



HAL
open science

Uncontrolled manifolds and short-delay reflexes in speech motor control: a modeling study

Andrew Szabados

► **To cite this version:**

Andrew Szabados. Uncontrolled manifolds and short-delay reflexes in speech motor control: a modeling study. *Neurons and Cognition [q-bio.NC]*. Université Grenoble Alpes, 2017. English. NNT: 2017GREAS039 . tel-01731536

HAL Id: tel-01731536

<https://theses.hal.science/tel-01731536>

Submitted on 14 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : PCN - Sciences cognitives, psychologie et neurocognition

Arrêté ministériel : 25 mai 2016

Présentée par

Andrew SZABADOS

Thèse dirigée par **Pascal PERRIER (EDISCE)**, Professeur, G-INP

préparée au sein du **Laboratoire Grenoble Images Parole Signal Automatique**
dans l'**École Doctorale Ingénierie pour la santé la Cognition et l'Environnement**

Uncontrolled manifolds et réflexes à courte latence dans le contrôle moteur de la parole : une étude de modélisation

Uncontrolled manifolds and short-delay reflexes in speech motor control: a modeling study

Thèse soutenue publiquement le **27 novembre 2017**,
devant le jury composé de :

Monsieur PASCAL PERRIER

PROFESSEUR, GRENOBLE INP, Directeur de thèse

Madame SUSANNE FUCHS

DIRECTRICE DE RECHERCHE, LEIBNIZ CENTRE LINGUISTIQUE
BERLIN, Examineur

Madame BEATRICE VAXELAIRE

PROFESSEUR, UNIVERSITE DE STRASBOURG, Examineur

Monsieur RAFAEL LABOISSIERE

CHARGE DE RECHERCHE, CNRS DELEGATION ALPES, Président

Monsieur TAKAYUKI ITO

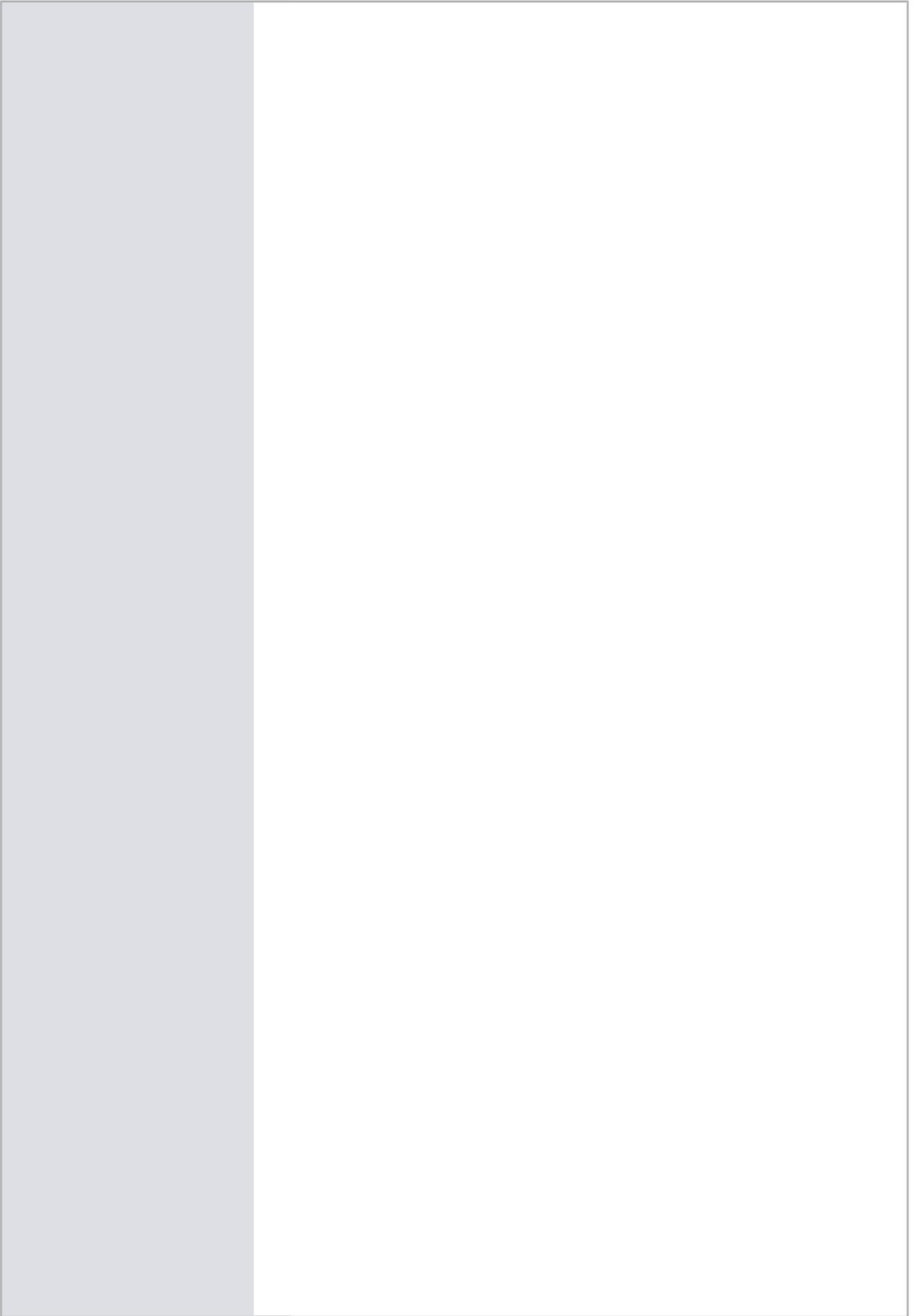
CHARGE DE RECHERCHE, CNRS DELEGATION ALPES, Examineur

Monsieur GREGOR SCHONER

PROFESSEUR, UNIV. DE LA RUHR A BOCHUM - ALLEMAGNE,
Rapporteur

Monsieur YVES LAPRIE

DIRECTEUR DE RECHERCHE, CNRS DELEGATION CENTRE-EST,
Rapporteur



Abstract

Uncontrolled manifolds and short-delay reflexes in speech motor control: a modeling study

Keywords: speech motor control, biomechanical modeling, motor equivalence, perturbation, uncontrolled manifolds

This work makes use of a biomechanical model of speech production as a reference subject to address several phenomena related to the adaptability and stability of speech motor control, namely motor equivalence and postural stability.

The first part of this thesis is related to the phenomenon of motor equivalence. Motor equivalence is a key feature of speech motor control, since speakers must constantly adapt to various phonetic contexts and speaking conditions. The Uncontrolled Manifold (UCM) idea offers a theoretical framework for considering motor equivalence in which coordination among motor control variables is separated into two subspaces, one in which changes in control variables modify the output and another one in which these changes do not influence the output.

This concept is developed and investigated for speech production using a 2D biomechanical model. First, a representation of the linearized UCM based on orthogonal projection matrices is proposed. The UCMs of various vocal tract configurations of the 10 French oral vowels are then characterized using their command perturbation responses. It is then investigated whether each phonetic class such as phonemes, front/back vowels, rounded/un-rounded vowels can be characterized by a unique UCM, or whether the UCMs vary significantly across representatives of these different classes. It was found that linearized UCMs, especially those that are specifically computed for each configuration, but also across many of the phonetic classes allow for a command perturbation response that is effective. This suggests that similar motor equivalence strategies can be implemented within each of these classes and that UCMs provide a valid characterization of an equivalence strategy. Further work is suggested to elaborate which classes might be used in practice.

The second part addresses the question of the degree to which postural control of the tongue is accomplished through passive mechanisms - such as the mechanical and elastic properties of the tongue itself - or through reflexes - such as the stretch reflex.

An external force perturbation, was applied to the 2D biomechanical model, specifically one in which the tongue is pulled anteriorly using specific force profile exerted on the tongue body using a force effector attached to the superior part of the tongue blade. Simulation results were compared to experimental data collected at Gipsa-lab under similar conditions.

This perturbation was simulated with various values of the model's parameter modulating the reflex strength (feedback gain). The results showed that a perturbation rebound seen in simulated data is due to a reflex mechanism. Since a compatible rebound is seen in data from human subjects, this can be taken as evidence of a reflex mechanism being involved in postural stability of the tongue. The time course of the mechanisms of this reflex, including the generation of force and the movement of the tongue, were analyzed and it was determined that the precision of the model was insufficient to make any conclusions on the origin of this reflex (whether cortical or brainstem). Still, numerous experimental directions are proposed.

Résumé en Français

Uncontrolled manifolds et réflexes de court délai dans le contrôle moteur de la parole : une étude de modélisation

Mots clés: contrôle moteur de la parole, modélisation biomécanique, équivalence motrice, perturbation, uncontrolled manifolds

Ce travail exploite un modèle biomécanique de la production de la parole comme sujet de référence pour étudier plusieurs phénomènes liés à l'adaptabilité et à la stabilité du contrôle moteur de la parole, en particulier l'équivalence motrice et le contrôle postural.

La première partie de cette thèse s'intéresse au phénomène de l'équivalence motrice. L'équivalence motrice est une caractéristique essentielle du contrôle moteur de la parole, car les locuteurs doivent s'adapter constamment à des contextes phonétiques toujours différents et à conditions variables de production de la parole. Le concept de « Uncontrolled Manifold » (UCM) offre un cadre théorique pour comprendre les mécanismes sous-jacents à l'équivalence motrice : il propose de représenter la coordination entre les variables de contrôle moteur en deux sous-espaces séparés, un dans lequel tout changement des variables de contrôle affectent la sortie et un autre dans lequel ces changements n'influencent aucunement la sortie.

Ce concept est développé et étudié pour la production de la parole en utilisant un modèle biomécanique 2D du conduit vocal. D'abord, une représentation des UCM linéarisées basée sur des matrices de projection orthogonale est proposée. Les UCM de différentes configurations du conduit vocal des 10 voyelles orales françaises sont ensuite caractérisées en étudiant les réponses aux perturbations de leurs commandes. On étudie alors si chaque catégorie phonétique, telle que les phonèmes, les voyelles antérieures/postérieures, ou les voyelles arrondies/non-arrondies, peut être caractérisée par une UCM unique ou si les UCM varient considérablement entre les différents représentants de chacune de ces classes. On a constaté que les UCM linéarisées, celles qui sont spécifiquement calculées pour chaque configuration du conduit vocal, mais aussi celles, plus globales, des classes phonétiques, permettent une réponse efficace aux perturbations des commandes. Cela suggère que des stratégies équivalentes d'équivalence motrice peuvent être mises en œuvre dans chacune de ces classes et que les UCM en fournissent des caractérisations exploitables. Des suggestions sont faites pour de futurs travaux pour déterminer quelles classes pourraient être utilisées dans la pratique.

La deuxième partie étudie dans quelle mesure le contrôle postural de la langue exploite des mécanismes passifs - tels que les propriétés mécaniques et élastiques intrinsèques de la langue- ou des réflexes à faible latence - comme le réflexe d'étirement.

Une perturbation en force a été appliquée au modèle biomécanique 2D, dans laquelle la langue est tirée vers l'avant par une force exercée sur le corps de la langue à l'aide d'un robot relié à la partie supérieure de la lame de la langue. Les simulations ont été comparées à des données expérimentales recueillies au Gipsa-lab dans des conditions similaires.

Cette perturbation a été simulée avec différentes valeurs du paramètre qui dans le modèle module le feedback induit par l'étirement des fibres musculaires. Les résultats ont montré un effet de rebond dans les mouvements de la langue suite à la perturbation qui est imputable au mécanisme réflexe. Étant donné qu'un rebond similaire est observé dans les données expérimentales sur des sujets humains, ce résultat suggère qu'un mécanisme réflexe joue un rôle significatif dans la stabilité posturale de la langue. Les caractéristiques temporelles de ce réflexe ont été analysées et il s'avère que la précision du modèle est insuffisante pour tirer des conclusions sur l'origine, corticale ou spinale, de ce réflexe. Des pistes pour de futures études expérimentales sont proposées.

Uncontrolled manifolds et réflexes de court délai dans le contrôle moteur de la parole : une étude de modélisation

L'intérêt d'étudier le contrôle du moteur de la parole en général a été présenté. En particulier, la parole présente des caractéristiques intéressantes, telles que les cibles étant de formes multiples, dont certaines ne sont liées qu'à des transformations non linéaires de la configuration physique des articulateurs. Les motivations pour l'utilisation d'un modèle biomécanique de la production de la parole, par opposition à des méthodes articulatoires ou autres, ont ensuite été présentées.

À la suite de cette motivation, deux études ont été décrites, dans lesquelles un modèle biomécanique de la production de la parole a été utilisé pour répondre à des questions fondamentales concernant non seulement le contrôle du moteur de la parole, mais le contrôle du moteur en général. Aucune de ces études n'aurait été possible sans l'utilisation d'un modèle biomécanique.

Uncontrolled manifolds dans le contrôle moteur de la parole

Tout d'abord, la question a été posée sur la question de savoir si les uncontrolled manifolds, initialement proposés comme moyen de traiter l'équivalence motrice et d'expliquer les modèles de variabilité observés dans le mouvement, pourraient être utiles dans le contexte de la parole. Après avoir discuté d'une extension du concept de l'UCM pour aborder la nature hiérarchique de la parole et se concentrer sur l'UCM auditive, une représentation des UCM linéarisées basé sur des matrices de projection orthogonale a été développée. Il a ensuite été démontré qu'ils semblent être utiles dans le contexte de la parole en raison de 1) leur capacité à proposer des compensations à une perturbation de commande qui avait des niveaux d'erreur acceptables, malgré la non-linéarité de la correspondance entre les commandes du moteur et l'acoustique les caractéristiques qu'ils sont censés préserver et 2) le fait que ces compensations de commande étaient toujours plus proches du point perturbé que des commandes non perturbées. En particulier, en testant différentes granularités de localité dans l'espace de sortie acoustique, on a constaté qu'il y avait un léger avantage pour chaque phonème représenté par son UCM linéarisé, un avantage qui serait d'autant plus prononcé dans le cas où un nombre limité de configurations exemplaires auraient été utilisées pour chaque phonème.

Bien qu'il soit conclu que le concept de l'UCM, en particulier dans sa représentation linéarisée en tant que matrice de projection orthogonale sur l'UCM linéarisée, est approprié et même utile dans le contexte de la parole, plusieurs questions ouvertes concernant les UCM dans la parole restent.

L'une est celle de savoir si ces UCM caractérisent bien la variabilité observée dans la pratique pendant la parole naturelle. Cependant, la difficulté à aborder cette question est la nature "cachée" non seulement des commandes dans la parole, mais aussi des paramètres articulatoires. On pourrait imaginer cependant un travail futur dans lequel les commandes, un proxy de celles-ci, ou simplement une certaine représentation des paramètres articulatoires sont estimés à partir de signaux EMG ou à travers d'autres données observables (telles que la position de la langue à l'aide d'EMA,

d'échographie ou même d'IRM en temps réel) qui pourrait alors être adapté aux commandes en utilisant le modèle biomécanique ou une approximation de celui-ci. Une telle étude permettrait des preuves empiriques soutenant l'hypothèse selon laquelle la variabilité du mouvement de la langue ou des articulateurs impliqués dans la production de la parole est structurée de manière à refléter l'utilisation de collecteurs incontrôlés dans la planification du mouvement de la langue.

Une autre question ouverte est de savoir comment une telle représentation pourrait être intégrée dans un contrôleur. Comme on l'a vu dans le chapitre 2, les UCM sont compatibles avec l'approche du contrôle de rétroaction stochastique optimal, mais des travaux supplémentaires devraient être effectués afin de voir si une relation d'équivalence pourrait être établie entre une implémentation neuronale compatible avec OFC et une des UCM.

Réponses de perturbation dans le contrôle moteur de la parole

Deuxièmement, les mécanismes de contrôle postural de la langue ont été étudiés en utilisant le modèle biomécanique pour aider à interpréter les données empiriques acquises dans une étude parallèle chez des sujets humains. Grâce à des simulations et à leurs analyses, des preuves ont été présentées en faveur de l'interprétation des données empiriques comme contenant une signature d'un réflexe élastique. En particulier, en modulant la rétroaction proprioceptive dans l'équation de la force musculaire du modèle biomécanique, le rôle de ce paramètre dans la génération d'un effet de rebond après une perturbation externe a été démontré et on a observé, en particulier, que les données empiriques étaient cohérentes avec le cas dans lequel le feedback proprioceptif est non nulle. Grâce à une analyse des délais de différents aspects de cette simulation et à une comparaison de ces données avec la littérature sur les délais des mécanismes de réponse corticale et du tronc cérébral dans la langue et les mécanismes réflexes des lèvres, il a été déterminé que les données n'étaient pas concluantes et qu'ils sont compatibles, dans le cadre de la précision temporelle du modèle, avec un réflexe lent du cerveau ou un transcortical rapide.

Des travaux supplémentaires sont donc nécessaires pour éliminer complètement la possibilité que le réflexe soit dû à des mécanismes corticaux et l'établir en raison d'un réflexe du cerveau, ou vice versa. En outre, comme mentionné dans le chapitre 2 la preuve pour les fibres myotatiques chez l'humain, bien que existant, laisse place à la spéculation quant à savoir s'il existe un nombre suffisant qui est correctement situé pour être responsable d'un tel réflexe dans la langue humaine. D'autres études pourraient donc inclure un travail anatomique pour vérifier l'emplacement et la prévalence des fibres myotatiques, le cours et l'emplacement potentiellement précis de toute réponse du tronc cérébral, en particulier en ce qui concerne toute réponse corticale. On pourrait également imaginer des études plus complexes qui chercheraient à élucider la voie neurale précise et si la réponse est en effet basée sur la réponse des fibres myotatiques ou, éventuellement, des mécanorécepteurs.

Utilisation d'un modèle biomécanique

Dans les deux études décrites ci-dessus, l'utilisation du modèle biomécanique était essentielle aux conclusions spécifiques obtenues et représentait la nouveauté des approches présentées dans ce travail. Dans le premier cas, les collecteurs incontrôlés impliquant les commandes musculaires n'auraient pas pu être étudiés aussi facilement sans un tel modèle et la perspicacité obtenue dans l'application des UCM à la planification du mouvement aurait été limitée. En particulier, si un modèle articulaire avait été utilisé, la question de savoir si les UCM fonctionneraient si bien pour caractériser une relation plus complexe, avec une non-linéarité réaliste. De même, dans la deuxième étude, le modèle biomécanique des forces impliquées dans le mouvement de la langue a permis l'interprétation de données empiriques qui, autrement, auraient nécessité de l'expérimentation supplémentaire à évaluer, la différence avec l'état actuel étant que des hypothèses beaucoup plus précises peuvent être testées avec des expériences plus précises. Ces deux études ont ainsi permis, en plus d'aborder les hypothèses spécifiques discutées ci-dessus dans les sections correspondantes, l'utilité des modèles biomécaniques pour répondre à des questions complexes de contrôle du moteur de la parole.

En effet, cette relation d'efforts de modélisation biomécanique avec les efforts pour comprendre les questions théoriques de contrôle moteur est mutuellement bénéfique. Non seulement les modèles peuvent servir de proxies aux données humaines, en facilitant la conception de méthodes expérimentales en aidant à la génération et à la formulation d'hypothèses et à l'interprétation de données acquises expérimentalement, toutes les données expérimentales acquises de cette manière peuvent également être utilisées pour améliorer la fidélité et de même l'utilité des modèles biomécaniques eux-mêmes.

Perspective

Comme vu au chapitre 1, alors que le travail présenté ici se concentre sur le contrôle du moteur de la parole, la réalité est que l'étude de la parole, en raison de sa complexité et de son importance dans le langage et la connaissance, a le potentiel de donner un aperçu non seulement du langage et de la cognition, mais des systèmes adaptatifs complexes capables de générer des comportements complexes en général.

En particulier, les questions d'intérêt dans ce contexte sont les degrés auxquels la commande est exécutée en fonction de cibles de différents types (sensoriels ou moteurs) et auxquelles elle est centralisée ou réalisée par des boucles de contrôle spécialisées, voire potentiellement hiérarchiques. Grâce à ces études et à l'utilisation du modèle biomécanique, certaines preuves du rôle de la périphérie dans le contrôle du moteur de la parole ont été développées.

Tout d'abord, la périphérie elle-même offre une redondance physique, où différentes commandes et différentes configurations d'activation musculaire peuvent accomplir les mêmes tâches. Les uncontrolled manifolds montrent comment cela est une conséquence naturelle de la relation entre commande et position tandis que l'étude de perturbation a montré que non seulement la langue, elle-même élastique, a tendance à résister aux perturbations, mais il semble y avoir des mécanismes actifs qui utilisent cette élasticité, stratégiquement en le modifiant pour stabiliser la

langue après une perturbation.

Deuxièmement, la riche dynamique de la périphérie, en raison de sa complexité biomécanique, peut être utilisée pour minimiser les efforts physiques et neuraux pour faire face à l'incertitude, qu'il s'agisse d'instabilité ou de perturbations, pendant les tâches motrices. Les uncontrolled manifolds permettent de structurer la variabilité selon les nécessités d'une tâche, en évitant des efforts de compensation excessifs lorsque cela est possible, en particulier dans le contexte des commandes motrices. Ils, à la lumière de la dynamique des points d'équilibre, permettent également une représentation simple des objectifs qui peut être utilisée pour générer des commandes pour des trajectoires efficaces. La réponse aux perturbations observée dans l'étude de perturbation et de modélisation met en évidence les faits selon lesquels la périphérie a un effet de stabilisation lui-même et que des boucles de feedback du tronc cérébral rapide ou même corticales implémentent des réflexes qui résistent aux perturbations et stabilisent ainsi la maintenance de configurations articulées.

Bien que ces phénomènes aient été étudiés dans le contexte de la parole, ces mécanismes sont généraux pour tous les types de contrôle sensori-moteur, en particulier une périphérie qui maximise la stabilité et minimise la complexité du contrôle et une division du travail dans les différents circuits d'une hiérarchie de contrôle, tous sensible au feedback.

Contents

Abstract	i
Summary in French	ii
1 Introduction	1
1.1 Purpose	1
1.2 General Motivation	1
1.3 Why Speech?	1
1.3.1 The complexity of the speech motor system	1
1.3.2 Speech as a sensorimotor task	2
1.3.3 Speech is involved in cognition	3
1.3.4 Functional Complexity	3
Adaptiveness	3
Precision and Stability	4
1.3.5 Speech as a complex adaptive behavior	4
1.4 Use of a biomechanical model	5
1.5 Outline of thesis	5
2 Background elements	7
2.1 Neural and physiological overview of speech motor control	7
2.1.1 General organizational principles of sensorimotor control	7
Central versus peripheral	7
Voluntary, involuntary movement and reflexes	9
2.1.2 Divisions of central nervous system	9
Cortex	10
Subcortical structures	10
Spinal cord and brainstem	11
Cerebellum	11
Descending tracts	11
Lower motor neurons	11
Muscles	12
Sensory information	12
2.1.3 Peripheral structures involved in speech	13
Primary structures contributing to speech articulation	13
Other structures involved in speech	14
2.1.4 Motor commands	15
2.1.5 Between muscle commands and sensory feedback	16
2.1.6 Stretch reflex	16
2.1.7 Inseparability of speech, language, and cognition	17

2.2	Control - basic principles	17
2.2.1	Formulation of the control problem	18
	System/plant/process and state variables	18
	Control/input variables	18
	Output variables	19
	Objective/cost function	19
	Optimal control	19
2.2.2	Feedforward and feedback control	19
	Feedback control	20
	Feedforward control	20
	Feedforward vs. feedback control in motor control	20
	Stochastic optimal feedback control	21
	State feedback control	23
2.2.3	Internal models	23
	Forward model	24
	Inverse model	24
	Internal dynamical models	25
	Internal models in speech	25
2.2.4	Motor Equivalence	26
	Stochastic optimal feedback control and motor equivalence	26
	Uncontrolled manifolds and motor equivalence	28
2.3	Uncontrolled Manifolds	28
2.3.1	UCMs and Linearized UCMs	28
	Jacobian matrix	29
2.3.2	Uncontrolled manifolds in relation to inverse models	30
2.3.3	Uncontrolled manifolds and stochastic optimal feedback control	30
2.4	Biomechanical modeling of speech motor control	30
2.4.1	Tongue, jaw, and lip model	31
	Biomechanical tongue model	32
	λ -model and threshold parameters	32
	Geometric model of lips and jaw	33
	Time parameters	35
	Area function extraction	35
	Acoustic analysis	37
	Vowel identification	37
3	Uncontrolled manifolds in speech motor control	39
3.1	Uncontrolled manifolds in speech	39
3.1.1	Hierarchy of uncontrolled manifolds	40
	Anticipatory UCM	40
	Articulatory UCM	40
	Auditory UCM	42
	Phonemic UCM and beyond	42
	Limitations and considerations for other decompositions	42
3.1.2	Application to speech motor control	43
	Focus on auditory UCMs	43

	Potential roles of acoustic UCMs	43
	Linearized uncontrolled manifolds in speech	45
	Orthogonal projection matrix as representation of LUCM	46
3.2	Validation of utility of LUCMs in speech	46
3.2.1	Methodology	48
	Calculation of Jacobian matrices and projection matrices	48
	LUCM groupings	50
3.2.2	Results and Analysis	52
	Overall compensation error	52
	Presentation of Error Data	53
	Results by Grouping	54
	Source of Error	55
	Compensation effort and efficiency of LUCMs	56
3.2.3	Conclusion	56
3.3	Discussion and Perspective	59
4	Reflex mechanism in speech motor control	63
4.1	Postural control of the tongue	63
4.1.1	Previous work on orofacial reflexes	64
4.1.2	Short-delay reflex loops in the tongue	66
4.2	Tongue perturbation modeling study	67
4.2.1	Experimental setup	68
	Preliminary results in human subjects	68
	Experimental question	68
4.2.2	Methodology	70
	Addition of external force perturbation to biomechanical model	70
	Simulations - Part A	71
	Addition of local feedback gain parameter to model	72
	Simulations - Part B	73
	Analyzing timecourse of perturbation response	73
4.2.3	Results and Analysis	73
	Numerical stability of model and representativeness of results	73
	Part A	75
	Part B	79
	Cortical or brainstem reflex	81
	Experimental hints from biomechanical model	84
4.2.4	Conclusion	87
4.3	Discussion and Perspective	89
5	Conclusion	91
5.1	Summary	91
5.1.1	Uncontrolled manifolds in Speech Motor Control	91
5.1.2	Perturbation responses in Speech Motor Control	92
5.1.3	Use of a biomechanical model	93
5.2	Perspective	93
5.3	List of contributions	94
5.4	List of experimental directions	95

List of Figures

2.1	Neural and physiological components of speech motor control	8
2.2	Illustration of motor equivalence	27
2.3	Illustration of uncontrolled manifold concept	29
2.4	Diagram of biomechanical model usage	32
2.5	Tongue muscles in biomechanical model	33
2.6	Lips and jaw with effect of parameters	34
2.7	Area function extraction	35
2.8	Vowels	36
2.9	Acoustic feature extraction	37
2.10	Formant space with French vowels	38
3.1	Hierarchies in speech and language	41
3.2	UCM scenarios	44
3.3	Block diagram of Study 1	48
3.4	Vowels obtain from dictionary	49
3.5	Depiction of Groupings 3, 4, and 5	51
3.6	Method for calculating tolerances	53
3.7	Histograms of residual formant error	54
3.8	Error by grouping	55
3.9	Error by phoneme	58
3.10	Error by perturbation location	59
3.11	Error by perturbation location and type	60
4.1	Gomi et al. 2002 latency data	65
4.2	Ito et al. 2005 TMS data	65
4.3	Weber and Smith 1987 EMG data	66
4.4	Tongue perturbation study setup	68
4.5	Perturbation data from one subject	69
4.6	Block diagram of Study 2	70
4.7	Anchor and sensor placement	71
4.8	Simulation result variable diagram	74
4.9	Command and force level histograms	74
4.10	Tongue, jaw, lip configurations for /e/	76
4.11	Simulated perturbation responses	77
4.12	Horizontal sensor displacement in simulations	77
4.13	Peak horizontal displacement and rebound magnitude	78
4.14	Force level and maximum displacement	79
4.15	0-deflection intercept delay and deflection velocities	80
4.16	Effect of cFactor on example configuration (I)	81

4.17	Effect of cFactor on example configuration (II)	82
4.18	Effect of cFactor on perturbation response variables	83
4.19	Histograms of temporal perturbation response values	84
4.20	Muscle activation levels after perturbation	85
4.21	Force levels after perturbation	86
4.22	Contribution of feedback to post-perturbation force levels	86
4.23	Displacement due to perturbation and due to feedback	87
4.24	Verticalis muscle in biomechanical model	88

List of Tables

2.1	Motor command variables in biomechanical model	34
3.1	French oral vowels in dictionary	49
3.2	Description of groupings	50
3.3	Command perturbation summary	52
3.4	LUCM performance results by data subset	57
3.5	Compensation effort by grouping and perturbation type	60
4.1	Orofacial reflex latencies	67
4.2	Simulation failure statistics	75

Chapter 1

Introduction

1.1 Purpose

The purpose of this thesis is to gain insight into the mechanisms underlying speech motor control. The originality of our approach lies in the use of a biomechanical model of speech production as a *reference-subject* that we can thoroughly manipulate in order to assess a number of hypotheses and basic principles. In particular, this will be done through two studies, the first of which investigates the phenomenon of motor equivalence while the second looks into the role of the periphery, specifically reflexes, in maintaining stable control in the presence of perturbations.

1.2 General Motivation

The interest of studying speech motor control is in its status as a prime example of a complex sensorimotor task, which in turn is an interesting category of behavior to study as an example of a complex adaptive behavior.

The hope is then, by understanding speech motor control, to develop insight into the functional principles not just of human cognition but of complex adaptive behavior in general.

1.3 Why Speech?

The justification for studying speech lies not just in its important role in cognition, but also in the complexity of speech as a sensorimotor behavior, notably one which exhibits particularly well-defined hallmarks of complex behavior in its adaptiveness, stability, and precision despite relatively high speed.

1.3.1 The complexity of the speech motor system

Speech is realized through the coordination of a large number of muscles of different size, in different locations, of different types and structural complexity. Moreover this coordination involves multiple timescales from those of the fast movements of the primary articulators such as the tongue in the production of individual sounds to the slower variations in breathing and tension responsible for prosody.

Indeed the count of the number muscles involved can be as high as 80, depending on how the divisions of the muscles of the tongue, jaw, lips diaphragm, larynx, etc. are counted (Bhatnagar, 2002). These structures reflect a great variety in location, morphology, and complexity as well.

The motor gestures of speech, including those of individual phonemes, can last on the order of 30-200 ms (Umeda, 1977; Ziółko and Ziółko, 2009) while prosodic variations and other elements of discourse-level planning, such as coordination of pauses and breathing, pitch resets, and boundary tones, not to mention planning of discourse content can be on the scale of multiple seconds (Swerts, 1997).

1.3.2 Speech as a sensorimotor task

Not only is speech of a high complexity from a strictly motor point of view, but it is a prime example of a sensorimotor behavior, that is a motor behavior which is highly integrated with sensory representations and feedback. The control of the articulation of speech aims to reach primarily acoustic targets although processing makes use of not just auditory information, but also somatosensory (both proprioceptive and tactile) information and even visual information, all of which is of a different nature and involves differing temporal delays.

Indeed the fundamental representations underlying both speech motor control and speech perception have been hypothesized to be in terms of not just sensory or motor representations, but of combined sensorimotor representations (Schwartz et al., 2012). Moreover the representations of sensory systems have recently been shown to encode information in motor terms just as the motor areas encode information in terms of sensory variables (Bouchard et al., 2013).

There are also numerous examples of multimodal sensory integration and its influences on both speech perception and production. The best known among these is the McGurk effect (McGurk and MacDonald, 1976) where visual information can influence the auditory perception. Additional examples include somatosensory perturbations of the lips and soft-tissue of the cheek influencing speech perception (Ito, Tiede, and Ostry, 2009) and perturbation of auditory feedback influencing speech production (Houde and Jordan, 1998) even for clinical purposes as a way to prevent stuttering (Vaxes, 1963; Cai et al., 2012). Moreover, top-down influences in terms of cognitive control also influence speech production and perception, even beyond the scope of volitional control, allowing context to influence speech in ways from priming effects for the articulation or mis-articulation of individual sounds (Goldinger, Luce, and Pisoni, 1989) to adaptation effects in terms of prosody or speaking rate (Pardo, 2013), even as a function of complex social factors (Natale, 1975).

The information involved in different sensory modalities also represents different physical processes and is necessarily processed at different times, making its integration into the processes of speech production and perception all the more impressive.

So, speech is an example of a particularly complex sensorimotor behavior as evidenced by not just its motor complexity, but the underlying sensorimotor representations and degrees to which information from multiple sensory modalities and executive processes is able to influence its production and perception.

1.3.3 Speech is involved in cognition

As the sensorimotor phenomenon in which language is grounded, speech is an integral component of cognition.

Sensorimotor processes give rise to a phenomenon of categorical perception by which speech gestures are perceived in discrete units, phonemes. Through the process of double articulation (Martinet, 1957; Hockett, 1960) these discrete units and combinations become associated with meaning which is in turn grounded in multi-modal representations formed through experience and interaction with the world. The combinatorial rules of morphology and syntax underly the combination of units of meaning allowing for symbolic and abstract processing.

The ability to speak is thus if not the basis at least an integral component for many aspects of higher cognition including mathematical reasoning, memory, etc. as well as social cognition. It is even suggested that the ability to speak had a central role in the evolution of human cognition (Deacon, 1998).

1.3.4 Functional Complexity

Speech exhibits several interesting characteristics of functionally complex behavior, including its adaptiveness, stability, and high precision even in light of its relatively high speed.

Adaptiveness

Speech is highly adaptive. Articulation can be adapted to different linguistics contexts, acoustic environments, or speaking rates, with phonemes or other characteristics being approximated or modulated in order to both facilitate comprehension and minimize effort (Lindblom, 1990). The set of phonemes making up language is not fixed and indeed varies highly between different languages, just as the specific set of devices - for example pitch or consonant length - varies. Moreover, the functions of speech itself can be taken over entirely through an analogous use of the hands in sign language (Hickok, Bellugi, and Klima, 1996).

It has even been suggested that the adaptiveness of speech is due to its operation in the space of possible reconstructions in the mind of other speakers instead of the physical space in which most actions take place (Grimme et al., 2011). Indeed the double articulation mentioned above results in a hierarchy of goals (Ingram, 2007, p. 21) which, while ultimately being grounded in a physical gesture, allows for a great degree of flexibility across a wide range of conditions.

Precision and Stability

Speech is relatively fast - both in the sense of the articulatory velocities which can be on the order of 10-20 cm/s for the dorsum and tip of the tongue (Ostry and Munhall, 1985; Adams, Weismer, and Kent, 1993; Stoll, Hoole, and Harrington, 2016) and in the short duration of movements which can be around 100 ms (Adams, Weismer, and Kent, 1993). It is also relatively precise with certain dimensions of articulatory targets being on the order of several millimeters (Perkell et al., 1993).

Despite the phenomenon of a speed-accuracy trade-off which is ubiquitous in biological systems (Heitz, 2014) and is most famously known in the context of motor control in the form of Fitts's law (Fitts, 1954), speech motor control provides an interesting example where the complexity of the underlying sensorimotor system seems to be adapted to minimize the functional effects of this compromise, although it is still present (Lammert et al., 2016).

As will be seen in Section 2.2.4 and Chapter 3, one phenomenon which can be exploited to minimize functional effects of this trade-off is that of motor equivalences and structuring of variability in such a way that precise targets can be reached more reliably. Uncontrolled manifolds will be investigated as such a mechanism in the context of speech motor control in Chapter 3.

Also, as suggested by Perrier, Ostry, and Laboissière (1996) the neural, muscular, and biomechanical dynamics of speech production seem to implement a partial solution to this trade-off in the context of stabilization of speech production by implementing fast, peripheral feedback loops in order to overcome the delay that would otherwise be introduced by cortical processing of feedback. The existence of such a process in speech will be investigated in Chapter 4.

1.3.5 Speech as a complex adaptive behavior

These examples highlight several characteristics - adaptiveness and interesting solutions to minimize the speed-accuracy trade-off while providing for stability - that motivate a general interest one could have in the speech-language-cognition system - that of it being a prime example of a complex adaptive system (Steels, 2000; Beckner et al., 2009). That is, by studying speech we potentially gain insight into general principles of complex organisms and how they control their complex behaviors (Holland, 1992).

As mentioned above, speech is grounded in a complex network of neural and musculoskeletal structures, exhibiting dynamics on numerous time scales, and behavior which is adaptive and relatively precise and fast. There are certainly general principles at work in the organizational complexity of this system. For example, the patterns of connectivity and structure of dependencies, including whether there are any hierarchically organized components, are likely to reveal much about the way in which information is processed in these systems.

As seen in the example of sign language and in the general resilience of language and especially speech, for everything from the articulation of specific phonemes to the encoding of meaning in complex social environments, this system is certainly adaptive and certainly its adaptive nature and flexibility can be better understood

by looking at the complex structuring of its components. Thus, increased knowledge about the underlying abilities of adaptiveness or their implementation has the potential to provide information about the general principles of complex motor behavior.

1.4 Use of a biomechanical model

A model embodies certain hypotheses, and in particular a biomechanical model embodies hypotheses regarding the materials and dynamics of the system modeled.

We would like to investigate several questions related to the above described phenomena of adaptiveness and stability, but there are numerous obstacles to doing so directly on human subjects. One alternative is to attempt these analyses on a model representing the relevant functionality of the underlying speech production system. Since the questions involve the dynamics of the articulation of speech, using simpler kinematic or articulatory models is insufficient and we must use a biomechanical model if we wish to avoid studying speech production on human subjects directly.

Indeed there are numerous advantages to making use of a biomechanical model as a reference subject. One is that the study need not be limited to data or parameters accessible in human subjects. For example, in the case of the tongue, the data that can be extracted from the model is not limited to the location of sensors and the experiment can be repeated a virtually unlimited number of times with a wide range of precisely controlled parameters. Also, as will be seen in Chapter 4, such use of a biomechanical model can be made in parallel to studies on human subjects in a way that the biomechanical model and the analysis of its results aids in the analysis of the experimental data.

Despite the advantages and the utility of using a biomechanical model as a reference subject, there are some potential limitations when compared to acquiring data directly from human subjects. Most importantly, the accuracy of the results when using a biomechanical model is limited to the accuracy of the model not just in terms of the accuracy of its parameters but especially in terms of the accuracy and completeness of the hypotheses which it embodies. Such an issue will also be looked at in the context of Chapter 4.

1.5 Outline of thesis

The work presented in the following chapters thus is meant to be seen in this context. Looking at speech, in particular with a biomechanical model, is meant to take a closer look at several specific questions of speech motor control while at the same time knowing that the motivation behind these efforts includes the potential for this work to bring insight into similar questions in terms of both sensorimotor processes and complex adaptive processes in general.

In that spirit, first an overview of specific background knowledge that will be useful for understanding the following work will be presented in Chapter 2. This

will include presenting relevant details about the neural and musculoskeletal system and its control, uncontrolled manifolds, and the biomechanical model used.

One specific question requiring further study is the degree to which this adaptiveness is a consequence of neurally-implemented algorithms which exploit the structure of the problem. In Chapter 3, a study involving uncontrolled manifolds seeking to characterize the redundancy inherent to the sensorimotor phenomenon of speech and the appropriateness of uncontrolled manifolds as a representation of this redundancy will be presented.

Another question is the the degree to which the complex structuring of the neuromuscular system itself provides for stabilization and enables its fast, yet accurate use. The study presented in Chapter 4 looks at the role of reflexes in the system's response to external perturbations.

Lastly, in Chapter 5 the results will be discussed in the context of these general goals and a summary will be presented of both the experimental results and future directions.

Chapter 2

Background elements

The purpose of this chapter is to establish specific elements of the context within which the work presented in Chapters 3 and 4 was undertaken. Central to this work is the "physical substrate" of speech, which includes the musculoskeletal and neural components. Moreover it is these properties that have been modeled in a biomechanical model of the vocal tract. Using this model several questions related to control will be explored, including motor equivalence. All of these topics are addressed below.

2.1 Neural and physiological overview of speech motor control

In this section an overview¹ of the neural and physiological aspects of speech motor control relevant to the following work will be presented. First, it should be noted that there are numerous neural and physical aspects of speech motor control which go well beyond the scope of the present work, from details at the lowest level of analysis including neural dynamics to the complexities of neural circuitry. In particular, the purpose of this section will be to lay out the basic framework for understanding the questions implicated by the study in Chapter 4.

2.1.1 General organizational principles of sensorimotor control

Despite the specificities of speech motor control, there are a few general principles of sensorimotor control which are a relevant to speech and which are worth mentioning.

Central versus peripheral

First, there is a division of labor between the central nervous system (CNS) and the periphery. The CNS includes the brain and spinal cord, while the peripheral nervous system (PNS) is defined to be the nerves and ganglia outside of the CNS. The PNS is further broken down into the somatic nervous system (SoNS), which innervates skeletal muscles and is involved in their voluntary control, and the autonomic

¹Except where noted by specific references, general neuroanatomical information is based on Latash, 2008, Crossman and Neary, 2014 and Kemmerer, 2014 and information regarding the articulators of speech is based on McDonald and Creanor, 2016.

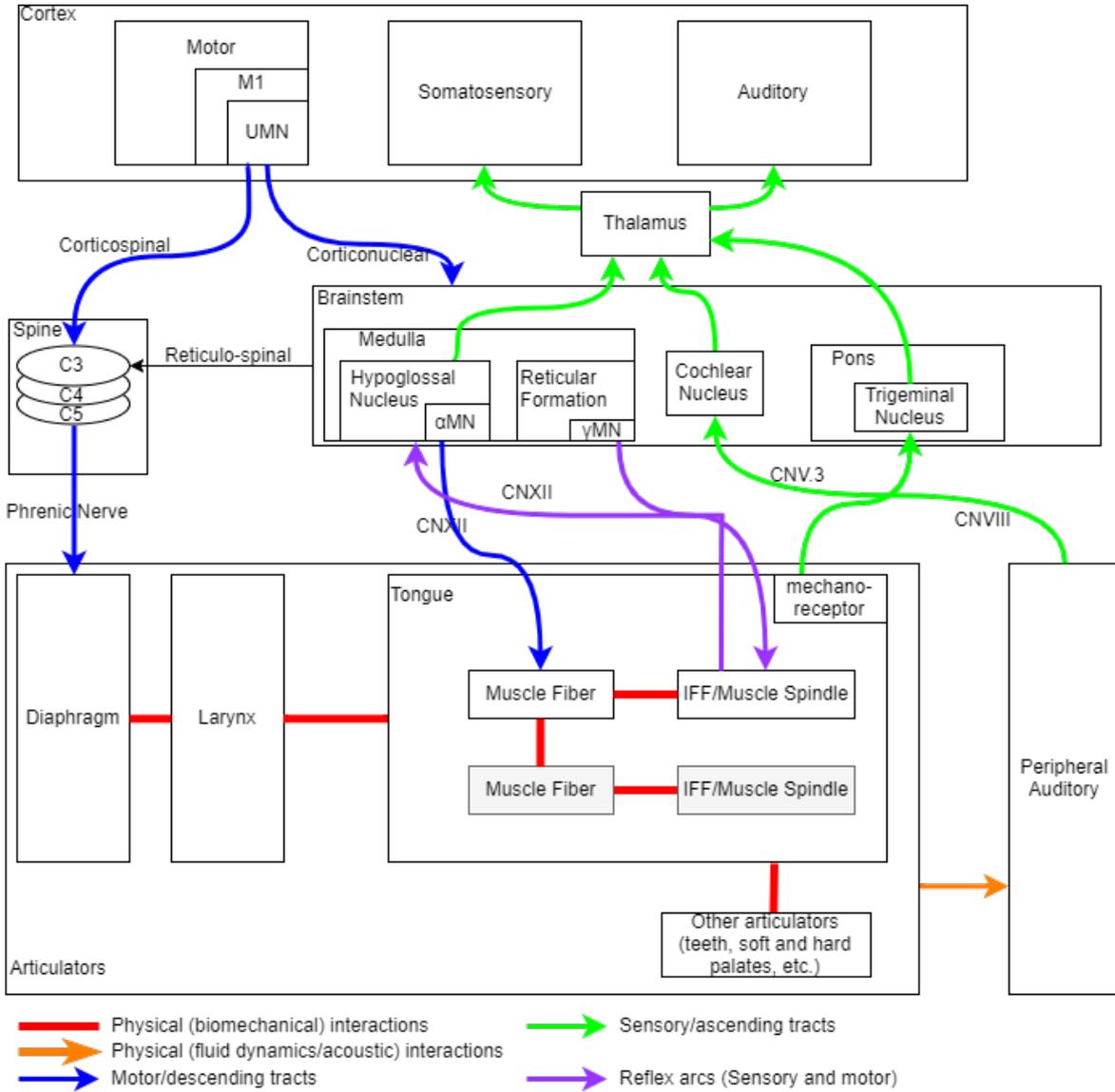


FIGURE 2.1: Overview schematic of neural and physiological components of speech motor control

nervous system (ANS), which deals with smooth muscles and glands. The notion of periphery relevant here is wider and distinct from that of the PNS, because it also includes the muscles and articulated effectors themselves. These structures - the muscles, joints, bones, and cartilage - which make up the orofacial system, are controlled by the CNS via the PNS and contribute to the dynamics of speech and its motor control. The present work thus seeks to better understand some aspects of the apparent division of labor, and one of the primary questions of interest in the study in Chapter 4 will be to elucidate the precise nature of this distinction.

Voluntary, involuntary movement and reflexes

Another important distinction made in the context of neural control of movement is voluntary and involuntary movement. Voluntary movement is typically defined as that which is perceived to be under conscious control or corresponds to goal-directed behavior. Involuntary or automatic movement consists then of movement that is done without conscious awareness and includes reflexes. The challenge in differentiating these two classes of control comes from the fact that while certain actions are executed according to conscious volition and others are executed in a way, such as through fast reflexes, that precludes their being considered voluntary, there are numerous components of movement whose origin and control are not clearly voluntary or involuntary in this sense.

Instead, in order to avoid such ambiguity, the term *voluntary movement* will be used to refer to that which is executed intentionally, while *reflex movement* will refer to the movement which is the direct result of sensory feedback and which happens too quickly for intervention of higher-level cognitive processes. More complex cases of voluntary movement which has become automated through repetition and learning will not be addressed.

Importantly, while voluntary movement is necessarily initiated in the cortex, reflex responses can be initiated both in the brainstem (or spinal cord) or in the cortex but always in response to an afferent signal originating from the PNS. So, cortical involvement does not imply voluntary and reflex does not imply a non-cortical response. Reflexes will be discussed in more detail below (see Section 2.1.6) and in Chapter 4.

2.1.2 Divisions of central nervous system

The structures of the CNS itself are divisible into several categories, including cortical, subcortical, brainstem, hindbrain, and spinal cord.

The primary structures of the CNS include those which arise developmentally as part of the telencephalon, namely the cerebral cortex and the main subcortical structures such as the thalamus and basal nuclei (also known as basal ganglia).²

Rhombencephalic, or hindbrain structures, namely the pons, medulla, and cerebellum are also essential to motor control, in particular to speech. The pons and

²While the term *basal ganglia* is certainly more common, here the term *basal nuclei* will be used due to its coherence with the distinction of ganglia and nuclei according to which ganglia are groups of cell bodies outside of the CNS and nuclei are such groups within the CNS (Anatomical Terminology, 1998).

medulla (along with the midbrain) are part of the anatomical structure referred to as the brainstem.

Additionally, the spinal cord itself is considered a structure of the CNS.

All of these structures contribute to the learning and execution of motor actions.

Cortex

The cortex is made up of the grey matter containing the cell bodies of its approximately 30 billion neurons, laid out in a layered sheet several millimeters thick, which can essentially be seen as having been folded up in order to fit inside the cranium.

The cortex is generally considered to be divided into functionally distinct areas by functions - motor, somatosensory, auditory, executive - and there are also cytoarchitectural and anatomical distinctions which give rise to cortical maps. Moreover there are complex patterns of connectivity and thus interaction between regions and the notion of functionally distinct areas should be seen as a simplification of a much more complex organization.

For example, the primary areas, namely motor and somatosensory, share numerous organizational aspects due to their primary subdivisions (primary motor and somatosensory cortices) being adjacent to each other across the central sulcus. Also interactions between areas, namely motor and auditory, give rise to organizational principles being shared. For example, coding of auditory features has been observed in motor regions as has the opposite (Chang, 2016).

For the purposes of the present work, the essential knowledge is that the cytoarchitecturally distinct area known as the primary motor cortex is the location of upper motor neurons (UMN) whose axons terminate at synapses on the dendrites of lower motor neurons in either the brainstem or the spinal cord and form the corticonuclear and corticospinal tracts respectively.

These UMNs are influenced by complex organizational principles and numerous structures - including the premotor areas and supplementary motor area - the details of which are beyond the scope of this work.

Subcortical structures

The primary subcortical structures involved in speech motor control and learning are the basal nuclei and the thalamus.

The *basal nuclei* are a set of structures including the striatum, globus pallidus, substantia nigra, and subthalamic nucleus whose primary role in motor control can be seen as that of learning action selection by either applying inhibition or disinhibiting motor behaviors, in particular in voluntary movement (Kemmerer, 2014). Since these structures and circuits are mostly implicated in action selection, the details of their function are not relevant to the present work.

The *thalamus* is a structure whose primary role can be thought of as relaying information, especially afferent sensory information to different cortical areas. It can then be involved in the reflex loop that will be studied in Chapter 4, but further details are not relevant to this work.

Together with other structures, notably the cortex, these structures are essential to the learning of motor actions and to the relaying of descending motor commands and incoming feedback processed by the cortex.

Spinal cord and brainstem

For the activation of skeletal muscles located anywhere below the neck, lower motor neurons are located in the ventral horn of the spinal column and receive projections from the motor cortex via the descending corticospinal tract.

For the activation of muscles located at or above the neck, lower motor neurons are located in various brainstem nuclei corresponding to the origins or terminations of the various cranial nerves.

The brainstem is composed of the medulla oblongata, the pons, as well as the midbrain. The primary motor nucleus for the muscles of the tongue is the hypoglossal nucleus located in the medulla whose axons comprise CN XII - hypoglossal nerve. Sensory innervation from the anterior two-thirds of the tongue converge on the trigeminal nucleus extending through the midbrain, pons, and medulla.

Cerebellum

The cerebellum is implicated in motor control, in particular in sensorimotor adaptation (Golfinopoulos et al., 2011; Guediche et al., 2014), error-driven learning and the internal models necessary for movement planning (Wolpert, Miall, and Kawato, 1998; Kawato, 1999; Parrell et al., 2017). It also seems to be involved in specifying kinematic parameters and contributing to the coordination or sequencing of vocal tract movements (Ackermann, Mathiak, and Riecker, 2007). It is certainly involved in learning responses to motor perturbations, but it is not clear what its role might be in perturbation responses such as those studied in Chapter 4.

Descending tracts

Muscle commands corresponding to activation of upper motor neurons travel down the axons of these neurons via the corticospinal or corticonuclear tracts, synapse with lower motor neurons in the spinal column or brainstem which in turn propagate this activity through their axons which comprise different nerves up until the neuromuscular junction at which point they recruit muscular activity.

Accordingly, the motor cortex can be thought of as influencing muscular activation through the intermediary of the lower motor neurons.

Lower motor neurons

There are several types of lower motor neurons. First, those primarily responsible for the activation of muscles are alpha motor neurons. Alpha motor neurons receive projections from UMNs, interneurons and sensory fibers from intrafusal fibers. Located in the cortical nuclei corresponding to the appropriate cranial nerve, for example in the hypoglossal nucleus (HGN) in the case of the tongue.

Second, there are gamma motor neurons which synapse on intrafusal fibers in muscle spindles. They are innervated by cortical motor neurons and through their influence on intrafusal fibers set the sensitivity of muscle spindles to changes in muscle length. The presence and amount of muscle spindles in the tongue have been questioned and even the evidence in favor of their existence does not clarify how they are used in speech, if at all (Cooper, 1953; Walker and Rajagopal, 1959). This point will be discussed further in Chapter 4.

Additional neurons located in the cranial nuclei include beta motor neurons and interneurons which project onto both alpha and gamma motor neurons. The role of these neurons, especially interneurons, include that hypothesized by Bizzi to be the coordination of muscle synergies, or groups of muscle fibers which are often activated together (Cheung et al., 2005).

Muscles

The neuromuscular junction is the point at which LMN axons synapse with the muscle fibers themselves. At this point axons release acetylcholine which crosses the synaptic cleft and is received by receptors in muscle fibers which begin the process of muscular contraction.

Sensory information

Motor control in general - and speech motor control in particular - incorporates numerous types of sensory information. While the question of the degree to which different forms of sensory information are taken into account during movement execution remains open (see Section 2.2), it is certainly taken into account during learning.

In speech there are two broad types of sensory information that are relevant: somatosensory and auditory.

Somatosensory information includes proprioceptive information about muscle length and movement from afferent sensory fibers originating from muscle spindles or Golgi tendon organs as well as tactile information (including vibration) from mechanoreceptors. All of this information is relayed through the spinal column or brainstem through the thalamus into the contralateral somatosensory cortex in the parietal lobe. It could be used to reconstruct the state of the configuration of the motor structures and, especially in the case of vibrational information from mechanoreceptors, to provide information about voicing and speech through a different channel than the auditory system. It is not known however to what extent this information is used in practice.

When speech is produced, auditory information relays information about the vibration of the air caused by the vocal folds and/or the airflow itself and the modulation of these vibrations by the vocal tract. In general it is considered that this information is the primary target of control in the context of speech (see Blumstein and Stevens, 1979; Perkell et al., 1995; Savariaux, Perrier, and Orliaguet, 1995) but alternative proposals have also been influential (Browman and Goldstein, 1995; Fowler, 1996). This information is relayed via excitation of frequency-band specific hair cells

of the cochlea through the cochlear nerve, cochlear nucleus, and then the thalamus to the auditory cortex.

Sensory information of all these types is used to varying degrees in different speech motor tasks. Part of the complexity of the sensorimotor control of speech is reconciling the source of this information despite its different nature and the different latencies of arrival to the cortex depending on sensory modality.

2.1.3 Peripheral structures involved in speech

Numerous peripheral structures are involved in speech, from those involved in the buildup of pressure and creation of airflow to the vibration that serves as sound source to those that shape the vocal tract. Additionally some are actively involved while others are involved passively or even statically. Below the main structures involved in speech production and articulation are outlined.

Primary structures contributing to speech articulation

The primary active structures involved in the articulation of speech are the tongue, lips, jaw, and, to a lesser extent, velum. There are also several passive structures, namely the teeth and hard palate, which are integral to speech production.

The *tongue* is the primary speech articulator. It is composed of muscles whose activation gives rise to the tongue's movement and deformation. These muscles form two sets: intrinsic muscles contained entirely within the tongue and extrinsic muscles which are attached to bones outside the tongue (Brand, Isselhard, and Satin, 2013). For more information on the muscles of the tongue, see the description below in the the section describing the biomechanical model (Section 2.4.1).

Depending on the type of sound produced and the specific point of articulation, different regions of the tongue need to be controlled with differing levels of precision. For example, for dental consonants the apex of the tongue must be placed precisely, while for palatal or velar consonants the blade or dorsum of the tongue might need to be controlled precisely. Additionally, for certain sounds, such as simple vowels or nasal consonants the tongue might need to be held in a static configuration for a certain amount of time, while in other cases such as diphthongs or certain consonants the varying acoustic target implies a trajectory of positions.

The motor innervation of all muscles of the tongue is provided by the hypoglossal nerve (CN XII), with the exception of the extrinsic palatoglossus muscle whose motor innervation is supplied by the vagus nerve (CN X; Baker, 1954).

Sensation of the tongue is generally considered to be supplied by the lingual branch of the trigeminal nerve (CN V3) for general somatic information coming from the anterior 2/3 of the tongue, by the glossopharyngeal nerve (CN IX) for the posterior third of the tongue, and by the internal laryngeal branch of the vagus nerve (CN X). While numerous cell types including different types of mechanoreceptors and nociceptors primarily innervate the tongue via the trigeminal nerve (Haggard and Boer, 2014), certain proprioceptive information seems to be communicated along the hypoglossal nerve as well (Adatia and Gehring, 1971). The specific pathway for information originating from muscle spindles, the details of any potential complex

pathways, including branches from the hypoglossal nerve ultimately arriving at the trigeminal nucleus, and the role any of these might play in speech motor control are not known.

Beyond the question of sensory versus motor innervation of the tongue, there is evidence that the sensory and motor networks are tightly connected, such that trigeminal nerve blocks result in numerous effects on EMG activity in tongue and other (mylohyoid and anterior digastric) muscles (Borden, Harris, and Catena, 1973) even if proprioceptive innervation may involve both the hypoglossal and trigeminal nerves. The *jaw* modulates the cross-sectional area of the vocal tract by opening and closing the mouth. The mandible is moved in relation to the temporal bone of the skull at the temporomandibular joints allowing both hinging and sliding movements.

The lowering movement of the jaw is then actually a combination of a protrusion (actuated by the lateral pterygoid muscle) and retention from below by the digastric muscle. Elevation of the jaw is the result of the coordinated activation of the three other muscles of the jaw - the masseter, medial pterygoid, and the temporalis (Tuller, Harris, and Gross, 1981). All muscles of the jaw (which the digastric is not) are innervated by the mandibular division of the trigeminal nerve (CN V3). The digastric muscle is actually composed of posterior and anterior bellies. The posterior belly is innervated by the facial nerve (CN VII), while the anterior belly is innervated by the mandibular division of the trigeminal nerve (CN V3), but through the mylohyoid nerve, a branch of the inferior alveolar branch of the mandibular division.

The *lips* are a complex of muscles - orbicularis oris - which serve to constrict the vocal tract via closing (in a way which can be said to have a rotational component) or prolong it via protrusion. Additionally, they interact with the teeth or tongue for labiodental or labiolingual sounds. The orbicularis oris is innervated by the buccal branch of the facial nerve (CN VII).

The *velum* or soft palate is used in speech production to control the flow of air into the nasal cavity to make nasal sounds. It is moved through the activation of five muscles (Tensor veli palatini, Palatoglossus, Palatopharyngeus, Levator veli palatini, Musculus uvulae) which are innervated by the pharyngeal plexus via the vagus nerve (CN X) with the exception of the Tensor veli palatini which is innervated by the mandibular division of the trigeminal nerve (CN V3).

Additionally there are several structures of which the primary are the teeth and hard palate which are integral to the articulation of speech but which are static. The tongue interacts with both of these structures to produce, for example, dental and palatal consonants. Even without direct interaction with the tongue, the shape and position of these structures contribute to the shape of the vocal tract and consequently the acoustics of the sounds produced (Stone and Vatikiotis-Bateson, 1995).

Other structures involved in speech

The diaphragm, larynx, and pharynx are essential to speech production too, but do not contribute to the articulatory variations investigated here. Still, their precise coordination, especially in terms of the voicing originating in the larynx, is a prerequisite for the articulation that will be studied.

The *diaphragm* is the primary muscle which creates the pressure responsible for inhalation and exhalation. It must be coordinated with speech production, since the majority of human speech sounds are produced exclusively during exhalation. The diaphragm is an exception in the context of the mechanisms of speech production, since its descending control passes through lower motor neurons in the cervical vertebrae (C3-C5) of the spinal column via the corticospinal tract and then the phrenic nerve, the other structures involved in speech production having their lower motor neurons in the brainstem.

The role of the *larynx* in speech can be seen primarily as that of coordinating voicing, which itself only occurs in the presence of buildup of pressure caused by the diaphragm. There are six intrinsic and 11 extrinsic muscles, which must be coordinated with precise timing since speech sounds can be both voiced and unvoiced. The larynx is innervated by the vagus nerve (CN X) for both sensory and motor aspects. In particular the motor innervation is supplied by the recurrent laryngeal nerve except for the cricothyroid muscle which is innervated by the external branch of the superior laryngeal nerve. Sensory innervation to the subglottis is by the recurrent laryngeal nerve while the vocal fold's sensory innervation is supplied by the superior laryngeal nerve and that of the glottis and laryngeal vestibule is supplied by the internal branch of the superior laryngeal nerve (Sasaki and Weaver, 1997).

The *pharynx* is composed of three outer and three inner muscles all of which are innervated by the pharyngeal branch of the vagus nerve (CN X) except the stylopharyngeus muscle which is innervated by the glossopharyngeal nerve (CN IX). The pharynx is involved in swallowing, and the passage of both food and air. The configuration of the pharyngeal muscles as well as the shape and size of the three pharyngeal regions (nasal, oral, and laryngeal) contribute to the acoustic properties of speech.

The present work deals only with the muscles of the jaw, lips and tongue.

2.1.4 Motor commands

Upper motor neurons involved in speech are located in the primary motor cortex and their axons, which form the corticonuclear tract, synapse on lower motor neurons located in brainstem nuclei for all structures mentioned above except the diaphragm. The lower motor neurons of the diaphragm are located in the spinal column (C3-C5 in particular) and receive projections from UMNs whose axons form the corticospinal tract.

In the case of the tongue, alpha motor neurons in the hypoglossal nucleus (HGN) project onto the neuromuscular junction of tongue muscles through the hypoglossal nerve (CN XII) while gamma motor neurons similarly project onto intrafusal fibers of muscle spindles.

As mentioned above, the lower motor neurons activating muscles in the diaphragm are located in the spinal column. Their axons form part of the phrenic nerve and they also receive input from the reticular formation of the brainstem through the reticulospinal tract.

So, upper motor neurons located in the cortex do not directly influence muscles, but indirectly do so through the intermediary of either brainstem or spinal lower motor neurons. This point is essential since it is for this reason reflex arcs are effective even though they do not necessarily modify cortical activity and, when they do, they do so indirectly.

2.1.5 Between muscle commands and sensory feedback

Once a muscle is activated, it begins producing a force. This force interacts with connected tissue and skeletal structures and gives rise to movement. The resulting forces and movement of any articulator and its interactions with other structures are determined by biomechanical principles.

At any given point in time, the articulators will be in a given position and have a certain velocity depending on the circumstances. As a result of that configuration or movement, sensory fibers will respond and transmit proprioceptive information to both the brainstem and the cortex.

Additionally, the air moving through the vocal tract (and its excitation by the vocal folds in the case of voiced vowels and consonants) gives rise, through acoustic principles which are essentially consequences of laws of fluid dynamics, to sound. Sound is then perceived via the auditory system which includes the peripheral auditory system, the brainstem, thalamus, and auditory cortex.

Vibrations induced by the sound can also be perceived by other sensory organs sensitive to vibration, such as the skin.

2.1.6 Stretch reflex

As outlined above, sensory information of numerous types is available and potentially used as feedback in the execution and learning of movement. There are also numerous ways that information could be taken into account. One way is for it to be relayed to the cortex where it can be processed and influence the update of cortical output. While this can be a relatively slow process (i.e. longer than 50 ms), it can also occur through highly optimized cortical reflexes. Another, faster route, especially for proprioceptive information, is that of fast reflex arcs through monosynaptic connections in the spinal cord or brainstem nuclei.

The simplest reflex is the classical stretch reflex which is based on a muscle spindle's response to lengthening of a muscle fiber. The muscle spindle's sensitivity is determined by the corresponding gamma motor neuron's activity level. When the muscle length is increased, activity in the muscle spindle causes the lengthening muscle's alpha motor neuron to be activated via a monosynaptic connection. This alpha motor neuron's increased activity causes the muscle to activate further and resist the lengthening (Schmidt, 1988).

This monosynaptic stretch reflex loop is typically contained in the spinal cord or brainstem, depending on the location of the muscle's lower motor neuron. There are however other analogous reflex loops including the long-latency stretch reflex (LLSR) for which a cortical origin has been shown to be unlikely despite a seemingly

polysynaptic pathway (Pearce et al., 2003)) as well as a compensatory reflex which has been shown to be cortical (Gomi et al., 2002; Ito, Kimura, and Gomi, 2005). In the case of a transcortical reflex, the afferent information is relayed through a polysynaptic chain through the thalamus, somatosensory cortex, and motor cortex.

While both the standard stretch reflex and LLSR have been studied in the jaw (for example Cooker, Larson, and Luschei, 1980 and Pearce et al., 2003) where responses even include multi-articulator coordination (Gomi et al., 2002), evidence of a stretch reflex in the human tongue is lacking (Bratzlavsky and Vander Eecken, 1974; Neilson et al., 1979). There is evidence for the existence of sensory fibers, in particular, muscle spindles, in the human tongue (Kubota, Negishi, and Masegi, 1975; Cooper, 1953; Walker and Rajagopal, 1959; Sanders et al., 2013), even if to differing degrees in different locations (Sanders et al., 2013; Kubota, Negishi, and Masegi, 1975). For example, Sanders et al., 2013 claim the superior longitudinalis and transversalis muscles have the highest concentration of muscle spindles in the tongue.

A review of different types of reflex loops involved in speech, along with their latencies, will be presented in Chapter 4

2.1.7 Inseparability of speech, language, and cognition

As mentioned in the general motivation (Chapter 1), one interest of speech is its relation to language and cognition. But more than simply related, speech is an integral part of the phenomenon of speech-language-cognition. That is they are all components of the same system and must be understood in relation to one another.

In concrete terms, this means that speech cannot be studied or understood in total isolation from language or cognition. Indeed, speech is not the same as simple vocalization and although there may be correlates in animals, the highly structured, complex motor act with which the present work is concerned exists only in humans and therefore in the presence of certain linguistic and cognitive capacities. Speech is inherently influenced by the structure of language and cognition and this is reflected even at the neural and physiological levels.

Most relevant to the present work is the perspective that while the CNS is certainly specifying through various indirect and direct pathways activity levels of muscles involved in speech and thus the physical configuration of the speech apparatus, it is also concerned with the immediate sensory consequences of these actions, in terms of both proprioceptive and auditory modalities, as well as less immediate consequences at a cognitive and even social level of analysis.

2.2 Control - basic principles

While the neurobiological and peripheral mechanisms implementing the articulatory aspects of speech and their control have been outlined, the concept of control will now be formalized.

Control as intended in this work is simply behavior which has as its purpose the realization of certain goals. In the case of speech, as will be outlined in Chapter 3, these goals can be of numerous types including articulatory, acoustic, semantic, or even pragmatic.

Given the physiological complexity of the speech apparatus and its equally complex neural circuitry, the question arises whether the organization of the CNS and the control strategies used by it can be understood in terms of any higher level computational or algorithmic principles.

Much work has been done to study how the motor system controls movement and numerous important concepts have emerged, a few which will now be outlined and will help clarify the context of the work presented later.

2.2.1 Formulation of the control problem

While there are numerous related conceptual frameworks in which the control problem can be formulated, an attempt will be made at outlining the generic terminology necessary for the present work.

System/plant/process and state variables

First, there is the system to be controlled, which is often known as the plant or process in control theory. This notion corresponds to the physical components of the system which respond to control. In the case of generation of movement in biological systems, this corresponds to the notion of the periphery introduced above (see Section 2.1.1).

This system has a space of potential states, known as the *state space* each point of which corresponds to a different state of the system - for example the state of a system composed of a wheel could be described in terms of the current angle and rotational velocity of the wheel. With physical systems, the state space is often thought of in terms of the number of elements, number of dimensions and position and velocity vectors of these elements. However there may be constraints on individual elements or between groups of elements of the system such that not every point in a simple Euclidean state space composed of, for example, position and velocity vector components corresponds to a viable state of the system.

Control/input variables

The system then can be controlled through a number of control or input variables which are acted on by the controller. These do not necessarily correspond to the state variables of the underlying physical system, but can influence the temporal evolution of this system and therefore its corresponding state variables. For example, a control variable could be the force exerted by a force-generating component of the system (such as a muscle) or a parameter which modulates such a structure's functioning (for example an input current to a motor neuron, or, some proxy to a motor neuron's activity level including instantaneous firing rate).

Output variables

There are also output variables which correspond to a function of internal state variables. For example the position and velocity of a component could be converted into a joint angle which is the output variable. Often the output variables are those which are subject to control by the system; that is, a controller often has goals which are formulated in terms of output variables or functions of these variables. If the output variables are, for example, the joint angles of a finger, the controller could be interested in controlling those joint angles or the relative angle of the finger tip relative to the base of the finger. One can also think in terms of *goal variables*, which the relative angle of the finger tip to the base of the finger would be in this example.

Objective/cost function

There is then an objective function or cost function, which the controller uses to relate the current state of the system to potential goals or objectives it intends on achieving. The cost function conveys a penalty for performance that does not achieve a goal perfectly. Such a function can be used to drive online control or learning over time as well, a classic example being that of error backpropagation used to train a controller, including those based on artificial neural networks, to perform a task (Dreyfus, 1990). Since such formulations and their implementations have proven to be very powerful computationally, attempts have been made to study biologically plausible formulations which could be at work in the brain (O'Reilly, 1996).

Optimal control

Additionally, this cost function may be subject to minimization, which results in the notion of optimal control. A control input that when applied to a system minimizes the cost, i.e. maximizes the system's performance, is the optimal control policy and the controller which provides this control policy for any situation is an optimal controller.

Such a formulation of optimality has been widely studied in the field of motor control and, although not without controversy (Friston, 2011), has been used to explain patterns of movement observed in actions such as locomotion (Chow and Jacobson, 1971), jumping (Zomlefer, Levine, and Zajac, 1974; Zajac, Zomlefer, and Levine, 1981), reaching and grasping (Uno, Kawato, and Suzuki, 1989), and even speech (Perrier, Ma, and Payan, 2006).

2.2.2 Feedforward and feedback control

One of the major distinctions in models of motor control is that of whether feedback is taken into account (Loeb, Levine, and He, 1990). Indeed there has been much discussion about the merits and adequacy of different models of motor control of both types and especially given the reality of both biological and engineered systems being much more complex than these basic concepts, the present section is only intended to be a rapid overview.

Feedback control

Most generally and intuitively the problem of motor control is that of how a system in a certain state/situation can execute a movement which accomplishes a specified goal. An important part of this intuition is that there is an environment with which the system interacts and on which it depends for choosing contextually relevant actions. Even in the simplistic example of a task which consists of motion of an articulator, for example the hand, one can consider the periphery itself to be the relevant part of the environment.

It is important to note that in any case the action of the system depends on information about the current state of the system and the environment being presented to the controller. Since this information also contains clues to how well the action's execution is unfolding, it can be fed back into the controller and potentially result in updated control input. This information can be useful for correcting erroneous commands or reacting to circumstances in the environment which were not planned for. Controllers which make use of such information perform what is known as feedback or closed-loop control, since the loop is closed by providing information about the environment back to the controller.

Feedforward control

A controller which determines a sequence of commands based on an initial state and executes those commands without taking into account ongoing feedback is known as a feedforward controller.

It may seem, given the arguments for feedback control, that feedback control would be preferred to feedforward control in biological systems, especially where it is clearly advantageous to be able to update commands during execution.

The arguments in favor of feedforward control however are quite pertinent to the constraints on biological systems as well. For example, feedforward controllers can be simpler and therefore faster since they do not require the structure, complexity or time necessary to integrate feedback and they avoid the potential instability caused by integrating delayed feedback.

Feedforward vs. feedback control in motor control

In fact, due not just to the mathematical simplicity but also their efficacy in explaining certain aspects of the motor actions mentioned above - locomotion, jumping, reaching and grasping, and speech - the first models of optimal motor control, including all the works cited above, are feedforward models of motor execution.

Additional work by Nelson looked at skilled movements including arm movement during violin playing and jaw movement during speech and showed them similarly to be consistent with a feedforward control model trained using an optimality criterion (Nelson, 1983). Similarly, Flash and Hogan modeled kinematic aspects of multi-joint arm movement using a smoothness criterion, namely jerk minimization (Flash and Hogan, 1985). Still, even later the explanatory power of such feedforward models continued to be highlighted, the study of Harris and Wolpert

showing that stereotyped trajectories of both eye and arm movements can be explained in terms of signal-dependent noise and a minimization constraint on the variance of the final eye or arm position.

A second wave of models explaining similar phenomena recognized a role for the integration of feedback during movement execution. For example, Meyer et al. showed that rapid aimed movements can be explained in terms of stochastic optimized-submovements, where a secondary submovement, subsequent to a primary submovement, integrated visual and potentially kinesthetic feedback (Meyer et al., 1988), while Hoff showed that a single delayed-feedback model could account for stereotypical reaching movements (Hoff, 1992) and Kuo presented a model of human postural balance which integrated feedback in the form of state feedback (Kuo, 1995).

The success of these approaches and advances in control theory, led to the more generic formulation of motor control as a stochastic optimal feedback control problem (Todorov and Jordan, 2002; see below for further details), which gained support for its compatibility with neuroscientific data (Scott, 2004), and has inspired numerous lines of research in both the fields of motor control and robotics.

One specific work which builds on the framework proposed by Todorov and Jordan is that of Houde and Nagarajan who proposed state feedback control as a model of speech production (Houde and Nagarajan, 2011; see below for further details) wherein an optimal observer in the form of a Kalman filter - the same as that proposed by Todorov and Jordan in the case of a linear system³ - was used to integrate sensory - auditory, in their example - feedback in the estimation of the current state.

In any case, as stated by Loeb, Levine, and He, “most complex systems require a mixture of the two types of control”. Accordingly, numerous models exist which combine the two by nesting fast feedforward controllers inside of feedback controllers and it has even been suggested that the type of control implemented in the nervous system is best thought of as a “hybrid model of motor control in which pre-planning and feedback control are both used” (Desmurget and Grafton, 2000), a perspective advocated by Todorov (2004) as well.

Stochastic optimal feedback control

As mentioned above, an important work was the generic formulation of motor control as a stochastic optimal feedback control problem by Todorov and Jordan.

The essential premise is that by formulating the control problem in terms of stochastic dynamics, the knowledge the controller has about average trajectories

³This is an oversimplification made for the purpose of clarity. In reality, Todorov and Jordan propose a framework for dealing with non-linear systems and detail an example of a linear system in which a *separation principle* holds allowing the optimal control problem to be divided into a separate optimal observer and an optimal regulator. In many systems this principle does not hold. Nonetheless Houde and Nagarajan focus on the construction of an observer the optimality of which they extend in later work to non-linear systems using non-linear methods including the extended Kalman filter (EKF) and unscented Kalman filter (UKF). Their point is not however to construct an optimal controller but to demonstrate how such an observer could provide a way for the CNS to perform inference from all available sensory information, thus integrating it into a state estimation which is then made available to a controller, the details of which are not discussed.

allows it to 1) infer the state of the system from delayed and noisy sensory feedback, 2) estimate an expected cost to go from a given state to a goal state, 3) and react accordingly to deviations in the expected trajectory and its effect on expected cost, as soon as information about these deviations becomes available by integrating feedback of any type available. The advantage of doing so is described by Todorov and Jordan (2002, p. 1227) as follows:

"Whenever the task allows redundant solutions, movement duration exceeds the shortest sensorimotor delay, and either the initial state of the plant is uncertain or the consequences of the control signals are uncertain, optimal performance is achieved by a feedback control law that resolves redundancy moment-by-moment - using all available information to choose the best action under the circumstances. By postponing decisions regarding movement details until the last possible moment, this control law takes advantage of the opportunities for more successful task completion that are constantly created by unpredictable fluctuations away from the average trajectory."

Stochastic optimal feedback control thus offers itself as a computational description (in the sense of Marr (1982)) of how the CNS executes motor control. Most significant is its compatibility with and even ability to explain patterns of task-relevant variability (which will be discussed below in Section 2.2.4) as a natural consequence of optimality.

Even with their observations of its predictions' compatibility with observed data (Liu and Todorov, 2007) and efforts to clarify its potential implementation in a neural system (Scott, 2004), it remains a high-level description - a computational one in the sense of Marr - and conceptual framework for how the neural control of movement might be implemented.

There are several issues with this approach however. One, as raised by Scott (2004) is that the biomechanics of realistic systems introduce a great deal of complexity, with the physics of even a two-joint system being complex. More precisely, the perspective expressed by Scott (2004, p. 543) is that:

"There might be many ways to use optimal feedback control to guide neurophysiological research, although several challenges remain. First, the mathematics that is required to identify optimal feedback control laws is extremely challenging even for the simplest of linear systems. This limits the conditions under which formal solutions can be used to predict the properties of an optimal feedback controller Further theoretical work is also required to break down the processes of optimal feedback control into more biologically plausible algorithms and processes [(Marr, 1982)] that can help to guide experimental studies. However, it is unlikely that such efforts will attain the level of detail that is present in oculomotor models of brainstem circuitry."

This question of the biomechanical aspects was even mentioned by Todorov and Jordan as one point, along with the "computational mechanisms and learning algorithms available to the nervous system [which] must impose information-processing constraints", requiring further study (Todorov and Jordan, 2002, p. 1228).

Thus, stochastic optimal feedback control is a promising computational description of the processes of motor control, one advantage of which will be discussed further below, but whose complexity in terms of implementation and formal solutions provide interesting challenges, especially in the case of complex biomechanical systems, such as those involved in speech. Nonetheless, the theory provides a useful conceptual framework in which to attempt to understand the functioning of neural control of complex sensorimotor phenomena such as speech.

Moreover the present work will be undertaken precisely in the interest of better understanding the challenges posed by the complexity of the biomechanical system to be controlled.

State feedback control

Another specific concept of notable importance that has been adapted not just to motor control, in particular its adaptation to speech motor control is that of state feedback control (Houde and Nagarajan, 2011). As mentioned above, the basic premise of state feedback is that the variation over a recent time windows of the state of a system carries information about its history as well as future evolution and thus provides a controller with the possibility to perform optimally. In particular, state feedback is combined with state estimation to reconstruct the state from a limited subset of observable variables.

Specifically the framework proposed by Houde and Nagarajan involved the estimation of the underlying system state from information available to the sensory systems, namely auditory information in their first iteration. Importantly, this model supposes the ability to make predictions about the sensory states based on descending commands, i.e. through a forward model (see below for further discussion on internal models), and then make corrections to the state estimate of the system based on discrepancies between the prediction and the observed values. This is accomplished, at least for linear systems, through a Kalman Filter (KF) and through extensions such as the Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF) for non-linear systems (Parrell et al., 2017).

Just as with stochastic optimal feedback control, state feedback control offers a compelling conceptual framework for understanding how sensory information could be integrated into a control process, especially given the noisy nature of sensory information and the need for it to inform the control system of the current state not just of the sensory feedback, but in terms of its relevance to the state of the system as a whole. In particular, the work of Houde and Nagarajan, including later attempts at using more complex models of speech production and even multiple, hierarchical types of feedback (Ramanarayanan et al., 2016; Parrell et al., 2017), provides interesting insight into the potential neural mechanisms of integrating sensory feedback into the control of speech.

2.2.3 Internal models

In addition to the high-level distinction between feedback and feedforward control, another important question revolves around the information or knowledge the

controller has about the controlled system. This knowledge can take the form of an internal model, the existence of which has been debated. Kawato, Furukawa, and Suzuki proposed the concept of internal models as a strategy for overcoming limitations of control relying on feedback when that feedback is noisy and delayed (Kawato, Furukawa, and Suzuki, 1987).

Forward model

First, the controller could have up to complete knowledge of the controlled system's dynamics, which would essentially allow the controller to simulate the results of any command input. In practice, the model is never assumed to be a perfect reconstruction/simulation of the system, and there is always the possibility that a forward model's predictions do not hold out, even for reasons such as un-modeled external forces/perturbations.

The utility of a forward model can be in the prediction of the consequences of a control sequence in terms of output variables against which observed consequences can be compared to calculate a sensory prediction error (SPE). Indeed is an important concept implicated in many processes of sensorimotor learning, including learning of control sequences, as well as in terms of fine-tuning the forward model itself (Shadmehr, Smith, and Krakauer, 2010). Additionally, as mentioned above in the context of feedback control, forward models and their ability to predict sensory consequences can also be used to estimate the state of the underlying system (Shadmehr, Smith, and Krakauer, 2010; Todorov and Jordan, 2002; Houde and Nagarajan, 2011).

While the information provided from a forward model could be considered a form of internal feedback and used accordingly in updating commands or driving learning, the question of the existence of a forward model can be considered independently from the mechanisms of online control of movement; for example the role of such feedback could be limited to offline learning, leaving online control to be executed through a feedforward controller which does not make use of a forward model. It has however been argued that fast internal feedback of this type is used in motor control (Miall and Wolpert, 1996).

Inverse model

A controller could also have an internal model which maps a desired state of the system (or some approximation of it or its output variables) to commands which would bring it into that state. Such a model could also be sensitive to the current state of the system. This is not trivial though since in many cases the forward model or approximation thereof is not invertible for such a reason as there being multiple ways (See below, section 2.2.4 for a discussion of this) or no way to arrive at a desired state from the current state. An inverse model could potentially choose amongst multiple options (even potentially implementing some form of optimality constraint) or choose a best approximation of a desired but unreachable state (Jordan and Rumelhart, 1992; Kawato et al., 1990).

More commonly an internal model is framed in terms of either product of learning over repeated executions of a movement or repeated simulations of movement using an accurate forward model.

Internal dynamical models

Various control schemes have been proposed which implicate representations of the dynamics of the underlying controlled system in both forward and inverse models. In many of these models however the question of complexity arises, given that the mechanical complexity of the system would require large amounts of detail in order for general dynamic models to be learned. While originally discussing dynamical internal models in the context of trajectory planning (Kawato, Furukawa, and Suzuki, 1987), later revisions have included optimality constraints and a relaxation of the notion of the entire trajectory being planned - for example Kawato et al. suggest that just key points of a trajectory could be planned, thus relaxing the requirements of a dynamical internal model to be able to accurately characterize the system's dynamics in a general way (Kawato et al., 1990).

Indeed the move has been towards models which allow for differing degrees of locality in any representation of dynamics. First, Wolpert and Kawato proposed use of modules of paired forward and inverse models among which the most appropriate would be chosen based on its ability to model the dynamics of the current use of the system and its interaction with the environment (Wolpert and Kawato, 1998). It was then proposed that local dynamical models informed by experience could form the basis of simple, generalized models constructed through some sort of interpolation process (Mattar and Ostry, 2010). Still, the degree to which knowledge of the dynamics of the underlying system must be represented is not clear (Perrier, 2012).

Internal models in speech

One area where the necessity of internal dynamical models is least evident is that of speech. While limited in number, studies have failed to establish a clear use of dynamical models in speech planning or learning and have even shown, in the case of the jaw, that intrinsic mechanical characteristics, namely stiffness, which does not require dynamical representation can explain observed phenomena (Shiller, Ostry, and Gribble, 1999; Shiller, Laboissière, and Ostry, 2002; Tremblay, Houle, and Ostry, 2008; Perrier and Fuchs, 2008; Buchaillard, Perrier, and Payan, 2009).

Indeed Perrier claims 2012 that intrinsic dynamics of speech articulators give rise to the patterns which could otherwise be compatible with an optimal control process requiring representation of dynamics and that alternative explanations involving either equilibrium point dynamics (Feldman and Levin, 1995), memory (Ganesh et al., 2010), or active inference processes (Friston, 2011).

In any case, several open questions remain about the existence, type, and granularity of internal models in speech motor control. The most pertinent to the present work are the questions of the inverse models and their potential granularity in the context of speech, which will be the subject of the study presented in Chapter 3.

2.2.4 Motor Equivalence

A recurring theme in the context of motor control is that of motor equivalence. Motor equivalence is the capacity to accomplish the same motor task with different control strategies, associated with different body positions, or the use of different effectors, or different strategies of coordination between some specific effectors. This capacity is essential for the Central Nervous System to be able to deal with changes in external constraints applied to the body before or during the execution of a movement, or with perturbations of the normal achievement of the task, for example, due to pathologies.

In the context of speech, numerous motor equivalence phenomena have been discussed in the literature (Hughes and Abbs, 1976; Perkell et al., 1993; Savariaux, Perrier, and Orliaguet, 1995) and they are indeed crucial for speech production, since they enable speakers to produce intelligible speech under a broad spectrum of conditions, including various phonetic contexts, speaking with a pen in the mouth, speaking while eating, or speaking while running, etc. Motor equivalence is at the basis of the phenomenon of coarticulation that makes speech segments highly variable even for an individual speaker depending on the preceding and following sounds (Kent, 1977; Whalen, 1990; Vaxelaire, 1993; Vaxelaire et al., 2010) and on the speaking rate (Matthies et al., 2001).

One consequence of motor equivalence is seen in the patterns of variability that have been observed as early as Bernstein (Bernstein, 1966) in various motor tasks. The hallmark of this pattern is that variability is highly structured, specifically in a way that enables the control of its interference with task-dependent objectives. An example is that of a metal smith (see Figure 2.2) whose task it is to hit a precise target. The trajectory that the hammer might take to arrive at this point (and similarly the specific motor commands) could vary considerably, but that variability does not interfere with reaching the target with a high precision and might even reflect a process whose primary effect is to increase this precision.

Stochastic optimal feedback control and motor equivalence

As mentioned above stochastic optimal feedback control is claimed to provide a generic framework in which to study motor control phenomena. In particular, if effort is required to compensate for a perturbation and is penalized in the cost function, then an optimal controller will only compensate what it is minimally necessary to compensate and will accept variations which do not increase the estimated cost of achieving the desired goal state.

Stochastic optimal feedback control thus provides a computational solution to the motor equivalence problem. It nevertheless leaves unspecified how such a controller or, in particular, how the internal forward models containing the implicit or explicit knowledge of dynamics of the controlled system, usually modeled as some sort of Kalman filter (see Todorov, 2004), might be implemented in the combination of the CNS and skeletomuscular or orofacial systems.



FIGURE 2.2: An illustrative example of the concept of motor equivalence. A metal smith seeks to hit a precise target, but might arrive at that target through numerous different trajectories.

Uncontrolled manifolds and motor equivalence

One potential algorithmic/representational implementation of a solution to the motor equivalence problem that would be entirely compatible with the computational principles of the solution proposed in the context of stochastic optimal feedback control is that of Uncontrolled Manifolds. Even without formalizing them in order to implement such a controller, they provide a solution to the motor equivalence problem in their own right and will be discussed in detail in the next section and studied further in Chapter 3.

2.3 Uncontrolled Manifolds

The Uncontrolled Manifold (UCM) concept was proposed by Scholz and Schöner (1999) as a way to address the question of motor equivalence, in particular, as a way to understand the structure of the patterns of variability mentioned above. Working in the context of arm movement, Scholz and Schöner defined the uncontrolled manifold as “the set of [shoulder, elbow, and wrist] joint configuration[s] that leaves the end-effector position invariant”.

Specifically, as the name suggests, the uncontrolled manifold is defined in contrast to a “controlled manifold”, known as the *orthogonal space* where variability is precisely controlled. The uncontrolled manifold is the set of configurations that tolerate variability. See Figure 2.3 for a simple explanatory example of this concept.

Up until now the use of uncontrolled manifolds have been as an analytical technique, in particular, a hypothesis against which observed patterns of variability have been tested. While UCMs were originally discussed in the context of arm motor control, they have been evoked and used to explain the patterns of variability observed in various motor tasks, including bimanual pointing (Domkin et al., 2005), standing up from the seated position (Reisman, Scholz, and Schöner, 2002; Scholz and Schöner, 1999), pistol-drawing (Scholz, Schöner, and Latash, 2000), and walking where it has been used to differentiate the control strategies employed in walking by healthy preadolescents compared to those with Down Syndrome (Black et al., 2007) and between healthy subjects and a stroke patient (Papi, Rowe, and Pomeroy, 2015).

Although Schöner et al. proposed in 2008 that uncontrolled manifolds could be useful in explaining motor equivalence in speech, up until present such research has not been done. Still, recent results in the use of uncontrolled manifolds in sequential pointing tasks motivate even further the potential utility of this concept in explaining not just motor equivalence but in particular how it might give rise to coarticulation effects (Hansen et al., 2015). A goal of the present work has been to study their appropriateness in the context of speech motor control. Chapter 3 presents a study with this as its primary objective.

2.3.1 UCMs and Linearized UCMs

While the uncontrolled manifold is the set of all points in the command space that give rise to the same output as some reference point, the difficulty of representing

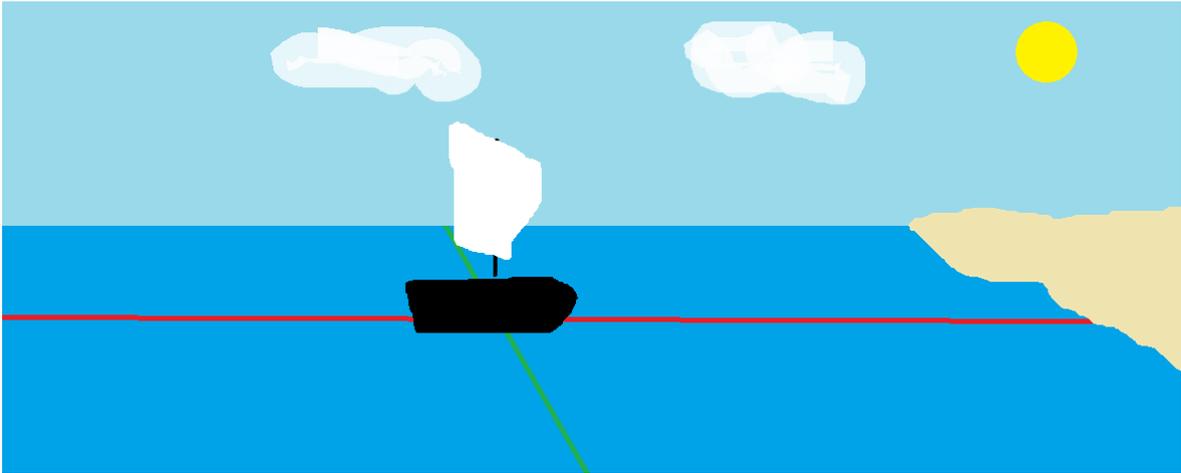


FIGURE 2.3: An explanatory example of the concept of uncontrolled manifold. If the goal of someone on the boat is to reach the shore, then they will be more interested in movements (wind or current) that move the boat closer or further from the shore, while movements parallel to the shore can be ignored. Counteracting these lateral movements would be energy wasted. The direction parallel to the shore where variability is tolerated (the green line) is the uncontrolled manifold, while the direction perpendicular to the shore (the red line) is the controlled manifold in this example. Note that the movement takes place in two dimensions and the sum of the dimensions of the uncontrolled and controlled manifolds is also two.

this space leads to the linearized uncontrolled manifold (LUCM)⁴ around each reference point being used in practice instead.

The LUCM is defined as the null space of the Jacobian matrix of the mapping, M , from command space, Λ , to output space, S :

$$M : \Lambda \mapsto S \quad (2.1)$$

where $\Lambda = [\lambda_1 \dots \lambda_a]$ is a vector in the a -dimensional command space and is composed of the different commands and $S = [s_1 \dots s_b]$ is a vector in the b -dimensional space of sensory consequences to be controlled. See Chapter 3 for a description of this mapping in the context of the biomechanical model introduced below.

Jacobian matrix

The Jacobian matrix represents the sensitivity of the output variables to local changes to the commands. The Jacobian of the vector-valued mapping M defined above is

⁴Note on terminology: while in other works (including Scholz and Schöner, 1999) the term uncontrolled manifold is taken to refer to both the non-linear and linearized versions, for sake of clarity we introduce here and will use consistently the term *linearized uncontrolled manifold* for any linear structure which serves to locally approximate a potentially non-linear uncontrolled manifold; this is also for terminological precision since if we were dealing with truly linear structures instead of linear approximations of non-linear ones then the proper term would be "uncontrolled linear subspace", although the LUCM is itself, strictly speaking, a linear subspace.

the $a \times b$ matrix each element i, j of which is the partial derivative of the respective elements of S with respect to the elements of Λ , $\frac{\partial s_i}{\partial \lambda_j}$, which is the amount that the output varies for a change in the input value. It is a way to capture the sensitivity of the output to changes in the input.

$$J_M = \begin{bmatrix} \frac{\partial s_1}{\partial \lambda_1} & \cdots & \frac{\partial s_1}{\partial \lambda_a} \\ \vdots & \cdots & \vdots \\ \frac{\partial s_b}{\partial \lambda_1} & \cdots & \frac{\partial s_b}{\partial \lambda_a} \end{bmatrix} \quad (2.2)$$

2.3.2 Uncontrolled manifolds in relation to inverse models

Beyond the question of their appropriateness in the context of speech, another goal of this work is to use speech as a case study on the theoretical question of UCMs as an implementation of an inverse model.

As mentioned above, an inverse model potentially represents the physical behavior, including the dynamics, of the physical system, but might not do so explicitly. Since the uncontrolled manifold contains information about the structure of the physical system being controlled, it can be seen as a form of an internal model. However, especially in the present work where control variables are formulated in terms of the lambda-model and equilibrium point hypothesis (see below for details about this model), the uncontrolled manifold does not contain detailed information about the dynamics and would therefore not be considered a dynamical inverse model.

Still, as a type of inverse model, the question of accuracy and comprehensiveness necessary for acceptable performance could also be posed and indeed will be studied in Chapter 3.

2.3.3 Uncontrolled manifolds and stochastic optimal feedback control

The framework proposed by Todorov and Jordan and described above in section 2.2.1 may seem to be an alternative to uncontrolled manifolds, but in reality the two are compatible, since one could see UCMs as one potential algorithmic/representational implementation of an optimal feedback controller. That is, even if combination of the CNS and skeletomuscular or orofacial systems can be thought of as implementing a stochastic optimal feedback controller, that does not exclude the possibility that the representational form of this process be of a type along the lines of uncontrolled manifolds.

2.4 Biomechanical modeling of speech motor control

Models of the mechanisms of speech are not new. Indeed models which produce speech-like sounds are attested to even as early as the 18th century (Kempelen, Fügler, and Mansfeld, 1791). While models of the acoustic and articulatory aspects of human speech have a long history (e.g. Coker and Fujimura, 1966; Coker, 1976)

and are still by far the most common models used in speech production research (Tourville and Guenther, 2011; Maeda, 1990; Iskarous et al., 2003; Saltzman and Munhall, 1989), models which incorporate information regarding the biomechanics of human speech, in particular of the muscles of the tongue, are relatively recent (beginning in the 1970s with Perkell, 1974 and Kiritani et al., 1975) and currently represent a minority of models used in speech motor control research.

After these early models, there was interest in studying the biomechanics of different articulators involved in speech including the tongue (e.g. Hashimoto and Suga, 1986; Otten, 1987; Wilhelms-Tricarico, 1995). This body of work eventually evolved into models of the tongue that were capable of modeling the movement these articulators in the context of speech (Laboissière, Ostry, and Feldman, 1996; Payan and Perrier, 1997; Sanguineti, Laboissière, and Payan, 1997; Sanguineti, Laboissière, and Ostry, 1998).

Of particular importance among this work is the 2D biomechanical tongue model of Yohan Payan and Pascal Perrier (1997), which was later enhanced with additional features including parameters allowing for different levels of jaw opening and lip protrusion and rounding, although these latter aspects are not modeled biomechanically (Zandipour, 2006; Zandipour et al., 2006; Perrier et al., 2003; Perrier and Winkler, 2015). This original 2D work has served as the basis for additional models including three-dimensional models (e.g. Gerard et al., 2003; Buchaillard, Perrier, and Payan, 2009; Rohan et al., 2014), which have been used, among other things, in an implementation in the ArtiSynth project (Vogt et al., 2006; Lloyd, Stavness, and Fels, 2012).

The present work makes use of the 2D biomechanical tongue and vocal tract model with lip and jaw parameters originally described in Perrier et al. (2003). Despite the lesser degree of realism due to this model being two-dimensional, its use is justified in several ways. First, the complexity of this model is sufficient for addressing the phenomena of interest. Second, its extensive use (for example in Perrier, Ma, and Payan, 2006) has led to its accuracy in terms of biomechanical properties being well-accepted. Third, its computational speed, especially relative to the most realistic 3D models, allows us to easily run the numerous simulations necessary for the studies that will be described.

2.4.1 Tongue, jaw, and lip model

The 2D model of the vocal tract used in the work presented here includes a 2D biomechanical finite element model of the muscles of the tongue (Perrier et al., 2003; Payan and Perrier, 1997), and 2D geometrical models of the lips and jaw (Zandipour et al., 2006) which together simulate the tongue-jaw-lip configuration due to the application of 10 motor commands and the resulting acoustic characteristics of the vocal tract, including formants and phoneme identity (Badin and Fant, 1984).

The basic component processes which make up the model are depicted in Figure 2.4 and each described below.

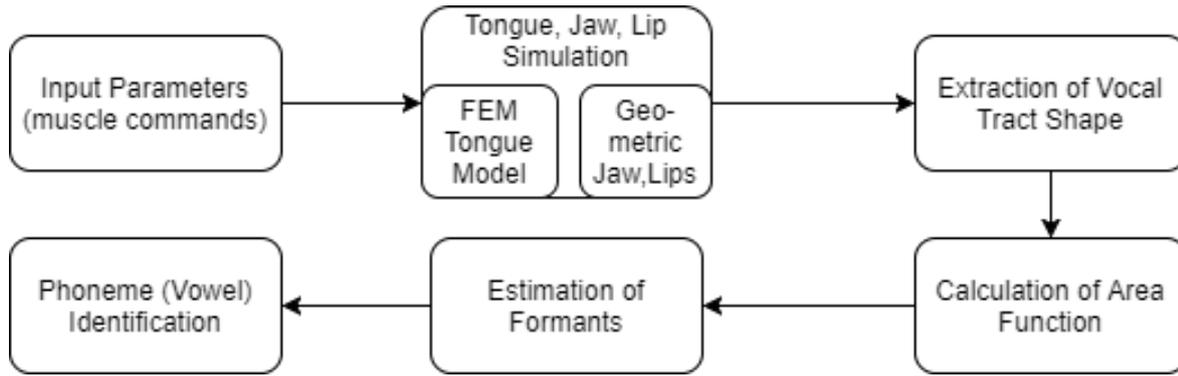


FIGURE 2.4: Block diagram depicting the usage of the biomechanical model and its component processes.

Biomechanical tongue model

The biomechanical tongue model is a 221-node finite element mesh in which the 7 primary intrinsic and extrinsic muscles of the human tongue are modeled. These muscle, which significantly influence the tongue shape in the mid-sagittal plane, are the Posterior and Anterior Genioglossus, the Hyoglossus, the Styloglossus, the Inferior Longitudinalis, Superior Longitudinalis, and the Verticalis (see Figure 2.5). In practice, the Superior Longitudinalis was not used in this study since it often results in numerical instabilities which would have posed a challenge to the extensive use of the model that has been required for the purposes of this work.

λ -model and threshold parameters

Forces corresponding to the contraction of muscle fibers are applied to appropriate nodes while overall force of a muscle fiber is calculated based on the λ -model proposed by Feldman in the context of the Equilibrium Point Hypothesis for motor control (Feldman, 1966; Feldman and Levin, 1995).

According to this model, muscle force, F , is calculated as a function of muscle activation level

$$F = \rho(e^{cA} - 1) \quad (2.3)$$

where A is the activation level of the muscle, ρ is a parameter specifying the amount of force it can generate (and is related to the muscle fiber's cross-section area), and c is the feedback gain parameter.

Muscle activation, A , in turn depends on the difference between the current length L of a given muscle fiber and its corresponding lambda threshold parameter λ .

$$A = L - \lambda + \mu\dot{L} \quad (2.4)$$

where μ is a parameter which determines the amount of damping of the reflex.

Notably, force is only generated (ignoring the $\mu\dot{L}$ component which is zero for any settled configuration) when the threshold parameter (λ) is smaller than the muscle fiber's current length (L), which corresponds to the muscle activation level being greater than 0. Thus, force is generated by a muscle only when the muscle length (L)

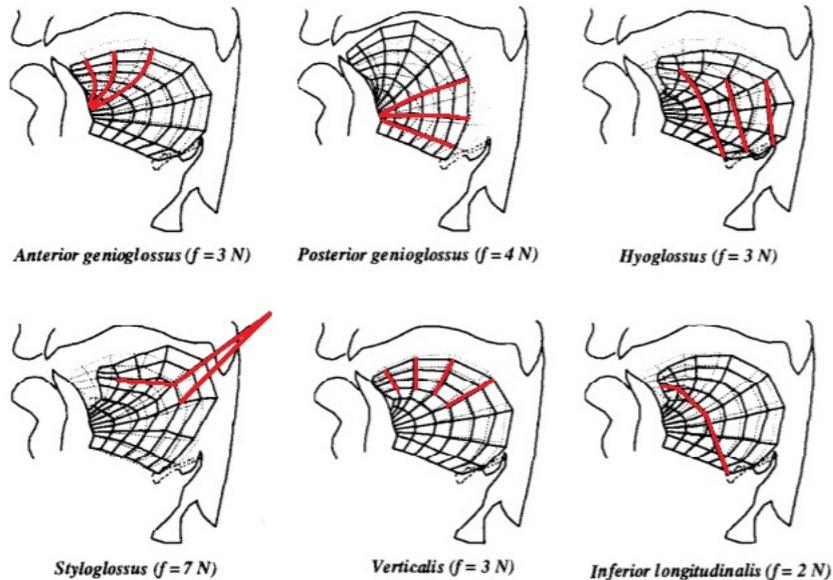


FIGURE 2.5: Depictions of the different muscles modeled in the biomechanical tongue model and the effects of their individual activation. Reproduce from Payan and Perrier, 1997. © 1997 Elsevier BV.

is *above* the certain threshold specified by λ . Also, for any given set of lambda threshold values and a constant load, there will be a unique equilibrium configuration of the tongue which does not depend on its starting position (ignoring numerical effects); for this reason the lambda threshold values are taken as the commands of the muscle fibers of the tongue.

This model was originally designed to account for the role of the muscle spindles and the stretch reflex (see above Sections 2.1.5 and 2.1.6) in posture control. However, in response to criticism of the generality of the model, in particular in the case of eye motor control in which the involvement of muscle spindles is excluded, Feldman and Levin (1995) suggested that the model can account for any kind of reflex or combination of reflex involving the motoneuron pool in the spinal cord or the brainstem. In the case of grip-movements with the hand, Pilon, De Serres, and Feldman (2007) have also suggested that “pressure-sensitive receptors in the finger pads” contribute to the modulation of the reflex activity accounted for in the model. It will be seen below that in the case of orofacial motor control it will be assumed that the reflex activity included in the model could be associated with cutaneous receptors, mechanoreceptors, or tactile receptors.

Geometric model of lips and jaw

Besides the lambda threshold values which serve as tongue muscle commands, additional commands control the lip protrusion, lip rounding, and the jaw aperture of the model by specifying their target positions (See Figure 2.6).

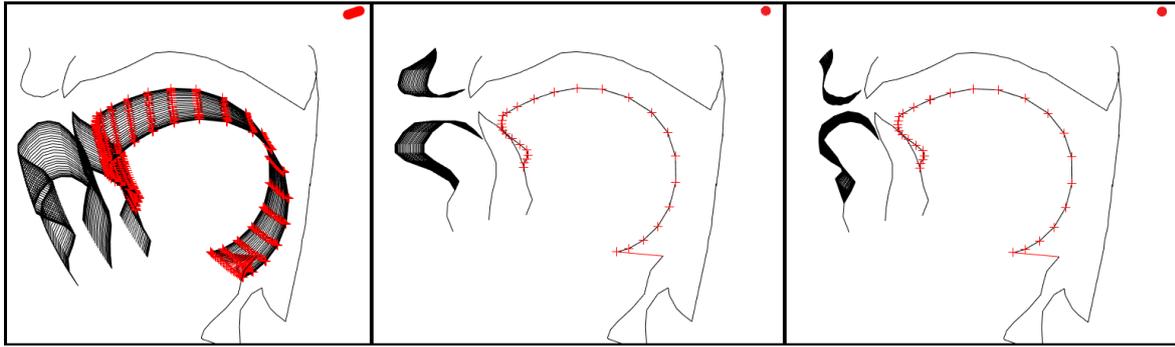


FIGURE 2.6: Depictions of the effect of jaw and lip parameters: jaw rotation (left), lip protrusion (center), and lip rotation (right). In all three the black lines are the outlines of the articulators of the vocal tract, the red dots along the tongue are the surface nodes of the finite element model of the tongue, and the red dots in the upper right corners are the condyle, i.e. the point around which the jaw rotates. Notice that with jaw rotation (opening) this point moves forward slightly due to protrusion of the mandible.

TABLE 2.1: List of motor command variables in biomechanical model

Motor Command	Units	Dynamics
Posterior genioglossus (GGP)	mm	EP
Anterior genioglossus (GGA)	mm	EP
Hyoglossus (Hyo)	mm	EP
Styloglossus (Sty)	mm	EP
Verticalis (Vert)	mm	EP
Inferior longitudinalis (IL)	mm	EP
Jaw opening	degrees	set point
Lip rounding	degrees	set point
Lip protrusion	10xmm	set point

Contrary to the tongue, which is the focus of the model, the jaw and the lips are not biomechanically modeled. They were included in the model in order (1) to realistically account for the influence of the jaw position on the tongue position, and (2) to correctly shape the lips at the external extremity of the vocal tract. Thus the jaw and the lips are modelled as kinematic parameters (jaw height, lip protrusion, and lip aperture) whose variation over time corresponds to half the period of a sinusoid from the initial and final positions. This second-order account of articulatory trajectories is a rough approximation of experimental trajectories of orofacial articulators during speech production that has been used in different studies (see among many others Munhall, Ostry, and Parush, 1985; Kelso et al., 1985; Ostry, Cooke, and Munhall, 1987).

The characteristics of the 9 control parameters are listed in Table 2.1.

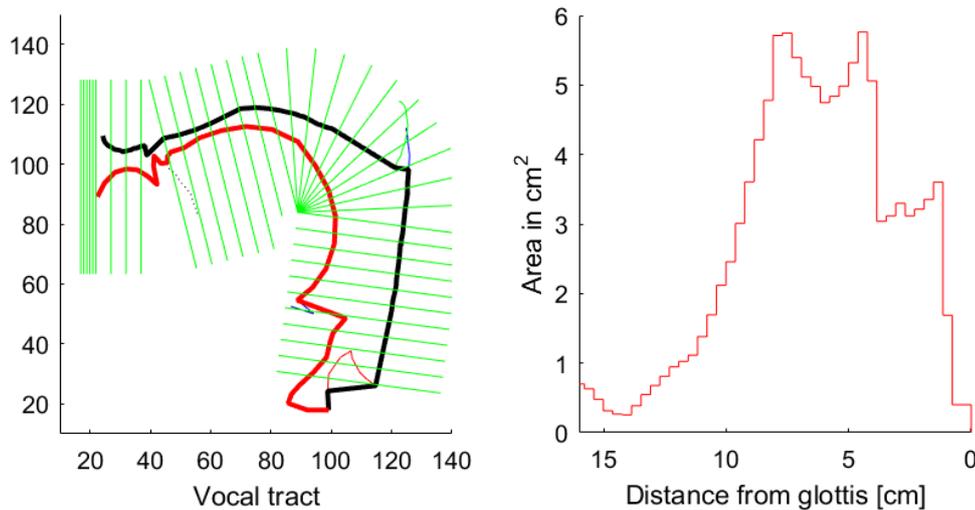


FIGURE 2.7: Left: the outlines of the vocal tract as extracted from a specific configuration of the tongue, lips, and jaw are depicted in (in black and red lines), while the grid which is used to extract the area function is superimposed and shown in green. Right: the extracted area function as estimated from the vocal tract depicted on the left.

Time parameters

In addition to the command parameters, the model also includes temporal parameters which determine the time over which the commands change from the default resting values to those specified and for how long the simulation should continue with these commands held at their target values. For example, the command-transition time could be specified as 150 ms while the command-hold time could be 300 ms, which would result in a total of 450 ms of model-time being simulated.

Area function extraction

At any point in the simulation, the 2D shape of the vocal tract formed by the current configuration of the tongue, lips, and jaw can be extracted and used to estimate the volume of the vocal tract through what is called the area function (Perrier, Boë, and Sock, 1992; Figure 2.7), which is simply a representation of the area of the vocal tract in relation to the distance from the glottis.

Formants and Vowels

A *formant* is a peak on the spectral envelope of a vowel sound, which corresponds to a resonance frequency of the vocal tract. They are usually called F1, F2, F3, etc along the increasing axis of frequencies. They are distinct from the fundamental frequency, usually designated as F0, which corresponds to the vibration frequency of the vocal folds in voiced sounds, and is related to the perceived pitch of speech. Multiples of the fundamental frequencies are called harmonics. The energy of these harmonics is amplified in the neighborhood of the formant frequencies, and this gives rise to the timbre that is characteristic of each vowel.

Vowels are defined as sounds which are produced without constriction of the vocal tract which plays the role of an acoustic resonator. The spectral characteristics, in particular the formants, of this resonator are closely related to the relevant articulatory features of vowel production, namely how open the vowel tract is due to jaw height or tongue height, the position along the vocal tract where there is maximal constriction caused by the tongue, and the configuration of the lips. The first two formants can be roughly associated respectively with aperture of the vocal tract and frontness of the constriction, with open vowels having higher F1 and front vowels having higher F2 values. Figure 2.8 depicts schematically the relation between articulatory and acoustic properties of vowels.

According to Figure 2.8 and Figure 2.10, vowels seem to be uniquely defined by their two formants (Schwartz et al., 1997; Carlson, Fant, and Granström, 1974).

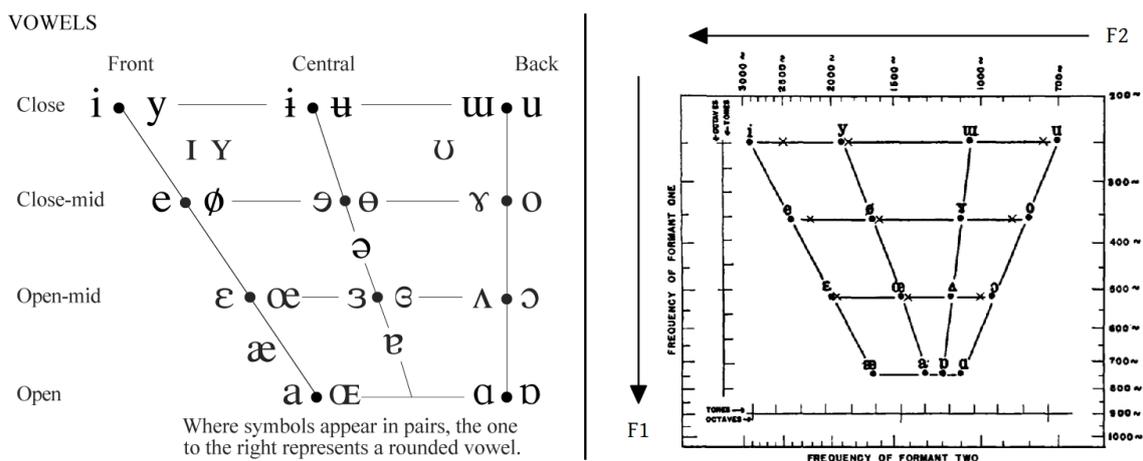


FIGURE 2.8: (Left) IPA Vowel Chart⁵ depicting vowels based on articulatory characteristics of openness, frontness, and rounding. (Right) A schematic vowel triangle based on formants.

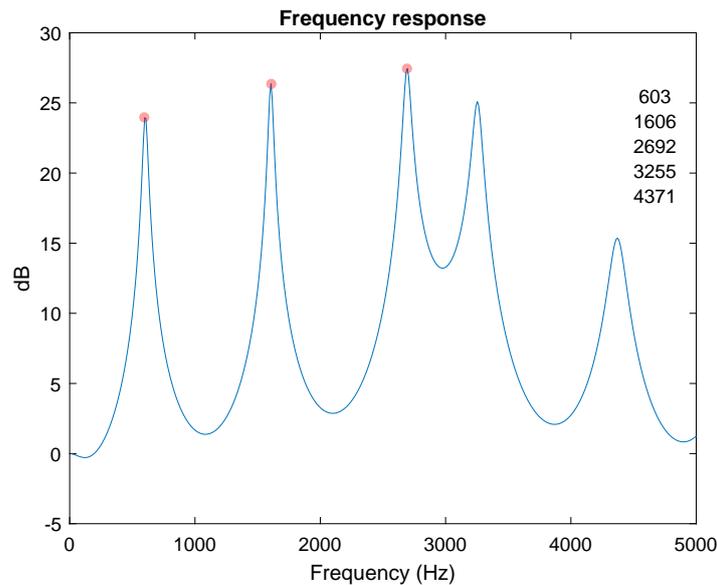


FIGURE 2.9: Example of frequency response resulting from acoustic analysis of the area function of a vowel /a/. The values of the spectral peaks are shown and the first three - F1, F2, F3 - are indicated in red.

Acoustic analysis

The area function can then be used with a multiple-segment tube acoustic model to estimate the transfer function that represents the effect of the series of tubes that approximate the vocal tract on the spectral energy of an acoustic source (Figure 2.9).

Since the first three formants characterize the vocalic phonemes of a language (Figure 2.10), these are extracted from the frequency response (Badin and Fant, 1984) and serve as the output variables of the biomechanical model.

Vowel identification

The formants extracted from the transfer function can then be used to identify which, if any, French oral vowel they correspond to. As mentioned above, vowels are characterized by their first three formants. Perception of vowels is not limited to a specific formant value, but a set of values which can be described as an ellipsoid in the 3-dimensional F1-F2-F3 space, assuming a Gaussian distribution in each of the planes of the F1-F2-F3 space.

The process by which vowel labels are found is then just a verification of which, if any, reference ellipsoid the formants fall within. Figure 2.10 depicts the projection of the reference ellipsoids of the 10 French oral vowels onto the F1-F2 and F3-F2 planes. The parameters of these reference ellipsoids were compiled from different works in which the average vowels' characteristics as well as permissible variable were

⁵Left: IPA Chart, <http://www.internationalphoneticassociation.org/content/ipa-chart> available under a Creative Commons Attribution-Sharealike 3.0 Unported License. ©2015 International Phonetic Association; Right: Vowel triangle with formant values, adapted from Delattre et al. (1952).

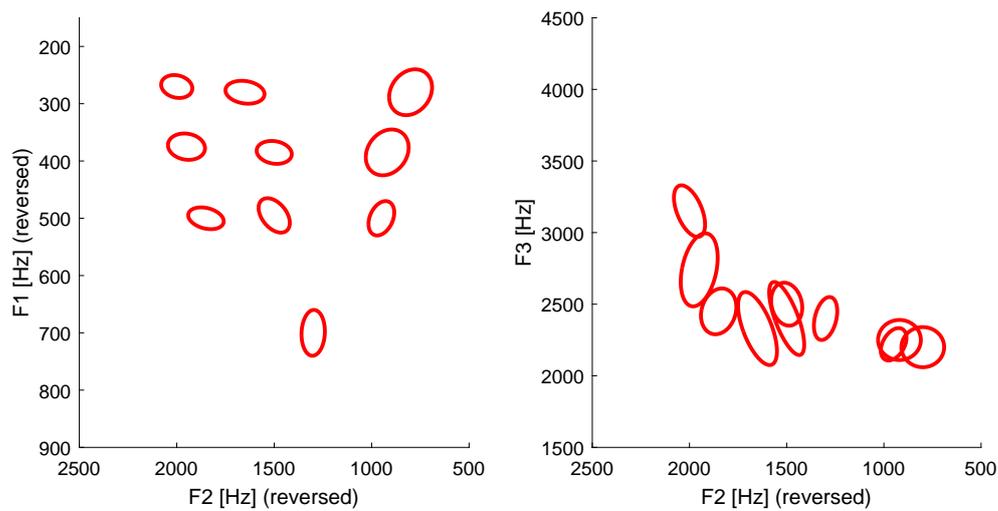


FIGURE 2.10: Formant space with regions corresponding to French oral vocalic phonemes in F1-F2 (left) and F2-F3 (right) planes. As can be seen in the F1-F2 plane on the left, position in these two dimensions uniquely identifies French vocalic phonemes, but, as shown on the right, F3 must fall within specific, although not unique, ranges for these vowels.

reported (Calliope, 1989; Ménard, 2002; Robert-Ribes, 1995). The specific algorithm used to detect a match was adapted from tools developed by Ralf Winkler in the context of the work presented by Winkler, Ma, and Perrier (2010).

It should be noted that the choice of French vowels is just used as an example partitioning of the formant space.

While the output variables can be calculated at any point in the simulation, the standard use of the model involves applying the commands, allowing the model to settle into its equilibrium configuration and then obtaining the formant values. The biomechanical model in this sense can be thought of as a mapping between input motor command parameters and output acoustic values, with an optional vowel label. It will be used in this way in the studies presented in Chapter 3 and Chapter 4 although in the latter it will be modified to simulate an external perturbation as well.

Chapter 3

Uncontrolled manifolds in speech motor control

In this chapter an extension of the concept of uncontrolled manifolds to speech will be presented, including a discussion of the necessity of considering the hierarchical nature of speech in doing so, as well as a proposal for a specific representation for the case of what will be introduced and referred to as the acoustic UCM.

An attempt will then be made to experimentally validate the claimed utility of UCMs in the context of speech motor control through a study in which they are used in conjunction with the biomechanical model described in Chapter 2.

3.1 Uncontrolled manifolds in speech

For a given mapping of commands to sensory consequences and a given point in command space, the original notion of uncontrolled manifold is simply that set of possible values in command space which result in the same sensory consequences.

As mentioned in Section 2.3, while useful in numerous motor control tasks, the question arises as to whether this is the case for speech. A first issue with the example of speech is that, unlike a task such as reaching where the sensory space can be characterized in terms of spatial orientations of the involved components (even joint angles or muscle activation parameters ultimately map onto spatial values), speech is inherently a multi-level and hierarchical phenomenon and the mappings between muscle activations and spatial configurations of the articulators, especially the tongue, are much less straightforward than in cases where the controlled system is composed entirely of rigid bodies. That is, first, the specific physical configuration of the tongue is only one thing that might be controlled for. The sensory consequences in the context of speech are articulatory, acoustic, auditory, and phonetic phenomena and there are numerous dependencies between them. Second, the mapping between the control variables and any potential output variables, whether the physical configuration of the tongue or higher-level goals such as auditory or phonetic aspects, is highly non-linear. Indeed, speech is a multi-level phenomenon of which each level has its own important and inseparable role.

Consequently, in order to define uncontrolled manifolds in the context of speech motor control, we must address the issue of a complex task. The solution we adopt is to propose that with a complex sensorimotor task having various levels of analysis and representation, there is similarly a hierarchy of uncontrolled manifolds which

reflects these different levels of analysis. Here we outline a way to address the challenge of applying the notion of uncontrolled manifold to speech in light of its complex, hierarchical nature taking the example of the articulation of vowels.

3.1.1 Hierarchy of uncontrolled manifolds

As mentioned in Chapter 1, speech is itself a component of the complex speech-language-cognition phenomenon. The degree to which this is hierarchical is reflected in the various descriptions of the components of this system including those based on neural/physical or linguistic terminology (Figure 3.1). Based on such a decomposition of the problem, a similarly hierarchical approach will be proposed in order to overcome one of the main obstacles of adapting the uncontrolled manifold approach to speech, namely that of being obliged to resolve precisely what variables are referred.

In the example uses of UCMs cited in Section 2.3, the mappings considered have been straightforward, such as from joint angles to spatial coordinates of the end effector or centers of mass (Latash, Scholz, and Schöner, 2002). In speech, at least for the tongue, there is no immediate analogue to joint angle and instead muscular activation parameters will be considered as the control variables, necessitating the use of the biomechanical model introduced in Section 2.4. This leaves the question as to the appropriate output variables, which will be considered among the options described below.

Anticipatory UCM

First, for a given set of λ -commands, there is a “trivial” uncontrolled manifold composed of all command sets which result in the same effective force for a given muscle. This occurs when no force is being generated for a given muscle due to the corresponding λ -command being above the current length of the muscle. Any change to that command which remains above the equilibrium length will not change the muscle activation levels or the forces and therefore will have no effect on the configuration of the tongue or any acoustic or other sensory consequences of that configuration. This UCM could be useful to the CNS, since it allows different levels of reactivity to anticipated perturbations or interactions, without affecting the current tongue configuration. For this reason, it is referred to this as the *anticipatory uncontrolled manifold*. While this level of the hierarchy is described here in line with the λ -model used in the biomechanical model we make use of, this concept is not exclusive to the λ -model. Indeed, other models of motor neuron recruitment could have an equivalent ability to “hide” preparatory aspects in the planning of movement at the level of their neural representation.

Articulatory UCM

Next, there is the set of all commands which give rise to the same physical configuration of the tongue with various levels of muscle forces, which includes all isometric changes in muscle activation levels and forces. This could be called the positional, isometric, or *articulatory uncontrolled manifold* and includes as a subset

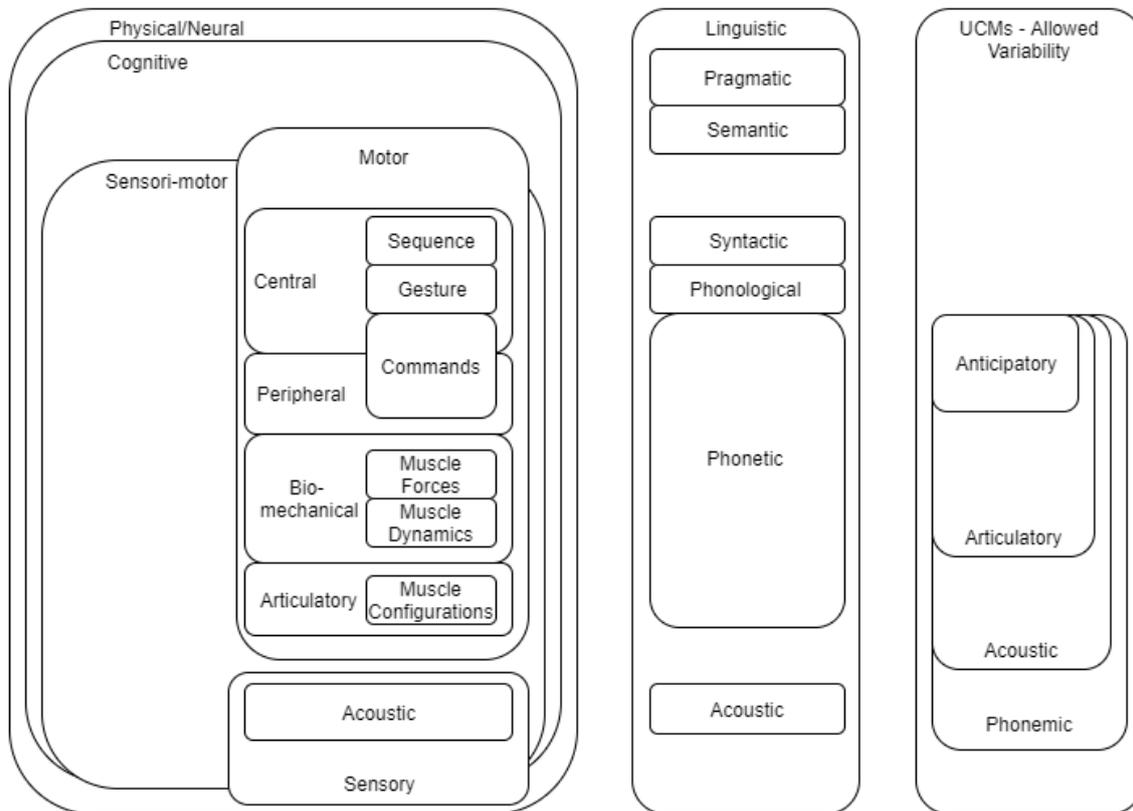


FIGURE 3.1: Hierarchies in speech. Left is the neural/physical hierarchy including the sensorimotor aspects of speech. Center is the standard linguistics decomposition of speech and language. Right is the proposed hierarchy of UCMs aligned vertically with the other hierarchies showing which aspects are allowed to vary in a given UCM. Phonemic UCM allows sub-phonemic acoustic variation; Acoustic UCM allows muscular and articulatory variation that does not result in acoustic variation (at least in terms of the first 3 formants); Articulatory UCM allows muscle variation that does not change the articulatory configuration; Anticipatory UCM allows for central command changes (e.g. gamma-activation level changes) that do not change muscle forces (but might change reactive dynamics).

the anticipatory UCM. In the example of a standard joint this would correspond to co-contraction of flexor and extensor muscles, but in the tongue would involve more complex patterns of muscular coactivation. Since the physical configuration of the tongue in an equilibrium position would not change, then neither would any sensory consequence. The tongue would still have different dynamical behavior within this set of commands, since the tongue stiffness and force amplitudes would be different.

Auditory UCM

In the case of vocalic phonemes, phonetic identity is determined entirely by the first three formants – the spectral peaks (See Chapter 2). Accordingly for the production of vowels there is a three dimensional acoustic feature space. While these features, the formants, are closely related to certain aspects of the vocal tract, this mapping is non-linear and there exist different configurations of the tongue for which the changes in formants are none or very small (Atal et al., 1978). This means there is an *auditory uncontrolled manifold* in which changes in commands do not change the acoustic features as extracted by the auditory system or the vowel identity which depends on these.

Phonemic UCM and beyond

Lastly, since configurations falling within the same ellipsoidal region in formant space give rise to the same vowel identity (see Chapter 2, Figure 2.10) there is a set of points in the command space which all give rise to the same vowel being perceived, even with different acoustic features. This is the *phonemic uncontrolled manifold*.

Similar reasoning could be applied to other linguistic features, and one could imagine uncontrolled manifolds representing the equivalence of phoneme sequences from lexical, syntactic, semantic, pragmatic perspectives.

Limitations and considerations for other decompositions

This hierarchy is not meant to be exhaustive or complete. For example, the above cases were discussed in terms of the commands of the muscles of the tongue, but could very well include the muscles of the jaw and lips which also influence acoustic features related to vowel production. Also, the muscles of the diaphragm and larynx are involved too, but perhaps in a way such that their values would be relevant to an even higher-level distinction such as a voiced/whispered (whispered vowels are indeed perceived as vowels and defined by the same acoustic features as voiced ones, despite a very different fundamental frequency).

Moreover, speech, and in particular the production of vowels, is just one example of a complex phenomenon with components having at least partially hierarchical relations. Many complex tasks can be broken down into action hierarchies and the same type of reasoning in terms of a hierarchical relationships between UCMs could apply.

$$UCM_{anticipatory} \subset UCM_{articulatory} \subset UCM_{auditory} \subset UCM_{phonemic} \dots$$

3.1.2 Application to speech motor control

As described above, there is a hierarchy of uncontrolled manifolds which could be exploited by the CNS in planning speech. Of the various levels of the hierarchy, that one which seems particularly interesting in the case of vowel production is the auditory UCM.

Focus on auditory UCMs

The reason for this choice is that the acoustic aspects of speech, especially in the case of vowel production, exhibit several interesting characteristics.

First, auditory representations of vowels are highly related to articulatory aspects. As described in Section 2.4, in particular Figure 2.8, the formants in terms of which vowels are characterized acoustically are related to articulatory parameters modulating physical characteristics of the vocal tract.

Second, these same acoustic aspects are highly important to the categorical, phonetic perception of vocalic sounds as members of different classes of sounds.

Lastly, feedback and auditory perception both convey information in these terms, making them useful for both corrective actions taken in terms of feedback as well as a learning signal.

Since the acoustic information which makes up these auditory representations is highly related to both lower, articulatory aspects and higher, phonetic ones, and is readily available in sensory feedback and perception, it is a good candidate for object of primary control of the CNS.

Potential roles of acoustic UCMs

While the standard use of UCMs has been as an analytical tool in the form of a hypothesis on the structure of variability, here several roles are considered by which such structuring of variability implicating UCMs could arise. For example, an acoustic UCM could conceivably be useful for updating commands in response to auditory feedback. It could also be exploited in movement planning as a way to know which regions in command space are best to consider given a target acoustic state and a current configuration. Both of these mechanisms if exploited in practice by a controller would give rise to the structured variability UCMs were developed to explain.

In fact, this notion leads one to imagine a potential utility of a UCM as that of serving as a structure onto which current command values are projected in order, in the example of an orthogonal projection, to make the minimal change in commands in order to reach a desired acoustic target. Moreover, it suggests a potential representation of the UCM in the form of its orthogonal projection operator.

This representation can also be thought of in terms of the reaction to command perturbations, where an original point in command space is perturbed resulting in a new set of commands and where motor control has the goal of returning to the same output values as the unperturbed point without necessarily going back to it.

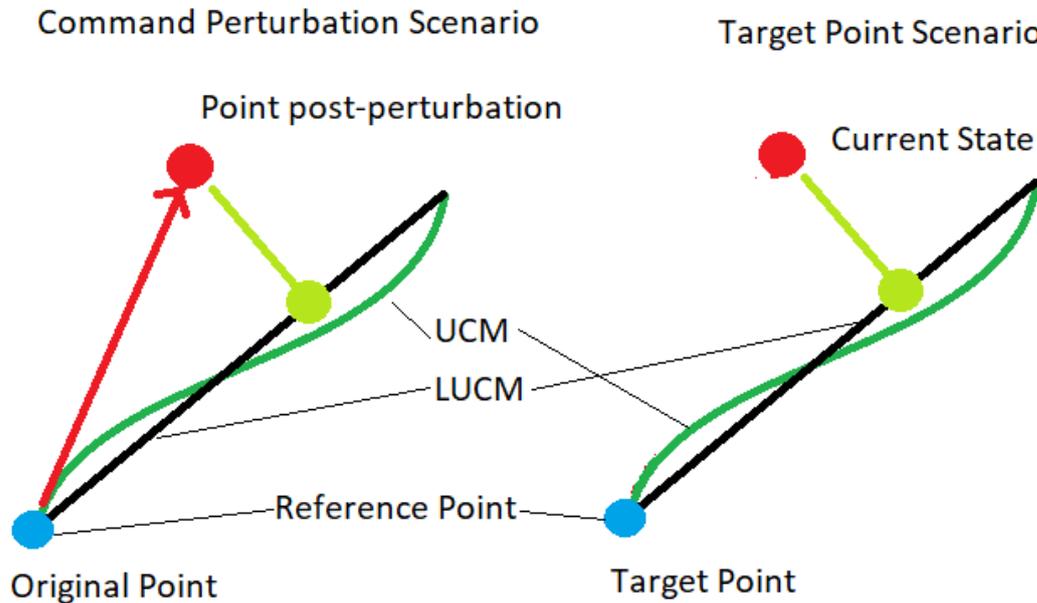


FIGURE 3.2: Two scenarios in which the UCM concept could be seen as useful and the underlying representing is equivalent. In both cases, the red circle is the location in command space where the system current is and the blue circle is the state where the controller wants to bring the system to (or back to). In the command perturbation scenario (left), the commands have been perturbed (red arrow) in order to arrive at this state. In both cases, reference point is important not for its position in the command space, but for the sensory consequences of that point and for this reason the UCM concept is useful in finding a different, potentially closer point in command space which gives rise to the same sensory consequences - the light green circle in both cases. The dark green line is the potentially non-linear Uncontrolled Manifold (UCM), while the black line is the linearized uncontrolled manifold (LUCM).

In both of these scenarios, the common situation arises where the sensory consequences are known at given point in command space, either the target (in situation 1) or the unperturbed state (in situation 2). The target or unperturbed state are for the controller a "reference point" whose corresponding UCM it is useful to know (see Figure 3.2). A different point, for example resulting from a perturbation, could then be the current state from which the model needs to arrive at any point in command space which gives rise to the same sensory consequences as the original unperturbed state (defined entirely by its sensory consequences; see Figure 3.2).

What follows is an investigation into whether the auditory uncontrolled manifold can be of practical use in the case of vowel production. In particular, it will be studied whether linearized methods give rise to acceptable behavior in a control

scenario.

Linearized uncontrolled manifolds in speech

As introduced in Section 2.3, the LUCM is defined as the null space of the Jacobian matrix of the command space to output space mapping. The auditory uncontrolled manifold described above, in the case of the biomechanical model described in Chapter 2, corresponds to a mapping of the 9-dimensional command space into the 3-dimensional acoustic space composed of the first three formants. Thus equation 2.1 becomes:

$$M = \Lambda \mapsto F \quad (3.1)$$

where $F = [F_1 F_2 F_3]$, is the first three formants, and $\Lambda = [\lambda_1 \dots \lambda_9]$ corresponds to commands of the 6 muscles of the tongue, 2 parameters setting the protrusion and rotation of the lips, and 1 parameter setting the rotation of the jaw, as described in Section 2.4 (in particular the control variables described in Table 2.1).

In the case of a 9-dimensional command space and a 3-dimensional acoustic space and the corresponding command-to-acoustic mapping M defined above, the Jacobian matrix, J_M , is then the 3x9 matrix each element $J_{M(i,j)}$ of which is the partial derivative of F with respect to Λ , $\frac{\partial F_j}{\partial \lambda_i}$ reflecting the local change in Hz to the respective formant for a local change in the corresponding command parameter.

$$J_M = \begin{bmatrix} \frac{\partial F_1}{\partial \lambda_1} & \dots & \frac{\partial F_1}{\partial \lambda_9} \\ \frac{\partial F_2}{\partial \lambda_1} & \dots & \frac{\partial F_2}{\partial \lambda_9} \\ \frac{\partial F_3}{\partial \lambda_1} & \dots & \frac{\partial F_3}{\partial \lambda_9} \end{bmatrix} \quad (3.2)$$

Since the mapping M is actually the complex, biomechanical model, the partial derivatives $\frac{\partial F_j}{\partial \lambda_i}$ must be estimated numerically at a given point, which makes the Jacobian matrix a linear approximation of the behavior of the mapping around the point in question.

Specifically, the elements of the Jacobian matrix can be calculated numerically using a finite difference method with an appropriate step size whereby each command parameter is perturbed by a fixed amount s and the difference between the formants of the perturbed commands is compared with the unperturbed commands' formants.

$$\frac{\partial F_j}{\partial \lambda_i} \approx \frac{(F_j(\lambda_i + s) - F_j(\lambda_i))}{|s|} \quad (3.3)$$

where s is the step size used.

Once the Jacobian has been calculated numerically as described, its null space can easily be obtained. The most common method for doing so involves using the algorithm for singular value decomposition which provides the null space as a set of orthonormal basis vectors.

Singular Value Decomposition

Any $m \times n$ matrix J can be decomposed into a factorization of the form $J = U\Sigma V^*$ where Σ is a diagonal matrix with diagonal entries σ_i , the r non-zero ones of which are the *singular values* and the value r is the *rank* of the matrix J . By the rank-nullity theorem, the *nullity* - or dimensionality of the null space - is $n - r$, and U and V are respectively left and right *singular vectors*. For our purposes it is only important to know that the last $n - r$ columns of V form a basis of the null space of J (corresponding to the $n-r$ singular values equal to zero).

In practice, the algorithm used for calculating the basis of the null space numerically - the `svd` routine in MATLAB - provides this as an orthonormal basis. The MATLAB routine `null` performs the above-described process for Jacobian J_M such that in practice the LUCMs were obtained as follows: $\text{LUCM} = \text{null}(J_M)$.

Orthogonal projection matrix as representation of LUCM

Above the conceptual motivation was outlined according to which a UCM could be represented by its orthogonal projection operator. In the case of a linearized UCM this operator is the *orthogonal projection matrix*.

The orthogonal projection matrix takes any point in the space in question (the command space) and projects it onto the subspace (the LUCM) at its closest point.

Letting r be the rank of the Jacobian J_M and $k = 9 - r$ its nullity, the SVD algorithm provides, on one hand, the r orthonormal eigenvectors of the Jacobian matrix, which define the directions along which a change in the commands generates a change in the formants, and on the other hand the k basis vectors $u_i \in \mathbb{R}^9$ of the null-space of the Jacobian, in which a change in the command has no effect on the formants. Let U be the $9 \times k$ matrix the k columns of which are made of the coordinates of the basis vectors of the null-space, i.e. the UCM, in the 9-dimensional motor control space.

The 9×9 orthogonal projection matrix P on the LUCM is then calculated as

$$P = UU^T \quad (3.4)$$

where UU^T is the matrix product of U and the transpose of U .

In fact, this representation has an advantage over the set of orthonormal vectors which can otherwise be used to represent the space; it is unique, since it does not depend on the order of the basic vectors in U , and can be used in a straightforward manner to compare the similarity between different LUCMs.

3.2 Validation of utility of LUCMs in speech

While above a procedure for obtaining and representing linearized UCMs in speech has been proposed, the question as to their utility is still open.

This utility depends on both the accuracy of the linear representations and also the variability of the UCM across configurations of a same vowel and across vowels.

The underlying mapping from commands to formants could simply be too highly non-linear. This could mean that the LUCMs calculated using this procedure might not be representative of the behavior in the area around the configurations used or, even if they do capture this behavior well locally, that the knowledge of one configuration's LUCM is not informative about another configuration's LUCM; that is the granularity required of the representation of information in the proposed form may be impractically small in order to achieve acceptable performance (See Section 2.2.3 for discussion of the question of granularity in the context of internal models).

In order to test the utility of UCMs, an experiment will be undertaken whereby first the LUCMs will be calculated and then used to test their behavior with different changes in commands, which will be referred to as *command perturbations*. As mentioned above, such variations in commands could be seen as either a command-side perturbation or alternatively a way to use the current position in command space and a target configuration in command and acoustic space to find a suitable alternative to the known target configuration which has the same acoustic position but which is closer to the current position in command space.

One purpose of this experiment will be to investigate the necessary level of granularity of LUCMs in order to have acceptable performance. In order to accomplish this a test will be performed of the ability of different groupings of LUCMs, ranging from those calculated for each configuration to a global "average" LUCM, to respond to a series command perturbations. The groupings will be inspired by phonetic considerations also to test whether variations in certain phonetic features seem to be more important in determining a command-space perturbation response than others.

In a first part, this experiment will be run with a series of simple perturbations – those used to numerically estimate the Jacobians, i.e. perturbations of -1 to each command variable (muscle), one at a time. It should be noted that even though the perturbations will be the same as those used in the calculation of the Jacobian (and therefore the LUCM), this does not mean they should work perfectly. Indeed, the performance in this case is indicative of the local linearity of the mapping M around the points in question.

In a second part, this analysis will be extended to commands of a greater magnitude and of a greater complexity.

While the above-described approach addresses the utility of LUCMs in speech motor control in terms of their appropriateness despite the clear non-linearities in the mapping between motor commands and acoustic features, it does not address another important aspect regarding their utility, namely their ability to provide for solutions to the motor equivalence problem by suggesting the closest configuration. In a third part, we analyze the utility of this approach in terms of its effective advantage using a measure of *compensation effort* (see Equation 3.6 below).

This will be defined as the Euclidean distance between the perturbed point and the point in the UCM onto which it is projected divided by the Euclidean distance between the perturbed point and the unperturbed point in case in the command perturbation scenario or between the point at which the UCM is centered and the current point in the UCM target scenario. A value of 1 in the perturbation scenario indicates that the way to compensate for the perturbations to the motor commands

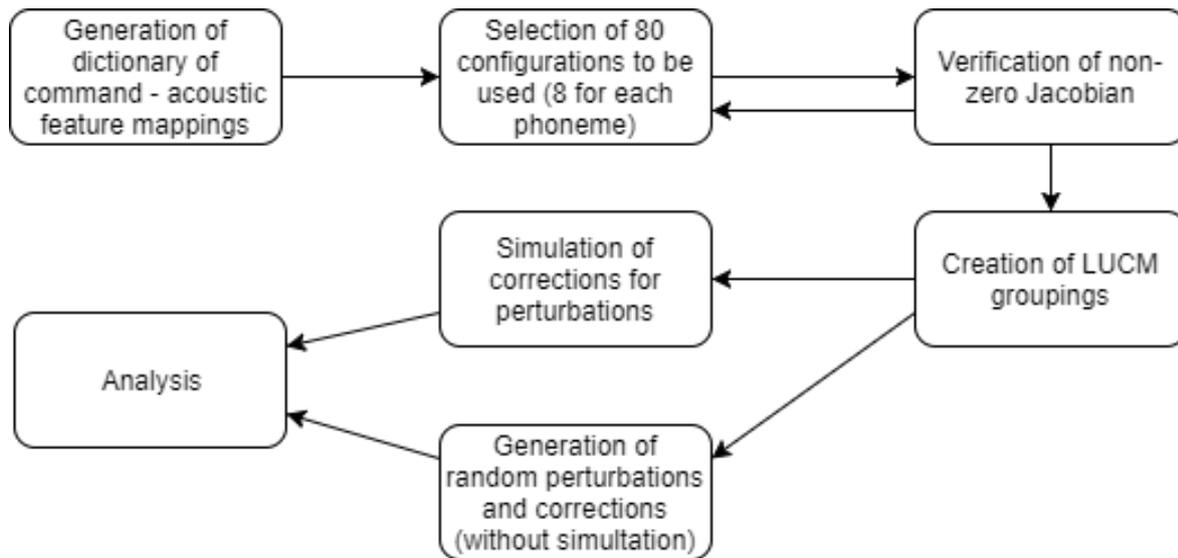


FIGURE 3.3: Block diagram of Study 1

is to move the commands back to their values in the unperturbed condition, and shows that the UCM is ineffective in suggesting an alternative besides the original unperturbed commands, while a value of 0 means that the perturbed point is already on the UCM, i.e. the perturbation caused no change in formants.

In addition to addressing the hypotheses regarding the existence, utility, and efficacy of LUCMs in the context of speech motor control, this study has as a secondary purpose that of demonstrating the value of a biomechanical model in the investigation of phenomena relating to speech motor control.

3.2.1 Methodology

Calculation of Jacobian matrices and projection matrices

The biomechanical model was used to generate a dictionary of 50000 configurations covering the space of possible combinations of command values. (See Perrier, Ma, and Payan, 2006 and Winkler, Ma, and Perrier, 2010 for further discussion of the method by which the command space was sampled). Of these attempted commands, 49188 simulations completed successfully, i.e. did not fail due to issues of numerical stability. For each configuration the acoustic characteristics of the resulting vocal tract shape and vowel identity (corresponding to the 10 French oral (non-nasal) vocalic phonemes - see Figure 3.4 and Table 3.1) were determined. 3211 were labeled as one of the 10 phonemes.

8 configurations from each of the 10 phonemes were then randomly selected. For each of these configurations the Jacobian matrix was calculated using the finite difference method and a step-size of -1 for each of the commands. Configurations which encountered simulation errors due to numerical issues with this step size were discarded and replaced.

¹Source: [Wikimedia Commons](#). Creative Commons License

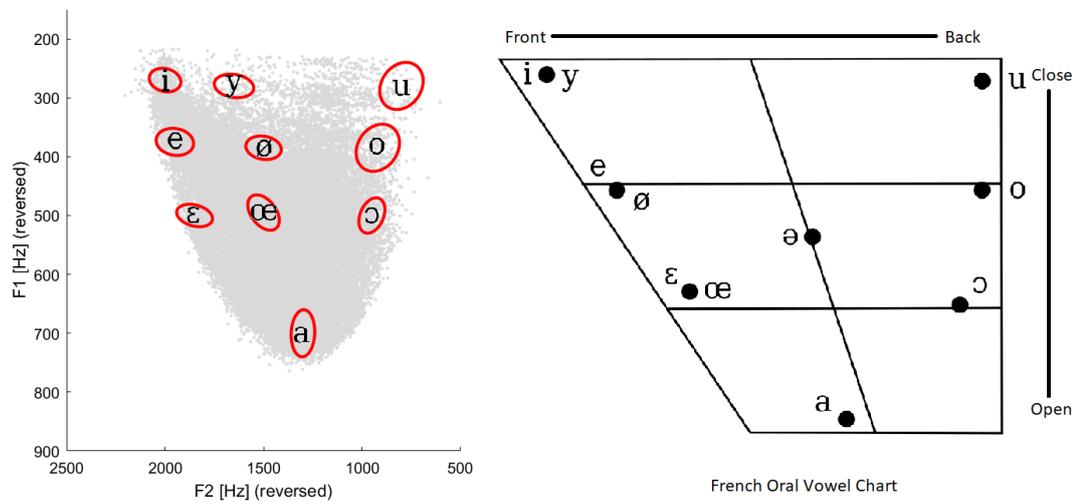


FIGURE 3.4: Left - Plot of dictionary F1-F2 values, with phonemes shown. See Table 3.1 for description of vowels. Right - French Oral Vowel Chart¹ based on articulatory features. For overlapping phonemes in vowel chart (front vowels), the left vowel is unrounded and the right vowel is unrounded. This articulatory presentation corresponds to that of the F1-F2 plane, with the rounded vowels having a slightly decreased F2 compared to the unrounded pair.

TABLE 3.1: Description and number of occurrences of French oral vowels in dictionary

IPA	Description	Example	Nr in dictionary
i	close front unrounded	fini	73
e	close-mid front unrounded	beauté	332
ε	open-mid front unrounded	bête	130
a	open front unrounded	chat	863
y	close front rounded	tu	57
ø	close-mid front rounded	peu	256
œ	open-mid front rounded	jeune	1131
u	close back rounded	où	50
o	close-mid back rounded	réseau	147
ɔ	open-mid back rounded	sort	172

TABLE 3.2: List of groupings

Grouping Nr	Nr of Groups	Groupings
1	80	individual configurations
2	10	phoneme average
3	5	{i e}, {ε a}, {y ø}, {ɔ œ}, {u o}
4	3	{i e y ø}, {ε a}, {ɔ u o œ}
5	3	{i e ε a}, {y ø u o}, {ɔ œ}
6	2	front {i y e ø ε œ} vs. back {u o ɔ a}
7	2	open {ɔ a ε œ} vs. close {i y u e ø o}
8	2	rounded {y ø u o ɔ œ} vs. unrounded {i e ε a}
9	1	global average

Since commands were determined randomly in order to cover the command space, it was possible that some were above the corresponding muscle length in the equilibrium configuration, meaning that no force was generated by that muscle and that small changes in command values would not result in any change in output and the corresponding partial derivative in the Jacobian matrix would be zero. In the terms introduced earlier, this would correspond to the point being in the anticipatory UCM and would not reflect the command pattern's sensitivity to below-threshold changes. Each command was therefore adjusted as follows such that any above-threshold commands were set just below threshold:

1. If with command perturbation of -1 no change to F1 was noticed, the command perturbation was increased in magnitude by 0.1 (i.e. -0.1) until a change in F1 was noticed.
2. It was verified that the new commands gave the same phoneme labels. Indeed, formant values were mostly within fractions of a Hz of the original values.

Using Jacobian matrices now verified as being non-zero, orthogonal projection matrices were then calculated using the method described above. Moreover it was verified that all Jacobian matrices had the same rank and consequently the dimension of the null space – the nullity – was the same as well.

At this point the linear uncontrolled manifold has been calculated for each configuration (8 examples for each of the 10 French vocalic phonemes).

LUCM groupings

Nine different groupings were then created ranging from a grouping made of LUCMs of individual configurations to a single global average LUCM. Intermediary groupings were made using different phonetic features such as phonemes and distinctions including front/back, open/close and rounded/unrounded. The details of each grouping is described in Table 3.2 and shown, for certain cases, in Figure 3.5.

The idea is that for each group in each grouping the average LUCM should be calculated. First though, the arithmetic mean of a set of projection matrices is not itself a projection matrix. Nor is it straightforward to take the mean of orthonormal

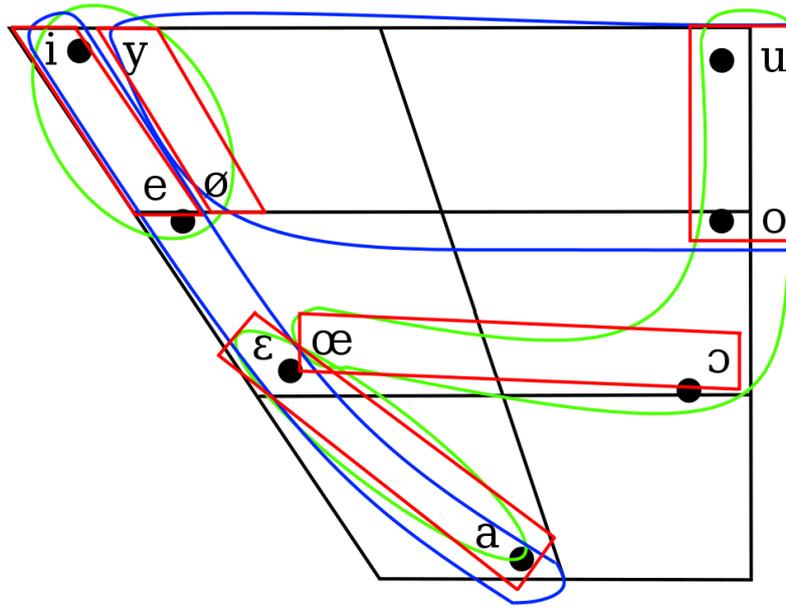


FIGURE 3.5: Groupings 3, 4, and 5 depicted with red, green, and blue respectively, in the F1-F2 plane. The x-axis is F2 increasing from right to left and the y-axis is F1 increasing from top to bottom. Rounded vowels are depicted on the right side of rounded/unrounded pairs.

basis sets of different linear subspaces of a vector space. Instead, Jacobian matrices, where each element represents a continuous variation value, can be averaged together and should give rise, for not-too-drastic changes in values, to “average” LUCMs. Accordingly, for each grouping, the group-average projection matrix was calculated by averaging the group constituents’ Jacobian matrices and recalculating the projection matrix on the basis of the Singular Value Decomposition applied to these “average Jacobian Matrices”.

For each grouping, each configuration of each phoneme was perturbed with a series of command perturbations as described in Table 3.3 and then corrected by having the perturbation vector (linking the unperturbed and perturbed points) multiplied by projection matrix, resulting in a command perturbation which 1) was in the linear UCM of the unperturbed condition and (2) was as close as possible to the perturbed configuration. The resulting F1, F2, and F3 values were recorded as F_c (Formants-compensated).

$$F_c = M(\Lambda + P\Pi) \quad (3.5)$$

where M is the model, Λ is the set of commands of the unperturbed configuration, P is the orthogonal projection matrix onto the LUCM of the unperturbed condition, and Π is the perturbation vector.

The perturbations tested included three perturbation types, three types and either 6 or 9 perturbation locations, depending on the type. For the first, simple type, only one of the 9 command variables was perturbed at a time and was always a subtractive perturbation (only decreased the command value, in order to make sure

TABLE 3.3: Summary of command perturbations

Type	Nr of Perturbations	Range of each element	Commands Perturbed
A	9	$\pi_i = -1, \forall j \neq i, \pi_j = 0$	all 9
B	50	$\pi_i \in [-1, 0]$	tongue muscles (1-6)
C	50	$\pi_i \in [-1, 1]$	tongue muscles (1-6)

that it has an effect on the muscle force). For the second type, 50 random vectors of length 6, each component of which was on the range $[-1,0]$ were obtained, padded with three zeros to become length-9 vectors after which they were normalized to magnitude 1. Again these perturbations only decreased the command values. The third type, 50 more random 6-d vectors were obtained, but with components each on the interval $[-1,1]$. They were also padded with three zeros, to become length-9 vectors and normalized to magnitude 1. For the latter two types, only the λ parameters pertaining to the tongue were perturbed, with each jaw and lip parameter left at 0. The perturbations of all three types were then applied at three different magnitudes – 1, 3, and 5 and then compensated using each of the 9 previously obtained groupings of projection matrices.

Lastly, in order to answer the question of the potential utility of LUCMs to reduce effort resulting from a command perturbation or to reduce effort to go to a target acoustic configuration, the compensation effort was calculated as

$$C_e = \frac{|\Pi - P\Pi|}{|\Pi|} \quad (3.6)$$

where $|\Pi|$ is the magnitude of the perturbation, and $|\Pi - P\Pi|$ is the distance between the point in command space following perturbation and the projection of this point onto the LUCM ($P\Pi$).

Since this measure depends entirely on the perturbation vector and the projection matrix, it does not require simulation to be calculated. Accordingly, it was calculated for the original type A perturbations and compensations and for 10000 random type B and type C perturbations and compensations.

3.2.2 Results and Analysis

Overall compensation error

First, the absolute error after compensation in Hz was obtained for each simulated configuration and the median value was calculated over all the vowels and all the configurations. Histograms in Figure 3.7 show that the errors after projection of the perturbed configuration onto the LUCM of the unperturbed configuration were fairly small, especially when compared to the average half-width of the projection onto the F1, F2, and F3 axes of the phonemic ellipsoid for each phoneme, which was defined as its tolerance (see Figure 3.6 for a depiction of how this was determined). Only 3.98%, 2.22%, 1.52% of the compensated F1, F2, and F3 values were above the tolerance threshold.

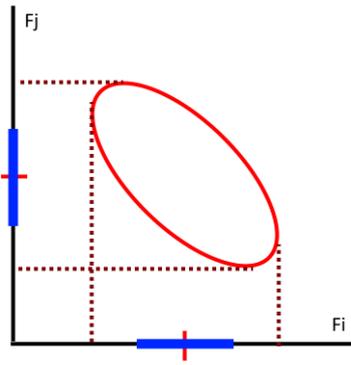


FIGURE 3.6: Diagram of the calculation of tolerances, θ_i , used in calculation of above-threshold rates. The vowel ellipsoid is projected onto each axis of the formant space (representing a variability of 2 standard-deviation around the mean value, assuming a gaussian distribution of the data) and the half the length between the two extremes is taken as θ_i .

Additionally, the acoustic features of the compensated configurations were evaluated to see whether they lie within the correct phonemic category and the fraction of simulations for which this was not the case was calculated to 17.5%.

The difference between the very low error rates (both in Hz and fractions of simulations exceeding the defined tolerance threshold) and the category failure rate was due to this latter measurement's sensitivity to covariation among F1, F2, F3 dimensions in the definition of vowel ellipsoids, but also to the number of simulations whose unperturbed formant values were not centered in the corresponding ellipsoid. Indeed this result is an example of the hierarchical nature of speech motor control and justification for the present approach wherein the auditory UCM has been distinguished from a related but different phonemic UCM.

Presentation of Error Data

The following procedure was used to normalize the error data within each subset of simulations (groupings, phonemes, etc.):

1. the average value of the error, e_{mean} , was calculated;
2. the error values were centered $e_{cent} = e - e_{mean}$;
3. the maximum of the absolute value of the centered error, $e_{cent_{max}}$, was calculated;
4. the centered error was normalized by $e_{cent_{max}}$.

As a result, a zero-error corresponds to value -1 of the centered normalized error and all the errors below average in the considered subset of data are plotted below the zero level in the normalized centered error. Hence in Figures 3.8 to 3.11, the lower the value, the better the result.

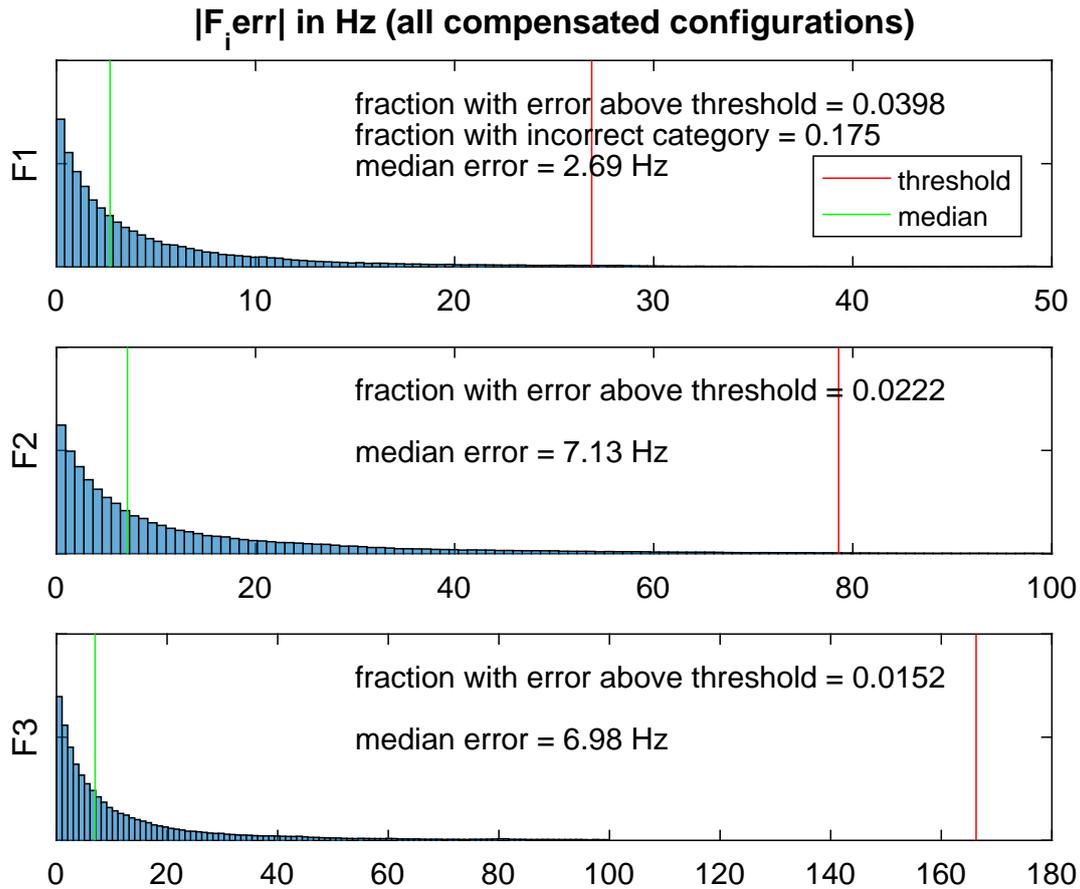


FIGURE 3.7: Histograms of formant error ($F_i \text{err}$) after projection of the perturbed configuration onto the LUCM of the unperturbed configuration. Horizontal axis is $F_i \text{err}$ in Hz. Results are shown by formant for all compensated configurations. Tolerance threshold values, as determined in Figure 3.6 are depicted with a red line and median values are depicted with a green line as well as overlaid textually.

Results by Grouping

Median absolute error values, the rate of category failure, and the rates of formants exceeding the tolerance thresholds described above were calculated for each grouping. These values are presented numerically in the first 10 lines of Table 3.4 and visually in Figure 3.8.

From this several observations can be made:

- Grouping 1 which corresponds to the case of individual configurations has, as could be expected, by far the least error and best performance in almost every error measure.
- Groupings 7 (based on the open vs close distinction) and 9 (with one global UCM) have the highest error values across measures

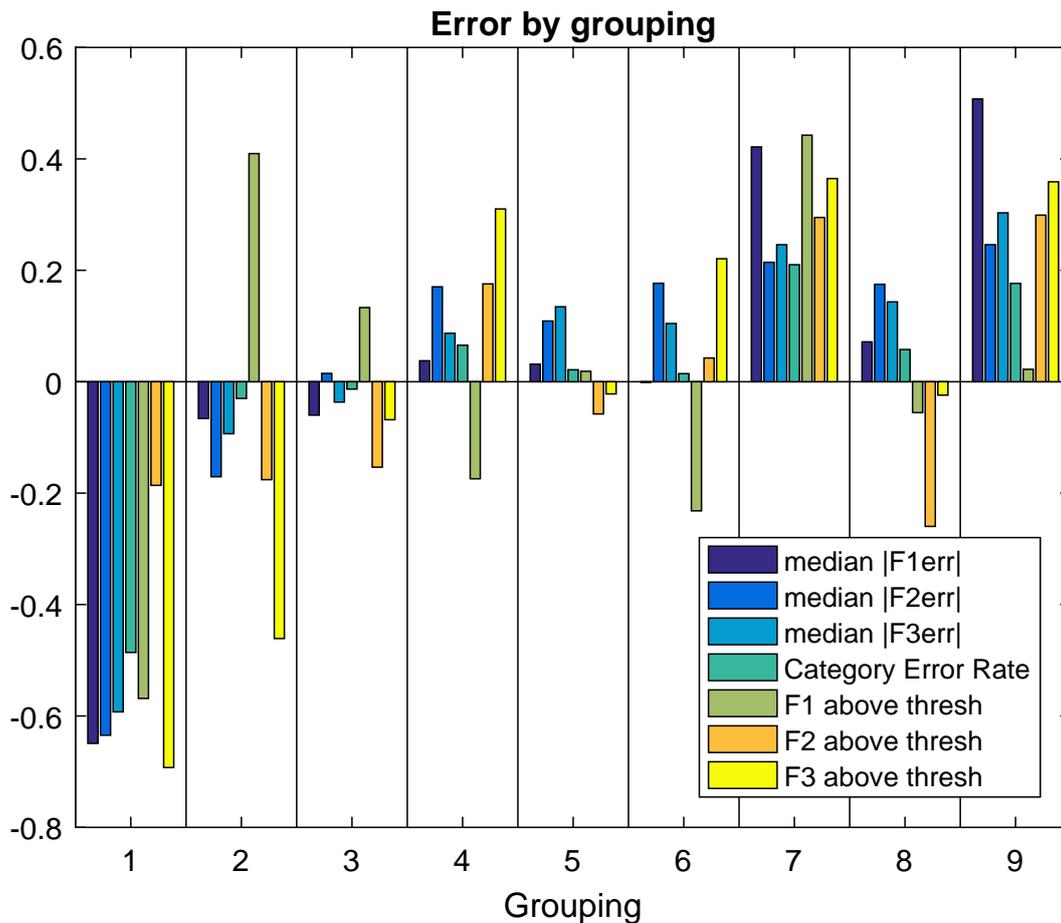


FIGURE 3.8: Normalized error by grouping - lower is better. Data were normalized by centering and scaling such that 0 is the average error level across all groupings and -1 is error-free performance. See Section 3.2.2 for details of normalization.

- Groupings 2 (individual phonemes) and 3 (5-way partitioning of the vowel space) are the best performing, with grouping 2 performing better than grouping 3.
- the only metric according to which grouping 2 seems to perform badly - F1 threshold error rate - does not seem, when looking at the numerical data (Table 3.4) to indicate significant problems in performance.

Source of Error

Median absolute error values, the rate of category failure, and the rates of formants exceeding the tolerance thresholds were calculated for each subset of the dataset in order to see which factor influenced the error among group (Figure 3.8), phoneme (Figure 3.9), perturbation location (Figure 3.10), perturbation type, and perturbation magnitude (both in Figure 3.11). The data is presented numerically in Table 3.4 as well.

For all measures, the following observations can be made:

- perturbation complexity (type) and perturbation magnitude are both positively correlated to acoustic error, although the impact of perturbation magnitude is much greater than perturbation complexity.
- perturbation type, location, and group (other than group 1), have a rather limited impact on error, with limited spread, while perturbation magnitude and phonemic category have the most impact.
- which phonemic categories are responsible for greater amounts of error depends on the measure with each F1, F2, F3, category giving different values (with the exception that both F1 and mis-categorization measures suggest that phoneme 10 (/o/) has the highest error).
- the close front spread vowels /i,e/ are associated with significantly larger error on F3 and the close front rounded vowel /y/ is associated with significantly larger error on F2. This suggest that the LUCM does account well for the impact of the change in position of the constriction for the front close vowels in general.

Compensation effort and efficiency of LUCMs

Lastly, as can be seen in Table 3.5, the mean compensation effort was quite low, meaning that on average the point on the UCM onto which the commands were projected was considerably closer to the perturbed configuration than the reference point of the UCM. Also these results show that on average the compensation effort did not change much by grouping or perturbation type. The overall average of 0.39 means that on average the closest point on the LUCM to the perturbed commands was 39% of the distance of going back to the unperturbed commands.

3.2.3 Conclusion

The three questions of interest addressed with this study were:

1. whether LUCMs are useful in the context of speech motor control,
2. whether there is any evidence that average LUCMs constructed using the procedure described above and representing different ways to partition the acoustic space based on different linguistic features, exhibit different levels of performance,
3. and whether there is any structure to the error observed, i.e. whether there are certain factors which contribute more than others to the observed formant and category errors.

Regarding the first question, a case for utility of LUCMs in the context of speech motor control can be made as long as there are no formal issues to their construction, they are well-behaved, and there is some interest in their use. The reasonable error behavior shown above (primarily in Figure 3.7) at the same time as the results

TABLE 3.4: Results after projection onto the LUCM determined for the different subsets of the data. Columns 2-4 display the median of the absolute formant error in Hz. Column 5 displays the proportion of the distribution not in the phonetic category of the unperturbed (or target) configuration. Columns 6 to 8 display the proportion of the distribution that lies above the tolerance threshold determined as in Figure 3.6.

Subset	$ F_1err $	$ F_2err $	$ F_3err $	cat	$ F_1 > \theta_1$	$ F_2 > \theta_2$	$ F_3 > \theta_3$	
(overall)	2.69	7.13	6.98	0.18	0.040	0.022	0.015	
Grouping	1	0.94	2.61	2.84	0.09	0.017	0.018	0.005
	2	2.51	5.91	6.33	0.17	0.056	0.018	0.008
	3	2.53	7.24	6.73	0.17	0.045	0.019	0.014
	4	2.79	8.34	7.59	0.19	0.033	0.026	0.020
	5	2.77	7.91	7.92	0.18	0.041	0.021	0.015
	6	2.69	8.39	7.71	0.18	0.031	0.023	0.019
	7	3.82	8.66	8.70	0.21	0.057	0.029	0.021
	8	2.88	8.38	7.98	0.19	0.038	0.016	0.015
	9	4.06	8.88	9.10	0.21	0.041	0.029	0.021
Phoneme	i	1.83	2.99	11.82	0.09	0.001	0.000	0.049
	e	2.65	6.47	14.66	0.11	0.003	0.000	0.087
	ε	2.22	5.95	11.44	0.13	0.000	0.006	0.002
	a	1.70	7.00	4.75	0.26	0.017	0.019	0.005
	y	2.43	15.96	9.44	0.21	0.003	0.116	0.001
	\emptyset	3.01	8.33	6.19	0.26	0.039	0.021	0.004
	$\text{\textcircled{ae}}$	2.86	5.17	4.19	0.08	0.019	0.000	0.000
	u	3.29	10.09	5.17	0.08	0.019	0.026	0.001
	o	3.37	8.50	4.49	0.21	0.061	0.013	0.000
$\text{\textcircled{v}}$	5.99	8.15	5.14	0.32	0.239	0.022	0.000	
Location	GGP	2.52	7.01	6.95	0.18	0.036	0.020	0.016
	GGA	3.20	8.52	8.05	0.21	0.051	0.029	0.012
	HYO	3.05	7.47	7.86	0.18	0.042	0.023	0.017
	STY	2.42	7.13	6.35	0.16	0.033	0.022	0.012
	VERT	2.62	6.73	6.71	0.16	0.039	0.016	0.016
	IL	2.52	6.66	6.38	0.16	0.035	0.021	0.018
Magnitude	1	1.13	2.97	2.80	0.06	0.013	0.004	0.009
	3	3.41	9.30	8.89	0.18	0.038	0.018	0.015
	5	5.66	15.36	15.33	0.29	0.070	0.046	0.022
Type	A	2.49	6.60	6.68	0.16	0.038	0.025	0.017
	B	2.66	7.07	6.84	0.16	0.037	0.022	0.014
	C	2.80	7.46	7.34	0.19	0.042	0.022	0.016

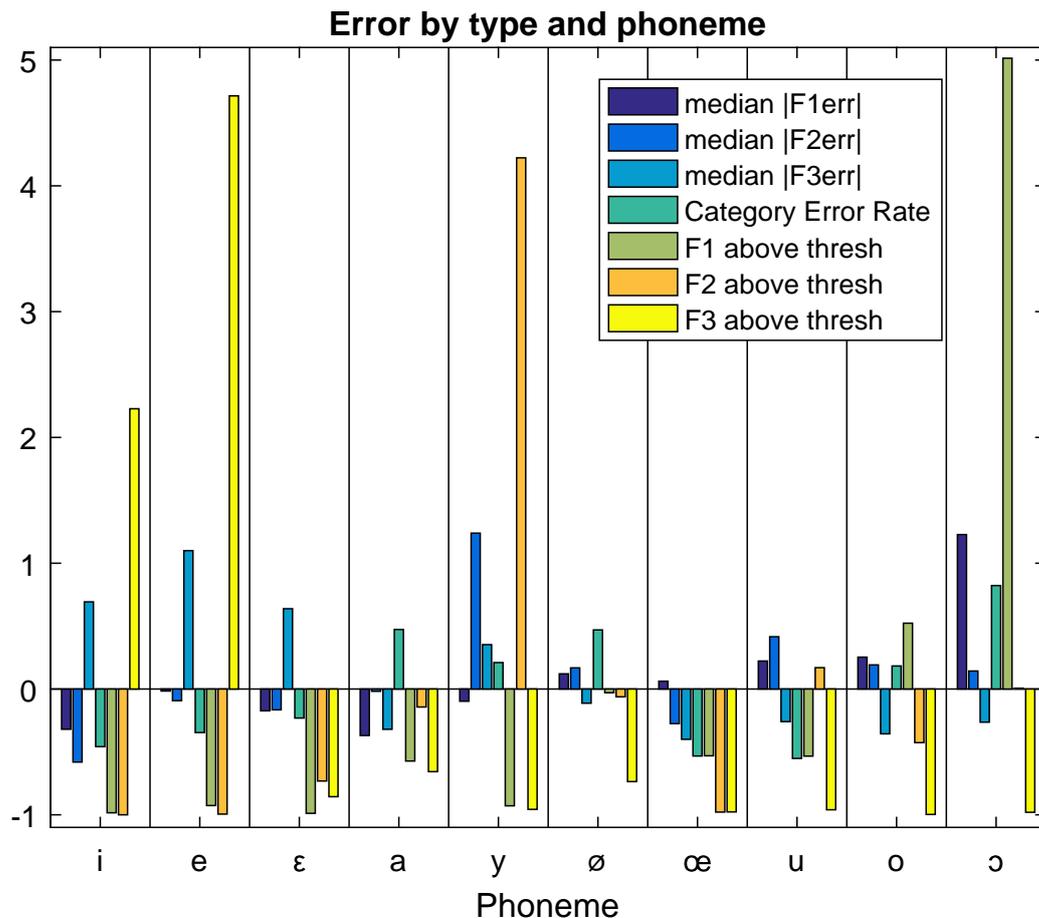


FIGURE 3.9: Normalized error by phoneme - lower is better. Data was normalized by centering and scaling such that 0 is the average error level across all phonemes and -1 is error-free performance. See Section 3.2.2 for details of normalization.

showing that LUCMs propose considerably closer solutions than returning to the reference point (Table 3.5) is strong evidence in favor of accepting the hypothesis that LUCMs do make sense and are useful in the context of speech motor control.

As for the second question, the above analysis shows that average performance is not highly affected by grouping and LUCMs specifically determined for each configuration allow for considerably better performance. Additionally, there are several groupings (notably 7 and 9) which perform consistently worse than the others, although only slightly, and several (2, 3) which appear to perform better than average, with grouping 2 being slightly better than grouping 3. For this reason we conclude that local - configuration specific - representations perform the best, but the degree to which the other methods perform worse is not defined functionally here. In other words, whether the errors produced by the other groupings are acceptable would depend on the context of their potential use.

The last question addressed by our analysis shows that perturbation type, location, and group (other than group 1), have a rather limited impact on error, while

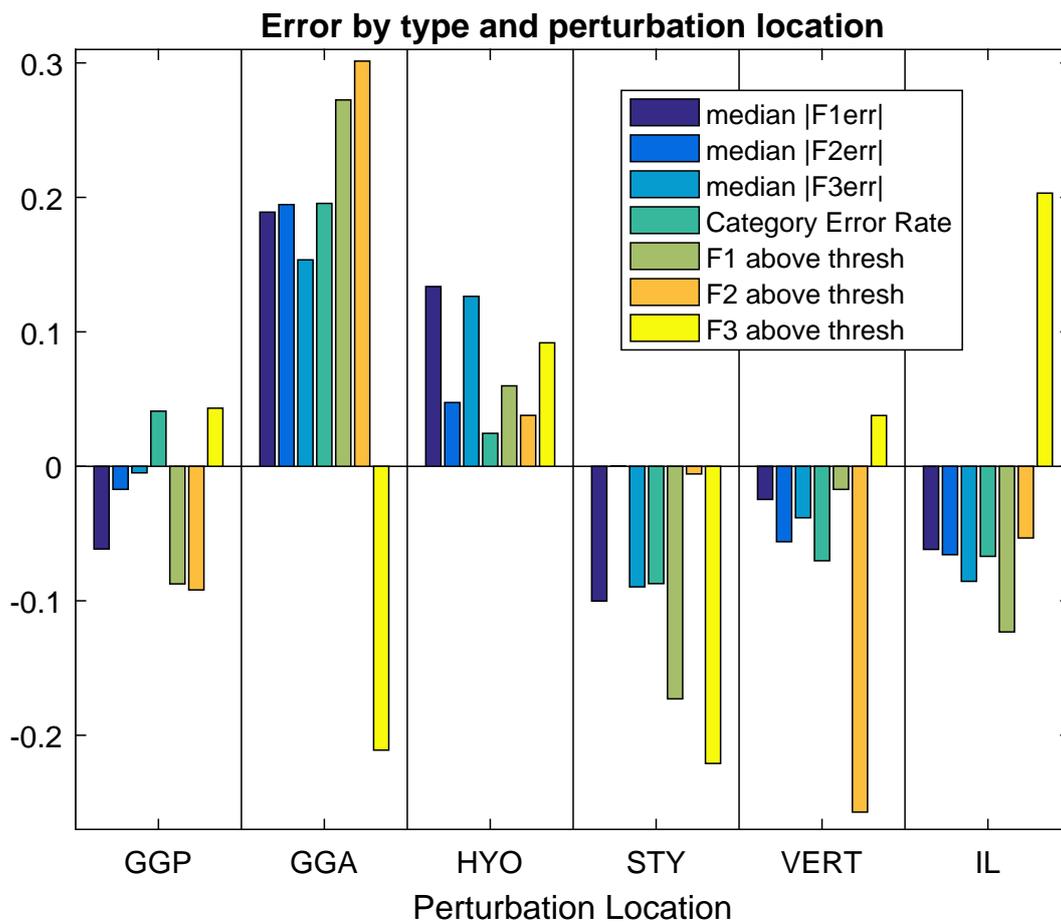


FIGURE 3.10: Normalized error by perturbation location - lower is better. Data was normalized by centering and scaling such that 0 is the average error level across all perturbation locations and -1 is error-free performance. See Section 3.2.2 for details of normalization.

perturbation magnitude and phonemic category have the most impact. This observation however is complicated by the fact that the specific phoneme which contributes most to the error depends on the specific measure and only in the case of median F1 absolute error and categorical failure rate did the analysis show that phoneme 10 was worst.

3.3 Discussion and Perspective

In this chapter it was discussed how to make use of the uncontrolled manifold concept in the case of a hierarchical control problem such as speech motor control. By having a corresponding hierarchy of UCMs and a dependency relationship, any relevant control variable - output variable relationship can be captured in a single UCM which may be related to UCMs at other levels of the hierarchy.

The conceptual utility of this distinction of hierarchical characteristics is evident in the distinction seen in terms of results between the Hz and categorical error rates.

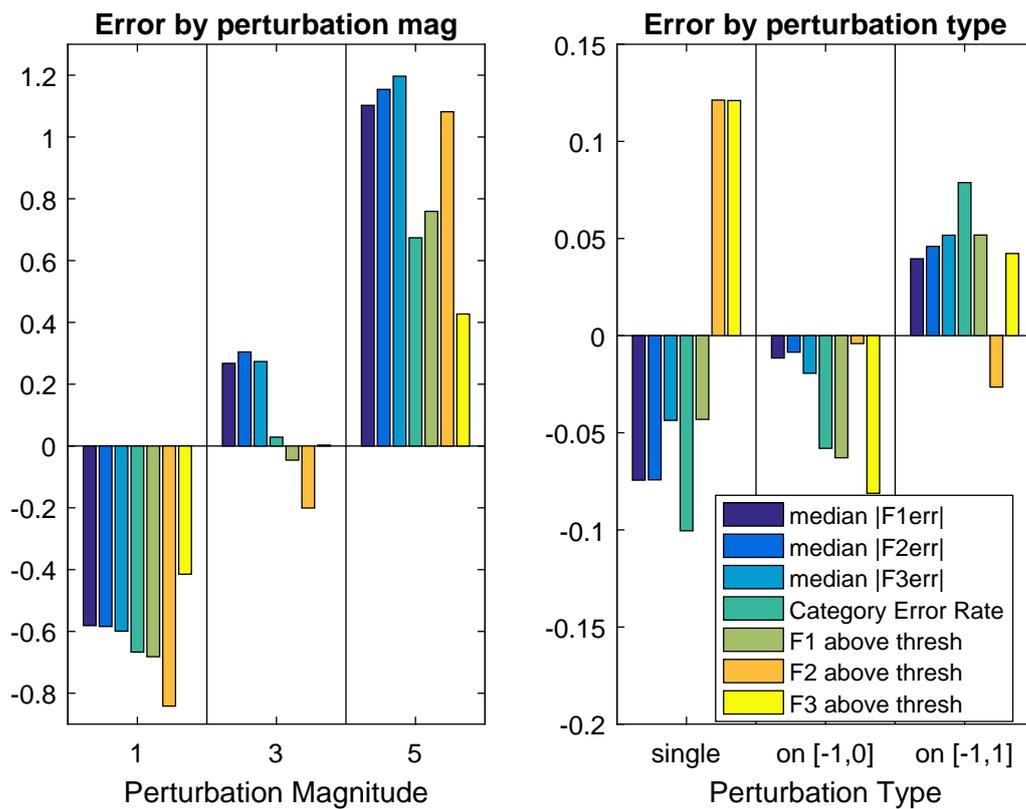


FIGURE 3.11: Normalized error by perturbation magnitude (left), perturbation type (right) - lower is better. Data was normalized by centering and scaling such that 0 is the average error level across all perturbation types and magnitudes and -1 is error-free performance. See Section 3.2.2 for details of normalization.

Indeed if the goal were to maintain phonemic category, the appropriate UCM to be investigated would have been the phonemic UCM. Since the notion of phoneme is much more categorical (although there it is not perfectly binary as evidenced by perceptual difference curves (Harnad, 2003)), the formulation of the phonemic UCM might not be compatible with the linearized uncontrolled manifolds used in this study or in particular with their representation as an orthogonal projection matrix and further work is necessary to study this question.

A specific representation of the linear uncontrolled manifold was proposed in the form of the orthogonal projection matrix and finally strong evidence was obtained for this representation's utility using a biomechanical model.

TABLE 3.5: Compensation effort by grouping and perturbation type

	overall	Grouping									Type		
		1	2	3	4	5	6	7	8	9	A	B	C
C_{eff}	0.39	0.42	0.38	0.39	0.37	0.4	0.37	0.41	0.39	0.41	0.36	0.36	0.42

From the above it can be concluded that the UCM concept is indeed useful in the context of speech and suggest that further work is necessary for studying which level of granularity is most appropriate for any given application. Although the results presented were inconclusive in this respect, they can be interpreted as suggestive of the phonemic level being an appropriate starting point, especially if exemplar phonemes are used and about which configuration specific UCM-type information could be known.

As for methodological concerns which perhaps impeded more precise or conclusive results, it should be pointed out that the individual configurations used within each phoneme might not be representative of any real knowledge or information a real speaker might have. It is entirely possible for an individual to make use of a limited number of exemplar or prototypical representations. This means the approach used here wherein groupings of different configurations giving rise to the same vowels might have little relevance to the kind of knowledge a controller might make use of.

In this interpretation the present experimental approach captures more of the possible variations in motor parameters which could be employed and thus reflects a question more along the lines of the degree to which similar articulatory phenomena (to the extent there is a high correlation in vowels between articulatory and acoustic properties) with varying motor control parameters (stiffness, coactivation strategies, etc) could be represented as a whole. Said otherwise, this study, in this light, goes in the direction of asking whether a structure, such as a UCM, can represent a group of possible motor realizations of a sensory phenomenon. If future work should be done in this direction, a suggestion would be for a larger number of individual configurations for each phoneme and controlling the distribution of these configurations.

Even in light of this interpretation, the results can be seen as promising for linear methods potentially being employed by the CNS in motor planning and perturbation responses, especially if a limited number of exemplar configurations is used.

Lastly, as mentioned in Chapter 2, UCMs and optimal feedback control are closely related. In light of our findings and in particular the proposal of a potential utility of UCMs being as suggesting directions towards targets defined in terms of sensory values, further work should be done to detail the compatibility of these two approaches.

Chapter 4

Reflex mechanism in speech motor control

Speech and complex behaviors in general are not accomplished through monolithic feedforward controllers which simply send commands to the periphery, hoping that the commands are precise enough to accomplish the desired action faithfully. Instead, faithful execution of desired actions is done through a hierarchical process involving nested feedback loops and a periphery which itself plays an important role.

From a neurobiological point of view, top-down cortical commands, generated taking into account both goals and sensory feedback, influence lower neurons which in turn work in a tighter proprioceptive feedback loop with the periphery to modulate the execution of movement. There are numerous ways in which sensory information - proprioceptive all the way to auditory - can be integrated into the process of control. While understanding the neural circuitry is certainly informative, as is looking at the time course of neural responses at different levels, one piece of information missing from the current understanding is the importance of the periphery itself, the lowest level of the control hierarchy.

In this chapter a review will be presented of some previous work looking into the mechanisms of postural control in the context of speech, followed by a study in which perturbations of the tongue are modeled and analyzed in relation to data acquired in human subjects in a separate study.

4.1 Postural control of the tongue

Postural control is the stabilized maintenance of a configuration of a part of the body, for example the tongue. Several classes of mechanisms are responsible for it.

First, the passive mechanical and elastic characteristics of the periphery – the tongue itself – provide for stabilization against perturbations. The physical material itself can have a tendency to resist perturbations or react in a way to minimize their effect.

Second, connections between muscle fibers and brainstem nuclei implement fast reflex arcs, such as the stretch reflex (see Section 2.1.6), that allow faster responses to perturbations than cortical mechanisms would allow.

Lastly the central nervous system actively controls movement and specifies target positions of the tongue via descending pathways going from the primary motor

cortex (M1) to muscles in the tongue via the brainstem. In response to afferent sensory information, these control mechanisms can stabilize the tongue after perturbation or otherwise compensate for its effects.

4.1.1 Previous work on orofacial reflexes

Previously Ito, Gomi, Honda, et al. (Gomi et al., 2002; Ito, Gomi, and Honda, 2004; Ito, Kimura, and Gomi, 2005) investigated a series of related questions in the context of compensatory articulation. Their experiments focused on the reaction of the lips during a bilabial production task to perturbations of the jaw. They had speakers produce the bilabial fricative / Φ /, which requires precise control of the lip aperture and thus precise control of the position of both lips. Since a jaw perturbation modifies the position of the lower lip and the lip aperture, compensation for the perturbation requires an increase of the movement amplitude of lips. In practice, experimental observations have shown that the downward movement of the upper lip increases in reaction to this perturbation.

First, Gomi et al. (2002) observed that the compensation in terms of lip movement was greater for / Φ / than it was for /a/, which does not require lip aperture to be precisely controlled, meaning that the compensation was task-dependent. They also observed that the upper lip displacement often preceded the EMG response (See Figure 4.1), suggesting that muscle activation would not be the only cause of lip movement.

From this they hypothesized that the passive stiffness of the mechanical coupling between the jaw/lower lip and the upper lip could be important in generating the fast compensatory response, which they then validated through simulations of a numerical dynamical model of this coupling.

Second, Ito, Gomi, and Honda (2004), using their dynamical model and EMG data, further hypothesized that this stiffness parameter itself is under the control of the CNS implying that the control strategy of the CNS is not just for movement but also dynamic stability.

In addition to the mechanical nature of the linkage between the upper and lower lips and responsibility of passive dynamics in the initial response of the upper lip, they observed a muscle activation response in the EMG signal occurring afterwards. As shown in Figure 4.1, this initial, mechanical response occurs within 20 ms and the muscle activation response is seen around 50 ms after perturbation onset.

Citing Abbs and Gracco (1983) and Weber and Smith (1987) who had observed a brainstem-mediated perioral reflex with a latency of between 12 and 18ms, Ito, Kimura, and Gomi (2005) postulated that the response the onset of which was observed around 50ms was cortical in origin. They performed an experiment using transcranial magnetic stimulation (TMS) to facilitate cortical activation and found that TMS applied 10 ms before the observed EMG activity's peak (they had determined 10 ms to be the latency of TMS induced motor evoked potentials (MEPs)) enhanced the EMG response. They additionally verified that similar enhancement did not occur in the case of the brainstem-mediated perioral reflex and from these experiments concluded that the response was due to a cortical reflex loop.

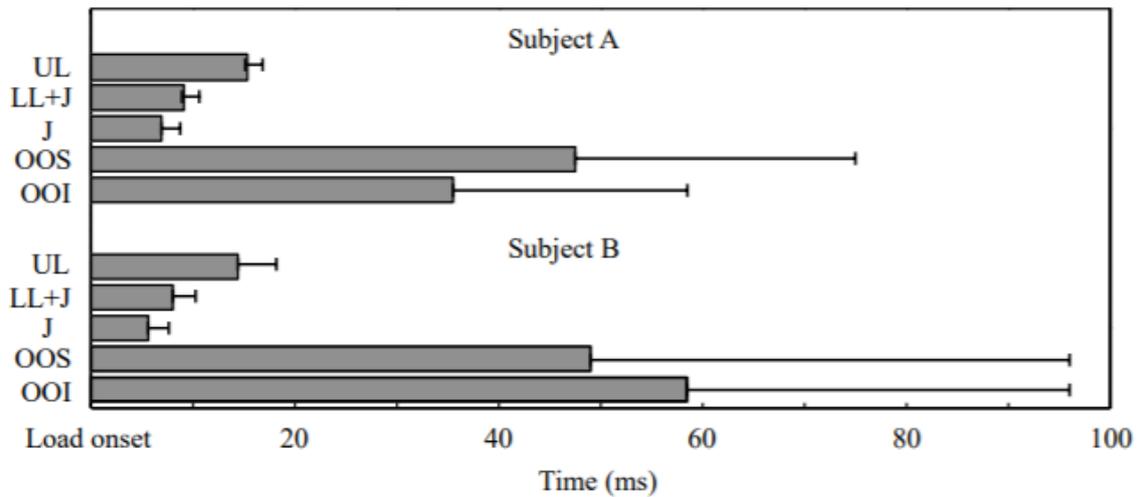


FIGURE 4.1: Latencies of displacements of upper lip (UL), lower lip + jaw (LL+J), jaw alone (J), and EMG changes of orbicularis oris superior (OOS) and orbicularis oris inferior (OOI) during sustained production of $/\Phi/$ of two subjects (error bars depict standard deviation). Reproduced from Gomi et al., 2002. © 2002 Elsevier Science Ltd.

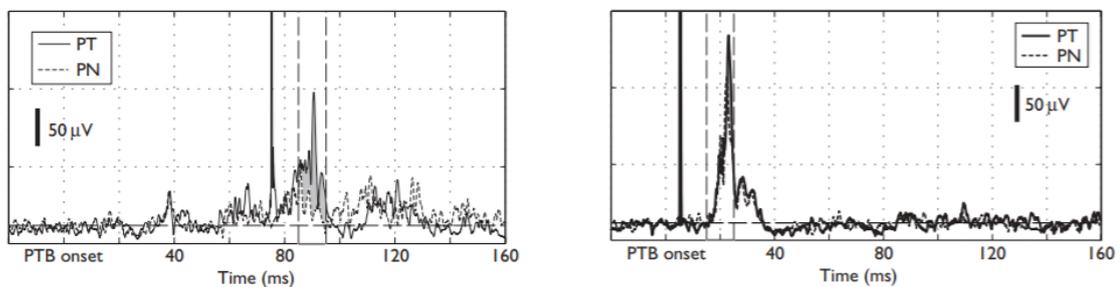


FIGURE 4.2: (Left) Typical reflexive compensatory response with or without TMS in rectified and averaged EMGs of the orbicularis oris superior (one participant). Time zero is the jaw perturbation onset with TMS pulse applied 75 ms after that. (Right) Perioral reflex by mechanically stimulating the vermilion of the upper lip with or without TMS. (Both) The two vertical dashed lines indicate 10 and 20 ms after TMS onset. The shaded area (gray) indicates the EMG activity enhanced by TMS. The solid line denotes the response in the 'perturbation with TMS' (PT) condition; the dotted line gives the response in the 'perturbation alone' (PN) condition. The horizontal dashed line denotes the background EMG activity level (BK). Reproduced from Ito, Kimura, and Gomi, 2005. © 2005 Lippincott Williams & Wilkins.

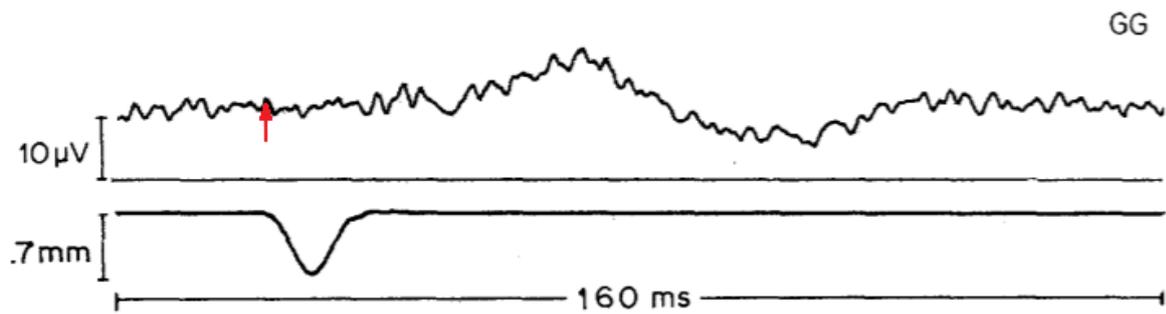


FIGURE 4.3: (Upper) Autogenic response of genioglossus (GG) muscle as averaged EMG record. Arrow indicates onset of stimulus. (Lower) Stimulus displacement. Reproduced from Weber and Smith, 1987. ©1987 American Speech Language Hearing Association.

In the 2005 paper however, using the same production task and perturbation as in the 2002 paper, the latency reported (and used for the calculation of the window of focus for the TMS) was 75 ms instead of ~48 ms. The reason of this discrepancy is that in the 2002 paper the latency reported was for the EMG pattern's onset while the 2005 paper used the value of 75 ms for the activity's peak (personal communication). This does mean that the activity for which they claim a cortical origin began around 48 ms post perturbation, which is longer than the evidence they cited from Pearce et al. (2003) of a brainstem-mediated long-latency stretch reflex with a latency of 34 ms.

The studies presenting latency information about various orofacial reflexes are summarized in Table 4.1.

4.1.2 Short-delay reflex loops in the tongue

Postural control in skeletal muscles is mediated primarily through a stretch reflex generated by muscles spindles. Although, as presented in Chapter 2 there is evidence of the muscle spindles (Cooper, 1953; Walker and Rajagopal, 1959) that could provide the basis for such a reflex response, but they were not found in each muscles and their density could be low, with the consequence that it is not clear whether they could significantly contribute to the postural control of the tongue.

Previous work done by Weber and Smith (1987) investigated reflex responses of "innocuous mechanical stimuli" applied to the tongue dorsum, lips, and jaw on the EMG activity of the genioglossus (GG) of the tongue, the masseter of the jaw, and the orbicularis oris inferior (OOI) of the lips. Smith et al. (1985) had previously found that such stimulation of the tongue would elicit activity in the muscles of the jaw and similar activity was found in the 1987 study. Additionally they observed what they called "autogenic" reflex activity in the GG when the tongue was stimulated that they claim had not been observed previously. In particular, they found that the EMG deflection comprising the reflex occurred between 31 and 43 ms after the stimulation.

TABLE 4.1: Latencies in orofacial reflex literature

latency (ms)	claimed origin	details	study
10.5±0.2	brainstem	SLSR - jaw masseter	pearce2003
12-18	(brainstem)	perioral	abbs1983
14-17	(brainstem)	perioral	weber1987
~15	mechanical	upper lip displacement	gomi2002
17.75±0.92	brainstem	perioral	ito2005
31-43	-	genioglossus	weber1987
34.0±1.4	brainstem	LLSR - jaw masseter	pearce2003
48.25	(cortical)	lip OOS EMG onset	gomi2002
~75	cortical	lip OOS EMG peak	ito2005
315.7±98.4	cortical (voluntary)	OOS EMG	ito2005

Citing Lowe, Gurza, and Sessle (1977) and Bratzlavsky and Vander Eecken (1974), Weber and Smith (1987) claim this was likely to be due to "superficial mechanoreceptors innervated by the lingual nerve rather than muscle spindle receptors innervated by the hypoglossal nerve". Indeed Bratzlavsky and Vander Eecken (1974) reported they were unable to "record by electromyography any reflex response in the genioglossal muscle of healthy subjects upon stretching the tongue or upon electrical stimulation of the hypoglossal nerve". Weber and Smith concede however that their experimental methods could not rule out the possibility of a stretch reflex.

It is still not clear whether tongue postural control is achieved with the contribution of short reflex loops to the brainstem, mediated either by muscle spindles or mechanoreceptors (Trulsson and Essick, 1997; Fitzgerald and Sachithanandan, 1979) or both.

4.2 Tongue perturbation modeling study

The biomechanical model described in Section 2.4, especially along with efforts to perform tongue perturbations in human subjects, presents an opportunity to make progress in answering questions about the existence of a stretch-like reflex in the human tongue. In fact the combined use of the two approaches has the potential for providing advantages over each approach used individually. For example, modeling alone has the disadvantage of having results of unknown fidelity to the real phenomena. That is, anything observed in the biomechanical model might be questioned as to whether it is an accurate representation of what really happens. Using a parallel study provides a baseline against which the model's behavior can be evaluated. Similarly, performing just a study on human subjects limits the experiments to what can be performed on a limited number of subjects and trials as well as limits the information which can be observed. Also, a realistic biomechanical model enables establishing a link between measured EMG activity and movement. Indeed the model allows for the time response of the physical system to be studied and an understanding of the delay between the onset of an EMG activity and movement.

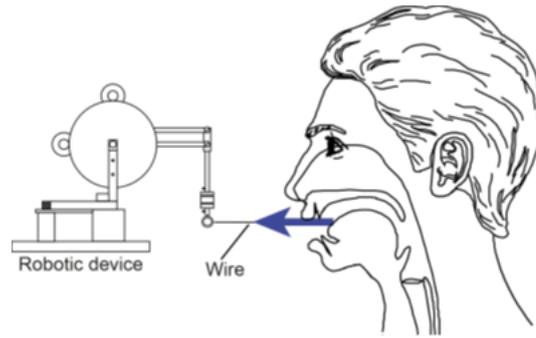


FIGURE 4.4: Graphical depiction of the tongue perturbation study on human subjects described in Caillet (2017) run in parallel to simulation of similar perturbation. The robotic device is attached via a wire to subject's tongue. Image courtesy of Takayuki Ito.

Using both in parallel allows for additional data to be analyzed, within the limits of accuracy of the model.

4.2.1 Experimental setup

Recently Jean-Loup Caillet, in his master's study, supervised by Takayuki Ito and Pascal Perrier, performed tongue perturbation experiments (Caillet, 2017; see Figure 4.4), in which the tongue is pulled anteriorly using a constant force of 1N exerted on the tongue body using a force effector attached to the superior part of the tongue blade of human subjects. For his study, subjects are instructed to produce and maintain a specified vowel (/i,e,E/) prior to, during, and after the perturbation is applied.

Preliminary results in human subjects

Preliminary data from these experiments show a phenomenon which is a good candidate for evidence of a reflex being involved in the perturbation response, notably a rebound in the horizontal displacement of the tongue. This rebound is visible in Figure 4.5 after the initial displacement caused by the perturbation. While it is clear that the tongue resists the perturbation, it is not clear whether this resistance is due entirely due to the passive, elastic characteristics of the tongue or whether it is due to a reflex mechanism, involving either the cortex or the brainstem.

Experimental question

In order to answer the question of whether the phenomenon observed in Ito's data is due to passive, elastic characteristics of the tongue or a reflex mechanism, we adapt the biomechanical model described in Chapter 2 to accurately model the same external perturbation used by Ito in his experiment.

Using this adapted model we will first simulate the basic conditions of the experiment. We will then modify the model so that we can modulate the level of

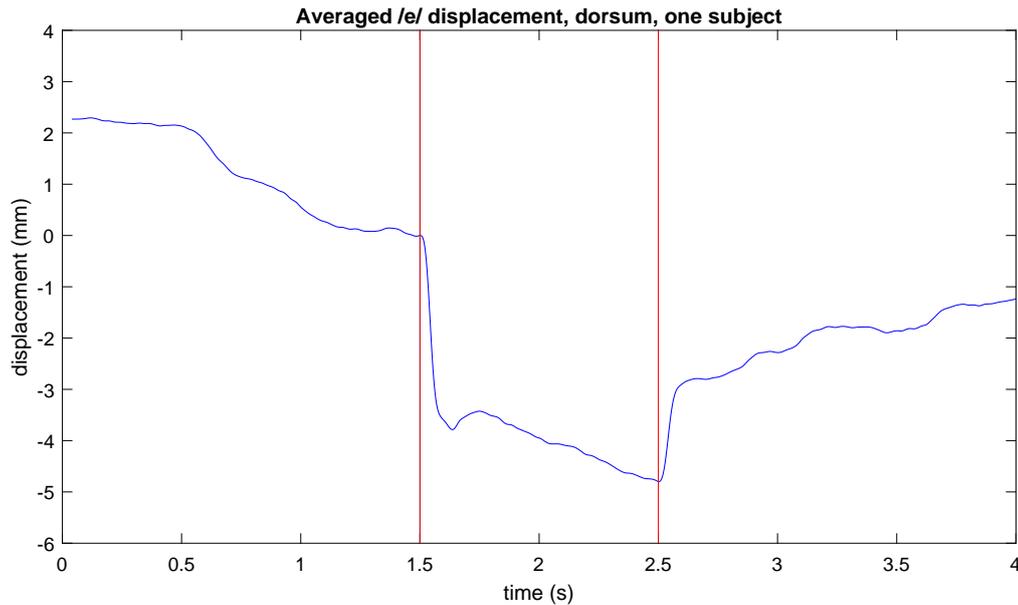


FIGURE 4.5: Data from one subject (averaged across instances of same conditions) showing the displacement of the tongue dorsum sensor in response to a force perturbation occurring during sustained production of the vowel /e/. The beginning and end of the perturbation are indicated with red lines. Displacement is zeroed at perturbation onset.

proprioceptive feedback, upon which a reflex would be dependent.

Three potential scenarios can be imagined:

1. rebound persists in absence of feedback;
2. rebound disappears with absence of feedback;
3. rebound persists in absence of feedback, but is modulated by feedback amount.

The first case would be highly suggestive of the rebound being mechanical in origin and due to physical properties of tongue alone. The second case would provide strong evidence in favor of the rebound being due entirely to a reflex mechanism, and the third case would suggest that a reflex exists and modulates physical characteristics to increase its effect.

In light of the work by Ito, Gomi, and Honda (2004), described above, which showed that stiffness characteristics were controlled for by muscular activation levels and that this was beneficial in perturbation responses to the orofacial musculature, our hypothesis is that the tongue response is similarly mediated by stiffness parameters, but the experimental question is whether the stiffness values on which the observed rebound depends is related only to the configuration-dependent activation, is entirely related on the updates to this activation via feedback, or whether it is related to the configuration-dependent activation but modulated by the feedback.

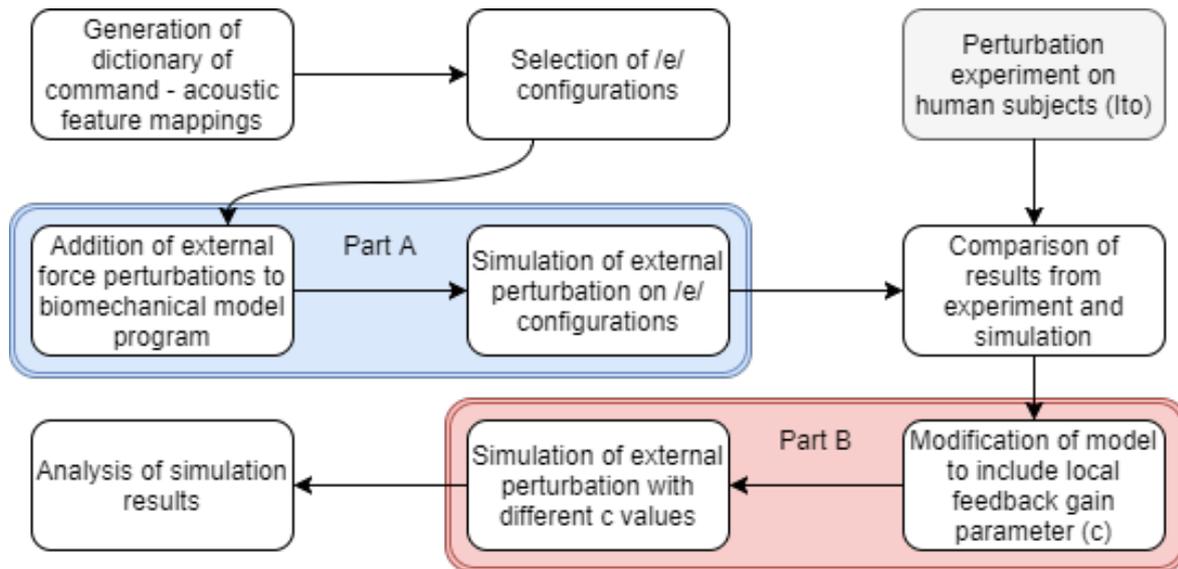


FIGURE 4.6: Block diagram of Study 2

4.2.2 Methodology

In order to address these questions, a two-part study will be performed, as outlined in Figure 4.6 and described below. In the first part, the biomechanical model will be used to simulation external perturbations and in the second part simulations will be performed with different levels of feedback, in order to characterize the relative contributions of reflexes and passive mechanical characteristics of the tongue in the observed behavior.

Addition of external force perturbation to biomechanical model

Modifications were first made to the biomechanical model program so that the application of an external force could be simulated. Parameters of both the force and its location of application were chosen in order to best approximate the study done in parallel on human subjects by Takayuki Ito and Jean-Loup Caillet, described and depicted above. Specifically these modifications allow for the application of a 1 N force for any duration to specified nodes on the surface of the tongue.

In order to distribute the force in a realistic manner, it was evenly divided between adjacent nodes on the upper contour and the two corresponding nodes one layer below this contour, so that each of these four nodes had a force of .25 N applied to it. Specifically, the surface nodes were located in the tongue blade region (Figure 4.7). Additionally, in order to have data from the simulation that can be compared with that from the the experiment, tongue movements were assessed in the regions anterior and posterior to the tongue blade region to which the force was applied.

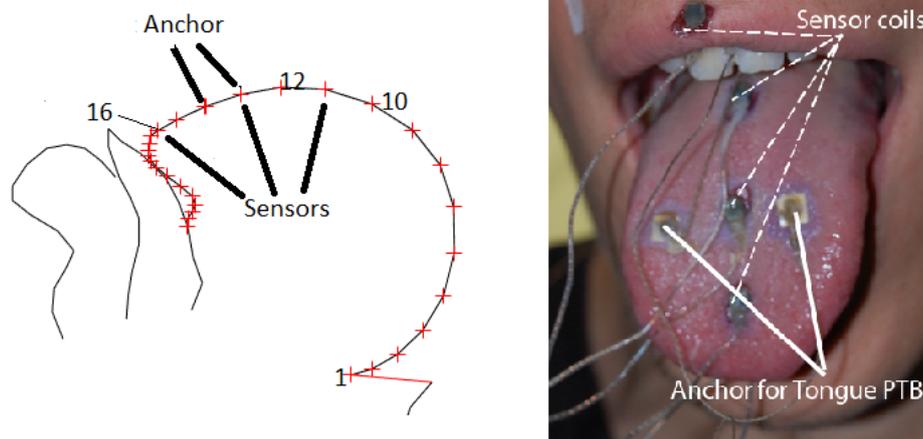


FIGURE 4.7: Depiction of anchor and sensor placement in both the simulation (left) and in Ito and Caillet's study (right). In the simulation, sensors correspond to the specific surface nodes indicated, while anchor positions indicated are where the force is applied on both the surface and first subsurface layer nodes. Image on the right is courtesy of Takayuki Ito.

Simulations - Part A

As in the study described in Chapter 3, the command space was sampled to obtain a number of configurations corresponding to different French vowels. Although Ito has performed the perturbation experiment in different vocalic conditions, in this preliminary study we have focused on the vowel /e/ to which 332 examples in our dictionary correspond.

While in Ito and Caillet's experiment described above the force perturbation was applied for a duration of 1 second, numerical stability concerns¹ motivated our reduction in duration of the applied force to 200 ms. It was verified however that the displacement stabilized within this time in most of the successful simulations.

For each configuration a simulation was run in which the transition time (time across which the commands changed from those giving the rest configuration to those which give the desired equilibrium position) was set at 150 ms, after which the commands were held constant. 100 ms later, at which point the majority of simulations had reached a stable equilibrium, the perturbation was applied instantaneously and held for 200 ms before being released, also instantaneously. 200 additional milliseconds of simulation were performed (unless the simulation failed, in which case the simulation ended at the release of perturbation) in which the tongue was allowed to return to its stable equilibrium position.

¹Finite element modeling requires solving differential equations with numerical methods, such as the adaptive Runge-Kutta method, which compute the state of the system at regular time steps, which are automatically determined. When the deformation of the Finite Element mesh is very slow these methods will choose to use a very small time step which can induce numerical instabilities. This problem is known as a "stiff" problem in ODE solving.

Using this strategy 194 of 332 configurations completed successfully and were identified as numerically stable.

Due to the large number of simulation failures or instability issues, and in order to verify the results where not heavily biased we generated command histograms for all simulations attempted and for failed simulations as well as force level histograms of the successful simulations. In particular, we used the sum of the modulus of the force matrix of the simulation at equilibrium as the measure of global force level of a given configuration.

Addition of local feedback gain parameter to model

In a second part, in order to investigate the role of proprioceptive feedback in the form of the stretch reflex as it is implemented in the λ -model, we made additional modifications to the model to be able to control how much of an impact the feedback had on the system's response to external perturbations occurring while in a stabilized equilibrium configuration. This was done by implementing a local feedback gain parameter.

As shown in Chapter 2, the force generated by a muscle fiber depends, in the model, on the muscle activation, A , and the feedback gain parameter c as:

$$F = \rho(e^{cA} - 1) \quad (4.1)$$

where $A = L - \lambda + \mu\dot{L}$ and L is the current length of the muscle fiber.

These can be combined as:

$$F = \rho(e^{c(L-\lambda+\mu\dot{L})} - 1) \quad (4.2)$$

In order to study the effect of perturbations on a particular configuration giving rise to a specific vowel, it was necessary to implement a modification to the feedback gain mechanism which would not modify the equilibrium configuration. The feedback gain parameter could have been modified from the beginning of the simulation, but in order to perturb configurations of a specific vowel, then configurations corresponding to the vowel of interest would first have to be found for each value of c , a process which requires exploring the command space and excessive computational time. Moreover, this technique would not easily allow for configuration-specific comparisons of the perturbation response.

Accordingly a modified force equation was introduced as follows, which allows for variation around a local equilibrium configuration:

$$F(t) = \begin{cases} \rho(e^{c(L-\lambda+\mu\dot{L})} - 1) & \text{if } t < t_{target} \\ \rho(e^{c(L_{target}-\lambda)+c_{modified}(L-L_{target}+\mu\dot{L})} - 1) & \text{if } t \geq t_{target} \end{cases} \quad (4.3)$$

where $c_{modified}$ is the local feedback gain parameter which takes effect after a specified point in the simulation, t_{target} , chosen for when the tongue has settled into the target vowel configuration, in practice chosen to be the moment the perturbation is applied. At this point in time, the muscle activation level is also frozen, by saving

and freezing L_{target} . From this time onward any modifications in length of the muscles will not change the force that the muscle is generating through the original c value, but instead through the new $c_{modified}$ value that is only applied to deviations from the target position. Since the new parameter only affects local changes around the equilibrium position, we call it the *local* feedback gain parameter.

Indeed the same equation can be written as:

$$F(t) = \begin{cases} \rho(e^{c(L-\lambda+\mu\dot{L})} - 1) & \text{if } t < t_{target} \\ \rho(A_{target}e^{c_{modified}(L-L_{target}+\mu\dot{L})} - 1) & \text{if } t \geq t_{target} \end{cases} \quad (4.4)$$

where A_{target} more clearly shows that the first component of that case serves to provide the activation level necessary to reach the equilibrium configuration with the "normal" gain factor in the first step of the movement.

Simulations - Part B

Simulations were then performed with the modified model across a range of values of the local feedback gain parameter for each of the configurations mentioned above, using the same time parameters as before. In practice the original value of c was 0.112 and $c_{modified}$ was set as $c \times c_{factor}$. Below and in the analysis c_{factor} will be referred to how much c changed from the original value.

Analyzing timecourse of perturbation response

Additionally, in order to further investigate the nature of the observed patterns in the experimental data and the performance of the biomechanical model, several measures were extracted from the data. This information is depicted in Figure 4.8:

- time and magnitude of peak displacement (red in Figure 4.8)
- time and magnitude of perturbation rebound (green in Figure 4.8)
- deflection velocity (slope of orange line in Figure 4.8)
- deflection 0-intercept (pink point in Figure 4.8)

4.2.3 Results and Analysis

Numerical stability of model and representativeness of results

First, by comparing the command histograms (Figure 4.9), and in particular the results of the Wilcoxon rank sum test, it can be concluded that in 4 of the 9 cases there is a statistically significant difference in the distributions, meaning there is a correlation between failure and certain command values. Nonetheless it only seems to be the case that for the Verticalis there is a region of the command space which is not covered by the successful commands. Almost all of the simulations with Verticalis commands below 10 failed, just as those with GGA below 35 had a significant tendency to fail. This seems to be due to high stiffness in these configurations or to the

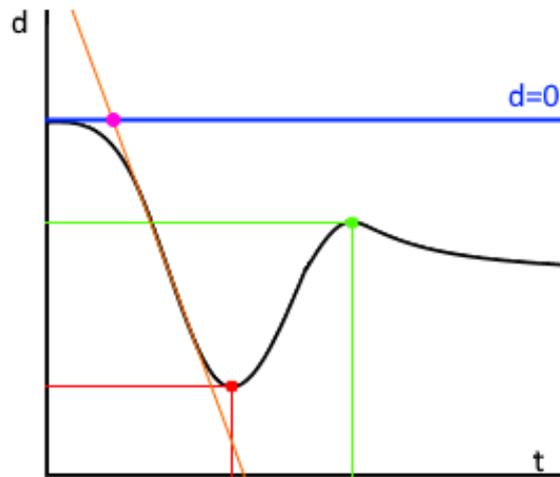


FIGURE 4.8: Depiction of information extracted from simulation results. X-axis is time and y-axis is deflection distance (lower is forward). Red dot represents the point of maximum deflection due to the perturbation. Green dot represents the point of maximum rebound. Orange line represents the linear interpolation of the deflection, the slope of which corresponds to deflection velocity. The pink dot is the intersection of this line with the 0-deflection line and is a measure of the time of deflection onset.

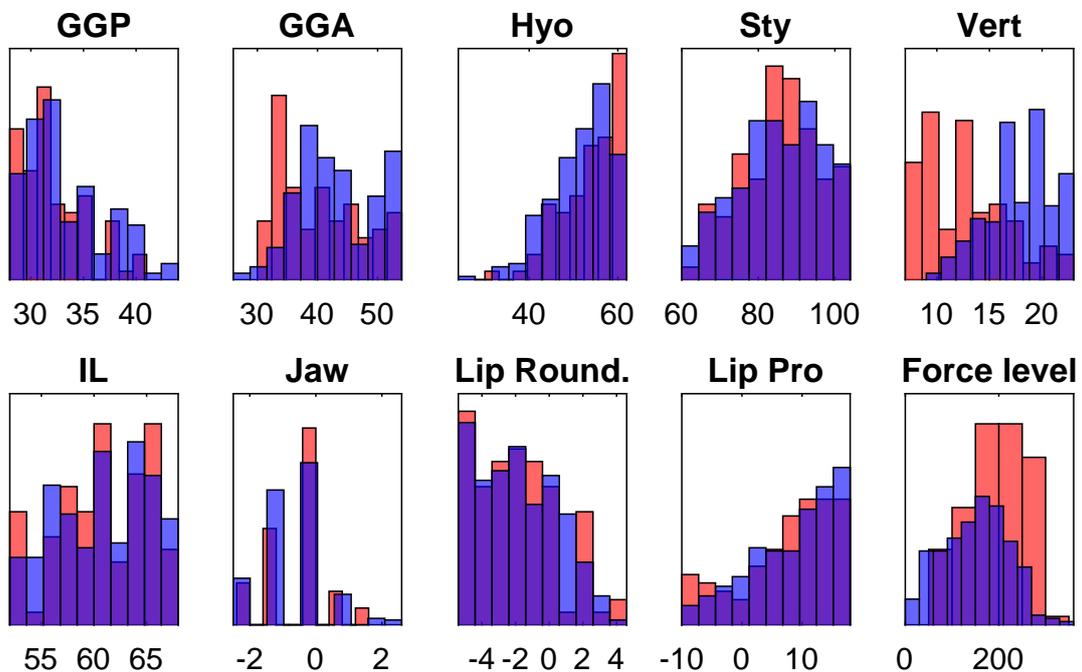


FIGURE 4.9: Command (and goal force level, bottom right) normalized histograms for successful simulations (blue) and failed simulations (red).

TABLE 4.2: Results of Wilcoxon rank sum test of distributions depicted in Figure 4.9. The null hypothesis is that data samples corresponding to failed simulations and successful simulations are from distributions with equal medians.

Value	null hypothesis	p-value
GGP	reject	0.0157
GGA	reject	0.0008
Hyo	reject	0.0075
Sty		0.9908
Vert	reject	2.2e-21
IL		0.9500
Jaw		0.1164
Lip Round.		0.7546
Lip Prot		0.3842
Force level	reject	9.9e-6

possibility they could be cases in which perturbation induced contact of the tongue against the lower teeth, with both cases increasing the likelihood of numerical instability.

Although the distribution of commands differs slightly between the successful and failed cases, other than the lower end of the Verticalis commands, the examples cover the entire range of values. This interpretation is confirmed by looking at the force levels. A reasonable range of commands and force levels is represented in the data, although it cannot be claimed that the commands of successful simulations are a random subset of the command space; there are correlations between specific command combinations and simulation failure. This observation will serve as a caveat in the statistical interpretation of the following results.

To summarize, simulations with a high level of force, in particular in the GGA or Verticalis muscles, and the consequent dental contacts, were more likely to encounter numerical issues. This is not a problem however since there remains good coverage of the command space and force levels in the successful simulations. Moreover, since it is expected (and shown below in Figure 4.14) that higher force levels involve lower displacements and displacements seen in simulations were already lower than in experimental data, it was decided to proceed with at least good coverage of the command space.

Part A

The results of the 187 numerically stable simulations (Figure 4.10) done without modification to the feedback gain parameter were analyzed. As can be seen in Figures 4.11 and 4.12, the simulation data qualitatively match those of human subjects, most significantly in the existence of a post-perturbation rebound.

Additionally it was observed that the magnitude of rebound is related to peak displacement by what appears to be a linear relationship (Figure 4.13) and that peak displacement and stable displacement seem related to force level (Figure 4.14).

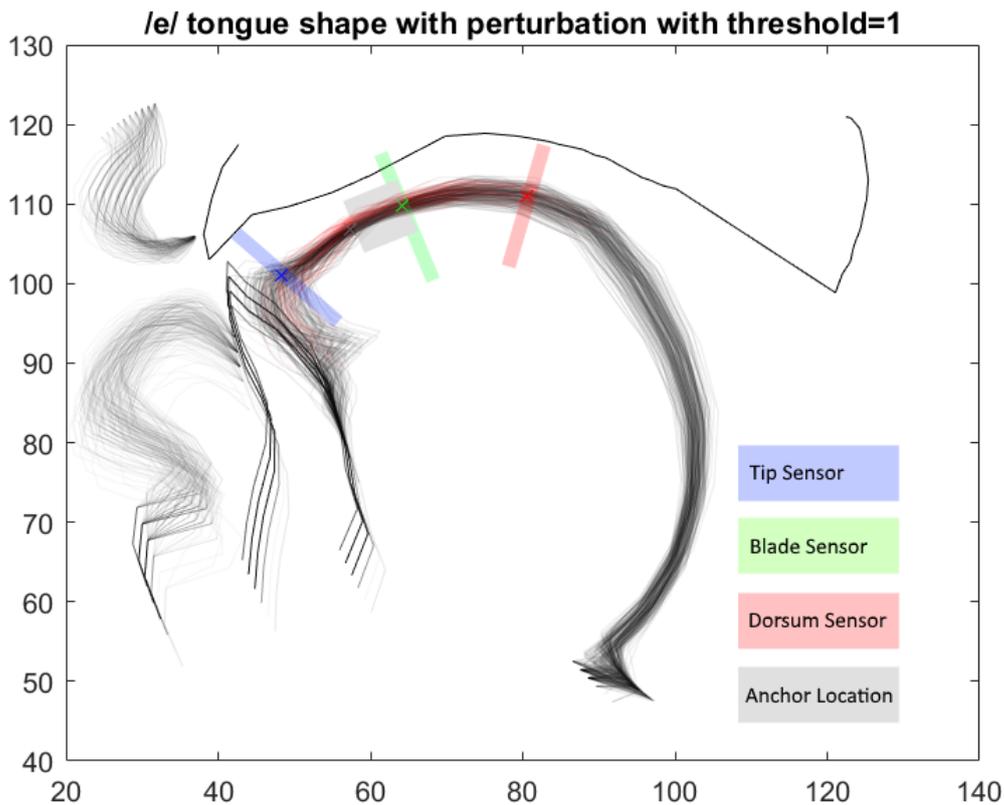


FIGURE 4.10: Depiction of tongue, jaw, and lip configurations giving rise to /e/ along with sensor locations and effect of perturbation (red shading on tongue contour indicates location of displacement of tongue surface resulting from perturbation).

Despite this qualitative coherence with the experimental data, there were several quantitative issues with the simulation results.

First, displacement was typically smaller than seen in preliminary data from human subjects. For example, as can be seen in comparing the data presented in Figure 4.5 with that in Figures 4.11 and 4.12, the displacement in the case of human subjects is approximately 3.5 mm while in the case of the simulation it is often below 1 mm. Looking at Figure 4.12 though we see that there are numerous cases where the peak displacement is greater than one, even if most are not. This could be because peak displacement and stable displacement are related to sumforce (Figure 4.14).

Second, the deflection velocity seen in simulation does not seem to match that calculated from the experimental data. As can be seen in the bottom row of Figure 4.15, deflection velocity in the simulations was approximately 10-30 mm/s, depending on sensor, while it was 62 mm/s for all three sensors in the experimental data.

Lastly, there are questions regarding deflection onset timing and force propagation velocities. As visible in the top row of Figure 4.15, the deflection onset timing (estimated by calculating the time at which the deflection slope intercepts

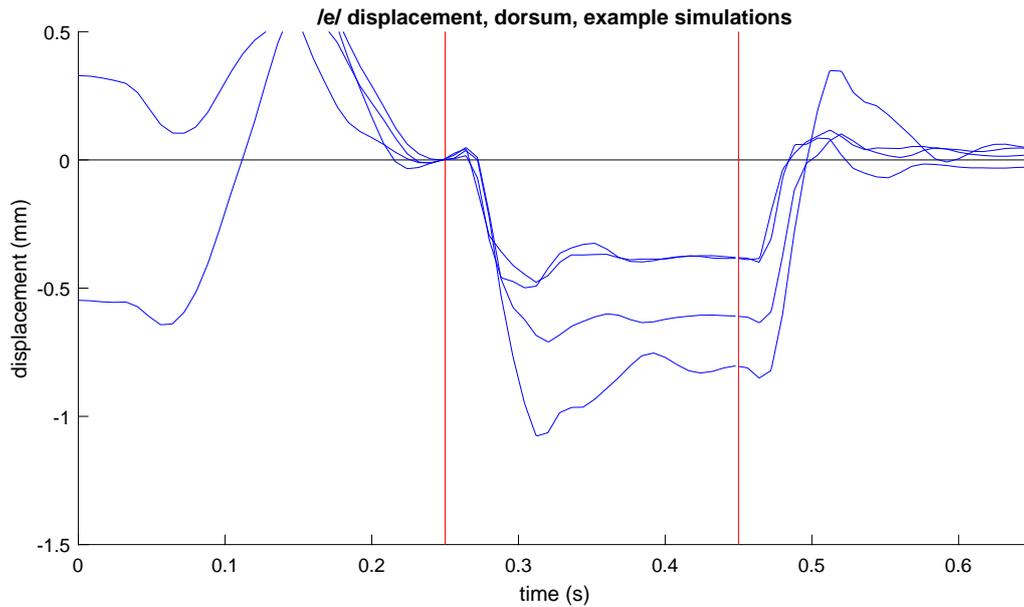


FIGURE 4.11: Simulated perturbation responses for 4 different configurations with perturbation beginning and end marked with red lines and displacement zeroed at perturbation onset. Large movement before perturbation onset reflects tongue moving into position to produce /e/.

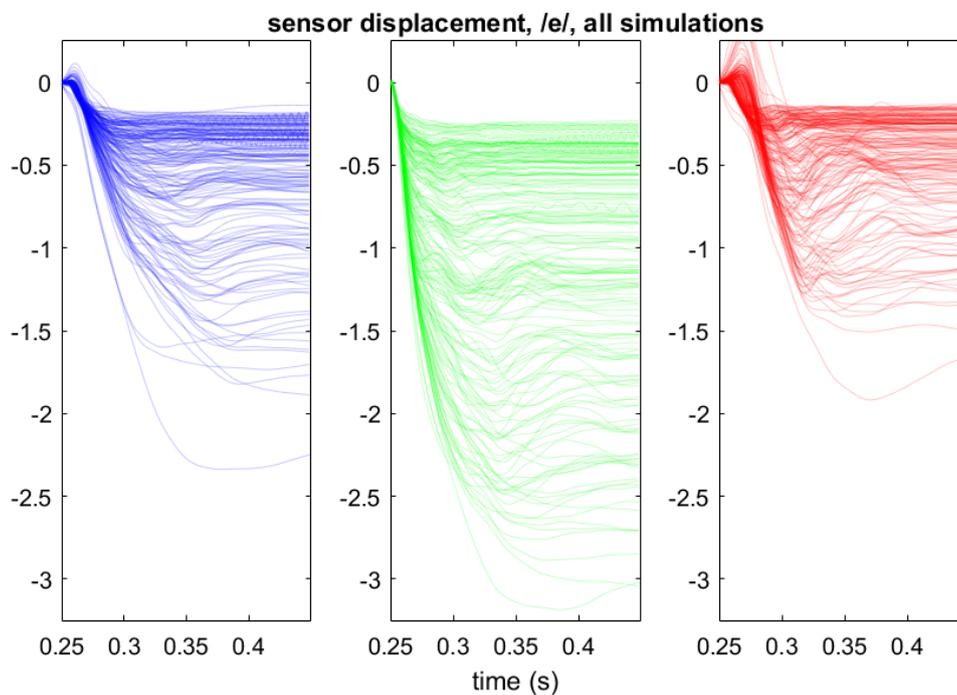


FIGURE 4.12: Horizontal sensor displacement for all successfully simulated configurations of /e/ (negative values correspond to forward displacement). Sensors are tip (left in blue), blade (center in green), dorsum (right in red). Value of x-axis is simulation time, where perturbation begins at 0.25 seconds.

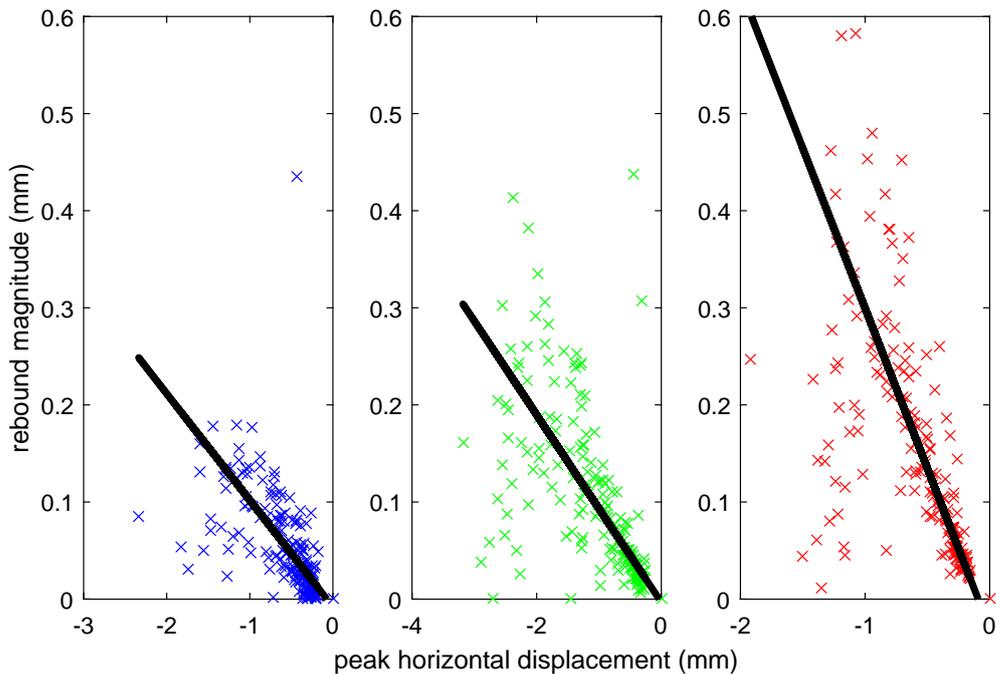


FIGURE 4.13: Relation between peak horizontal displacement and rebound magnitude for tip (left in blue), blade (center in green), dorsum (right in red). Black line is robust linear regressor (calculated using Matlab `robustfit` routine).

the level of zero deflection) in the simulations only seems to correspond to the experimentally-obtained data for the case of the tongue tip sensor. For the other cases the onset seems to happen too quickly after perturbation onset, especially in the case of the blade sensor. Since this sensor is located at the point of force application and the tongue dorsum is located approximately 20 mm anterior to the blade sensor and force anchors and the deflection does not appear here until 5-10 ms later, one could question whether forces are propagated through the material of the tongue at the correct velocity.

Further work would be needed to definitely determine what has caused this discrepancy between the experimental data and the simulations, but one possible explanation that would be consistent with the observations is that the tongue stiffness used in the model seems higher than in human subjects. Indeed higher stiffness would reduce the amplitude of the deflection and shortens the period of the oscillations caused by the abrupt application of the force to the tongue.

Other possible causes include that the experimental data obtained in subjects is itself not precise enough to make such conclusions about the time of deflection onset between different sensors (it would suffice that the sensor data not be perfectly synchronized or that the post-processing assume equal deflection onset time to cause such a discrepancy), that some parameters of the biomechanical simulation are incorrect, that the underlying model of the tongue as an incompressible fluid is not sufficiently accurate at this level of analysis, or that the simplifications made in the

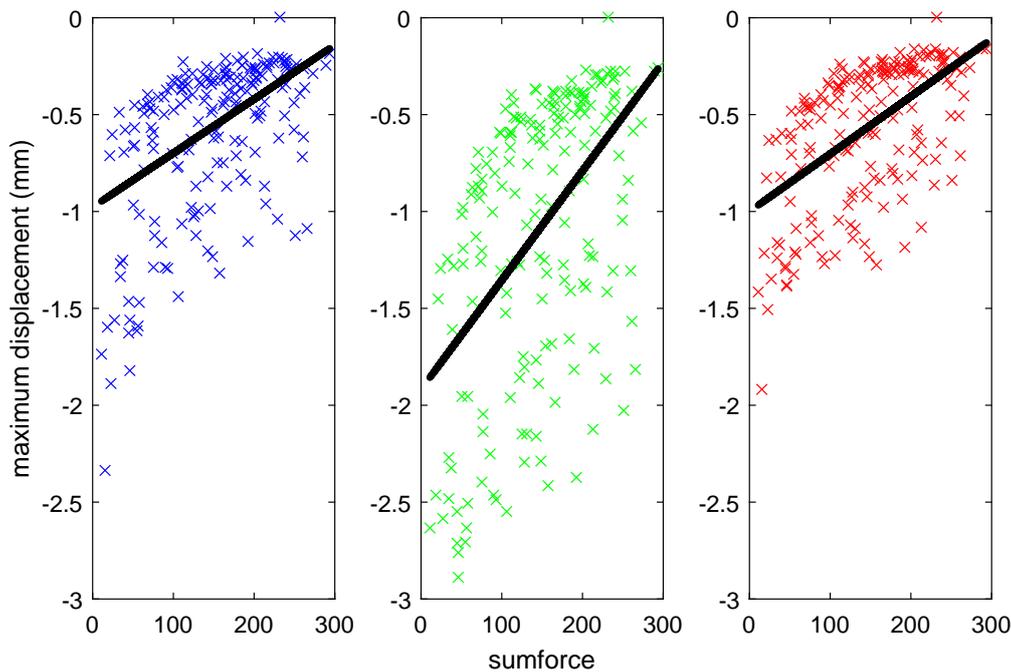


FIGURE 4.14: Relation between force level measure (sumforce) and maximum displacement for tip (left in blue), blade (center in green), dorsum (right in red). Black line is robust linear regressor (calculated using Matlab robustfit routine).

2D model prevent the model from capturing possible 3D deformations.

One issue is that there has not been such precise data (or at least data that deals with such precise parameters) regarding the tongue against which to compare the physical accuracy of the biomechanical model. Further experimental data and confirmation of the values extracted from the preliminary experimental data will be useful in narrowing the search for an explanation of these discrepancies.

For this reason, the following analysis contains an additional caveat regarding these potential limits of the precision of the experimental data and of the biomechanical model.

Part B

Figures 4.16 and 4.17 show the simulations across different $cFactor$ values for two example configurations.

For the first round of simulations (with $cFactor = 1$), a linear relationship was found between the peak displacement and rebound magnitude (Figure 4.13) meaning the greater in magnitude the displacement, the greater the magnitude of the rebound.

It was observed that as $cFactor$ increases, there is a decrease in peak displacement (magnitude of maximum deflection of sensor before rebound; Figure 4.18-Left)

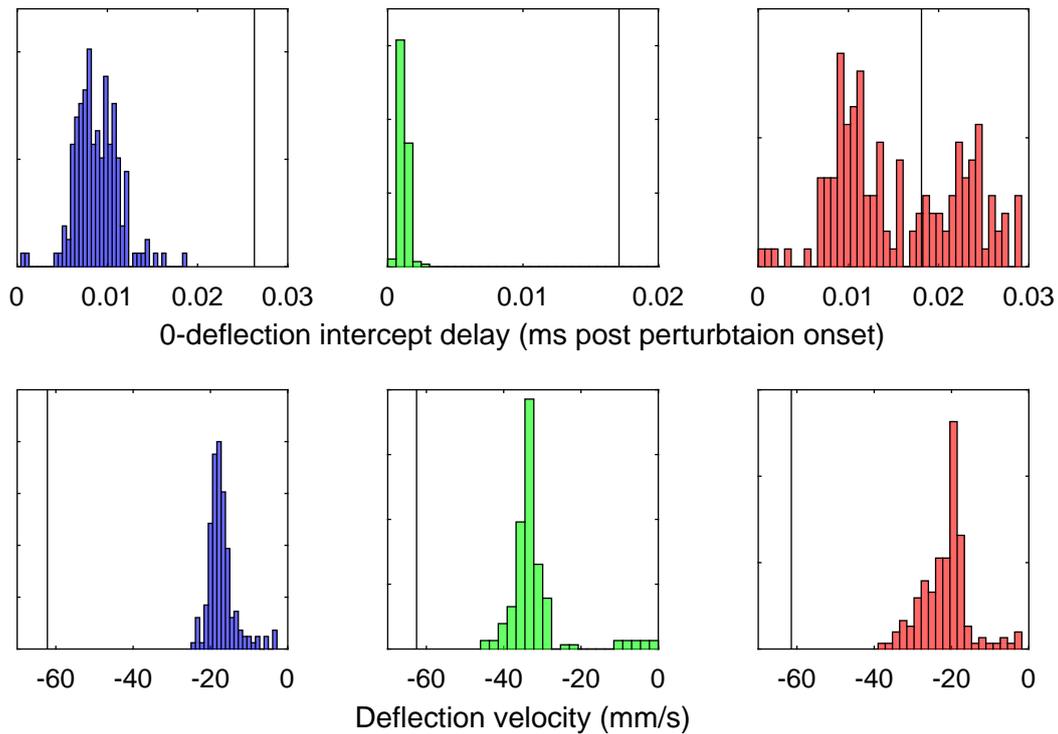


FIGURE 4.15: Normalized histograms of simulated 0-deflection intercept delay (top row) and deflection velocity (bottom row) for tip (left in blue), blade (center in green), dorsum (right in red). Vertical black line is value from experimental data. 0-deflection intercept delay is the intercept of the linearized deflection curve with the 0-displacement line and provides an estimate of displacement onset. Deflection velocity is estimated as slope of linearized displacement curve. The significance of the values is explained in Figure 4.8

at the same time as an increase in the rebound magnitude for all sensors (Figure 4.18-Center). This is confirmed by the fact that as $cFactor$ increases so does the ratio between rebound magnitude and peak displacement magnitude (Figure 4.18-Right). This means that the increase in rebound magnitude is not due to the change in peak displacement and suggests that it is modulated, or more precisely, amplified by c .

In the case where $cFactor$ was zero there is a small rebound and considerably increased peak displacement. This activity is due to the underlying mechanical and elastic characteristics of the tongue, in particular, the passive stiffness of the tongue which determines the frequency of any oscillations resulting from the sudden application of force and the damping factor which determines the settling rate of these oscillations. As seen in Figure 4.17, even in the case where $cFactor = 0$, there is a rebound which is actually an oscillation that continues due to insufficient damping.

Generally it was observed that increasing c has not only the effect of amplifying the rebound but also of limiting the deflection. As the force is applied, the deflection of the tongue in the direction of the force is limited and the rebound increases. This

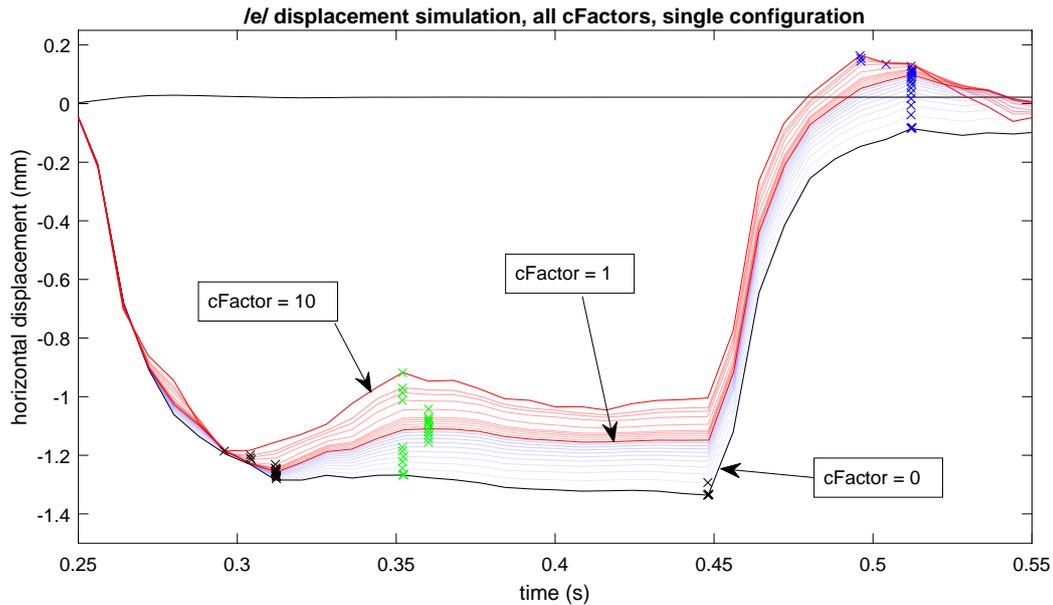


FIGURE 4.16: Effect of cFactor on perturbation response of single configuration. Perturbation occurs at .25 seconds. Red lines depict simulations for which cFactor was 1 or greater (cFactor = 1 and cFactor = 10 are labelled) and blue lines depict simulations for which cFactor was less than 1, except for cFactor = 0 which is depicted in black. Black Xs indicate peak displacement on [0.25,0.45] interval, green Xs indicate peak rebound, and blue Xs indicate peak release rebounds.

is consistent with a reflex which seeks to counteract the force applied to the tongue beyond just the mechanical characteristics of the tongue.

Since the rebound occurs within the same time window in both the experimental data (Figure 4.5) and the simulations (Figures 4.11, 4.12, and 4.19) and this rebound activity is modulated by proprioceptive feedback, the interpretation of the experimental data as being compatible with a reflex mechanism is supported.

Cortical or brainstem reflex

While the above analysis suggests the phenomenon observed in the experimental data is indeed compatible with a reflex mechanism, the question of whether it is of a cortical or brainstem origin remains open.

In order to further investigate this question, the time course of the reflex responses observed in the model must be investigated more closely. Since the model provides an estimate of the reaction time of the tongue in accordance with the hypotheses of its dynamical properties, the analysis of the reflex responses in the model will provide insight into the question of the origin of that which is interpreted as a reflex phenomenon in the experimental data.

It should be noted that the model embodies biomechanical hypotheses and not mechanistic hypotheses with regard to the pathways of proprioceptive feedback enabling a reflex response. More precisely, the model, in particular its use of the

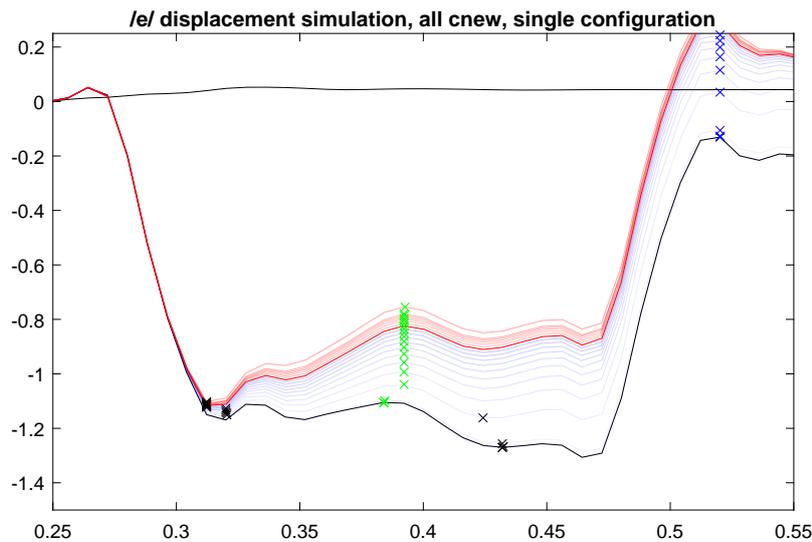


FIGURE 4.17: Effect of cFactor on perturbation response on an additional configuration, one in which the step response to force perturbation is an oscillation which persists even without feedback.

lambda model, can be seen as implementing the stretch reflex with no delay. Due to this lack of delay, it can be stated that the model implements a fast reflex mechanism.

Nonetheless, in modeling the dynamics of the tongue in a realistic manner, the model allows for the estimation of the time of the rebound which is due purely to the dynamical properties of the tongue. Since the data acquired from human subjects is in the form of the displacement of the tongue, having the additional knowledge of the time of the dynamics of the tongue will allow for constraints to be placed on the origin of the muscular activity.

In Figure 4.19 (top row) it can be seen that within 30-80 ms of the onset of the force perturbation peak displacement of the tongue tip and blade sensors occurs, and within 40-100 ms this is seen for the dorsum sensor as well. The rebound arrives 10-80 ms later (Figure 4.19 bottom row).

While this information serves as an end-to-end temporal constraint, the advantage of the biomechanical model is that it allows for the time course of the different processes responsible for the generation of these movements - muscular activation, force generation, and displacement - to be looked at in detail.

In order to investigate the time course of these processes, the relative effect of feedback must be isolated, since the passive characteristics of the tongue cause changes in the force and displacement in response to a perturbation as well. This can be done by comparing these values against the case when cFactor=0, since in that case there is no proprioceptive feedback.

For the case of the muscle activation level, this is relatively straightforward. The muscle activation levels shown in Figure 4.20 in response to the perturbation, correspond to the activation levels due to feedback, since, according to Equation 4.4

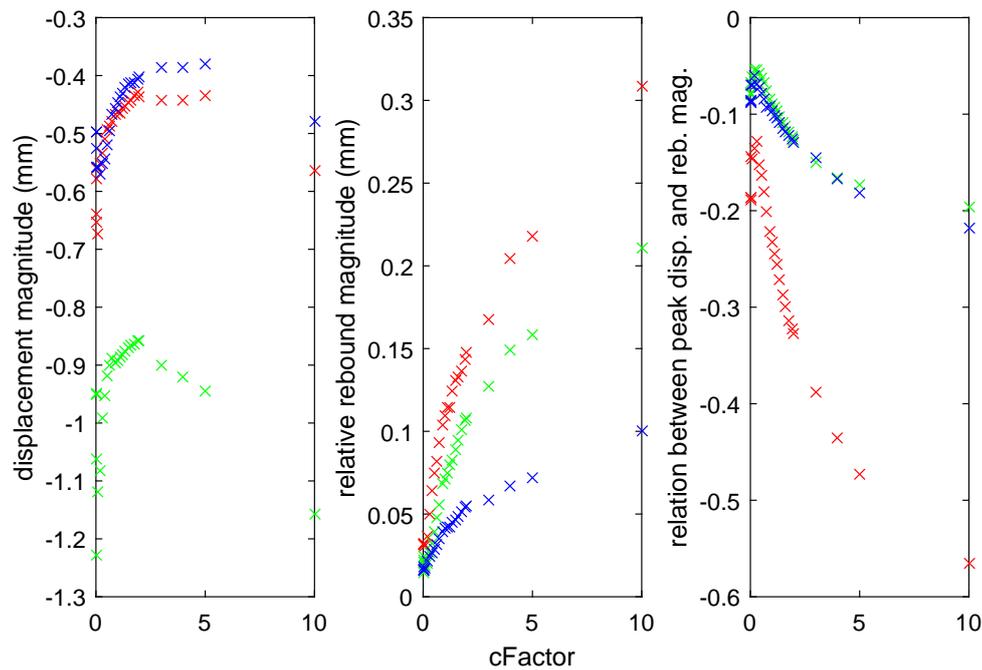


FIGURE 4.18: Effect of cFactor on displacement magnitude (Left), relative rebound magnitude (Center), and slope of robust linear regressor characterizing relation between peak horizontal displacement and rebound magnitude (See Figure 4.13 for calculation of this slope for cFactor = 1) for tip (blue), blade (green), dorsum (red) sensors.

when cFactor=0 the activity level is zero.

Accordingly, it can be extracted that muscle activation levels increase as a result of the external force perturbation in less than 10 ms for the GGP, GGA, and Verticalis, within around 20 ms for the Sytloglossus and IL. It appears that the activation of the Hyoglossus might also begin within 10 ms, but its activation level is relatively low and this cannot be conclusively shown without further analysis.

For force levels, even without changes in the level of muscle activation due to feedback, the muscles are active and generating force in the configuration to which the perturbation is applied. In Figure 4.21 the onset of forces resulting from the perturbation is seen in blue to occur quite early, while in Figure 4.22 the relative contribution of feedback is shown by subtracting the force present in the simulation with cFactor=0. From this it can be seen that feedback-induced force increases occur around 10 ms for numerous muscle fibers.

Lastly, the same analysis can be done of the displacement of the sensors. As can be seen in Figure 4.23, although the displacement occurs quickly, especially for the blade sensor, the displacement which is due to feedback does not appear to be significant until approximately 30 ms after perturbation, and mostly reaching maximal values between 50 and 100 ms.

Since there is no feedback delay in the model, the main utility of the model is

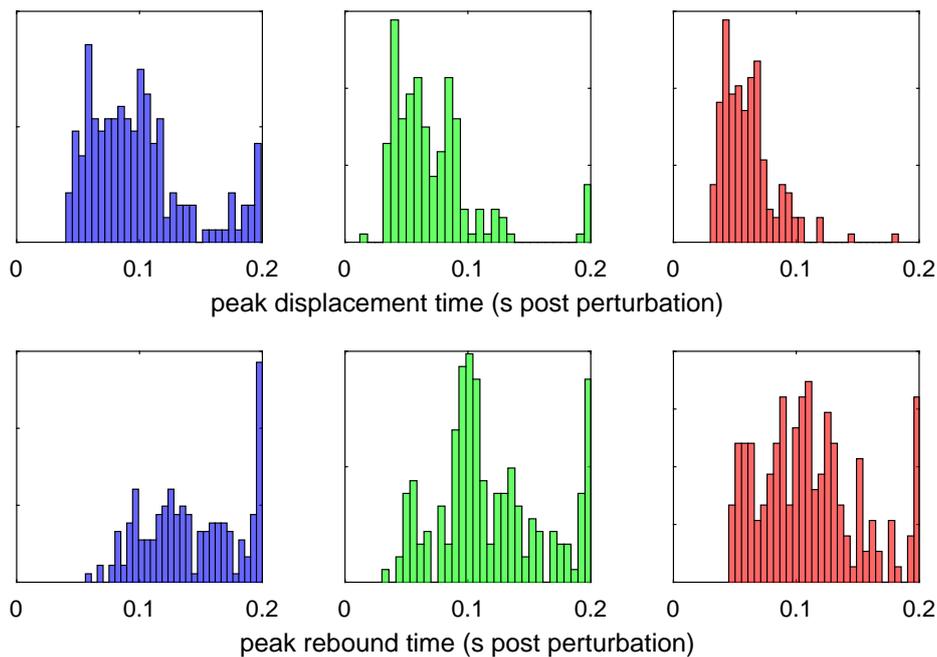


FIGURE 4.19: Normalized histograms of time of peak displacement (top row) and time of peak rebound (bottom row) for tip (left in blue), blade (center in green), dorsum (right in red) sensors.

to provide an estimate of the amount of time it takes that feedback to have an effect on observable values, such as displacement. According to the analysis above, where muscle activation increases were seen within 5 ms for some muscles and a measurable impact of feedback on displacement was observed starting around 30 ms and increasing until around 100 ms post perturbation, the time it would take for feedback to leave a measurable sign on displacement is approximately 30-90 ms.

Since deceleration of the displacement, which could be due to feedback, in the experimental data begins around 50 ms after the force perturbation begins, and the rebound peaks around 180ms, the range of possibilities for feedback onset given these constraints is from approximately 20 - 90 ms. Due to this large range it is not possible to establish with the existing data whether the reflex leading to the rebound is cortical or not. The fact that deceleration occurs in a time frame compatible with a fast reflex is however suggestive that there is at least a possibility of an initial rapid response. Even if this initial modulation of the original displacement happens too quickly for it to be the result of any cortical influence, it could be followed by a long latency or transcortical reflex.

Experimental hints from biomechanical model

Since further research will be necessary, it is useful to look at the muscle activation levels as suggested by the biomechanical model in order to gain insight into which muscles' activation might contribute most significantly to a the reflex movement.

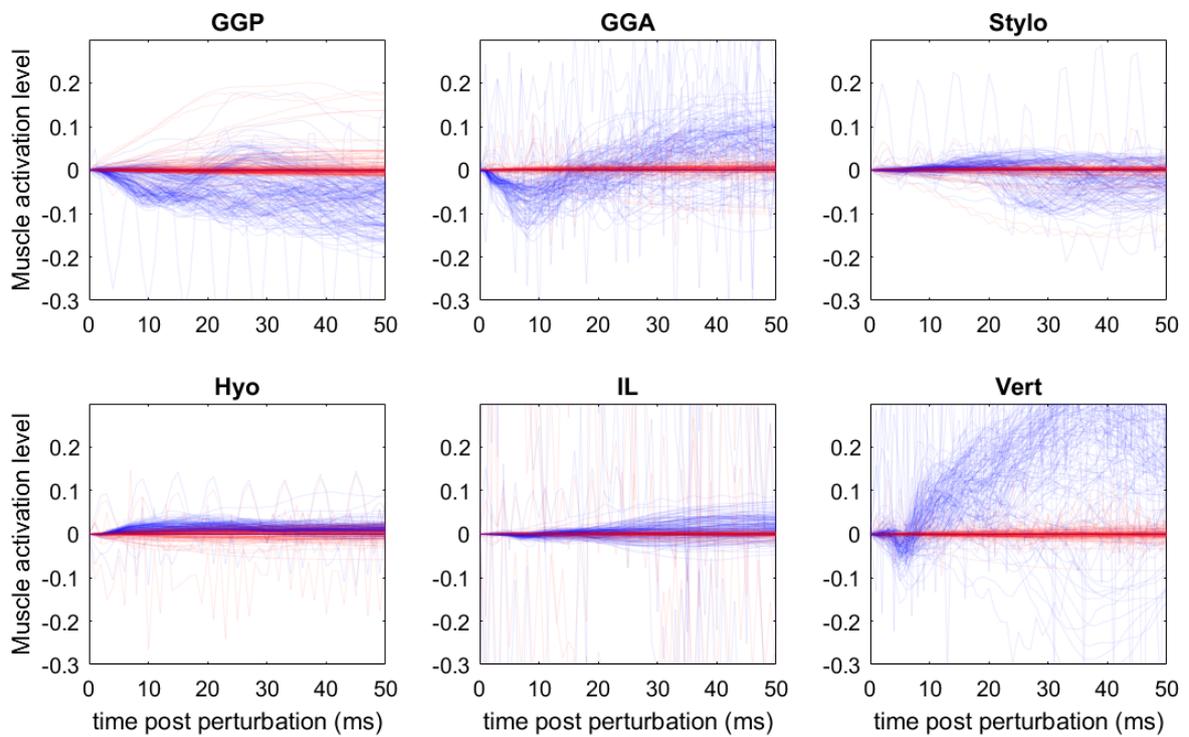


FIGURE 4.20: Muscle activation level with perturbation (blue) and without perturbation (red) by muscle. Activation levels for GGP, GGA, and Verticalis muscles are clearly significantly affected by the perturbation within 10 ms of perturbation.

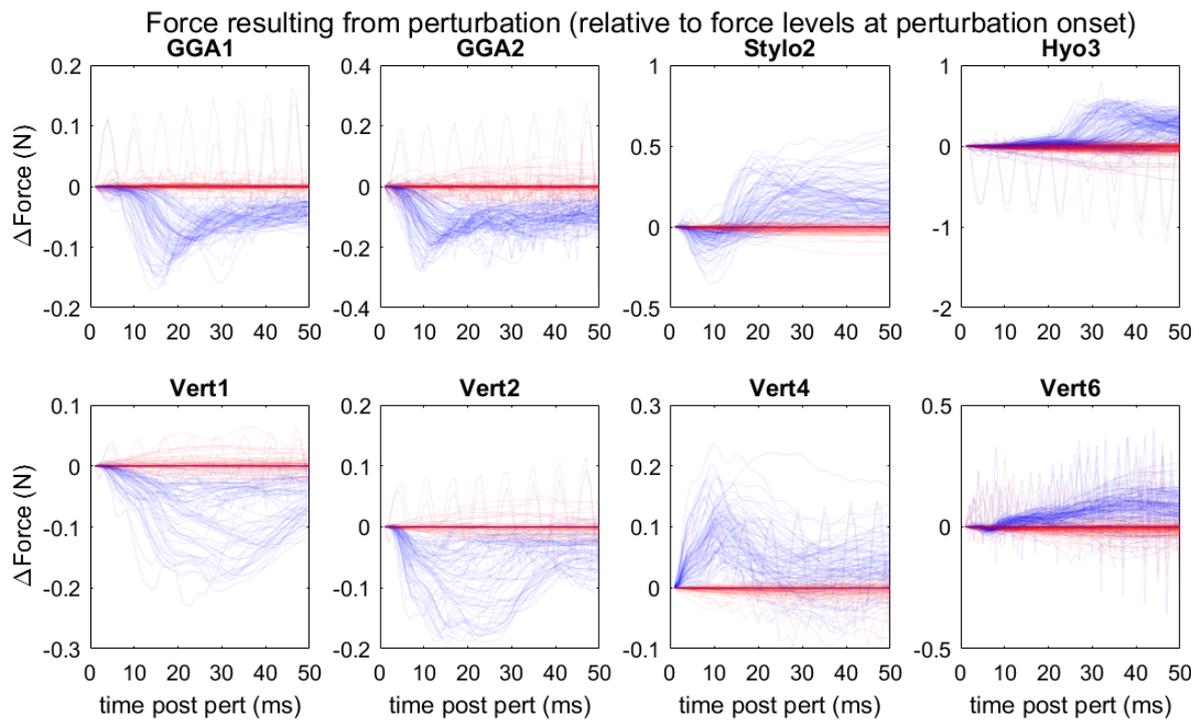


FIGURE 4.21: Force levels for specific fibers in case of perturbation (blue) and without perturbation (red). Forces are shown as relative to force level at perturbation onset.

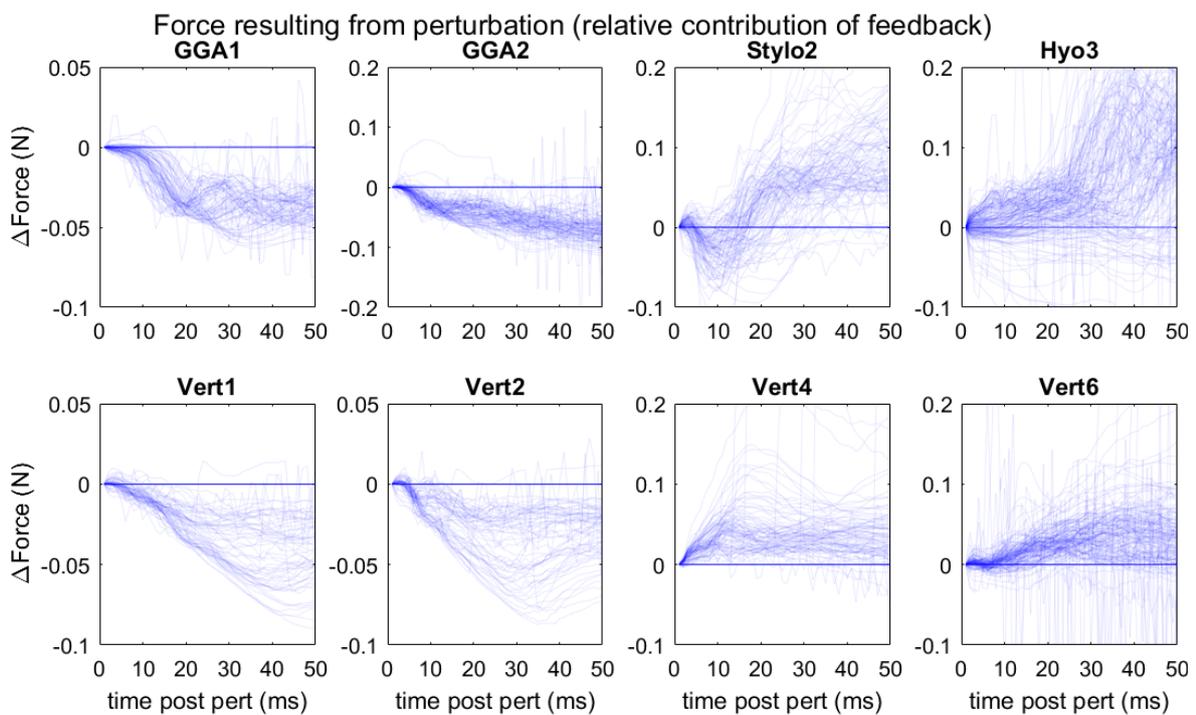


FIGURE 4.22: Relative contribution of feedback to force levels for specific fibers in case of perturbation.

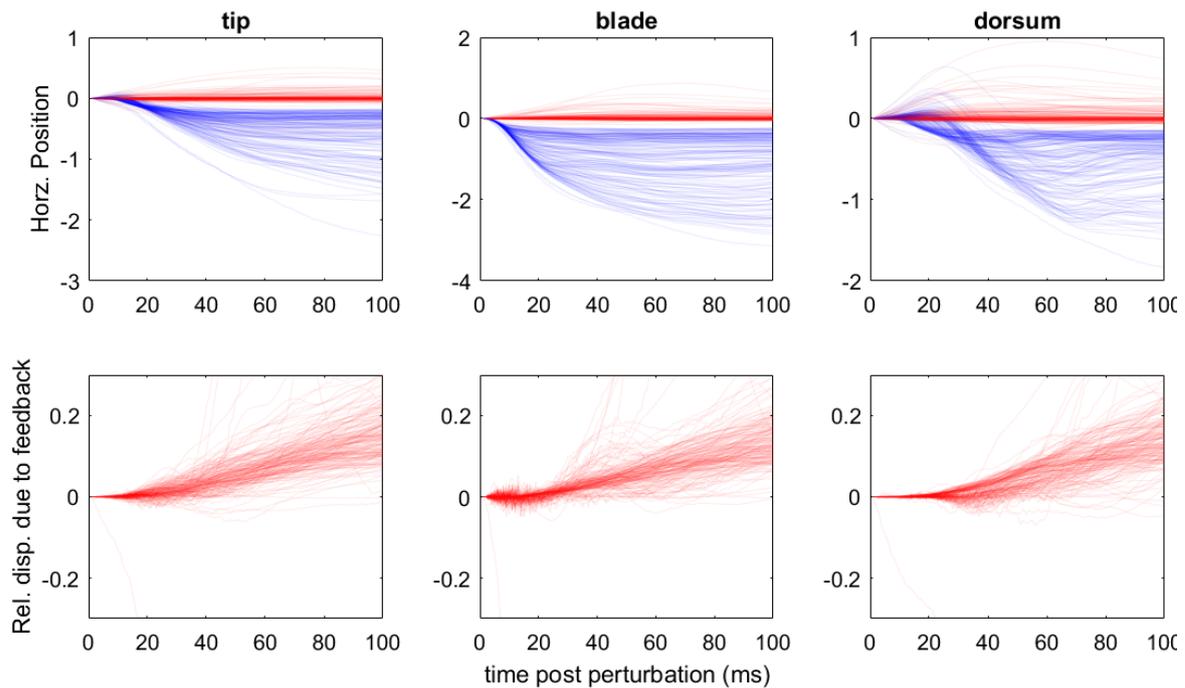


FIGURE 4.23: Upper row: Displacement resulting from perturbation (blue) and without perturbation (red). Bottom row: Displacement due to feedback, calculated by subtracting displacement seen in case with $cFactor=0$; Negative values indicate anterior movement of tongue. Positive values in bottom row thus indicate effect of feedback is to oppose anterior movement.

Figure 4.20 shows that the Verticalis seems to have the highest level of activity as a response to perturbation. It also reacts quickly and in an interesting way; its activation level first seems to decrease for a few milliseconds before increasing significantly. The GGP and GGA also have significant responses, with the GGP having a relatively lower activation level following perturbation and the GGA initially having lower activation but around 20ms post perturbation that level increasing above baseline.

The Verticalis muscle as modeled seems to react most significantly to perturbations of this location, which is coherent with the configuration of the tongue at the time of perturbation (Figure 4.10) and the Verticalis muscle (Figure 4.24).

This suggests that if electromyographical recordings are attempted in future work they should be done in such a way as to capture the activity of the Verticalis.

4.2.4 Conclusion

Experimental perturbation data show a rebound effect and the biomechanical model modified to simulate the same perturbation produces qualitatively similar results using baseline parameters.

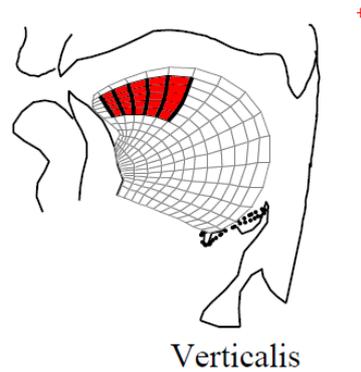


FIGURE 4.24: Verticalis muscle in biomechanical model. Reproduced from Perrier et al. (2003). ©2003 AIP Publishing LLC

Manipulation of the feedback gain parameter c indicates that the rebound phenomenon is modulated by c . Simulations with $c = 0$ show minor rebound which can be interpreted to be the mechanical-elastic characteristics of the tongue.

These facts suggest that feedback is responsible for the presence of this rebound phenomenon.

Consistent with the findings of Ito, Gomi, and Honda (2004) and their interpretation of previous experimental data, the present results show that the rebound persisted in the absence of feedback in many cases, providing evidence for configuration-dependent activation which establishes a baseline stiffness of the tongue. The proprioceptive feedback would then activate a reflex mechanism which modulates the stiffness.

Feedback is implemented in the λ -model without delay, and although conceptually it is meant to represent the stretch reflex, the λ -model can only be considered phenomenological in this sense i.e. feedback consistent with the λ -model could be due either to a brainstem reflex arc (where α motor neuron's activity is increased directly by sensory feedback) or the result of cortical feedback (sensory feedback causing either increase in α motor neuron activity directly or indirectly through modulation of γ motor neuron activity). As implemented it is also not specific to a particular type of receptor and such a reflex could result from afferent information originating in either muscle spindles or mechanoreceptors (Trulsson and Essick, 1997).

The only way to establish that the observed phenomenon is due to a brainstem reflex arc from positional data obtained experimentally is to demonstrate that it occurs too quickly for it to be a cortical process.

While the data do indeed suggest a fast reflex mechanism begins quickly after perturbation, lending evidence to the argument in favor of the existence of sensory receptors which are able to quickly react to deformations of the tongue. However, further study is necessary to determine definitely whether the rebound phenomenon observed is a cortical or brainstem reflex.

4.3 Discussion and Perspective

While the above study already has provided useful information in the interpretation of experimentally observed data, methodological limitations were apparent. Part of the benefit of the approach involving parallel modeling and experimental studies is that of the experimental data being able to constrain model parameters and overall development. There are certainly several directions that must be explored in order to assure the fidelity of the model.

One such item to consider modifying the muscle activation equation so that the component theoretically resulting from a fast reflex is delayed in a way consistent with at least a brainstem reflex (~10-20 ms).

Another item to consider is modifying the parameters of the current model in order to assure temporal accuracy in terms of fitting the experimental data, especially as it pertains to the propagation of force along the surface of the tongue. Before making such modifications though, it may be beneficial to attempt such simulations on 3d models (such as Rohan et al., 2014; Gérard et al., 2006; Buchaillard, Perrier, and Payan, 2009) to see whether the greater realism of these models, despite the inconveniences of their greater complexity, allows for more accurate results.

In any case, the use of the biomechanical model has already been useful in interpreting the experimental data and suggesting that a reflex mechanisms is made use of in order to stabilize the tongue following an external perturbation. It has also been useful in suggesting which muscles in particular to focus EMG recording efforts on.

In light of this preliminary work on parallel modeling and experimental perturbation studies, there is still work to be done in order to definitively address the question of whether a stretch reflex exists and is used in the human tongue during speech. Indeed this question is of interest to more than just speech, but to motor control in general, since knowing whether a reflex mechanism is involved in such a complex motor task as speech would help reveal the control strategies of complex motor tasks, specifically the question of the degree to which central control relies on peripheral control loops such a reflexes.

Specific questions still outstanding are thus whether the observed reflex phenomenon is due to a cortical reflex mechanism, and in any case, what its precise neural pathways and timing are. Additionally, despite the consensus in the literature as to the existence of muscle spindles in the human tongue, more information regarding their location and functional role would be useful to understanding the specific control strategies in use in speech.

Chapter 5

Conclusion

5.1 Summary

The interest of studying speech motor control in general was discussed. In particular speech has some interesting characteristics, such as the targets being of multiple forms, some of which are related only to non-linear transformations of the physical configuration of the articulators. The motivations for using a biomechanical model of speech production, as opposed to articulatory or other methods were then presented.

Following this motivation two studies were described, in which a biomechanical model of speech production was used in order to answer fundamental questions regarding not just speech motor control but motor control in general. Neither of these studies would have been possible without the use of a biomechanical model.

5.1.1 Uncontrolled manifolds in Speech Motor Control

First, the question was addressed of whether uncontrolled manifolds, originally proposed as a means by which to deal with motor equivalence and explain the variability patterns observed in movement, could be useful in the context of speech. After discussing an extension of the UCM concept to address the hierarchical nature of speech and focusing on the auditory UCM, a representation of linearized UCMs based on orthogonal projection matrices was developed. It was then demonstrated that they indeed appear to be useful in the context of speech due to 1) their ability to propose compensations to a command perturbation that had acceptable levels of error, despite the non-linearity of the mapping between motor commands and the acoustic features which they are intended to preserve and 2) the fact that these command compensations were consistently closer to the perturbed point than to unperturbed commands. In particular, by testing different granularities of locality in the acoustic output space, evidence was found for there being a slight advantage to each phoneme being represented by its own linearized UCM, an advantage which would be all the more pronounced in the case where a limited number of exemplar configurations were used for each phoneme.

While it was concluded that the UCM concept, in particular in its linearized representation as a orthogonal projection matrix onto the linearized UCM, is appropriate and even useful in the context of speech, several open questions regarding UCMs in speech remain.

One is that of whether these UCMs do indeed characterize the variability observed in practice during natural speech. The difficulty in addressing this question however is the “hidden” nature of not only the commands in speech, but even the articulatory parameters. One could imagine however future work in which commands, a proxy thereof, or simply some representation of the articulatory parameters are estimated from EMG signals or through some other observable data (such as tongue position using EMA, ultrasound, or even real-time MRI) which could then be fit with commands using the biomechanical model or an approximation of it. Such a study would allow for empirical evidence supporting the hypothesis that variability of movement of the tongue or the articulators involved in speech production is structured in a way that reflects the use of uncontrolled manifolds in the planning of the movement of the tongue.

Another open question is how such a representation could be integrated into a controller. As discussed in Chapter 2, UCMs are compatible with the approach of stochastic optimal feedback control but further work would need to be done in order to see if an equivalence relationship could be made between a neurally-compatible implementation of OFC and one of UCMs.

5.1.2 Perturbation responses in Speech Motor Control

Second, the mechanisms of postural control of the tongue were studied using the biomechanical model to help interpret empirical data acquired in a parallel study in human subjects. Through simulations and their analyses, evidence was presented in favor of the interpretation of the empirical data as containing a signature of a stretch-like reflex. In particular, by modulating proprioceptive feedback in the muscle force equation of the biomechanical model, this parameter’s role in the generation of a rebound effect after external perturbation was demonstrated and it was observed, specifically, that the empirical data were consistent with the case in which the proprioceptive feedback is non-zero. Through an analysis of the time course of various aspects of this simulation and a comparison of this data with the literature regarding the time course of cortical and brainstem response mechanisms in the tongue and reflex mechanisms of the lips, it was determined that the data are inconclusive and that they are compatible, within the temporal accuracy of the model, with either a slow brainstem reflex or a fast transcortical one.

Further work is thus needed to fully eliminate the possibility of the reflex being due to cortical mechanisms and establish it being due to a brainstem reflex, or vice versa. Moreover, as mentioned in Chapter 2 Section 2.1.6 the evidence for muscle spindles in the human, although existing and accepted leaves room for speculation as to whether there is a sufficient number which are properly located to be responsible for such a reflex in the human tongue. Further studies could thus include anatomic work to verify the location and prevalence of muscle spindles, the time course and potentially precise location of any brainstem response, in particular in relation to any cortical response. More complex studies could also be imagined which would seek to elucidate the precise neural pathway and whether the response is indeed based on the response of muscle spindles or, alternatively, mechanoreceptors.

5.1.3 Use of a biomechanical model

In both of the above described studies, the use of the biomechanical model was essential to the specific conclusions reached and represented the novelty of the approaches presented in this work. In the first case, uncontrolled manifolds involving the muscle commands could not have been studied as easily without such a model and the insight gained into the application of UCMs to movement planning would have been limited. In particular, had an articulatory model been used the question of whether the utility of UCMs to characterize a more complex mapping of realistic non-linearity were used would persist. Similarly, in the second study, the biomechanical model of the forces involved in the movement of the tongue allowed for the interpretation of empirical data which would otherwise have required further experimentation to evaluate, the difference with the current state being that much more precise hypotheses can be tested with more precise experiments. Both of these studies have thus, in addition to addressing the specific hypotheses discussed above in the corresponding sections, highlighted the utility of biomechanical models to help respond to complex questions of speech motor control.

Indeed this relationship of biomechanical modeling efforts with efforts to understand theoretical questions of motor control is mutually beneficial. Not only can the models serve as proxies to human data, facilitating the design of experimental methods by aiding in the generation and formulation of hypotheses and interpretation of experimentally acquired data, any experimental data acquired in this way can also be used to improve the fidelity and likewise the utility of the biomechanical models themselves.

5.2 Perspective

As discussed in Chapter 1, while the work presented here focuses on speech motor control, the reality is that the study of speech, due to its complexity and significant importance in language and cognition, has the potential to give insight into not just language and cognition, but complex adaptive systems capable of generating complex behavior in general. In particular, the questions of interest in this context are the degrees to which control is executed based on targets of different types (sensory or motor) and to which it is centralized or performed through specialized, even potentially hierarchical control loops.

Through both of these studies and the use of the biomechanical model some evidence for the role of the periphery in speech motor control was developed.

First, the periphery itself offers physical redundancy, where different commands and different configurations of muscle activation can accomplish the same tasks. Uncontrolled manifolds show how this is a natural consequence of the command to position mapping while the perturbation study showed that not only is the tongue itself elastic with a tendency to resist perturbations, but there seem to be active mechanisms which make use of this elasticity, strategically modifying it to stabilize the tongue after perturbation.

Second, the rich dynamics of the periphery, due to its biomechanical complexity, can be taken advantage of to minimize physical and computational effort in dealing with uncertainty, whether instability or perturbations, during speech motor tasks. Uncontrolled manifolds allow variability to be structured according to necessities of a task, avoiding excessive compensation effort when possible, in particular in the context of the motor commands. They, in light of the equilibrium point dynamics, also allow a simple representation of goals which can be used to generate commands for efficient trajectories. The perturbation response observed in the perturbation and modeling study highlighted the facts that the periphery has a stabilizing effect itself and that fast brainstem or even cortical feedback loops implement reflexes which resist perturbations and thus stabilize the maintenance of an articulated configurations.

While these phenomena were investigated in the context of speech, such mechanisms are general to all types of sensorimotor control, in particular a periphery which maximizes stability and minimizes control complexity and a division of labor across a hierarchy of feedback-sensitive control circuits.

5.3 List of contributions

The following summarizes the contributions made during the work presented:

1. In order to address speech using uncontrolled manifolds, first the treatment of hierarchical phenomena had to be address. This was done by proposing a hierarchy of different uncontrolled manifolds in the context of speech (see Section 3.1)
2. In order to exploit the information stored in a linearized uncontrolled manifold and to store it a potentially useful way, the use of projection matrix for representation of linearized UCMs was proposed (see Section 3.1.2).
3. The ability of the LUCM represented by projection matrix to suggest responses to command-perturbations was addressed (see study presented in Chapter 3, in particular Section 3.2).
4. The 2d biomechanical model presented in Section 2.4 was modified to simulate external perturbation (see Section 4.2.2).
5. The model was additionally modified to allow for control of the feedback gain around stable configurations (see Section 4.2.2).
6. Simulations with different feedback gain values were used to determine that a positive feedback loop is responsible for the rebound seen in simulations (see Section 4.2.2) and suggest that experimental data do then contain evidence of some sort of stretch reflex.
7. The time course of the perturbation response was analyzed to see whether convincing evidence for it being a cortical or brainstem reflex (Section 4.2.2) and this evidence was determined to be inconclusive.

5.4 List of experimental directions

From the work presented here, numerous directions of future work are suggested. Among these:

1. The auditory UCM was used in the form of a continuous 3 dimensional space. Further work could include the mathematical characterization of categorical uncontrolled manifolds, such as that which phonemes might present.
2. A control algorithm could be implemented in which UCMs are used to suggest commands giving rise to desired target states.
3. The compatibility of UCM-like representations with neurally-plausibly stochastic feedback control algorithms could be investigated.
4. Similar work to that presented here on uncontrolled manifolds could be done but focused on the optimal use of exemplar configurations.
5. The role of uncontrolled manifolds in the context of perturbations responses where there are constraints present could be studied.
6. A comprehensive study of the configurations covering articulatory space could be performed.
7. Electromagnetic articulometry, ultrasound, fMRI, x-ray or other imagery could be studied to look for task-related variability patterns in practice.
8. The perturbation studies presented here could be performed on potentially more accurate 3D models.
9. The realism of the perturbation simulation could be improved through modifying the biomechanical model's parameters to match observed data.
10. The biomechanical model, including the muscle activation equations, could be modified towards millisecond scale temporal accuracy, by including, for example, realistic nerve transmission delays.
11. Anatomical work could be performed to verify how a stretch reflex might be implemented in the tongue: location/distribution of muscle spindles, sensory innervation, including high-resolution tractography, etc.
12. Perturbation-related evoked potentials (EMG/EEG/MEG) could be studied in order to better understand the time course of stretch reflex generation.
13. Perturbations could be performed during articulator movement and not just stabilized vowel production.

Bibliography

- Abbs, James H and Vincent L Gracco (1983). "Sensorimotor actions in the control of multi-movement speech gestures". In: *Trends in Neurosciences* 6, pp. 391–395.
- Ackermann, Hermann, Klaus Mathiak, and Axel Riecker (2007). "The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data". In: *The Cerebellum* 6.3, pp. 202–213.
- Adams, Scott G, Gary Weismer, and Raymond D Kent (1993). "Speaking rate and speech movement velocity profiles". In: *Journal of Speech, Language, and Hearing Research* 36.1, pp. 41–54.
- Adatia, AK and EN Gehring (1971). "Proprioceptive innervation of the tongue." In: *Journal of anatomy* 110.Pt 2, p. 215.
- Anatomical Terminology, Federative Committee on (1998). *Terminologia anatomica*. Georg Thieme Verlag.
- Atal, Bishnu S, Jih Jie Chang, Max V Mathews, and John W Tukey (1978). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique". In: *The Journal of the Acoustical Society of America* 63.5, pp. 1535–1555.
- Badin, Pierre and Gunnar Fant (1984). "Notes on vocal tract computation". In: *STL QPSR* 2.3, pp. 53–108.
- Baker, RE (1954). "The tongue and dental function". In: *American Journal of Orthodontics* 40.12, pp. 927–939.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H Christiansen, William Croft, Nick C Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann (2009). "Language is a complex adaptive system: Position paper". In: *Language learning* 59.s1, pp. 1–26.
- Bernstein, Nikolai (1966). "The co-ordination and regulation of movements". In: *The co-ordination and regulation of movements*.
- Bhatnagar, Subhash Chandra (2002). *Neuroscience for the study of communicative disorders*. Lippincott Williams & Wilkins.
- Black, David P, Beth A Smith, Jianhua Wu, and Beverly D Ulrich (2007). "Uncontrolled manifold analysis of segmental angle variability during walking: preadolescents with and without Down syndrome". In: *Experimental brain research* 183.4, pp. 511–521.
- Blumstein, Sheila E and Kenneth N Stevens (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants". In: *The Journal of the Acoustical Society of America* 66.4, pp. 1001–1017.
- Borden, Gloria Jones, Katherine S Harris, and Lome Catena (1973). "Oral feedback. An electromyographic study of speech under nerve-block anesthesia". In: *Journal of Phonetics* 1, pp. 297–308.

- Bouchard, Kristofer E, Nima Mesgarani, Keith Johnson, and Edward F Chang (2013). "Functional organization of human sensorimotor cortex for speech articulation". In: *Nature* 495.7441, p. 327.
- Brand, Richard W, Donald E Isselhard, and Elaine Satin (2013). *Anatomy of orofacial structures: a comprehensive approach*. Elsevier Health Sciences.
- Bratzlavsky, M and H Vander Eecken (1974). "Afferent influences upon human genioglossal muscle". In: *Journal of neurology* 207.1, pp. 19–25.
- Browman, Catherine P and Louis Goldstein (1995). "Dynamics and articulatory phonology". In: *Mind as motion*, pp. 175–193.
- Buchaillard, Stéphanie, Pascal Perrier, and Yohan Payan (2009). "A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning". In: *The Journal of the Acoustical Society of America* 126.4, pp. 2033–2051.
- Cai, Shanqing, Deryk S Beal, Satrajit S Ghosh, Mark K Tiede, Frank H Guenther, and Joseph S Perkell (2012). "Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation". In: *PloS one* 7.7, e41830.
- Caillet, Jean-Loup (2017). "Sensorimotor control of tongue during speech production". Internship Report, Master 1 EEA (Energie électrique, électronique et automates). MA thesis. Université Grenoble Alpes.
- Calliope, ed. (1989). *La parole et son traitement automatique*. Masson Paris.
- Carlson, Rolf, Gunnar Fant, and Björn Granström (1974). "Two-formant models, pitch, and vowel perception". In: *Acta Acustica united with Acustica* 31.6, pp. 360–362.
- Chang, Edward (2016). *The Human Speech Cortex*. Keynote 2 at Interspeech 2016, San Francisco.
- Cheung, Vincent CK, Andrea d'Avella, Matthew C Tresch, and Emilio Bizzi (2005). "Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors". In: *Journal of Neuroscience* 25.27, pp. 6419–6434.
- Chow, CK and DH Jacobson (1971). "Studies of human locomotion via optimal programming". In: *Mathematical Biosciences* 10.3-4, pp. 239–306.
- Coker, Cecil H (1976). "A model of articulatory dynamics and control". In: *Proceedings of the IEEE* 64.4, pp. 452–460.
- Coker, Cecil H and Osamu Fujimura (1966). "Model for Specification of the Vocal-Tract Area Function". In: *The Journal of the Acoustical Society of America* 40.5, pp. 1271–1271.
- Cooker, Harry S, Charles R Larson, and ES Luschei (1980). "Evidence that the human jaw stretch reflex increases the resistance of the mandible to small displacements." In: *The Journal of Physiology* 308.1, pp. 61–78.
- Cooper, Sybil (1953). "Muscle spindles in the intrinsic muscles of the human tongue". In: *The Journal of physiology* 122.1, pp. 193–202.
- Crossman, Alan R and David Neary (2014). *Neuroanatomy E-Book: An Illustrated Colour Text*. Elsevier Health Sciences.
- Deacon, Terrence W (1998). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Company.

- Delattre, Pierre, Alvin M Liberman, Franklin S Cooper, and Louis J Gerstman (1952). "An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns". In: *Word* 8.3, pp. 195–210.
- Desmurget, Michel and Scott Grafton (2000). "Forward modeling allows feedback control for fast reaching movements". In: *Trends in cognitive sciences* 4.11, pp. 423–431.
- Domkin, Dmitry, Jozsef Laczko, Mats Djupsjöbacka, Slobodan Jaric, and Mark L Latash (2005). "Joint angle variability in 3D bimanual pointing: uncontrolled manifold analysis". In: *Experimental brain research* 163.1, pp. 44–57.
- Dreyfus, Stuart E (1990). "Artificial neural networks, back propagation, and the Kelley-Bryson gradient procedure". In: *J. Guidance, Control, Dynamics* 13.5, pp. 926–928.
- Feldman, Anatol G (1966). "Functional tuning of the nervous system with control of movement or maintenance of a steady posture. II. Controllable parameters of the muscle." In: *Biophysics* 11, pp. 565–578.
- Feldman, Anatol G and Mindy F Levin (1995). "The origin and use of positional frames of reference in motor control". In: *Behavioral and brain sciences* 18.4, pp. 723–744.
- Fitts, Paul M (1954). "The information capacity of the human motor system in controlling the amplitude of movement." In: *Journal of experimental psychology* 47.6, p. 381.
- Fitzgerald, MJ and SR Sachithanandan (1979). "The structure and source of lingual proprioceptors in the monkey." In: *Journal of anatomy* 128.Pt 3, p. 523.
- Flash, Tamar and Neville Hogan (1985). "The coordination of arm movements: an experimentally confirmed mathematical model". In: *Journal of neuroscience* 5.7, pp. 1688–1703.
- Fowler, Carol A (1996). "Listeners do hear sounds, not tongues". In: *The Journal of the Acoustical Society of America* 99.3, pp. 1730–1741.
- Friston, Karl (2011). "What is optimal about motor control?" In: *Neuron* 72.3, pp. 488–498.
- Ganesh, Gowrishankar, Masahiko Haruno, Mitsuo Kawato, and Etienne Burdet (2010). "Motor memory and local minimization of error and effort, not global optimization, determine motor behavior". In: *Journal of neurophysiology* 104.1, pp. 382–390.
- Gerard, Jean-Michel, Reiner Wilhelms-Tricarico, Pascal Perrier, and Yohan Payan (2003). "A 3D dynamical biomechanical tongue model to study speech motor control". In: *Recent Research Developments in Biomechanics* 1, pp. 49–64.
- Gérard, Jean-Michel, Reiner Wilhelms-Tricarico, Pascal Perrier, and Yohan Payan (2006). "A 3D dynamical biomechanical tongue model to study speech motor control". In: *arXiv preprint physics/0606148*.
- Goldinger, Stephen D, Paul A Luce, and David B Pisoni (1989). "Priming lexical neighbors of spoken words: Effects of competition and inhibition". In: *Journal of memory and language* 28.5, pp. 501–518.

- Golfinopoulos, Elisa, Jason A Tourville, Jason W Bohland, Satrajit S Ghosh, Alfonso Nieto-Castanon, and Frank H Guenther (2011). "fMRI investigation of unexpected somatosensory feedback perturbation during speech". In: *Neuroimage* 55.3, pp. 1324–1338.
- Gomi, Hiroaki, Masaaki Honda, Takayuki Ito, and Emi Z Murano (2002). "Compensatory articulation during bilabial fricative production by regulating muscle stiffness". In: *Journal of Phonetics* 30.3, pp. 261–279.
- Grimme, Britta, Susanne Fuchs, Pascal Perrier, and Gregor Schöner (2011). "Limb versus speech motor control: A conceptual review". In: *Motor control* 15.1, pp. 5–33.
- Guediche, Sara, Lori L Holt, Patryk Laurent, Sung-Joo Lim, and Julie A Fiez (2014). "Evidence for cerebellar contributions to adaptive plasticity in speech perception". In: *Cerebral Cortex* 25.7, pp. 1867–1877.
- Haggard, Patrick and Lieke de Boer (2014). "Oral somatosensory awareness". In: *Neuroscience & Biobehavioral Reviews* 47, pp. 469–484.
- Hansen, Eva, Britta Grimme, Hendrik Reimann, and Gregor Schöner (2015). "Carry-over coarticulation in joint angles". In: *Experimental brain research* 233.9, pp. 2555–2569.
- Harnad, Stevan (2003). "Categorical Perception". In: *Encyclopedia of Cognitive Science*. Vol. 67. 4. MacMillan: Nature Publishing Group.
- Harris, Christopher M and Daniel M Wolpert (1998). "Signal-dependent noise determines motor planning". In: *Nature* 394.6695, p. 780.
- Hashimoto, Kiyoshi and Shinsuke Suga (1986). "Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue". In: *Journal of the Acoustical Society of Japan (E)* 7.1, pp. 39–46.
- Heitz, Richard P (2014). "The speed-accuracy tradeoff: history, physiology, methodology, and behavior". In: *Frontiers in neuroscience* 8.
- Hickok, Gregory, Ursula Bellugi, and Edward S Klima (1996). "The neurobiology of sign language and its implications for the neural basis of language". In: *Nature* 381.6584, p. 699.
- Hockett, Charles F (1960). "The origin of speech". In: *Scientific American* 203.3, pp. 88–97.
- Hoff, Bruce Richard (1992). "A Computational Description of the Organization of Human Reaching and Prehension". PhD thesis.
- Holland, John H (1992). "Complex adaptive systems". In: *Daedalus*, pp. 17–30.
- Houde, John F and Michael I Jordan (1998). "Sensorimotor adaptation in speech production". In: *Science* 279.5354, pp. 1213–1216.
- Houde, John F and Srikantan S Nagarajan (2011). "Speech production as state feedback control". In: *Frontiers in human neuroscience* 5.
- Hughes, Olive M and James H Abbs (1976). "Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence". In: *Phonetica* 33.3, pp. 199–221.
- Ingram, John CL (2007). *Neurolinguistics: An introduction to spoken language processing and its disorders*. Cambridge University Press.

- Iskarous, Khalil, Louis Goldstein, Douglas H Whalen, Mark Tiede, and Philip Rubin (2003). "CASY: The Haskins configurable articulatory synthesizer". In: *International Congress of Phonetic Sciences, Barcelona, Spain*, pp. 185–188.
- Ito, Takayuki, Hiroaki Gomi, and Masaaki Honda (2004). "Dynamical simulation of speech cooperative articulation by muscle linkages". In: *Biological Cybernetics* 91.5, pp. 275–282.
- Ito, Takayuki, Toshitaka Kimura, and Hiroaki Gomi (2005). "The motor cortex is involved in reflexive compensatory adjustment of speech articulation". In: *Neuroreport* 16.16, pp. 1791–1794.
- Ito, Takayuki, Mark Tiede, and David J Ostry (2009). "Somatosensory function in speech perception". In: *Proceedings of the National Academy of Sciences* 106.4, pp. 1245–1248.
- Jordan, Michael I and David E Rumelhart (1992). "Forward models: Supervised learning with a distal teacher". In: *Cognitive science* 16.3, pp. 307–354.
- Kawato, Maeda, Y Maeda, Y Uno, and R Suzuki (1990). "Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion". In: *Biological cybernetics* 62.4, pp. 275–288.
- Kawato, Mitsuo (1999). "Internal models for motor control and trajectory planning". In: *Current opinion in neurobiology* 9.6, pp. 718–727.
- Kawato, Mitsuo, Kazunori Furukawa, and R Suzuki (1987). "A hierarchical neural-network model for control and learning of voluntary movement". In: *Biological cybernetics* 57.3, pp. 169–185.
- Kelso, JA SCOTT, Eric Vatikiotis-Bateson, Elliot L Saltzman, and Bruce Kay (1985). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling". In: *The Journal of the Acoustical Society of America* 77.1, pp. 266–280.
- Kemmerer, David (2014). *Cognitive neuroscience of language*. Psychology Press.
- Kempelen, W. von, H. Fügler, and J.G. Mansfeld (1791). *Wolfgang von Kempelen k.k. wirklichen Hofraths Mechanismus der menschlichen Sprache: nebst der Beschreibung seiner sprechenden Maschine*. J.V. Degen. URL: <https://books.google.ch/books?id=W75CAQAAMAAJ>.
- Kent, RD (1977). "Coarticulation in recent speech production". In: *Journal of Phonetics* 5.1, pp. 15–133.
- Kiritani, S, K Miyawaki, O Fujimura, and JE Miller (1975). "Computational model of the tongue". In: *The Journal of the Acoustical Society of America* 57.S1, S3–S3.
- Kubota, K, T Negishi, and T Masegi (1975). "Topological distribution of muscle spindles in the human tongue and its significance in proprioception." In: *The Bulletin of Tokyo Medical and Dental University* 22.3, pp. 235–242.
- Kuo, Arthur D (1995). "An optimal control model for analyzing human postural balance". In: *IEEE transactions on biomedical engineering* 42.1, pp. 87–101.
- Laboissière, Rafael, David J Ostry, and Anatol G Feldman (1996). "The control of multi-muscle systems: human jaw and hyoid movements". In: *Biological cybernetics* 74.4, pp. 373–384.

- Lammert, Adam C, Christine H Shadle, Shrikanth S Narayanan, and Thomas F Quatieri (2016). "Investigation of Speed-Accuracy Tradeoffs in Speech Production Using Real-Time Magnetic Resonance Imaging." In: *INTERSPEECH*, pp. 460–464.
- Latash, Mark L (2008). *Neurophysiological basis of movement*. Human Kinetics.
- Latash, Mark L, John P Scholz, and Gregor Schöner (2002). "Motor control strategies revealed in the structure of motor variability". In: *Exercise and sport sciences reviews* 30.1, pp. 26–31.
- Lindblom, Björn (1990). "Explaining phonetic variation: A sketch of the H&H theory". In: *Speech production and speech modelling*. Springer, pp. 403–439.
- Liu, Dan and Emanuel Todorov (2007). "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control". In: *Journal of Neuroscience* 27.35, pp. 9354–9368.
- Lloyd, John E, Ian Stavness, and Sidney Fels (2012). "ARTISYNTH: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation". In: *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*. Springer, pp. 355–394.
- Loeb, GE, WS Levine, and Jiping He (1990). "Understanding sensorimotor feedback through optimal control". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 55. Cold Spring Harbor Laboratory Press, pp. 791–803.
- Lowe, AA, SC Gurza, and BJ Sessle (1977). "Regulation of genioglossus and masseter muscle activity in man". In: *Archives of oral biology* 22.10-11, pp. 579–584.
- Maeda, Shinji (1990). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model". In: *Speech production and speech modelling*. Springer, pp. 131–149.
- Marr, David (1982). *Vision: A Computational Investigation Into*. WH Freeman.
- Martinet, André (1957). "Arbitraire linguistique et double articulation". In: *Cahiers Ferdinand de Saussure* 15, pp. 105–116.
- Mattar, Andrew AG and David J Ostry (2010). "Generalization of dynamics learning across changes in movement amplitude". In: *Journal of neurophysiology* 104.1, pp. 426–438.
- Matthies, Melanie, Pascal Perrier, Joseph S Perkell, and Majid Zandipour (2001). "Variation in anticipatory coarticulation with changes in clarity and rate". In: *Journal of Speech, Language, and Hearing Research* 44.2, pp. 340–353.
- McDonald, Stuart and Stephen Creanor (2016). "The tongue". In: *Essential Clinical Oral Biology*, p. 120.
- McGurk, Harry and John MacDonald (1976). "Hearing lips and seeing voices". In: *Nature* 264.5588, pp. 746–748.
- Ménard, Lucie (2002). "Production et perception des voyelles au cours de la croissance du conduit vocal: variabilité, invariance et normalisation". PhD thesis. Grenoble 3.
- Meyer, David E, Richard A Abrams, Sylvan Kornblum, Charles E Wright, and JE Keith Smith (1988). "Optimality in human motor performance: Ideal control of rapid aimed movements." In: *Psychological review* 95.3, p. 340.
- Miall, R Christopher and Daniel M Wolpert (1996). "Forward models for physiological motor control". In: *Neural networks* 9.8, pp. 1265–1279.

- Munhall, Kevin G, David J Ostry, and Avraham Parush (1985). "Characteristics of velocity profiles of speech movements." In: *Journal of Experimental Psychology: Human Perception and Performance* 11.4, p. 457.
- Natale, Michael (1975). "Convergence of mean vocal intensity in dyadic communication as a function of social desirability." In: *Journal of Personality and Social Psychology* 32.5, p. 790.
- Neilson, Peter D, Gavin Andrews, Barry E Guitar, and Peter T Quinn (1979). "Tonic stretch reflexes in lip, tongue and jaw muscles". In: *Brain research* 178.2, pp. 311–327.
- Nelson, Winston L (1983). "Physical principles for economies of skilled movements". In: *Biological cybernetics* 46.2, pp. 135–147.
- Ostry, David J, James D Cooke, and Kevin G Munhall (1987). "Velocity curves of human arm and speech movements". In: *Experimental Brain Research* 68.1, pp. 37–46.
- Ostry, David J and Kevin G Munhall (1985). "Control of rate and duration of speech movements". In: *The Journal of the Acoustical Society of America* 77.2, pp. 640–648.
- Otten, E (1987). "A myocybernetic model of the jaw system of the rat". In: *Journal of neuroscience methods* 21.2, pp. 287–302.
- O'Reilly, Randall C (1996). "The Leabra model of neural interactions and learning in the neocortex". PhD thesis.
- Papi, Enrica, Philip J Rowe, and Valerie M Pomeroy (2015). "Analysis of gait within the uncontrolled manifold hypothesis: Stabilisation of the centre of mass during gait". In: *Journal of biomechanics* 48.2, pp. 324–331.
- Pardo, Jennifer S (2013). "Measuring phonetic convergence in speech production". In: *Frontiers in psychology* 4.
- Parrell, Benjamin, Zarinah Agnew, Srikantan Nagarajan, John Houde, and Richard B Ivry (2017). "Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration". In: *Journal of Neuroscience*, pp. 3363–16.
- Payan, Yohan and Pascal Perrier (1997). "Synthesis of VV sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis". In: *Speech communication* 22.2-3, pp. 185–205.
- Pearce, Sophie L, Timothy S Miles, Philip D Thompson, and Michael A Nordstrom (2003). "Is the long-latency stretch reflex in human masseter transcortical?" In: *Experimental brain research* 150.4, pp. 465–472.
- Perkell, Joseph S (1974). "A physiologically-oriented model of tongue activity in speech production." PhD thesis. Massachusetts Institute of Technology.
- Perkell, Joseph S, Melanie L Matthies, Mario A Svirsky, and Michael I Jordan (1993). "Trading relations between tongue-body raising and lip rounding in production of the vowel /u: A pilot 'motor equivalence' study". In: *The Journal of the Acoustical Society of America* 93.5, pp. 2948–2961.
- (1995). "Goal-based speech motor control: A theoretical framework and some preliminary data". In: *Journal of Phonetics* 23.1, pp. 23–35.
- Perrier, Pascal (2012). "Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control". In:

- Speech Planning and Dynamics*. Ed. by D. Pape S. Fuchs M. Weirich and P. Perrier. Frankfurt, Germany: Peter Lang Publishers, pp. 191–238.
- Perrier, Pascal, Louis-Jean Boë, and Rudolph Sock (1992). “Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients”. In: *Journal of Speech, Language, and Hearing Research* 35.1, pp. 53–67.
- Perrier, Pascal and Susanne Fuchs (2008). “Speed–curvature relations in speech production challenge the 1/3 power law”. In: *Journal of neurophysiology* 100.3, pp. 1171–1183.
- Perrier, Pascal, Liang Ma, and Yohan Payan (2006). “Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue”. In: *arXiv preprint physics/0610170*.
- Perrier, Pascal, David J Ostry, and Rafael Laboissière (1996). “The equilibrium point hypothesis and its application to speech motor control”. In: *Journal of Speech, Language, and Hearing Research* 39.2, pp. 365–378.
- Perrier, Pascal and Ralf Winkler (2015). “Biomechanics of the orofacial motor system: Influence of speaker-specific characteristics on speech production”. In: *Individual differences in speech production and perception*. Ed. by C. Petrone S. Fuchs D. Pape and P. Perrier. Frankfurt, Germany: Peter Lang Publishers, pp. 223–254.
- Perrier, Pascal, Yohan Payan, Majid Zandipour, and Joseph Perkell (2003). “Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study”. In: *The Journal of the Acoustical Society of America* 114.3, pp. 1582–1599.
- Pilon, Jean-François, Sophie J De Serres, and Anatol G Feldman (2007). “Threshold position control of arm movement with anticipatory increase in grip force”. In: *Experimental brain research* 181.1, pp. 49–67.
- Ramanarayanan, Vikram, Benjamin Parrell, Louis Goldstein, Srikantan Nagarajan, and John Houde (2016). “A New Model of Speech Motor Control Based on Task Dynamics and State Feedback.” In: *INTERSPEECH*, pp. 3564–3568.
- Reisman, Darcy S, John P Scholz, and Gregor Schöner (2002). “Coordination underlying the control of whole body momentum during sit-to-stand”. In: *Gait & posture* 15.1, pp. 45–55.
- Robert-Ribes, Jordi (1995). “Modèles d’intégration audiovisuelle de signaux linguistiques: de la perception humaine à la reconnaissance automatique des voyelles”. PhD thesis.
- Rohan, P-Y, Claudio Lobos, M Ali Nazari, Pascal Perrier, and Yohan Payan (2014). “Finite element modelling of nearly incompressible materials and volumetric locking: a case study”. In: *Computer methods in biomechanics and biomedical engineering* 17.sup1, pp. 192–193.
- Saltzman, Elliot L and Kevin G Munhall (1989). “A dynamical approach to gestural patterning in speech production”. In: *Ecological psychology* 1.4, pp. 333–382.
- Sanders, Ira, Liancai Mu, Asif Amirali, Hungxi Su, and Stanislaw Sobotka (2013). “The human tongue slows down to speak: muscle fibers of the human tongue”. In: *The Anatomical Record* 296.10, pp. 1615–1627.

- Sanguineti, Vittorio, Rafael Laboissière, and David J Ostry (1998). "A dynamic biomechanical model for neural control of speech production". In: *The Journal of the Acoustical Society of America* 103.3, pp. 1615–1627.
- Sanguineti, Vittorio, Rafael Laboissiere, and Yohan Payan (1997). "A control model of human tongue movements in speech". In: *Biological cybernetics* 77.1, pp. 11–22.
- Sasaki, Clarence T and Edward M Weaver (1997). "Physiology of the larynx". In: *The American journal of medicine* 103.5, 9S–18S.
- Savariaux, Christophe, Pascal Perrier, and Jean Pierre Orliaguet (1995). "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production". In: *The Journal of the Acoustical Society of America* 98.5, pp. 2428–2442.
- Schmidt, Richard A (1988). *Motor learning and control*. Human Kinetics.
- Scholz, John P and Gregor Schöner (1999). "The uncontrolled manifold concept: identifying control variables for a functional task". In: *Experimental brain research* 126.3, pp. 289–306.
- Scholz, JP, G Schöner, and ML Latash (2000). "Identifying the control structure of multijoint coordination during pistol shooting." In: *Experimental brain research* 135.3, p. 382.
- Schöner, G, V Martin, H Reimann, J Scholz, R Sock, S Fuchs, and Y Laprie (2008). "Motor equivalence and the uncontrolled manifold". In: *Proceedings of the International Seminar on Speech Production (ISSP 2008) in Strassbourg*, pp. 23–28.
- Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry (1997). "The dispersion-focalization theory of vowel systems". In: *Journal of phonetics* 25.3, pp. 255–286.
- Schwartz, Jean-Luc, Anahita Basirat, Lucie Ménard, and Marc Sato (2012). "The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception". In: *Journal of Neurolinguistics* 25.5, pp. 336–354.
- Scott, Stephen H (2004). "Optimal feedback control and the neural basis of volitional motor control". In: *Nature reviews. Neuroscience* 5.7, p. 532.
- Shadmehr, Reza, Maurice A Smith, and John W Krakauer (2010). "Error correction, sensory prediction, and adaptation in motor control". In: *Annual review of neuroscience* 33, pp. 89–108.
- Shiller, Douglas M, Rafael Laboissière, and David J Ostry (2002). "Relationship between jaw stiffness and kinematic variability in speech". In: *Journal of Neurophysiology* 88.5, pp. 2329–2340.
- Shiller, Douglas M, David J Ostry, and Paul L Gribble (1999). "Effects of gravitational load on jaw movements in speech". In: *Journal of Neuroscience* 19.20, pp. 9073–9080.
- Smith, Anne, Christopher A Moore, Christine M Weber, David H McFarland, and Jerald B Moon (1985). "Reflex responses of the human jaw-closing system depend on the locus of intraoral mechanical stimulation". In: *Experimental neurology* 90.3, pp. 489–509.
- Steels, Luc (2000). "Language as a complex adaptive system". In: *Parallel Problem Solving from Nature PPSN VI*. Springer, pp. 17–26.
- Stoll, Taja, Philip Hoole, and Jonathan Harrington (2016). "Influence of palatalization on tongue-tip velocity in trills". In:

- Stone, Maureen and Eric Vatikiotis-Bateson (1995). "Trade-offs in tongue, jaw, and palate contributions to speech production". In: *Journal of Phonetics* 23.1, pp. 81–100.
- Swerts, Marc (1997). "Prosodic features at discourse boundaries of different strength". In: *The Journal of the Acoustical Society of America* 101.1, pp. 514–521.
- Todorov, Emanuel (2004). "Optimality principles in sensorimotor control". In: *Nature neuroscience* 7.9, p. 907.
- Todorov, Emanuel and Michael I Jordan (2002). "Optimal feedback control as a theory of motor coordination". In: *Nature neuroscience* 5.11, p. 1226.
- Tourville, Jason A and Frank H Guenther (2011). "The DIVA model: A neural theory of speech acquisition and production". In: *Language and cognitive processes* 26.7, pp. 952–981.
- Tremblay, Stéphanie, Guillaume Houle, and David J Ostry (2008). "Specificity of speech motor learning". In: *Journal of Neuroscience* 28.10, pp. 2426–2434.
- Trulsson, Mats and Gregory K Essick (1997). "Low-threshold mechanoreceptive afferents in the human lingual nerve". In: *Journal of neurophysiology* 77.2, pp. 737–748.
- Tuller, Betty, Katherine S Harris, and Bob Gross (1981). "Electromyographic study of the jaw muscles during speech". In: *Journal of Phonetics* 9, pp. 175–188.
- Umeda, Noriko (1977). "Consonant duration in American English". In: *The Journal of the Acoustical Society of America* 61.3, pp. 846–858.
- Uno, Yoji, Mitsuo Kawato, and Rika Suzuki (1989). "Formation and control of optimal trajectory in human multijoint arm movement". In: *Biological cybernetics* 61.2, pp. 89–101.
- Vaxelaire, Béatrice (1993). "Etude comparee des effets des variations de debit-lent, rapide-surles parametres articulatoires, a partir de la cineradiographie (sujets francais)". PhD thesis. Strasbourg 2.
- Vaxelaire, Béatrice, Camille Fauth, Fabrice Hirsch, Fayssal Bouarourou, and Rudolph Sock (2010). "Anticipatory and sustained coupling of jaw and tongue gestures in VCV sequences: an x-ray study". In: *Summerschool Cognitive and Physical Models of Speech Production, Speech Perception and Production-Perception Interaction*.
- Vaxes, AUBREY J (1963). "Delayed auditory feedback". In: *Psychological bulletin* 60.3.
- Vogt, Florian, John E. Lloyd, Stephanie Buchaillard, Pascal Perrier, Matthieu Chabanas, Yohan Payan, and Sidney S. Fels (2006). "Efficient 3D Finite Element Modeling of a Muscle-Activated Tongue". In: *Proceedings of ISBMS 06*. Vol. 4072. Lecture Notes in Computer Science. Springer, pp. 19–28.
- Walker, LB and MD Rajagopal (1959). "Neuromuscular spindles in the human tongue". In: *Anatomical Record*. Vol. 133. 2, pp. 438–438.
- Weber, Christine M and Anne Smith (1987). "Reflex responses in human jaw, lip, and tongue muscles elicited by mechanical stimulation". In: *Journal of speech and hearing research* 30.1, pp. 70–79.
- Whalen, DH (1990). "Coarticulation is largely planned 7/3". In: *Journal of Phonetics* 18, pp. 3–35.
- Wilhelms-Tricarico, Reiner (1995). "Physiological modeling of speech production: Methods for modeling soft-tissue articulators". In: *The Journal of the Acoustical Society of America* 97.5, pp. 3085–3098.

- Winkler, Ralf, Liang Ma, and Pascal Perrier (2010). "A model of optimal speech production planning integrating dynamical constraints to achieve appropriate articulatory timing." In: *Cognitive and Physical Models of Speech Production, Speech Perception and Production-Perception Interaction-Part III Planning and Dynamics (CPMSP2 2010)*, pp. 44–48.
- Wolpert, Daniel M and Mitsuo Kawato (1998). "Multiple paired forward and inverse models for motor control". In: *Neural networks* 11.7, pp. 1317–1329.
- Wolpert, Daniel M, R Chris Miall, and Mitsuo Kawato (1998). "Internal models in the cerebellum". In: *Trends in cognitive sciences* 2.9, pp. 338–347.
- Zajac, Felix E, Michael R Zomlefer, and William S Levine (1981). "Hindlimb muscular activity, kinetics and kinematics of cats jumping to their maximum achievable heights". In: *Journal of Experimental Biology* 91.1, pp. 73–86.
- Zandipour, Majid (2006). "Modeling Investigation of Vowel to Vowel Movement Planning in Acoustic and Muscle Spaces". PhD thesis. Boston University.
- Zandipour, Majid, JS Perkell, Frank Guenther, Mark Tiede, Kiyoshi Honda, and Emi Murano (2006). "Speaking with a bite-block: Data and modeling". In: *Proceedings of the 7th International Seminar on Speech Production*, pp. 361–368.
- Ziółko, Bartosz and Mariusz Ziółko (2009). "Time durations of phonemes in Polish language for speech and speaker recognition". In: *Language and Technology Conference*. Springer, pp. 105–114.
- Zomlefer, MR, WS Levine, and FE Zajac (1974). "The use of optimal control in the study of a normal physiological movement". In: *Decision and Control including the 13th Symposium on Adaptive Processes, 1974 IEEE Conference on*. Vol. 13. IEEE, pp. 195–196.