fermeture de la glotte on parle de jeu détaché ou *staccato*. Lorsque le flux est maintenu, alors la variation de fréquence entre les deux notes est continue. On parle alors de *legato*, ou de *portamento* lorsque la variation est relativement

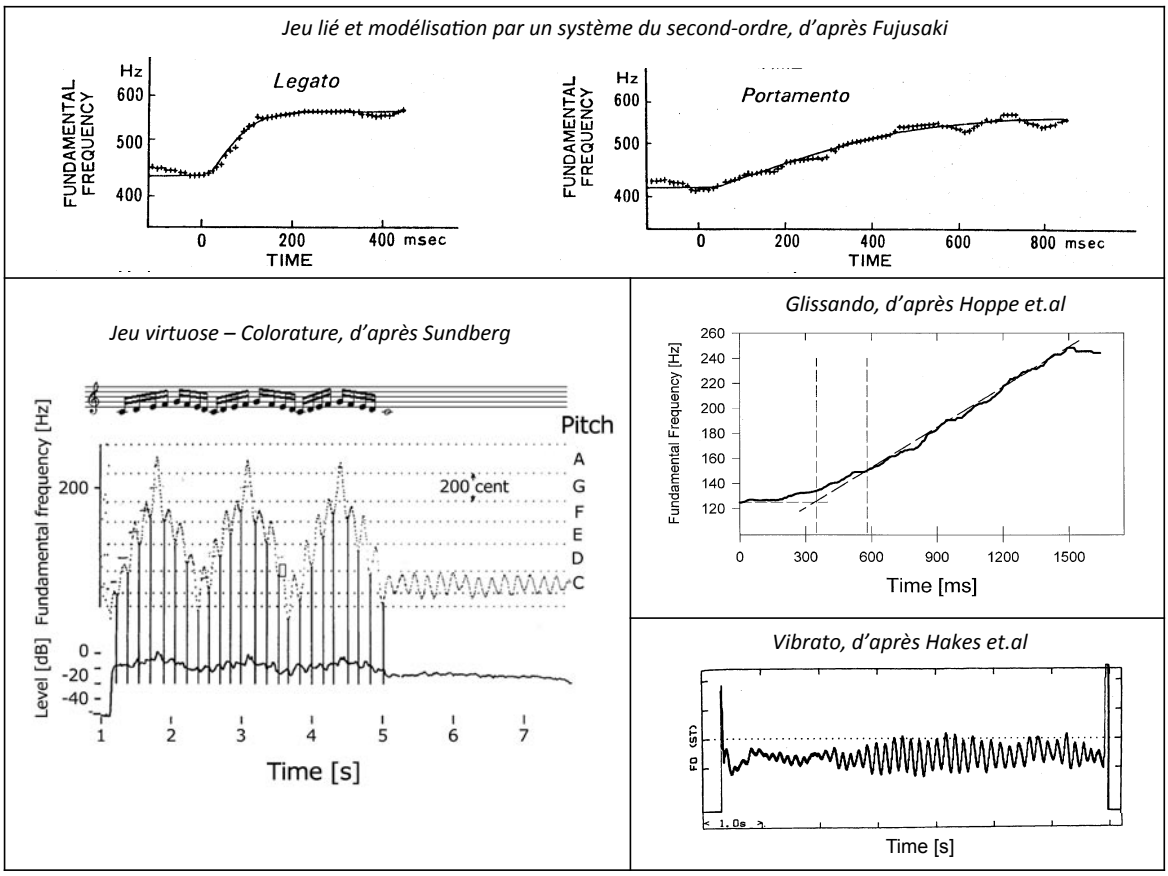

FIGURE 6.7 – *Exemples d'effets musicaux : jeu lié legato et portamento d'après [Fuj81], jeu virtuose d'après [Sun06], jeu glissando d'après [HRD+03] et jeu vibrato d'après [HSD87].*

lente. Lorsque la variation est très lente, où l'intention musicale réside plus dans la transition que dans les notes de départ et d'arrivée on parle de *glissando*. Lorsque l'enchaînement des notes est très rapide ou virtuose, le jeu choisi est souvent *legato*, par le manque de temps nécessaire pour la fermeture et réouverture de la glotte. Enfin pendant les notes tenues il est courant d'introduire une oscillation de la hauteur appelée *vibrato*. Chacun des effets est détaillé par la suite et un exemple pour chacun est présenté en figure 6.7.

Le jeu détaché ou *staccato*

Le jeu détaché consistant à introduire un silence entre chaque note, la trajectoire de hauteur obtenue présente simplement des paliers. La subtilité et l'expressivité du jeu détaché résident donc plus dans l'intensité du son que dans la hauteur. Peu d'études ont été conduites sur les propriétés des profils d'intensité dans la voix. En synthèse sonore, le profil d'intensité est généralement modélisé par une enveloppe constituée de quatre phases : l'attaque, le déclin, le soutien, et le relâchement (ADSR). Une attaque très marquée différencie un jeu staccato d'un simple jeu détaché.

Le jeu lié *legato* et le *portamento*

Le jeu lié ou *legato* implique une transition de hauteur continue dans le cas du chant. La forme de la transition a suscité l'intérêt de plusieurs équipes de recherche. Fujisaki [Fuj81] propose la modélisation du contour de hauteur par la réponse à un système du second-ordre. Celle-ci est basée sur les propriétés physiques des muscles de l'appareil phonatoire. Il modélise le cartilage thyroïde comme une masse en rotation autour du cartilage cricoïde, retenue par les muscles crico-thyroïde (reliant thyroïde et cricoïde), et les cordes vocales sont considérées comme des ressorts. Le système obtenu fournit une réponse du second-ordre de l'élongation des cordes vocales proportionnelle à la perception logarithmique de la hauteur.

Saitou *et al.* vont plus loin dans la modélisation en identifiant quatre caractéristiques d'une transition de note : le vibrato, le dépassement, la préparation et les irrégularités [SUA02], [SUA05]. Le dépassement, comme son nom l'indique consiste à atteindre une hauteur au-delà de la cible, et de revenir vers celle-ci après rectification. La préparation est l'effet inverse, consistant à s'éloigner de la cible avant la transition, à l'image d'une prise d'élan. Les notes tenues sont alors ornées d'un vibrato et d'irrégularités aléatoires reflétant les instabilités de l'appareil vocal (jitter). Saitou modélise le vibrato par une réponse à un deuxième ordre oscillant, et la préparation et le dépassement par des réponses à des systèmes du second ordre oscillants amortis. Il montre alors par ajout successif de ces effets sur de la voix synthétisée que le dépassement suivi de la préparation et enfin du vibrato sont indispensables à l'aspect naturel de la voix chantée.

Le *glissando*

Le *glissando* est une transition lente pendant laquelle la hauteur évolue linéairement de la note de départ à la note cible. A la différence d'une transition *legato* où les notes de départ et d'arrivée importent plus que la transition, ici la vitesse de transition est contrôlée précisément. Il est montré que chez un chanteur non entraîné, l'intensité du son tend à augmenter avec la fréquence dans les glissandos montants [HRD+03]. De plus, si un changement de registre a lieu pendant le glissando (de voix de poitrine à voix de tête), une baisse d'intensité est perçue au moment du changement. Pour une pratique musicale, un entraînement est indispensable afin de pouvoir chanter des glissandos les plus homogènes possibles.

Le jeu virtuose

La rapidité d'enchaînement des notes est limitée par l'appareil vocal et il est toujours impressionnant d'entendre des chanteurs repousser sans cesse leurs limites. Les spécialistes du chant virtuose sont les sopranos coloratures enchaînant des séries de notes dans l'extrême aigu de leur tessiture comme c'est le cas par exemple dans le célèbre *Air de la Reine de la Nuit* de la *Flûte Enchantée* de Mozart.

Une analyse de la fréquence fondamentale des coloratures montre que les notes cibles ne sont jamais atteintes, mais que la hauteur oscille autour de ces dernières [Sun06]. Plus précisément, la hauteur se trouve environ un demi-ton au-dessus de la note cible au début de la note, puis un demi-ton en dessous à la moitié de la note avant d'évoluer vers la cible suivante. Ce phénomène est observé car il n'est physiologiquement pas possible de faire varier la tension des cordes vocales précisément au-delà d'une certaine vitesse. C'est par des pulsations de pression sous-glottique que les chanteurs obtiennent de tels motifs.

Le *vibrato*

Le *vibrato* est une oscillation de la hauteur autour d'une note stable, très fréquemment utilisé car il apporte un certain dynamisme aux notes tenues. Il est décrit par quatre paramètres : sa fréquence d'oscillation, son amplitude, sa régularité, et la forme d'une période [Sun94]. La fréquence d'oscillation d'un vibrato vocal dépend du sexe et de l'âge du chanteur, mais aussi de la hauteur chantée. En général on relève un vibrato oscillant de 5 à 8 Hz avec une moyenne de 6 Hz [HSD87], [Sun94]. Il a aussi été mesuré que la fréquence d'oscillation dépend du tempo [MM87]. L'amplitude du vibrato est corrélée à l'intensité du chant, pouvant aller de 0.5 demi-ton pour des nuances faibles à 2 demi-tons pour des nuances fortes [HSD87], [Sun94]. Un vibrato est de forme sinusoïdale mais présente des irrégularités [Sun94]. Une irrégularité importante du vibrato est caractéristique d'un mauvais chanteur [Sun06]. Le vibrato est en général déclenché en phase avec les transitions de hauteur [DHAT00], [MM87].

Lorsque les paramètres du vibrato respectent les ordres de grandeurs cités plus haut, sa hauteur est perçue par sa moyenne [dC94]. A l'inverse, une oscillation en dessous de 4 Hz ou d'amplitude supérieure à 2 demi-tons n'est plus perçue comme un vibrato. On entend alors la hauteur osciller [Sun94]. Enfin, il a été montré que la combinaison de vibrato et de transitions continues introduit une discrétisation perceptive de la hauteur [dC91], [Sun94].

6.3.3 Propositions de gestes pour le jeu du *Cantor Digitalis*

L'étude des différents coûts fournissant des trajectoires optimales pour le geste est très théorique. Une approche dans la proposition de gestes pour le contrôle du *Cantor Digitalis* pourrait être d'associer chaque effet musical à un coût, puis d'en déduire la trajectoire optimale et de considérer celle-ci comme le geste le mieux adapté pour tel effet.

Une approche beaucoup plus pragmatique est la découverte des gestes par la pratique intensive de l'instrument. Le *Chorus Digitalis* est un ensemble de joueurs de *Cantor Digitalis* créé au sein du laboratoire, dont les membres sont des chercheurs possédant aussi une expérience musicale. Quatre des membres du *Chorus Digitalis* participent activement au développement de l'instrument. Plus de détails sur cet ensemble sont donnés au chapitre 7. L'ensemble se produit d'une à trois fois par an depuis 2011 lors d'évènements art-sciences de la région parisienne ou de conférences internationales. Il en résulte un nombre conséquent d'heures de répétitions (environ 10h par concert) pendant lesquelles l'instrument est pratiqué de manière intensive. Par conséquent, pour chaque effet musical de chaque morceau, un travail d'exploration des gestes les plus adaptés a été effectué afin d'identifier le ou les gestes à la fois les plus confortables et proposant le contrôle le plus fin de la dimension expressive concernée. Après plus d'une cinquantaine d'heure de pratique, des gestes relativement consensuels ont émergé pour le contrôle de chaque effet musical.

Dans cette partie, nous présentons les différents gestes identifiés comme les plus adaptés pour chaque effet selon notre pratique de l'instrument. Chaque geste est ensuite confronté à la trajectoire minimisant le coût du mouvement associé à l'effet concerné, afin de comparer approche pratique et théorique.

La tablette, une surface à deux dimensions

L'intonation est une variable d'une dimension qui s'exprime en hertz ou en demi-tons. Le geste vocal pour contrôler l'intonation est lui aussi d'une dimension, c'est la tension des plis vocaux qui est directement reliée à l'intonation. Par analogie, un contrôle chironomique simple de l'intonation est donc réalisé sur une unique dimension spatiale. Dans ce cas, les

mouvements d'intonations sont directement liés aux mouvements gestuels. Une accélération gestuelle se traduit immédiatement par une accélération intonative par exemple. Contrôler avec précision la trajectoire intonative demande une grande maîtrise du geste manuel.

Afin de faciliter le contrôle chironomique, il est possible de décorréler le mouvement du stylet et le mouvement intonatif en étendant l'espace de contrôle chironomique à deux dimensions, sur la surface entière de la tablette. La hauteur est contrôlée uniquement par la position horizontale du stylet mais ce dernier peut se déplacer sur toute la surface de la tablette. Le contrôle par deux dimensions d'une grandeur à une dimension permet une plus grande variété de mouvement. Par exemple un mouvement horizontal à vitesse constante donnera une intonation évoluant à la même vitesse. Un mouvement vertical à vitesse constante conduira quant à lui à une intonation stable. Il est donc possible de contrôler finement la vitesse d'intonation non plus par la seule vitesse du geste manuel, mais aussi par sa direction. Nous verrons par la suite que l'extension du geste à deux dimensions a été grandement exploitée dans le jeu d'effets mélodiques.

Le *legato* et le *portamento*

Une transition de hauteur *legato* ou *portamento* se doit d'être continue et lisse. Théoriquement, la trajectoire rectiligne minimisant le coût de secousse, soit un polynôme d'ordre 5 (équation 6.16), semble être le meilleur candidat pour jouer un *legato*.

En pratique, une trajectoire rectiligne semble être adoptée aux premiers contacts avec l'instrument puisque seule la direction horizontale du stylet contrôle l'intonation. Néanmoins, ces trajectoires entraînent plusieurs inconforts. Tout d'abord, il n'est pas naturel de dessiner des segments sur une même droite. Les articulations de notre main, poignet et bras étant des rotations, cela nous incite à tracer des mouvements courbes comme c'est le cas dans l'écriture cursive. C'est pourquoi après une certaine pratique, une partie des musiciens adoptent une trajectoire en arc pour la transition continue entre deux notes.

La figure 6.8 montre un exemple de trajectoire extraite d'une répétition du *Kyrie* de la *Messe de Nostre Dame* de Machaut. L'abscisse de la trajectoire en fonction du temps est montrée en haut, sous la partition, et l'ordonnée de la trajectoire est montrée au milieu. Deux portions de la trajectoire sont colorées et affichée en bas selon les deux coordonnées spatiales et en fonction du temps. Toutes les grandeurs sont indiquées en centimètres, soit le déplacement réel du stylet sur la tablette. On remarque que le stylet se déplace sur un intervalle d'un centimètre selon l'axe vertical. La représentation en trois dimensions nous montre les arches réalisées par le musicien pour les transitions de notes colorées. On remarque que les transitions de notes descendantes sont faites par des arches "par le bas" (en vert), et les transitions de notes montantes par des arches "par le haut" (en bleu). Cela indique que le sens de rotation du stylet est constant, propriété que l'on observe en partie dans l'écriture.

Deux avantages ressortent de la trajectoire en arc. Celle-ci permet d'atteindre la cible verticalement et non plus horizontalement. En cas de dépassement, la coordonnée horizontale de la trajectoire en arc ne sera pas déviée et permet donc une plus grande justesse et précision. Ensuite, effectuer une trajectoire sur la dimension verticale implique d'incliner le stylet. On peut supposer que cette inclinaison entraîne une chute de pression et permet des transitions d'une intensité moindre que les notes cibles.

Toutefois, ces avantages ne ressortent que dans le ressenti des musiciens. Il serait intéressant de comparer à grande échelle des trajectoires arquées et rectilignes afin de voir si des mesures objectives de justesse, précision ou d'intensité corroborent ces hypothèses.

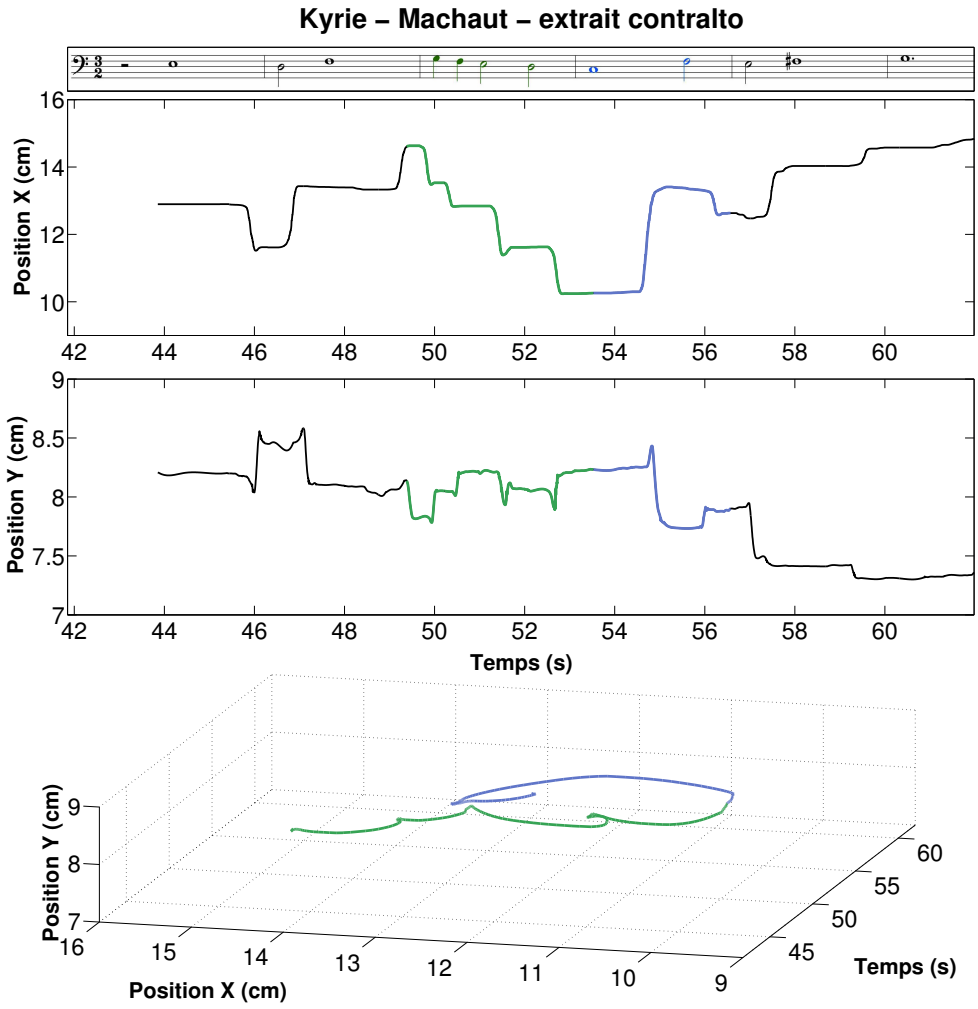

FIGURE 6.8 – *Exemple de tracés en arches pour le legato.*

Le jeu virtuose (colorature)

Le jeu virtuose consiste à jouer une succession de notes de manière rapide. Le geste le plus simple et optimal d'un point de vue théorique est de jouer chaque transition par une trajectoire rectiligne minimisant la durée entre chaque note (équation 6.14). La contrepartie de cette trajectoire est qu'elle maximise le coût de force, limité par les contraintes physiologiques du musicien. Le même problème que les sopranos coloratures est rencontré : la limite de dextérité du geste le plus simple doit être compensée par un geste alternatif. Les sopranos ne peuvent pas contrôler finement la tension de leurs cordes vocales au-delà d'une certaine vitesse et réalisent alors des impulsions de la pression sous-glottique, résultant à une hauteur oscillant autour des cibles (voir figure 6.7).

Concernant le geste, la difficulté consiste à stopper le stylet sur chaque cible, introduisant des décélérations et accélérations successives de grandes amplitudes. L'alternative serait donc de conserver une vitesse la plus constante possible. Néanmoins, une vitesse constante dans un mouvement rectiligne entraînerait un glissement de hauteur et les notes seraient imperceptibles individuellement. Une solution est donc d'exploiter à nouveau les deux dimensions de la tablette en traçant des boucles.

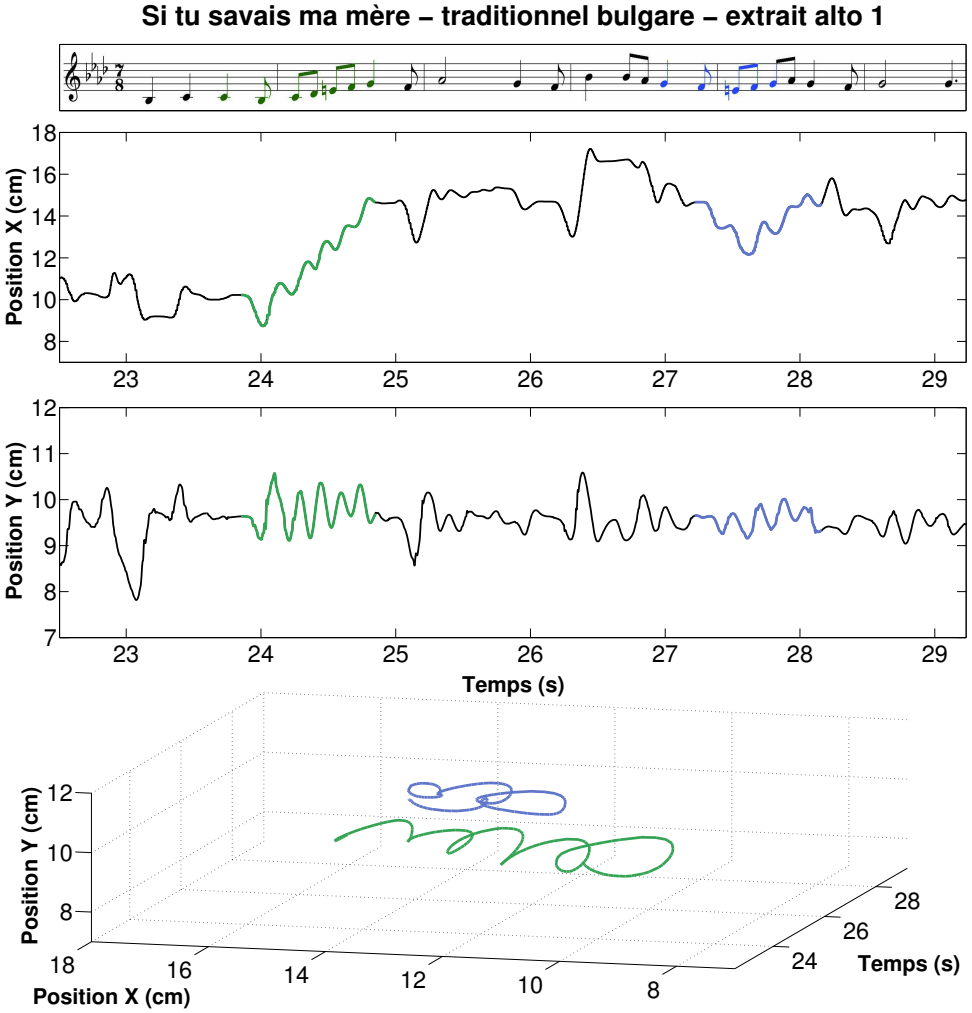

Figure 6.9 – *Exemple de tracés en boucles pour le jeu virtuose.*

Un exemple de tracé est montré en figure 6.9. Ce geste a été enregistré lors d'une répétition du morceau traditionnel bulgare *Si tu savais, ma mère*. Ce morceau est écrit avec une mesure de 7 croches avec une division 2+2+3. Le rythme le plus rapide est la croche avec un tempo de 200 noires par minutes. On observe une grande excursion de la trajectoire suivant l'axe vertical, allant jusqu'à 2 cm. Deux suites de croches rapides ont été colorées en vert et en bleu, et les trajectoires résultantes sont montrées en bas selon les deux dimensions de la tablette. On observe bien des boucles, ou dans une moindre mesure, des arcs comme pour le portamento.

Du point de vue pratique, le tracé rapide de boucles ou d'arcs est omniprésent dans l'écriture et est donc un geste relativement naturel pour l'utilisateur. On peut citer comme exemple les lettres *e*, *a*, *o* pour les boucles ou *u*, *w*, *m*, *n* pour les arcs. Du point de vue musical, la boucle introduit un retour en arrière autour de la note cible conduisant à une oscillation de la hauteur typique des contours observés chez les sopranos coloratures. On peut effectivement observer que la trajectoire de hauteur résultant des boucles montrée en haut de la figure 6.9 est très similaire à celle montrée en figure 6.7.

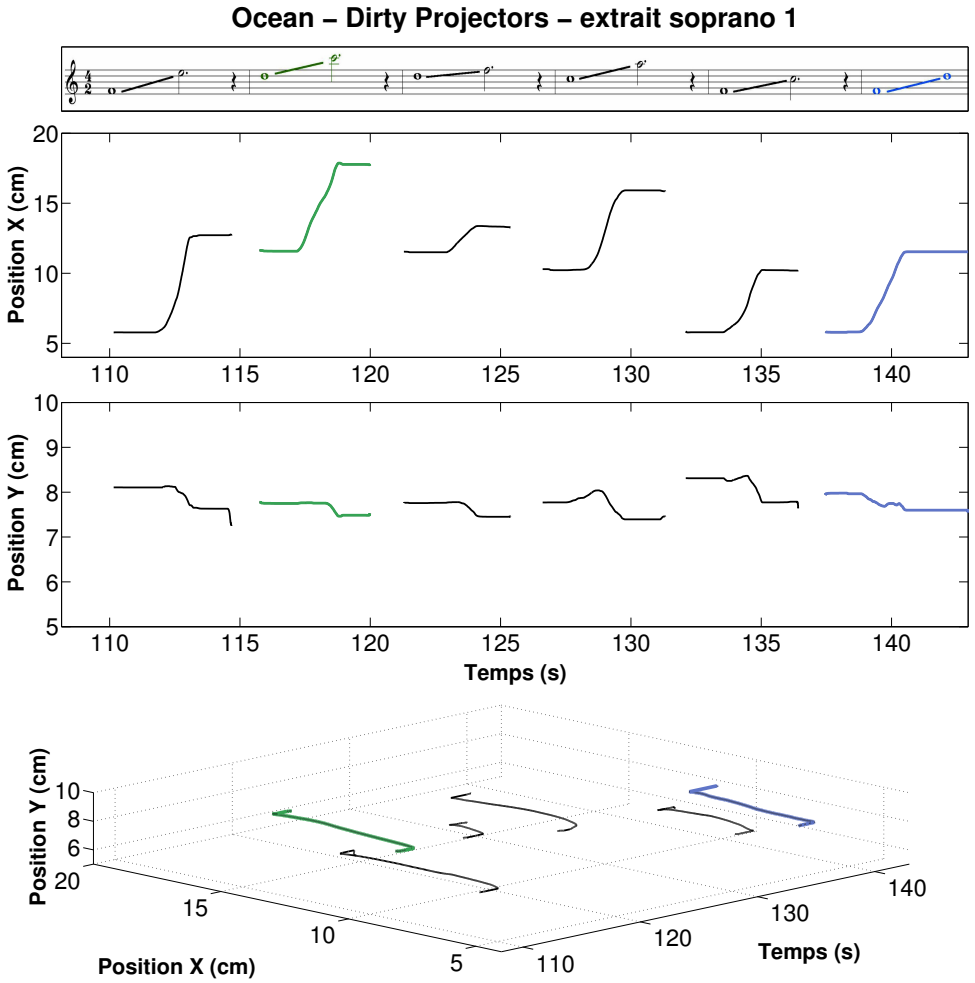

FIGURE 6.10 – *Exemple de tracés linéaires pour le glissando.*

Bien qu'introduisant des imprécisions de hauteur, cette technique gestuelle permet une perception précise des notes jouées, comme c'est le cas de la technique vocale des coloratures.

Le *glissando*

Le glissando est une évolution linéaire de la hauteur d'une note de départ vers une note cible. La trajectoire la plus adaptée à cet effet est donc une rampe, minimisant le coût d'impulsion et donnée par l'équation 6.12. La figure 6.10 donne un exemple de tracé pour le morceau *Ocean* de *Dirty Projectors*. Il s'agit d'un enchaînement de glissandos jouées entre chaque ronde et blanche pointée. Le tempo est de 50 blanches par minutes, soit légèrement plus d'une seconde par glissando. On observe effectivement une évolution linéaire de la trajectoire. Le geste est donc maîtrisé pour fournir une vitesse constante au stylet pendant la transition de notes. Pour cet effet, la trajectoire théorique optimale minimisant le coût d'impulsion, semble être adoptée par les musiciens.

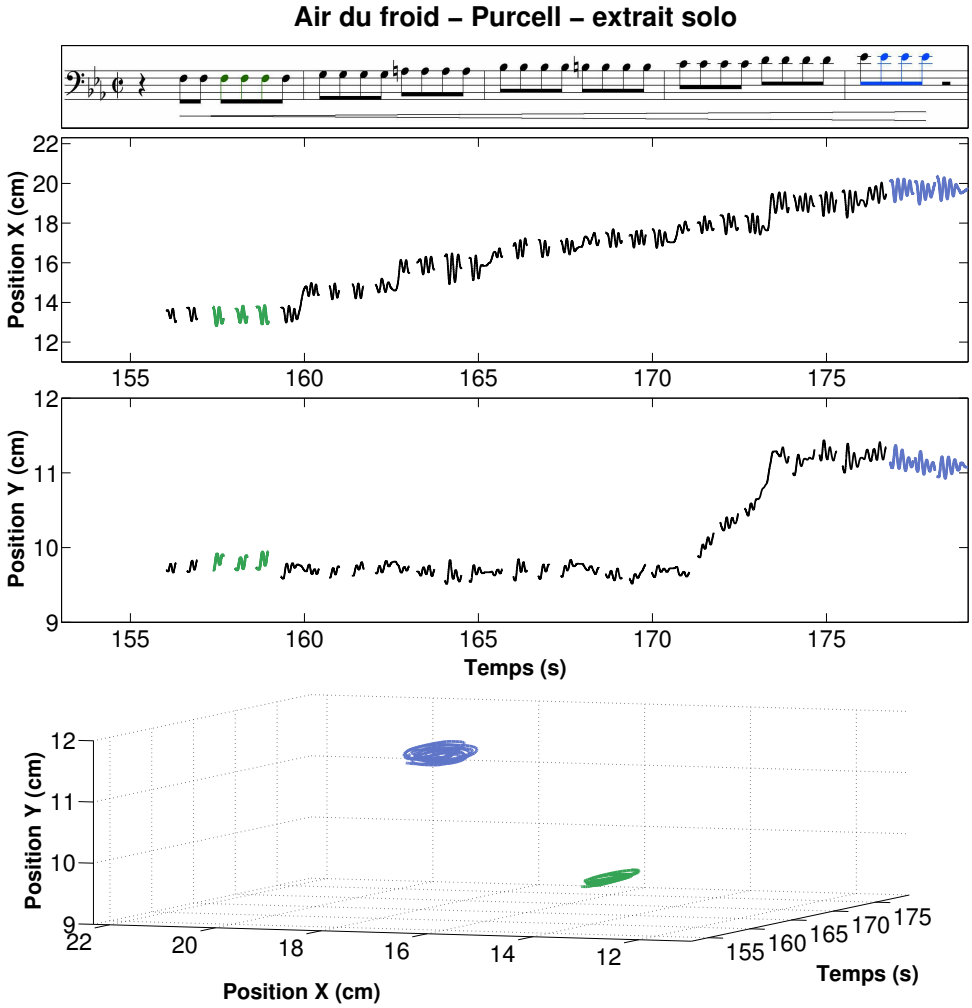

Figure 6.11 – *Exemple de tracés circulaires pour le vibrato.*

Le *vibrato*

Pour moduler la hauteur autour d'une note cible, il est nécessaire d'effectuer un geste oscillant. Le premier réflexe d'un utilisateur naïf est d'effectuer des va-et-vient horizontaux, comme le ferait un système masse-ressort rectiligne non amorti. Toutefois, cette technique entraîne des changements de directions donc des vitesses nulles aux extrémités du vibrato qui se traduisent par un coût de force plus élevé. De plus, le changement permanent de vitesse est difficile à reproduire de manière similaire d'une période à l'autre, entraînant des irrégularités.

Afin de conserver une vitesse constante pendant le vibrato, il est préférable encore une fois d'exploiter la deuxième dimension de la tablette en réalisant des cercles autour de la cible. La trajectoire résultante est sinusoïdale, et plus régulière qu'avec un mouvement sur une dimension. La fréquence du vibrato se traduit donc en nombre de cercles par secondes, et l'amplitude par le rayon du cercle.

La figure 6.11 montre un exemple de tracé pour un vibrato. Il s'agit d'un extrait de la partie soliste de *L'Air du Froid* de l'opéra *Roi Arthur* de Purcell, présentant des notes répétées jouées très vibrées. En effet, l'excursion du vibrato est de 1 à 2 cm sur la tablette, soit de 1.7 à

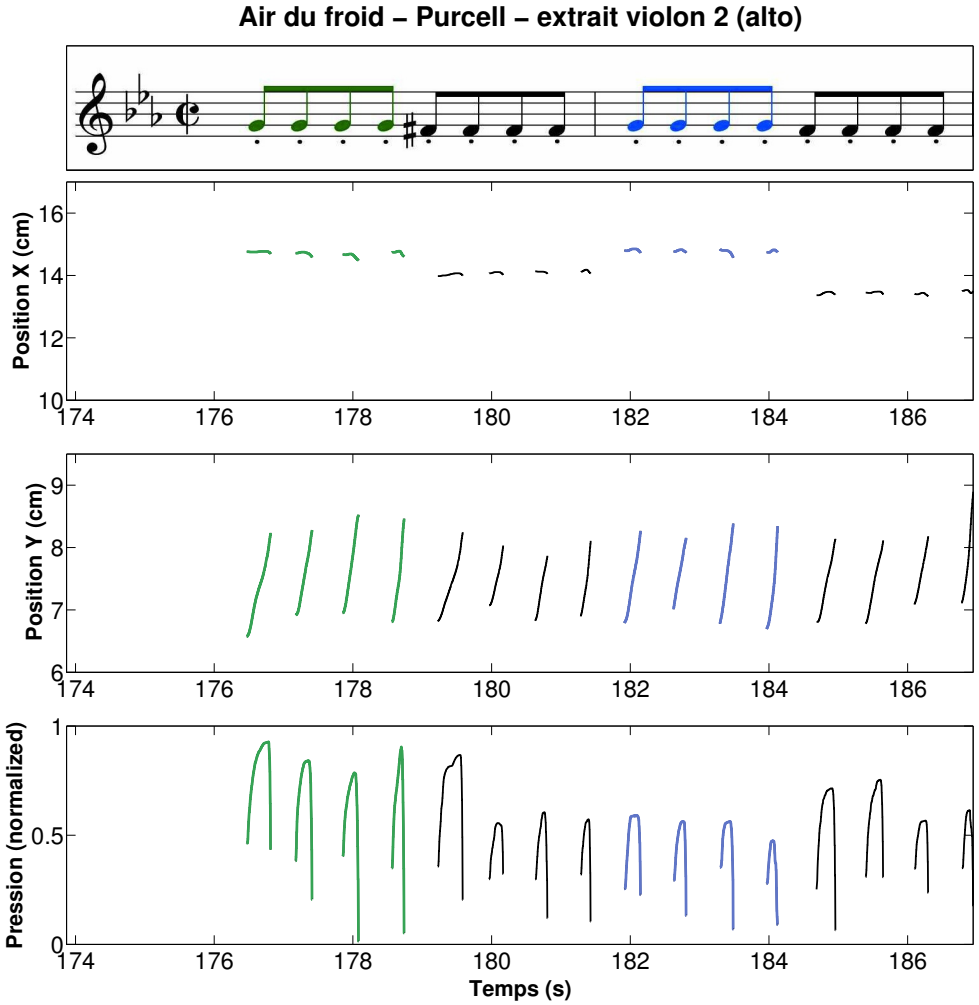

FIGURE 6.12 – *Exemple de tracés verticaux pour le staccato.*

3.3 demi-tons d'excursion totale. Deux morceaux de trajectoire ont été colorés et représentés selon les deux dimensions de la tablette en bas de la figure. La trajectoire bleue est un vibrato ample, joué au paroxysme de la pièce. On observe une excursion verticale importante (1 cm) qui conduit à un tracé elliptique. La trajectoire verte est un vibrato très serré, proche d'un tremblement et d'amplitude importante. La durée d'une période de vibrato étant plus courte, l'ellipse tracée par le musicien est aplatie en réduisant l'excursion selon l'axe vertical, afin de diminuer la distance à parcourir sur la tablette.

Ce geste circulaire a fait consensus parmi les musiciens pour son confort (car il s'agit simplement d'écrire un *o* sur la tablette) et sa régularité. En pratique, plus la fréquence du vibrato ou son amplitude augmente, plus le tracé est aplati vers une ellipse, ou à l'extrême vers un mouvement rectiligne.

Le détaché et le *staccato*

Le détaché consiste à jouer des notes isolées les unes à la suite des autres. Un détaché esthétique réside dans une grande maîtrise de l'intensité du son, qui est contrôlé par la pression du stylet sur notre instrument. Il suffit donc de placer le stylet au-dessus de la note ciblée,

puis de l'appuyer sur la tablette et le relâcher pour obtenir un jeu détaché. Bien qu'intuitive, cette technique ne permet pas de grandes libertés de contrôle sur l'intensité car l'enfoncement de la mine du stylet n'est que de quelques millimètres.

La figure 6.12 montre un exemple de jeu staccato dans l'accompagnement de *L'Air du Froid*, originalement joué au violon et transcrit pour une voix d'alto pour le *Cantor Digitalis*. L'abscisse de la trajectoire nous montre un geste très court d'excursion minimale. En revanche, la figure du milieu nous indique que chaque note est en fait jouée par un tracé rectiligne vertical sur la tablette. La figure du bas montre le profil de pression normalisé associé à l'enfoncement de la pointe du stylet. On remarque que les 5 premières notes sont jouées fortes, et les suivantes plus douces.

Le geste vertical entraîne une inclinaison du stylet et fournit un plus grand contrôle sur l'enfoncement de la mine. Le profil résultant observé est une augmentation graduelle de la pression, et donc de l'intensité, jusqu'à un maximum, puis une chute brutale. Ce motif d'intensité est aussi pratiqué par les violonistes. Musicalement, l'augmentation graduelle de l'intensité se fait avant le temps, introduisant une légère anticipation de la note à jouer. Jouer un tel motif en ne contrôlant que l'enfoncement du stylet demande une grande précision du geste. En revanche, ce geste vertical crée naturellement le profil d'intensité recherché en jouant sur l'inclinaison du stylet.

6.3.4 Discussion et conclusion

Cette étude nous a permis d'aborder le contrôle d'effets musicaux selon deux points de vue. Tout d'abord, chaque mouvement peut être caractérisé en termes de coûts, et pour chacun de ces coûts (minimisation de la durée, de la vitesse, de l'énergie, de la secousse), il existe une trajectoire permettant d'optimiser ce dernier.

Parallèlement, suite à une pratique intensive de l'instrument, des gestes ont émergé pour le contrôle d'effets musicaux. On retiendra une trajectoire arquée pour des transitions continues telles que le legato ou le portamento, qui se transforment en boucles lors d'un jeu plus rapide. Le vibrato s'effectue plus aisément par des mouvements circulaires autour des cibles et le glissando par un mouvement rectiligne constant entre deux notes. Enfin, un mouvement rectiligne vertical permet d'obtenir un profil d'intensité proche du staccato.

La comparaison de ces gestes aux trajectoires théoriques montre que les gestes musicaux appliqués sur le *Cantor Digitalis* ne sont pas optimaux. En effet, alors que la plupart des trajectoires optimales suggèrent un mouvement en une dimension, l'expérience montre qu'il est plus aisé de jouer sur un espace à deux dimensions. Seul le glissando fait exception, en présentant une trajectoire rectiligne qui minimise le coût d'impulsion. Une hypothèse expliquant cette non-compatibilité pourrait être l'opposition optimalité / expressivité. En effet, un geste expressif, un ornement, n'a pas pour but d'être optimal. Au contraire, un geste optimal serait un geste sans ornement. Le problème de compatibilité entre lois du mouvement et geste expressif se retrouve dans la modélisation automatique de transitions pour la synthèse vocale en temps différé. L'absence de trajectoires théoriques décrivant les effets vocaux incite à modéliser ces derniers par des courbes paramétriques. Ardaillon *et al.* proposent une modélisation des transitions intonatives par des B-splines par exemple [ADR15].

6.4 Discussion générale et conclusion

Ce chapitre a abordé deux aspects du geste de contrôle de l'intonation sur le *Cantor Digitalis* : sa temporalité, et sa forme. Pour chacune des deux études, aucune expérience n'a été conduite. Chaque réflexion s'est construite par la confrontation d'éléments théoriques caractérisant le mouvement (lois temporelles, lois de coûts) avec des observations de tracés sur la tablette mesurés en conditions expérimentales (données des chapitres 3 et 4) ou en jeu libre (répétition du *Chorus Digitalis*).

L'étude de la temporalité du geste a montré que le jeu musical du *Cantor Digitalis* s'affranchissait de la loi de Fitts souvent observée pour des tâches de pointages. L'étude sur la forme des gestes a montré que les mouvements effectués en pratique pour le contrôle d'effets musicaux n'étaient pas optimaux en termes de coûts. Il semble donc que lorsque le geste devient musical, ou plus généralement porteur d'expression, celui-ci ne respecte plus les lois de mouvement. En effet, ces dernières décrivent des mouvements efficaces, spontanés, dont seules les amplitudes, et les durées du geste sont prises en compte dans la représentation des lois. Dans le cas du geste expressif, la forme de la transition est aussi importante que la cible ou la note à atteindre. Ainsi, le musicien exerce un contrôle permanent sur son geste pendant la transition et peut lui faire prendre des directions non prédites par les lois du mouvement.

Finalement, la proposition des gestes de contrôle d'effets musicaux pour le *Cantor Digitalis* tire sa richesse de la pratique intensive de l'instrument plus que des considérations théoriques entourant les gestes. Bien que non supportés par des grandeurs objectives, ces gestes sont optimaux au sens des musiciens les ayant pratiqués, pour leur confort de jeu et la finesse du contrôle qu'ils apportent pour chacun des effets. Cette liste de geste s'agit donc plus d'une proposition de méthode pour le jeu de l'instrument qui serait fournie aux musiciens débutant l'instrument que d'une étude objective sur le geste.

Deux pistes sont possibles dans la poursuite de ces travaux. D'abord, il serait intéressant d'approfondir et de compléter cette liste de gestes adaptés au contrôle de l'instrument. Par exemple, une méthode de déplacement du poignet sur la tablette selon des positions permettrait aux musiciens de mieux appréhender le contrôle de la hauteur de l'instrument, comme il a été fait sur le Thérémine. Ensuite, la plupart des gestes proposés sont très liés aux tracés de lettres cursives dans l'écriture. Il serait donc possible d'aller plus loin dans l'étude théorique des gestes en considérant des lois décrivant l'écriture comme la loi de puissance 2/3. De plus, l'écriture ou la signature ont des caractéristiques propres à chaque individu. Le tracé de l'écriture se manifestant dans le contrôle de l'instrument, on peut s'interroger sur l'apport d'une identité intonative par chaque musicien, corrélée à sa graphie.

Chapitre 7

Evaluation de l'instrument par sa pratique au sein du Chorus Digitalis

Sommaire

7.1 Le Chorus Digitalis

Le *Chorus Digitalis* est un chœur de *Cantor Digitalis* initié en 2011 à l'occasion du workshop *Performative Speech and Singing Synthesis* (P3S) à Vancouver [FLBd11]. Parmi les divers ensembles d'instruments de musique numériques tels que PLOrk (*Princeton Laptop Orchestra*) [TCSW06], le méta-orchestre de Puce Muse [1], ou ChoirMob [dPW+12], le *Chorus Digitalis* est à notre connaissance le premier chœur de voix de synthèse au monde [LBFd11].

L'intérêt de la création d'un ensemble de *Cantor Digitalis* est double. D'abord du point de vue du développement, le jeu intensif de l'instrument constitue le meilleur des tests. Des milliers de combinaisons de paramètres d'entrée sont testées à la fois par les développeurs de l'instrument mais aussi par des musiciens plus naïfs, amenant à la détection et correction de nombreux bogues, indispensables à la diffusion du logiciel.

Ensuite, cet ensemble définit un cadre de pratique de l'instrument. Ces répétitions ont permis d'explorer l'instrument à la fois au niveau des gestes de contrôle comme discuté au chapitre 6 mais aussi par rapport au répertoire musical à aborder. Le principe 5 énoncé par Cook à propos de la création d'instruments de musique numériques [Coo01] : "*Make a piece, not an instrument or controller*" n'a pas été suivi pour notre instrument. Celui-ci a été créé avant d'envisager les pièces à jouer avec. C'est pourquoi une étape importante d'exploration du répertoire vocal (ou non vocal) a été effectuée au fil de nos concerts. Enfin, la réalisation de concerts permet de diffuser notre instrument lors de conférences scientifiques ou au grand public dans des cadres moins formels, tels que des évènements art-sciences organisés dans la région. L'ensemble se produit d'une à trois fois par an depuis 2011. Les répétitions ne sont pas régulières, ayant lieu seulement dans les semaines précédent un concert.

Le *Chorus Digitalis* est composé de 4 à 8 musiciens, membres du LIMSI-CNRS et possédant des expériences musicales diverses autant du point de vue de l'instrument de musique joué (claviers, percussions, cordes, musique assistée par ordinateur (MAO), ...) que des styles de musiques préférés (classique, jazz, musique du monde, musiques pop, musique électronique, ...). Quatre des membres de l'ensemble participent activement au développement de l'instrument et forment le noyau du groupe. Lors de concerts proches du laboratoire, des chercheurs extérieurs au projet se joignent à l'ensemble pour atteindre jusqu'à 8 musiciens. Chaque musicien du *Chorus Digitalis* est équipé d'une tablette graphique Wacom Intuos 4M ou 5M, connectée à un ordinateur sur lequel est installé le moteur de synthèse, lui même relié à une enceinte active Genelec. Ainsi, chaque musicien est capable de jouer et de produire du son indépendamment des autres, comme s'il s'agissait d'instruments acoustiques. Sur scène, les musiciens sont disposés en arc de cercle, et chaque enceinte est placée derrière son musicien, afin de placer les sources sonores proche de leurs interprètes.

Afin d'aider le public à comprendre le maniement de ce nouvel instrument, un système de visualisation des données de la tablette a été mis au point, projetant celles-ci sur un écran lors de nos représentations [2]. L'implémentation ainsi que les retours obtenus sur cette visualisation sont présentés en section 7.2. Au fil des années, la mise en scène et le répertoire du *Chorus Digitalis* a évolué. La section 7.3 fait état des transformations scéniques et musicales de l'ensemble. La chronologie de ces transformations est donnée en dernière section, suivant les différents concerts effectués depuis 2013, année de mon entrée dans l'ensemble.

1. http://www.pucemuse.com/portfolio/meta-orchestre/ (vérifié le 22 octobre 2015)
2. Ces travaux sont parus dans [Pd14].

7.2 Outil de visualisation du jeu de l'instrument

7.2.1 Performances multimodales

Magique, secrète, expressive : le choix d'une performance

Deux aspects sont essentiels lors d'une performance artistique faisant l'usage d'un ordinateur : la manipulation de l'interface associée aux gestes effectués par l'interprète, et les effets découlant de cette manipulation. En fonction de la quantité d'information montrée aux spectateurs, une taxonomie des performances a été établie par Reeves [RBOF05] et présentée en figure 7.1. Alors qu'une performance magique masquant toute manipulation est parfois grandiose, une performance expressive nécessite quant à elle la révélation, voir l'amplification des manipulations. En effet, la compréhension décroissante de la relation entre un contrôleur et ses effets auditifs tend à réduire l'intérêt d'une performance musicale.

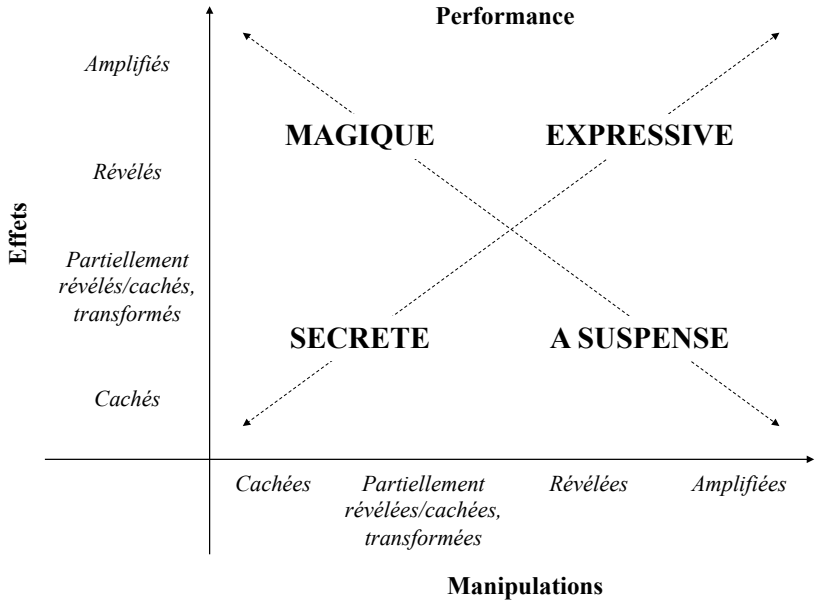

Figure 7.1 – *Classification des types de performances en fonction de la quantité de manipulations et d'effets montrés aux spectateurs, d'après [RBOF05].*

Ce problème récurent dans la création de nouveaux instruments de musique numériques s'est manifesté lors des performances du *Chorus Digitalis*. Malgré une amélioration de la visibilité des gestes sur la tablette par une évolution de la disposition scénique (voir section 7.3), la tablette reste petite et difficile à observer depuis une certaine distance. Afin d'augmenter la perception des différentes manipulations des musiciens, nous proposons l'introduction de retours visuels au sein de la chaîne d'évolution des paramètres de l'instrument. Affichés sur un écran d'ordinateur placé devant les musiciens et projetés sur un écran large au fond de la scène, ceux-ci bénéficient à la fois à l'interprète et aux spectateurs dans l'identification du fonctionnement de l'instrument.

Visualisation des gestes et du son

Plusieurs types de visualisation ont été développés à différents points dans la chaîne d'évolution des paramètres. Tout d'abord, la technique de visualisation la plus répandue est d'afficher le contrôleur directement en filmant les gestes du musicien. Cela amplifie la perception de la manipulation mais n'ajoute pas d'information. Pour cela, certains choisissent de représenter les paramètres associés au contrôleur comme par exemple le dessin résultant de la trajectoire d'un stylet sur une tablette graphique contrôlant un séquenceur [ZS06].

En allant plus loin dans la chaîne d'évolution des paramètres, la représentation de paramètres intermédiaires entre contrôle et son montre les intentions musicales de l'interprète facilitant la compréhension de la production sonore [ACK05]. Le *Voicer* [Kes04b] par exemple affiche les paramètres liés au geste pour assister le contrôle de l'instrument. Dans une optique plus esthétique, le MelodicBrush [HTL+12] affiche sur un écran les tracés d'une brosse dont les mouvements sont captés par une Kinect et associés à des sons musicaux. Dans le cas de relations complexes entre les différents paramètres, celles-ci sont parfois représentées de manière simplifiée par des connexions directes entre les composants de contrôle et sonores [BMSH13]. D'autres instruments proposent une représentation visuelle interactive, où les composants de contrôle et sonores ainsi que leurs liens sont manipulables par l'utilisateur [Jor03].

Ces représentations se concentrent sur l'interaction entre l'utilisateur et l'interface, mettant en lumière les paramètres de contrôle et les paramètres liés au geste. En revanche, peu de travaux ont été réalisés dans la représentation des paramètres sonores, qui pourtant aide à se concentrer sur certaines caractéristiques acoustiques (timbre, hauteur, amplitude) [PK13]. De plus, afin de conserver une trace des mouvements précédents, l'utilisation du visuel permet d'afficher l'évolution des paramètres en fonction du temps [BK07]. Ces exemples de visualisation soulignent l'importance de la représentation des paramètres sonores. Cependant, ils ont été développés dans un contexte de production de parole et n'ont jamais été appliqués lors de performances instrumentales.

Le but de cette étude est d'utiliser le *Cantor Digitalis* comme support d'exploration des possibilités de retour visuel, en fonction de leurs position dans la chaîne d'évolution des paramètres. Chaque visualisation est présentée sous deux aspects : son apport d'information et sa qualité esthétique.

7.2.2 Visualisation du *Cantor Digitalis*

Espaces de représentation du *Cantor Digitalis*

Un instrument de musique numérique transforme les paramètres bruts du contrôleur en paramètres de synthèse sonore par une ou plusieurs transformations, constituant la chaîne d'évolution des paramètres. Le *Cantor Digitalis* intègre une transformation des paramètres à trois couches (figure 2.6), comme le propose le modèle de [HW02]. En considérant la normalisation des paramètres de contrôles effectuée par le module de récupération des données de la tablette, un modèle à quatre couches peut être appliqué [ACKV02] (figure 1.10).

La figure 7.2 illustre cette organisation sur l'exemple du *Cantor Digitalis* et montre l'évolution des paramètres issus de la tablette vers les paramètres régissant la synthèse à travers divers espaces de représentation. Deux chaînes sont représentées : la première décrit l'évolution des paramètres associés au contrôle de la source vocale, et la deuxième se rapporte au contrôle du conduit vocal. Le contrôle de la source commence par la mesure des coordonnées du stylet sur la tablette et de sa pression. La première transformation transpose ces paramètres de bas niveau dans le domaine haut niveau de la gestuelle en les normalisant. Les

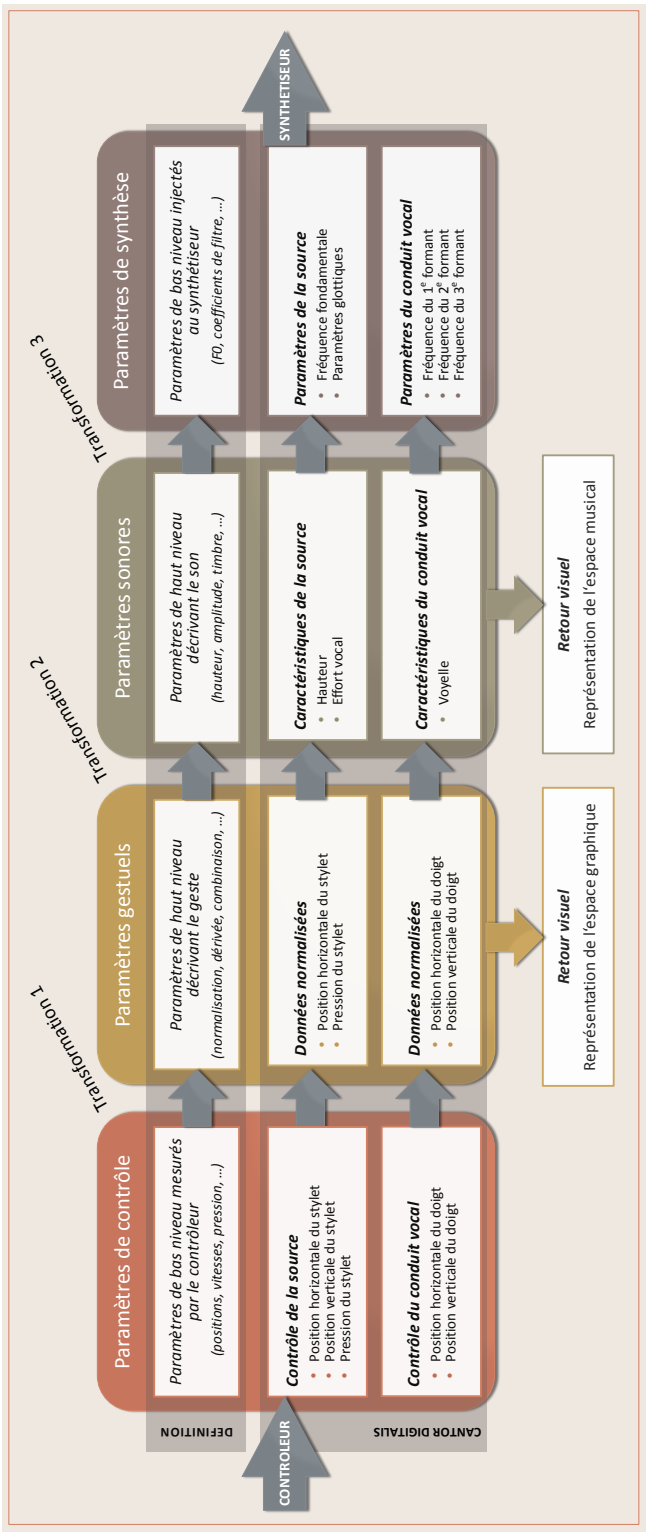

FIGURE 7.2 – *Evolution des paramètres gestuels à travers les couches successives du Cantor Digitalis pour contrôler la source et le conduit vocal du synthétiseur vocal à l'aide d'une tablette graphique.*

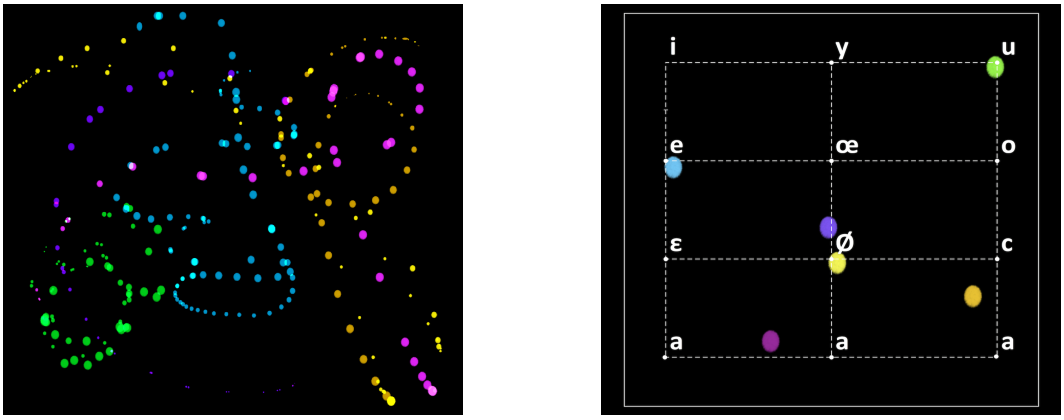

Figure 7.3 – *Captures d'écran de la représentation des paramètres liés au geste dans le domaine graphique. L'état du stylet contrôlant la source est montré à gauche, et l'état du doigt contrôlant l'articulation est présenté à droite. Chaque couleur est associée à un musicien.*

paramètres gestuels sont ensuite transformés en paramètres sonores de haut niveau : la hauteur et l'effort vocal. Finalement, la dernière transformation associe ces paramètres sonores aux paramètres bas niveau du synthétiseur, soit la fréquence fondamentale et les différents paramètres glottiques (quotient d'ouverture, coefficient d'asymétrie, ...). La deuxième chaîne débute par la mesure des coordonnées du doigt sur la tablette. Celles-ci sont normalisées par la première transformation et renseignent alors sur le geste du musicien. Elles sont ensuite transformées en paramètres sonores qui sont simplement le choix de la voyelle à jouer. Enfin, ces paramètres sont transformés en fréquences centrales, bandes passantes et amplitudes des filtres formantiques utilisés dans la synthèse.

Cette représentation met en évidence deux étapes de paramètres haut niveau : les paramètres gestuels et sonores. Avec le *Cantor Digitalis*, les gestes sur la tablette s'assimilent à une tâche d'écriture ou de dessin. La représentation visuelle des paramètres gestuels se fera donc dans ce qu'on appellera *l'espace graphique*, construit sur les deux dimensions spatiales de la tablette et une dimension temporelle. La représentation visuelle des paramètres sonores se fera quant à elle dans un *espace musical*, constitué des dimensions de hauteur, d'effort vocal, d'articulation et de temps.

Représentation dans l'espace graphique

La manière la plus simple de représenter les gestes du musicien dans l'espace graphique est d'afficher une forme sur un écran (point, cercle, tâche) correspond à l'état du stylet ou du doigt. Il suffit pour cela de faire correspondre la position normalisée de ces derniers à la position de la forme à l'écran, et d'associer la pression normalisée du stylet à la largeur de la forme. Il s'agit du même principe que des applications de dessin à main levée. Plusieurs contextes sont proposés en fonction du contrôle.

Source : Bien que seule la position horizontale du stylet soit nécessaire pour contrôler la hauteur, le musicien a la liberté d'explorer la dimension verticale à des fins d'expressivité. Cela lui donne la possibilité d'effectuer aussi bien des lignes droites que des courbes dans le jeu de la hauteur (chapitre 6). Un exemple d'une visualisation du contrôle de la source dans l'espace graphique est montré à gauche de la figure 7.3, où chaque couleur est associée à un

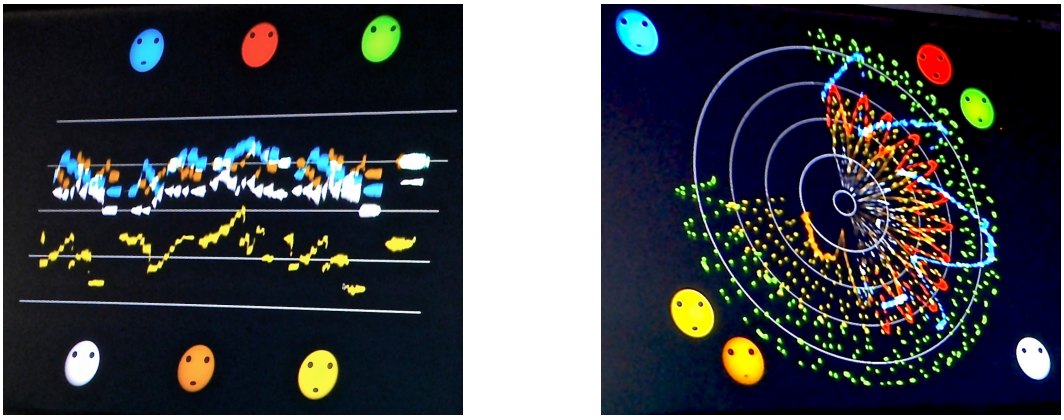

Figure 7.4 – *Images de la visualisation des paramètres musicaux prises durant une performance. Les lignes représentent la hauteur en fonction du temps de manière linéaire (gauche) ou circulaire (droite). Les avatars imitent l'articulation choisie par chaque musicien. Chaque couleur correspond à un joueur.*

musicien. Pour l'audience, cette représentation est une amplification du geste sur la tablette. Contrairement à une simple caméra qui filmerait les mains du musicien, la visualisation de l'espace graphique apporte des informations supplémentaires telles que la pression du stylet ou les superpositions des traces de chaque musicien. Il s'agit d'une métaphore calligraphique reflétant l'habilité des musiciens à dessiner des formes dans l'espace, permettant une expressivité multimodale, mélangeant dessin et musique. Une plus grande part à la réflexion sur les modes de représentation a été accordée dans cette étude que dans l'implémentation des visuels eux-mêmes. Il serait possible d'aller beaucoup plus loin dans l'esthétique du rendu en s'inspirant de travaux tels que [HTL+12] qui en associant calligraphie et musique cherchent à représenter numériquement et le plus fidèlement les coups de pinceaux effectués dans le dessin d'alphabet Chinois.

Conduit vocal : La représentation graphique du contrôle du conduit vocal utilise la disposition des voyelles en triangle vocalique dans un plan de dimensions (position de langue × ouverture de la mâchoire). L'ouverture des lèvres est une troisième dimension projetée sur ce plan. Cette représentation est fortement liée aux paramètres de contrôle de l'instrument puisque les positions verticale et horizontale du doigt correspondent implicitement à l'ouverture de la mâchoire et à la position de la langue respectivement. Le triangle vocalique est donc représenté à l'écran, sur lequel sont superposés des disques de couleurs dont la position est associée à la position du doigt de chaque musicien. Comme l'articulation varie moins que la hauteur, on ne représente ici que la position courante chaque doigt et non une trace de chaque trajectoire. L'image de droite de la figure 7.3 montre un exemple d'une telle représentation.

Représentation dans l'espace musical

La visualisation de l'espace graphique n'apporte que des informations sur les gestes sur la tablette et n'est pas directement liée à la performance musicale. Pour aider à la compréhension de la production sonore, il est intéressant d'afficher les paramètres musicaux.

Source : Les deux paramètres de contrôle temps réel de la source sur le *Cantor Digitalis* sont la hauteur et l'effort vocal. De plus, bien que la notion temporelle ait peu d'importance en

Préfixe	Description
pitch	*Hauteur de la note en demi-tons*
vocal_effort	*Effort vocal normalisé [0,1]*
xstyl	*Position horizontale du stylet normalisée [0,1]*
ystyl	*Position verticale du stylet normalisée [0,1]*
xfing	*Position horizontale du doigt normalisée [0,1]*
yfing	*Position verticale du doigt normalisée [0,1]*
F1	*Position du premier formant en Hz*
F2	*Position du deuxième formant en Hz*
F3	*Position du troisième formant en Hz*
register	*Registre (entier)* *1 – Noisy soprano* *2 – Soprano* *3 – Noisy Alto* *4 – Alto* *5 – Tenor* *6 – Bass* *31 – Bulgarian style* *21 – Child 1* *22 – Child 2* *23 – Baby* *11 – Extreme Voice 1* *12 – Extreme voice 2* *13 – Extreme voice 3*

Figure 7.5 – *Liste des données transmises au programme de retour visuel par le Cantor Digitalis.*

dessin, celle-ci constitue la base de la musique. C'est pourquoi nous choisissons de représenter les paramètres sonores en fonction du temps. L'instant présent est placé à une position fixe de telle sorte que les trajectoires de chaque musicien défilent sur l'écran. L'effort vocal est associé à l'épaisseur de chaque trace. Le temps peut aussi bien être représenté linéairement et défiler de droite à gauche, comme montré à gauche de la figure 7.4, ou défiler circulairement comme sur une horloge [BK07] (droite de la figure 7.4). Afin d'aider l'audience à associer chaque son produit à chaque musicien, les couleurs des trajectoires de hauteur sont assorties avec les tenues des ces derniers.

Conduit vocal : Tandis que le triangle vocalique est lié à des notions de traitement de signal, la production sonore découle directement des mouvements physiques effectués dans la prononciation des voyelles. On visualise ici une représentation simplifiée de l'ouverture de la mâchoire, de la position de la langue et de l'ouverture des lèvres par des avatars construits très simplement à l'aide de sphères de couleurs différentes (une pour chaque musicien). Ceux-ci sont visibles sur la figure 7.4.

Implémentation

Pour des raisons pratiques, la représentation visuelle a été développée sous Max avec la librairie Jitter. Cela permet une bonne compatibilité avec les paramètres du *Cantor Digitalis*. Le programme est indépendant du logiciel de synthèse et fonctionne sur un ordinateur annexe afin de ne pas surcharger en calcul les ordinateurs utilisés par l'instrument. Chaque *Cantor Digitalis* envoie alors ses données en temps réel par protocole UDP sur un réseau Wi-Fi privé. Les données du *Cantor Digitalis* formatées et envoyées au retour visuel sont données en figure 7.5. Le registre de voix est nécessaire afin de transformer les positions des formants en gestes articulatoires. Tous les réglages pour la visualisation (type de représentation, couleur des traces, épaisseur moyenne des traces) sont accessibles uniquement dans le programme de visualisation et non dans le *Cantor Digitalis* pour éviter des erreurs de manipulation pendant les performances.

	Gauche	Centre	Droit
Fond	2	3	1
Milieu	5	1	0
Devant	3	0	3

Table 7.1 – *Position dans la salle du public ayant répondu au questionnaire.*

7.2.3 Evaluation qualitative

Une évaluation informelle de retour visuel a été effectuée en distribuant des questionnaires au public lors d'un concert du *Cantor Digitalis* dans le cadre du festival CuriositAS [3]. Les premières 15 min ont été jouées sans retour visuel. Les représentations des espaces musicaux linéaire et circulaire ont été ensuite affichées alternativement pendant les 30 min restantes. La relation entre la couleur des traces et la couleur des tenues des musiciens n'a pas été annoncée avant le spectacle. Chaque question posée devait être évaluée sur une échelle de 1 (non) à 5 (tout à fait). Les questions étaient :

- Avez-vous trouvé les voix naturelles ?
- Pouviez-vous voir les musiciens sur scène ?
- Les gestes des musiciens étaient-ils visibles ?
- Avez-vous compris comment les voix étaient contrôlées ?
- Pouviez-vous distinguer qui contrôlait chaque voix avant la projection ?
- Pouviez-vous distinguer qui contrôlait chaque voix après la projection ?

18 personnes ont pris le temps de répondre au questionnaire. La moyenne d'âge était de 38 ans, et seules 4 personnes avaient déjà vu une représentation du *Chorus Digitalis*, bien que la projection visuelle était nouvelle à ce concert. 8 sujets étaient musiciens, 2 d'entre eux étaient chanteurs, et 5 sujets étaient familiers avec la voix de synthèse.

La performance s'est déroulée dans une salle où le public était assis en rang sur des chaises au même niveau. Les musiciens étaient assis par terre devant l'audience, les tablettes placées sur des supports de 20 cm de haut. Par conséquent la visibilité de la scène n'était pas toujours bonne pour les personnes assises au fond de la salle. Toutefois, l'écran montrant l'affichage visuel était placé en hauteur, au dessus de la scène, et était visible par l'ensemble du public. Les spectateurs ayant répondu au questionnaire étaient équitablement répartis dans la salle, comme l'indique le tableau 7.1. La figure 7.6 indique que deux tiers de l'audience pouvait voir les musiciens plus que convenablement, ainsi que les gestes de ces derniers. Bien que la visibilité était mauvaise pour le tiers restant, 78% du public a déclaré comprendre plus que convenablement comment les voix étaient contrôlées. C'est dans de telles situations que l'apport d'un retour visuel visible par tout le public gagne un intérêt.

On constate d'après les réponses aux questions que le retour visuel a permis principalement l'identification des musiciens. Deux tiers de l'audience ont déclaré n'être que faiblement capable de distinguer les différents musiciens avant l'affichage, et 44% ne pouvaient pas du tout. Après l'application du retour visuel, tous sauf deux ont pu différencier convenablement les joueurs. Finalement, bien que seulement qualitatif, ce retour d'expérience nous a permis de valider l'utilité d'un retour visuel dans la compréhension du maniement de l'instrument par le public.

3. http://www.youtube.com/watch?v=9XpnDiJJyMk (vérifié le 22 octobre 2015)

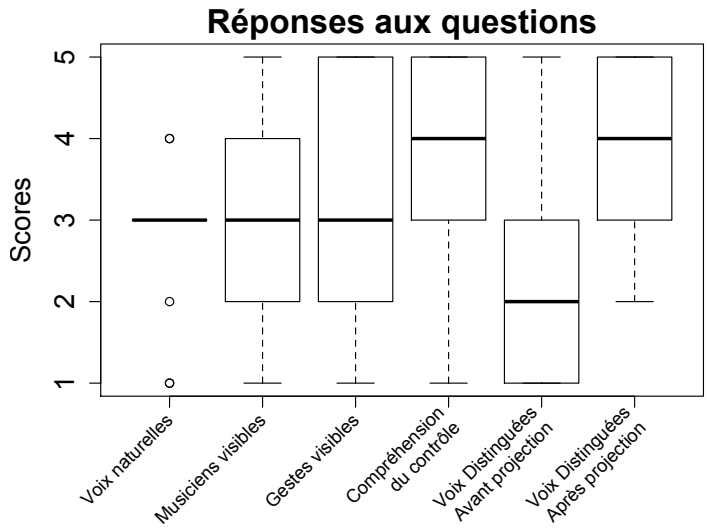

Figure 7.6 – *Réponses de l'audience à notre questionnaire.*

7.2.4 Callisonography - un retour audio à la performance visuelle

Le but principal de la visualisation est d'apporter des informations supplémentaires à un contexte musical. Cependant, il est possible d'imaginer un processus inverse où la performance consiste à dessiner sur l'affichage visuel et où le *Cantor Digitalis* fournit un retour auditif. C'est d'autant plus intéressant que la performance doit rester musicalement cohérente. Une façon facile d'improviser depuis le visuel est d'utiliser des motifs périodiques. L'affichage circulaire permet alors la création de rosaces à l'écran, et le son produit est joué en boucle, produisant une sorte de boîte à rythmes. Une improvisation visuelle a été tentée durant cette même performance (figure 7.7). Le principal retour du public suggérait de faire durer l'improvisation plus longtemps, et d'exploiter de manière plus significative les possibilités de dessin sur toute la performance. Par ailleurs, 4 personnes ont considéré la représentation visuelle comme la partie la plus intéressante du concert, alors que l'objet du spectacle était avant tout la présentation de la synthèse vocale performative. Cela montre l'intérêt du public dans ce nouveau mode de création musicale.

7.2.5 Conclusions

Pour conclure, la visualisation de la chaîne d'évolution des paramètres a pour but d'améliorer la compréhension de l'audience au fonctionnement d'un instrument de musique numérique lors d'une performance. La décomposition de la chaîne d'évolution met en évidence différents espaces de représentation : l'espace gestuel (graphique pour le *Cantor Digitalis*) et l'espace musical. Nous avons utilisé uniquement la représentation de l'espace musical en concerts, car moins utilisée que l'espace gestuel. De bons retours nous sont revenus par le public, confortant l'utilité de l'affichage visuel lors d'une performance.

Pour compléter la visualisation des paramètres à chaque étape de la chaîne, la représentation des paramètres du synthétiseur comme les fréquences des formants pourraient être explorée dans un travail futur. De plus, une plateforme plus propice au développement graphique ainsi que l'aide d'un artiste graphiste permettrait d'améliorer grandement l'esthétique du visuel proposé ici.

Figure 7.7 – *Images de la représentation visuelle des paramètres musicaux prise comme support pour une improvisation. Chaque couleur correspond à un musicien.*

7.3 Evolution du Chorus Digitalis

7.3.1 De la table aux pupitres, l'évolution de la mise en scène

Contrairement au *Handsketch* dont la tablette se tient verticalement sur les genoux [dD09b], nous avons fait le choix de poser la tablette à plat sur un support pour le *Cantor Digitalis*, comme c'est le cas pour une tâche d'écriture ou de dessin. Trois différents supports ont été testés au fil des concerts.

La première solution fut de disposer tablettes et ordinateurs sur des tables, comme montré en figure 7.8. Cette disposition résultant de la position naturelle de l'écriture en Occident, celle-ci permet un jeu confortable. Néanmoins, du point de vue du spectateur, il est difficile d'apprécier l'activité de chaque interprète. En effet, comme le montrent les images du concert du Printemps de la Culture, chaque musicien semble absorbé dans une tâche qui lui est propre. Les tables ainsi que les écrans d'ordinateurs ouverts devant chaque tablette forment une barrière entre musiciens et audience. Ce manque de communication avec le public a été soulevé lors de ce concert, et la recherche d'une disposition plus ouverte a été entreprise.

Une disposition au sol a été proposée pour la première fois lors d'une conférence internationale en Corée (*NIME 13*). Chaque musicien est alors assis par terre sur un coussin, et dispose devant lui d'un support d'une vingtaine de centimètres de haut. Tablettes et partitions sont placées à plat sur le support, et l'ordinateur posé par terre sur le côté du musicien. Cette disposition a été conservée pendant une année de quatre concerts, pour des ensembles allant de quatre à sept personnes. Deux exemples de concerts sont montrés en figure 7.9. Cette disposition a permis de supprimer la barrière entre musiciens et public imposée par les tables et les écrans. En effet, la position basse des supports permet à une audience surélevée (simplement assise sur des chaises ou dans un amphithéâtre) de voir les tablettes et donc le maniement de l'instrument. L'inconvénient majeur de cette mise en scène est l'inconfort.

Figure 7.8 – *Disposition des musiciens derrière des tables - Concert du Printemps de la culture à Orsay, 2012.*

Figure 7.9 – *Disposition des musiciens assis au sol - Gauche : Journées du Développement (JDEV) à Palaiseau, 2013. Droite : Festival CuriositAS à Orsay, 2013.*

D'abord, la position assise au sol est difficile à tenir pendant deux longues heures de répétition. Ensuite, ne pouvant pas placer les jambes sous le support, les tablettes sont à une distance relativement éloignées du corps. Cela demande de s'incliner en avant entraînant une posture fermée mauvaise pour le dos et nuisible à l'ouverture vers le public.

La troisième solution adoptée pour le concours Guthman d'instruments de musique s'inspire des ensembles de musique de chambre. Chaque musicien est assis sur une chaise, et tient sa tablette sur ses genoux. L'ordinateur est posé sur un support placé à côté du musicien, et les partitions sont posées sur un pupitre placé de biais, afin de ne pas gêner la vue des spectateurs sur la tablette. La figure 7.10 montre une telle disposition. Celle-ci combine à la fois l'ouverture vers le public grâce à la position basse de la tablette sur les genoux, et le confort d'être assis sur une chaise avec la tablette près du corps. Comme il est toujours nécessaire de s'incliner légèrement, il est possible de poser ses pieds sur un support afin de surélever la tablette.

Ces dispositions on été testées pour l'ensemble des musiciens jouant en chœur. Il est de tradition de jouer à chaque concert un *Raga*, qui est un type de pièce d'Inde du Nord. Chaque Raga implique un soliste qui se place au centre de l'arc de cercle. Celui-ci a fait le choix de s'asseoir en tailleur et de placer la tablette sur ces genoux comme le montre la figure 7.11. Cela permet un jeu plus intimiste avec le joueur de tablas qui l'accompagne à ses côtés.

Figure 7.10 – *Disposition de type quatuor de musique de chambre - Gauche : Concours Guthman d'instruments de musique à Atlanta, 2015. Droite : Concert Surchauffe à Metz, 2015.*

Figure 7.11 – *Disposition pour un Raga - Gauche : JDEV, 2012. Droite : Concert Surchauffe à Metz, 2015.*

7.3.2 Des chorals baroques aux voix de zombies, l'évolution du répertoire musical

La conception de l'instrument a été initiée dans un contexte scientifique. Par conséquent, aucun *a priori* sur le répertoire musical n'a été défini avant d'avoir l'instrument prêt à jouer. Une exploration des styles de musiques vocales a donc été effectuée lors de nos répétitions.

Le répertoire vocal classique

La création d'un chœur, même digital, incite en premier lieu à interpréter des pièces chorales. Celles-ci sont des pièces à quatre voix : *soprano*, *alto*, *ténor* et *basse* où les deux premières sont des voix de femmes, et les deux autres des voix d'hommes. Ecrites au XV ou XVI^{e} siècle, ces pièces étaient chantées lors de cérémonies religieuses protestantes. Elles sont écrites de manière verticale, les quatre voix étant souvent en homorythmie. Deux chorals ont été joués : *Alta Trinita*, anonyme italien du XV^{e} siècle, et *Wie schön leuchtet der Morgenstern* de Bach. Un autre type de musique religieuse a été travaillé en répétition mais jamais joué en concert. Il s'agit du *Kyrie* de la *Messe de Notre Dame* de Machaut.

Une deuxième forme de chant dans le répertoire classique est le chant soliste. Nous avons choisi l'opéra et proposé une interprétation de *l'Air du Génie du Froid* extrait de l'opéra *Roi*

Arthur de Purcell. Bien que la partie de soliste se chante parfois avec une voix de contre-ténor (homme chantant en voix de tête), nous avons opté pour l'interprétation originale par une voix de baryton (registre grave d'une voix d'homme). Au lieu de chanter le texte, le soliste improvise librement sur les voyelles en fonction des nuances de la pièce. Pour imiter l'accompagnement écrit pour ensemble baroque composé de cordes et clavecin, les musiciens du *Chorus Digitalis* jouent chaque partie de manière très *staccato*. La voyelle /e/ est utilisée pendant l'accompagnement du soliste, et la voyelle /a/ pendant les passages *forte* entre les interventions du soliste.

Toutes les pièces jouées en concert (chorals et *Air du froid*) ont été accompagnées de clavicorde lorsqu'il était possible d'amener l'instrument. Le choix du répertoire classique est risqué. Bien qu'il permette de démontrer les capacités de l'instrument dans l'imitation de voix réelles, les attentes du public pour ce type de musique sont extrêmement élevées de par le cadre très strict imposé par la musique savante. Généralement, le peu de temps disponible pour répéter ne nous permet pas de perfectionner chaque pièce comme nous le ferions avec nos instruments acoustiques respectifs. Ces considérations nous ont donc amené à étendre le répertoire à d'autres styles de musiques.

Musiques du monde

Une des directions choisies a été de franchir les frontières occidentales pour interpréter des musiques du monde. Le premier type de musique testé et joué à chaque concert est le *Raga* qui est la musique vocale savante d'Inde du Nord. Un *Raga* est un cadre de jeu auquel est associé un mode musical de sept notes, ce dernier évoquant des sentiments particuliers comme la joie ou la tristesse, ou un moment de la journée. Chaque raga est construit sur un thème, par dessus lequel le chanteur soliste improvise librement, et est traditionnellement accompagné de tablas et d'une tampura. Les tablas sont deux percussions sous la forme de petits tambours. Le plus grave est accordé sur la tonique du morceau, et le plus aigu sur le quatrième ou cinquième degré du mode. La tampura est un instrument à cordes jouant un bourdon donnant la tonique au musicien soliste.

Les ragas que nous avons interprétés se découpent en deux parties. La première, *l'Alap* est une partie lente, arythmique et sans percussions pendant laquelle le chanteur expose le mode du morceau. Vient ensuite le *Bandish*, où une pulsation est introduite par les tablas, appelée *Tala*. De nombreux cycles rythmiques existent parmi lesquels nous avons interprété un *Ektal* (douze temps) et un *Teental* (seize temps). Un dialogue se met alors en place entre le chanteur et le joueur de tabla. Le *Bandish* aboutit sur le climax de la pièce, puis à sa dissolution brutale, laissant le soliste conclure sur une base de tampura.

Le jeu soliste du raga nécessite des effets de portamentos et une virtuosité importante que la tablette graphique permet de réaliser. Au prix d'une pratique soutenue, Boris Doval, le musicien soliste, a pu développer une technique de jeu propre à ce style en utilisant un calque dédié sur la tablette (figure 2.5). Les tablas sont joués par Lionel Feugère, et la tampura est imitée par le reste du chœur, tenant des bourdons dans le registre grave d'une voix d'homme en changeant lentement de voyelles permettant des déplacements audibles des formants dans l'aigu. Deux ragas ont été interprétés au fil de nos concerts, un *Raga Yaman* en *Teental* et un *Raga Miyan Ki Milhar* en *Ektal*. De nombreux retours positifs nous sont parvenus de spectateurs familiers à ce type de musique.

Les possibilités de modifications du modèle vocal nous ont permis de créer un type de voix tendue couramment utilisée dans la musique traditionnelle bulgare. Nous avons proposé une

interprétation du traditionnel *Si tu savais, ma mère*. Ce morceau est écrit en mesures incomplètes (7 croches), se divisant en deux temps courts et un temps long, soit le regroupement de croches 2+2+3. L'enregistrement de référence sur lequel nous avons construit l'arrangement contient trois voix de voix femmes écrites en homorythmie, et un accompagnement de guitare, contrebasse, violon, flûte et percussions. Lors de notre première interprétation du morceau, les chanteurs virtuels étaient accompagnés de violon, de guitare et de percussions. La partie de contrebasse était jouée par une voix de basse et trois autres *Cantor Digitalis* jouaient les voix bulgares. Les parties de chants alternaient l'usage de voyelles /e/ pour le premier thème, et /o/ et /a/ pour les répétitions du deuxième thème *piano* et *forte* respectivement. Dans notre dernière interprétation, seules une percussion et une guitare accompagnaient les trois voix. Les intermèdes de violons étaient joués à la voix sur une voyelle ouverte /a/ accompagnée de notes tenues par le reste du chœur sur la voyelle /u/. Le tempo rapide et le motif rythmique particulier de cette pièce ont fait d'elle une des plus dynamiques de notre répertoire.

Répertoire contemporain

La dernière partie de notre répertoire regroupe sept morceaux occidentaux du 20^{e} et 21^{e} siècle. *North Star* est un morceau écrit par Philip Glass en 1977, faisant partie du mouvement de la musique minimaliste. Il s'agit d'une superposition progressive de six voix bouclant des structures de quatre mesures, accompagnées d'un orgue électrique. Une des interprétations de cette pièce est réalisée par des voix ne chantant que des voyelles. Il s'agit donc d'une pièce bien adaptée à notre instrument. La difficulté du jeu réside principalement dans la synchronisation entre musiciens des enchaînements de croches jouées à 120 b.p.m. présentant des intervalles relativement grands (sixtes) pour certaines voix.

Ocean est une pièce de Dirty Projectors et de Björk écrite en 2009. Il s'agit de trois voix de femme enchaînant des glissandos longs entre une note de départ commune et des notes d'arrivées divergentes, sur un bourdon grave. La première répétition de la pièce est effectuée sur une voyelle /u/ et la deuxième sur /ɛ/. A nouveau, l'adaptation pour *Chorus Digitalis* est immédiate, permettant une interprétation convaincante de cette pièce.

Valse de Bruno Lecossois est un morceau écrit pour l'ensemble vocal *a capella Les Grandes Gueules* composé de quatre voix de femmes, et deux voix d'hommes. Cette pièce est construite sur un motif rythmique réalisé par quatre des voix. Cette cellule est conservée tout le morceau, s'adaptant aux changements d'harmonie. Le thème est effectué par deux voix soprano, à l'unisson d'abord puis à la tierce. Comme précédemment, aucune parole n'est prononcée dans cette pièce, seulement des enchaînements d'onomatopées. Il s'agit à nouveau d'un morceau bien adapté pour notre ensemble. Néanmoins, l'absence de consonnes sur le *Cantor Digitalis* limite la possibilité de réaliser des attaques percussives présentes dans l'interprétation originale, rendant notre version moins dynamique.

Lil Darlin est une pièce de jazz écrite par Neal Hefti en 1957. Elle a été reprise par de nombreux musiciens comme Monty Alexander ou Ray Charles. Son interprétation vocale la plus célèbre est celle d'Henri Salvador en 1964. L'adaptation pour *Chorus Digitalis* reprise de la version originale est écrite pour soprano, alto, deux ténors et une basse. Le thème est donné à la soprano. Alto et ténors complètent l'harmonie en homorythmie, la voix de basse joue une ligne de basse et l'ensemble est accompagné d'une caisse claire. La difficulté de ce morceau est d'apporter une dynamique malgré le tempo très lent de la pièce.

Les Profondeurs est une pièce écrite par Boris Doval pour le *Chorus Digitalis* dans le cadre du festival CuriositAS 2013 sur le thème de l'eau. Ecrite pour quatre voix, alto, ténor et basse répètent un motif rythmique très lent faisant progresser l'harmonie dans le morceau. Par dessus, la partie de soprano joue des longues tenues dans l'aigu.

Circlesong 6 est une pièce de Bobby McFerrin extraite de l'album *Circlesongs* paru en 1997. Le morceau est construit sur des cellules rythmiques courtes enregistrées par le chanteur et jouées en boucles. Celui-ci réalise par dessus une longue improvisation vocale. On distingue quatre voix d'accompagnement réparties en soprano, alto, ténor et basse pour les musiciens du *Chorus Digitalis*. Lors de l'unique représentation de ce morceau, l'improvisation a été confiée à un chanteur invité, Olivier Chardin, dialoguant alternativement avec les membres du chœur. Un accompagnement de percussions corporelles était réalisé par une des membres de l'ensemble.

Le Lion est Mort ce Soir est une chanson traditionnelle africaine composée en 1939 par Solomon Linda. Elle est devenue un succès mondial après de nombreuses reprises. En France, on peut citer celle d'Henri Salvador en 1962. C'est la version *a capella* du groupe *Pow Wow* qui a inspiré l'arrangement pour le *Chorus Digitalis*. Notre version se traduit en une voix soliste ténor, accompagnée d'un quatuor soprano, alto, ténor et basse. Ces derniers jouent en homorythmie le célèbre motif "Ohimbawé", simplifié en /oiae/ sur la tablette. Cette pièce possède de nombreux avantages démonstratifs. D'abord, sa popularité permet de vulgariser la pratique du *Cantor Digitalis* au grand public. Ensuite, les motifs vocaliques /oiae/ sont parfaitement reconnaissables et permettent de montrer la capacité articulatoire de notre instrument, bien que réduite aux seules voyelles.

L'apport d'instruments acoustiques

Comme indiqué précédemment, nous avons souvent fait le choix d'introduire des instruments acoustiques pour accompagner le chœur, profitant de l'expérience musicale de chacun des membres. L'apport d'instruments acoustiques permet d'intégrer le *Cantor Digitalis* dans un contexte musical déjà existant, et montre sa compatibilité avec d'autres sonorités. Par ailleurs, la virtuosité comparable du joueur soliste de *Cantor Digitalis* dans le raga et du joueur de tabla a été apprécié du public, démontrant que notre instrument permet d'atteindre des virtuosités comparables à celles d'instruments acoustiques.

Exploration du modèle vocal

Le répertoire proposé précédemment utilise le *Cantor Digitalis* comme imitateur de la voix chantée. Cela a permis de démontrer la capacité de l'instrument à reproduire le répertoire vocal existant, avec un succès partagé selon le style de musique. La critique majeure que nous avons reçue est la frustration de certains spectateurs percevant notre concert comme une reproduction de mauvaise qualité d'une pièce vocale, découlant de la qualité moindre de la voix de synthèse, comparée à la voix réelle.

Afin de contourner ce problème et d'apporter une plus-value à l'instrument, nous avons tenté de produire des sons depuis un appareil vocal numérique non réalisables à partir d'un appareil vocal réel. Cela consiste en des réglages extrêmes des paramètres de source tels que la taille du conduit vocal, l'aspiration, la tension, l'apériodicité et la tessiture de la voix. Par de nombreux tests nous avons pu élaborer un ensemble de sons vocaux extrêmes résumés dans

Texture sonore	Tessiture (MIDI)	Taille du conduit vocal [0,1]	Souffle [0,1]	Tension [0,1]	Rugosité [0,1]	Phonation
Lion	8→43	1	1	1	0.8	Oui
Desert breeze	68→103	1	0.9	0.85	1	Oui
Wood bells	8→43	0	1	1	1	Oui
Ring modulator	68→103	0	1	1	1	Oui
Pitched noise	68→103	1	1	0	1	Oui
Mouette	56→91	0.65	0.3	0.8	0	Oui
Corne de brume	8→43	1	0	0	0	Oui
Vent 1	56→91	0.65	1	0.1	0	Non
Vent 2	8→43	1	1	0	0	Non
Ping-pong	-16→19	0.47	1	1	0	Oui
Didjeridoo	8→43	1	1	0	0	Oui
Zombie 1	56→91	0	0.5	0	0	Oui
Zombie 2	44→79	0.47	0.8	0	1	Oui
Zombie 3	32→67	0.47	1	0	1	Oui
Zombie 4	56→91	0	0.75	0.5	0	Oui

Table 7.2 – *Résumé des différents réglages définis pour des voix extrêmes. Les valeurs sont normalisées entre 0 et 1 correspondant respectivement aux positions basse et haute des curseurs sur l'interface.*

le tableau 7.2. Les paramètres donnés ici sont les valeurs normalisées choisies sur l'interface du *Cantor Digitalis*. Des exemples sonores sont donnés à l'url en bas de page [4].

La première, d'abord baptisée *Extreme voice* puis renommée en *Lion* allonge le conduit vocal d'une voix d'homme, et rajoute de la tension et de l'aspiration. On se rapproche alors d'une voix animale. Celle-ci a été utilisée pour rugir à la fin du *Lion et Mort ce Soir*.

Les quatre textures suivantes ont été développées pour le concours Guthman d'instruments de musique. Le but était de démontrer les possibilités de l'instrument, et une improvisation de quelques minutes a été réalisée avec ses sons. Il s'agit de textures très éloignées de sons vocaux, bien que calculées avec un modèle vocal. *Desert breeze* et *Pitched noise* sont des sont très bruités. *Wood bells* utilise une tessiture très grave permettant de distinguer les coups de glottes individuellement. On obtient alors une sorte d'instruments à percussions.

Nous nous sommes ensuite intéressés à la reproduction de sons existants non vocaux par notre modèle. Les sons de *Mouette*, *Corne de brume* et de *Vent* ont alors été créés pour une très courte improvisation maritime. La *Mouette* est une voix de soprano tendue. La corne de brume est une voix de basse munie d'un très grand conduit vocal et d'une importante réverbération. Le vent dérive d'une déconstruction du modèle vocal, ou seul le bruit de source est filtré par un ou deux formants. Un son de *Ping-pong* proche de *Wood Bells* isole les coups de glottes pour simuler le rebond d'une balle sur une raquette.

Enfin, quatre voix qu'on appellera *zombie* combinent des taux d'aspiration, de tension et des tailles de conduits vocaux anormalement élevés pour des êtres humains. Celles-ci ont été crées pour répondre à la critique portant sur notre *Air du Froid* trop synthétique. Lors de notre dernière interprétation de ce morceau, l'accompagnement vocal a été remplacé par un accompagnement zombie. Cela a permis de s'éloigner d'une interprétation classique, tout en mettant plus en valeur la partie soliste toujours jouée avec une voix de baryton. Une recherche plus subtile dans la définition de ces voix serait probablement nécessaire, mais cette expérience nous ouvre des portes vers l'adaptation de pièces vocales vers des pièces pour conduits vocaux extrêmes, qui fait la particularité du *Cantor Digitalis*.

4. https://perso.limsi.fr/operrotin/these.fr.php#Annexes (vérifié le 22 octobre 2015)

7.4 Liste des concerts

Depuis mon entrée dans le *Chorus Digitalis* en 2013, nous avons pu nous produire six fois, en France et à l'étranger. Cette partie présente chacun de ces concerts et leurs particularités. Les vidéos des différents concerts peuvent être visualisées sur notre site web [5].

Vox Tactum Meets Chorus Digitalis : Seven Years of Singing Surfaces

Pour la conférence *New Interfaces for Musical Expression (NIME)* 2013 à Daejeon, en Corée, un projet de réunion entre *Chorus Digitalis* et *Vox Tactum* a été entrepris [ddF+13]. *Vox tactum* est un ensemble créé en 2011 par des membres de l'université de Mons et de British Columbia (Vancouver) réunissant les instruments *Handsketch* [dD09b], *ChoirMob* et le système *Vuzik*[6]. Finalement, les musiciens présents en Corée pour mener le projet à bien étaient trois membres du *Chorus Digitalis*, et Nicolas d'Alessandro et son *Handsketch*. Ce fut l'occasion de montrer l'évolution de deux instruments découlant du même projet initié au workshop eNTERFACE 05 [ddLB+05]. Une "Cantate" d'une dizaine de minutes a été préparée, enchaînant plusieurs éléments musicaux, comme du chant diphonique, un dialogue d'onomatopées, des rires, et le choral *Wie schön leuchtet der Morgenstern* de Bach. Ce concert fût aussi l'occasion de tester pour la première fois la position assise au sol avec les supports.

Chorus Digitalis - saison 3.0

Pendant l'année 2013-2014, l'ensemble s'est produit à trois reprises lors d'évènements art-sciences sur le campus de l'université Paris-Sud. Le premier concert a eu lieu à l'école Polytechnique pour les Journées du Développement JDEV en septembre 2013. Nous avons joué en quatuor, en position assise sur le sol. La première partie était un Raga joué par Boris Doval et Lionel Feugère. La deuxième partie était à nouveau une forme Cantate, combinant morceaux classiques : le choral *Wie schön leuchtet der Morgenstern* de Bach et *l'Air du Froid* de Purcell avec des discours d'onomatopées, des rires, et des glissandos infinis.

Chorus Digitalis - saison 3.1

Le deuxième concert de la saison 2013-2014 s'est déroulé en octobre au bâtiment Sciences ACO à Orsay lors du festival CuriositAS. Deux musiciennes ont complété le quatuor du dernier concert. En plus du Raga et de la Cantate présentés précédemment, la *Valse*, *Lil Darlin*, *Les Profondeurs* et *Alta Trinita Beata* ont été interprétés. Ce concert a aussi été l'occasion de présenter le retour visuel qui se fondait relativement bien dans le décor de la scène à la fois rétro et futuriste.

Chorus Digitalis - saison 3.2

Le dernier concert de la saison a eu lieu au PROTO 204 en juin 2014 à Orsay, à l'occasion du festival Futur en Seine. Sept musiciens ont participé à ce concert et trois nouveaux morceaux ont été introduits : *Su tu savais, ma mère*, *Le lion est mort ce soir* et *Circlesong 6* avec la participation du chanteur Olivier Chardin. *North Star* qui avait été joué une première fois en 2012 a été de nouveau interprété. Le retour visuel était aussi projeté, mais derrière l'audience pour des contraintes pratiques. Cette fois-ci, ce dernier n'a eu que peu d'impacts puisque le public l'a très peu regardé.

5. https://cantordigitalis.limsi.fr/chorusdigitalis_fr.php (vérifié le 22 octobre 2015)
6. http://www.nicolasdalessandro.net/choirmob/ (vérifié le 22 octobre 2015)

Compétition Margaret Guthman d'instruments de musique

L'année 2014-2015 s'est caractérisée par la participation à la compétition Margaret Guthman d'instruments de musique organisé par Georgia Tech aux Etats-Unis. Dans une optique de concours où l'instrument serait évalué, un travail de préparation beaucoup plus conséquent a été effectué. Pour des raisons budgétaires, seuls les quatre membres développant aussi l'instrument ont participé à ce projet. Le format demandé était un créneau de 20 minutes pendant lesquelles nous étions libre de parler, d'expliquer le fonctionnement de l'instrument, et de jouer. Nous avons pris parti de combiner les deux en sélectionnant les morceaux les plus démonstratifs des capacités du *Cantor Digitalis*. Un morceau de chaque répertoire a été choisi. *L'air du froid* a été interprété avec accompagnement de voix normales pour démontrer la capacité de soliste de l'instrument dans de la musique savante occidentale. Un raga a ensuite été présenté pour démontrer la virtuosité intonative de l'instrument. S'en est suivi une improvisation de quelques minutes de voix extrêmes afin de dévoiler la flexibilité de notre modèle, et les nouvelles perspectives offertes par l'instrument. Enfin, nous avons terminé par le *Lion est mort ce soir* à la fois pour la démonstration du contrôle des voyelles, et montrer que la musique populaire est accessible par notre instrument. La compétition a été aussi l'occasion de repenser notre disposition scénique, faisant émerger la disposition de type musique de chambre. Le *Cantor Digitalis* a finalement remporté le premier prix de la compétition.

Soirée Surchauffe

Suite à une collaboration de l'antenne de Georgia Tech en Lorraine et l'Arsenal de Metz, les gagnants du concours Guthman ont été invité à jouer en première partie du Concert Surchauffe joué par l'orchestre National de Lorraine, introduisant un nouvel instrument de percussion acoustique. Il nous a été accordé un créneau de 20 minutes pendant lesquelles nous avons alterné morceaux et intermèdes en commençant par un dialogue d'onomatopées suivi d'un Raga, d'une improvisation maritime utilisant les voix *mouette*, *vent* et *corne de brume*, de *l'Air du Froid* accompagné de voix de zombies, d'un match de ping-pong et concluant par *Si tu savais, ma mère*.

7.5 Conclusion

Ce chapitre a présenté la partie artistique de ce travail de thèse. Bien que plus informelle du point de vue scientifique, la pratique de l'instrument en contexte de concert a permis à la fois de développer un logiciel robuste, de démontrer les capacités de notre instrument, et de le faire connaître du grand public. La participation à la compétition Guthman d'instrument de musique a été un évènement essentiel dans l'évolution de la pratique de l'instrument, tant dans l'exploration de notre modèle d'appareil vocal que dans l'amélioration de notre disposition scénique, et nous faisant connaître du milieu de la musique numérique. Le premier prix nous a permis une petite couverture médiatique entraînant des centaines de téléchargements du logiciel et l'adaptation de notre moteur de synthèse à d'autres interfaces telles que le *Continuum Fingerboard* ou le *Soundplane*[7].

Finalement, l'exploration du modèle vocal a ouvert des portes à de nombreuses possibilités musicales et a permis d'identifier la plus-value de notre instrument, comparé à de la voix réelle ou des moteurs de synthèses par corpus par exemple. La malléabilité de notre modèle suggère d'ajouter des contrôles des paramètres de sources plus accessibles que par la souris de l'ordinateur, afin de pouvoir en faire usage plus finement lors de prestations.

7. https://cantordigitalis.limsi.fr/use_en.php (vérifié le 22 octobre 2015)

Conclusion générale et perspectives

Contributions de la thèse

Ces travaux nous ont permis d'explorer l'usage de la tablette graphique pour le contrôle de la voix chantée dans le cadre de l'instrument *Cantor Digitalis*. Trois aspects ont émergé de ces recherches : l'adéquation de la tablette graphique augmentée d'une aide à la justesse pour un contrôle mélodique continu ; l'importance de la vision dans le maniement de la tablette graphique ; l'exploration de gestes pour un jeu expressif de l'instrument.

Une interface adaptée pour le contrôle de l'intonation

Le contrôle de l'intonation a été abordé sous deux angles. D'abord, une expérience d'imitation de motifs mélodiques a été conduite afin de comparer la précision permise par la tablette graphique avec les possibilités offertes par l'appareil vocal. Les résultats ont montré une plus grande facilité à jouer juste et précis avec un stylet sur une tablette graphique qu'avec la voix. Dans le cas de sujets présentant une forte expérience musicale, leurs justesses chironomiques (jeu avec la tablette) sont comparables à leurs justesses vocales. Dans le cas de sujets non musiciens, la tablette permet d'atteindre des seuils de justesses et précisions bien en deçà de leurs performances vocales.

Cette expérience a mis en valeur deux caractéristiques de la tablette graphique, et plus particulièrement des indices visuels présents sur la surface de cette dernière. D'abord, ces indices font de la tâche de contrôle de l'intonation une tâche de visée de cibles absolue, contrastant avec la perception relative d'intervalles prédominante en musique. Ensuite, la présence d'indices visuels a inhibé l'attention que portent les sujets sur le rendu auditif de leur performance. En effet, aucunes différences de justesse et précision n'ont été constatées entre imitations avec et sans retour audio. Ce problème a été traité dans une autre étude.

Après la validation de la tablette graphique comme candidat au contrôle de l'intonation, diverses méthodes automatiques de correction de justesse ont été explorées afin d'aider le jeu de l'instrument. Deux méthodes de déformation dynamiques sont proposées appelées DPW (*Dynamic Pitch Warping*). La première présente une fonction de correction fixe appelée *notes étendues*. La deuxième est adaptative et appelée *élastique*. Ces méthodes ont été testées dans des contextes d'attaques de notes et de contours continus. La correction d'attaque est immédiate et très efficace. L'ajustement de contours mélodiques est plus délicat et demande la détection de zones stables dans la trajectoire par un choix de trois paramètres : l'intervalle de détection et le temps critique qui définissent la sensibilité de la correction par rapport la stabilité de trajectoire ; le temps de transition qui définit le lissage de l'application de la correction. Une observation empirique d'effets musicaux a permis de proposer des ordres de grandeurs pour ces paramètres, permettant de conserver certaines formes d'expressivité vocales telles que le vibrato, le portamento ou le glissando.

L'efficacité de la correction est démontrée par une expérience d'imitation où les valeurs de justesse et précision sont significativement réduites avec l'application de la correction. Une expérience perceptive a montré que l'effet de la correction est perçu par les auditeurs. La comparaison de l'efficacité de la correction sur des sujets musiciens et non musiciens montre qu'une expérience musicale minimale est nécessaire pour tirer profit de l'ajustement. En effet, la note cible doit être atteinte à une erreur raisonnable (plus proche de la cible que de la note voisine) et la trajectoire doit être stable autour des notes cibles. La correction est aussi validée par son usage lors de concerts avec l'ensemble *Chorus Digitalis*. Un défaut de justesse est d'autant plus audible qu'il y a de musiciens, et la correction apporte un grand confort pour le jeu d'ensemble.

Finalement, la combinaison de ces deux études permet de proposer une interface idéale pour le contrôle de l'intonation. La précision permise par le maniement du stylet sur la tablette ainsi que les indices visuels placés sur sa surface rendent accessible le jeu de mélodies aux non-musiciens. L'ajout d'un mapping dynamique pour ajuster automatiquement la hauteur des utilisateurs ajoute un confort dans le jeu de l'instrument, facilitant le jeu juste et permettant de se concentrer sur l'expressivité de la mélodie, non altérée par la correction.

La modalité visuelle, inhérente à l'art calligraphique

L'ajout d'un retour visuel sur la tablette pour aider au contrôle de l'intonation introduit l'usage de la perception visuelle dans le maniement de l'instrument, contrairement à la plupart des instruments de musique acoustiques. L'étude de la justesse chironomique a montré qu'en présence de retour visuel, l'audio n'avait pas d'influence significative sur la précision du contrôle. Une étude cognitive a donc été réalisée pour quantifier l'influence des modalités visuelles et auditives sur le contrôle de l'intonation. Des retours discordants ont été soumis à des sujets effectuant des tracés rectilignes sur la tablette, simulant le contrôle de l'intonation. Les résultats ont montré que le tracé des sujets est influencé par le retour auditif lorsque ce dernier est présenté seul, mettant en évidence la sensibilité des joueurs parfois non musiciens à leurs environnements sonores. En revanche, lorsque les retours visuel et auditif sont présentés simultanément, seul le retour visuel a un effet sur le contrôle, comme il a été observé dans l'expérience précédente.

Ces résultats démontrent l'importance de la vue dans le maniement de la tablette graphique, développée initialement pour les arts visuels (calligraphie, dessin). Cela soulève donc la question de la pertinence d'une telle interface pour le contrôle d'un instrument de musique. L'expérience réalisée ici uniquement pour le contrôle de l'intonation ne quantifie pas l'apport de chaque modalité pour les autres tâches musicales telles que le jeu expressif ou le jeu d'ensemble. Bien qu'on puisse supposer que la modalité auditive soit indispensable à ces tâches, une expérience complémentaire apporterait une base solide sur les processus cognitifs engagés dans l'utilisation de la tablette graphique dans un contexte musical au sens large.

Développement du jeu de l'instrument

L'exploration des gestes dans le contrôle expressif s'est déroulée en deux étapes : l'étude de la temporalité du contrôle de l'intonation, et la proposition de gestes pour le contrôle d'effets expressifs vocaux. L'observation des temps de transitions entre deux notes dans le contrôle de l'intonation permet de retrouver certaines lois temporelles démontrées dans le domaine de l'interaction homme-machine. Une évolution logarithmique du temps de transition en fonction de l'intervalle à parcourir apparaît dans un contexte de pointage telles que les expériences

d'imitation proposées pour l'évaluation de la justesse de jeu (loi de Fitts). Une évolution linéaire du temps de transition en fonction de l'intervalle se manifeste dans un contexte vocal et musical. Néanmoins, l'hétérogénéité des données utilisées n'a pu que suggérer des tendances. Par ailleurs, aucune loi n'a pu être extraite de tâches chironomiques musicales extraites des répétitions du *Chorus Digitalis*. On peut toutefois émettre l'hypothèse qu'une tâche musicale même chironomique tend à uniformiser les temps de transition entre chaque intervalle se démarquant ainsi de la loi de Fitts tant observée dans des tâches de pointage. Une expérience serait à conduire pour confirmer ou infirmer cette hypothèse.

Parallèlement, cinq gestes de contrôle pour jouer *legato*, *staccato*, *glissando*, *vibrato* ou *virtuoso* sont suggérés par expérience de jeu. Le répertoire musical exploré avec l'ensemble *Chorus Digitalis* a permis de travailler successivement chacune des techniques vocales et de trouver à chaque fois un geste adapté, à la fois naturel pour le musicien et porteur d'expressivité. La confrontation de ces gestes aux lois du coût du mouvement a montré qu'aucun n'était optimal en terme de minimisation des coûts. En effet, ce n'est pas l'efficacité d'un mouvement qui est recherchée ici, mais son contrôle fin permettant au musicien de faire preuve d'expressivité. Par ailleurs, la majorité des gestes identifiés sont utilisés dans l'écriture. Un transfert de compétence entre écriture et contrôle mélodique a donc été effectué afin de profiter à la fois de l'expertise et de l'expressivité du geste d'écriture.

La pratique de l'instrument au sein du *Chorus Digitalis* a permis de développer à la fois l'aspect visuel et l'aspect auditif d'une représentation. L'aspect visuel se traduit d'abord par la recherche d'une présentation scénique confortable pour les musiciens et ouverte au public, mettant en évidence la manipulation de l'instrument. Trois dispositions ont été testées au fil des concerts, en minimisant à chaque fois la taille du support pour la tablette. De plus, un retour visuel a été développé et proposé pour deux concerts, permettant d'observer les contours d'intonation ainsi que l'articulation de chaque musicien. Un retour d'expérience a montré un certain intérêt du public pour ce visuel qui permet une meilleure compréhension du maniement de l'instrument, mais un travail de graphisme est à accomplir pour inclure ce retour dans une dimension artistique plus qu'informative.

L'aspect auditif se caractérise par l'association d'un répertoire à l'instrument, lui conférant une identité musicale. Une partie conséquente des différents styles du répertoire vocal a été essayée, des chorals baroques aux chansons populaires occidentales en passant par les ragas d'Inde du nord ou les traditionnels Bulgares. Parmi ces styles, les ragas d'Inde du nord demandent une virtuosité mélodique pour laquelle la tablette est particulièrement bien adaptée. Ensuite, l'exploration de voix non réelles dévoile une valeur ajoutée de l'instrument, permettant de produire des sons avec un modèle de conduit vocal qui n'existe pas dans la nature. Ces deux styles semblent donc correspondre particulièrement à l'instrument, le premier adapté à l'interface, l'autre au moteur de synthèse.

Perspectives

Les résultats obtenus dans cette thèse incitent à poursuivre des travaux dans deux directions, vers une étude plus poussée de la gestuelle de contrôle, et l'adaptation à d'autres systèmes de synthèse.

Evaluation de l'apport du geste expressif sur le rendu sonore

L'exploration de la gestuelle induite par la tablette graphique pour un contrôle vocal expressif a été effectuée de manière empirique pendant cette thèse. La caractérisation de ces gestes par des études scientifiques apporterait une justification plus solide de leur pertinence pour le contrôle vocal. Une première étude consisterait à mettre en relation gestes empiriques et lois théoriques tel qu'il a été tenté dans ces travaux. Une des lois biologiques peu abordée ici est la loi de puissance 2/3 associée à l'écriture qui semblerait plus adaptée au type de gestes choisis. Par ailleurs les styles d'écritures très variés de chaque individu incitent à étudier la présence d'un style de jeu intonatif propre à chaque musicien, en relation avec leurs graphies respectives.

Lors du choix du moteur de synthèse et de l'interface pour l'instrument, il a été affirmé que la qualité moindre d'une synthèse à formant comparée à d'autres méthodes serait compensée par l'expressivité apportée par le contrôle en temps réel du musicien. La pratique de l'instrument montre en effet que le son produit par le moteur de synthèse est d'autant meilleur que le musicien est entraîné à manier l'instrument. Néanmoins, aucune expérience n'a été conduite pour valider ces résultats. Une expérience perceptive de comparaison de stimuli générés par des trajectoires chironomiques ou prédéfinies et synthétisés par des moteurs divers serait à envisager. Une telle expérience justifierait complètement le choix d'un tel moteur de synthèse, et permettrait de comparer aussi le contrôle expressif du musicien proposé par une tablette graphique, et le contrôle différé réalisé par un ensemble de règles prédéfinies à l'avance.

Adaptation du contrôle vers d'autres modes de synthèses

Avec la puissance de calcul croissante des processeurs, le temps réel n'est presque plus une contrainte pour le choix du moteur de synthèse. Un des types de synthèse phares aujourd'hui est la synthèse par concaténation. Celle-ci utilise des segments de parole très courts pré-enregistrés et permet la reconstruction d'un signal de qualité excellente. En contrepartie, chaque son synthétisé doit avoir été enregistré au préalable. Dans le cas contraire, des traitements sont réalisés pour extrapoler les enregistrements, réduisant la qualité de la synthèse. Il est donc difficile d'obtenir des effets vocaux très variés tels que différentes tensions de voix, différentes aspirations, ou des gammes d'effort vocal très importantes. Néanmoins, l'excellente qualité vocale et la possibilité d'articuler et de prononcer de la parole intelligible font de cette synthèse un candidat de choix pour des applications commerciales destinées au grand public.

Ces travaux ont démontré l'adéquation de la tablette graphique pour le contrôle de la voix chantée. Par conséquent, son adaptation pour le contrôle d'autres moteurs tels que la synthèse par concaténation est à envisager. Le problème majeur sera alors le contrôle des consonnes en temps réel, encore irrésolu. En allant encore plus loin, l'exploration de synthèses d'autres instruments de musique est aussi envisageable. Cela placerait la tablette graphique dans un contexte musical plus vaste, posant la question de la pertinence du geste chironomique pour le contrôle d'instruments de musiques numériques à espace de contrôle continu.

ANNEXES

Annexe A

Calculs des coûts théoriques de trajectoires polynomiales

A.1 Trajectoire polynomiale d'ordre 5 - Minimisation de la secousse

A.1.1 Expression de la trajectoire

Soit un mouvement d'une dimension d'amplitude A, de durée T et démarrant en position x_0 à l'instant t_0. La trajectoire minimisant le coût de secousse est un polynôme d'ordre 5 défini comme :

$$\begin{cases} x_{poly5}(t) &= A_0 + A_1(t-t_0) + A_2(t-t_0)^2 + A_3(t-t_0)^3 + A_4(t-t_0)^4 + A_5(t-t_0)^5 \\ \dot{x}_{poly5}(t) &= A_1 + 2A_2(t-t_0) + 3A_3(t-t_0)^2 + 4A_4(t-t_0)^3 + 5A_5(t-t_0)^4 \\ \ddot{x}_{poly5}(t) &= 2A_2 + 6A_3(t-t_0) + 12A_4(t-t_0)^2 + 20A_5(t-t_0)^3 \\ \dddot{x}_{poly5}(t) &= 6A_3 + 24A_4(t-t_0) + 60A_5(t-t_0)^2 \end{cases} \tag{A.1}$$

Avec les conditions initiales suivantes :

$$\begin{cases} \leq x_{poly5}(t_0) &= x_0 & \text{Position initiale} \\ x_{poly5}(t_0+T) &= x_0 + A & \text{Position finale} \\ \dot{x}_{poly5}(t_0) &= 0 & \text{Vitesse initiale} \\ \dot{x}_{poly5}(t_0+T) &= 0 & \text{Vitesse finale} \\ \ddot{x}_{poly5}(t_0) &= \Gamma_0 & \text{Accélération initiale} \\ \ddot{x}_{poly5}(t_0+T) &= -\Gamma_0 & \text{Accélération finale} \end{cases}$$

Après résolution du système on obtient :

$$\begin{cases} x_{poly5}(t) &= x_0 + \frac{\Gamma_0}{2}(t-t_0)^2 + \frac{10A-2\Gamma_0T^2}{T^3}(t-t_0)^3 + \frac{-15A+\frac{5}{2}\Gamma_0T^2}{T^4}(t-t_0)^4 + \frac{6A-\Gamma_0T^2}{T^5}(t-t_0)^5 \\ \dot{x}_{poly5}(t) &= \Gamma_0(t-t_0) + \frac{30A-6\Gamma_0T^2}{T^3}(t-t_0)^2 + \frac{-60A+10\Gamma_0T^2}{T^4}(t-t_0)^3 + \frac{30A-5\Gamma_0T^2}{T^5}(t-t_0)^4 \\ \ddot{x}_{poly5}(t) &= \Gamma_0 + \frac{60A-12\Gamma_0T^2}{T^3}(t-t_0) + \frac{-180A+30\Gamma_0T^2}{T^4}(t-t_0)^2 + \frac{120A-20\Gamma_0T^2}{T^5}(t-t_0)^3 \\ \dddot{x}_{poly5}(t) &= \frac{60A-12\Gamma_0T^2}{T^3} + \frac{-360A+60\Gamma_0T^2}{T^4}(t-t_0) + \frac{360A-60\Gamma_0T^2}{T^5}(t-t_0)^2 \end{cases} \tag{A.2}$$

A.1.2 Cas particuliers :

– Minimisation du coût de secousse : $\Gamma_0 = \frac{5A}{T^2}$

$$\begin{cases} x_{poly5}(t) &= x_0 + A\left[\frac{5}{2}\left(\frac{t-t_0}{T}\right)^2 + \frac{-5}{2}\left(\frac{t-t_0}{T}\right)^4 + \left(\frac{t-t_0}{T}\right)^5\right] \\ \dot{x}_{poly5}(t) &= \frac{5A}{T}\left[\left(\frac{t-t_0}{T}\right) - 2\left(\frac{t-t_0}{T}\right)^3 + \left(\frac{t-t_0}{T}\right)^4\right] \\ \ddot{x}_{poly5}(t) &= \frac{5A}{T^2}\left[1 - 6\left(\frac{t-t_0}{T}\right)^2 + 4\left(\frac{t-t_0}{T}\right)^3\right] \\ \dddot{x}_{poly5}(t) &= \frac{60A}{T^3}\left[-\left(\frac{t-t_0}{T}\right) + \left(\frac{t-t_0}{T}\right)^2\right] \end{cases} \tag{A.3}$$

– Accélération nulle au début et à la fin du mouvement : $\Gamma_0 = 0$

$$\begin{cases} x_{poly5}(t) &= x_0 + A\left[10\left(\frac{t-t_0}{T}\right)^3 - 15\left(\frac{t-t_0}{T}\right)^4 + 6\left(\frac{t-t_0}{T}\right)^5\right] \\ \dot{x}_{poly5}(t) &= \frac{30A}{T}\left[\left(\frac{t-t_0}{T}\right)^2 - 2\left(\frac{t-t_0}{T}\right)^3 + \left(\frac{t-t_0}{T}\right)^4\right] \\ \ddot{x}_{poly5}(t) &= \frac{60A}{T^2}\left[\left(\frac{t-t_0}{T}\right) - 3\left(\frac{t-t_0}{T}\right)^2 + 2\left(\frac{t-t_0}{T}\right)^3\right] \\ \dddot{x}_{poly5}(t) &= \frac{60A}{T^3}\left[1 - 6\left(\frac{t-t_0}{T}\right) + 6\left(\frac{t-t_0}{T}\right)^2\right] \end{cases} \tag{A.4}$$

A.1.3 Coût d'impulsion

Le coût d'impulsion égal au maximum de la vitesse est donné par :

$$\begin{cases} V_{poly5} &= \frac{15}{8}\frac{A}{T} - \frac{\Gamma_0 T}{16} \\ T_{V_{poly5}} &= t_0 + \frac{T}{2} \end{cases} \tag{A.5}$$

On a les cas particuliers suivants :

– Minimisation du coût de secousse : $\Gamma_0 = \frac{5A}{T^2}$

$$\begin{cases} V_{poly5} &= \frac{25}{16}\frac{A}{T} \\ T_{V_{poly5}} &= t_0 + \frac{T}{2} \end{cases} \tag{A.6}$$

– Accélération nulle au début et à la fin du mouvement : $\Gamma_0 = 0$

$$\begin{cases} V_{poly5} &= \frac{15}{8}\frac{A}{T} \\ T_{V_{poly5}} &= t_0 + \frac{T}{2} \end{cases} \tag{A.7}$$

A.1.4 Coût de force

Le coût de force égal au maximum de l'accélération est donné par :
Pour $\Gamma_0 < \frac{5A}{T^2}$

$$\begin{cases} \Gamma_{poly5} &= \left|\ddot{x}_{poly5}(t_0 + \frac{T}{2} \pm \frac{T\sqrt{\Delta}}{2(30A-5\Gamma_0 T^2)})\right| \\ T_{\Gamma_{poly5}} &= t_0 + \frac{T}{2} \pm \frac{T\sqrt{\Delta}}{2(30A-5\Gamma_0 T^2)} \end{cases} \tag{A.8}$$

avec $\Delta = 5T^4(\Gamma_0 - \frac{6A}{T^2})(\Gamma_0 - \frac{10A}{T^2})$.

Pour $\Gamma_0 \geq \frac{5A}{T^2}$

$$\begin{cases} \Gamma_{poly5} &= \Gamma_0 \\ T_{\Gamma_{poly5}} &= t_0 \end{cases} \tag{A.9}$$

On a les cas particuliers suivants :

– Minimisation du coût de secousse : $\Gamma_0 = \frac{5A}{T^2}$

$$\begin{cases} \Gamma_{poly5} &= \frac{5A}{T^2} \\ T_{\Gamma_{poly5}} &= t_0 \end{cases} \tag{A.10}$$

– Accélération nulle au début et à la fin du mouvement : $\Gamma_0 = 0$

$$\begin{cases} \Gamma_{poly5} &= \frac{10}{\sqrt{3}}\frac{A}{T^2} \\ T_{\Gamma_{poly5}} &= t_0 + \frac{3\pm\sqrt{3}}{6}T \end{cases} \tag{A.11}$$

A.1.5 Coût de secousse

Le coût de secousse égal à la puissance de la secousse est donné par :

$$J_{poly5} = \frac{12}{T^5}\left[30A^2 - 10A\Gamma_0 T^2 + \Gamma_0^2 T^4\right] \tag{A.12}$$

On a les cas particuliers suivants :

– Minimisation du coût de secousse : $\Gamma_0 = \frac{5A}{T^2}$

$$J_{poly5} = 60\frac{A^2}{T^5} \tag{A.13}$$

– Accélération nulle au début et à la fin du mouvement : $\Gamma_0 = 0$

$$J_{poly5} = 360\frac{A^2}{T^5} \tag{A.14}$$

A.2 Trajectoire polynomiale d'ordre 2 - Minimisation de la durée

A.2.1 Expression de la trajectoire

Soit un mouvement d'une dimension d'amplitude A, de durée T et démarrant en position x_0 à l'instant t_0. La trajectoire minimisant la durée en maximisant l'accélération possède idéalement une accélération positive constante sur la première moitié du mouvement, et une accélération négative constante (décélération) sur la deuxième moitié du mouvement. La vitesse est donc une rampe croissante (resp. décroissante) sur la première (resp. deuxième) moitié du mouvement. La trajectoire est donc une concaténation de deux paraboles, soit un polynôme d'ordre 2.

Seulement, ce modèle entraîne des secousse nulles, ne traduisant pas le changement brusque d'accélération au milieu du mouvement. On introduit donc T_t le temps de transition nécessaire pour passer d'une accélération constante positive à une accélération constante négative. On considère que sur le segment $[t_0 + \frac{T-T_t}{2}, t_0 + \frac{T+T_t}{2}]$ l'accélération évolue de A_{max} à $-A_{max}$ linéairement. On choisit donc un ordre 1 pour l'accélération, 2 pour la vitesse et 3 pour la trajectoire sur ce segment. On en déduit :

$$
\left\{
\begin{array}{l}
\left\{
\begin{array}{lcll}
x_{poly2}(t) & = & A_0 + A_1(t-t_0) + A_2(t-t_0)^2 & \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
x_{poly2}(t) & = & B_0 + B_1(t-t_0) + B_2(t-t_0)^2 + B_3(t-t_0)^3 & \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
x_{poly2}(t) & = & C_0 + C_1(t-t_0) + C_2(t-t_0)^2 & \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lcll}
\dot{x}_{poly2}(t) & = & A_1 + 2A_2(t-t_0) & \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
\dot{x}_{poly2}(t) & = & B_1 + 2B_2(t-t_0) + 3B_3(t-t_0)^2 & \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
\dot{x}_{poly2}(t) & = & C_1 + 2C_2(t-t_0) & \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lcll}
\ddot{x}_{poly2}(t) & = & 2A_2 & \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
\ddot{x}_{poly2}(t) & = & 2B_2 + 6B_3(t-t_0) & \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
\ddot{x}_{poly2}(t) & = & 2C_2 & \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lcll}
\dddot{x}_{poly2}(t) & = & 0 & \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
\dddot{x}_{poly2}(t) & = & 6B_3 & \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
\dddot{x}_{poly2}(t) & = & 0 & \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right.
\end{array}
\right.
\tag{A.15}
$$

Avec les conditions initiales suivantes :

$$
\left\{
\begin{array}{lcll}
x_{poly2}(t_0) & = & x_0 & \text{Position initiale} \\
x_{poly2}(t_0+T) & = & x_0 + A & \text{Position finale} \\
\dot{x}_{poly2}(t_0) & = & 0 & \text{Vitesse initiale} \\
\dot{x}_{poly2}(t_0+T) & = & 0 & \text{Vitesse finale} \\
\ddot{x}_{poly2}(t_0) & = & -\ddot{x}_{poly2}(t_0+T) & \text{Mouvement symétrique}
\end{array}
\right.
$$

Et les continuités à $x_{poly2}(t_0 + \frac{T-T_t}{2})$, $x_{poly2}(t_0 + \frac{T+T_t}{2})$, $\dot{x}_{poly2}(t_0 + \frac{T-T_t}{2})$, $\ddot{x}_{poly2}(t_0 + \frac{T-T_t}{2})$, $\ddot{x}_{poly2}(t_0 + \frac{T+T_t}{2})$.

Après résolution du système on obtient :

$$
\left\{
\begin{array}{l}
\left\{
\begin{array}{lll}
x_{poly2}(t) &=& x_0 + \frac{6A}{3T^2-T_t^2}(t-t_0)^2 \qquad \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
x_{poly2}(t) &=& x_0 + \frac{6A}{3T^2-T_t^2}\left[\frac{(T-T_t)^3}{12T_t} - \frac{(T-T_t)^2}{2T_t}(t-t_0) + \frac{T}{T_t}(t-t_0)^2 - \frac{2}{3T_t}(t-t_0)^3\right] \qquad \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
x_{poly2}(t) &=& x_0 + \frac{6A}{3T^2-T_t^2}\left[-\frac{3T^2+T_t^2}{6} + 2T(t-t_0) - (t-t_0)^2\right] \qquad \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lll}
\dot{x}_{poly2}(t) &=& \frac{6A}{3T^2-T_t^2}2(t-t_0) \qquad \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
\dot{x}_{poly2}(t) &=& \frac{6A}{3T^2-T_t^2}\left[-\frac{(T-T_t)^2}{2T_t} + \frac{2T(t-t0)}{T_t} - \frac{2(t-t0)^2}{T_t}\right] \qquad \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
\dot{x}_{poly2}(t) &=& \frac{6A}{3T^2-T_t^2}2\left[T-(t-t_0)\right] \qquad \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lll}
\ddot{x}_{poly2}(t) &=& \frac{12A}{3T^2-T_t^2} \qquad \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
\ddot{x}_{poly2}(t) &=& \frac{12A}{3T^2-T_t^2}\left[\frac{T-2(t-t_0)}{T_t}\right] \qquad \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
\ddot{x}_{poly2}(t) &=& -\frac{12A}{3T^2-T_t^2} \qquad \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lll}
\dddot{x}_{poly2}(t) &=& 0 \qquad \text{si } t \leq t_0 + \frac{T-T_t}{2} \\
\dddot{x}_{poly2}(t) &=& \frac{6A}{3T^2-T_t^2}\frac{-4}{T_t} \qquad \text{si } t_0 + \frac{T-T_t}{2} < t \leq t_0 + \frac{T+T_t}{2} \\
\dddot{x}_{poly2}(t) &=& 0 \qquad \text{si } t > t_0 + \frac{T+T_t}{2}
\end{array}
\right.
\end{array}
\right. \tag{A.16}
$$

A.2.2 Cas particulier :

En prenant $T_t \to 0$ dans A.16 on obtient :.

$$
\left\{
\begin{array}{l}
\left\{
\begin{array}{lll}
x_{poly2}(t) &=& x_0 + \frac{2A}{T^2}(t-t_0)^2 \qquad \text{si } t \leq t_0 + \frac{T}{2} \\
x_{poly2}(t) &=& x_0 + \frac{2A}{T^2}\left[-\frac{T^2}{2} + 2T(t-t_0) - (t-t_0)^2\right] \qquad \text{si } t > t_0 + \frac{T}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lll}
\dot{x}_{poly2}(t) &=& \frac{4A}{T^2}(t-t_0) \qquad \text{si } t \leq t_0 + \frac{T}{2} \\
\dot{x}_{poly2}(t) &=& \frac{4A}{T^2}\left[T-(t-t_0)\right] \qquad \text{si } t > t_0 + \frac{T}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lll}
\ddot{x}_{poly2}(t) &=& \frac{4A}{T^2} \qquad \text{si } t \leq t_0 + \frac{T}{2} \\
\ddot{x}_{poly2}(t) &=& -\frac{4A}{T^2} \qquad \text{si } t > t_0 + \frac{T}{2}
\end{array}
\right. \\
\left\{
\begin{array}{lll}
\dddot{x}_{poly2}(t) &=& 0 \qquad \text{si } t \leq t_0 + \frac{T}{2} \\
\dddot{x}_{poly2}(t) &=& 0 \qquad \text{si } t > t_0 + \frac{T}{2}
\end{array}
\right.
\end{array}
\right. \tag{A.17}
$$

A.2.3 Coût d'impulsion

Le coût d'impulsion égal au maximum de la vitesse est donné par :

$$
\left\{
\begin{array}{lll}
V_{poly2} &=& \frac{3A(2T-T_t)}{3T^2-T_t^2} \\
T_{V_{poly2}} &=& t_0 + \frac{T}{2}
\end{array}
\right. \tag{A.18}
$$

En prenant $T_t \to 0$

$$\begin{cases} V_{poly2} &= \frac{2A}{T} \\ T_{V_{poly2}} &= t_0 + \frac{T}{2} \end{cases} \tag{A.19}$$

A.2.4 Coût de force

Le coût de force égal au maximum de l'accélération est donné par :

$$\Gamma_{poly2} = \frac{12A}{3T^2 - T_t^2} \tag{A.20}$$

En prenant $T_t \to 0$

$$\Gamma_{poly2} = 4\frac{A}{T^2} \tag{A.21}$$

A.2.5 Coût de secousse

Le coût de secousse égal à la puissance de la secousse est donné par :

$$J_{poly2} = \frac{288A^2}{T_t(3T^2 - T_t^2)^2} \tag{A.22}$$

En prenant $T_t \to 0$

$$J_{poly2} = \infty \tag{A.23}$$

A.3 Trajectoire polynomiale d'ordre 1 - Minimisation de la vitesse

A.3.1 Expression de la trajectoire

Soit un mouvement d'une dimension d'amplitude A, de durée T et démarrant en position x_0 à l'instant t_0. La trajectoire minimisant la vitesse possède idéalement un profil de vitesse constant égal à la vitesse moyenne de la trajectoire $V_{moy} = A/T$. La trajectoire est donc une rampe, soit un polynôme d'ordre 1.

$$\begin{cases} x_{poly1}(t) &= \frac{A}{T}t \\ \dot{x}_{poly1}(t) &= \frac{A}{T} \\ \ddot{x}_{poly1}(t) &= 0 \\ \dddot{x}_{poly1}(t) &= 0 \end{cases} \tag{A.24}$$

Seulement, l'expression donnée en A.24 entraîne des accélération et secousse nulles, ne traduisant pas les transitions de début et fin de mouvement. En effet, on considère la vitesse nulle à $t = t_0$ et $t = t_0 + T$. On introduit donc $\frac{T_t}{2}$ le temps de transition nécessaire pour atteindre une vitesse constante. On considère que sur le segment $[t_0, t_0 + \frac{T_t}{2}]$ (resp. $[t_0 + T - \frac{T_t}{2}, t_0 + T]$) la vitesse évolue de 0 à V_{max} (resp. de V_{max} à 0) et l'accélération évolue de A_{max} à 0 (resp. 0 à $-A_{max}$). On choisit donc un ordre 1 pour l'accélération et 2 pour la vitesse sur ces segments. On en déduit :

$$\begin{cases} \begin{cases} x_{poly1}(t) &= A_0 + A_1(t-t_0) + A_2(t-t_0)^2 + A_3(t-t_0)^3 & \text{si } t \leq t_0 + \frac{T_t}{2} \\ x_{poly1}(t) &= B_0 + B_1(t-t_0) & \text{si } t_0 + \frac{T_t}{2} < t \leq t_0 + T - \frac{T_t}{2} \\ x_{poly1}(t) &= C_0 + C_1(t-t_0) + C_2(t-t_0)^2 + C_3(t-t_0)^3 & \text{si } t > t_0 + T - \frac{T_t}{2} \end{cases} \\ \begin{cases} \dot{x}_{poly1}(t) &= A_1 + 2A_2(t-t_0) + 3A_3(t-t_0)^2 & \text{si } t \leq t_0 + \frac{T_t}{2} \\ \dot{x}_{poly1}(t) &= B_1 & \text{si } t_0 + \frac{T_t}{2} < t \leq t_0 + T - \frac{T_t}{2} \\ \dot{x}_{poly1}(t) &= C_1 + 2C_2(t-t_0) + 3C_3(t-t_0)^2 & \text{si } t > t_0 + T - \frac{T_t}{2} \end{cases} \\ \begin{cases} \ddot{x}_{poly1}(t) &= 2A_2 + 6A_3(t-t_0) & \text{si } t \leq t_0 + \frac{T_t}{2} \\ \ddot{x}_{poly1}(t) &= 0 & \text{si } t_0 + \frac{T_t}{2} < t \leq t_0 + T - \frac{T_t}{2} \\ \ddot{x}_{poly1}(t) &= 2C_2 + 6C_3(t-t_0) & \text{si } t > t_0 + T - \frac{T_t}{2} \end{cases} \\ \begin{cases} \dddot{x}_{poly1}(t) &= 6A_3 & \text{si } t \leq t_0 + \frac{T_t}{2} \\ \dddot{x}_{poly1}(t) &= 0 & \text{si } t_0 + \frac{T_t}{2} < t \leq t_0 + T - \frac{T_t}{2} \\ \dddot{x}_{poly1}(t) &= 6C_3 & \text{si } t > t_0 + T - \frac{T_t}{2} \end{cases} \end{cases} \tag{A.25}$$

Avec les conditions initiales suivantes :

$$\begin{cases} x_{poly1}(t_0) &= x_0 & \text{Position initiale} \\ x_{poly1}(t_0+T) &= x_0+A & \text{Position finale} \\ \dot{x}_{poly1}(t_0) &= 0 & \text{Vitesse initiale} \\ \dot{x}_{poly1}(t_0+T) &= 0 & \text{Vitesse finale} \end{cases}$$

Ainsi que les continuités à respecter à $x_{poly1}(t_0+\frac{T_t}{2})$, $x_{poly1}(t_0+T-\frac{T_t}{2})$, $\dot{x}_{poly1}(t_0+\frac{T_t}{2})$, $\dot{x}_{poly1}(t_0+T-\frac{T_t}{2})$, $\ddot{x}_{poly1}(t_0+\frac{T_t}{2})$, $\ddot{x}_{poly1}(t_0+T-\frac{T_t}{2})$.

Après résolution du système on obtient :

$$\begin{cases} \begin{cases} x_{poly1}(t) &= x_0+\frac{6A}{T_t(3T-T_t)}\left[(t-t_0)^2-\frac{2}{3T_t}(t-t_0)^3\right] & \text{si } t\le t_0+\frac{T_t}{2} \\ x_{poly1}(t) &= x_0+\frac{6A}{T_t(3T-T_t)}\left[\frac{T_t}{2}(t-t_0)-\frac{T_t^2}{12}\right] & \text{si } t_0+\frac{T_t}{2}<t\le t_0+T-\frac{T_t}{2} \\ x_{poly1}(t) &= x_0+\frac{6A}{T_t(3T-T_t)}\left[\frac{(T-T_t)(3T^2+(T-T_t)^2)}{6T_t}-\frac{2T(T-T_t)(t-t_0)}{T_t}+\frac{(2T-T_t)(t-t_0)^2}{T_t}-\frac{2(t-t_0)^3}{3T_t}\right] & \text{si } t>t_0+T-\frac{T_t}{2} \end{cases} \\ \begin{cases} \dot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\left[2t-\frac{2(t-t0)^2}{T_t}\right] & \text{si } t\le t_0+\frac{T_t}{2} \\ \dot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\frac{T_t}{2} & \text{si } t_0+\frac{T_t}{2}<t\le t_0+T-\frac{T_t}{2} \\ \dot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\left[\frac{-2T(T-T_t)}{T_t}+\frac{2(2T-T_t)(t-t_0)}{T_t}-\frac{2(t-t_0)^2}{T_t}\right] & \text{si } t>t_0+T-\frac{T_t}{2} \end{cases} \\ \begin{cases} \ddot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\left[2-\frac{4(t-t0)}{T_t}\right] & \text{si } t\le t_0+\frac{T_t}{2} \\ \ddot{x}_{poly1}(t) &= 0 & \text{si } t_0+\frac{T_t}{2}<t\le t_0+T-\frac{T_t}{2} \\ \ddot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\left[\frac{2(2T-T_t)}{T_t}-\frac{4(t-t0)}{T_t}\right] & \text{si } t>t_0+T-\frac{T_t}{2} \end{cases} \\ \begin{cases} \dddot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\frac{-4}{T_t} & \text{si } t\le t_0+\frac{T_t}{2} \\ \dddot{x}_{poly1}(t) &= 0 & \text{si } t_0+\frac{T_t}{2}<t\le t_0+T-\frac{T_t}{2} \\ \dddot{x}_{poly1}(t) &= \frac{6A}{T_t(3T-T_t)}\frac{-4}{T_t} & \text{si } t>t_0+T-\frac{T_t}{2} \end{cases} \end{cases} \tag{A.26}$$

En prenant $T_t \to 0$ dans A.26 on retrouve bien la solution simple donnée en A.24.

A.3.2 Coût d'impulsion

Le coût d'impulsion égal au maximum de la vitesse est donné par :

$$V_{poly1} = \frac{3A}{3T-T_t} \tag{A.27}$$

En prenant $T_t \to 0$

$$V_{poly1} = \frac{A}{T} \tag{A.28}$$

A.3.3 Coût de force

Le coût de force égal au maximum de l'accélération est donné par :

$$\Gamma_{poly1} = \frac{12A}{T_t(3T-T_t)} \tag{A.29}$$

En prenant $T_t \to 0$

$$\Gamma_{poly1} = \infty \tag{A.30}$$

A.3.4 Coût de secousse

Le coût de secousse égal à la puissance de la secousse est donné par :

$$J_{poly1} = \frac{288A^2}{T_t^3(3T - T_t)^2} \tag{A.31}$$

En prenant $T_t \to 0$

$$J_{poly1} = \infty \tag{A.32}$$

Annexe B

Liste des publications scientifiques

B.1 Communications

Revues internationales

O. Perrotin, C. d'Alessandro
Target Acquisition vs. Expressive Motion : Dynamic Pitch Warping Correction for Digital Musical Instruments [Pd]
ACM TOCHI (accepté)

L. Feugère, C. d'Alessandro, B. Doval, O. Perrotin
Cantor Digitalis : Chironomic Parametric Synthesis of Singing [FdDP]
Journal of Transactions on Audio, Language and Signal Processing (en soumission)

C. d'Alessandro, L. Feugère, S. Le Beux, O. Perrotin, A. Rilliard (ordre alphabétique)
Drawing melodies : Evaluation of Chironomic Singing Synthesis [dFLB+14]
J. Acoust. Soc. Am. 135, 3601 (2014)

Conférences internationales avec actes

O. Perrotin, C. d'Alessandro
Visualizing Gestures in the Control of a Digital Musical Instrument [Pd14]
Proceedings of the 2014 International Conference on New Interfaces for Musical Expression (NIME14), Goldsmiths, University of London, UK, 30 Juin - 4 Juillet, 2014, pp. 605-608.

O. Perrotin, C. d'Alessandro
Adaptive mapping for improved pitch accuracy on touch user interfaces [Pd13]
Proceedings of the 2013 International Conference on New Interfaces for Musical Expression (NIME13), KAIST, Daejeon + Seoul, Korea Republic, 27-30 Mai, 2013, pp. 186-189.

Conférence francophone avec actes

O. Perrotin, C. d'Alessandro
Quel ajustement de hauteur mélodique pour les instruments de musique numériques ? [Pd15]
Journées d'Informatique Musicale (JIM), Faculté de musique, Université de Montréal, QC, Canada, 5-7 Mai, 2015.

Conférence francophone sans acte

O. Perrotin
Chironomie et interfaces pour les instruments de musique virtuels
Journées Jeunes Chercheurs en Audition, Acoustique musicale, et Signal audio (JJCAAS), Marseille, France, 5-7 Décembre, 2012

B.2 Diffusion logicielle

C. d'Alessandro, B. Doval, L. Feugère, S. Le Beux, O. Perrotin (ordre alphabétique)[1]
Logiciel *Cantor Digitalis* sous licence CeCILL
Logiciel, 21 octobre, 2014.

O. Perrotin, L. Feugère
Cantor Digitalis v1.1 - Manuel d'utilisation
Cantor Digitalis v1.1 - Documentation technique
Documentation, 21 octobre, 2014.

B.3 Prix

C. d'Alessandro, B. Doval, L. Feugère, S. Le Beux, O. Perrotin (ordre alphabétique)[2]
Premier prix de la Compétition Margaret Guthman d'Instruments de Musique
Compétition Margaret Guthman d'Instruments de Musique, Georgia Tech, Atlanta, GA, USA, 19-20 Février, 2015

C. d'Alessandro, B. Doval, L. Feugère, S. Le Beux, O. Perrotin (ordre alphabétique)
Finalistes du concours international du Logiciel Musical (Lomus) de l'AFIM
Journées d'Informatique Musicale, Bourges, France, 23 Mai, 2014

B.4 Concerts art-science

C. d'Alessandro, B. Doval, L. Feugère, O. Perrotin
Soirée *Surchauffe*
Music Tech Metz, Arsenal, Metz, France, 12 Mai, 2015

C. d'Alessandro, B. Doval, L. Feugère, O. Perrotin
Compétition Margaret Guthman d'instruments de musique
Georgia Tech, Atlanta, GA, Etats-Unis, 19-20 Février, 2015

O. Perrotin, L. Feugère, C. d'Alessandro, B. Doval, A. Braffort, S. Delalez, S. Jacquin, Bolivier
Chorus Digitalis - saison 3.2
Futur en Seine, PROTO 204, Orsay, France, 19 Juin, 2014

O. Perrotin, L. Feugère, C. d'Alessandro, B. Doval, A. Braffort, H. Maynard
Chorus Digitalis - saison 3.1
Festival CuriositAS, Sciences ACO, Orsay, France, 7 Octobre, 2013

L. Feugère, O. Perrotin, C. d'Alessandro, B. Doval
Chorus Digitalis - saison 3.0
JDEV 2013, Ecole polytechnique, Palaiseau, France, 5 Septembre, 2013

N. d'Alessandro, C. d'Alessandro, L. Feugère, M. Astrinaki, J. Wang, O. Perrotin, A. Pon, B. Doval
Vox Tactum Meets Chorus Digitalis : Seven Years of Singing Surfaces
13th International Conference on New Interfaces for Musical Expression (NIME13), Daejeon + Seoul, Korea Republic, 27-30 Mai, 2013

1. https://cantordigitalis.limsi.fr (vérifié le 22 octobre 2015)
2. http://guthman.gatech.edu/2015winners (vérifié le 22 octobre 2015)

Liste des tableaux

Table des figures

Bibliographie

[ACDCH12] Jonathan Aceituno, Julien Castet, Myriam Desainte-Catherine et Martin Hachet : Improvised interfaces for real-time musical applications. *In Proceedings of the International Conference on Tangible and Embedded Interaction (TEI)*, pages 197–200, Kingston, Ontario, Canada, February 19-22 2012.

[ACK05] Daniel Arfib, Jean-Michel Couturier et Loïc Kessous : Expressiveness and digital musical instrument design. *Journal of New Music Research*, 34(1):125–136, 2005.

[ACKV02] Daniel Arfib, Jean-Michel Couturier, Loïc Kessous et Vincent Verfaille : Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organized Sound*, 7(2):127–144, 2002.

[AdD12] Maria Astrinaki, Nicolas d'Alessandro et Thierry Dutoit : Mage - a platform for tangible speech synthesis. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '12, Ann Arbor, MI, USA, May 21-23 2012.

[ADR15] Luc Ardaillon, Gilles Degottex et Axel Roebel : A multi-layer f0 model for singing voice synthesis using a b-spline representation with intuitive controls. *In Proceedings of Interspeech*, Dresden, Germany, September 6-10 2015. ISCA.

[AGZ10] Georg Apitz, François Guimbretiere et Shumin Zhai : Foundations for designing and evaluating user interfaces based on the crossing paradigm. *ACM Transactions on Computer-Human Interactions (TOCHI)*, 17(2), 2010.

[AH71] B. S. Atal et Suzanne L. Hanauer : Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, 50(2B):637–655, 1971.

[AMH95] Motoyuki Akamatsu, I. Scott MacKenzie et Thierry Hasbrouc : A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38(4):816–827, 1995. PMID : 7729406.

[Ard13] Luc Ardaillon : Synthèse du chant. Mémoire de D.E.A., UPMC (Université Pierre et Marie Curie), 2013.

[AZ97] Johnny Accot et Shumin Zhai : Beyond fitts' law : Models for trajectory-based hci tasks. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 295–302, Atlanta, Georgia, USA, 1997. ACM.

[AZ10] Tue Haste Andersen et Shumin Zhai : "writing with music" : Exploring the use of auditory feedback in gesture interfaces. *ACM Trans. Appl. Percept.*, 7(3):17 :1–17 :24, juin 2010.

[Bal04] Ravin BALAKRISHNAN : "beating" fitts' law : Virtual enhancements for pointing facilitation. *Int. J. Hum.-Comput. Stud.*, 61(6):857–874, décembre 2004.

[BCL+01] Jordi BONADA, Òscar. CELMA, Àlex LOSCOS, Jaume ORTOLÀ et Xavier SERRA : Singing voice synthesis combining excitation plus resonance and sinusoidal plus residual models. *In Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba, September 17-22 2001. Singing Voice.

[Ber95] Gunilla BERNDTSSON : The kth rule system for singing synthesis. Quarterly Progress and Status Report 1, Royal Institute of Technologies - Dept. for Speech, Music and Hearing, 1995.

[BGBL04] Renaud BLANCH, Yves GUIARD et Michel BEAUDOUIN-LAFON : Semantic pointing : Improving target acquisition with control-display ratio adaptation. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 519–526, Vienna, Austria, 2004. ACM.

[BJ14] Michael V. BUTERA et Jack JENKINS : Ergonomic electronic musical instrument with pseudo-strings, octobre 2 2014. US Patent App. 14/306,818.

[BK07] Tony BERGSTROM et Karrie KARAHALIOS : Conversation clock : Visualizing audio patterns in co-located groups. *In Hawaii International Conference on Systems Science (HICSS)*, page 78, Hawaii, USA, January 3-6 2007.

[BL03] Jordi BONADA et Àlex LOSCOS : Sample-based singing voice synthesizer by spectral concatenation. *In Proceedings of the Stockholm Music Acoustics Conference (SMAC)*, Stockholm, Sweden, August 6-9 2003.

[BL04] Michel BEAUDOUIN-LAFON : Designing interaction, not interfaces. *In Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 15–22, Gallipoli, Italy, May 25-28 2004. ACM.

[BMSH13] Florent BERTHAUT, Mark T. MARSHALL, Sriram SUBRAMANIAN et Martin HACHET : Rouages : Revealing the mechanisms of digital musical instruments to the audience. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '13, Daejeon, South Korea, May 27-30 2013.

[Cad88] Claude CADOZ : Instrumental Gesture and Musical Composition. *In Proceedings of the International Computer Music Conference (ICMC)*, pages 1–12, Cologne, Germany, février 1988. http ://computermusic.org/.

[Cad94] Claude CADOZ : Le geste canal de communication homme-machine. la communication 'instrumentale'. *Science Informatiques, numéro spécial : Interface homme-machine*, 13(1):31–61, 1994.

[CB05] Andy COCKBURN et Stephen BREWSTER : Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48(9):1129–1150, July 2005.

[CBBL07] Olivier CHAPUIS, Renaud BLANCH et Michel BEAUDOUIN-LAFON : Fitts' law in the wild : A field study of aimed movements. Rapport technique, LRI, Univ. Paris-Sud, 2007.

[CBS10] Baptiste CARAMIAUX, Frédéric BEVILACQUA et Norbert SCHNELL : Towards a gesture-sound cross-modal analysis. *In* Stefan KOPP et Ipke WACHSMUTH, éditeurs : *Gesture in Embodied Communication and Human-Computer Interaction*, volume 5934 de *Lecture Notes in Computer Science*, pages 158–170. Springer Berlin Heidelberg, 2010.

[CF03] Andy COCKBURN et Andrew FIRTH : Improving the acquisition of small targets. *In* Eamonn O'NEILL, Philippe PALANQUE et Peter JOHNSON, éditeurs : *People and Computers XVII — Designing for Society*, pages 181–196. Springer London, 2003.

[Cho73] John M. CHOWNING : The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7):526–534, 1973.

[CJDS10] Eakachai CHAROENCHAIMONKON, Paul JANECEK, Mattew N. DAILEY et Atiwong SUCHATO : A comparison of audio and tactile displays for non-visual target selection tasks. *In Proceedings of the International Conference on User Science and Engineering (i-USEr)*, pages 238–243, Shah Alam, Malaysia, December 13-15 2010.

[CLP09] Olivier CHAPUIS, Jean-Baptiste LABRUNE et Emmanuel PIETRIGA : Dynaspot : Speed-dependent area cursor. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1391–1400, Boston, MA, USA, 2009. ACM.

[CMR91] Stuart K. CARD, Jock D. MACKINLAY et George G. ROBERTSON : A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst.*, 9(2):99–122, avril 1991.

[Coo90] Perry R. COOK : *Identification of Control Paramters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing.* Thèse de doctorat, Center for Computer Research in Music and Acoustics (CCRMA), 1990.

[Coo93] Perry R. COOK : Spasm : A real-time vocal tract physical model controller ; and singer the companion software synthesis system. *Computer Music Journal*, 17(1):30–44, 1993.

[Coo96] Perry R. COOK : Singing voice synthesis : History, current work, and future directions. *Computer Music Journal*, 20(3):38–46, 1996.

[Coo98] Perry R. COOK : Toward the perfect audio morph ? singing voice synthesis and processing. *In Proceedings of the Conference on Digital Audio Effects (DAFX)*, 1998.

[Coo01] Perry R. COOK : Principles for designing computer music controllers. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '01, pages 1–4, Seattle, Washington, USA, April 1-2 2001.

[Coo05] Perry R. COOK : Real-time performance controllers for synthesized singing. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '05, pages 236–237, Vancouver, Canada, May 26-28 2005.

[Coo09] Perry R. COOK : Re-designing principles for computer music controllers : a case study os squeezevox maggie. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '09, pages 218–221, Pittsburgh, PA, United States, June 3-6 2009.

[Cou02] Jean-Michel COUTURIER : A scanned synthesis virtual instrument. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '02, Dublin, Ireland, May 24-26 2002.

[Cra13] Michael J. CRAWLEY : *The R Book*. Wiley, 2013.

[d'A01] Christophe D'ALESSANDRO : 33 ans de syntèse de la parole à partir du texte : une promenade sonore (1968-2001). *Traitement Automatique des Langues*, 42(1):297–321, 2001.

[d'A06] Christophe D'ALESSANDRO : Voice source parameters and prosodic analysis. *Method in Empirical Prosody Research*, pages 63–87, 2006.

[DBGP07] Simone DALLA BELLA, Jean-François GIGUÈRE et Isabelle PERETZ : Singing proficiency in the general population. *Acoustical Society of America*, pages 1182–1189, November 22 2007.

[dC91] Christophe D'ALESSANDRO et Michèle CASTELLENGO : Etude, par la synthèse, de la perception du vibrato vocal dans les transitions de notes. *Bulletin d'Audiophonologie*, 7(5):551–564, 1991.

[dC94] Christophe D'ALESSANDRO et Michèle CASTELLENGO : The pitch of short-duration vibrato tones. *Acoustical Society of America*, 95(3):1617–1630, November 1994.

[DCM99a] Myriam DESAINTE-CATHERINE et Sylvain MARCHAND : Structured additive synthesis : Towards a model of sound timbre and electroacoustic music forms. *In Proceedings of the International Computer Music Conference (ICMC)*, Beijing, China, October 1999.

[DCM99b] Myriam DESAINTE-CATHERINE et Sylvain MARCHAND : Vers un modèle pour unifier musique et son dans une composition multiéchelle. *In Actes des Journées d'Informatique Musicale (JIM)*, Paris, France, May 17-19 1999.

[dD09a] Nicolas D'ALESSANDRO et Thierry DUTOIT : Advanced techniques for vertical tablet playing a overview of two years of practicing the handsketch 1.x. *In* Noel ZAHLER, Roger B. DANNENBERG et Tom SULLIVAN, éditeurs : *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '09, pages 173–174, Pittsburgh, PA, United States, June 3-6 2009.

[dD09b] Nicolas D'ALESSANDRO et Thierry DUTOIT : Handsketch : Bi-manual control of voice quality dimensions and long term practice issues. Quarterly Progress and Status Report 2, Numediart Research Program, June 2009.

[dDD08] Nicolas D'ALESSANDRO, Thomas DRUGMAN et Thomas DUBUISSON : Trans-voice table. Rapport technique 1, Numediart Research Program, 2008.

[ddF+13] Nicolas D'ALESSANDRO, Christophe D'ALESSANDRO, Lionel FEUGÈRE, Maria ASTRINAKI, Johnty WANG et Olivier PERROTIN : Vox tactum meets chorus digitalis : Seven years of singing surfaces. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '13, Daejeon, South Korea, May 27-30 2013.

[DdH03] Boris DOVAL, Christophe D'ALESSANDRO et Nathalie HENRICH : The voice source as a causal/anticausal linear filter. *In VOQUAL'03*, Geneva, August 27-29 2003.

[DdH06] Boris DOVAL, Christophe D'ALESSANDRO et Nathalie HENRICH : The spectrum of glottal flow models. *Acta Acustica*, 92(6):1026–1046, 2006.

[ddLB+05] Christophe D'ALESSANDRO, Nicolas D'ALESSANDRO, Sylvain LE BEUX, Juraj SIMKO, Feride CETIN et Hannes PIRKER : The speech conductor : Gestu-

ral control of speech synthesis. Rapport technique, eNTERFACE, July 18 – August 12 2005.

[dDLB+06] Nicolas d'Alessandro, Boris Doval, Sylvain Le Beux, Pascale Woodruff et Yohann Fabre : Ramcess : Realtime and accurate musical control of expression in singing synthesis. Rapport technique, eNTERFACE, July 17–August 11 2006.

[ddLB+08] Christophe d'Alessandro, Nicolas d'Alessandro, Sylvain Le Beux, Juraj Simko, Feride Cetin et Hannes Pirker : The speech conductor : Gestural control of speech synthesis. Rapport technique, eNTERFACE, August 4-29 2008.

[ddLBD06a] Christophe d'Alessandro, Nicolas d'Alessandro, Sylvain Le Beux et Boris Doval : Comparing time domain and spectral domain voice source models for gesture controlled vocal instruments. *In Proceedings of the International Conference on Voice Physiology and Biomechanics*, 2006.

[ddLBD06b] Nicolas d'Alessandro, Christophe d'Alessandro, Sylvain Le Beux et Boris Doval : Real-time calm synthesizer new approaches in hands-controlled voice synthesis. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '06, Paris, France, June 4-8 2006. IRCAM ; Centre Pompidou.

[DdR84] F. Dechelle, Christophe d'Alessandro et Xavier Rodet : Synthèse en temps réel sur microprocesseur tms 320. *In Proceedings of the International Computer Music Conference (ICMC)*, Paris, France, 1984.

[dFLB+14] Christophe d'Alessandro, Lionel Feugère, Sylvain Le Beux, Olivier Perrotin et Albert Rilliard : Drawing melodies : Evaluation of chironomic singing synthesis. *Acoustical Society of America*, 135(6):3601–3612, June 2014.

[DGR94] Philippe Depalle, G. Garcia et Xavier Rodet : A virtual castrato (!?). *In Proceedings of the International Computer Music Conference (ICMC)*, pages 357–360, Aarhus, Denmark, 1994.

[DHAT00] Peter Desain, Henkjan Honing, Rinus Aarts et Renee Timmers : *Rythm Perception and Production*, chapitre Rhytmic Aspects of Vibrato, pages 203–216. Lisse : Swets and Zeitlinger, 2000.

[dLG06] Serge de Laubier et Vincent Goudard : Meta-instrument 3 : A look over 17 years of practice. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '06, pages 288–291, Paris, France, June 4-8 2006. IRCAM ; Centre Pompidou.

[DPP+96] Thierry Dutoit, Vincent Pagel, N. Pierret, F. Bataille et Van der Vrecken Olivier : The mbrola project : Towards a set of high quality speech synthesizers free of use for non commercial purposes. *In Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, volume 3, pages 1393–1396, Philadelphia, PA, USA, October 3-6 1996.

[dPW+12] Nicolas d'Alessandro, Aura Pon, Johnty Wang, David Eagle, Ehud Sharlin et Sidney Fels : A digital mobile choir : Joining two interfaces towards composing and performing collaborative mobile music. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '12, Ann Arbor, MI, USA, May 21-23 2012.

[dRLB11] Christophe d'Alessandro, Albert Rilliard et Sylvain Le Beux : Chironomic stylization of intonation. *Acoustical Society of America*, 129(3):1594–1604, 2011.

[DRW39] Homer Dudley, R.R. Riesz et S.S.A. Watkins : A synthetic speaker. *Journal of the Franklin Institute*, 227(6):739 – 764, 1939.

[dWF+07] Nicolas d'Alessandro, Pascale Woodruff, Yohann Fabre, Thierry Dutoit, Sylvain Le Beux, Boris Doval et Christophe d'Alessandro : Real-time and accurate musical control of expression in singing synthesis. *Journal on Multimodal User Interfaces*, 1(1):31–39, March 2007.

[EF87] Shimon Edelman et Tamar Flash : A model of handwriting. *Biological Cybernetics*, 57(1–2):25–36, 1987.

[Fan70] Gunnar Fant : *Acoustic Theory of Speech Production*. Mouton, 1970.

[Fd13] Lionel Feugère et Christophe d'Alessandro : Performative voice synthesis for edutainment in acoustic phonetics and singing : a case study using the "cantor digitalis". *In In Intelligent Technologies for Interactive Entertainment*, volume 124, pages 169–178. Springer 2013 Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Mons, Belgium, July 3-5 2013.

[FdDP] Lionel Feugère, Christophe d'Alessandro, Boris Doval et Olivier Perrotin : Cantor digitalis : Chironomic parametric synthesis of singing. *Journal of Transactions on Audio, Language and Signal Processing*, Soumis.

[Feu13] Lionel Feugère : *Synthèse par règles de la voix chantée contrôlée par le geste et applications musicales*. Thèse de doctorat, Université Pierre et Marie Curie (UPMC), September 26 2013.

[FH85] Tamar Flash et Neville Hogan : The coordination of arm movements : An experimentally confirmed mathematical model. *Journal of Neuroscience*, 5(7):1688–1703, 1985.

[FH92] Sidney Fels et Geoffrey Hinton : Glove-talk - a neural-network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, 3(6):1–7, 1992.

[FH98] Sidney Fels et Geoffrey Hinton : Glove-talk ii - a neural-network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, 9(1):205–212, January 1998.

[Fit54] Paul M. Fitts : The information capacity of the human motor system in controlling the amplitude movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954.

[FLBd11] Lionel Feugère, Sylvain Le Beux et Christophe d'Alessandro : Chorus digitalis : Polyphonic gestural singing. *In International Workshop on Performative Speech and Singing Synthesis*, March 14-15 2011.

[FLL85] Gunnar Fant, Johan Liljencrants et Q. Lin : A four-parameter model of glottal flow. Quarterly Progress and Status Report 4, Royal Institute of Technologies - Dept. for Speech, Music and Hearing, 1985.

[FS58] James L. Flanagan et Michael G. Saslow : Pitch discrimination for synthetic vowels. *Acoustical Society of America*, 30(5):435–442, 1958.

[Fuj81] H. FUJISAKI : Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations. Rapport technique, Royal Institute of Technologies - Dept. for Speech, Music and Hearing, 1981.

[GGF14] Vincent GOUDARD, Hugues GENEVOIS et Lionel FEUGÈRE : On the playing of monodic pitch in digital music instruments. *In* Anastasia GEORGAKI et Giorgos KOUROUPETROGLOU, éditeurs : *Proceedings of the International Computer Music Conference (ICMC)*, pages 1418–1425, Athens, Greece, September 2014. National and Kapodistrian University of Athens.

[GGGD11] Vincent GOUDARD, Hugues GENEVOIS, Emilien GHOMI et Boris DOVAL : Dynamic intermediate models for audiographic synthesis. *In Proceedings of Sound and Music Computing*, Padova, Italy, July 6-9 2011.

[GIKGK13] Roni Y. GRANOT, Rona ISRAEL-KOLATT, Avi GILBOA et Tsafrir KOLATT : Accuracy of pitch matching significantly improved by live voice model. *Journal of Voice*, 27(3):390.e13–390.e20, 05 2013.

[GKP04] Sylvie GIBET, Jean-François KAMP et Franck POIRIER : Gesture analysis : Invariant laws in movement. *In* Antonio CAMURRI et Gualtiero VOLPE, éditeurs : *Gesture-Based Communication in Human-Computer Interaction*, volume 2915 de *Lecture Notes in Computer Science*, pages 1–9. Springer Berlin Heidelberg, 2004.

[GRDC00] Claude GHEZ, Thanassis RIKAKIS, R. Luke DUBOIS et Perry R. COOK : An auditory display system for aiding interjoint coordination. *In Proceedings of the International Conference on Auditory Display (ICAD)*, 2000.

[Gre70] Anthony G. GREENWALD : Sensory feedback mechanisms in performance control : With special reference to the ideo-motor mechanism. *Psychological review*, 77(2):73–99, 1970.

[GRH12] Krzysztof Z. GAJOS, Katharina REINECKE et Charles HERRMANN : Accurate measurements of pointing performance from in situ observations. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 3157–3166, Austin, Texas, USA, 2012. ACM.

[GST+09] Anke GRELL, Johan SUNDBERG, Sten TERNSTRÖM, Martin PTOK et Eckart ALTENMÜLLER : Rapid pitch correction in choir singers. *Acoustical Society of America*, 126(1):407–413, May 2009.

[Hak09] Lippold HAKEN : Position correction for an electronic musical instrument, novembre 17 2009. US Patent 7,619,156.

[Hen01] Nathalie HENRICH : *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception.* Thèse de doctorat, Université Paris 6, Novembre 2001.

[HF87] Neville HOGAN et Tamar FLASH : Moving gracefully : quantitative theories of motor coordination. *Trends in Neurosciences*, 10(4):170 – 174, 1987.

[HHW00] Andy HUNT, David M. HOWARD et Jim WORSDALL : Real-time interfaces for speech and singing. *In Proceedings of the Euromicro Conference*, pages 356–361, Maastricht, The Netherlands, September 5-7 2000.

[Hil99] Harold Andy HILDEBRAND : Pitch detection and intonation correction apparatus and method, octobre 26 1999. US Patent 5,973,252.

[HMAP01] Bernhard HOMMEL, Jochen MÜSSELER, Gisa ASCHERSLEBEN et Wolfgang PRINZ : The theory of event coding (tec) : A framework for perception and action planning. *Behavioral and Brain Sciences*, 24:849–937, 2001.

[Hog84] Neville HOGAN : An organizing principle for a class of voluntary movements. *Journal of Neuroscience*, 4(11):2745–2754, 1984.

[Hol83] J.N. HOLMES : Formant synthesizers : Cascade or parallel ? *Speech Communication*, 2(4):251 – 273, 1983.

[HRD+03] Ulrich HOPPE, Frank ROSANOWSKI, Michael DÖLLINGER, Jörg LOHSCHELLER, Maria SCHUSTER et Ulrich EYSHOLDT : Glissando : laryngeal motorics and acoustics. *Journal of Voice*, 17(3):370 – 376, 2003.

[HSD87] Jean HAKES, Thomas SHIPP et E. Thomas DOHERTY : Acoustic properties of straight tone, vibrato, trill, and trillo. *Journal of Voice*, 1(2):148 – 156, 1987.

[HTL+12] Michael Xuelin HUANG, Will W. W. TANG, Kenneth W. K. LO, C. K. LAU, Grace NGAI et Stephen CHAN : Melodicbrush : A novel system for cross-modal digital art creation linking calligraphy and music. *In Proceedings of the Designing Interactive Systems Conference*, DIS '12, pages 418–427, Newcastle Upon Tyne, UK, June 11-15 2012. ACM.

[HTW98] Lippold HAKEN, Ed TELLMAN et Patrick WOLFE : An indiscrete music keyboard. *Computer Music Journal*, 22(1):30–48, 1998.

[HW02] Andy HUNT et Marcelo M. WANDERLEY : Mapping performer parameters to synthesis engines. *Organized Sound*, 7(2):97–108, 2002.

[HWP03] Andy HUNT, Marcelo M. WANDERLEY et Matthew PARADIS : The importance of parameter mapping in electronic instrument design. *Journal of New Music Research*, 32(4):429–440, 2003.

[IRDC14] Léonidas IOANNIDIS, Jean-Luc ROUAS et Myriam DESAINTE-CATHERINE : Caractérisation et classification automatique des modes phonatoires en voix chantée. *In Actes des Journées d'Etudes sur la Parole (JEP)*, Le Mans, France, June 23-27 2014.

[JBB06] Jordi JANER, Jordi BONADA et Merlijn BLAAUW : Performance-driven control for sample-based singing voice synthesis. *In Proceedings of the Conference on Digital Audio Effects (DAFX)*, Montreal, Canada, September 18-20 2006.

[JGAK07] Sergi JORDÀ, Günter GEIGER, Marcos ALONSO et Martin KALTENBRUNNER : The reactable : Exploring the synergy between live music performance and tabletop tangible interfaces. *In Proceedings of the International Conference on Tangible and Embedded Interaction (TEI)*, TEI '07, pages 139–146, Baton Rouge, Louisiana, 2007. ACM.

[JL03] Patrik N. JUSLIN et Petri LAUKKA : Communication of emotions in vocal expression and music performance : Different channels, same code ? *Psychological Bulletin*, 129(5):770–814, 2003.

[Jor03] Sergi JORDÀ : Interactive music systems for everyone exploring visual feedback as a way for creating more intuitive, efficient and learnable instruments. *In Proceedings of the Stockholm Music Acoustics Conference (SMAC)*, Stockholm, Sweden, August 6-9 2003.

[KB08] Bernd J. KRÖGER et Peter BIRKHOLZ : Articulatory synthesis of speech and singing : State of the art and suggestions for future research. *In COST Action 2102 and euCognition Internation School*, Vietri sul Mare, Italy, 2008.

[KBLH08] Roi Cohen KADOSH, Warren BRODSKY, Michal LEVIN et Avishai HENIK : Mental representation : What can pitch tell us about the distance effect ? *Cortex*, 44(4):470 – 477, 2008. Special Issue on Numbers, Space, and Action.

[Kes04a] Loïc KESSOUS : *Contrôles Gestuels Bi-Manuels de Processus Sonores*. Thèse de doctorat, Université Paris VIII, November 2004.

[Kes04b] Loïc KESSOUS : Gestural control of singing voice, a musical instrument. *In Proceedings of Sound and Music Computing*, Paris, France, October 20-22 2004.

[KK90] Dennis H. KLATT et Laura C. KLATT : Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Acoustical Society of America*, 1990.

[KL62] J. L. Jr. KELLY et C. C. LOCHBAUM : Speech synthesis. *Proceedings of the Stockholm Speech Communication Seminar*, pages 1–4, 1962.

[Kla80] Dennis H. KLATT : Software for a cascade/parallel formant synthesizer. *Acoustical Society of America*, 67(3):971–995, 1980.

[KMKdC99] Hideki KAWAHARA, Ikuyo MASUDA-KATSUSE et Alain de CHEVIGNÉ : Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency- based f0 extraction : Possible role of a repetitive structure in sounds. *Speech Communication*, 27:187–207, 1999.

[KO07] Hideki KENMOCHI et Hayato OHSHITA : Vocaloid - commercial singing synthesizer based on sample concatenation. *In Proceedings of Interspeech*, pages 4009–4010, Antwerp, Belgium, August 27-31 2007. ISCA.

[Lar77] Bjorn LARSSON : Music and singing synthesis equipment (musse). Rapport technique 1, Royal Institute of Technologies - Dept. for Speech, 1977.

[LB09] Sylvain LE BEUX : *Contrôle Gestuel de la Prosodie et de la Qualité Vocale*. Thèse de doctorat, Université Paris-Sud, 2009.

[LBdRD10] Sylvain LE BEUX, Christophe D'ALESSANDRO, Albert RILLIARD et Boris DOVAL : Calliphony : A system for real-time gestural modification of intonation and rhythm. *In Speech Prosody*, Chicago, IL, US, 2010.

[LBFd11] Sylvain LE BEUX, Lionel FEUGÈRE et Christophe D'ALESSANDRO : Chorus digitalis : Experiments in chironomic choir singing. *In Proceedings of Interspeech*, Florence, Italy, August 27-31 2011. ISCA.

[LBRd07] Sylvain LE BEUX, Albert RILLIARD et Christophe D'ALESSANDRO : Calliphony : A real-time intonation controller for expressive speech synthesis. *In 6th ISCA Worksop on Speech Synthesis*, Bonn, Germany, August 22-24 2007. ISCA.

[LMLS+13] Pauline LARROUY-MAESTRI, Yohana LÉVÊQUE, Daniele SCHÖN, Antoine GIOVANNI et Dominique MORSOMME : The evaluation of singing voice accuracy : A comparison between subjective and objective methods. *Journal of Voice*, 27(2):259.e1 – 259.e5, 2013.

[LMM12] Pauline LARROUY-MAESTRI et Dominique MORSOMME : Criteria and tools for objectively analysing the vocal accuracy of a popular song. *Logopedics Phionatrics Vocology*, pages 1–8, April 2012.

[LMMM13] Pauline LARROUY-MAESTRI, David MAGIS et Dominique MORSOMME : The effect of melody and technique on the singing accuracy of trained singers. *Logopedics Phionatrics Vocology*, pages 1–4, 2013.

[LMMM14] Pauline LARROUY-MAESTRI, David MAGIS et Dominique MORSOMME : Effects of melody and technique on acoustical and musical features of western operatic singing voices. *Journal of Voice*, 28(3):332 – 340, 2014.

[LR11] Roland LAMB et Andrew N. ROBERTSON : Seabord : a new piano keyboard-related interface combining discrete and continuous control. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '11, pages 503–506, Oslo, Norway, May 30 - June 1 2011.

[LSG67] Erich LUSCHEI, Carol SASLOW et Mitchell GLICKSTEIN : Muscle potentials in reaction time. *Experimental Neurology*, 18(4):429 – 442, 1967.

[LSM12] Stefan LADWIG, Christine SUTTER et Jochen MÜSSELER : Crosstalk between proximal and distal action effects when using a tool. *Journal of Psychology*, 220(1):10–15, 2012.

[LSM13] Stefan LADWIG, Christine SUTTER et Jochen MÜSSELER : Intra- and intermodal integration of discrepant visual and proprioceptive action effects. *Experimental Brain Research*, 231(4):457–468, 2013.

[LTV83] Francesco LACQUANITI, Carlo TERZUOLO et Paolo VIVIANI : The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54:115–130, 1983.

[Mac92] I. Scott MACKENZIE : Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139, mars 1992.

[Mar10] Martin MARIER : The sponge. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '10, pages 356–359, Sidney, Australia, June 15-18 2010.

[MB92] I. Scott MACKENZIE et William BUXTON : Extending fitts' law to two-dimensional tasks. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 219–226, Monterey, California, USA, May 3-7 1992. ACM.

[MB02] Michael MCGUFFIN et Ravin BALAKRISHNAN : Acquisition of expanding targets. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 57–64, Minneapolis, Minnesota, USA, 2002. ACM.

[MB05] Michael J. MCGUFFIN et Ravin BALAKRISHNAN : Fitts' law and expanding targets : Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interactions (TOCHI)*, 12(4):388–422, décembre 2005.

[MDCGPS13] Pauline MOUAWAD, Myriam DESAINTE-CATHERINE, Anne GÉGOUT-PETIT et Catherine SEMAL : The role of the singing acoustic cues in the perception of broad affect dimensions. *In Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR)*, Marseille, France, October 15-18 2013.

[MGS13] Andrew P. MCPHERSON, Adrian GIERAKOWSKI et Adam M. STARK : The space between the notes : Adding expressive pitch control to the piano keyboard. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2195–2204, Paris, France, 2013. ACM.

[MHT14] Ville MÄKELÄ, Tomi HEIMONEN et Markku TURUNEN : Magnetic cursor : Improving target selection in freehand pointing interfaces. *In Proceedings of The International Symposium on Pervasive Displays*, PerDis '14, pages 112 :112–112 :117, Copenhagen, Denmark, June 3-4 2014. ACM.

[MHWL09] Mark T. MARSHALL, Max HARTSHORN, Marcelo M. WANDERLEY et Daniel J. LEVITIN : Sensor choice for parameter modulations in digital musical instruments : Empirical evidence from pitch modulation. *Journal of New Music Research*, 38(3):241–253, 2009.

[MKA+88] David E. MEYER, Sylvan KORNBLUM, Richard A. ABRAMS, Charles E. WRIGHT et J. E. KEITH SMITH : Optimality in human motor performance : Ideal control of rapid aimed movements. *Psychological review*, 95(3):340–370, 1988.

[MKSW82] David E. MEYER, J. E. KEITH SMITH et Charles E. WRIGHT : Models for the speed and accuracy of aimed movements. *Psychological review*, 89(5):449–482, 1982.

[MM87] Denise MYERS et John MICHEL : Vibrato and pitch transitions. *Journal of Voice*, 1(2):157 – 161, 1987.

[MOKH04] Hiroki MORI, Wakana ODAGIRI, Hideki KASUYA et Kiyoshi HONDA : Transitional characteristics of fundamental frequency in singing. *In Internal Congress on Acoustics (ICA)*, pages 499–500, Kyoto, Japan, 2004.

[Moo73] B. C. J. MOORE : Frequency difference limens for short-duration tones. *Acoustical Society of America*, 54(3):610–619, 1973.

[MPHS02] Dirk MÜRBE, Friedemann PABST, Gert HOFMANN et Johan SUNDBERG : Significance of auditory and kinesthetic feedback to singers' pitch control. *Journal of Voice*, 16(1):44 – 51, 2002.

[MR94] I. Scott MACKENZIE et Stan RIDDERSMA : Effects of output display and control—display gain on human performance in interactive systems. *Behaviour and Information Technology*, 13(5):328–337, 1994.

[MS09] Jochen MÜSSELER et Christine SUTTER : Perceiving one's own movements when using a tool. *Consciousness and Cognition*, 18:359–365, 2009.

[MW06] Mark T. MARSHALL et Marcelo M. WANDERLEY : Evaluation of sensors as input devices for computer music interfaces. *In Proceedings of the International Conference on Computer Music Modeling and Retrieval (CMMR)*, CMMR'05, pages 130–139, Pisa, Italy, 2006. Springer-Verlag.

[Nag89] H. NAGASAKI : Asymmetric velocity and acceleration profiles of human arm movements. *Experimental Brain Research*, 74(2):319–326, 1989.

[Nel83] W. L. NELSON : Physical principles for economies of skilled movements. *Biological Cybernetics*, 46(2):135–147, February 1983.

[OGMGS14] Laura ORTEGA, Emmanuel GUZMAN-MARTINEZ, Marcia GRABOWECKY et Satoru SUZUKI : Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Atten Percept Psychophys*, 76:1485–1502, 2014.

[OMNT12] Keiichiro. OURA, A. MASE, Yoshihiko NANKAKU et Keiichi TOKUDA : Pitch adaptive training for hmm-based singing voice synthesis. *In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5377–5380, Kyoto, Japan, March 25-30 2012.

[OSW01] Nicola ORIO, Norbert SCHNELL et Marcelo M. WANDERLEY : Input devices for musical expression : Borrowing tools from hci. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '01, pages 1–4, Seattle, Washington, USA, April 1-2 2001.

[PB07] Peter Q. PFORDRESHER et Steven BROWN : Poor-pitch singing in the absence of "tone deafness". *Music Perception*, 25(2):99–115, 2007.

[PBM+10] Peter Q. PFORDRESHER, Steven BROWN, Kimberly M. MEIER, Michel BELYK et Mario LIOTTI : Imprecise singing is widespread. *Acoustical Society of America*, 128(4):2182–2190, July 18 2010.

[Pd] Olivier PERROTIN et Christophe D'ALESSANDRO : Target acquisition vs. expressive motion : Dynamic pitch warping for intonation correction of digital musical instruments. *ACM Transactions on Computer-Human Interactions (TOCHI)*, Accepted.

[Pd13] Olivier PERROTIN et Christophe D'ALESSANDRO : Adaptive mapping for improved pitch accuracy on touch user interfaces. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '13, pages 186–189, Daejeon, South Korea, May 27-30 2013.

[Pd14] Olivier PERROTIN et Christophe D'ALESSANDRO : Visualizing gestures in the control of a digital musical instrument. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '14, pages 605–608, London, UK, June 30 - July 4 2014.

[Pd15] Olivier PERROTIN et Christophe D'ALESSANDRO : Quel ajustement de hauteur mélodique pour les instruments de musique numériques ? *In Actes des Journées d'Informatique Musicale (JIM)*, Montréal, Canada, 2015.

[PF06] Bob PRITCHARD et Sidney FELS : Grassp : Gesturally-realized audio, speech and song performance. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '06, pages 272–276, Paris, France, June 4-8 2006. IRCAM ; Centre Pompidou.

[PF08] Pascal PERRIER et Susanne FUCHS : Speed–curvature relations in speech production challenge the 1/3 power law. *Journal of Neurophysiology*, 100:1171–1183, June 2008.

[PK13] Maria PIETROWICZ et Karrie KARAHALIOS : Sonic shapes : Visualizing vocal expression. *In Proceedings of the International Conference on Auditory Display (ICAD)*, pages 157–164, July 6-10 2013.

[PKS+12] Ondrej POLACEK, Martin KLIMA, Adam J. SPORKA, Pavel ZAK, Michal HRADIS, Pavel ZEMCIK et Vaclav PROCHAZKA : A comparative study on distant free-hand pointing. *In Proceedings of the European Conference on Interactive Tv and Video*, EuroiTV '12, pages 139–142, Berlin, Germany, 2012. ACM.

[PMNI05] J. Karen PARKER, Regan L. MANDRYK, Michael N. NUNES et Kori M. INKPEN : Tractorbeam selection aids : Improving target acquisition for pointing input on tabletop displays. *In* MariaFrancesca COSTABILE et Fabi PATERNÒ, éditeurs : *Human-Computer Interaction - INTERACT*, volume 3585 de *Lecture Notes in Computer Science*, pages 80–93. Springer Berlin Heidelberg, Rome, Italy, 2005.

[PTIK13] Alexandros PINO, Evangelos TZEMIS, Nikolaos IOANNOU et Georgios KOUROUPETROGLOU : Using kinect for 2d and 3d pointing tasks : Performance evaluation. *In Proceedings of the International Conference on Human-Computer Interaction : Interaction Modalities and Techniques*, volume 4 de *HCI'13*, pages 358–367, Las Vegas, NV, July 21-26 2013. Springer-Verlag.

[RBOF05] Stuart REEVES, Steve BENFORD, Claire O'MALLEY et Mike FRASER : Designing the spectator experience. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 741–750, Portland, Oregon, USA, 2005. ACM.

[RD85] Xavier RODET et Philippe DEPALLE : High quality synthesis-by-rule of consonants. *In Proceedings of the International Computer Music Conference (ICMC)*, pages 91–96, Burnaby, Canada, 1985.

[RKG+06] Elena RUSCONI, Bonnie KWAN, Bruno L. GIORDANO, Carlo UMILTÀ et Brian BUTTERWORTH : Spatial representation of pitch height : the smarc effect. *Cognition*, 99(2):113–129, 3 2006.

[RKP05] Martina RIEGER, Günther KNOBLICH et Wolfgang PRINZ : Compensation for and adaptation to changes in the environment. *Experimental Brain Research*, 163(4):487–502, 2005.

[Rod02] Xavier RODET : Synthesis processing of the singing voice. *In Proceedings of the IEEE Benelux Workshop in Model based Processing and Coding of Audio (MPCA)*, pages 99–108, Leuven, Belgium, November 15 2002.

[RPB84] Xavier RODET, Yves POTARD et Jean-Baptiste BARRIERE : The chant project : From the synthesis of the singing voice to synthesis in general. *Computer Music Journal*, 8(3):15–31, 1984.

[SM04] R. William SOUKOREFF et I. Scott MACKENZIE : Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *Int. J. Hum.-Comput. Stud.*, 61(6):751–789, décembre 2004.

[SMB11] Christine SUTTER, Jochen MÜSSELER et L. BARDOS : Effects of sensorimotor transformations with graphical input devices. *Behaviour and Information Technology*, 30(3):415–424, 2011.

[SRC10] Minghui SUN, Xiangshi REN et Xiang CAO : Effects of multimodal error feedback on human performance in steering tasks. *Information Processing Society of Japan Journal*, 51(12):2375–2383, 2010.

[SSI90] Xavier SERRA et J. SMITH III : Spectral modeling synthesis : A sound analysis/synthesis based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14:12–24, 1990. SMS.

[SSRM13] Christine SUTTER, Sandra SÜLZENBRÜCK, Martina RIEGER et Jochen MÜSSELER : Limitations of distal effect anticipation when using tools. *New Ideas in Psychology*, 31(3):247–257, 2013.

[STUA04] Takeshi SAITOU, Naoya TSUJI, Masashi UNOKI et Masato AKAGI : Analysis of acoustic features affecting "singing-ness" and its application to singing-voice synthesis from speaking-voice. *In Proceedings of Interspeech*, Jeju, Korea, October 4-8 2004. ISCA.

[SUA02] Takeshi SAITOU, Masashi UNOKI et Masato AKAGI : Extraction of f0 dynamic characteristics and develoment of f0 control model in singing voice. *In Pro-*

ceedings of the International Conference on Auditory Display (ICAD), Kyoto, Japan, July 2–5 2002.

[SUA05] Takeshi SAITOU, Masashi UNOKI et Masato AKAGI : Development of an {F0} control model based on {F0} dynamic characteristics for singing-voice synthesis. *Speech Communication*, 46(3–4):405 – 417, January 2005. Quantitative Prosody Modelling for Natural Speech Description and Generation International Conference on Speech Prosody.

[Sun73] Johan SUNDBERG : Data on maximum speed of pitch changes. Quarterly progress and status report, Royal Institute of Technologies - Dept. for Speech, Music and Hearing, 1973.

[Sun94] Johan SUNDBERG : Acoustic and psychoacoustic aspects of vocal vibrato. Quarterly Progress and Status Report 2-3, Royal Institute of Technologies - Dept. for Speech, Music and Hearing, 1994.

[Sun06] Johan SUNDBERG : The kth synthesis of singing. *Advances in Cognitive Psychology*, 2(2–3):131–143, 2006.

[SZH+79] Richard A. SCHMIDT, Howard N. ZELAZNIK, Brian HAWKINS, James S. FRANK et John T. QUINN : Motor-output variability : A theory for the accuracy of rapid motor acts. *Psychological review*, 86(5):415–451, September 1979.

[TAB+14a] Etienne THORET, Mitsuko ARAMAKI, Christophe BOURDIN, Lionel BRINGOUX, Richard KRONLAND-MARTINET et Solvi YSTAD : Audio-motor synchronization : the effect of mapping between kinematics and acoustic cues on geometric motor features. *In Sound, Music and Motion*, pages 234–245. Springer Berlin Heidelberg, 2014.

[TAB+14b] Etienne THORET, Mitsuko ARAMAKI, Lionel BRINGOUX, Richard KRONLAND-MARTINET et Solvi YSTAD : When acoustic stimuli turn visual circles into ellipses : sounds evoking accelerations modify visuo-motor coupling. *International Multisensory Research Forum (IMRF)*, June 11-14 2014.

[TAKM+14] Etienne THORET, Mitsuko AMARAKI, Richard KRONLAND-MARTINET, Jean-Luc VELAY et Solvi YSTAD : From sound to shape : Auditory perception of drawing movements. *Journal of Experimental Psychology*, page January, 2014.

[TCSW06] Daniel TRUEMAN, Perry R. COOK, Scott SMALLWOOD et Ge WANG : Plork : The princeton laptop orchestra year 1. *In Proceedings of the International Computer Music Conference (ICMC)*, New Orleans, USA, 2006.

[TDW03] Caroline TRAUBE, Philippe DEPALLE et Marcelo M. WANDERLEY : Indirect acquisition of instrumental gesture based on signal , physical and perceptual information. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '03, pages 42–47, Montreal, Canada, May 22-24 2003.

[tea13] R Core TEAM : *R : A Language and Environment for Statistical Computing*. The R Foundation - http ://www.r-project.org, 2013.

[TNT+13] Keiichi TOKUDA, Yoshihiko NANKAKU, Tomoki TODA, Heiga ZEN, Junichi YAMAGISHI et Keiichiro OURA : Speech synthesis based on hidden markov models. *In Proceedings of the IEEE*, volume 101, pages 1234–1252, May 2013.

[TRZ12] Huawei TU, Xiangshi REN et Shumin ZHAI : A comparative evaluation of finger and pen stroke gestures. *In Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems, CHI '12, pages 1287–1296, Austin, Texas, USA, 2012. ACM.

[TS88] Sten TERNSTRÖM et Johan SUNDBERG : Intonation precision of choir singers. *Acoustical Society of America*, pages 59–69, February 1988.

[TW04] Stephen M. TASKO et John R. WESTBURY : Speed-curvature relations for speech-related articulatory movement. *Journal of Phonetics*, 32(1):65 – 80, 2004.

[VC85] Paolo VIVIANI et Marco CENZATO : Segmentation and coupling in complex movements. *Journal of Experimental Psychology*, 11(6):828–845, December 1985.

[Vel98] Raymond VELDHUIS : A computationally efficient alternative for the lf model and its perceptual evaluation. *Acoustical Society of America*, 1998.

[VF95] Paolo VIVIANI et Tamar FLASH : Minimum-jerk, two-thirds power law, and isochrony : Converging approaches to movement planning. *Journal of Experimental Psychology*, 21(1):32–53, 1995.

[VM83] Paolo VIVIANI et G. MCCOLLUM : The relation between linear extent and velocity in drawing movements. *Neuroscience*, 10(1):211 – 218, 1983.

[VT82] Paolo VIVIANI et Carlo TERZUOLO : Trajectory determines movement dynamics. *Neuroscience*, 7(2):431 – 437, 1982.

[VUK96] Roel VERTEGAAL, Tamas UNGVARY et Michael KIESLINGER : Towards a musician's cockpit : Transducers, feedback and musical function. *In Proceedings of the International Computer Music Conference (ICMC)*, pages 308–311, Hong Kong, China, 1996.

[Wan97] Marcelo M. WANDERLEY : Les nouveaux gestes de la musique. Rapport technique, IRCAM, April 1997.

[WD99] Marcelo M. WANDERLEY et Philippe DEPALLE : Contrôle gestuel de la synthèse sonore. *In* H. VINET et F. DELALANDE, éditeurs : *Interfaces Homme-Machine et Creation Musicale*. Hermès Science Publishing, 1999.

[WD04] Marcelo M. WANDERLEY et Philippe DEPALLE : Gestural control of sound synthesis. *In Proceedings of the IEEE*, volume 92, pages 632–644, April 2004.

[WM83] Charles E. WRIGHT et David E. MEYER : Conditions for a linear speed-accuracy trade-off in aimed movements. *Quarterly Journal of Experimental Psychology*, 35A:279–296, 1983.

[WOL11] Ge WANG, Jieun OH et Tom LIEBER : Designing for the ipad : Magic fiddle. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '11, pages 197–202, Oslo, Norway, May 30 - June 1 2011.

[WSM+12] Lei WANG, Christine SUTTER, Jochen MÜSSELER, Ronald Josef Zvonimir DANGEL et Catherine DISSELHORST-KLUG : Perceiving one's own limb movements with conflicting sensory feedback : The role of mode of movement control and age. *Frontiers in Psychology*, 3(289):1–7, August 2012.

[WSS14] Nike WENDKER, Oliver S. SACK et Christine SUTTER : Visual target distance, but not visual cursor path length produces shifts in motor behavior. *Frontiers in Psychology*, 5, March 2014.

[WVIR00] Marcelo M. WANDERLEY, Jean-Philippe VIOLLET, F. ISART et Xavier RODET : On the choice of transducer technologies for specific musical functions. *In Proceedings of the International Computer Music Conference (ICMC)*, volume 2000, Berlin, Germany, September 2000.

[WWBH97] Aileen WORDEN, Neff WALKER, Krishna BHARAT et Scott HUDSON : Making computers easier for older adults to use : Area cursors and sticky icons. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 266–271, Atlanta, Georgia, USA, 1997. ACM.

[WWF97] Matthew WRIGHT, David WESSEL et Adrian FREED : New musical control structures from standard gestural controllers. *In Proceedings of the International Computer Music Conference (ICMC)*, pages 387–390, Thessaloniki, Greece, 1997.

[XS00] Yi XU et Xuejing SUN : How fast can we really change pitch ? maximum speed of pitch change revisited. *In Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, pages 666–669, Beijing, China, October 16-20 2000.

[YTK+99] Takayoshi YOSHIMURA, Keiichi TOKUDA, Takao KOBAYASHI, Takashi MASUKO et Tadashi KITAMURA : Simultaneous modeling of spectrum, pitch and duration in hmm-based speech synthesis. *In Proceedings of Eurospeech*, Budapest, Hungary, September 5-9 1999.

[ZGS84] Jan ZERA, Jan GAUFFIN et Johan SUNDBERG : Synthesis of selected vcv-syllables in singing. Rapport technique 2-3, Royal Institute of Technologies - Dept. for Speech, 1984.

[ZS06] Mark ZADEL et Gary SCAVONE : Different strokes : A prototype software system for laptop performance and improvisation. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '06, pages 168–171, Paris, France, June 4-8 2006. IRCAM ; Centre Pompidou.

[ZTB09] Heiga ZEN, Keiichi TOKUDA et Alan W. BLACK : Statistical parametric speech synthesis. *In Speech Communication*, volume 51, pages 1039–1064, 2009.

[ZWMC07] Michael ZBYSZYNSKI, Matthew WRIGHT, Ali MOMENI et Daniel CULLEN : Ten years of tablet musical interfaces at cnmat. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '07, pages 100–105, New York, NY, USA, June 6-10 2007. ACM.