Global loudness processing of time-varying sounds. The case of rising and falling intensity stimuli

Emmanuel Ponsot

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Ecole Doctorale ED3C

Global loudness processing of time-varying sounds

The case of rising and falling intensity stimuli

Emmanuel Ponsot

Soutenue le 14 décembre 2015

devant le jury composé de :

M. Laurent Demany, Rapporteur
M. Wolfgang Ellermeier, Rapporteur
M. Romain Brette, Examinateur
M. Daniel Pressnitzer, Examinateur
M. Massimo Grassi, Examinateur
M. Patrick Susini, Directeur de thèse
Mme. Sabine Meunier, Co-Directrice de thèse
M. Daniel Oberfeld, Membre invité

Ircam,
Equipe Perception et Design Sonores,
1, place Igor Stravinsky, 75004 Paris, France

Abstract

The *global loudness* evaluation of a time-varying sound is based on the transformation of our momentary loudness sensations into a single scalar value that reflects our overall impression a posteriori. The aim of this thesis is to deepen the understanding of this dynamical process in the case of specific stimuli that rise or fall in level over a few seconds. Starting from the observation that rising tones are perceived louder than falling tones while both contain the same energy over the same duration, this thesis is structured around four experimental chapters that scrutinize the origins of this perceptual asymmetry from various perspectives. Chapter 1 explores the extent to which this asymmetry relies on contextual or methodological factors. Various measurement procedures and contexts of presentation yielding comparable results, evidence is provided that a steady and robust mechanism is at the basis of this perceptual asymmetry. In Chapter 2, the influence of the sound spectral structure is examined and it is shown that asymmetries are significantly diminished when broadband noises rather than tonal stimuli are used. Chapter 3 examines the effects of the temporal profile characteristics of the stimuli (slope, duration, dynamics) on their *global loudness* judgments. Results indicate that listeners' judgments rely on a temporal integration of the loudest portion but that other mechanisms are involved; asymmetries whose magnitude depends on the profile characteristics and the level of the sounds in a complex and nonlinear manner are found. In Chapter 4, the *global loudness* evaluation process is examined locally in a loudness discrimination task. The respective contributions of the different temporal portions of rising and falling-intensity stimuli as well as their interactions are derived from reverse-correlation analyses. First-order analyses

show that listeners exclusively focus on the loudest portions of these sounds and that similar temporal weighting patterns apply to rising and falling profiles; second-order analyses indicate the presence of nonlinear components in the integration process that have never been reported so far. Current loudness models cannot directly account for the effects highlighted throughout these studies, which supports the idea that the mechanism(s) under study might take place in high auditory or decisional stages not considered in these models. From the use of very basic stimuli, this work brings to light key temporal mechanisms in global loudness processing that might be shared with other sensory attributes. This research offers new perspectives to investigate the coding and the processing of time-varying stimuli by sensory systems from identified perceptual asymmetries.

Résumé

L'évaluation de la *sonie globale* d'un son non-stationnaire repose sur la transformation de nos sensations de sonie momentanées en une valeur scalaire unique reflétant notre impression globale a posteriori. L'objectif de cette thèse est d'approfondir la compréhension de ce processus dynamique dans le cas particulier de stimuli présentant des profils d'intensité croissants ou décroissants de quelques secondes. Partant de l'observation que les sons croissants sont perçus plus forts que les sons décroissants alors qu'ils contiennent la même énergie pendant la même durée, cette thèse s'articule autour de quatre chapitres expérimentaux qui abordent la question des origines de cette asymétrie perceptive sous différentes perspectives. Le Chapitre 1 évalue dans quelle mesure cette asymétrie repose sur des facteurs contextuels ou méthodologiques. Différentes méthodes de mesures et différents contextes de présentation conduisant à des résultats similaires, il en ressort qu'un mécanisme stable et robuste est très probablement impliqué dans cette asymétrie perceptive. Dans le Chapitre 2, l'influence de la structure spectrale du son est examinée, faisant apparaître des asymétries significativement réduites lorsque des bruits large bande plutôt que des sons tonaux sont utilisés. Le Chapitre 3 s'intéresse aux effets des paramètres décrivant le profil temporel des stimuli (pente, durée, dynamique) sur leurs jugements de *sonie globale*. Les résultats indiquent que les jugements des auditeurs se fondent sur une intégration temporelle de la région la plus intense mais que d'autres mécanismes sont impliqués; des asymétries dépendant de façon complexe et non-linéaire des caractéristiques du profil temporel et du niveau des stimuli sont observées. Le Chapitre 4 étudie localement le processus d'évaluation de la *sonie globale* à partir d'une tâche de discrimination en sonie. Les contributions respectives des différentes portions temporelles des sons croissants et des sons décroissants sont obtenues à partir d'analyses "reverse-correlation". Au premier ordre, les analyses font apparaître le fait que les auditeurs se basent exclusivement sur les parties les plus intenses des sons et que des pondérations temporelles similaires (symétriques) sont appliquées aux profils croissant et décroissant; les analyses au second ordre mettent en évidence la présence de composantes non-linéaires dans le processus d'intégration qui n'avaient encore jamais été rapportées jusqu'à présent. Les modèles de sonie actuels ne permettent pas directement d'expliquer ces résultats, ce qui soutient l'idée selon laquelle les mécanismes étudiés pourraient provenir de traitements auditifs ou décisionnels "haut-niveau" n'étant pas encore pris en considération dans ces modèles. A partir de stimuli élémentaires, cette étude met en lumière des mécanismes temporels fondamentaux dans le traitement de la sonie globale, possiblement partagés par d'autres attributs sensoriels. De nouvelles perspectives de recherche pour l'étude du codage et du traitement de stimuli non-stationnaires par les systèmes sensoriels à partir d'asymétries perceptives avérées sont ainsi offertes.

Remerciements

Mes remerciements les plus chaleureux vont tout d'abord à mes deux directeurs de thèse, Patrick Susini et Sabine Meunier. Merci à tous les deux de m'avoir ouvert au monde de la psychoacoustique et de la psychophysique dès mon premier stage à vos côtés. Merci pour votre confiance, votre écoute et pour vos encouragements réguliers ainsi que pour avoir su me laisser emprunter librement les diverses pistes de recherche aperçues pendant ces trois ans. Un grand merci pour ces années de travail si agréables avec vous dont je conserverai une emprunte certaine.

Je tiens également à remercier les membres de mon jury. Merci à Laurent Demany et Wolfgang Ellermeier d'avoir accepté d'être rapporteur de cette thèse, merci aussi à Romain Brette, Massimo Grassi, Daniel Oberfeld et Daniel Pressnitzer d'avoir pris le temps d'évaluer ce travail.

Merci aux chercheurs extérieurs au projet LoudNat avec qui j'ai eu la chance d'interagir, et qui ont fait de ces années de thèse une période particulièrement enrichissante et stimulante sur le plan scientifique. Merci à Daniel Oberfeld pour m'avoir introduit à l'aventure passionnante de la psychophysique moléculaire, pour son soutien scientifique irremplaçable et son dynamisme constant. Merci à Brice Bathellier ainsi qu'à son équipe pour nos échanges réguliers, ses idées, et pour cette collaboration fructueuse qui est née, ainsi que pour m'avoir permis de reconsidérer mes questions de recherche différemment, en me mettant dans la peau (des neurones) d'une souris. Je remercie également Peter Neri pour son aide précieuse et son expertise scientifique lors des derniers mois de thèse concernant les analyses non linéaires et la modélisation des données moléculaires. Merci à Åke Hellström de m'avoir permis de venir à bout des TOEs du Chapitre 1. Enfin, merci à Guillaume Saint Pierre pour ses explications et sa collaboration sur certains aspects des analyses statistiques menées au cours de cette thèse.

Je souhaiterais bien sûr remercier les différentes personnes de l'Ircam et du LMA que j'ai côtoyées pendant ces années. Merci à toute l'équipe PDS (Perception et Design Sonores), Olivier Houix, Nicolas Misdariis, Mondher Ayari, Jean-Julien Aucouturier, Guillaume Lemaitre, Eric Boyer et Maxime Carron ainsi qu'aux personnes arrivées plus récemment dans le cadre du projet CREAM. Merci enfin aux différentes personnes impliquées dans le projet LoudNat pour nos interactions scientifiques et amicales, en particulier Jacques Chatron, Gaston Hilkhuysen, Sophie Savel et Guy Rabeau. Je souhaite remercier particulièrement mon compagnon de route de la sonie et ami Michaël Vannier, pour nos échanges réguliers sur nos travaux ainsi que pour avoir permis que ces années non-stationnaires se déroulent dans une bonne ambiance stationnaire.

Je voudrais également exprimer de sincères remerciements aux personnes de Radio France que j'ai eu la chance de rencontrer dès mes premières armes dans le monde du son. Un grand merci à Martine Chicault de m'avoir introduit dans cet univers fantastique qu'est la maison de la radio et de m'avoir toujours encouragé et soutenu dans mes projets. Merci à Edwige Roncière de m'avoir accueilli si sympathiquement dans son bureau et de m'avoir permis de m'intéresser aux problématiques pratiques de production/diffusion radiophonique liées à la sonie. Merci aussi aux experts des oreilles, Hervé Déjardin et mon ami Lasse Munk pour nos riches discussions sur l'avenir du binaural.

Je remercie Jean-Julien Aucouturier pour ses précieuses relectures pédagogiques et constructives de plusieurs parties de cette thèse.

Je remercie Hugues Vinet, directeur scientifique de l'Ircam, pour son accueil ainsi que pour ses remarques et son enthousiasme sur ce travail.

Merci à mes amis de longue date, Kuj, Cot, Vidounz, Mich, Van, à mes copains ingénieurs Mariane et Féfé ainsi qu'aux thésards Seb et Edgar, tous indispensables.

Merci à mes parents pour leur amour, leurs encouragements et leur soutien depuis toujours.

Enfin, je souhaiterais dire un très grand merci à Indiana pour avoir partagé toutes les étapes de ce travail et pour m'avoir guidé dans cette aventure intellectuelle au quotidien, et bien plus. Merci infiniment d'avoir été là et de m'avoir toujours épaulé et conseillé.

Contents

List of Figures

List of Tables

General Introduction

When a light is too intense, we can close our eyes. The hearing system does not comprise such a mechanism that would allow to turn on and off the incoming acoustical information, which highlights the survival aspect of sound perception and more particularly of intensity perception, *loudness*. Traditionally, psychoacoustics mostly studied loudness from a *static* perspective, through the use of stationary signals. However, the last two decades have seen a growing number of psychophysical, physiological and neuroscientific studies investigating different *dynamical* characteristics of loudness (e.g., Irino and Patterson 1996; Zhang and Zeng 1997; Patterson and Irino 1998; Neuhoff 1998; Stecker and Hafter 2000; Lu et al. 2001; Neuert et al. 2001; Phillips et al. 2002; Teghtsoonian et al. 2005; Lentz and Shen 2011). The use of time-varying stimuli has put the spotlight on novel sensory and cognitive mechanisms (e.g., Neuhoff 1998; Maier and Ghazanfar 2007; Bach et al. 2008) that had been overlooked by previous research focused on static aspects of loudness. Investigating loudness with non-stationary signals is crucial to understand its coding and processing in a natural context, given that most of the sounds in our everyday listening environment present physical features that constantly change through time (Theunissen and Elie 2014; Santoro et al. 2014).

One important aspect of dynamic loudness processing concerns how *global loudness*, namely the overall evaluation of loudness of stimuli that vary in loudness over time, is processed from the signal entering the auditory system. In particular, very little is known about how momentary loudness sensations are transformed into a scalar value that reflects the overall impression a posteriori. This thesis breaks through this question by considering the case of specific diotic stimuli (i.e. identical signals to both ears) with linear increasing or decreasing levels of a few seconds. We find that such stimuli induce perceptual "asymmetries" and recruit auditory mechanisms that had not yet been identified by the static loudness literature.

0.1 What is loudness?

0.1.1 Definition

When a sound wave reaches our ears, it is processed by our auditory system, which in turns gives rise to an auditory sensation. From this sensation, one can for instance tell "how loud" the sound is. Loudness is the perceived intensity of a sound, and like its pitch, it constitutes one of the main dimensions of auditory perception. *Loudness* is thus a psychological – not a physical – quantity, often considered as the "*primary* psychological correlate of physical [sound] level" (Florentine, 2011). Many attempts were made to define loudness in unequivocal ways. In the recent book *Loudness*, a few of these definitions are reported (Florentine, 2011). Loudness is said to be "the attribute of a sound that changes most readily when sound intensity is varied" (Scharf, 1978). The American National Standards Institute also proposed a standardized definition of loudness as being the "attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud" (ANSI, 1973). These definitions are straightforward but need in fact further clarification, because sounds are generally composed of several frequency components that vary over time. Different aspects of loudness can be introduced; they are illustrated in Figure 1.

0.1.2 Spectral aspects of loudness

Consider a stationary[1] complex tone, composed of two different largely spaced frequencies clearly distinguishable – so that there is a priori no interaction between the two components feeding different critical bands –, such as a two-tone complex made of one component at 100 Hz and another component at 3 kHz. How will one judge the loudness of this sound? Will one consider the loudness of the lowest tone in pitch (the 100-Hz component), the loudness of the highest tone in pitch (the 3-kHz component), the loudness of the loudest

[1]By definition, a *stationary* sound is a sound whose overall physical characteristics (i.e., spectro-temporal properties and intensity) do not vary through time. Pure tones and noises are examples of stationary sounds.

component or the overall loudness of the sound made by the percept of the two merged components? In such contexts, the concepts of *specific loudness* and *overall loudness*[2] were proposed to better specify what one means. *Specific loudness*[3] refers to the loudness of a specific component of a complex tone, while *overall loudness* refers to the loudness of the complex tone as a whole (cf. Fig. 1 (a)). Whether listeners primarily attend to *specific* or *overall loudness* when listening to such sounds, and how different *specific loudness* components are merged into a single *overall loudness* component are issues that have been raised since long[4], but have been rarely specifically investigated. Attentional strategies and criterion decision aspects have to be considered when measuring the loudness of such complex tones. This issue has also been addressed in the "auditory scene analysis" literature (e.g., Grimault et al. 2007).

0.1.3 Temporal aspects of loudness

In time as well, loudness cannot be limited to a *unitary* concept: when one starts considering non-stationary sounds, whose physical characteristics – e.g., sound pressure level, frequency – vary in time. As an example, consider a simple 1-kHz tone whose physical level is varying during several seconds. How will one evaluate the loudness of this sound? Providing that the level fluctuations are not too fast, our sensation of loudness is also varying through

[2]Sometimes also referred to as *global loudness*.

[3]Here, *specific loudness* refers to the perceptual evaluation of a specific frequency component. Note that *specific loudness* is also the term employed elsewhere in loudness models for the intermediate variable referring to the loudness inside each critical band of the ear, but in that case, it cannot directly be interpreted as a perceptual quantity.

[4]When studying loudness summation, Zwicker et al. (1957) already observed different participants' behaviors with respect to loudness evaluation of two-tones complexes. In some conditions where the two-tones complexes were composed of frequencies widely separated, in which "loudness summation" was not recruited *per se*, he observed that while some listeners were indeed integrating the two components in their judgments (what they were asked to), other listeners only focused on one single component. Scharf (1969) also reported similar effects when investigating dichotic summation of loudness: when two tones of widely separated frequencies were presented to each ear, respectively, they could be heard as "two more or less distinct images" and loudness could then be splitted into two components.

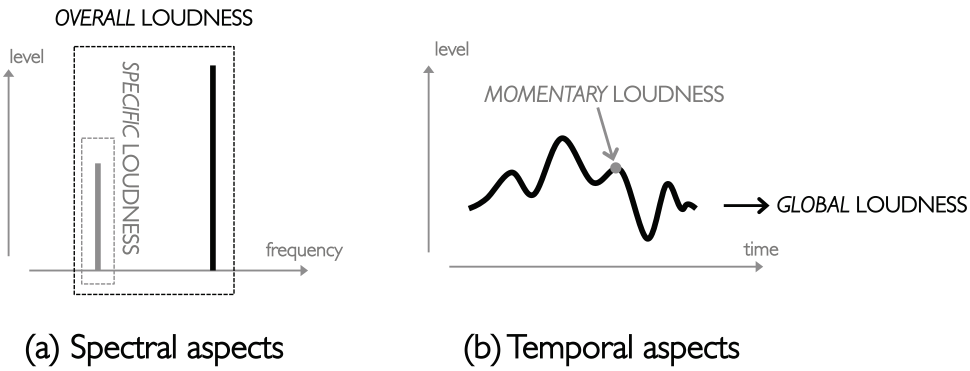

Figure 1: Different loudness components can be defined when considering its spectral (a) and temporal (b) aspects. (a) For example, a stationary two-tones complex can be judged by considering its *specific* or *overall* loudness. (b) A pure tone with a time-varying level profile can be judged in terms of *momentary* or *global* loudness.

time and appears to be *dynamic*. It is thus necessary to distinguish our momentary sensations from what can be experienced and evaluated at the end of the sound, as a retrospective summary. The concepts of *momentary loudness* and *global loudness* were introduced to this aim (Glasberg and Moore, 2002). *Momentary loudness* refers to the instantaneous sensation of loudness in time available at any instant. The whole temporal loudness pattern induced by a given sound can thus be defined as a vector made of these "momentary" loudness snapshots over time. *Global loudness*[5] corresponds to the "overall loudness" perceived after the end of the sound, when the whole momentary loudness pattern is merged into one scalar value (cf. Fig. 1 (b)).

[5]Finally, this definition of *global loudness*, which accounts for the temporal aspects of loudness, could be compared with what was introduced to define the *overall pitch* of frequency-modulated tones (Gockel et al., 2001) and frequency glissandos (d'Alessandro and Castellengo, 1994) – although these studies did not address the link between momentary pitch and overall pitch.

0.1.4 Loudness of natural sounds in natural environments

Both the spectral and temporal aspects of loudness detailed above take on their full meaning when considering everyday environmental or musical sounds, which are not stationary and whose level and spectral characteristics are constantly changing through time. In natural listening situations, as opposed to laboratory conditions (e.g., when the sounds are diotically presented through headphones), it is quite obvious but easily overlooked, that (i) human listeners have two ears receiving different signals and that (ii) natural auditory scenes are often composed of several, potentially moving, auditory objects. First, sound waves coming from either static or moving sound sources which are not frontally presented induce different loudness over both ears, however the loudness of a sound is always perceived as a whole. Our auditory system does what is called “binaural loudness summation” (for a recent review, see Sivonen and Ellermeier 2011). Second, when dealing with soundscapes composed of several auditory objects, loudness can either refer to the loudness of one specific object or to the loudness of the whole soundscape. While one clearly is able to separately judge the loudness of different sources presented in the same environment, it is not known yet how the auditory system can “zoom” in and out over time and integrate the loudness of specific objects into the whole “picture” (see the “cocktail party problem”; McDermott 2009). In short, just as stationary complex tones can give rise to different loudness depending on whether *specific* or *global* loudness is considered (see above), similar issues exist in complex ecological listening situations composed of multiple sources, another subject of study for the “auditory scene analysis” domain (Bregman, 1994).

0.1.5 The dimension of loudness investigated in this thesis

In summary, loudness is a complex multi-faceted dimension (we could even say multi-dimensional attribute) of auditory perception, whose various aspects all have to be considered. The present thesis specifically addresses temporal aspects of loudness and will focus on *global loudness* perception and evaluation of non-stationary *diotic* stimuli that vary in level but whose spectral characteristics remain unchanged. In that context, *global loudness* will always refer to the overall judgment of loudness after the sound has ended, indicating how these *time-varying levels*, identical over both ears, are integrated into one scalar value. This will allow various aspects of auditory temporal processing which take place at different timescales to be considered. By exclusively studying temporal variations in level, this work will bring to light processes specific to the intensity dimension. These processes, later considered in their interaction with time-varying spectral and spatial characteristics will pave the way toward the understanding of global loudness processing of more complex natural scenes, often composed of several, spatially distributed and potentially moving sound sources.

0.2 Measuring loudness of time-varying sounds

0.2.1 Generalities

Loudness is a perceptual quantity, which can be accessed by means of psychophysical[6] experiments. Measuring loudness consists in determining how the perceived intensity of sounds (as judged by humans) relates to their physical properties. Psychophysics provides well-established measurement methods able to control for many potential judgment "biases", which allow to establish these relationships as accurately as possible (see Baird 2014). According to its various definitions, loudness as a psychological quantity can be ordered on some perceptual "scale", ranging from soft to loud: one can compare two sounds in loudness (relative judgment) but one can also evaluate their respective magnitude independently (absolute judgment) (Florentine, 2011).

By convention, loudness is most often expressed on a linear scale directly proportional to the perceived intensity, the *sone* scale (Stevens, 1956). The *sone* scale is a ratio scale[7], which has been defined with a unity reference taken as the sensation of loudness produced by a 40-dB SPL, 1-kHz tone frontally presented in free field[8]. Loudness can also be expressed on a *phon* scale, which has been introduced to provide an easier way to make correspondence with the measurable physical level (as compared to the *sone* scale where loudness values are compared in terms of ratios) and which also accounts for loudness dependence on tone frequency[9]. By definition, a sound

[6]Psychophysics aims to understand the relationships existing between the physical world and the psychological (i.e. perceptual) world of humans, accessible through his senses. In practice, this means characterizing the whole transformation chain of the physical input entering a sensory system (e.g., the auditory system, the visual system) and the perceptual outcome that can be experienced and judged by humans.

[7]This scale has been built using direct evaluation methods such as magnitude estimation methods, where subjects evaluate the perceived loudness of sounds using numbers.

[8]Thus, a sound of 2 sones is perceived twice as loud as the 1-kHz tone at 40 dB SPL, and a sound of 4 sones is perceived twice as loud as the 2-sones sound (and consequently four times louder than the 1-kHz tone at 40 dB SPL).

[9]The phon scale has been established using comparison methods (this type of methods

with a value of x phons indicates that it is perceived as loud as a 1-kHz tone presented at x dB SPL in free field in front of the listener.

The loudness of stationary sounds has been thoroughly investigated since the early 30's (e.g., Fletcher and Munson 1933), and its relationships with the physical attributes of the stimuli, such as intensity, frequency, spectral bandwidth and duration, have been widely described and characterized in the literature. These relationships will not be presented in this thesis (for a recent review, see Jesteadt and Leibold 2011). In what follows, we rather focus on the methods and issues directly regarding loudness measurement of time-varying stimuli.

0.2.2 Methodological considerations

The purpose of loudness measurement for time-varying sounds is the same as for stationary sounds, that is to say, to characterize the relationships between the loudness perception (and evaluation) and the physical attributes of the sounds. Both the *sone* and the *phon* scales introduced for stationary sounds keep their full meaning with time-varying sounds. As discussed in the introduction, when dealing with time-varying sounds, different aspects of loudness can however be considered, and thus measured: *momentary loudness* and *global loudness* have to be distinguished.

Various studies have attempted to measure *momentary loudness* sensations of time-varying sounds (e.g., Canévet and Scharf 1990; Teghtsoonian et al. 2000; Susini et al. 2002; Susini et al. 2007; Olsen et al. 2014; for a review see Kuwano and Namba 2011). Momentary loudness of time-varying sounds cannot be accessed in the same way as the loudness induced by a stationary stimulus, therefore new "real-time" methods were introduced to have access to this "continuous" sensation. In these studies, subjects were

will be considered later in this section), where the loudness of a tone at a certain frequency is matched in loudness against a reference 1-kHz tone (Fletcher and Munson, 1933). This scale is well known today through iso-loudness contours (e.g., Suzuki and Takeshima 2004), which are used in many applications (for example, A-weighting corresponds to the averaged iso-loudness curve obtained at mid levels).

asked to report their momentary sensation of loudness in real-time during the listening by judging it on a scale with different categories (the method of continuous judgment by category, e.g., Kuwano and Namba 1985), by applying a force proportional to instantaneous sensation on a specifically-designed proprioceptive device (the cross-modal matching method, e.g., Susini et al. 2002) or by evaluating it at different moments using numbers[10] (the magnitude estimation method, which was used for example in the investigation of loudness "decruitment" induced by tones continuously decreasing in level; Canévet and Scharf 1990; Schlauch 1992; Olsen et al. 2014). The measurement of momentary loudness typically has two purposes. First, when dealing with complex stimuli, it can be more accurate to measure a loudness profile like it is actually perceived by listeners rather than using a loudness model to evaluate it, because the model may always provide wrong predictions[11]. Second, momentary loudness can be used to complement a measure of global loudness, as it will be discussed hereafter.

Contrary to momentary loudness, measuring *global loudness* consists in asking participants to evaluate the loudness of a sound taken as a whole, which only requires one unique judgment, at the end. Therefore, the measurement methods which can be used are essentially the same as those employed for the measurement of stationary sounds, for which many psychophysical "standards" are available (for a review of all these methods and their applications to stationary loudness measurement, see Marks and Florentine 2011). In what follows, we will also consider different studies that did not measure *global loudness* but investigated another overall quantity called *loudness change*. *Loudness change* refers to the overall perceived amount of loudness variation, a construct that primarily makes sense for time-varying sounds varying monotically in loudness, i.e. increasing or decreasing in level. While this percept relates more to the perceived *dynamics* of a sound than

[10] Note that this procedure had also been used to explore loudness adaptation with long stationary tones, by asking participants to continuously report their loudness sensation at different times using numbers.

[11] For instance, current loudness models cannot account for long-term loudness adaptation effects (e.g., Hellman et al. 1997) or "decruitment" effects induced by long decreasing-intensity stimuli (e.g., Teghtsoonian et al. 2000).

to its *loudness* per se, the methods and issues related to its measurement are similar to those used for *global loudness* measurement, so they must be considered.

Among the standard psychophysical methods, one can distinguish evaluation methods and comparison methods, which have both been employed to measure the *global loudness* or the *loudness change* of time-varying sounds. In evaluation methods (e.g., the method of magnitude estimation), listeners evaluate the magnitude of a sound presented alone with a number or a visual-analog / category scale (e.g., Neuhoff 1998; Susini et al. 2007; Susini et al. 2010), or report the loudness ratio either to another sound (e.g., Pastore and Flint 2011). In comparison methods, listeners have to match two sounds in loudness, either by adjusting (e.g., with a slider or a potentiometer) the loudness of one comparison sound to match that of one reference, fixed sound (e.g., Ries et al. 2008), or have to indicate which sound in a pair is louder (i.e., when comparisons are made in terms of proportions, e.g. Olsen et al. 2010, but also when adaptive loudness-matching procedures are employed, e.g. Grassi and Darwin 2006).

0.2.3 Issues related to global loudness measurement

The main challenge of *global loudness* measurement is to understand the mechanisms and the processes that underpin its computation from *momentary loudness* experiences. There are two ways of tackling this problem. The first approach consists in measuring both momentary loudness and global loudness and then look for a transformation from the former (a vector) to the latter (a scalar); i.e. a transformation that connects two perceptual variables. Several studies in the literature have adopted this approach, employing different types of stimuli, either long sequences of natural sounds such as traffic noise (Hellbrück 2000; Gottschling 1999; Kuwano and Namba 1985) or simple 1-kHz stimuli varying in level (Nakajima et al. 1983; Susini et al. 2007). The second approach consists in measuring global loudness only. With this approach, the relationship under study concerns the overall transformation of the time-varying physical properties of the sound into the overall judgment (i.e., a transformation from the physical to the perceptual domain). This second approach has been primarily employed with very basic stimuli such as tones, vowels or noises increasing or decreasing in level over a few second (e.g., Neuhoff 1998; Susini et al. 2010), to ensure the spectral characteristics of the sounds remain unchanged. At first sight, the first approach has the advantage to attempt to link two perceptual variables, a momentary loudness vector and a global loudness scalar value. The second approach is more complex given that it looks for a larger transformation, which is actually made of two transformations: (1) the level profile mapped into a momentary loudness profile and (2) the momentary loudness profile converted into global loudness. However, some studies reported that when listeners are asked to judge both their continuous perception and their global impression, their global judgment can be tainted by their continuous report. Indeed, listeners produce different estimations of global loudness when they are asked to report their continuous loudness sensation too[12] (Susini et al., 2002). In the context of annoyance evaluation (which is in some way close to loudness evaluation), Ariely and Zauberman (2000) also demonstrated that measuring perceived continuous momentary intensities tend to moderate the relationships and to downsize the heuristic processes between the pattern and the overall annoyance evaluation made at the end of the sound, presumably because subjects then tend to self-segment the experience. As a result, when investigating the link between continuous perception and spontaneous overall evaluation, measurements should be conducted with different subjects.

Regardless of these two approaches, *global loudness* of time-varying sounds can be evaluated either with sounds presented alone or with sounds presented in pairs and matched in loudness. The former type of measurement is most commonly employed when the sounds last more than a few seconds, because in that case, comparing pairs become very time-consuming and even infeasible. However, it should always be kept in mind that different ex-

[12]In analogy to the measurement problem in quantum mechanics, continuous measurement of loudness could be said to *corrupt* the integration processes that are taking place during the listening, and thus modifying the resulting overall judgments.

perimental conditions often yield different results, because the underlying sensory integration and evaluation processes related to the task are not the same. Even though methodological and contextual issues have been widely addressed concerning static loudness (for a review on "contextual" effects in static loudness measurement, see Arieh and Marks 2011), much less is known with regard to dynamic loudness. These effects are nevertheless critical because they can provide important information about underlying integration processes. Ariely and Carmon (2000) already addressed some of the measurement issues with respect to the evaluation of time-varying sensory information. In that context, the debates, regarding *loudness change* perception, where different studies reached different conclusions because the measurement methods were not the same (Neuhoff 1998; Canévet et al. 1999; see the discussion in Olsen 2014) can also be taken as an instructive example.

0.2.4 Methods considered in this thesis

This thesis will exclusively be concerned with the measurement of global loudness when it is not combined with momentary loudness evaluations. The most standard and established methods to *quantitatively* measure loudness will be employed. The timescales considered ranging from one to several seconds, both magnitude estimation methods (which provide loudness estimates that can be expressed in sones) and paired comparison methods with adaptive procedures (which provide points of subjective equality and thus differences in level (dB) between two sounds at equal-loudness) will be used.

0.3 Predicting (time-varying) loudness with current models

Different loudness models were developed based on the findings from loudness research over more than 50 years. Throughout this section, we will only consider the models available for non-impulsive sounds that are presented diotically (i.e. the same signal reaching the two ears) to normal-hearing listeners. Other models exist for binaural dichotic signals (i.e. different signals between the two ears) (Moore and Glasberg 2007; Sivonen and Ellermeier 2008; Glasberg and Moore 2010), for impulsive signals (e.g., Boullet 2005) or when considering hearing-impaired listeners (e.g., Chalupper and Fastl 2002), but they are not under the scope of this thesis. Moreover, we will only focus on auditory models whose outputs reflect loudness *per se*, sometimes also referred to as “psychological models” (Launer, 1995). There also exist auditory “physiological” models, which can help go further into the functioning of the different elements composing the first (peripheral) stages of the auditory system (e.g., Patterson et al. 1995; Zilany et al. 2009; see also Launer 1995). However, it is still unclear how the outputs of these models relate to the psychological quantity of loudness. They will thus not be considered either.

0.3.1 The simplest loudness model: Stevens’ power law

Loudness models are all derived from the most fundamental relationship known in psychoacoustics, which is the one relating loudness to sound level. The simplest model that has been proposed to account for this relationship is known as Steven’s power law (Stevens, 1955):

$\mathrm{S} = \mathrm{kp}^{\alpha}$,

where S is the loudness (in sones), p is the sound pressure (in Pa) and k is a constant. At 1-kHz, the exponent α is found to be, on average[13] around 0.6[14]. The value of this exponent varies with stimulus frequency, bandwidth and duration. Stevens' power law indicates that the relationship between the physical level and loudness is nonlinear and that, at 1-kHz, loudness is doubled for each increase of 10 dB. This law is a model, which means that it approximates the relationships that can be observed experimentally, and in particular, holds true mostly for sound levels in the range 30-90 dB SPL (see Jesteadt and Leibold 2011). Indeed, loudness growths more rapidly at low (< 30 dB SPL) and high levels (> 90 dB SPL) as compared to mid levels. Several studies have proposed to account for this effect by applying a correction factor on the exponent (see Buus et al. 1998) or by using more complex polynomial functions such as the "InEx" (Inflected Exponential) function (Epstein and Florentine, 2006), which has steeper slopes at low and high levels than at mid levels.

0.3.2 Loudness models for time-varying sounds

0.3.2.1 Early stages taken from stationary loudness models

The two models most frequently used to predict the loudness of time-varying sounds are Chalupper and Fastl's model (2002) known as "DLM"[15] (Dynamic Loudness Model), and Glasberg and Moore's model (2002) known as "TVL" (Time-Varying Loudness). These models are based on earlier loudness models developed for stationary sounds. DLM is based on Zwicker's

[13]This 0.6 exponent value is the one adopted by the psychophysical community. However, it should be noted that (1) several studies reported smaller exponents for 1-kHz tones, closer to 0.4 (Canévet et al. 2003; Teghtsoonian et al. 2005; Susini et al. 2010; Marks 1988; Marks 1992; Marks 1993); and (2) this value significantly varies across listeners such that neither the 0.6 nor the 0.4 value holds true at an individual level (e.g. Canévet et al. 1986; Gescheider and Hughson 1991).

[14]If sound intensity instead of sound pressure is used, the mean value of the exponent would be 0.3.

[15]DLM is an extended version of the time-varying loudness model developed by Zwicker (1977) but it also includes hearing-impaired listening.

model (1958), which was adopted as the International standard method for the calculation of loudness of stationary sounds ISO-532B (1975) and TVL comes from Moore et al.'s model (1997), which led to the American standard ANSI-S3.4 (2007).

The first five computational stages of these models are directly imported from earlier stationary versions. In these stages, the input signal is (1) filtered to match the frequency response of the outer and middle ear, (2) band-pass filtered (to model auditory filters), (3) transformed into excitation levels using the power level of the signal feeding each critical band[16] or auditory filter[17], (4) converted into specific loudness[18] (using power functions with corrections factors) and, finally (5) integrated to provide a unique loudness value. Although the first stages of these two time-varying loudness models are conceptually comparable, some notable differences exist which have been depicted in Rennies et al. (2010) (for further details about these models, see also Marozeau 2011). In the present work, we will rather focus on the subsequent stages of both models (i.e. after the five computational stages just mentioned) which make them appropriate for predicting loudness of time-varying sounds.

0.3.2.2 Late temporal integration stages

In both models, the output of the fifth stage is called "instantaneous loudness" (see Figure 2) and is assumed to be a temporary variable "which is not available for conscious perception" (Glasberg and Moore, 2002). This is a temporal vector which contains loudness values that have been computed over very brief time periods on a sliding-window basis (every 1 or 2 ms), as if they constituted a series of brief stationary sounds. The dynamic characteristics of the models come at the next stage, where this "instantaneous loudness" vector is "smoothed". The smoothing structures of both models are schematized in Figure 2. In the DLM, "instantaneous loudness"

[16] In DLM, the excitation levels are calculated by integrating the power inside each critical band and by considering frequency masking.

[17] In TVL, the excitation levels are based on the patters given at the output of the different auditory filters.

[18] Not to be confounded with specific loudness as defined in the introduction.

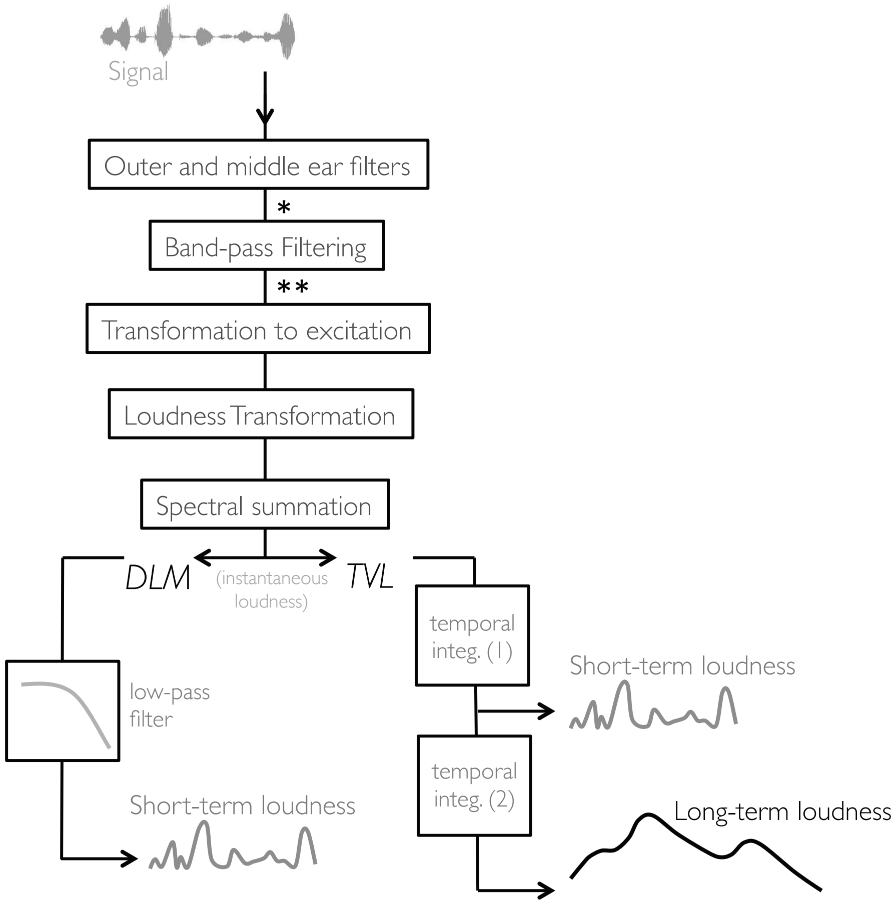

Figure 2: Schematic representation of the respective structures of the two main current loudness models for time-varying sounds, which are TVL (Glasberg and Moore, 2002) and DLM (Chalupper and Fastl, 2002). The time-varying signal is analyzed with time-steps of 1-ms in TVL (before the second stage, as indicated by one asterisk) or 4-ms in DLM (at the output of the second stage, as indicated by two asterisks). While the former five stages do analogous operations, the latter stage(s) compute the short-term and the long-term loudness in noticeable different ways. In DLM (left side), the instantaneous loudness is low-pass filtered to compute the short-term loudness. In TVL (right side), one or two stages of temporal integration (as described in the text) are applied to instantaneous loudness time-series to compute short-term loudness and long-term loudness patterns, respectively.

is simply low-pass filtered using a first-order filter having a cut-off frequency equals to 8 Hz. The output of this filtering is called “short-term loudness” and is assumed to reflect the “loudness perceived at any instant”, i.e. what we also called “momentary loudness”. In the TVL, two temporal integration stages are used, which mimic automatic gain-control circuits having different time constants for increasing and decreasing periods. The first stage derives short-term loudness from instantaneous loudness, as follows:

$$\mathrm{STL}(\mathrm{t}_{i+1}) = (1-\alpha)IL(t_{i+1}) + \alpha STL(t_i),$$

where STL indicates “short-term loudness” and IL instantaneous loudness. $\alpha = e^{\frac{-dt}{\tau_\alpha}}$ and $\tau_\alpha = 22$ ms when $IL(t_{i+1}) > STL(t_i)$ (i.e. loudness is growing) or $\tau_\alpha = 50$ ms when $IL(t_{i+1}) \leq STL(t_i)$ (i.e. loudness is decreasing or constant). The term $dt = t_{i+1} - t_i$ is the time step (equals to 1 ms).

The second integration stage then computes long-term loudness from short-term loudness:

$$\mathrm{LTL}(\mathrm{t}_{i+1}) = (1-\beta)STL(t_{i+1}) + \beta LTL(t_i),$$

where LTL holds for "long-term loudness". $\beta = e^{\frac{-dt}{\tau_\beta}}$ and, $\tau_\beta = 100$ ms when $STL(t_{i+1}) > LTL(t_i)$ or $\tau_\beta = 2000$ ms when $STL(t_{i+1}) \leq LTL(t_i)$.

As in DLM, the “short-term loudness” output of TVL is assumed to reflect the perceived loudness at any instant. However, Glasberg and Moore suggest using it only for short signals and rather to use the “long-term loudness” prediction for slowly time-varying stimuli. In particular, the “long-term loudness” output variable was introduced to account for the results obtained for amplitude-modulated stimuli. It is important to note that the first values of the time constants in TVL (used to derive short-term loudness) as well as the frequency cut-off in the DLM have been chosen to account for temporal integration (e.g. Poulsen 1981) and forward masking (e.g., Zwicker 1984) results obtained in the literature, mostly collected using stationary sounds of various durations or from signal detection thresholds inferred from temporal masking experiments, respectively.

0.3.2.3 Computing global loudness

A final operation is necessary to predict *global loudness* from the output of these models. Indeed, a unique scalar variable is needed whereas the models provide continuous “momentary loudness” patterns: short-term loudness (STL) and long-term loudness (LTL) are time series. For this final stage, in line with Zwicker’s recommendation, Chalupper and Fastl (2002) suggest to use the maximum value of STL given by the DLM to account for global loudness; another possibility is to use the 95th percentile of the short-term loudness, often referred to as N_5 (Zwicker et al., 1999). Glasberg and Moore (2002) too propose to use the maximum value of STL to predict the global loudness of brief time-varying sounds. For longer amplitude modulated sounds, they suggest using the average value of the LTL pattern [19] For other types of “long” stimuli varying slowly in loudness, the maximum value of LTL is rather employed in practice (e.g. Ries et al. 2008).

This thesis will specifically consider the maxima of short-term and long-term loudness patterns provided by the models to account for global loudness.

[19] Indeed, as noted in Moore (2014), an amplitude-modulated sound can be heard as fluctuating, but at the same time listeners can report an overall impression of loudness. It is proposed that the fluctuations can be accounted for by the short-term loudness patterns while the global loudness impression can be accounted for by the average of the long-term loudness pattern.

0.4 Context of this thesis

This section aims at presenting different issues that have emerged in the literature regarding the loudness or the global loudness of different types of time-varying sounds at various timescales.

0.4.1 Current issues related to the loudness of time-varying sounds

0.4.1.1 Global loudness of long time-varying sounds

Some studies investigated *global loudness* of time-varying sounds using long and complex sequences lasting from a few seconds to a few minutes (cf. Fig. 3 (c)) often made of uncontrolled stimuli such as recordings of natural sounds (e.g., Kuwano and Namba 1985; Hellbrück 2000; Gottschling 1999). There are only a few exceptions where long synthesized sequences and thus "controlled" long time-varying sound sequences were employed (e.g., Susini et al. 2007) (see also Chapter 3). All these studies provided converging evidence that memory processes such as "recency effects" (e.g., Susini et al. 2002, Susini et al. 2007) seem to be involved in global loudness judgments of long time-varying sounds, but we are still far from a full comprehension of the parameters that govern these processes. The advantage of using such long and complex sequences is that the findings of these experiments can be directly interpreted in the context of everyday sound perception and evaluation. However, the main disadvantage is that such complex stimuli might recruit various mechanisms, so they are probably not the best candidates for psychophysical experiments that aim to disentangle, identify and scrutinize the underlying mechanisms.

0.4.1.2 Sounds with rising vs. falling amplitude envelopes: Perceptual "asymmetries" at different timescales

The last two or three decades have seen a growing interest in the psychophysical community for what we will call in this thesis "*perceptual asymmetries*". This term refers to the fact that auditory stimuli with the same overall

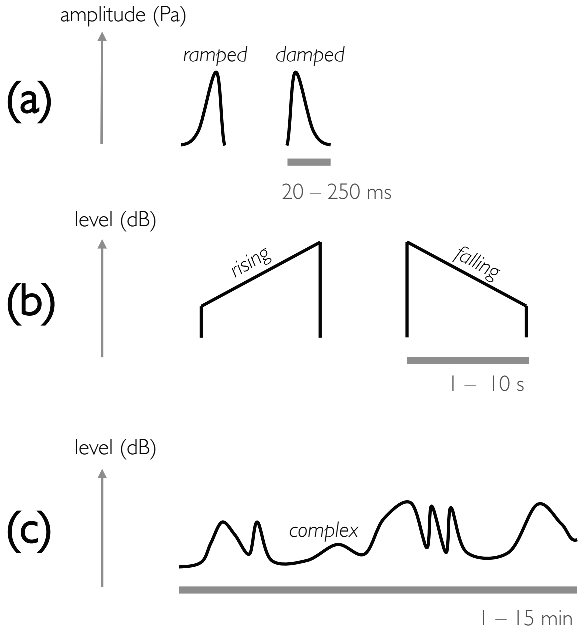

Figure 3: Examples of sounds with time-varying level profiles at different timescales.

physical characteristics (i.e., level and spectral characteristics) but different asymmetric amplitude envelopes are often perceived differently.

This phenomenon has been considered in the case of short-duration stimuli (ranging from 20 ms to 250 ms) with asymmetrical amplitude envelopes. In particular, *ramped* or *damped* stimuli made of exponential increase or decrease in amplitude (cf. Fig. 3 (a)) were found to be with different loudness[20] (e.g., Stecker and Hafter 2000), timbre (e.g., Irino and Patterson 1996) and duration (e.g., Schlauch et al. 2001) (see Chapter 1 for a more detailed review). Different mechanisms were proposed to account for these effects, but some pieces of the puzzle are still missing to understand the whole process (see Ries et al. 2008). It is interesting to note that these short asymmetric stimuli induce perceptual "asymmetries" not only concerning loudness, but

[20]Note that with such stimuli varying very rapidly in loudness, the continuous loudness profiles cannot be accessed and reported consciously by the listeners such that there is no need to distinguish *global loudness* from a simple loudness definition of the perceived intensity.

also other auditory dimensions. This suggests that an intrinsic component of the auditory system might processes sounds having asymmetric amplitude envelopes differently, thus causing consistent perceptual asymmetries between sounds whose amplitude is either rising or falling. Recent studies indicate that these effects would not only involve low-level peripheral mechanisms but also more central stages of auditory processing (e.g., Lentz and Shen 2011). It should be noted however, that although *damped* stimuli can be compared to impulsive sounds produced in reverberant environments or to musical notes of struck string instruments such as those produced by a piano (see Schutz and Vaisberg 2014), *ramped* stimuli cannot be found in daily life; they merely correspond to *damped* sounds played backward. Therefore, we argue that, even such stimuli are particularly good candidates to recruit and thus to investigate nonlinear auditory mechanisms, the outcome of these psychophysical investigations cannot be transposed directly to what occurs with natural sounds.

Other studies have investigated perceptual "asymmetries" between longer stimuli (generally, from 1000 ms to 10000 ms) varying monotically and linearly in level, either rising of falling (cf. Figure 3 (b)). Both rising and falling stimuli that vary slowly in level constitute a first approximation of the level profiles that everyone can encounter in his everyday auditory environment, such as the sound of a car passing-by (a first portion where the level increases - approaching period - and a second portion where the level decreases - receding period). These studies were mostly concerned with either (i) the perception and the evaluation of the perceived dynamics of these sounds, called *loudness change*, which was measured directly (e.g., Neuhoff 1998; Teghtsoonian et al. 2005; Olsen et al. 2010) or indirectly (e.g., Canévet et al. 2003) or (ii) with the continuous perceived loudness (i.e., "momentary loudness") (e.g., Susini et al. 2007; Olsen et al. 2014). Despite the fact that one needs to understand the temporal integration mechanisms occurring at this time-scale in order to understand the perception of natural sounds, these types of basic sounds that simply increase or decrease in level have almost never been studied in terms of *global loudness* (to our knowledge, only one study addressed this aspect; Susini et al. 2007). Such stimuli that vary slowly in loudness over a few seconds would arguably not recruit the nonlinear processing characteristics of the auditory system as much as short asymmetrical ramped and damped stimuli. Nevertheless, a number of neuroscience studies have shown that rising ramp sounds increasing in level over 1 or 2 seconds recruit brain neural circuits associated to moving, approaching sounds (e.g., Bach et al. 2008) and emotional (fearful) stimuli (e.g., Tajadura-Jiménez et al. 2010) (see Chapter 1 and 2 for more details), which were proposed to be at the basis of the asymmetries observed at the behavioral level (e.g., Neuhoff 1998; Neuhoff 2001). Thus, investigating how *global loudness* is perceived and judged for these stimuli constitutes an interesting (and direct) way to explore how our auditory system and our brain process natural sounds.

0.4.1.3 Asymmetries and current loudness models

Current loudness models can be used as *tools* to investigate loudness asymmetries, because they can help identify whether the observed results can be predicted by already identified mechanisms. In that context, it has been shown that the loudness asymmetries found between short exponential asymmetric stimuli (Stecker and Hafter, 2000) might be predicted by current loudness models, because the temporal integration (or the low-pass filtering) stage(s) of the models react differently to attack and decay periods and thus provide short-term loudness patterns that do not reach the same maxima (see Rennies et al. 2010). However, as noted by Rennies et al. (2010), one of the main challenge for future loudness models is probably to accurately predict the global loudness of long (i.e. a few seconds) time-varying sounds, such as those employed in Susini et al. (2007). Indeed, in that case, the maximum of the short-term loudness cannot predict the observed "recency" effects.

0.4.1.4 A novel perspective: Temporal weighting of loudness

Another more recent line of research on loudness is concerned with temporal weighting underlying *global loudness* judgments of time-varying sounds. As compared to studies focused on perceptual asymmetries, which employed stimuli with asymmetrical amplitude envelopes, studies on temporal weighting of loudness primarily investigated the case of global loudness evaluations of sounds varying over *flat* profiles (e.g., Pedersen and Ellermeier 2008; Rennies and Verhey 2009). They used a particular but not so recent methodology to investigate this aspect, the so-called "molecular psychophysics" (Green 1964; Berg 1989). This methodology is based on the use of stimuli, about 1-s long, that are time-segmented into 100-ms portions and which levels are randomly varied on a trial-by-trial basis. Listeners are asked to judge their perceived global loudness over hundreds of trials (for example, in loudness discrimination tasks) and a relationship between their judgments and the variations introduced on each trial is then assessed by means of correlational analyses. These studies showed that listeners' global loudness judgments result from the use of non-uniform temporal weighting strategies, a finding that is particularly interesting because it is incompatible with what is assumed in all current loudness indicators, namely that all the different portions of the sounds contribute similarly to global loudness (see Oberfeld et al. 2012). More specifically, the results found by the "molecular" studies on 1-s flat stimuli indicate that the first portions and, to a lesser extent the final portions are weighted more heavily than the portions in the middle, a result that is consistent with what was observed concerning longer sequences showing the presence of memory effects such as "recency" effects (e.g., Susini et al. 2002). A study conducted by Oberfeld & Plank (2011) has begun to address the case of sounds with *non-flat* level profiles but did not yet specifically examine the case of basic stimuli rising or falling in level such as those the present thesis is concerned with.

All together, these results may suggest that asymmetries in global loudness judgments may already arise with temporally asymmetric stimuli of a few seconds (like rising and falling sounds) because non-uniform temporal weighting strategies are deployed.

0.4.1.5 Investigating intensity coding as a complementary approach

Investigating sound intensity *coding* at the different stages of the auditory system might also be considered a complementary approach to understanding the mechanisms that underlie loudness perception. Indeed, studying intensity coding can help to clear up the relationships inside the "black box" (inferred from psychophysical experiments) and relate them with biomechanical, physiological and neural properties of the auditory system. Such studies were mostly conducted with stationary sounds and have characterized the main and fundamental functions of sound level coding in the auditory system, from low-level peripheral stages to high-level cortical stages (see Clarey et al. 1992, pp. 245-251; Moore 2003; Schreiner and Malone 2015). However, the relationship between loudness, i.e. a "psychological" quantity, and physiological measurements that can be made along the auditory pathway is still a sensitive issue (Langers et al. 2007; Schreiner and Malone 2015). Only a small number of studies examined neural coding and processing with time-varying sounds (e.g., Lu et al. 2001; Seifritz et al. 2002), and almost nothing is known about the neural basis of *global loudness* perception of sounds that vary over a few seconds. A better understanding of the cortical processing of time-varying stimuli would certainly constitute a worthwhile asset to determine the temporal mechanisms that can be highlighted in psychophysical studies.

0.4.2 Approach of the present thesis

The present thesis explores *global loudness* processing of diotic time-varying sounds. The case of stimuli composed of rising or falling intensity profiles with linear increase and decrease of level (dB) over a few seconds[21] is considered in particular, in agreement with numerous past studies that investigated the percepts associated to these profiles (e.g. Neuhoff 1998; Teghtsoonian et al. 2005; Olsen and Stevens 2010; Meunier et al. 2014). We will specifically focus on the *asymmetries* in global loudness judgments between these profiles. Indeed, while previous studies investigating increasing-intensity and decreasing-intensity sounds of a few seconds were exclusively focused on their asymmetries in *loudness change*, recent experiments conducted in the lab also reported asymmetries in *global loudness* (Trapeau, 2009). We believe that studying the *global loudness asymmetries* induced by such controlled stimuli, i.e. varying monotically in level over a few seconds, constitutes the necessary step before investigating longer time-varying sequences composed of multiple sound events, which are made of both increasing and decreasing intensity profiles.

As indicated earlier, the terms "asymmetry" will be used throughout the manuscript to refer to the global loudness *difference* between rising and falling stimuli. The asymmetry is thus a signed measure of the global loudness difference between rising and falling sounds, which can be evaluated both from magnitude estimation and loudness-matching tasks, as indicated in Figure 4. By convention, a positive asymmetry of x dB, will always indicate that the rising sound was perceived louder than the falling sound, and that the falling sound thus needs to be presented x dB higher in level than the rising sound to be perceived as equally loud.

[21] Throughout this thesis, the stimuli will always be made of linear increasing or decreasing-level profiles. Although such profiles differ from those of natural moving auditory objects (e.g., the sound of a car moving at a constant velocity from a fixed observer), linear level changes (in dB) constitute a first reasonable approximation when considering sounds increasing and decreasing intensity in general (e.g., musical sounds, moving objects in natural, etc.), as shown in Ponsot (2012).

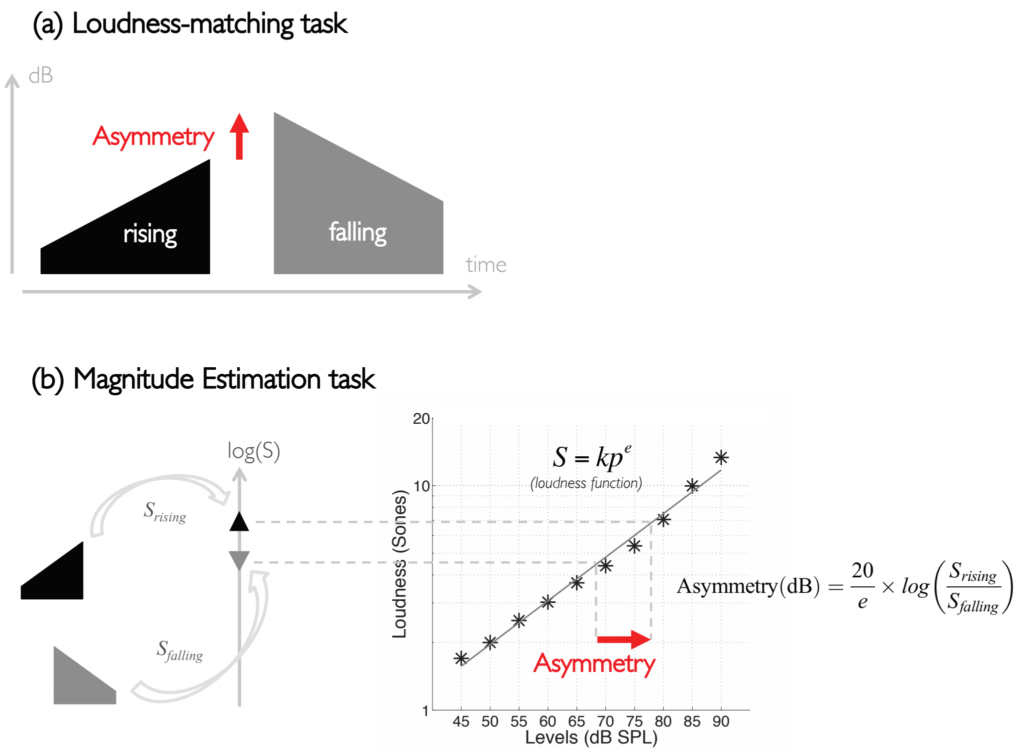

Figure 4: Schematic representation of the methods employed to derive a global loudness asymmetry between rising and falling-intensity sounds from different measurement conditions (as done throughout this thesis). (a) The asymmetry is directly obtained in dB units in a loudness-matching task (see Chapter 2 for details). (b) The asymmetry is derived from the global loudness estimates given to rising and falling sounds in a magnitude estimation task. The difference of the logarithm of these estimates is converted into dB units using the exponent e of the loudness function derived from judgments of constant-intensity sound (cf. the formula; see Chapter 1 for details).

A systematic approach is adopted in this thesis to investigate the sensitivity of global loudness *asymmetries* to a number of methodological, contextual and physical variables that were brought to light in previous related studies. Using approaches and methodologies primarily inspired from the literature on perceptual asymmetries at shorter timescales, our goal is to characterize and attempt to identify the mechanisms underlying global loudness processing and evaluation of rising and falling sounds.

The first study of this thesis investigates the robustness of the asymmetry to different measurement methods (results from both magnitude estimation and loudness-matching experiments being compared, as shown in Figure 4) and explores potential contextual or sequential effects that have been reported at shorter timescales (Stecker and Hafter, 2000), or concerning loudness change asymmetries (Olsen et al., 2010). The other studies aim to characterize the properties of the underlying mechanisms by manipulating the physical characteristics of such stimuli. In particular, the second study evaluates the influence of the spectral content of the sounds on the size of the asymmetry, as investigated by Stecker and Hafter (2000) on the loudness of short asymmetrical stimuli, and also considered in previous studies on loudness change (e.g. Neuhoff 1998). The third study manipulates the temporal profile characteristics of these sounds made of linear ramps of levels, namely their *slope*, *dynamics* and *duration*, in order to reach a better understanding of the *overall* underlying integration processes. The fourth study attempts to gain further insight into the *local* properties of the temporal integration processes in place with rising and falling intensity tones by investigating their temporal weighting of loudness using "molecular" psychophysics (i.e. psychophysical reverse correlation).

Our results will be compared with predictions from current loudness models in order to evaluate the extent to which they could be accounted for by already identified mechanisms. In particular, this will allow to examine the assumption made by Zwicker, Glasberg and Moore, that *global loudness* may be predicted by the peak value (or any percentile) of the short-term loudness (or the long-term) loudness given by the models. While this assumption constitutes a particularly parsimonious transformation of "momentary sensations" (time-series) into an "overall impression" (scalar), it has never been empirically validated in the case of sounds that rise or fall in level over of a few seconds.

Although this thesis is exclusively concerned with a *psychophysical* investigation of the phenomenon, we will attempt to put our findings into perspective with both psychophysical and neuroscientific studies. The general discussion will also mention the main results obtained from the collaboration initiated with Brice Bathellier[22] and his team who investigate the cortical bases of rising and falling intensity sounds in mice.

0.4.3 The LoudNat project

This thesis fits in the framework of a larger project called *LoudNat*[23], (Loudness of Natural sounds), which is concerned with the loudness of natural sounds in real listening conditions. The present thesis focuses on the mechanisms underlying *global loudness* of non-stationary *diotic* stimuli and more particularly, those involved with sounds increasing and decreasing in intensity. Another aspect of the project is concerned with "binaural integration of loudness", by considering loudness of stationary sounds in binaural listening situations involving strong interaural differences. This issue was addressed in the PhD thesis of Michaël Vannier (Vannier, 2015). Finally, the third part of the project aims to bridge these two parts by considering global loudness of non-stationary signals in binaural listening conditions. The LoudNat project involves three different French labs: LVA (Insa, Lyon), LMA (CNRS, Marseille) and Ircam (STMS lab, Paris).

0.4.4 Organization of the manuscript

The present manuscript is organized around four experimental chapters that explore the mechanisms underlying global loudness asymmetries between rising and falling sounds from the different perspectives specified above. Each chapter contains an already published article (Chapters 1, 2 and 4) or a paper currently in preparation (Chapter 3). The studies presented in these chapters are preceded by a preface and followed by a partial synthesis to link them with the general approach of the thesis. The order in which the studies appear in this manuscript has been chosen to provide an evolving comprehension of the investigated phenomenon. Complementary analyses

[22]"Cortical dynamics and multisensory processing" Team at UNIC, UPR CNRS 3293, Gif-sur-Yvette, France

[23]The LoudNat project is supported by the French National Research Agency (ANR-11-BS09-016-01).

and ongoing works related to the PhD but not directly connected to the asymmetries between rising and falling sounds are presented in appendices.

Chapter 1 was concerned with the characterization of the asymmetry as a function of methodological and contextual factors; Chapter 2 focused on the influence of the spectral structure and the intensity-region on the size of the asymmetry; Chapter 3 examined how temporal profile characteristics are integrated in global loudness judgments of rising and falling tones and finally, Chapter 4 explored the temporal integration processes underlying global loudness using reverse correlation analysis. Finally, the findings of these four chapters are synthesized in a general discussion section and several research perspectives for future works are proposed.

Chapter 1

A robust asymmetry in loudness between rising and falling intensity tones

Emmanuel Ponsot, Patrick Susini, and Sabine Meunier. A robust asymmetry in loudness between rising-and falling-intensity tones. *Attention, Perception & Psychophysics*, 77(3):907–920, 2015.

Preface

Judgment asymmetries have been widely reported in the literature between stimuli made of rising vs. falling intensity profiles in the context of *loudness change* perception (e.g., Neuhoff 1998; Olsen et al. 2010). However, despite the importance of considering the overall perception of loudness *per se* instead of the overall perception of loudness dynamics, much less has been done to investigate the mechanisms underlying *global loudness* perception and evaluation of these basic stimuli (Susini et al., 2010), except a recent work that reported perceptual asymmetries also occurred in *global loudness* evaluations of rising vs. falling tones (Trapeau, 2009).

This chapter presents the first experimental study exclusively focused on *global loudness* asymmetries arising with simple stimuli either rising or falling in intensity. Because different measurement methods may lead to different loudness judgments (like it was found for *loudness change* judgments, e.g. Susini et al. 2010), we conducted two psychophysical experiments using different tasks (absolute magnitude estimation in Exp. 1, loudness matching in Exp. 2). In addition, each task was tested within different contexts of stimuli presentation (mixed vs. separate contexts in Exp. 1; direct vs. indirect comparisons in Exp. 2A and Exp. 2B, respectively). Global loudness of the same 1-kHz, 2-s rising and falling ramps stimuli varying over 15-dB dynamics was measured for each context of presentation in both tasks and their resulting asymmetries were computed. This study had two main goals:

(1) To test the robustness of the magnitude of the asymmetry depending on the measurement paradigm and the context of stimuli presentation

(2) To examine how current loudness models may account for the asym-

metries in different experimental configurations

This work was published in *Attention, Perception, & Psychophysics*: Ponsot, E., Susini, P., Meunier, S. (2015). A robust asymmetry in loudness between rising-and falling-intensity tones. *Attention, Perception, & Psychophysics*, 77(3), 907-920.

Further analyses of these asymmetries on an individual basis are presented in Appendix A. They are discussed together with the main findings of the paper in the partial synthesis of this chapter.

Abstract: Tones rising in intensity over a few seconds are perceived as louder than symmetrical tones falling in intensity. However, the causes for such perceptual asymmetry, as well as its magnitude and dependency on contextual and methodological factors remain unclear. In this paper, two psychophysical experiments were conducted to measure the magnitude of this asymmetry for 2-s, 15-dB intensity-varying tones in different conditions. In the first experiment, participants assessed the global loudness of rising and falling-intensity sounds with an absolute magnitude estimation procedure (AME); in the second experiment, they compared sounds relatively in an adaptive, two-interval, two-alternative forced-choice task (2I-2AFC). In both experiments, the region of intensity change, the design of experimental blocks and the type of comparison stimulus were systematically manipulated to test for contextual and methodological factors. Remarkably, the asymmetry was virtually unaffected by the different contexts of presentation and similar results with 2I-2AFC and AME measurements were obtained. In addition, the size of the effect was comparable over all but the highest intensity regions (80-90 dB SPL), at which it was significantly smaller. All together, these results indicate that the loudness asymmetry is preserved under different measurement methods and contexts, and suggest that the underlying mechanism is strong and robust. In short, falling tones have to be about 4 dB higher in level than symmetrically rising tones in order to be perceived with the same global loudness, a finding that is still not predicted by current loudness models.

1.1 Introduction

While most of psychoacoustical research has been primarily concerned with the perception of stationary signals, a number of studies in the past decade have started investigating more more complex, time-varying stimuli (e.g., Chi et al. 2005; Lu et al. 2001; Neuhoff 1998; 2012). Undoubtedly, time variation is essential to completely understand the perception of environmental or musical sounds, for which spectral and energetic characteristics constantly change over time. Processing such changes involves particular attentional and cognitive mechanisms, and recruit specific neural circuits. As pointed out recently by Schutz and Vaisberg (2014), although too often neglected in the past, a sound's temporal shape plays a crucial role in its perception. In particular, perceptual differences in loudness, timbre or perceived duration have been observed between tones that have identical energy and long-term spectrum, but opposite temporal profiles. Because they occur between temporally asymmetric stimuli – the term "temporal asymmetry" (Patterson, 1994) was used to describe the sounds with different attack and decay times –, these arising perceptual differences were often referred to as perceptual "asymmetries" (e.g. see Grassi and Pavan 2012; Meunier et al. 2014; Ries et al. 2008; Susini et al. 2007). As a result, this term will also be used throughout the rest of this paper.

The present work is specifically concerned with asymmetries in global loudness. Listeners asked to make global loudness judgments of rising and falling-intensity sounds of a few seconds, i.e. "to evaluate the overall loudness of the sound over its entire duration", give greater estimates to rising sounds compared to their time-reversed versions, falling sounds (Ponsot et al. 2013; Susini et al. 2007). The causes and mechanisms underlying these asymmetries remain unclear.

Most of previous studies on auditory perceptual asymmetries have relied on short sounds (from 10 to 250 ms) with temporally asymmetric amplitude envelopes. This concerns the domains of physiology (e.g. Neuert et al. 2001), psychophysics (e.g. Meunier et al. 2014; Schlauch et al. 2001; Stecker and Hafter 2000) as well as neurosciences (Wang et al., 2014). Temporally asymmetric stimuli were used to model the amplitude envelope of environmental and musical sounds. Percussive-like also called damped stimuli have been modeled using envelopes with fast attacks and slow decays. Similarly, bowing-like also called ramped stimuli (with slow attacks and fast decays) were mimicked using temporally reversed versions of the percussive-like sound envelopes. Psychophysical studies have shown strong perceptual asymmetries between these two types of stimuli for timbre (Irino and Patterson, 1996), perceived duration (e.g. Meunier et al. 2014; Schlauch et al. 2001) and loudness (e.g. Stecker and Hafter 2000). Physiological measures have also supported these psychological findings. Investigating the response characteristics of primary auditory cortex neurons to ramped and damped 80-ms stimuli in awake cats, Wang et al. (2014) showed that distinct cells were involved in the processing of ramped and damped sounds, and that cells activated by ramped stimuli had longer, more persistent responses than those stimulated by damped stimuli. That ramped stimuli be associated with a relative “persistence of excitation” was taken to corroborate the theory that “persistence of perception” at the behavioral level (DiGiovanni and Schlauch 2007; Ries et al. 2008) explains why ramped sounds are perceived as longer and louder than damped sounds. Cognitive factors were also found to play a role in these perceptual asymmetries between short sounds. Stecker and Hafter (2000) proposed the existence of a "decay suppression" mechanism, with which listeners would ignore the decay part of damped stimuli, because it represents reverberation from the acoustical environment, resulting in sounds judged softer than ramped stimuli. The same authors, observing that the loudness asymmetry between 250-ms ramped and damped sounds was significantly reduced and even removed when ramped stimuli were presented as priors, proposed that this “local context” effect was based on a cognitive mechanism that favored the subjects processing in a particular representation-mode. Damped prior stimuli would bring the subjects to process in a “constancy mode” and to eliminate the decay portion (reverberation part) of the following stimulus. Later, DiGiovanni and Schlauch (2007) demonstrated that the size of the asymmetry in perceived duration between ramped and damped stimuli was modulated by the instructions given to the participants. When the participants were explicitly asked to consider all aspects of the sounds, and to not ignore decay portions, the asymmetry was found to be significantly reduced. These results demonstrate that cognitive aspects of sound evaluation can modulate the perceptual asymmetries that occur with temporally asymmetric sounds of short duration.

Although less documented, the perception of non-stationary sounds of longer durations has also been explored, in particular with rising and falling-intensity sounds lasting a few seconds. While asymmetries in perceived duration appear to fade-out after 1s (Grassi and Darwin 2006; Grassi and Pavan 2012; Meunier et al. 2014), asymmetries in loudness are still observed with longer sounds. Especially, Neuhoff (1998) observed that 1.8-s rising-intensity tones were perceived with a greater *loudness change*[1] than falling-intensity tones. He argued that the overestimation of rising sound sources would provide a "selective advantage" to an organism's preparation for approaching objects, characterized by their rising-intensity profiles, compared to receding objects that produce falling-intensity profiles. This effect was also demonstrated with tones, chords and vowels stimuli lasting 1.8 and 3.6 s (Olsen et al. 2010; Olsen and Stevens 2010). Neuroscience and psychophysiological experiments on humans showed that rising-intensity tonal stimuli from 750 ms to 2 s produce neural activity in the amygdala, recruit specific attentional and physiological resources and induce higher emotional ratings compared to falling-intensity sounds (Bach et al. 2008; Bach et al. 2009; Seifritz et al. 2002; Tajadura-Jiménez et al. 2010). Similar results are observed in non-human primates, for whom behavioral and neural differences are also found between the two types of stimuli (Ghazanfar et al. 2002; Maier and Ghazanfar 2007). This bias in primates is considered adaptive, serving as a warning cue to process looming sound sources with higher priority. Furthermore, recent behavioral studies reported a perceptual asymmetry between 2-s rising and falling-intensity tones when participants were instructed to evaluate the global loudness of these sounds (Ponsot et al. 2013; Susini et al. 2007). Although less studied than the loudness change, the global loudness is of particular interest in many fields of applications concerned with loudness measurement, e.g. loudness normalization of audio data in the media (Glas-

[1] *Loudness change* refers to the overall perceived amount of loudness change for rising or falling-intensity sounds, as first defined by Neuhoff (1998).

berg and Moore, 2002) or loudness assessment of vehicle and road-traffic noise in industry (e.g. Kaczmarek and Preis 2010). Yet, the magnitude and the robustness of the global loudness asymmetry between rising and falling-intensity sounds of a few seconds have never been systematically assessed. In particular, it remains unclear how much this asymmetry could rely on contextual factors, as previously demonstrated for shorter-duration stimuli.

The present study was designed for that purpose. The magnitude of the asymmetry was estimated from loudness measurements of rising and falling-intensity sounds obtained in several experimental contexts and using different psychophysical methods. We examined both "global context" effects, by manipulating the type of stimuli presented within an experimental block, and "local context" effects, by manipulating the type of stimuli that immediately precedes a sound. In addition, we measured asymmetries for ramps presented at different intensity regions, to explore potential level-dependent effects.

The "global context" of stimuli presentation was rarely considered as a factor in previous studies. In most of them, measuring either the loudness change or the global loudness of rising and falling-intensity sounds, the two types of sounds were either presented in a random order in one-interval paradigms (Neuhoff 1998; Susini et al. 2007; Teghtsoonian et al. 2005) or compared one with another in two-interval paradigms (Neuhoff 2001; Olsen et al. 2010; Olsen and Stevens 2010; Ponsot et al. 2013). However, it is now recognized that the range, the spacing, and the distribution of a set of stimuli that only vary in intensity influence the loudness evaluation of each stimulus (Arieh and Marks, 2011). Thus, the way in which different types of stimuli are presented plausibly also affects their evaluation. In a recent study investigating the perceived duration of damped and flat-intensity stimuli, Vallet, Shore, and Schutz (2014) showed that the strategies underlying duration estimation for the two types of sounds can be modulated by the "blocking procedure", namely whether the two types of stimuli are presented in "separate" blocks or "mixed" within a same block. In "separate" presentations where each type of stimulus is presented in a specific block, participants can guess the type of the stimulus that follows with 100% certainty, and thus can switch from one evaluation strategy to another every time a new block starts. On the contrary, in "mixed" presentations where stimuli types are randomly interleaved, participants cannot predict the type of the following stimulus. In this case, listeners seem to adopt a common, low-cost evaluation strategy for both profiles (Vallet et al., 2014). In addition, the magnitude of the asymmetry in perceived duration measured between ramped and damped stimuli in Grassi and Pavan (2012) was found to be larger in a “mixed” context compared to a “separate” context. The authors argued that the “mixed context” emphasizes the perceptual distinctiveness of the stimuli, resulting in larger perceptual asymmetries (p. 1329; Grassi and Pavan 2012). We can thus reasonably assume that a similar phenomenon might impact the loudness asymmetry between rising and falling-intensity sounds. This hypothesis was tested in a first experiment using a magnitude estimation procedure (Experiment 1): asymmetries in loudness between rising and falling stimuli were compared between two “global contexts”, using magnitude estimates obtained for two groups of participants, performing the experiment in either a “mixed” or in a “separate” context of presentation.

Besides “global context”, “local context” effects on the loudness asymmetry have also rarely been considered. However, several of such “sequential” effects have been documented for constant-intensity stimuli (for a review, see Arieh and Marks 2011), and two of them may worth to be mentioned. First, with “loudness enhancement”, the loudness of a tone is increased when a preceding tone is presented within a short temporal window under certain level conditions (for details, see Oberfeld 2007). Second, with “Induced Loudness Reduction” (ILR), the opposite effect is reached when the preceding tone is relatively loud (Epstein, 2007). Despite comforting the view that the estimation of a stimulus does not only reflect its perception at a given time but also depends on the acoustical characteristics of previous stimuli (Lockhead, 1992), such “local context” effects have received only little attention regarding non-stationary sounds. With temporally asymmetric sounds, the mechanism of decay suppression (Ries et al. 2008; Stecker and Hafter 2000) has only been observed for short-duration sounds. The influence of prior stimulus direction remains to be explored with longer sounds and no specific assumption can be made concerning the potential mechanisms that could modulate their loudness perception and evaluation. The influence of “local context” was thus explored in Experiment 1 by considering the influence of the prior stimuli direction on the loudness estimates given by the "mixed context" group. In addition, a second experiment was conducted (Experiment 2) to examine whether the type of comparison stimulus, in a two-interval paradigm, affects the loudness asymmetry between rising and falling-intensity tones. This was done by comparing the asymmetry assessed with two groups of participants that were asked (i) to match the loudness of rising and falling-intensity sounds presented in a same pair (Experiment 2A), or (ii) to match the loudness of each ramp relatively to constant-intensity tones (Experiment 2B). Note that this latter condition did not allow subjects to make any direct opposition when comparing the two "objects"(rising vs. falling), contrary to what was done in previous studies. If the global loudness judgments of the ramps inferred from these different matching configurations provide similar asymmetries, it would indicate that listeners use a common and robust loudness evaluation strategy. Moreover, by ruling out simple explanations of the effect based on judgment bias only, this would suggest that rising and falling tones contain an intrinsic characteristic, which results in directional-specific loudness judgments.

In the end, the robustness of the loudness asymmetry is discussed with respect to the quantitative psychophysical estimates obtained in the present experiments (Experiment 1, Experiment 2A and Experiment 2B) and the magnitude of the effect is compared with predictions from current loudness models.

1.2 Experiment 1

In this experiment, the loudness of 2-s rising and falling intensity tones with 15-dB dynamics (changing in level over 15 dB) was measured in a one-interval paradigm using an absolute magnitude estimation procedure (AME). The influence of "global context" was examined using a between-subject experimental design. In one group of participants (the "mixed context" group), rising and falling ramps were interleaved and presented in the same blocks; in the other group (the "separate context" group), they were presented in two distinct blocks. The influence of "local context" was examined as a within-subject factor in the "mixed context" group only, by considering the influence of prior stimuli direction (rising/falling) on loudness estimates.

1.2.1 Method

1.2.1.1 Participants

A total of thirty-two participants (ages between 20 and 34 years) were recruited for this experiment. They were assigned to one of the two groups (N=16 and sex-ratio =1 in each group), which performed the experiment under different conditions, as described below. All reported normal hearing. The participants were paid for their participation. They were all naïve with respect to the hypotheses under test.

1.2.1.2 Apparatus

The stimuli were generated at a sampling rate of 44.1 kHz with 16-bit resolution using Matlab. Sounds were converted using a RME Fireface 800 soundcard, amplified using a Lake People G-95 Phoneamp amplifier and presented diotically through headphones (Beyerdynamic DT 770 PRO). The levels were calibrated using a Brüel & Kjær artificial ear (type 4153) coupled with the mounting plate provided for circumaural headphones. Each participant was tested in a double-walled IAC sound-insulated booth.

1.2.1.3 Stimuli

All the stimuli were 1-kHz pure tones with 10-ms linear rise and fall times (on amplitude). Ten constant-intensity tones of 500 ms (between 45 and 90 dB SPL) were used to measure the individual loudness functions. Otherwise, linear (in decibels) rising and falling-intensity ramps with 15-dB dynamics and 2-s duration were employed in the experiment. These ramps were presented at seven different intensity regions: [45-60], [50-65], [55-70], [60-75], [65-80], [70-85] and [75-90 dB SPL] (in what follows, the ramps varying, for example, in the [45-60 dB SPL] region are simply referred to as 60-dB SPL ramps).

1.2.1.4 Procedure

The Absolute Magnitude Estimation (AME) procedure was used for this experiment, based on the instructions of Hellman (1982). No standard was given to the participants, whose task was simply to give a number proportional to the global loudness of each sound regardless to the numbers assigned to previous stimuli. We asked participants to judge the global loudness of the sounds, as in previous studies (Ponsot et al. 2013; Susini et al. 2007; Susini et al. 2010), defined as "the loudness of the sound over its entire duration". For each group of participant, the experiment was divided into two similar sessions scheduled on different days. Each session started with the measurement of the loudness function. As in Canévet, Teghtsoonian, and Teghtsoonian (2003), the ten constant-intensity tones were presented in a "quasi-random" order to reduce assimilation effects, as proposed by Cross (1973). These tones were evaluated 18 times each in a session. The loudness function measurement was followed by the main experiment, which was composed of the 14 rising and falling ramps presented 13 times each. For one group of participants (the "separate context" group), rising and falling ramps were presented in two distinct blocks whereas for the other group (the "mixed context" group), rising and falling ramps were randomly mixed in the same experimental block. Thus, each participant performed the magnitude estimation task in the same global context ("mixed" or "separate") over the two experimental sessions. A between-subject design was adopted in order to avoid any "asymmetrical transfer effect" between the two conditions (see Poulton and Freeman 1966; Poulton 1982). In each block, the ramps were presented in random order. For the "separate context" group, the order of presentation of the two blocks (rising / falling) was counterbalanced between subjects. Before each session, participants were instructed not to change their scale between the preliminary measurement of the loudness function, where they rated the loudness of constant-intensity tones, and the following blocks that contained rising and falling-intensity sounds. Subjects became familiar with the procedure by giving their estimates for 20 constant-intensity sounds (between 45 and 90 dB SPL) before the beginning of each session, which lasted approximately 1 hour.

1.2.1.5 Normalization and data analysis

For each subject, the average perceived loudness of each stimulus was computed using the geometric mean of the ratings. Mean loudness estimates were then normalized individually for each session so that all estimates fall within the same range and could be expressed in loudness unit (sone). All the ratings given by a subject to constant tones and ramps in a daily session were divided by the mean rating assigned to the 50-dB SPL constant-intensity tone and multiplied by 2, to match the loudness of a 50-dB SPL, 1-kHz pure-tone, which corresponds to 2 sones (see Susini et al. 2010). Power loudness functions were then fitted to constant-intensity sound estimates to evaluate the loudness exponent of the loudness function of each subject in each session.

A mixed-ANOVA [1-between (Context) and 3-within (Direction $\times$ Session $\times$ Level) factors] using a univariate approach was performed on the logarithm of the normalized loudness ratings accorded to rising and falling-intensity stimuli. Since no significant differences between experimental sessions and no interactions with other experimental factors were found ($p > .05$), the data from the two experimental sessions were then pooled together and the analysis was rerun without the factor "Session". All the statistical analyses presented in this study were conducted using R (R Core Team 2015). Unless otherwise specified, all the tests were two-tailed and used a probability level of .05 to test for significance. The Huynh–Feldt correction for degrees of freedom was used where appropriate and $\tilde{\epsilon}$ correction factors are reported. Effect sizes are reported using generalized eta-squared η_g^2 (Bakeman, 2005).

1.2.2 Results

1.2.2.1 Loudness asymmetry

The mean loudness ratings given to rising and falling ramps are plotted as a function of the maximum level of their range of intensity-variation (between 60 and 90 dB SPL) in Figure 1.1, panel (a). The black symbols correspond to the estimates of the "mixed context" group and the grey symbols to the estimates of the "separate context" group. To evaluate quantitatively the asymmetry in loudness between rising and falling-intensity ramps for the two groups of participants, we converted loudness ratings given by each subject into phon units by using his own loudness exponent. For a given maximum level, the loudness estimated in sones (Figure 1.1, panel (a)) was converted into a loudness level in phon by finding the corresponding dB SPL value on the loudness function for both rising and falling tones. The difference between rising and falling tones loudness levels (in phons) provides a measure, in decibels, of the asymmetry. This calculation is made for each listener, at each intensity region. Mathematically, it corresponds to the use of the following index:

$$Asymmetry(dB) = \frac{20}{e} \times log(\frac{S_{rising}}{S_{falling}})$$

where e is the individual exponent of the loudness function and S_{rising} and $S_{falling}$ are the mean loudness ratings measured in the experiment for rising and falling tones, respectively. The asymmetries computed with this index are presented on panels (b) and (c) of Figure 1.1.

As it can be seen on the panel (a) of Figure 1.1, rising ramps were perceived as louder than falling ramps in the two "global" contexts. This claim was supported by the statistical analysis, which revealed a significant effect of stimuli Direction (rising / falling) $[F(1, 30) = 34.390, p < .001, \eta_g^2 =$

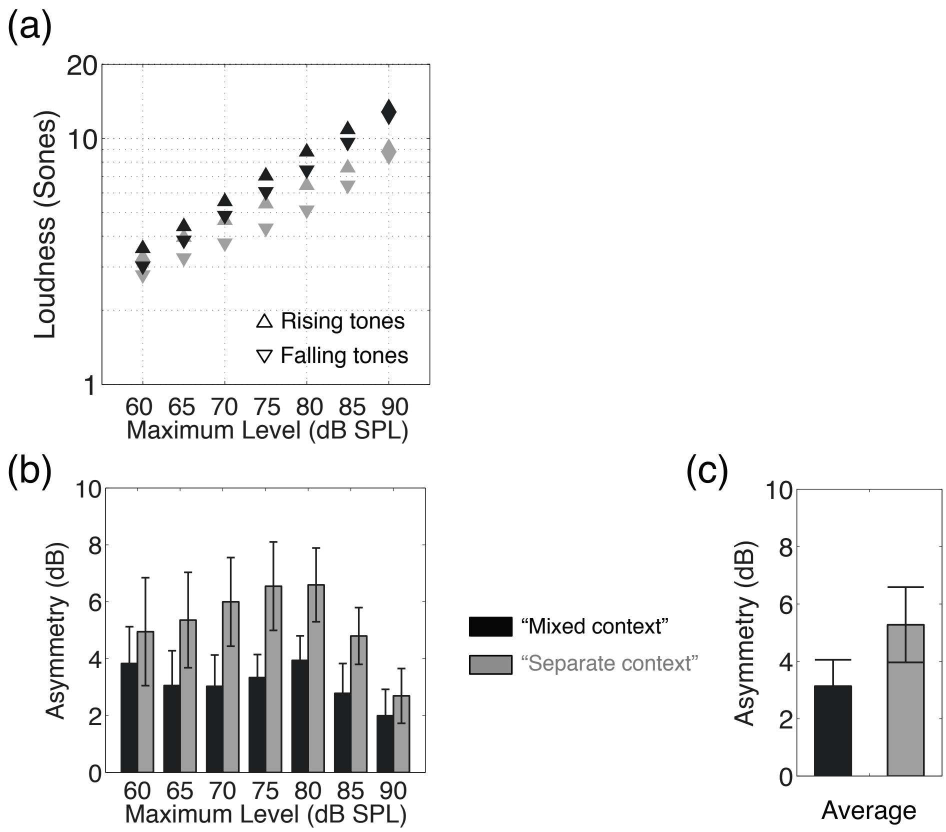

Figure 1.1: Results of Experiment 1. (a) Mean normalized loudness ratings, in sones, of the "mixed context" (in black) and the "separate context" (in grey) groups plotted as a function of the maximum level of the ramps. (b) Loudness asymmetries between these rising and falling ramps assessed for the different levels, using the introduced index, for each group (black bars, "mixed context"; grey bars, "separate context"). (c) Asymmetries averaged over the different intensity regions for each group. Note that positive values indicate that rising-intensity tones were perceived as louder than falling-intensity tones. Error bars indicate standard errors of the mean (SEM).

.01]. A small but significant interaction Level × Direction was also found [$F(6, 180) = 4.657, p < .001, \tilde{\epsilon} = .51, \eta_g^2 = .0005$], indicating that the size of the asymmetry is not the same at all levels. However, the same mixed-ANOVA but without the estimates for the 90-dB SPL ramps gives no significant interaction Level × Direction ($p > .05$), which shows that the asymmetry is in fact similar over the different regions from 60 to 85 dB SPL and that the observed interaction is due to the estimates given for the 90-dB SPL ramps.

These asymmetries were assessed quantitatively using the formula above and were averaged over the different intensity regions to evaluate the overall magnitude of the effect (see panel (c), Figure 1.1) One-tailed t-tests conducted on these data show that the mean asymmetries are significantly positive; 5.3 dB (SD=5.2) for the "separate context" group [$t(15) = 4.017, p < .001$] and 3.1 dB (SD=3.7) for the "mixed context" group [$t(15) = 3.4215, p < .01$].

1.2.2.2 Global context effect

The influence of the "global context" can be observed directly in the ANOVA by looking at the interaction Context × Level. This interaction turned out to be not significant ($p > .05$), which does not support the hypothesis that a "mixed context" would have caused a greater asymmetry. A second mixed-ANOVA [1-between (Context) and 1-within (Level) factors] was also conducted on the asymmetries computed with the proposed index to comfort this observation (i.e. based on the asymmetries plotted on Figure 1.1, panel (b)). Again, no influence of the context was found ($p > .05$) and a significant effect of the Level was obtained [$F(6, 180) = 4.667, p < .001, \tilde{\epsilon} = .42, \eta_g^2 = .031$], which was caused by a smaller asymmetry at 90 dB SPL compared to other levels, as confirmed by a post-hoc mixed-ANOVA conducted on the data but without the estimates for the 90-dB ramps, where the effect of the Level became non-significant ($p >$.05).

1.2.2.3 Local context effect

To examine potential "local context" effects, another ANOVA was conducted on the results of the "mixed context" group only, taking into account the direction of the preceding stimulus (PrecDir) on each rating [i.e. a within three-factor repeated-measure ANOVA (Direction × PrecDir × Level) was conducted]. The ratings of the "mixed context" group are plotted as a function of the maximum level of the ramp on the left panel of Figure 1.2, with the preceding stimulus direction as a parameter. A small but significant effect of the preceding stimulus direction was found [$F(1,15) = 11.661, p < .01, \eta_g^2 = .0014$], as well as significant two-order and three-order interactions [PrecDir × Direction: $F(1,15) = 8.945, p < .01, \eta_g^2 = .0003$; PrecDir × Direction × Level: $F(6,90) = 5.126, p < .001, \tilde{\epsilon} = .94, \eta_g^2 = .0019$]. As it can be observed on the left panel of Figure 1.2, the estimates given to falling tones (grey and black downward triangles) were virtually unaffected by the preceding stimulus direction, whereas the estimates given to rising tones were slightly increased when they were presented at low levels and preceded by rising tones (black upward triangles), compared to falling tones (grey upward triangles). Asymmetries computed for the different levels and priors with the index proposed in the paper are presented on the right panel of Figure 1.2. Greater asymmetries are obtained at 60 and 65 dB SPL when the preceding stimulus is a rising ramp compared to a falling ramp.

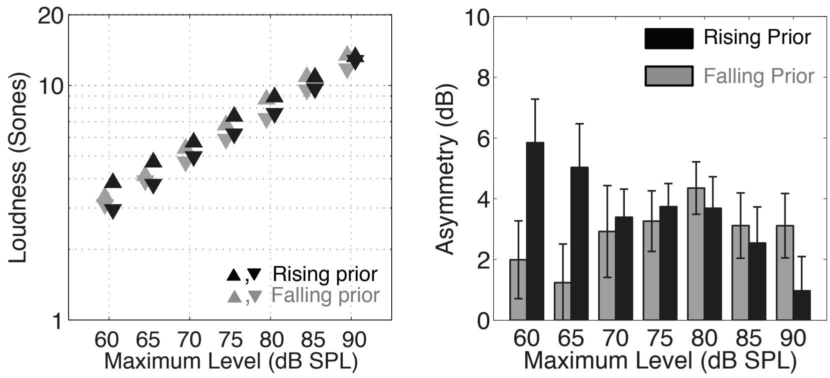

Figure 1.2: Left panel: Mean normalized loudness ratings, in sones, of the "mixed context" group specifically. Upward and downward triangles correspond to estimates given to rising and falling ramps, respectively, when the preceding tone was either a rising ramp (black triangles) or a falling ramp (grey triangles). Right panel: Asymmetries computed for each level and each configuration (rising prior with black bars, falling prior with grey bars). Error bars show standard errors of the mean (SEM).

1.2.3 Discussion

In this experiment, the loudness of rising and falling-intensity tones was measured using an AME procedure in two different experimental contexts. The present results confirm the existence of loudness asymmetries between rising and falling tones in both contexts.

Our first research question was whether the global context of stimulus presentation influences the magnitude of the loudness asymmetry. The results show no significant difference between the asymmetries measured in “mixed” and “separate” contexts. Our hypothesis that a “mixed context” would have reinforced the loudness asymmetry, compared to a “separate context”, cannot be verified here. These results indicate that listeners don’t need to compare the two types of stimuli, and have access to some *intrinsic* information, which lead them to produce different global loudness estimates. This provides evidence for the robustness of their evaluation strategy.

Our second research question was the influence of the local context of stimuli presentation. We found a significant effect of prior stimulus direction, but the size of this effect was about ten times smaller than the effect of interest (namely, the direction of the ramp). The estimates given to rising stimuli were slightly higher with rising priors than with falling priors, specifically at low sound levels (see Figure 1.2, right panel). This effect is different from that observed by Stecker and Hafter (2000) with short-duration stimuli, for which damped stimuli were judged lower when preceded by a damped stimulus than by a ramped stimulus. Because the stimuli of the present experiment were separated by silent intervals of several seconds, it is very unlikely that this level-dependent effect has sensory origins. Whether it relies on some cognitive or judgment-deviation phenomenon, induced by the end of rising ramps on following, low-level, rising ramps, still has to be determined. As a result, the asymmetry was found to be slightly greater at low levels (60 and 65 dB SPL) when assessed with rising priors than with falling priors, but remained similar at other levels. This small effect of “local context”, occurring only at low levels, suggests that the evaluation strategy was not specifically affected by the direction of prior stimuli.

Finally, the magnitude of this asymmetry was similar at all intensity regions, except for ramps culminating at 90 dB SPL, where it turned out to be significantly smaller. The reasons for this decrease only at the highest level remain undetermined. In sum, the results of this experiment reveal that the context of stimuli presentation does not particularly modulate the global loudness asymmetry between 2-s rising and falling-intensity tones. Overall, the present data show that the global loudness of falling stimuli is around 4 phons lower than the global loudness of rising stimuli.

1.3 Experiment 2

The second experiment was designed to examine the loudness asymmetry in a two-interval paradigm. Loudness of 2-s, 15-dB dynamics stimuli was measured using an interleaved adaptive 2I, 2AFC procedure. Rising and falling ramps were compared in loudness one to another by one group of subjects (Experiment 2A), while another group compared rising and falling stimuli respectively to constant-intensity stimuli (Experiment 2B). Loudness estimates were compared between the two groups, providing new insights into the sensitivity of this perceptual asymmetry to the type of comparison tone.

1.3.1 Method

1.3.1.1 Participants

Ten subjects (age 21-44 years, sex-ratio = 1) participated in Experiment 2A and twelve different subjects (age 18-33 years, sex-ratio = 1) took part in Experiment 2B. None reported having hearing problems. They gave their informed consent prior to the experiment and were paid for their participation. The participants were naïve with respect to the hypotheses under test.

1.3.1.2 Apparatus

The apparatus used was the same as described in Experiment 1.

1.3.1.3 Stimuli

The stimuli were 1-kHz pure tones with duration of 2 s. All the stimuli had 10-ms linear rise and fall times. In both experiments, the experimental design was developed to measure the loudness matches between rising and falling ramps having 15-dB dynamics, or between ramps and constant-intensity tones. For the ramps, the five lowest intensity regions tested in Experiment 1 were reemployed in this experiment: [45-60 dB SPL], [50-65 dB SPL], [55-70 dB SPL], [60-75 dB SPL] and [65-80 dB SPL]. The fixed-level test tones were rising or falling ramps (for details, see upper panels of Figures 1.3 and

1.4) and the variable comparison tones were either ramps (in Experiment 2A) or constant-intensity tones (in Experiment 2B). The maximum level of the comparison tones was limited to 95 dB SPL.

1.3.1.4 Procedure

In both Experiment 2A and Experiment 2B, a loudness matching task was employed to determine the loudness matches of the ramps in the 20 conditions [4 Configurations × 5 Levels] as presented in upper panels of Figures 1.3 and 1.4. The stimuli were presented in a 2I, 2AFC paradigm based on an interleaved adaptive procedure (Florentine et al. 1996; Jesteadt 1980). In each experiment, the loudness matches were obtained by using 20 interleaved tracks, i.e. one for each condition. On each trial, listeners heard two tones separated by 500 ms silent intervals. Their task was to indicate which sound was louder by pressing a button (labeled as “first” or “second”) on the interface. The response initiated the next trial after a 1-s delay. The level of the comparison tone for each track was adjusted according to a one-up, one-down procedure (Levitt, 1971). If the listener indicated that the comparison tone was louder than the standard its level was decreased; otherwise its level was increased. Different starting values and step sizes were chosen to make the difficulty of the task more homogenous between the different tracks. On the one hand, starting level values of comparison tones were determined in a preliminary experiment so as to be perceived either clearly louder or softer than the standard (Nieder et al., 2003). The comparison tone started above or below the expected equal-loudness level, with equal a priori probability. On the other hand, since the near miss to Weber’s law predicts an increase of sensitivity with sound level for 1-kHz tones (Rabinowitz et al., 1976), the initial step size was linearly adjusted (across the five intensity regions under study) from 3.6 dB for the ramps in the highest intensity region [65-80 dB SPL] to 5 dB for the ramps in the lowest intensity region [45-60 dB SPL].

To ensure that the tracks converge at roughly the same time, we used the same rule as in Grimm et al. (2002); each trial was assigned a track at random, but this choice was further restricted by requiring that each track be selected once before any track could be reselected. For each track, the step size was reduced by a ratio of 1.5 after every two reversals; and after six reversals, it was held constant. This procedure ended when 8 reversals was achieved for each of the 20 tracks. The level of the comparison tone thus converged towards the 50% point of the psychometric function (Levitt, 1971) for each track (i.e. each condition). The loudness matches were then defined as the average of the maximum level of the comparison tone in the last two reversals of each track. An experimental session lasted approximately 1 hour and required about 500 trials to end (i.e., 25 trials per track).

1.3.2 Results

The level differences between the second tone and the first tone of the pair ($L_2 - L_1$) at equal-loudness measured in the different configurations of Experiment 2A and Experiment 2B are presented on Figures 1.3 and 1.4, respectively, as a function of the maximum level of the ramps. Statistical analyses of each experiment were conducted on these level differences. In Experiment 2A, three factors were considered: the pair order (rising-falling / falling-rising), the position of the comparison tone in the pair (first / second) and the level of the ramp (5 levels, see above). This was done using a within three-factor repeated-measure ANOVA [Order × PosComp × Level] with a univariate approach. A similar analysis was conducted on the data obtained in Experiment 2B, where the influence of the position, the direction and the level of the standard was examined using a within three-factor repeated-measure ANOVA [PosStand × Direction × Level].

The results of Experiment 2A (Figure 1.3, lower panel) show that the level differences ($L_2 - L_1$) are positive in the rising-falling configurations – i.e., (a) and (b) – and negative in the opposite falling-rising configurations – i.e., (c) and (d). In other words, the level of the falling tone was matched higher than the level of the rising tone to produce equal-loudness. This asymmetry in loudness is supported by statistical analysis, where a significant effect of the Order is found [$F(1, 9) = 40.26, p < .001, \eta_g^2 = .35$]. However, as it can be observed on this panel, these differences depend strongly on the level considered; a significant effect of the Level is obtained [$F(4, 36) = 8.86, p < .001, \tilde{\epsilon} = .68, \eta_g^2 = .16$]. Besides, this level effect also

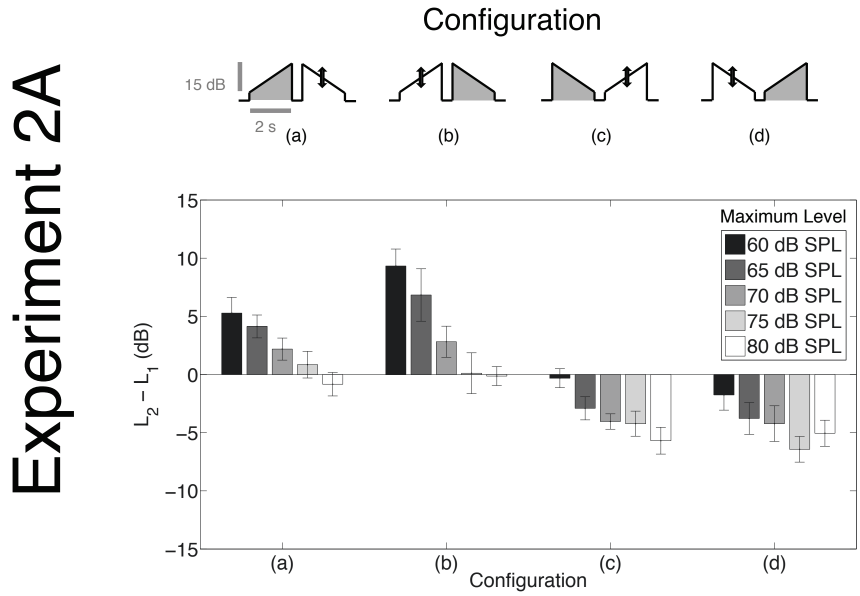

Figure 1.3: Upper panel: Experimental configurations used in Experiment 2A. The two types of standards (rising or falling ramps, filled in grey) were matched in loudness with their opposite ramps. Double arrows indicate that the levels of the comparison tones (unfilled) were varied by the adaptive procedure. Lower panel: Measured level differences between the second (L_2) and the first tone (L_1) of the pair at equal-loudness in the different configurations, derived from the obtained loudness matches, plotted as a function of the maximum level of the standards. Error bars show SEM.

depends on the pair order, as indicated by a significant Order × Level interaction [$F(4, 36) = 11.49, p < .001, \tilde{\epsilon} = .90, \eta_g^2 = .10$]. Finally, level differences measured in configurations (a) and (b) were similar, as well as in configurations (c) and (d)), as revealed by a non-significant effect of the position of the comparison tone and no significant interactions with other factors (all Ps > .05). This is due to the interleaved procedure, which ensured that subjects would not notice which configuration they were presented, and could thus not use different strategies to compare the stimuli based on which the first or the second tone was defined as the standard or the comparison tone.

The results of Experiment 2B, plotted in Figure 1.4 (lower panel), show

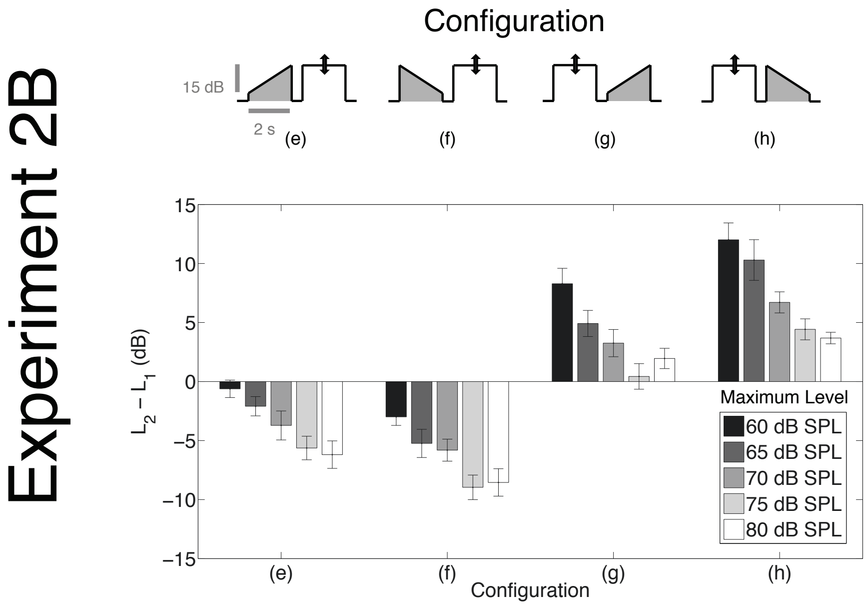

Figure 1.4: Upper panel: Experimental configurations used in Experiment 2B. The two types of standards (rising or falling ramps, filled in grey) were matched in loudness with constant tones. Lower panel: Measured level differences between the second (L_2) and the first tone (L_1) of the pair at equal-loudness in the different configurations, derived from the obtained loudness matches, plotted as a function of the maximum level of the standards. Error bars show SEM.

negative differences in the [ramp-constant] configurations –i.e., (e) and (f) – and positive differences in the [constant-ramp] configurations – i.e., (g) and (h). The analysis reveals a strong and significant effect of the position of the standard $[F(1, 11) = 124.65, p < .001, \eta_g^2 = .69]$. Thus, the level of the constant-intensity tone was matched lower than the maximum level of the ramp to produce equal loudness, i.e. the ramps were perceived as louder than constant-intensity tone presented at their maximum level. As in Experiment 2A, a strong effect of the Level can also be observed $[F(4, 44) = 39.52, p < .001, \tilde{\epsilon} = .62, \eta_g^2 = .33]$. Finally, the absolute magnitude of the level differences drawn between ramps and constant tones is larger in the conditions where the standard was falling – (f) and (h) – than where the standard was rising – (e) and (g); a significant interaction between the position of the standard and its direction (rising / falling) is found $[F(1,11) = 30.48, p < .001, \eta_g^2 = .33]$. Rising tones were thus matched at a higher level than falling tones, which provides clear evidence for a loudness asymmetry.

1.3.2.1 Time order errors

Before looking in closer detail at these loudness asymmetries and their relative magnitudes, one may ask first why the level differences were strongly influenced by the region of intensity, in both Experiment 2A and Experiment 2B. In fact, the decreasing patterns observed in all the configurations investigated in these two experiments (see lower panels of Figures 1.3 and 1.4) indicate that the first stimulus tended to be perceived more often as louder than the second stimulus when the level of these two tones felt in low intensity regions (dark grey and dark colors), compared to high intensity regions (white and light grey colors), where the opposite phenomenon is observed. Such level effects are common in experiments where the levels of the stimuli are roved trial-by-trial, or when different levels are tested – as in the present experiments – and are due to time-order errors "TOEs" (Berliner et al. 1977; Hellström 1979; Hellström 2003; Macmillan and Creelman 2004). In a roving-level paired comparison of loudness where two identical stimuli S_1 and S_2 are presented, they are almost never perceived as equal. One of the suggestions proposed to explain TOEs is that the two stimuli are weighted differently, a greater weight often being accorded to the second stimulus (Hellström 1979; Hellström 2003). As a consequence, when the levels of the two tones are in low intensity regions, the first stimulus is overestimated relative to the second (a positive TOE), and when they are in high regions, the first stimulus is underestimated relative to the second (a negative TOE). This particular perceptual weighting would be used by subjects to increase the detectability of level changes between the stimuli along an experiment where the level of the standard is roved or varied (for more details, see Patching et al. 2012). Thus, when the TOE is defined as the level difference between S_2 and S_1 at equal loudness, with short ISIs, negative TOEs are generally observed in high intensity regions and positive TOEs in low intensity regions (Hellström, 2003). In both Experiment 2A and Experiment 2B, the level differences reported in the different configurations are thus undoubtedly affected by this phenomenon, which explains why decreasing profiles of the level differences as a function of the maximum level of the ramps were observed (see Figure 1.3 and 1.4, lower panels).

1.3.2.2 Loudness asymmetries

Since the initial aim of this study was concerned with the potential loudness asymmetries between rising and falling tones and how they rely on the pair configuration and the level of presentation, a specific treatment of the data is proposed afterwards to compensate the TOEs. The only way to obtain data free of TOE-effects was to average the results in the four different configurations of each experiment. Thus, it still allowed assessing the magnitude of the loudness asymmetries for the different levels investigated in the two experiments, but a specific evaluation of the pair order influence on the asymmetry was not possible. The mean loudness asymmetries were then calculated for each intensity region by averaging the signed level difference between the second and the first stimulus of a pair matched in loudness in the different configurations, as follows:

In Experiment 2A:

$$Asymmetry(dB) = \frac{(a + b - c - d)}{4}$$

In Experiment 2B:

$$Asymmetry(dB) = \frac{(e - f + g - h)}{2}$$

The results of these averaged asymmetries are presented in Figure 1.5. The assessment of the influence of the type of comparison stimulus (TypeComp) and the level of the ramps was made by comparing these mean asymmetries in a mixed-ANOVA [1-between (TypeComp) and 1-within (Level) factors]. No influence of the type of comparison tone was found, as revealed by a non-significant difference between the two groups ($p > .05$).

However, the analysis showed a marginally significant effect of the Level $[F(4,80) = 2.697, p < .05, \tilde{\epsilon} = .49, \eta_g^2 = .06]$. This effect was caused by the significantly smaller size of the asymmetry at 80 dB SPL compared to other lower levels. Indeed, a post-hoc mixed-ANOVA conducted on the data of Experiment 2A and 2B without the estimates for the ramps at 80 dB level shows that the effect of the Level becomes non-significant. In addition, as in the first experiment, additional one-tailed t-tests were conducted on the asymmetries averaged over levels to demonstrate that these values were overall significantly higher than zero in both Experiment 2A [Mean = 3.4 dB (SD= 2.3); $t(9)4.826, p < .001$] and Experiment 2B [Mean = 3.2 dB (SD = 2.0); $t(11) = 5.521, p < .001$].

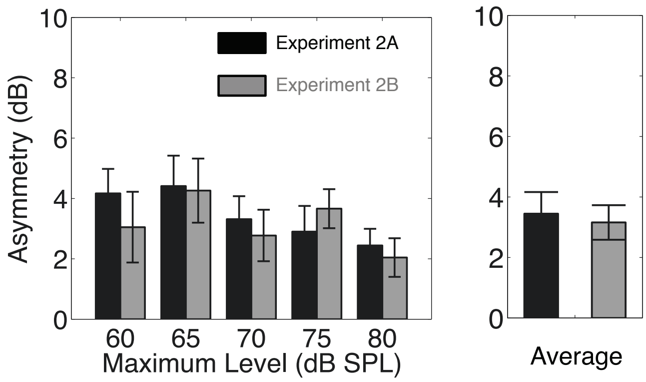

Figure 1.5: Loudness asymmetries measured when rising and falling ramps were compared within the same pair (Experiment 2A, black bars), or compared against constant-intensity tones (Experiment 2B, grey bars), presented as a function of their maximum level (left panel) and averaged over the different intensity levels (right panel). Error bars correspond to SEM.

1.3.3 Discussion

Asymmetries between 2-s rising and falling tones were assessed from direct loudness matching tasks in Experiment 2A and from indirect loudness matching tasks in Experiment 2B, where constant-intensity stimuli were employed as comparison tones. Strong TOEs were found in both tasks and were compensated by averaging the different configurations. Overall, the results show that the magnitude of these asymmetries were similar between the two experiments, providing clear evidence against any significant influence of the type of comparison stimulus on the asymmetry. If the asymmetry was (partly) due to some particular judgment bias, different values would likely have been found when the ramps were matched together compared to when the ramps were matched with constant-intensity tones. Because the data do not show such differences, this supports the results of Experiment 1, suggesting that rising and falling-intensity sounds contain an intrinsic information that lead to their asymmetry. As in Experiment 1, similar asymmetries were found between rising and falling ramps at lower levels (with peak levels between 60 and 75 dB SPL), but significantly smaller asymmetries were found between rising and falling ramps culminating at 80 dB SPL. The origins of this decrease at high levels remains to be examined.

Furthermore, one might wonder whether Induced Loudness Reduction (ILR) could have impacted the present results. In the configuration (g) of Experiment 2B only, that is where the falling ramp is preceded by a constant tone (see Figure 1.4), the conditions required for ILR to occur are met: (1) the two tones are identical in frequency, (2) the first tone is 10 to 20 dB louder than the second tone and (3) the two tones are separated by several hundred milliseconds (Epstein 2007; Epstein 2013; Oberfeld 2007). As a result, the loudness of the falling ramp (starting with at a low level) might have potentially been reduced by the preceding constant tone. If this were true, one direct consequence of this effect would be a greater asymmetry when the constant comparison tone was the second stimulus compared to when it was the first stimulus, especially at high levels. To test this, a comparison of the asymmetries between the conditions where the constant tone was presented first (calculated using the index A_1=(h) – (g)) or second (using A_2=(e) – (f)) was made. No significant effect or interactions with the level of the ramp were found (all Ps $>$ .05), suggesting that there was no particular influence of ILR on the present results.

1.4 General Discussion

In this study, we measured the global loudness of 2-s rising and falling-intensity tones using different measurement methods and contexts of presentation. The influence of “global” and “local” contexts on the perceptual asymmetry was examined in the first experiment (Experiment 1) using a one-interval paradigm. We found that “global context” of presentation had no significant influence on the loudness asymmetry, and that “local context” (i.e. prior stimuli direction) only had a small effect. This shows that no particular strong “representation mode” is involved in the processing of long stimuli, which contrasts with previous results obtained for short duration stimuli where a particular cognitive mechanism – the parsing of sounds into “source” and “reverberation” – was found to cancel the asymmetry under certain conditions (Stecker and Hafter, 2000). Based on two-interval loudness matching paradigms, Experiment 2A and 2B extended Experiment 1’s findings. We found no influence of the type of comparison stimulus on the magnitude of the asymmetry. In addition, over all three experiments, we measured asymmetries for ramps at different intensity regions, culminating between 60 and 90 dB SPL. The size of the asymmetry was virtually unaffected by the level of presentation, except for the loudest ramps (at 90 dB SPL in Experiment 1 and at 80 dB SPL in Experiment 2), where a significant decrease was observed. Whether this peculiarity for high-level stimuli relies on any contextual edge effect or has sensory origins remains to be determined in a future study. Overall, the mean size of the asymmetry was similar when the loudness was measured by AME (Experiment 1), when rising and falling sounds were compared one to another (Experiment 2A) and compared against constant-intensity tones (Experiment 2B); it was quantified and found to fall within the range 3-5 dB. Durlach, Braida and colleagues (e.g., see Braida et al. 1984) demonstrated that in one-interval paradigms such as in magnitude estimation tasks, subjects primarily judge the stimulus intensity by operating in a “context-coding” mode (that is by comparing the sensory representation to the stimulus context), whereas in two-interval paradigms like in roving-level discrimination tasks, an optimum combination of the “context-coding” mode and the “trace-mode” (the latter based on the direct sensory representation of the stimulus) is used. Taken in this framework, our results suggest that the perceptual asymmetry between 2-s asymmetrical stimuli is not constrained by any particular coding mode. Since two different measurement methods and various contexts of stimuli presentation yielded, on average, very similar values, our study provides strong evidence for the robustness of the mechanism responsible for the asymmetry in global loudness between 2-s symmetrical stimuli.

1.4.1 Predictions from current loudness models

It is interesting to compare the magnitude of this asymmetry with predictions from two loudness models applicable to time-varying stimuli (DLM, Chalupper and Fastl 2002; TVL, Glasberg and Moore 2002). Peak values of both short-term (proposed in DLM and TVL) and long-term (in TVL only) loudness time-series predicted by these models were used to evaluate the global loudness of our stimuli. Then, loudness ratios were used to compute the asymmetries at four different intensity regions ([45-60 dB SPL], [55-70 dB SPL], [65-80 dB SPL] and [75-90 dB SPL]) that cover the ranges examined in the present experiments. Asymmetries obtained using these short-term and long-term loudness predictions are reported in Table 1.1. Asymmetries evaluated with DLM and TVL short-term loudness (STL) lie between 0.4 dB and 0.5 dB, which predict very small or almost no difference in loudness between rising and falling tones. Asymmetries predicted with TVL long-term loudness (LTL) are slightly greater (around 1.3 dB) but still fall well below the loudness differences assessed psychophysically. These asymmetries can be seen with TVL time-series on Figure 1.6 where both STL and LTL of a rising-intensity tone reach higher peaks than for a falling-intensity tone. Note that these predictions are the consequence of the temporal integration stages employed in the models. Indeed, instantaneous loudness patterns are symmetrical for the two profiles. For example, Glasberg and Moore model (2002) uses two successive Automatic Gain Control (AGC) circuits to convert the instantaneous loudness to a short-term loudness pattern first, and then to a long-term loudness pattern. Moreover, while smaller perceptual asymmetries were found for the ramps in the highest regions of our experiments, the models give virtually identical predictions for the different levels (cf. Tab. 1.1).

	[45-60 dB]	[55-70]	[65-80]	[75-90]	**Mean**
DLM (dB)	0.46	0.46	0.43	0.43	**0.44**
STL (dB)	0.49	0.49	0.49	0.52	**0.5**
LTL (dB)	1.24	1.24	1.25	1.3	**1.26**

Table 1.1: Mean loudness asymmetries (in dB) calculated with DLM (Chalupper and Fastl, 2002) and TVL (STL and LTL; Glasberg and Moore 2002) models applicable to non-stationary sounds, for ramps with 15-dB dynamics at different intensity regions. These asymmetries are based on the ratio of loudness maxima obtained between rising and falling-intensity tones. The right column indicates the asymmetry averaged over these regions.

In these models, the time constants employed at the first integration stage (i.e. to compute the STL) were based only on data from psychophysical experiments on temporal masking and temporal integration of loudness (Glasberg and Moore, 2002). Although some studies showed that these constants surprisingly lead to fairly accurate predictions of loudness asymmetries between short duration rising and falling stimuli (Rennies et al. 2010; Ries et al. 2008), our results demonstrate that they are not sufficient to predict loudness asymmetries that occur between few seconds stimuli. In the view of our data, though it still underestimates the magnitude of the effect, the LTL of Glasberg and Moore's model seems to be the best candidate to evaluate these asymmetries. Increasing the time constants at the latter integration stage in LTL would be the only option to improve the predictions of the asymmetries. However, the downside would be a poorer loudness prediction for amplitude-modulated sounds, for which LTL was initially designed. Future investigations are thus needed to create new and appropriate descriptors that correctly predicts the loudness asymmetry between long-duration sounds, for example by using top-down modulated integration windows, as proposed recently to predict the pitch of short asymmetrical stimuli (Tabas et al., 2014).

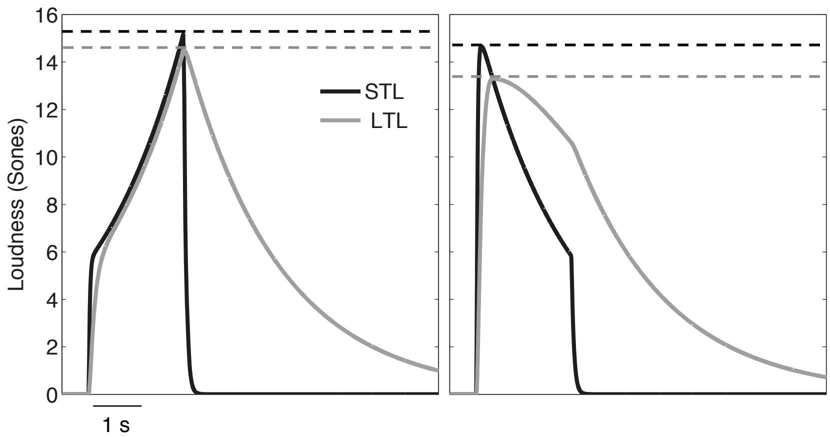

Figure 1.6: Loudness (in sones) of two symmetrical time-varying tones predicted by TVL loudness model (Glasberg and Moore, 2002). Left panel shows short-term loudness (STL, black lines) and long-term loudness (LTL, grey lines) of a 2-s [65-80 dB SPL] rising-intensity 1-kHz tone (with 10-ms linear rise and fall times) predicted by the TVL model. Right panel shows predictions for the time-reversed falling-intensity tone. Asymmetries between rising and falling-intensity loudness patterns can be observed considering either STL or LTL maxima (as indicated by dashed lines). A greater asymmetry is found with LTL compared to STL, due to the use of another stage of temporal integration.

1.4.2 Conclusions and perspectives

The *global loudness* judgment reflects a listener's overall evaluation of "a sound over its entire duration" (Ponsot et al., 2013). While the peak value of the long-term loudness proposed in Glasberg and Moore's model (2002) is typically used to evaluate the global loudness of time-varying sounds longer than 1 s (e.g. Rennies et al. 2013b; Rennies et al. 2013a; Ries et al. 2008), our results clearly point out the limits of that model. The fact that the loudness asymmetry between 2-s rising and falling-intensity sounds was consistent across different contexts of stimulus presentation and different measurement methods suggests that a strong underlying mechanism is involved. Further work is needed to determine the causes of the effect. In particular, the extent to which it relies on the "bias for rising tones" specified by Neuhoff

(1998) or on the “end level bias” specified by Susini et al. (2010) remains to be specifically addressed. Loudness models would certainly benefit from future research on the perceptual and cognitive processes involved into such asymmetries in order to improve the *global loudness* prediction of long non-stationary sounds.

Acknowledgements

The authors are especially grateful to Åke Hellström for his help and his expertise on the results of the second experiment. Indiana Wollman and Jean-Julien Aucouturier are particularly acknowledged for their comments upon this manuscript. Finally, we would like to thank Massimo Grassi and an anonymous reviewer, who provided fruitful comments and helpful suggestions. This work was supported by the project LoudNat funded by the French National Research Agency (ANR-11-BS09-016-01).

Partial synthesis

The two main goals of this study were (1) to explore to which extent the loudness asymmetry between tones with rising and falling-intensity profiles relies on contextual and methodological factors, and (2) to examine how well current loudness models predict it.

(1) **The asymmetry is robust to different methods and contexts. On average, rising tones are perceived about 4 dB louder than falling tones**. This quantification was obtained with the same stimuli (i.e., 2-s, 1 kHz ramps with 15-dB dynamics) on average across different experimental conditions. In Exp. 1, the global loudness of the stimuli was evaluated using absolute magnitude estimation procedures. The participants were divided into two groups, each being assigned a specific context of presentation; one group was presented with rising and falling stimuli that were interleaved in the same block ("mixed context"), the other group was presented with rising and falling stimuli that were separated in two different blocks ("separate context"). In Exp. 2A and Exp. 2B, the global loudness of these sounds was inferred from loudness-matching tasks with adaptive procedures. In Exp. 2A, rising and falling sounds were compared to each other whereas in Exp. 2B, the stimuli were compared against constant-intensity tones. The asymmetries derived from the estimates collected in these experiments were overall very similar, falling between 3.1 dB and 5.3 dB. This showed that both the task (absolute loudness judgment and relative judgment) and the context of stimuli presentation (mixed vs. separate and direct vs. indirect) had no or only little influence on global loudness asymmetries. Taken together, these results suggest that the ramps contain the intrinsic information responsible for the asymmetry.

A small but significant reduction of the asymmetry has been observed in the highest regions examined in the study, i.e. at [75-90 dB SPL] (Exp. 1) or at [65-80 dB SPL] (Exp. 2A and Exp. 2B). Moreover, a small but significant "local" contextual effect was also obtained in the Exp. 1: when a rising tone just preceded a falling tone, the number given to the latter was such that the asymmetry was overall significantly reduced (see Section 1.2.2.3). We had no explanation for these effects.

The fact that the asymmetry was on average about 4 dB does not hold at an individual level, as shown by the individual asymmetries values presented in Appendix A. Interindividual variability regarding the size of the asymmetry is quite large and the distributions of individual asymmetries derived in these experiments follow normal family distributions with standard deviations having about the same size as the effect, i.e. around 4 dB. However, with the present study, it is not possible to determine whether this variability reflects perceptual or evaluation strategy differences. Indeed, we deliberately recruited different subjects for the different experiments in order to avoid potential "habituation" effects or "transfer" effects (e.g., see Poulton and Freeman 1966; Poulton 1982). The asymmetry was thus only measured once for each listener in a given context using a particular method. Investigating the origins of interindividual differences constitutes a very interesting direction for future studies.

(2) **The asymmetry cannot be accounted for by current loudness models**. We demonstrated that neither the maximum of short-term loudness (STL) patterns nor the maximum of long-term loudness (LTL) patterns predicted by Glasberg and Moore loudness model (2002) could account for the asymmetries observed with simple rising and falling intensity sounds lasting 2 seconds. Although these indicators predict that rising sounds are perceived louder than falling tones, they significantly underestimate the size of this asymmetry. The STL maxima predict asymmetries lower than 0.5 dB and the LTL maxima predict asymmetries around 1.3 dB. Increasing the time constants to compute the LTL patterns would improve global loudness predictions for these 2-s rising and falling sounds but it would then fail to predict the global loudness of amplitude-modulated sounds (for which these

constants had been initially designed).

Chapter 2

Influences of spectral structure and intensity-region on the loudness asymmetry

Emmanuel Ponsot, Sabine Meunier, Abbes Kacem, Jacques Chatron and Patrick Susini (in press). Are rising sounds always louder? Influences of spectral structure and intensity-region on loudness sensitivity to intensity-change direction. *Acta Acustica united with Acustica.*

Preface

Numerous psychophysical studies on *loudness change* evaluations of rising and falling-intensity stimuli showed that the spectral structure of the sounds significantly influence the size of their resulting perceptual asymmetries. In particular, it was shown that tonal stimuli (e.g. pure tones, vowel sounds, musical chords) lead to greater *loudness change* asymmetries than non-tonal broadband noises (e.g., Neuhoff 1998; Olsen et al. 2010). Besides, in the neuroscience domain, this finding was supported by the fact that different neural regions are associated to the processing of tonal vs. non-tonal rising stimuli (e.g., Maier and Ghazanfar 2007; Bach et al. 2008).

In the present chapter, we investigate to which extent the spectral structure effect holds true in the context of *global loudness* asymmetries. To this end, global loudness judgments of rising and falling sounds composed of various frequencies as well as broadband noises were collected. In the previous chapter, we observed smaller asymmetries for pure tones in high intensity-regions (with maximum levels equal to or greater than 80 dB SPL) compared to other lower regions. This effect of the intensity-region is further investigated in the present chapter using both tones and noises.

The influences of the spectral structure and the intensity-region of the stimuli were scrutinized in two psychophysical experiments employing a magnitude estimation procedure and a loudness matching procedure, respectively. Magnitude estimation procedures being a fast way of collecting loudness estimates, the first experiment tested many different conditions (7 different spectral structures and 5 intensity-regions). On the other hand, adaptive loudness-matching procedures are longer procedures but provide more accurate measures of loudness differences. The second experiment, based on an adaptive procedure, revisited specific conditions of the first experiment in order to further investigate the conditions that led to the most extreme values of the asymmetries.

The manuscript presented in this chapter has been accepted for publication in a special issue on *Loudness* that will appear in *Acta Acustica united with Acustica*:

Ponsot, E., Meunier, S., Kacem, A., Chatron, J., Susini, P. (in press). Are rising sounds always louder? Influences of spectral structure and intensity-region on loudness sensitivity to intensity-change direction. *Acta Acustica united with Acustica.*

Appendix B presents additional data (not shown in the paper) collected in a complementary experiment that followed the second experiment presented in this chapter. This complementary experiment specifically examined whether the differences revealed between tones and noises with respect to the size of the asymmetry (the spectral structure effect) might be fully explained by differences of spectral bandwidth. Further correlation analyses of the data of the second experiment of this chapter were also carried out after the publication of the article, and are presented in Appendix C. Appendices B and C are discussed together with the main findings of the paper in the synthesis section of this chapter.

Abstract: In a previous study, a robust asymmetry in global loudness was observed between rising and falling-intensity 1-kHz tones, pointing out the existence of a mechanism specifically sensitive to sound intensity direction [Ponsot et al., *Attention Perception, & Psychophysics*, 77(3), 907-920 (2015)]. The properties of this "direction-sensitive" mechanism are further explored in the present study, where its dependence on two stimuli characteristics, the spectral content and the intensity-region is examined. In a first experiment, the global loudness of rising and falling-intensity sounds varying over 15-dB ranges was assessed in a magnitude estimation task. The ramps had various spectral contents (pure tones from 250 Hz to 8 kHz and broadband noises) and were presented in different intensity-regions (from 50-65 dB SPL to 70-85 dB SPL). Significant asymmetries were observed across the different frequencies but not for broadband noises. In addition, a significant interaction between the direction of change and the intensity-region was observed for tones and noises. This latter effect was specifically addressed in a second experiment using an adaptive loudness-matching procedure, in which asymmetries were inferred from pairwise comparisons, both for 1-kHz tones and for white noises presented in mid and high intensity-regions. As in Exp. 1, greater asymmetries were observed for tones compared to noises; however significant asymmetries were found for noises as well. Furthermore, for both tones and noises, the size of the asymmetries was significantly decreased with the intensity-region when the sound pairs were composed of a rising followed by a falling stimulus. These results are discussed in the light of recent physiological and neuroscience studies showing that spectrally structured looming sounds are treated specially by the brain.

2.1 Introduction

In the past, a few studies observed that, when listeners are required to evaluate the *global loudness* of time-varying sounds, defined as the overall loudness of a sound over its entire duration, they primarily focus on the loudest portions of the stimuli to construct their judgments (e.g., Kuwano and Namba 1985; Susini et al. 2002). In various recent psychophysical experiments on sound source discrimination or identification where psychophysical reverse correlation methods were employed, a somewhat related effect was found: the so-called "level-dominance" effect, reflecting the fact that listeners always tend to assign much greater weights to the loudest elements of the stimuli (e.g., to loudest frequency components in complex stimuli), even though these elements provide less information than others with respect to the required task (e.g., Turner and Berg 2007; Lutfi et al. 2008; Richards et al. 2013). Using reverse correlation methods to study *global loudness* judgments, we specifically demonstrated that for 2-s, 1-kHz tones with rising- or falling-intensity profiles, the temporal weighting of loudness was also entirely constrained by this "level-dominance" effect (Ponsot et al., 2013). However, although listeners' attention was exclusively focused on the loudest portions of the stimuli similarly for the two profiles, rising-intensity tones were consistently judged louder than falling-intensity tones (Ponsot et al., 2013). The robustness of this asymmetry in "global loudness" was later confirmed using various measurement methods and experimental contexts, but the origins of the effect still remain unknown (Ponsot et al., 2015). Therefore, the present study was conducted to better understand why and how loudness processing of time-varying sounds depends on the direction of intensity-change. In particular, we examined how this "direction-sensitive" mechanism is influenced by the spectral characteristics and the intensity-region of the stimuli.

Differences in the perception of looming and receding signals were already reported in various tasks and, interestingly, pure tones and sounds with harmonic spectral structures were found to elicit much larger differences than sounds having random spectral structure, such as white noises. One of the earliest reports of perceptual asymmetries between rising and falling-intensity sounds of a few seconds was made by Neuhoff in the context of *loudness change*, with rising sounds being judged as "changing more in loudness" than falling sounds (Neuhoff, 1998). This effect was shown to be larger for violin notes, chords or vowels (i.e., with harmonic structures) than for broadband noises[1] (Neuhoff 1998; Olsen et al. 2010; Olsen and Stevens 2010). Psychophysical studies investigating duration perception of auditory stimuli with rising or falling intensity profiles also yielded similar observations. Rising-intensity sounds are perceived as significantly longer than falling-intensity sounds (e.g., Grassi and Darwin 2006; Meunier et al. 2014) and this asymmetry was found to be greater with tones than with broadband noises (e.g., Schlauch et al. 2001). Recent imaging studies conducted on both human and non-human primates corroborated the aforementioned studies. In particular, it was shown that rising tones but not rising noises recruit specific neural networks and induce different behavioral asymmetries, as compared to their equivalent falling versions (Ghazanfar et al. 2002; Seifritz et al. 2002; Hall and Moore 2003; Maier and Ghazanfar 2007; Bach et al. 2008; for further details, see Ponsot et al. 2015). In addition, several auditory psychophysical tasks revealed that when the stimuli comprise a harmonic spectral structure, humans tend to underestimate time-to-contact (Rosenblum et al. 1993; Neuhoff 2001) and react faster to rising-intensity sounds compared to falling-intensity sounds (Bach et al. 2008; Bach et al. 2009; Ho et al. 2013). For example, reaction times to tone targets measured after the presentation of tonal "looming" sounds (i.e., rising in intensity) were found to be significantly faster than after the presentation of tonal "re-

[1]One should note that these results concern *direct* retrospective judgment measurements –i.e., the listeners were asked to produce overall evaluations of l*oudness change* at the end of the stimuli − which is the type of judgment investigated in the present study. Opposite findings were found where *indirect* loudness change measurement were used − i.e., when defined as the ratio of separate loudness "snapshots" estimates given by listeners at the start and the end of the stimuli − (e.g., Canévet and Scharf 1990; Canévet et al. 2003; Teghtsoonian et al. 2005); an effect known as "decruitment" (for a review, see Olsen 2014). However, these two types of judgments were shown to reflect different perceptual processes. Since the present study deals with the mechanisms underlying retrospective *global loudness* judgments, this paper will exclusively be focused on the asymmetries observed in various *direct* retrospective judgments.

ceding" sounds (i.e., falling in intensity) (Bach et al., 2009). This asymmetry was no longer observed for white noises. In a visual orientation discrimination task, a significant decrease in reaction time was observed when spatially congruent tonal auditory stimuli were presented at the same time as visual stimuli (Leo et al., 2011). In that study, the orientation of a visual Gabor patch presented on the left (right) part of the screen was detected faster when combined with a rising tonal auditory stimulus presented on the loudspeaker on the same side, i.e. on the left (right), than when combined with a falling tonal auditory stimulus. This facilitation of visual orientation discrimination was no longer observed when rising and falling-intensity stimuli made of white noises were employed. Moreover, concerning perceived duration, Grassi & Pavan (2012) showed that the "looming" / "receding" asymmetries in perceived duration might be specific to the auditory modality, since they were significantly reduced using audio-visual stimuli and even totally removed with purely visual stimuli.

All together, these studies indicate important differences in the processing of "looming" and "receding" auditory stimuli, not only regarding loudness. They suggest that a high-level neural mechanism specifically dedicated to "looming" spectrally structured sounds could be responsible for these different outcomes (e.g., Neuhoff 1998; Seifritz et al. 2002; Maier and Ghazanfar 2007). The effect is strong enough to consistently affect various behavioral responses, to facilitate audio-visual integration (e.g. Leo et al. 2011; Maier et al. 2004; Maier et al. 2008; Cappe et al. 2009; Cappe et al. 2012; Tyll et al. 2013) and even to produce crossmodal interactions (Liu et al., 2011) and remapping (Romei et al. 2009; Sutherland et al. 2014).

Another important factor that has to be considered when examining the perception of rising- and falling-intensity stimuli concerns their "intensity-region" (i.e., the region of intensity covered between the start and the end of the ramp). Few studies on direct *loudness change* first showed that the asymmetry between the two profiles increased with intensity-region (e.g., Neuhoff 1998; Olsen et al. 2010; Pastore and Flint 2011). However, it has now been demonstrated that this effect was mediated by an "end-level" recency mechanism biasing participant's loudness change estimates (Olsen and

Stevens 2010; Susini et al. 2010). On the other hand, a previous study on *global loudness* showed smaller asymmetries in highest regions (e.g., culminating at 80-90 dB SPL) compared to other lower regions (Ponsot et al., 2015). As compared to the spectral structure, this factor has been paid less attention so far. The causes for such potential intensity-region dependent effects remained to be addressed more specifically in the context of *global loudness* asymmetries.

As stated above, the aim of the present study was to examine the influence of two factors (spectral structure and intensity-region) on *global loudness* differences between rising and falling-intensity sounds; global loudness asymmetries have only been reported for 1-kHz tones so far (Ponsot et al., 2015). The main purpose of Exp. 1 was thus to collect quantitative estimates of these differences for other spectral contents in order to test whether the influence of the spectral structure could also be observed in the case of global loudness judgments. Pure tones with different frequencies (between 250 Hz and 8 kHz) were evaluated as well as two broadband noises (pink and white noises) in order to test if the asymmetry was larger for pure tones than for noises, whatever the frequency and the spectral envelop of the noises. We also wanted to investigate whether the magnitude of the asymmetries for tones would be reduced at higher frequencies as, for example the AIM model predicts (e.g., Patterson et al. 1995). The use of stimuli having various spectral contents was employed in previous studies examining perceptual asymmetries between short asymmetrical stimuli (<200 ms), both for loudness (Stecker and Hafter, 2000) and perceived duration (Schlauch et al., 2001). The main purpose of Exp. 2 was to explore further the effect of the intensity-region observed in Exp. 1. Based on the aforementioned studies, greater asymmetries were expected for pure tones compared to broadband noises, as well as smaller asymmetries in highest intensity-regions compared to other lower intensity-regions. The potential origins of these asymmetries are discussed with respect to the results obtained in the two experiments.

2.2 Experiment 1

In this experiment, an Absolute Magnitude Estimation (AME) procedure was used to measure the global loudness of rising and falling sounds in various conditions. Pure tones covering a wide range of the audible spectrum as well as broadband noises were used. The sounds were presented from mid up to high intensity-regions (50-65 dB SPL to 70-85 dB SPL).

2.2.1 Materials and method

2.2.1.1 Participants

Fifteen volunteer participants (5 women and 10 men; age 22-50 years, Mean = 28) took part in the experiment. They were not paid for their participation. All listeners had audiometric thresholds lower than 20 dB HL at octave frequencies of 125-8000 Hz. The participants were all naïve with respect to the hypotheses under test; only two of them (members of the lab) had prior experience in loudness rating experiments.

2.2.1.2 Stimuli

Stimuli were rising and falling-intensity ramps lasting 2 s, including 20-ms linear (on amplitude) rise and fall times at onset and offset. The main portion of these intensity-ramps lasting 1960-ms (onset and offset excluded) covered a region of 15-dB (i.e., the ramps had a linear variation – up- or downward – of 15 dB). They were presented at five different intensity-regions: 50-65, 55-70, 60-75, 65-80 and 70-85 dB SPL. Different spectral contents were used: pure tones at five frequencies (250Hz, 500Hz, 1kHz, 4kHz and 8kHz) and two broadband noises (white noise, pink noise).

2.2.1.3 Apparatus

The stimuli were generated at a sampling rate of 44.1 kHz with 16-bit resolution using Max/MSP Software. Sounds were converted using a Focusrite Saffire PRO24 soundcard and were presented diotically through headphones (Sennheiser HDA 200). The levels were calibrated using a Brüel & Kjær artificial ear (type 4153) coupled with the mounting plate provided for circumaural headphones with no free-field equalization. Each participant was tested in a double-walled sound-insulated booth at LMA, Marseille.

2.2.1.4 Procedure

The Absolute Magnitude Estimation (AME) procedure was employed, based on the instructions of Hellman (Hellman, 1982). No standard was given to the participants, whose task was to give a number proportional to the global loudness of each sound regardless of the numbers assigned to previous stimuli (Ponsot et al. 2013; Ponsot et al. 2015; Susini et al. 2010). The ramps were presented in 14 different blocks (2 directions $\times$ 7 spectral contents), depending on their direction of intensity (rising/falling) and spectral content. The order of presentation of these blocks was randomized. Each was made of 30 trials (5 intensity-regions $\times$ 6 repetitions) presented in random order, i.e., in a given block the stimuli simply differed with respect to their intensity-region (for example, a block was made of 500-Hz falling ramps presented 6 times at each of the five intensity-regions). The whole experiment lasted approximately 90 minutes for each participant, including several pauses.

2.2.2 Results

Mean global loudness estimates, taken as the geometric means of all individual estimates, are presented in Figure 2.1. Statistical analyses were conducted using R (R Core Team, 2015). We performed repeated-measures ANOVAs (rmANOVAs) with univariate approaches for the main analyses, using Bonferonni corrections where appropriate. Unless otherwise specified, all the tests were two-tailed and used a probability level of .05 to test for significance. Effect sizes are reported using partial eta-squared η_p^2. For this experiment, an overall $2 \times 5 \times 7$ (Direction, Intensity-Region, Spectral Content) rmANOVA was conducted, as well as planned contrasts to examine the expected tones/noise differences regarding the effect of the direction. These analyses were performed on linearized data, computed by taking the logarithm of the raw estimates (see Ponsot et al. 2015).

The main significant effect revealed by the analysis was, as expected, the

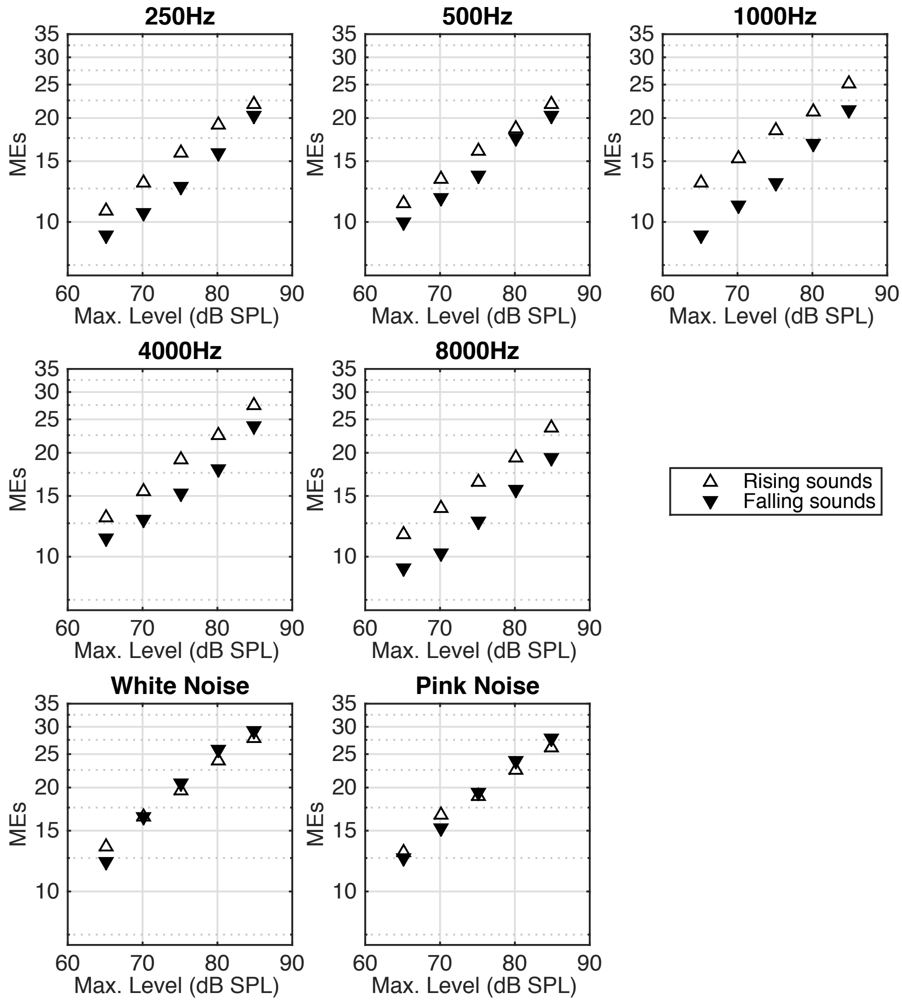

Figure 2.1: Geometric mean global loudness ratings for each frequency-spectrum (pure tones from 250Hz to 8kHz; white noise; pink noise), plotted as a function of the upper bound of the intensity-region at which the ramps were presented (i.e., these regions ranged from 50-65 dB SPL to 70-85 dB SPL but are referred to using their maximum only). Upward (unfilled) triangles correspond to loudness estimates attributed to rising-intensity sounds; downward (filled) triangles correspond to loudness estimates attributed to falling-intensity sounds.

intensity-region $[F(4,56) = 117.72, p < .001, \eta_p^2 = .894]$, meaning that the participants primarily considered the energy of the stimuli in their estimates. As it can be observed on Fig. 2.1 in most of the conditions, greater estimates were observed for rising tones compared to falling tones; this was supported by a significant effect of the Direction $[F(1,14) = 25.350, p < .001, \eta_p^2 = .644]$. However, the size of these asymmetries was found to depend on the type of stimuli (Spectral content), as shown by a significant Direction × Spectral Content interaction $[F(6,84) = 4.901, p < .001, \eta_p^2 = .259]$. Planned comparison showed that the magnitude of the asymmetries was significantly greater for tones (all frequencies combined) compared to noises (pink and white noises combined) $[F(1,14) = 23.217, p < .001]$. Although a look at Fig. 2.1 suggests that the overall size of these asymmetries was marginally greater for 1-kHz and 8-kHz tones compared to other frequencies, a later *post-hoc* rmANOVA conducted on the estimates given to pure tones only showed no significant differences across tone frequencies ($p > .05$). Similarly, no significant differences between white and pink noises were found ($p > .05$). Finally, a significant Direction × Intensity-Region interaction was found $[F(4,56) = 6.202, p < .001, \eta_p^2 = .307]$.

2.2.3 Discussion

Overall, greater global loudness estimates were obtained for rising tones as compared to falling tones when pure tones were used but not when broadband noises stimuli were used (significant Direction × Spectral Content interaction). The differences between rising and falling tones estimates were significant and similar across frequencies, and no significant differences were found for white noises and pink noises. These results support the hypothesis proposed in the introduction that the type of sound does play a role in the formation of these asymmetries, as it has previously been shown for *loudness change* judgments (e.g., Neuhoff 1998).

A significant Direction × Intensity-Region interaction was also found. As mentioned in the introduction, this interaction was already obtained in a previous study with 1-kHz tones: using either loudness matching or magnitude estimation procedures, loudness asymmetries for 1-kHz tones were significantly smaller in the highest regions examined, in 65-80 dB SPL and 75-90 dB SPL regions respectively (see Ponsot et al. 2015). In the case of broadband noises, the estimates given to rising tones appeared to be slightly greater than the estimates given to falling tones in lowest intensity-regions, whereas the opposite was observed in highest intensity-regions (however, post-hoc analyses between these specific conditions did not reveal any significant effect). The overall significant Direction × Intensity-Region interaction observed in the main rmANOVA of the present experiment reflects a common deviation in loudness judgments, where the effect of intensity direction (rising/falling) is mediated by the intensity-region in which the stimulus is presented. Since this intensity-dependent effect was also observed for 1-kHz tones in a loudness matching experiment (Ponsot et al., 2015), it can reasonably be assumed that it was not induced by a contextual effect related to the magnitude estimation procedure presently employed.

Two conclusions can be drawn from the present results. First, the presence of loudness asymmetries between rising and falling ramps depends on the tonal characteristic of the sounds. Second, the Direction × Intensity-Region interaction observed with pure tones at different frequencies and with broadband noises suggests that a specific intensity-dependent mechanism could be involved in the global loudness evaluation process of rising and falling sounds, regardless of their spectral content. This interaction was further explored in a second experiment with 1-kHz tones and white noises in particular. To assess the significance (or non-significance) of the asymmetries in specific intensity-regions, a loudness-matching procedure was employed where the asymmetries were directly measured in each region. Since 1-kHz tones and white noises presented at the same SPL are not perceived with the same loudness, these asymmetries were measured in different intensity-regions that allowed a control of their intrinsic loudness differences (the intensity-regions of the noises were translated with respect to a loudness equalization of their maximum intensities; for details, see below). From this design, the respective influences of spectral content and intensity-region on the results could be accurately distinguished (Gordon et al., 2013).

2.3 Experiment 2

In this experiment, loudness asymmetries were measured for 1-kHz tones and white noises using an adaptive loudness matching procedure. The ramps were presented in the lowest and highest intensity-regions used in Exp. 1 (i.e., 50-65 dB SPL and 70-85 dB SPL, respectively). The noises were presented in these two regions and in two additional regions judged as having the same loudness as the two regions at which the tones were presented. Thus, asymmetries were measured in both *non-equalized* and *equalized* regions (in loudness) to control the potential influence of loudness on the results.

2.3.1 Materials and method

2.3.1.1 Participants

Eighteen new volunteer participants (12 women and 6 men; age 19-31 years, Mean = 24) took part in this experiment. The participants gave their informed written consent prior to the experiment and were paid for their participation. All were naïve with respect to the hypotheses under test and had no prior experience with loudness rating experiments.

2.3.1.2 Stimuli

The stimuli were constructed in a similar fashion as in Exp. 1, using intensity-ramps varying positively (rising) or negatively (falling) over 15-dB during 1960 ms. Only 1-kHz pure tones and white noises were used. Since the aim was to examine more specifically the Direction × Intensity-Region interaction while controlling the potential influence of their intrinsic loudness differences, the tones were presented in the two most distant intensity-regions of the first experiment (i.e., 50-65 dB SPL and 70-85 dB SPL), whereas the white noises were presented in four intensity-regions: the same two intensity-regions as for the 1-kHz tones (i.e., 50-65 dB SPL and 70-85 dB SPL) and two additional regions producing the same maximum loudness as the tones presented in these regions (50-65 equal* and 70-85 equal*; an asterisk is used to indicate that it was specifically determined for each participant). These latter equal-loudness regions were obtained individually in a preliminary experiment (see section 2.3.1.4 below).

2.3.1.3 Apparatus

The stimuli were generated at a sampling rate of 44.1 kHz with 16-bit resolution using Matlab. Sounds were converted using a RME Fireface 800 soundcard and presented diotically through headphones (Sennheiser HD 250 Linear II). The level calibration process was similar to Exp. 1. There was no free-field equalization and the frequency characteristics of the headphones were not compensated. Participants were tested in a double-walled IAC sound-insulated booth at Ircam, Paris.

2.3.1.4 Preliminary experiment: Individual tone/noise loudness equalization

As stated above, the aim of this second experiment was to evaluate the influence of the intensity-region and the spectral content of the sounds on their asymmetries while controlling the potential influence of intrinsic loudness differences between tones and noises. To determine intensity-regions for noises that induced the same loudness as those for tones, preliminary measurements were conducted to evaluate for each participant the exact levels of white noise producing the same loudness levels as 65 and 85-dB SPL 1-kHz tones, that is, only obtained from the maximum intensity of each region. These individualized "equal-loudness" regions (also covering a 15-dB variation) were inferred from a tone/noise loudness-matching task conducted with each participant before the main experiment. The method employed was similar to the main experiment (see section 2.3.1.5 below). An adaptive 2I, 2AFC procedure with four interleaved tracks was used, comprising the two orders of presentation (tone-noise / noise-tone) and two repeated tracks for each order. Each participant matched in loudness 500-ms constant-intensity white noises with 500-ms constant-intensity 1-kHz tones. Tones were presented either at 65 dB SPL or at 85 dB SPL in distinct blocks. The levels of white noises producing the same loudness as 65 and 85-dB SPL 1-kHz tones (averaged across the four tracks) were saved and used individually in the main experiment to produce intensity-regions of equal-loudness.

Participant	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}	P_{18}
$D_{65}(dB)$	9.1	5.8	5.6	14.0	9.9	8.4	3.9	2.0	8.2	-3.3	7.6	0.3	0.5	3.5	7.7	8.3	12.0	2.0
$D_{85}(dB)$	9.2	8.7	11.2	9.3	6.8	5.1	3.2	3.2	10.4	-3.6	7.6	4.7	7.2	7.7	5.0	1.8	3.6	7.8

Table 2.1: Individual level differences between the 1-kHz tone and the white noise obtained in the preliminary loudness matching experiment; a positive value of x dB indicates that the noise was matched x dB lower than the tone to produce equal-loudness. These level differences (in dB) are specified for each participant (from P_1 to P_{18}) and for each level of the 1-kHz test tone, i.e. 65 dB SPL (D_{65}, first row) or 85 dB SPL (D_{85}, second row).

The averaged level differences between noises and tones that yielded equal-loudness judgments for each participant are presented in Table 2.1, as a function of the reference tone level (65 or 85 dB SPL). There was no significant difference between the results at 65 and 85 dB SPL [$t(17) = 0.184, p = 0.572$]. Large interindividual differences were found however, either when the tone was presented at 65 dB SPL (Mean = 5.8, SD = 4.5) or at 85 dB SPL (Mean = 6.1, SD = 3.6), covering a large range of almost 18 dB between the most extreme participants, as reported in studies on spectral loudness summation (e.g. Anweiler and Verhey 2006). This comforted our view that it was better to conduct this preliminary experiment in order to equalize tone/noise intensity-regions in an individual basis rather than referring to the mean difference inferred from any loudness model. It should be noted that participants were particularly consistent in their judgments, since the mean intraindividual differences between the four conditions (2 repetitions $\times$ 2 presentation orders) was remarkably small (Mean SD = 1.4 dB at 65 dB SPL; Mean SD =1.7 dB at 85 dB SPL), consistently with previous studies measuring equal-loudness points between noises of different bandwidths (Verhey and Kollmeier, 2002) and tones compared to noises with subcritical bandwidths (Hots et al., 2014a). One may note that one participant (P_{10}) surprisingly matched the noise in loudness around 3dB lower than the tone. A closer look at his/her results revealed a strong pair-order effect: the noise was matched clearly lower when it was the second sound of the pair (the mean loudness matching level averaged over the two intensities for this order was 0 dB, SD = 0.8 dB) than when the noise was the first sound of the pair (where it was -7.5 dB, SD = 1.6 dB). This pair-order effect was not observed with other participants. However, since his/her results were very similar between the two tracks with the same presentation order, this participant could not be considered as an outlier and was thus kept in subsequent analyses.

The individual loudness equalizations measured in this preliminary experiment were used to create two intensity-regions at which the noises were perceived as equally loud as the 1-kHz tones by each participant, i.e., the 50-65 equal* and 70-85 equal* regions.

2.3.1.5 Procedure

A loudness-matching task was employed to determine the loudness asymmetries between rising and falling ramps in the different conditions, as simply inferred from the difference between the measured point of subjective equal-loudness for rising and falling sounds, respectively. This procedure was already employed in a previous study (see Ponsot et al. 2015). The stimuli were presented in a 2I, 2AFC paradigm based on an interleaved adaptive procedure (Jesteadt 1980; Florentine et al. 1996) comprising four interleaved tracks with the two orders of presentation (rising-falling / falling-rising) and two types of varying sounds (rising varying / falling varying). The Inter-Stimulus Interval (ISI) was 500 ms. Listener's task was to indicate which sound of the pair was louder. The level of the variable tone, was adjusted according to a one-up, one-down rule (Levitt, 1971). An initial step size of 6 dB was chosen. The starting value was randomly selected either 5 dB higher or lower than the test stimulus. The step size was reduced by a ratio of 2 after every two reversals; after six reversals, it was then held constant. This procedure ended when 8 reversals were achieved for each track. The loudness matches were then defined as the average of the maximum level of the variable tone in the last two reversals of each track, thus tracking the 50% point (i.e., equal-loudness) on each underlying psychometric function (Levitt, 1971). The experiment was divided into 6 blocks; one block was used for each spectrum and for each intensity-region. A block was thus composed of the four interleaved tracks (2 orders and 2 types of stimuli varying). The blocks were presented in a random manner between participants. Since the order of the ramps could affect the subsequent loudness asymmetries, the loudness matches were derived for each spectral content, intensity-region and pair order, simply by averaging the loudness matches obtained in the two repetition tracks. In each condition, the asymmetry was calculated as the difference between the maximum level of the falling sound and the maximum level of the rising sound derived from the loudness matches (Ponsot et al., 2015). An experimental session (comprising the tone/noise preliminary loudness matching experiment) lasted approximately 1 hour for each participant.

2.3.2 Results

The loudness asymmetries obtained in the main part of the experiment in the different intensity-regions and orders are plotted in Figure 2.2, for tones (white bars) and noises (black bars). In the upper panel, the asymmetries are presented in *non-equalized* regions, i.e., using loudness matches obtained for both tones and noises in the 50-65 dB SPL and the 70-85 dB SPL intensity-regions. In the lower panel, the asymmetries are presented in *equalized* regions, i.e., using the matches for tones in the 50-65 dB SPL and the 70-85 dB SPL regions (from the upper panel) and the white noises presented in the 50-65 equal* and 70-85 equal* regions (obtained from the tone/noise loudness matching preliminary experiment; see above).

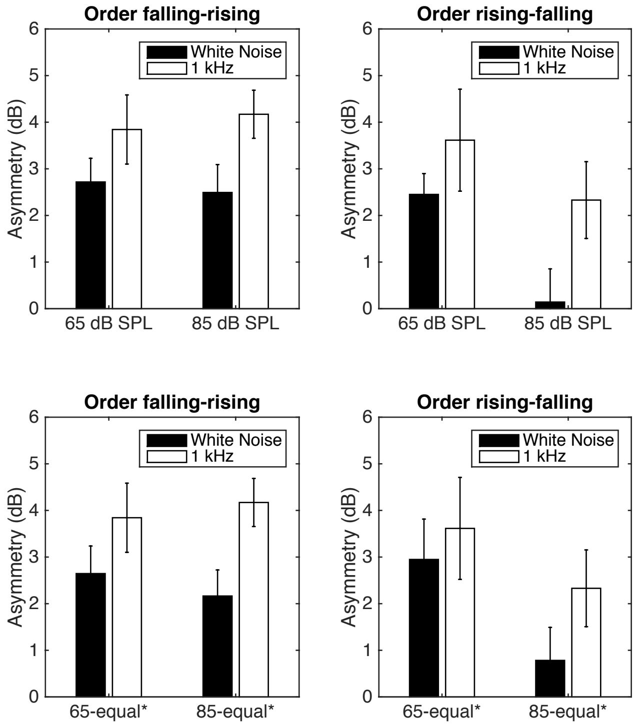

Figure 2.2: Loudness asymmetries derived from the loudness matches between rising and falling ramps, obtained for 1-kHz tones and for white noises. They correspond to the signed level differences between falling and rising tones at equal-loudness in each configuration (Asymmetry (dB) = $L_{falling}$ - L_{rising}); i.e., positive asymmetries indicate that the falling sound was matched higher in level than the rising sound to be perceived with equal-loudness. These asymmetries are presented as a function of the pair order (rising-falling or falling-rising, as indicated in the title of each plot) and the maximum of the intensity-region at which the ramps were presented. Upper panel: tones and noises varied in similar intensity regions (i.e., in 50-65 dB SPL and 70-85 dB SPL regions). Lower panel: tones and noises varied in intensity-regions equalized in loudness (i.e., in 50-65 equal* and 70-85 dB equal* regions; see the text for details). Error bars correspond to SEM.

Two rmANOVAs were conducted separately to compare these asymmetries in *non-equalized* regions and in *equalized* regions respectively, using 2 × 2 × 2 (Spectrum, Intensity-Region, Order) factorial designs. In *non-equalized* regions, the asymmetries were found to be significantly greater for 1-kHz tones than for white noises; Spectrum was significant $[F(1, 17) = 15.786, p < .001, \eta_p^2 = .481]$. The statistical analysis also showed a significant effect of Intensity-Region $[F(1, 17) = 6.362, p = .022, \eta_p^2 = .272]$ and a significant Intensity-Region × Order interaction $[F(1, 17) = 6.543, p = .020, \eta_p^2 = .278]$. The asymmetry was lower in the 85-dB SPL region compared to the 50-65-dB SPL region when the rising ramp was followed by a falling ramp. This latter interaction was found for both tones and noises, as indicated by a non-significant Order × Intensity-Region × Spectrum interaction ($p > .05$). In equalized regions, the same conclusions were reached: the analysis revealed a significant effect of Spectrum $[F(1, 17) = 9.622, p = .006, \eta_p^2 = .361]$, Intensity-Region $[F(1, 17) = 7.831, p = .012, \eta_p^2 = .315]$ and a significant Intensity-Region × Order interaction $[F(1, 17) = 4.515, p = .048, \eta_p^2 = .210]$. Again, the Order × Intensity-Region × Spectrum interaction was not significant ($p > .05$).

Additional one-tailed t-tests were conducted to determine whether the loudness asymmetries in each condition (Spectrum, Intensity-Region, Order) were significant or not. These analyses was done separately in equal and non-equal regions. The p-value threshold for significance was corrected using Bonferonni at the α level of .05 (.05/8 = .00625). The asymmetries were significantly higher than zero (i.e., Ps < .05) in all but two conditions, when the noises were presented in the rising-falling order in the 70-85 dB SPL region [Mean = 0.1 dB (SD = 3.0); $t(17) = 0.200, p = .422$] and in the 70-85-equal* region [Mean = 0.8 dB (SD = 3.0); $t(17) = 1.107, p = .142$].

2.3.3 Discussion

In the preliminary tone/noise loudness equalization experiment, considerable differences between participants' loudness tone/noise equalization judgments were found, as it was already reported in studies on spectral loudness summation (e.g., Anweiler and Verhey 2006; Verhey and Kollmeier 2002; Hots et al. 2014a). For most of the participants, the tone/noise level differences measured in the present experiment were lower than what has been reported in previous studies with broadband (environmental and pink) noises measured with adjustment methods (e.g., Schlittenlacher et al. 2011; Schlittenlacher et al. 2015). These differences probably primarily come from the fact that we did not account for the frequency characteristics of the headphones and that no free-field equalization was made. However, the only concern was that the sounds were matched in loudness on an individual basis, which is what was done.

The purpose of this second experiment was to further explore the effect of the intensity-region on the magnitude of the asymmetries, based on the results of Exp. 1. A two-interval loudness-matching task was employed in Exp. 2. A significant decrease of the asymmetry in the highest intensity-region was obtained when the sounds were presented in the rising-falling order, but not when the sounds were presented in the opposite order (i.e. in the falling-rising order) both in the analyses conducted on *equalized* and *non-equalized* regions. This decrease was obtained with both tones and noises, as indicated by non-significant Order × Intensity-Region × Spectrum interactions. In other words, this decrease occurred only in a specific context, i.e., when the two sounds were loud and matched in pairs in that particular order. In a previous study on 1-kHz tones, loudness asymmetries, inferred from a loudness matching roving-intensity experiment were found to be significantly reduced in high intensity-regions (65-80-dB SPL) compared to other lower regions (Ponsot et al., 2015). It is likely that this influence of the intensity-region on the asymmetry was actually caused by a significant decrease in the rising-falling presentation order but not in the other presentation order. However, we can unfortunately not check this assumption since Time-Order Errors affect loudness matches in roving-intensity paradigms (for details, see Ponsot et al. 2015). Which factors could be responsible for the present Order × Intensity-Region interaction? When loud rising and falling sounds separated by a small ISI (500 ms) are compared in loudness in a same pair and when the loudest portions of these sounds are contiguous (i.e., the sounds are presented in the rising-falling order), it can reasonably be assumed that participants are able to shift their attention to the loudest portions of the stimuli, thus paying less attention to their specific profile (rising or falling).

It is important to notice that the decrease of the asymmetry measured in this configuration cannot be attributed to any peripheral mechanism because the neural response of a sound decays within a few milliseconds after its offset (e.g., Smith and Zwislocki 1975). Thus, the intensity of the rising sound is no longer accessible at a peripheral stage when the second sound (falling) begins, suggesting that the effect could rather be attributed to a higher-level (cognitive or attentional) mechanism where the loudness comparisons of rising and falling sounds are mediated by their order of presentation when presented in high intensity-regions.

Furthermore, smaller but still significant asymmetries were observed for white noises compared to 1-kHz tones. The same outcome was reached whether the comparison was done in *non-equalized* regions or in *equalized* regions, which suggests that the spectral content of these sounds has a specific influence on their subsequent asymmetries, regardless of their inherent loudness difference at equal intensity[2]. Nevertheless, significant positive asymmetries were observed between rising and falling noises in many conditions, which contrast with the results of the first experiment where the asymmetry, evaluated by looking at the main effect of Direction in the ANOVA, was found to be significant for tones solely. This result suggests that the size of the asymmetries measured for broadband noises could actually be affected by the experimental procedure employed, since the loudness asymmetries were inferred from a direct loudness estimation task in the first experiment and from a loudness-matching task in the second experiment. Consistently, no significant asymmetries in loudness change were observed with direct scaling methods (Neuhoff, 1998), whereas smaller (compared to tones) but significant asymmetries were found in pairwise comparisons (Neuhoff, 2001). This would indicate that the measurement method and the presentation context could potentially affect the measured asymmetries for noises, as opposed to

[2]One could argue that a different conclusion would have been reached if equal-loudness regions had not been inferred from individualized loudness equalization experiment but rather from the use of current loudness models, because these models would have provided lower intensity-regions. However, this is not likely since the intensity had no effect on white noise asymmetries in the falling-rising order. It might only be possible that the asymmetries for white noises would have been slightly greater in the rising-falling order, more particularly in the high intensity-region.

what was reported for 1-kHz tones (Ponsot et al., 2015). Finally, it is worth mentioning that the asymmetries obtained for 1-kHz tones in the different conditions of this second experiment fall within a 2 –5 dB range, which is highly consistent with what was obtained in an equivalent experiment previously conducted (see Exp. 2 A presented in Ponsot et al. 2015).

2.4 General discussion and conclusion

The aim of the present study was to explore how the global loudness asymmetries observed between 2-s rising and falling sounds varied across different spectral contents and intensity-regions. Based on the large amount of literature exploring the perception and the coding of "looming" and "receding" sounds, showing that tonal "looming" stimuli are treated as special by the brain (e.g., Bach et al. 2009), greater asymmetries were expected for pure tones compared to noises. The two psychophysical experiments of this study supported this hypothesis at least in part, showing with different methods that loudness asymmetries for broadband noises (white noises or pink noises) were always significantly smaller than asymmetries for pure tones of various frequencies. More specifically, the first experiment showed similar asymmetries across pure tones ranging from 250Hz to 8kHz and no asymmetries for white noises and pink noises. This emphasizes the particular role of the spectral structure (structured vs. random) on global loudness ratings, similarly to what was observed on loudness change (Neuhoff 1998; Neuhoff 2001). The asymmetries obtained for the broadband noise in Exp. 2 were smaller than the asymmetries obtained for the 1-kHz tones; they were however significant in most of the conditions. This point is further discussed below. In short, rising / falling sounds having a random spectral structure, such as broadband noises, differ less in global loudness than do rising / falling tonal sounds[3].

[3]Another experiment conducted in the lab using a Magnitude Estimation procedure also showed significant asymmetries in global loudness between /ə/ vowels having rising versus falling-intensity profiles (the stimuli were created using Klatt synthesizer Klatt (1980) and were similar to those used in Neuhoff 's seminal study (Neuhoff, 1998), with a fundamental frequency at 100 Hz and formants at 450, 1450 and 2450 Hz), consistent with what was observed with pure tones, which emphasizes the present distinction made

Quantitative estimates of the asymmetries (in dB) were derived from the second loudness-matching experiment. Although the size of these asymmetries was found to depend on the condition, they were, on average, 3.3 dB for tones and 2.3 dB for noises. We evaluated to which extent a typical loudness model could account for such values, and more specifically how the observed spectral-content peculiarities could be predicted. As done in a previous study (Ponsot et al., 2015), the maxima of Long-Term-Loudness (LTL) patterns obtained from Glasberg & Moore model (Glasberg and Moore, 2002) were used as global loudness predictors. Asymmetries predicted for pure tones, across all frequencies and intensities, were always between 1.2 and 1.3 dB; asymmetries predicted for broadband noises were always around 1.1 dB. Thus, this model consistently underestimates the size of the asymmetries and cannot account for the significant dissimilarities observed between tonal stimuli and broadband noises (these asymmetries indeed come from the two contiguous AGCs employed in the model to compute LTL (see Ponsot et al. 2015)).

How can the present results then be related with Neuhoff's evolutionary interpretation? According to Neuhoff, rising tonal stimuli have to be treated specifically and to be given higher priority in order to provide us better preparation for potentially dangerous approaching objects. He argued that this effect was not observed for random spectral structures because most of the sounds produced by single approaching sound sources in our environment present a non-random, structured spectral content, whereas broadband noises present random spectral characteristics similar to "more dispersed phenomena such as wind or rain" (Neuhoff, 1998). In that sense, broadband noises, being less salient than spectrally structured sounds, would not be able to produce a "sufficiently naturalistic cue for a looming percept" (Leo et al., 2011). Based on the present results, several limits of this hypothesis can be pointed out:

(1) The asymmetry between rising and falling-intensity sounds seems to be reduced at higher compared to lower intensity-regions, a particularity that is not addressed in Neuhoff's hypothesis.

between tonal and broadband stimuli.

(2) Although the asymmetries obtained with noises in Exp. 2 were significantly smaller than the asymmetries obtained with 1-kHz tones, they were however significant in many conditions. Simply arguing that tonal stimuli produce asymmetries whereas noises do not is probably not sufficient. Moreover, many natural and environmental sounds exhibiting random spectral structures close to pink noises have also to be treated with high priority (e.g., natural sounds such as rocky screes or snow avalanches, sounds produced by approaching trains or cars). Thus, from a biological point of view, the potential underlying mechanisms responsible for the observed asymmetries are probably more complex than a simple harmonic/non-harmonic sound differentiation.

The first point needs to be studied in more details; in particular we still need to understand why the magnitude of the asymmetry in global loudness depends on the intensity-region of the stimuli as well as their presentation order when two-interval paradigms are employed. In the present study, smaller asymmetries were obtained when the ramps were in rather high intensity-regions (maximum intensities higher than 80 dB SPL). This effect was already observed in a previous study (Ponsot et al., 2015) but was not specifically addressed. Exp. 1 indicated that the effect could be observed both for tones and noises. In Exp. 2, the intensity-region had also a significant influence for both tones and noises but only when the pair was composed of a rising stimulus followed by a falling-intensity stimulus. The fact that those asymmetries can be reduced with increased intensity-region, but only in certain specific contexts, suggests that other high-level processes could potentially interact with the parent “direction-sensitive” mechanism. The second point stated above can also be further discussed using physical consideration of sound propagation. Indeed, the effect of atmospheric attenuation with distance on sound intensity depends on frequency. In that sense, potentially dangerous sources of natural broadband noises present intensity-variations that depend on the frequency regions: faster intensity-variations are observed at high spectral regions compared to low spectral regions (Gordon et al., 2013). In a study of Gordon et al. where time to arrival (TTA) and urgency ratings of rising narrow-band noises was investigated, narrow-band noises with center frequencies above 250 Hz were rated with greater urgency and were found to cause more TTA underestimation than lower-frequencies noises (Gordon et al., 2013). The authors argued that high-frequency components, being more affected by atmospheric conditions and distance than low-frequency components, have to be treated with greater priority because they provide a direct message informing the imminent arrival of a potentially dangerous approaching source. Therefore, one interesting perspective to further examine the evolutionary hypothesis proposed by Neuhoff would be to investigate (i) whether global loudness asymmetries for tones are reduced when frequencies below 250 Hz are used and (ii) whether ecologically structured intensity-varying broadband noises used by Gordon et al. induce greater global loudness asymmetries compared to random white noises or pink noises. This would suggest that behavioral asymmetries between rising- and falling-intensity sounds do not depend on the spectral content per se (noise/tone) but rather depend on the degree of "looming" that rising-intensity sounds might produce. The physical characteristics of the stimuli determining this impression of "looming" thus have to be assessed specifically.

All in all, the present results provide converging evidence to assume that the "direction-sensitive" mechanism, i.e., producing asymmetries in global loudness judgments, could indeed be due – at least in part – to the specific neural pathways followed by "looming" stimuli (e.g., Seifritz et al. 2002; Bach et al. 2008) because, as for many others but related phenomena, the degree of organization in the spectral content of the sound is found to reinforce these asymmetries. However, other, probably high-level processes related to the presentation order and the absolute intensity of the stimuli are found to mediate this mechanism and would need to be explored further. Although the view that an intrinsic dissimilar coding might be responsible for the behavioral asymmetries between rising and falling sounds lasting a few seconds is particularly interesting, a systematic study is needed to determine whether they fully arise from these inherent sensory or high-level (e.g. cortical) coding differences or whether they also reflect robust changes in observer's decision criteria or judgment strategies (e.g., DiGiovanni and Schlauch 2007).

Combining neuroimaging and behavioral experiments on both humans and animals to better explore the respective contribution of sensory biases and intrinsic biological coding differences underlying these asymmetries constitutes a promising direction for future research (e.g. Deneux et al. 2015). Current loudness models such as Glasberg and Moore's (Glasberg and Moore, 2002) cannot predict these asymmetries in global loudness, probably because they reflect low-level coding and peripheral auditory processes derived from the use of stationary sounds or amplitude-modulated sounds. As noted by Grimm et al. (2002) concerning the loudness of fluctuating sounds, even though these types of non-linear effects are rather small (less than 5 dB), these mechanisms deserve particular attention because they carry important information about the processes underlying global loudness evaluation of complex time-varying sounds. Further efforts have to be made to provide a full comprehension of how other features specific to rising and falling sounds, such as their duration and their slope (i.e., their rate of change), are integrated into global loudness judgments and affect this mechanism sensitive to the direction of sound intensity.

Acknowledgments

The authors would like to thank Guillaume Lemaitre and Michaël Vannier for helpful suggestions. This work was supported by the project LoudNat funded by the French National Research Agency (ANR-11-BS09-016-01).

Partial synthesis

The two main goals of this chapter were (1) to investigate the effect of the spectral structure on global loudness asymmetries and (2) to further explore the effect of the intensity-region.

(1) **Greater asymmetries are observed for pure tones as compared to broadband noises**. In the first experiment (magnitude estimation task), we found a significant effect of ramp direction (rising vs. falling) with pure tones and no interaction with ramp frequency, whereas there was no effect of ramp direction with broadband noises, both with white noise and pink noise. Quantitative measures of the asymmetry were obtained in the second experiment (loudness matching task): the asymmetries were, on average, 3.3 dB and 2.3 dB for 1-kHz tones and for white noises, respectively. Moreover, we found similar results whether the noises were equalized in loudness (as compared to tones) or not. Thus, we can conclude that the magnitude of the global loudness asymmetry between rising and falling sounds depends on their spectral structure (tonal vs. non-tonal) but also that the size of this discrepancy seems to rely on the employed measurement method.

One can ask why asymmetries for 1-kHz tones are greater than asymmetries for broadband noises. In the paper, we proposed a discussion based on Neuhoff's hypothesis, i.e. that this spectral structure effect might be caused by high-level auditory processing stages (e.g., cortical). Nevertheless, we raised several limits to this hypothesis. It is not clear where this tonal / non-tonal apparent discrepancy starts, so it cannot yet be ruled out that a part of the effect comes from the auditory periphery. As regard their physical properties, tones and noises differ only in terms of spectral bandwidth:

tones have an “infinitely” small bandwidth, whereas white noises have an “infinitely” large bandwidth. While broadband noises recruit all the ear’s critical bands, tones mainly activate one critical band[4]. As a consequence, it could be stated that the smaller the spectral bandwidth, the greater the asymmetry. To examine this hypothesis, a complementary experiment was conducted (see Appendix B for details) in which the global loudness of rising and falling ramps made of narrow-bandwidth noises covering either the [950-1050 Hz] spectral region or the [800-1200 Hz] spectral region was measured. We hypothesized that the asymmetries between ramps made of [950-1050 Hz] narrow-band noises (having a 100-Hz bandwidth, thus mostly recruiting the critical band centered on 1-kHz; Glasberg and Moore 1990) would be significantly greater that the asymmetries between ramps made of [800-1200 Hz] narrow-band noises (having a 400-Hz bandwidth and thus recruiting several critical bands). We did not find any significant difference between the asymmetries for noises having a 100-Hz bandwidth compared to noises having a 400-Hz bandwidth. It thus seems that the spectral structure effect observed in this chapter cannot simply be explained by the difference in spectral bandwidth between tones and noises, at least when considering the fact that asymmetries are comparable whether one or several critical bands (and hence nerve fibers) are recruited. Nevertheless, it is important to notice that the two narrow-bandwidth noises used in Appendix B present tonal aspects (a pitch could even be perceived with [800-1200 Hz] narrow-band noises). All together, these findings are consistent with the view that *tonalness* could be an important factor modulating global loudness asymmetries[5]. There were two other results of note. First, it is interesting to notice that the asymmetry value obtained in the second experiment for 1-kHz tones (3.3 dB) is highly consistent with the 3.2 dB value obtained Exp. 2A of Chapter 1 with the same experimental procedure. Second, contrary to what was found for 1-kHz tones in Chapter 1, a significant influence of the task (magnitude estimation task vs. loudness-matching task) was observed for white noises: no signifi-

[4]Due to ear nonlinearities, other critical bands are also activated, but to a much lesser extent.

[5]An experiment employing complex tones with unresolved harmonics could help determine whether tonalness is indeed the influential factor.

cant asymmetries were found in the first experiment (magnitude estimation task) whereas significant asymmetries were obtained in all the conditions of the second experiment (loudness-matching task), except when the sounds were compared in the rising-falling configuration around [70-85 dB SPL].

(2) **Asymmetries are reduced in high intensity-regions for both tones and noises**. In the first experiment based on a magnitude estimation task, results overall showed a significant intensity-region × direction interaction. In particular, the asymmetry was smaller in high intensity-regions compared to low intensity-regions. In the second experiment based on a loudness-matching paired comparison task, the asymmetries obtained for [70-85 dB SPL] ramps were lower than the asymmetries obtained for [50-65 dB SPL] ramps but only when these ramps were presented in the rising-falling order. Because this effect was found with 1-kHz tones, with white noises and also with narrow-band noises (Appendix B), it can be assumed that it does not depend on the spectral structure of the sound.

Following what was initiated in Chapter 1, we also compared the measured asymmetries with predictions from Glasberg and Moore's loudness model (2002). We showed that the model underestimates the mean size of the asymmetries and, in addition, that it could not account for the differences observed between tones and noises. Finally, Appendix C revealed significant correlations between the asymmetries measured with the same listeners for white noises and 1-kHz tones, suggesting that the same mechanism(s), or at least mechanism(s) sharing common properties, could be responsible for the asymmetries for tones and noises (at least in loudness-matching tasks). This result constitutes a first and preliminary attempt to explore the common properties of the mechanisms underlying asymmetries for tones and noises. It would be particularly challenging to investigate how such finding could be reconciled with the studies showing that a specific neural network is dedicated to rising tonal stimuli (e.g., Ghazanfar et al. 2002).

Chapter 3

How temporal profile characteristics of rising and falling tones shape their global loudness

Emmanuel Ponsot, Patrick Susini, and Sabine Meunier (in prep). Global loudness of rising- and falling-intensity tones: How temporal profile characteristics shape overall judgments.

Preface

Chapter 2 was primarily concerned with the influence of sound spectral structure on global loudness asymmetries between rising and falling intensity profiles. To this extent, the same 2-s rising and falling stimuli varying over 15-dB dynamics but made of different spectral characteristics (tones or noises) were considered. The present chapter is not concerned with the *spectral* but with the *temporal* characteristics of global loudness processing of rising and falling sounds, more specifically in the case of 1-kHz tones.

Based on previous studies conducted on time-varying sounds showing, on the one hand that global loudness judgments are heavily guided by the maximum level of the sounds but also by the energy contained in a certain temporal window located on that region (e.g., Susini et al. 2010), and on the other hand that "recency" effects appear to be involved in overall evaluations of sequences lasting several seconds (e.g., Susini et al. 2007), we explored the extent to which global loudness of ramps could be determined by two mechanisms. (1) An *integration mechanism*, which states that global loudness results from the integration of "momentary" loudness (i.e. the loudness perceived at any instant) within a fixed temporal window located around the loudest portion of the stimulus and (2) a *decay mechanism* that accounts for the "recency effect", in which global loudness is a function of the time elapsed between the loudest peak and the end of the sound. While global loudness evaluations of rising ramps are likely conditioned by the *integration mechanism* only (the peak always being at the end of the stimulus), we hypothesized that both the *integration mechanism* and the *decay mechanism* might be recruited by falling ramps.

The plausibility of these two mechanisms was assessed by means of two magnitude estimation experiments in which the global loudness of rising and falling tones with different temporal characteristics was measured. The *slope*, *duration*, and *dynamics* of these stimuli was manipulated to recruit the two hypothetical mechanisms in different ways. Lastly, we also examined to which extent Glasberg & Moore's model (2002) could account for the effects induced by these temporal profile manipulations.

The article presented in this chapter is currently in preparation.

Ponsot, E., Susini, P., Meunier, S. (in prep). Global loudness of rising- and falling-intensity tones: How temporal profile characteristics shape overall judgments.

The sizes of the asymmetries (in decibels) between rising and falling tones obtained in the different experiments presented in this chapter are presented in Appendix D. A partial synthesis presented at the end of this chapter summarizes the main outcomes of the paper.

Abstract: Two magnitude estimation experiments were conducted to further investigate the mechanisms underlying global loudness evaluations of 1-kHz tones rising or falling linearly in level by manipulating their temporal profile characteristics. In Experiment 1, ramps of fixed *dynamics* (15-dB) were stretched in time between 1 and 16 s, resulting in different [*slope, duration*] combinations. For both rising and falling tones, global loudness judgments increased until 6 seconds and remained relatively constant beyond that. The loudness growth between 2 and 6 s was however greater for rising than for falling tones. In Experiment 2, ramps of fixed *slope* (2.5 dB/s or 5 dB/s) were stretched in time between 2 and 12 s to create different [*dynamics, duration*] combinations. In that case, global loudness was greater for 2.5 dB/s ramps vs. 5 dB/s ramps, and the effect of duration became non significant. Overall, these results suggest that, qualitatively, global loudness of ramps could be seen as an integration of their loudest portion over a certain temporal window. However, this integration mechanism would not be sufficient to account for the asymmetries observed between rising and falling ramps in both experiments as well as their dependencies on the level, the dynamics and the duration of the stimuli. The present data show that the global loudness of time-varying stimuli is presumably due to several mechanisms not yet fully understood. These effects could not be accounted for by the peak values of short-term or long-term loudness provided by Glasberg and Moore 's model for time-varying sounds.

3.1 Introduction

Global loudness, which has been defined as the overall impression of loudness of stimuli varying in loudness through time (e.g., Susini et al. 2007), is an important psychoacoustical variable when dealing with time scales of several seconds. For instance, the industry and the media need to predict and therefore, to control, as accurately as possible, how loud a sound will be perceived (e.g., the sound produced by the passing-by of an airplane, or an advertisement to be broadcasted on the radio). These sound sequences are generally long (at least a few seconds) and strongly varying in loudness through time, such that overall indicators are required to normalize them in terms of *global loudness*. Psychoacoustical experiments investigating dynamic loudness perception at these time scales have shown that global loudness does not correspond to an average of momentary loudness, i.e. to the average loudness experienced during the stimulus, but is rather strongly influenced by the loudest events (e.g. Kuwano and Namba 1985; Gottschling 1999; Kuwano et al. 2003; Susini et al. 2002; Susini et al. 2007). Current indicators of global loudness are all based on this outcome. In the media, the overall loudness of a program is simply taken as the integration of its momentary loudness values (predicted by simplified auditory models) that exceed a certain threshold (see BS) 2006; EBU-Recommendation 2011). Other indicators employed in the industry, such as L_{Aeq}, also rely on the assumption that global loudness could be evaluated by averaging the physical energy of the stimulus. Even in the context of more basic psychoacoustics, Zwicker (1999) or Glasberg and Moore (2002) suggested to use the peak of the "short-term loudness" (STL) or the "long-term loudness" (LTL) time-series predicted by their models to estimate the global loudness of time-varying sounds. However, recent studies conducted with very basic sounds have pointed out a limitation to such assumptions. For instance, it has been shown that 1-kHz tones increasing linearly in level during 2-s over a 15-dB dynamics (i.e. range of level variation) are consistently judged about 3-4 dB louder than their time-reversed, falling versions, and it was demonstrated that this asymmetry could not be accounted for, neither by STL maxima nor by LTL maxima provided by current loudness models (Ponsot et al., 2015). These results show that, even with very basic 1-kHz stimuli ramping up or down in level, global loudness is not simply based on the maximum of their short-term loudness. The mechanisms that underpin global loudness evaluations of such ramps still remain undetermined.

Our listening environment contains lots of rising and falling-level events lasting a few seconds. Real moving sound sources (e.g. a car passing-by) present increasing and decreasing level profiles (induced by the approaching and the receding portions, respectively) and musical sequences are full of *crescendo* and *decrescendo* passages. It is thus particularly valuable to understand how intensity dynamics occurring at this timescale are processed and why asymmetries in global loudness judgments of simple rising vs. falling profiles occur. The purpose of the present study was twofold: (1) to better understand the mechanisms underlying the global loudness evaluation of basic 1-kHz rising and falling-intensity tones of a few seconds, and (2) to explore to which extent asymmetries between rising and falling tones are influenced by their temporal profile characteristics. These closely interrelated issues were addressed by means of two psychophysical experiments in which the parameters defining the temporal profiles of these linear rising and falling ramps of sound level, namely their *duration*, their *slope* (i.e. rate of change in dB/s) and their *dynamics* (i.e. difference between minimum and maximum levels), were manipulated to test specific hypotheses with regard to potential underlying mechanisms.

Most previous studies investigating *global loudness* of simple rising and falling sounds at this time scale employed ramps with the same combination of parameters (slope, duration, dynamics). Most often, the ramps were 2-s long and covered a dynamics of 15-dB, thus resulting in a slope of 7.5 dB/s (Ponsot et al. 2013; Ponsot et al. 2015; Ponsot et al. in press).

One study examined the effect of the dynamics on the global loudness of 1.8-s rising ramps (Susini et al., 2010). Greater global loudness estimates were found for rising ramps having 15-dB dynamics as compared to ramps having 30-dB dynamics and the same maximum level. Furthermore, global loudness estimates of these rising ramps were close but slightly lower than those of constant-intensity tones presented at their maximum level. To explain these effects, the authors proposed that global loudness evaluation of rising ramps might involve a certain integration of its level-profile over a temporal window located around the maximum of the stimulus (see Meunier et al. 2010). Such a mechanism, that we will call the *integration mechanism* in this paper, is indeed consistent, at least qualitatively, with the observation that when the dynamics of these ramps is increased while its maximum level remains the same, a greater amount of energy is contained under the temporal window, and hence global loudness increases. This *integration mechanism* would also explain why rising ramps are perceived softer than their maximum level. According to this hypothesis, if the duration of a ramp is increased but its dynamics is fixed, global loudness should also increase. No experiment directly tested this assumption with simple rising ramps, but there is one study that investigated the influence of the duration on global loudness judgments of time-varying 1-kHz tones, which consisted of sequences made of stationary tones plus ramps. For sound sequences made of a 3-s constant plateau followed by a rising ramp, global loudness was found to increase gradually when the duration of the ramp increased between 2 and 20 seconds while its dynamics was kept constant, equal to 20 dB (see Susini et al. 2007). This result supports the idea that a certain *integration mechanism* might be involved.

Similarly, this *integration mechanism* is assumed to be involved with falling ramps, taking place on their loudest portion, i.e. at the beginning of the stimuli (Meunier et al., 2010). Thus, at equal duration, the global loudness of a falling ramp should decrease when its dynamics is increased, whereas at equal dynamics, its global loudness should increase when its duration is increased. However, when dealing with falling-intensity stimuli of a few seconds, there is another phenomenon reflecting "memory process" that needs to be taken into account. Indeed, a number of studies observed that global loudness judgments were greater when the loudness peak was closer to the end of the sequence (Hellbrück 2000; Susini et al. 2002; Kuwano et al. 2003). These authors suggested that the loudness peak has a greater impact on the overall evaluation when its encoding in memory is more recent. “Recency" (Susini et al., 2002) was thus proposed as a candidate mechanism. However, it is not clear yet (and it has never been directly demonstrated) that such effects really involve memory mechanisms[1] or whether they simply reflect judgment heuristics such as the "peak-end" rule (Kahneman et al. 1993; Schreiber and Kahneman 2000). In what follows, we will thus simply refer to this phenomenon as the *decay mechanism* (i.e. reflecting either a "memory decay" or a "peak-end" rule - the fact that when the peak is not at the end, it has a smaller influence), because it seems to downsize the influence of a loudness peak as a function of the time lapse between its position and the end of the sound. It is however impossible yet to tell at which durations this mechanism really starts to be involved and what is its typical rate of decrease. We believe this mechanism is probably not responsible for the whole asymmetry observed between rising and falling tones at short durations, such as with 2-s ramps[2], but it might possibly play a role in global loudness evaluations of falling tones of longer durations (e.g., 10 s), since their loudness peak is then clearly further back away in time (Susini et al., 2007). Therefore, it could be hypothesized that, when the duration of a falling ramp is increased while its dynamics is fixed, global loudness judgments would result from the product of two mechanisms: (i) the *integration mechanism* that would increase global loudness, and (ii) the *decay mechanism* that would decrease global loudness. Whether the sum of these two processes would, in the end, lead to an increase or a decrease of global loudness as a function of the duration can however not be predicted[3]. The fact that these two mechanisms might potentially interact for falling

[1]As, for example, measured in terms of *d'* (e.g. Clément et al. 1999).

[2]Indeed, significant asymmetries are observed with 2-s ramps that are matched in loudness in pairwise comparisons in different orders of presentation (Ponsot et al. 2015; Ponsot et al. in press).

[3]One may argue that the results obtained in Susini et al. (2007) might indicate that these two mechanisms would have similar weights in the process, given that global loudness of sequences containing falling ramps followed by constant plateau was found to remain fairly constant when ramp duration was increased from 2 to 20 s and its dynamic was fixed. However, we believe that the presence of a plateau at the end of the sequence might have significantly affected the integration processes specifically related to the ramp itself (more than in the case of rising sequences where the plateau is located at the beginning), and hence these results might not be directly transposed to the present context.

tones can nevertheless be observed by looking at the evolution of the difference between rising and falling sounds (i.e., the so-called "asymmetry") as a function of their duration; this asymmetry should be increased.

Besides, from the same reasoning, increasing the duration of a rising ramp while keeping its slope constant should not affect global loudness because the *integration mechanism* would always integrate the same energy whatever the duration of the ramp is. Increasing the duration of a falling ramp while keeping its slope constant should this time decrease global loudness, directly reflecting the *decay mechanism*.

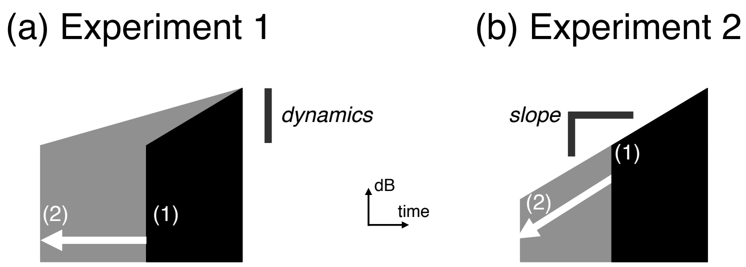

Figure 3.1: Schematic representation of the two experiments conducted in the present study, where the duration of rising and falling-intensity ramps was manipulated in different ways (see arrows, from (1) to (2)). (a). Experiment 1: the ramps were "stretched" in time while keeping their *dynamics* constant. (b) Experiment 2: the ramps were "stretched" in time while keeping their *slope* constant. A rising ramp is taken as an example on this figure. The analogous procedure was implemented for falling ramps.

These hypotheses remain however somewhat speculative since they are derived from a small number of studies, which in some cases, did not use simple rising and falling ramps but more complex sound sequences. The purpose of the present study was thus to directly address the plausibility that the two proposed mechanisms, namely the *integration mechanism* and the *decay mechanism*, might be involved in the global loudness processing of rising and falling tones. This was examined in particular in the context of direct global loudness judgments using magnitude estimation tasks. Two psychophysical experiments were designed to disentangle the two presumed mechanisms by manipulating[4] the parameters of the ramps. The stimuli were 1-kHz tones either rising or falling linearly in level (i.e. their slope, duration and dynamics), like those employed in our previous studies (Ponsot et al. 2015; Ponsot et al. in press). This manipulation is illustrated in Figure 3.1. In Experiment 1, the ramps were "stretched" in time in such a way that their dynamics was kept constant (cf. Fig. 3.1, panel (a)), resulting in different slope, duration combinations. This is similar to what was done in Susini et al. (2007) with more complex sequences. In this context, we hypothesized that (1) global loudness of rising tones would increase with duration because of the *integration mechanism* and (2) global loudness of falling tones would not grow as fast as for rising tones because both the *integration* and the *decay mechanism* would combine. In Experiment 2, the ramps were "stretched" in time in such a way that their slope was kept constant (cf. Fig. 3.1, panel (b)), resulting in different dynamics, duration combinations. In that context, we hypothesized that (1) global loudness of rising tones would not vary with duration because the same amount of energy will always be contained in a region located over the maximum level and (2) global loudness of falling tones would decay, providing a direct image of the *decay mechanism*.

Finally, these two experiments allow us to also assess the extent to which asymmetries between global loudness of rising and falling vary with the manipulated parameters, namely the slope, the dynamics and the duration of these ramps. Due to the presumed *decay mechanism* involved with falling tones, we were expecting an increase of the asymmetry with the duration of the stretching in both experiments (i.e. direction $\times$ duration interactions). Finally, the experimental design also attempted to determine to which extent the effects of the ramp parameters (slope, duration, dynamics) on both global loudness judgments and their resulting asymmetries would depend on the mean intensity of the stimuli. Indeed, we already observed in previous studies that the asymmetry between rising and falling ramps was significantly reduced when the maximum level of the stimuli was higher than 80

[4] This parameter manipulation was inspired from studies examining the influence of these factors to investigate the mechanisms underlying *loudness change* judgments (Canévet et al. 2003; Teghtsoonian et al. 2005).

dB SPL, an effect that remained unexplained so far (Ponsot et al. 2015; Ponsot et al. in press). Thus, in both experiments, ramps were presented in different intensity-regions, below and above 80 dB SPL. To complete this investigation, we examined to which extent different global loudness indicators derived from Glasberg and Moore's model outputs (2002) could account for the results collected in these two experiments.

3.2 Experiment 1

3.2.1 Materials and method

3.2.1.1 Participants

Forty-five participants were recruited for this experiment. They were divided into two groups, performing the experiment under different conditions (see Stimuli section below): Group A, 30 participants (15 women, 15 men; age 22-35 years); Group B, 15 participants (8 women, 7 men; age 18-32 years). All reported normal hearing. They gave their informed written consent according to the Declaration of Helsinki prior to the experiment and were paid for their participation. The participants were naïve with respect to the hypotheses under test.

3.2.1.2 Stimuli

The stimuli were 1-kHz pure tones with various durations and intensity profiles. The loudness function of each participant was measured prior to the experiment using 500-ms constant-intensity tones (presented at ten different levels equally spaced between 45 and 90 dB SPL). In the experiment, tones with rising and falling-intensity profiles were used; their sound level was linearly varied over 15 dB (i.e., 15-dB dynamics). They were presented in four regions: R_1 = [60–75], R_2 = [65–80], R_3 = [70–85] and R_4 = [75–90 dB SPL]. Participants of Group A were presented with ramps of five different durations (1, 2, 6, 9 and 12 s); participants of Group B were presented with another set of three durations (4, 8 and 16 s). The amplitude envelopes of the stimuli were all smoothed with 10-ms linear rise and fall times.

3.2.1.3 Apparatus

The stimuli were generated at a sampling rate of 44.1 kHz with 16-bit resolution using Matlab. Sounds were converted using a RME Fireface 800 soundcard, amplified using a Lake People G-95 Phoneamp amplifier and presented diotically through headphones (Sennheiser HD 250 Linear II). Sound level was calibrated using a Brüel & Kjær artificial ear (Type 4153, IEC318).

Participants were tested in a double-walled IAC sound-insulated booth at Ircam.

3.2.1.4 Procedure

An Absolute Magnitude Estimation (AME) procedure was used, based on the instructions of Hellman (1982). No standard was given to the participants. Their task was to give a number proportional to the global loudness of each sound, i.e. the overall impression loudness over the total sound duration (Ponsot et al. 2013; Ponsot et al. 2015; Susini et al. 2007; Susini et al. 2010). For each participant, the experiment was scheduled in one session lasting about one hour. The measurement of the loudness function was done at the beginning of the session. After 20 training trials, each tone was presented 9 times in a "pseudo-random" order in order to reduce sequential effects (Cross, 1973), as it was done previously (see Ponsot et al. 2015). The experiment continued with the presentation of rising and falling ramps at various durations and intensities. A blocked-duration design was adopted; i.e. each block was made of sounds having the same duration. Each block consisted of interleaved rising and falling ramps of equal-duration presented at the four different intensity-regions, as mentioned above. Each stimulus was presented 5 times. Thus, a total of 200 stimuli (2 Directions $\times$ 4 Intensity-Regions $\times$ 5 Durations $\times$ 5 Repetitions) were presented to the participants of Group A and a total of 120 stimuli to the participants of Group B (2 Directions $\times$ 4 Intensity-Regions $\times$ 3 Durations $\times$ 5 Repetitions). The order of presentation of the blocks was randomly chosen for each participant.

3.2.2 Results

For each listener of each group, the average perceived global loudness of each stimulus was computed using the geometric mean of all his / her ratings. These mean loudness estimates were then normalized individually. All the ratings given by a listener to both constant tones and ramps were divided by the mean rating assigned to the 60-dB SPL constant-intensity tone and multiplied by four in order to match the loudness of a 60-dB SPL, 1-kHz pure-tone, which corresponds to 4 sones (see Susini et al. 2010; Ponsot et al.

2015). After this normalization, magnitude estimates are supposed to be directly comparable with loudness values in sone units[5].

The collected data were analyzed separately for each group. Repeated-measure ANOVAs (Direction × Duration × Intensity-Region) with univariate approaches were performed on the logarithm of the normalized loudness ratings accorded to rising and falling ramps within each group. The statistical analyses were conducted using R 2015. All the tests were two-tailed and used a probability level of .05 to test for significance. The Huynh-Feldt corrections for degrees of freedom were used where appropriate. Effect sizes are reported using partial eta-squared η_p^2.

The normalized magnitude estimates obtained in each group are presented in Figures 3.2 and 3.3 respectively, on a y-log scale, as a function of the duration of the ramp. Overall, global loudness estimates of rising tones (upward unfilled triangles) appeared to be always greater than (or at least equal to) those given to their time-reversed versions, i.e. falling tones (downward filled triangles). This was supported by significant effects of the direction obtained both for Group A $[F(1,29) = 25.07, p < .001, \eta_p^2 = .464]$ and for Group B $[F(1,14) = 12.14, p = .004, \eta_p^2 = .464]$. Furthermore, the plots in Figures 3.2 and 3.3 showed that global loudness increased with duration for both ramp directions, at least until 6 s, but that the speed of this growth might differ between rising and falling tones. Beyond 6 s, global loudness tended to remain constant for rising-intensity tones whereas it seemed that there might be a slight decrease for falling-intensity tones. The analyses in Group A showed a significant effect of duration $[F(4,116) = 4.70, p = .013, \eta_p^2 = .139, \tilde{\epsilon} = .49]$ and a significant duration × direction interaction $[F(4,116) = 4.60, p = .009, \eta_p^2 = .137, \tilde{\epsilon} = .61]$. In Group B, a significant effect of duration was found $[F(2,28) = 5.82, p = .013, \eta_p^2 = .294, \tilde{\epsilon} = .81]$ but the duration × direction interaction was only marginally significant $[F(2,28) = 3.20, p = .056, \eta_p^2 = .186, \tilde{\epsilon} = 1.15]$.

[5]However, note that this is true only if (1) the exponent of the loudness function measured on the listeners is equal to 0.6, which is the mean value used in current loudness models, and if (2) listeners use the same scale (i.e. the same range of numbers) for the measurement of the loudness function and for the remaining part of the experiment where the ramps were presented.

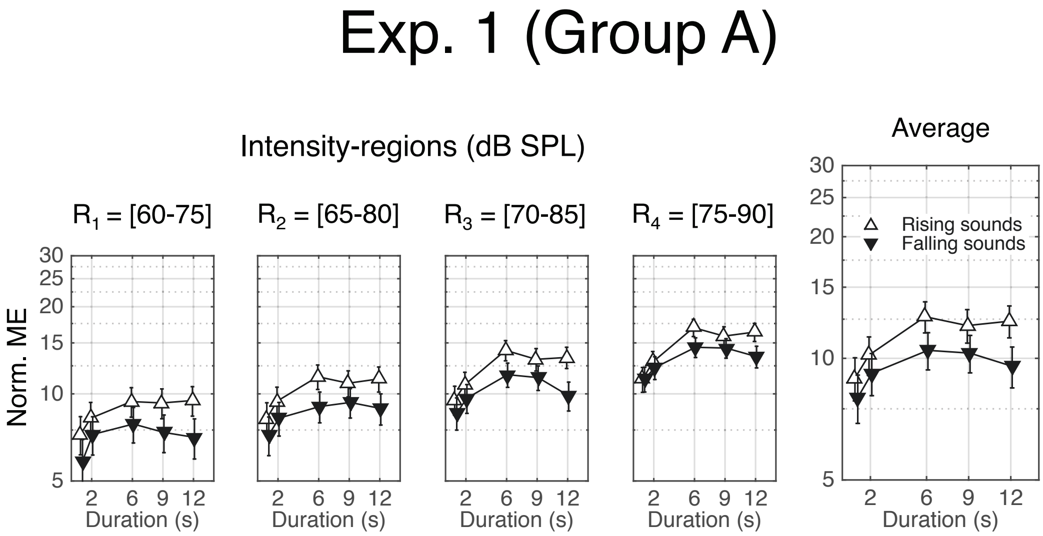

Figure 3.2: Normalized estimates of global loudness (i.e. normalized ratings) for rising (upward unfilled triangles) and falling sounds (downward filled triangles) obtained in Experiment 1 with Group A. Results are plotted on a y-log axis as a function of the duration of the ramp (from 1 to 12 s), for the different intensity-regions on the left panels (from R_1 to R_4) and averaged on the right-most panel. Error-bars show SEM in each configuration.

To gain further insight into the duration $\times$ direction interaction obtained in Group A, multiple *post-hoc* repeated-measures ANOVAs were conducted on pairs of adjacent durations to determine the duration at which this interaction appeared. The p-value threshold for significance was corrected using Bonferonni at the α level of .05 (.05 $\times$ 5/2 = .02). The only significant interaction was obtained in the analysis involving durations of 2 and 6 s $[F(1, 29) = 13.47, p = .001, \eta_p^2 = .317]$. The overall interaction obtained in the main ANOVA in Group A was thus due to what happened between 2 and 6 s. The slight decrease which could be observed at every intensity-regions for falling tones beyond 6 s was not significant.

Finally the overall difference between the curves presented in each panel (i.e. the asymmetry between rising and falling tones) was found to diminish with the intensity-region in Group A, as revealed by a significant direction $\times$ intensity-region interaction $[F(3, 87) = 5.94, p = .003, \eta_p^2 = .170, \tilde{\epsilon} = .78]$.

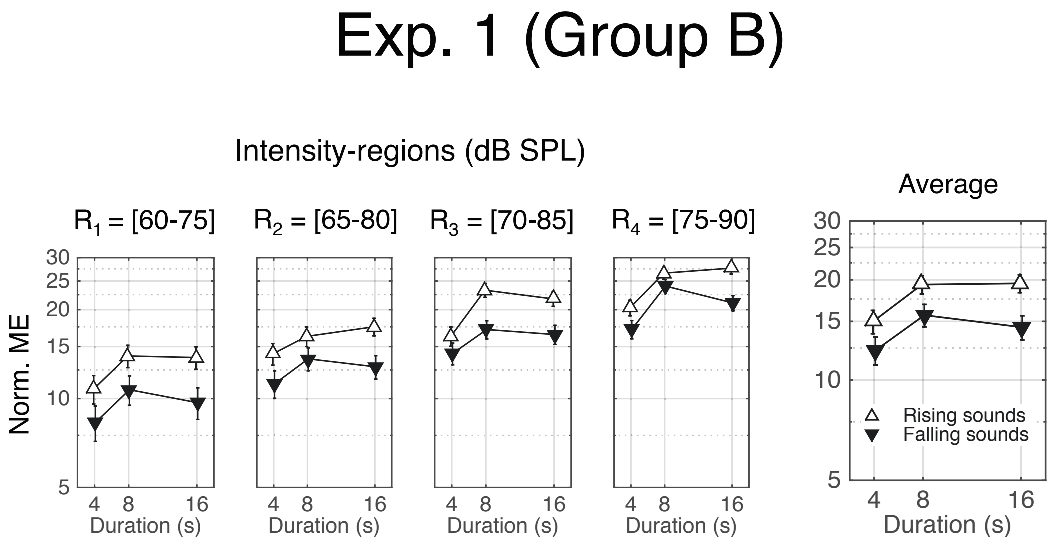

Figure 3.3: Normalized estimates of global loudness for rising and falling sounds obtained in Experiment 1, with Group B, for ramp durations of 4, 8 and 16 s. Same legend as Figure 3.2.

This interaction could also be observed in Group B (see Fig. 3.3) but was not significant $[F(3,42) = 1.85, p = .160, \eta_p^2 = .117, \tilde{\epsilon} = .90]$. Post-hoc analyses revealed that the interaction in Group A disappeared when ramps presented at R_4 (i.e. [75-90 dB SPL]) were discarded, indicating that the reduction occurred in this highest intensity-region.

3.2.3 Discussion

This first experiment examined how global loudness judgments of rising and falling ramps of a few seconds evolve when these sounds are "stretched" in time such that their *dynamics* is kept constant. We hypothesized that the *integration mechanism* would increase global loudness of rising tones and that its association with the *decay mechanism* will make global loudness of falling tones grow less rapidly. Our results only partially supported these hypotheses.

Global loudness of rising tones indeed increased with duration, as if there was some kind of loudness integration, but beyond 6 s appeared to "saturate"; a result that cannot – at first sight – be explained by this *integration mechanism*. Indeed, if one considers the area contained in a fixed window located under the level profile of a linear ramp, this area should grow logarithmically as a function of the duration of the "time-stretching". We have no clear explanation for this result, but given that 15-dB ramps longer than 6 s present slope which are smaller than 2.5 dB/s, it might be possible that this rate of change is not sufficient to involve an *integration mechanism* which is specific to dynamic stimuli, i.e. that listeners then process ramps as constant-intensity tones. Concerning falling tones, global loudness also increased with duration, until 6 s in Group A and 8 s in Group B. Beyond that, it remained constant and then appeared to slightly decrease after 9 s in Group A. The overall trend as a function of duration was different for falling tones compared compared to rising tones, as revealed by a significant interaction between the duration and the direction for Group A and a marginally significant interaction for Group B. It is likely that the duration × direction interaction was only marginally significant in Group B because the tested durations were more distant, namely, 4, 8 and 16 s, and also because the sample size was about half that of Group A. These interactions support our hypothesis that two mechanisms, an *integration mechanism* and a *decay mechanism* might add up, but post-hoc analyses indicated that this interaction in Group A was significant only between 2 and 6 s, i.e. the growth was slower for falling tones compared to rising tones. While we were expecting a gradual effect as a function of time, these results indicate that some *decay* *mechanism* might play a significant role but only between 2 s and 6 s (i.e. where the interaction was found). Although not significant, a decrease of the estimates of falling tones above 6 s in Group A and above 8 s in Group B could be noticed in every intensity-regions.

Overall, these results are different to those reported by Susini et al. (2007), where sequences of 1-kHz tones with time-varying intensity profiles were employed. In their study, the stimuli sequences were made of rising or falling ramps of different durations (from 2 to 20 s) having a fixed dynamics equal to 20 dB, combined with 3-s constant-intensity plateau presented either before or after the rising or falling ramp, respectively. Concerning [plateau-rising] sequences, they showed that global loudness increased gradually with duration, whereas we found that the global loudness of simple rising ramps reached a constant value at 6 s and then remained constant for longer durations, at least until 16 s. Concerning [falling-plateau] sequences, they showed that global loudness did not vary significantly with duration, whereas in the present study, global loudness of falling tones increased significantly with duration until 6 s. These noticeable differences between our study and Susini et al.'s (2007) could likely be attributed to the presence or the absence of a plateau before or after the ramp.

Another aspect of the present results concerns the asymmetry between rising and falling tones: greater global loudness judgments were obtained for rising ramps compared to falling ramps at all durations and, as expected, the size of this asymmetry increased with sound duration. This increase, which is assumed to be due to the *decay mechanism*, only occurred significantly between 2 and 6 s but one can see that the asymmetry also increased between 9 and 12 s (see also Appendix D. Lastly, the asymmetry was found to depend on the intensity-region of the ramps in Group A, and was significantly reduced in R_4 as compared to other regions. This decrease of the asymmetry in high intensity-regions was already observed in other studies on this topic (Ponsot et al. 2015; Ponsot et al. in press), but its causes still remain unclear.

The second experiment was designed to further assess the plausibility of the two proposed candidate mechanisms using the other experimental design presented in the introduction, i.e. using a "time-stretching" manipulation where the *slope* of the ramp is preserved. In addition to the ramp stimuli, we added constant-intensity tones presented at different levels and durations (corresponding to the maximum levels and the durations of the ramps), in order to compare their loudness with the global loudness of the ramps. This also allows to control that listeners do not deviate[6] with duration in their loudness evaluations for sounds lasting several seconds.

[6]The durations employed in this experiment being equal or greater than 1 s, we are well beyond the durations at which "temporal integration of loudness" occurs, which is generally assumed to be fully completed at 300 ms (see Rennies et al. 2010; Hots et al. 2014b). As a result, the loudness of constant tones should not be affected by its duration.

3.3 Experiment 2

3.3.1 Materials and method

3.3.1.1 Participants

Twenty-nine subjects took part in this experiment (13 women, 16 men; age 19-34 years). All reported normal hearing. They gave their informed written consent according to the Declaration of Helsinki prior to the experiment and were paid for their participation. The participants were naïve with respect to the hypotheses under test.

3.3.1.2 Stimuli

All the stimuli were 1-kHz pure tones. As in Experiment 1, 500-ms constant-intensity tones were used to measure the loudness function of each participant prior to the experiment (the same levels of presentation were used, equally spaced between 45 and 90 dB SPL). In the main part of the experiment, tones with constant, rising or falling-intensity profiles were employed. Different combinations of parameters [duration; slope] were used to create an appropriate set of rising and falling-intensity tones. Their slope (i.e. absolute rate of change) was either 2.5 dB/s or 5 dB/s. The ramps varying at 2.5 dB/sec were presented at four durations (2, 4, 6 and 12 s) and the ramps varying at 5 dB/sec were presented at three durations only (2, 4 and 6 s), in order to avoid too low (start or end) levels that would have been induced if the 12-s duration had also been used. All the ramps were presented with four different maximum levels ($M_1 = 75$, $M_2 = 80$, $M_3 = 85$ and $M_4 = 90$ dB SPL). Their minimum levels and consequently, their *dynamics*, resulted from the set [duration; slope] considered. The constant-intensity tones were presented at four durations (2, 4, 6 and 12 s) and at the same four levels as the maximum levels of the ramps (75, 80, 85, 90 dB SPL).

3.3.1.3 Apparatus

The apparatus were the same as described in Experiment 1.

3.3.1.4 Procedure

The procedure employed in this experiment was similar to the one described in Experiment 1, i.e. an AME procedure with a blocked-duration design. After the preliminary loudness function measurement (similar to Exp. 1), participants were presented with longer constant and ramp tones. Each block consisted of interleaved constant, rising and falling ramps of equal-duration presented at different levels, as mentioned above. Each stimulus was repeated 3 times. The order of presentation of the 4 blocks (one for each duration) was randomly varied between participants. A total of 48 constant-intensity tones (4 Levels × 4 Durations × 3 Repetitions), 96 ramps varying at 2.5 dB/sec (4 Max. Levels × 2 Directions × 4 Durations × 3 Repetitions) and 72 ramps varying at 5 dB/sec (4 Max. Levels × 2 Directions × 3 Durations × 3 Repetitions) were thus presented to the participants.

3.3.2 Results

The same normalization as in Exp. 1 was applied to the loudness ratings given by each listener. Different repeated-measure ANOVAs (rmANOVAs) were conducted to analyze the results in different ways because the ramps varying at 2.5 dB/s and those varying at 5 dB/s did not share the same set of durations. The *p*-values of these multiple analyses were not corrected since they were all planned prior to the experiment. The normalized loudness ratings are presented in Figures 3.4 and 3.5 for ramps having a slope of 2.5 dB/s and 5 dB/s, respectively. Loudness estimates for constant-intensity tones, which had the same maximum levels and durations as the ramps, are superimposed on each panel.

3.3.2.1 Analysis A - Loudness of constant tones with durations between 2 and 12 s

A first rmANOVA was conducted on the estimates given to constant tones only. A small increase of loudness estimates with duration until 6 s could be observed (see Figure 3.4), but the analysis revealed that the effect of duration was not significant ($p > .05$). There was no significant duration ×

level interaction ($p > .05$). The constant tones are not taking into account in the analyses that follow, which focused on the effects related to rising and falling ramps. However, it is important to observe that the loudness of these constant tones was always greater or at least equal to the global loudness estimates given to ramps with the same maximum level (see Fig. 3.4 and 3.5).

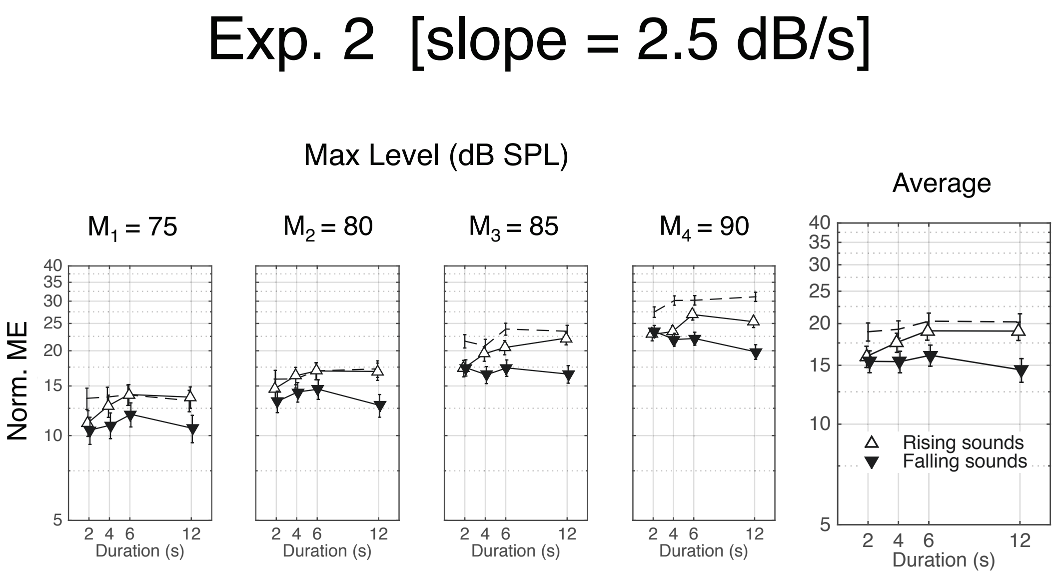

Figure 3.4: Normalized estimates of global loudness obtained in Experiment 2 for rising (upward unfilled triangles) and falling ramps (downward filled triangles) whose slope was equal to 2.5 dB/s, plotted as a function of their duration. Same layout as in previous figures; the data are presented for the different maximum level of the ramps on the left panels (from M_1 to M_4), and after an averaging over these different levels on the right-most panel. The loudness estimates obtained for constant-intensity tones having the same level as the maximum level of the ramp are superimposed in each panel (dashed lines). Error-bars correspond to SEM.

3.3.2.2 Analysis B - Global loudness of rising and falling ramps varying at 2.5 dB/s

A second analysis was performed to compare specifically rising and falling ramps varying at 2.5 dB/s, for which global loudness estimates are presented in Figure 3.4. Greater estimates were overall obtained for rising ramps compared to falling ramps, as supported by a significant effect of the direction $[F(1,28) = 11.39, p = .002, \eta_p^2 = .289]$. However, as it can be observed in Fig. 3.4, the size of this difference appeared to increase with the duration because the curves took somewhat different directions. Global loudness judgments of rising tones slightly increased with duration until 6 s and then reached a plateau, whereas global loudness judgments of falling tones remained fairly constant with duration and even appeared to decrease slightly between 6 and 12 s. This was supported by the fact that there was no significant main effect of the duration ($p > .05$) but a significant direction × duration interaction $[F(3,84) = 3.94, p = .034, \eta_p^2 = .123, \tilde{\epsilon} = .54]$. All these effects appeared to be similar at the different maximum levels tested, as supported by no significant interactions between the maximum level of the ramps and other factors ($p > .05$). We conducted *post-hoc* tests to determine whether the changes observed with duration for each profile separately (rising / falling) were significant or not. We found no significant effects of duration neither for rising tones, nor for falling tones ($p > .05$).

3.3.2.3 Analysis C - Global loudness of rising and falling ramps varying at 5 dB/s

This analysis was concerned with global loudness estimates of ramps varying at 5 dB/s. These data are presented in Figure 3.5. Overall, similar conclusions to those obtained with ramps varying at 2.5 dB/s were reached. A significant effect of the direction was found $[F(1,28) = 10.92, p = .003, \eta_p^2 = .281]$. As in Figure 3.4, we could observe a slight increase of the judgments of rising tones as a function of duration, and a slight decrease of the judgments of falling tones with duration. However, neither the effect of the duration nor the duration × direction interaction were significant ($p > .05$). Last, the overall difference between rising and falling tones estimates was slightly decreased

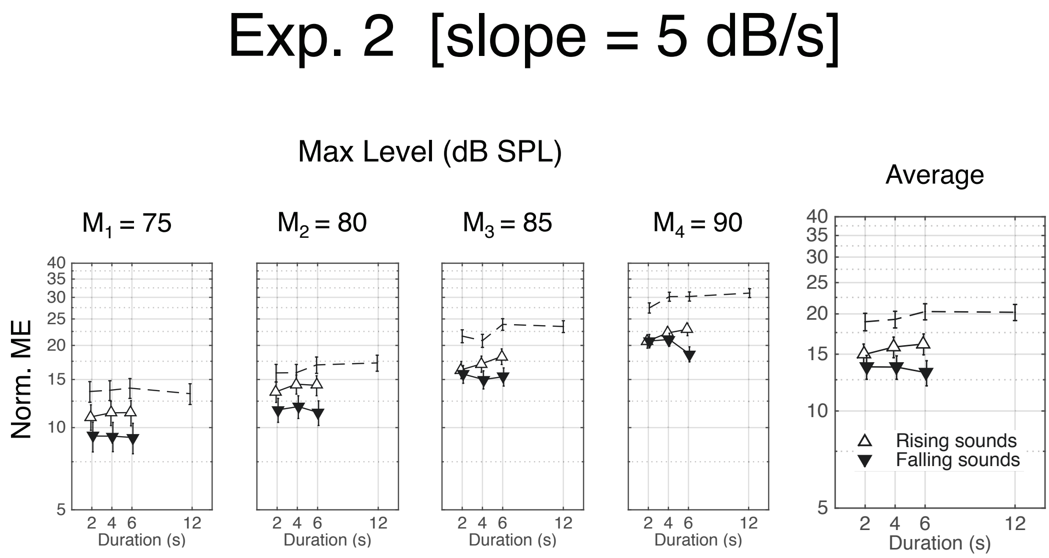

Figure 3.5: Normalized estimates of global loudness for rising and falling sounds obtained in Experiment 2, for ramps whose slope was 5 dB/s. Same legend as Figure 3.4.

as the maximum level of ramp increased; there was a significant direction × maximum level interaction $[F(3, 84) = 3.14, p = .031, \eta_p^2 = .101, \tilde{\epsilon} = .96]$. Post-hoc analyses revealed that this interaction disappeared when ramps presented at M_4 (i.e. [75-90 dB SPL]) were discarded, indicating that the reduction was caused by the ramps having a maximum level of 90 dB SPL.

In order to specifically examine the influence of the duration on each ramp direction and to assess the effect of the slope on their global loudness judgments, we conducted additional analyses on rising and falling ramps separately. Ramps varying at 2.5 dB/s and 5 dB/s were combined for the set of durations shared between these two groups, i.e., 2, 4 and 6 s (the 12-s ramps varying at 2.5 dB/s were thus not considered in these analyses). The data separated for rising and falling ramps are presented in Figure 3.6 (upper panels: rising tones; lower panels: falling tones).

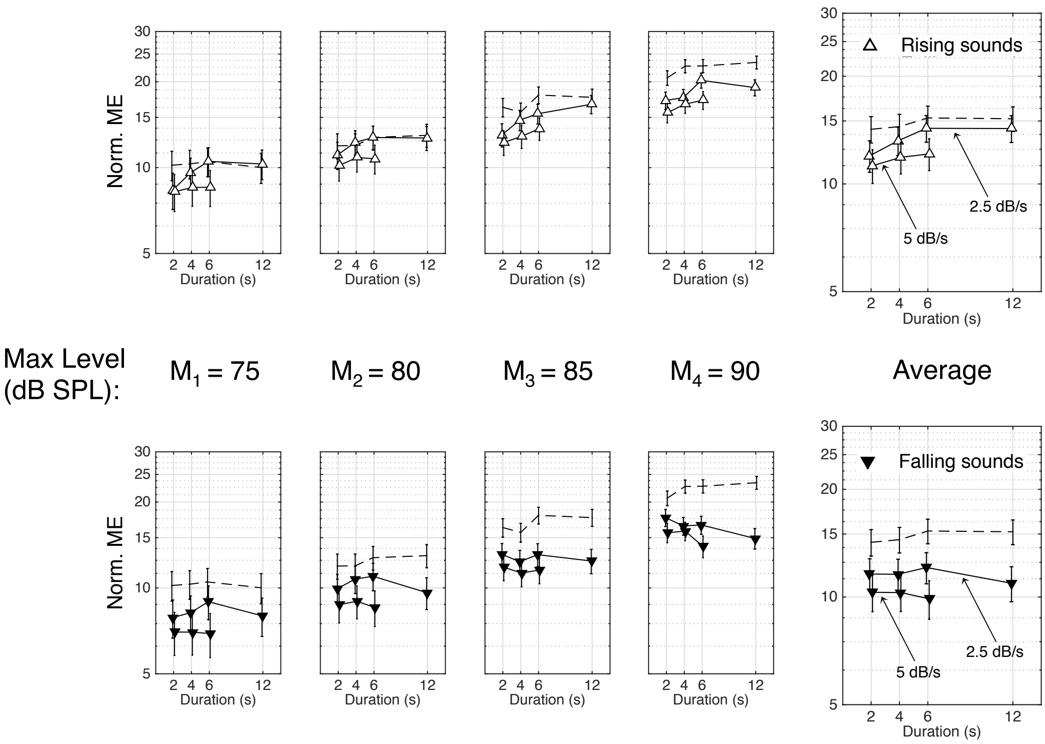

Figure 3.6: Normalized estimates of global loudness collected in Experiment 2 for rising sounds (upper panels) and falling sounds (lower panels). This figure combines the results plotted in Figures 3.4 and 3.5 to provide a clearer picture of the influence of the slope on rising and falling ramps respectively. It should be seen as a visual support to Analyses D and E. On each panel, highest triangles always correspond to the estimates given to 2.5 dB/s ramps (i.e. from Fig. 3.4) or to 5 dB/s ramps (i.e. from Fig. 3.5). Otherwise, the plotting convention is the same as in Figures 3.4 and 3.5.

3.3.2.4 Analysis D - Global loudness of rising ramps varying at 2.5 dB/s and 5 dB/s

This analysis only involved global loudness judgments of rising ramps lasting 2, 4 and 6 s and varying at 2.5 dB/s or 5 dB/s. Global loudness estimates of rising ramps varying at 2.5 dB/s were clearly higher than those given to rising ramps having the same maximum level and duration but varying at 5 dB/s; a large and significant effect of the slope supported this observation

$[F(1,28) = 31.92, p < .001, \eta_p^2 = .533]$. There was no significant effect of the duration ($p > .05$) but a significant duration × slope interaction $[F(2,56) = 4.57, p = .015, \eta_p^2 = .140, \tilde{\epsilon} = 1.06]$, revealing that the effect of the duration, although not significant as a main factor, was different for the two slopes.

3.3.2.5 Analysis E - Global loudness of falling ramps varying at 2.5 dB/s and 5 dB/s

This analysis involved only the global loudness judgments of falling ramps lasting 2, 4 and 6 s and varying at 2.5 dB/s or 5 dB/s. Global loudness estimates of falling ramps varying at 2.5 dB/s were again clearly higher than those given to falling ramps having the same maximum level and duration but which varied at 5 dB/s, as supported by a large and significant effect of the slope $[F(1,28) = 5.94, p < .001, \eta_p^2 = .680]$. There was no significant effect of duration ($p > .05$) but again, a significant duration × slope interaction $[F(2,56) = 4.38, p = .024, \eta_p^2 = .135, \tilde{\epsilon} = .83]$.

3.3.3 Discussion

The overall analyses (i.e. Analyses B and C) showed that the effect of duration as a main factor was not significant. Therefore, as compared to when the "time-stretching" was made at constant *dynamics* (Exp. 1), where a large main effect of the duration was found, a "time-stretching" at constant *slope* did not strongly affect global loudness. As discussed in the introduction, this result would be compatible with the *integration mechanism*. Moreover, Analyses D and E showed that the ramps varying at 2.5 dB/s were perceived clearly louder than those varying at 5 dB/s, a result that also appears to be consistent, at least at a qualitative level, with the *integration mechanism*, given that more energy is contained in an integration window for ramps varying at 2.5 dB/s as compared to ramps varying at 5 dB/s. The fact that the global loudness estimates of ramps were always below or equal to the loudness estimates of constant tones presented at their maximum level provide a substantial support to the *integration mechanism* too.

Although the effect of duration was not significant overall, a slight increase in the judgments of rising tones and a slight decrease in the judgments of falling tones as a function of duration could be noticed, respectively. This duration × direction interaction could be observed for both slopes but it was only significant for ramps varying at 2.5 dB/s. In what follows, we propose two scenarios, (a) and (b), to attempt to explain these observed interactions.

(a) A first way to interpret this interaction is to assume that they reflect the two proposed mechanisms to a certain extent. First, the slight decrease observed for falling tones would be compatible with the view that the loudness integrated on falling tones decay as a function of time between the loudness peak and the moment when the judgment is made; i.e. that a *decay mechanism* might be involved. However, the influence of duration was not significant when the falling ramps were analyzed separately (i.e. Analysis E). Second, the slight increase of global loudness with duration observed for rising tones could actually reflect a slight deviation of loudness judgments with duration. The fact that we also observed something comparable – although not significant (cf. Analysis A) – for constant tones between 1 and 6 s could be taken as a support to this assumption. Moreover, in a study of Canévet et al. (2003) in which listeners were asked to estimate not the global loudness but the final loudness of 1-kHz rising and falling ramps ([60-75 dB SPL]) lasting 1.8, 10 or 50 s, it was found that loudness estimates of the end levels of rising ramps significantly increased with duration from 1.8 s to 10 s and then remained constant between 10 and 50 s. Thus, the slight increase observed in the present experiment would not be specifically related to the global loudness of rising tones *per se* but would rather merely reflect a slight "deviation" of listeners' estimate in their loudness evaluation with duration. As for falling ramps, this small effect of duration on global loudness of rising ramps was however not significant when the ramps were analyzed separately (i.e. Analysis D).

(b) Another way to interpret the duration × direction interaction is to assume that it cannot be imputed neither to the *integration mechanism*, nor to the *decay mechanism*, but that it reflects how the (other) mechanism(s) responsible for the asymmetry between rising and falling sounds is (are) modulated when the ramps are stretched in time at constant slope. Indeed, the time stretching adopted in the present experiment also increases the *dynamics* of the ramps, which thus makes rising and falling profiles more and more different with duration. This would be consistent with the fact that the asymmetry was apparently increased with duration. At shortest durations, the dynamics of the ramps were very small (5 dB) and thus very close to constant tones, whereas when duration increased, their dynamics increased and then would increased the asymmetry between the two profiles. Furthermore, the significant duration $\times$ slope interactions revealed by Analyses D and E indicate that the effect of the stretching would not be the same whether the ramps vary at 2.5 or 5 dB/s, which could suggest that other mechanisms related to the asymmetry between the two profiles are involved and modulated.

All in all, with the present experiment, it is not possible to determine why the duration $\times$ direction interaction was observed and which scenario, (a) or (b), is more plausible. Therefore, this experiment did not allow to demonstrate that a *decay mechanism* was involved in global loudness evaluations of long falling ramps. The fact that significant interactions were obtained nevertheless attest the presence of significant differences in the processing of rising and falling profiles; and a decrease of a decrease of the estimates of falling sounds could be noticed between 4 and 6 s for 5-dB/s ramps and between 6 and 12 s for 2.5-dB/s ramps.

Last, a significant interaction between the direction of the ramps and their maximum level was found for ramps varying at 5 dB/s (with a maximum level of 90 dB SPL) but not for ramps varying at 2.5 dB/s, revealing that the reduction of the asymmetry at the highest intensity-region can be observed specifically for ramps varying at a certain rate of change. We have no explanation of this result.

3.4 General discussion and conclusion

3.4.1 Summary of the present findings

In two magnitude estimation experiments, we addressed whether two mechanisms, which had been proposed as potential candidates in previous studies, might indeed be involved in the global loudness evaluations of rising and falling ramps (Susini et al. 2007; Meunier et al. 2010). The first mechanism under study, called *integration mechanism*, relies on the assumption that the global loudness of a ramp could be determined by an integration of its loudest portion over a certain temporal window. The second mechanism investigated, called *decay mechanism*, is based on the assumption that the output of this loudness integration should gradually decrease until the end of the sound when the judgment is made. The plausibility of these mechanisms was determined by looking at the extent to which global loudness judgments of rising and falling tones (with linear level changes) were influenced by different manipulations of their parameters, namely their *slope*, their *dynamics* and their *duration*. It should be noted that disentangling perceptual mechanisms in place with such stimuli by "stretching" their parameters (slope, duration, and dynamics) is complex because these parameters are not independent; the slope is indeed equal to the dynamics divided by the duration. In both experiments, several parameters varied simultaneously and might thus have tainted the results such that it is not possible to make "clear-cut" conclusions. However, the two experiments yield various results that allow to further discuss the plausibility of the two proposed mechanisms.

Overall, the results obtained in this study provide some support to the hypothesis that an *integration mechanism* might be involved. First, the stretching adopted in Exp. 1 caused global loudness to increase for both rising and falling ramps, which is consistent with the fact that the energy contained within a fixed temporal window located around the peak stimulus is increased. Second, the stretching adopted in Exp. 2 led to (1) non-significant effects with respect to duration, consistent with the fact that the energy contained within a fixed temporal window remains unchanged, and

(2) a large and significant effect of the slope in Exp. 2 both for rising and falling tones (ramps varying at 2.5 dB/s were significantly louder than ramps varying at 5 dB/s). This is consistent with the fact that less energy is contained in a window located under ramps having steeper slopes. There are however two departures from this mechanism that can be noticed: (1) the "saturation" of the estimates of rising tones beyond 6 s observed in Exp. 1 and (2) the slope × duration interaction obtained in Exp. 2. As discussed earlier, since the stretching adopted in Exp. 1 induced both variations in duration and slope at the same time, it might be possible that another mechanism was involved in the evaluation of the ramps of long durations and that this mechanism was responsible for the "saturation" in the judgments beyond 6 s, because the ramps had very small slope and could possibly be assimilated to constant tones. In that sense, the saturation would finally not be imputable to the *integration mechanism* itself. Note that we only examined *qualitatively* the extent to which our results agreed with an integration mechanism; all these results remain to be verified *quantitatively*, for instance whether a loudness integration over a fixed temporal window is compatible with the rate of increase observed in Exp. 1.

The presence of the *decay mechanism* was assessed by comparing the estimates of falling tones with those of rising sounds for which only integration take place. We found small but significant duration × direction interactions both in Exp. 1 and Exp. 2 (only for the ramps varying at 2.5 dB/s), which could be taken as evidence that a certain *decay mechanism* takes place. However, these interactions were very small and only significant in certain cases (between 2 and 6 s in Exp. 1; only for ramps varying at 2.5 dB/s in Exp. 2). In particular, the estimates collected for falling ramps in Exp. 2 were supposed to be a direct image of this *decay mechanism* but the small decline as a function of duration that we could observed did not reach significance (using post-hoc tests). Therefore, although the presence of a *decay mechanism* in global loudness judgments of falling ramps cannot be ruled out, it would be too early to argue from this study that it is involved. For example, why would it be significant only between 2 and 6 s, as revealed by Exp. 1A? Such a question deserves further attention for future studies.

No fairly strong conclusions can be drawn from the present data yet, but the hypothesis that a certain *integration mechanism* is involved seems reasonable. Global loudness judgments of ramps appear to be primarily guided by their slope (in addition to their maximum level, obviously), much more than by their duration or dynamics. Besides, the size of the asymmetry between rising and falling ramps depends on the duration of the ramps (cf. Exp. 1, where asymmetries were increased with the duration) and on their dynamics (cf. Exp. 2, where the asymmetries were clearly reduced at shortest durations, corresponding to small dynamics). Therefore, the present study shows that the asymmetry between rising and falling tones is not specific to the 2-s, 15-dB ramps employed in previous studies (e.g., Ponsot et al. 2015; Ponsot et al. in press); it occurs in many other conditions but its magnitude depends on the parameters of the ramps. Finally, the asymmetries were again reduced in highest intensity-regions (maximum levels at 90 dB SPL) in some conditions (in Exp. 1 for Group A and in Exp. 2 for the ramps varying at 5 dB/s). We have no explanation why this reduction occurred and why it only occurred in certain cases.

3.4.2 Predicting global loudness from Glasberg and Moore's model outputs

We evaluated to which extent Glasberg & Moore's loudness model (2002) could account for the present results. Because of the numerous conditions investigated in the present study, we will only focus on [65-80 dB SPL] ramps, for durations ranging from 1 to 16 s, corresponding to those employed in Exp. 1. First, it should be recalled that Glasberg and Moore's model (Glasberg and Moore, 2002) does already produce asymmetries between rising and falling tones, as shown in Ponsot et al. (2015). However, the asymmetries given by the indicators proposed to predict global loudness so far, i.e. the maximum of Short-Term Loudness or Long-Term Loudness patterns, considerably underestimate what was measured by means of various psychophysical experiments (see Ponsot et al. 2015). The asymmetries produced by the model are due to the two temporal integration stages employed to derive, first, Short-Term Loudness (STL) and second, Long-Term

Loudness (LTL). To illustrate this aspect, we plotted the LTL outputs of the model obtained for the [65-80 dB SPL] ramps at the most extreme durations used in Exp. 1 (i.e. 1 s and 16 s) in Figure 3.7. From this plot, it can easily be observed that while the two temporal integration stages produce asymmetrical patterns for 1-s ramps, the asymmetry becomes much weaker for 16-s ramps.

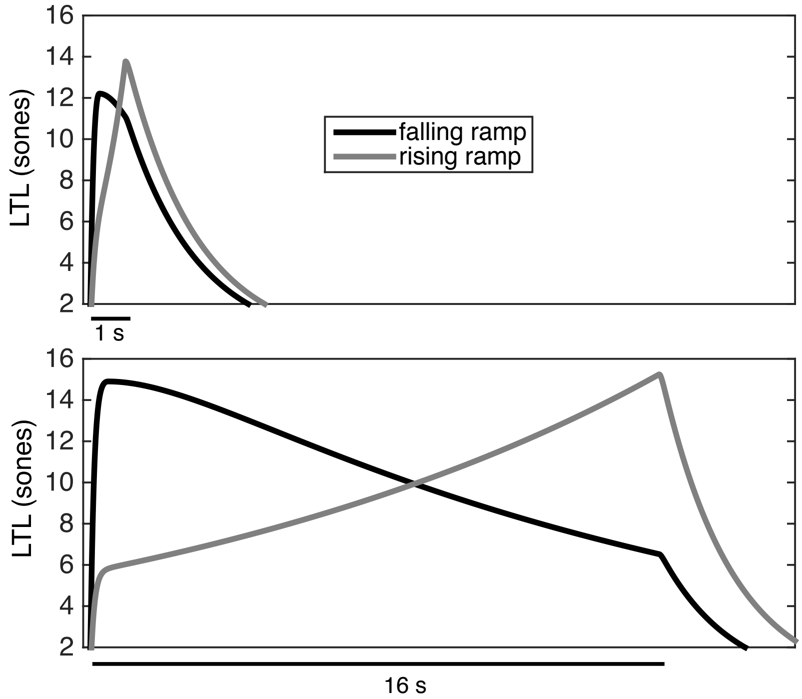

Figure 3.7: Long-term loudness (LTL, in sones) of rising-intensity (grey lines) and falling-intensity ramps (black lines) predicted by Glasberg & Moore's model (2002). The ramps tested as inputs are composed of 1-kHz tones with linear level profiles (in dB) covering the [65–80 dB SPL] region, with 10-ms linear rise and fall times. The upper panel shows the predictions for 1-s ramps. The lower panel shows the predictions for 16-s ramps.

In the present paper, we examined four basic indicators to predict global loudness, all based on the short-term and the long-term loudness time series outputs provided by Glasberg & Moore model. Two of them are those suggested by Glasberg & Moore, STL_{max} and LTL_{max}, namely the maximum of STL and LTL patterns, respectively, which were already tested in our previous studies (Ponsot et al. 2015; Ponsot et al. in press). The third indicator that we tested was based on the *integration mechanism* partially supported by the results of the present study. We thus introduced STL_{int}, corresponding to the short-term loudness (i.e. the loudness perceived at any instant) *averaged* over a fixed temporal window located around its maximum. Since we had no specific assumptions concerning the shape of this temporal window, we used a simple rectangular temporal window. The length of this window was chosen equal to 500 ms (in order to roughly account for the growth of global loudness estimates with duration obtained in Exp. 1). The fourth indicator, which was already introduced by Pedersen (2006), corresponds to the value of the long-term loudness at the moment where the stimuli end. This indicator was referred to as LTL_{end}. Indeed, because the long-term loudness might be seen as a variable reflecting the "overall loudness at any moment", it might be particularly relevant to take advantage of the fact that this value will decrease slowly over time for falling ramps. Thus, taking the value at the moment when the stimulus ends could mimic the effect of the decay mechanism and will consequently provide larger asymmetries than those obtained with indicators based on the maxima or on loudness integration only. The way these four indicators are computed from STL and LTL outputs, respectively, is presented in Figure 3.8.

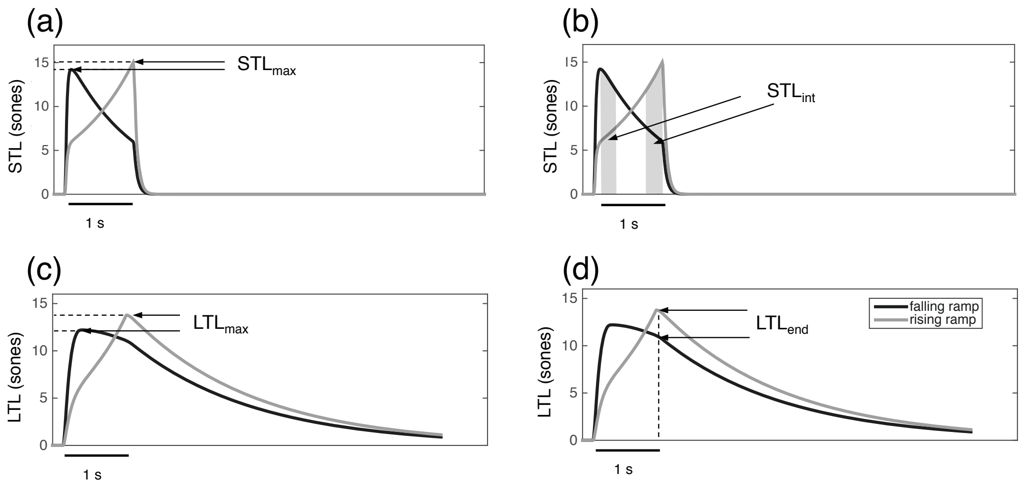

Figure 3.8: Presentation of the different indicators introduced in the present study to estimate global loudness from Glasberg and Moore's model (2002) outputs. Both short-term loudness (STL) and long-term loudness time series of 1-s, [65-80 dB SPL] rising-intensity (grey lines) and falling-intensity ramps (black lines) predicted by the model are considered. (a) Using the maximum of the short-term loudness (STL_{max}). (b) Using an average of short-term loudness over a fixed temporal window located around the maximum (STL_{int}) (note that the temporal windows are not to scale). (c) Using the maximum of the long-term loudness (LTL_{max}). (d) Using the long-term loudness value at the moment when the ramps end (LTL_{end}).

The global loudness predictions made by these four indicators at all ramp durations, ranging from 1 to 16 s, are plotted in Figure 3.9. As mentioned earlier, STL_{max} and LTL_{max} give rise to small asymmetries at short durations but these asymmetries disappear at longer durations, since they are caused by the temporal integration stages. The increase with duration with STL_{max} is negligible, whereas as expected a substantial and logarithmic increase is observed with STL_{int}. A significant increase with duration is also observed with LTL_{max} but only for falling ramps. The curves predicted by these indicators all seem to reach a plateau after 6 s. STLint does not predict any asymmetries between rising and falling tones. As expected, LTL_{end} products much larger asymmetries and also leads to a huge decrease of global loudness of falling tones as a function of their duration.

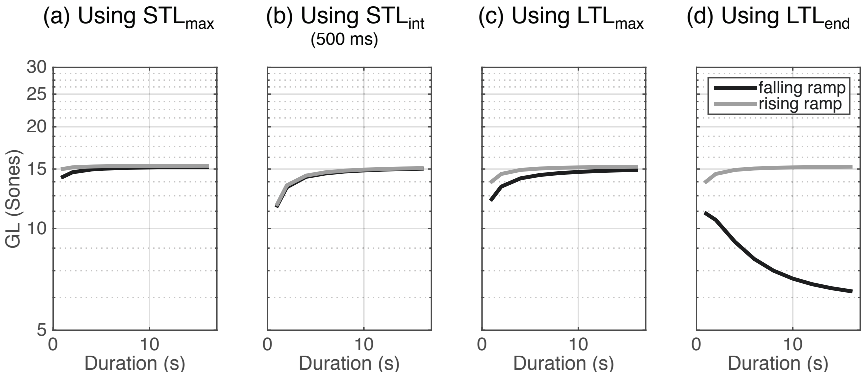

Figure 3.9: Global loudness of 1-kHz, [65-80 dB SPL] ramps (falling ramps, black lines; rising ramps, grey lines) of various durations (from 1 to 16 s) estimated with different indicators derived from STL and LTL outputs provided by Glasberg and Moore's model (2002). Global loudness predicted (a) using STL_{max}, (b) using STL_{int} with a 500-ms rectangular integration window, (c) using LTL_{max} and (d) using LTL_{end} (see the text and Fig. 3.8 for further details).

In order to compare the predictions provided by these indicators with the results obtained in Experiment 1, we superimposed our data and the predictions from these indicators in Figure 3.10. The normalized estimates collected in Experiment 1 were supposed to reflect loudness values in sones, but these values were conditioned by the fact that (1) the exponent of the loudness function measured on the listeners had to be equal to 0.6 and (2) listeners had to use the same scale (i.e. range of numbers) between the measurement of the loudness function and the ramps, which were conducted in two different blocks. Because these conditions could not be verified in the experiment, it is not possible to interpret the normalized estimates presented in Fig. 3.1 as loudness values in sones. Consequently, the discrepancies between the mean loudness provided by the model for these ramps (about 15 sones) and the mean loudness obtained in the experiments (about 10 sones) cannot be directly interpreted. Rather, we addressed whether the different indicators could account for the trends of the predictions as a function of duration. Global loudness values provided by the model were thus rescaled by an arbitrary factor of 1.4 to provide easier visual comparisons. The data and the rescaled model predictions are presented in Figure 3.10. First, with respect to the effect of the duration, it can be observed that the best fit is obtained with STL_{int}. Indeed, the increase between 1 and 2 s is comparable to what was measured and the logarithmic growth well approximates the fact that a plateau was observed experimentally beyond 6 s. It is clear that STL_{max} is not appropriate to account for the increase with duration and that, although LTL_{max} shows an increase that would fit reasonably well the data for falling sounds, it does not work for rising sounds. Obviously, LTL_{end} cannot reproduce the increase with duration. Second, with respect to the asymmetry rising vs. falling, it is clear that none of these indicators are able to account for the magnitude of the effect observed in the experiment. While LTL_{end} better predicts the asymmetries between 1-s ramps, it considerably overestimates the size of the effect for longer durations and predicts a significant decrease of loudness for the falling tone; which is opposite to what was observed experimentally. This type of indicator should be interesting to reflect a *decay mechanism* but the predicted decrease would likely be too abrupt.

Therefore, these analyses suggest that, apart from the asymmetry between the two profiles observed at all duration that cannot be reproduced by any of these indicators, STL_{int} based on an integration of the short-term loudness contained in a 500-ms temporal window seems to be able to reproduce the main trend observed with duration in Exp. 1 with Group A. However, since this 500-ms integration window predicts an increase of loudness mostly until 4 s, it would certainly not fit the data obtained in Exp. 1 with Group B where a clear increase of global loudness was obtained between 4 and 8 s. The 500-ms length of this integration window seems to be appropriate only in the present case (i.e. Exp 1, Group A) and should thus not be taken as a result per se.

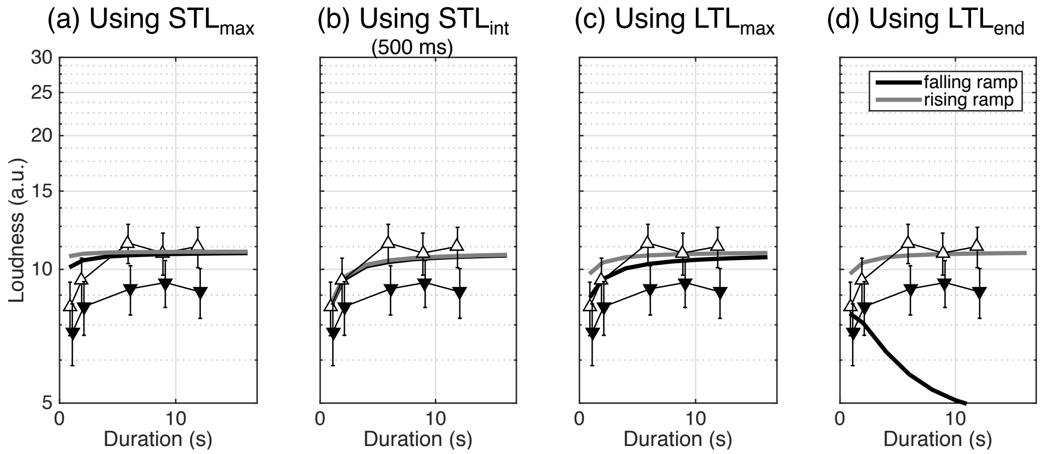

Figure 3.10: Global loudness estimates obtained in Experiment 1 (Group A) for [65-80 dB SPL] ramps superimposed with global loudness values given by the four indicators introduced in this article, which have been arbitrary rescaled (by a factor of 1.4) for a visualization purpose.

3.4.3 Conclusion and perspectives

Overall, the present study provided evidence to the hypothesis that a certain *integration mechanism* might be involved in the global loudness evaluation process of rising and falling intensity tones. We were however not able to support our hypothesis that a *decay mechanism* might also be involved. Analyses of different indicators based on Glasberg & Moore's model outputs show that the indicators most often used to predict global loudness, namely STL_{max}, LTL_{max}, are not able to reproduce most of the trends we observed experimentally. In particular, the increase with duration obtained with the time stretching at constant dynamics employed in Exp. 1 cannot directly be predicted by taking the maximum of short-term or long-term loudness time series provided by the model. We showed that STL_{int} can overall fit the increase obtained in Exp. 1, indicating that global loudness could be compatible with a certain integration of short-term loudness over a window, which was around 500-ms in that case (i.e. Exp. 1, Group A). However, further analyses have to be undertaken to confirm that the *integration mechanism* is indeed involved, and if this is the case, to determine the shape and the length of such temporal window. Indeed, our results suggest that the size of this integration window would depend on the ramp parameters or on the experimental context (given that it seems to differ between the different experiments of this study). Nevertheless, none of the four indicators investigated were able to account for the asymmetries observed between rising and falling tones and, consequently for the fact that this asymmetry depends on ramp parameters such as the dynamics. Therefore, even though an *integration mechanism* could be part of the global loudness evaluation process of rising and falling ramps, there are certainly other mechanisms involved that still remain to be determined, e.g. the *decay mechanism* should be further examined. According to Moore (2014), long-term loudness is supposed to reflect “relatively high-level cortical processes and involves memory”. Our results suggest this might not be exactly true. The computation of global loudness from short-term or long-term loudness patterns seems to be more complex than just considering their maximum value. It is very likely that the processes involved in global loudness evaluation of rising and falling tones of a few seconds are presumably part of high-level integration stage not yet totally reflected by LTL, which is simply based on a temporal integration of STL (reflecting the loudness at any instant). Therefore, further studies are required to determine the sensory and perceptual processes involved in global loudness evaluations of time-varying sounds and how they are governed by the physical characteristics of the stimuli.

Are these mechanisms specific to loudness evaluation? There are some studies in the literature that would suggest that the processes examined here in the case of global loudness evaluation might actually be involved in overall judgments of other types of sensory information. For example, although the time scale and the amount of sensory change are not comparable, the results obtained for the overall evaluation of increasing and decreasing sequences of pain yielded similar trends to those observed in the present study (Ariely and Carmon, 2000). In particular, increasing sequences are judged as more painful that decreasing sequences and the slope of the sequence play a significant role. Works addressing overall annoyance evaluation of aversive sounds (e.g. Kahneman et al. 1993) or the overall evaluation of image quality

(Hamberg and de Ridder, 1999) have brought to light important principles governing overall evaluation such as "peak-end" rules and "recency effects", which are also obtained in global loudness judgments of time-varying sounds (Västfjäll 2004; Susini et al. 2002). Could an *integration mechanism* be at the basis of any overall evaluation of every rising or falling pattern? Whether the mechanisms underlying global loudness evaluation are also involved in the overall evaluations of other sensory attributes is an aspect that deserves to be specifically addressed in future studies. As pointed out by Dan Ariely (1998): *" [. . .] although this work examines only one domain of experience (namely pain), one can speculate that the relationship between momentary and overall evaluations will apply to other domains as well."* Reinforcing our knowledge of psychoacoustics with the investigation of higher-level integration mechanisms related to general principles of time-varying information processing would constitute a considerable research advance, potentially providing important information as regards the mechanisms underlying global evaluation of time-varying loudness at this timescale.

Acknowledgments

The authors would like to acknowledge Anne-Laure Verneil for her help in conducting the experiments. This work was supported by the project *LoudNat* funded by the French National Research Agency (ANR-11-BS09-016-01). Some of the results contained in this article were presented at the 30th meeting of the International Society for Psychophysics, *Fechner Day*, in Lund, Sweden, August 2014.

Partial synthesis

The present chapter explored the relationships between listeners' global loudness evaluations of rising and falling-intensity 1-kHz tones and their temporal profile characteristics. More specifically, we manipulated different parameters (*duration*, *slope*, *dynamics*) characterizing the linear level profiles of these stimuli ramping up or down in order to determine the potential implication of two mechanisms presumed to be at the basis of their global loudness evaluations: the *integration mechanism* and the *decay mechanism*.

Overall, our results provided evidence that a certain *integration mechanism* might be involved in global loudness evaluations of both rising and falling sounds. Indeed, global loudness evaluations of rising and falling tones were primarily determined by their *slope* as shown in both Exp. 1 and Exp. 2, and thus closely linked to the amount of loudness contained in a temporal window located under their loudest portion. We found significant duration $\times$ direction interactions and observed slight decreases of global loudness judgments for falling tones in both experiments, suggesting that not only an integration but another mechanism might also be involved. It remains to be determined whether these observed interactions and decreases could indeed be attributed to a *decay mechanism*, e.g. from the use of another experimental procedure yielding greater statistical power.

Although the presence of an *integration mechanism* was overall supported by our results, this mechanism alone could not account for the asymmetries found between rising and falling tones. While in previous studies (see Chapter 1 and Chapter 2) the asymmetry was always determined for the same 2-s, ramps varying over 15-dB intensity-regions, the present study shows that the asymmetry also occurs for ramps of other durations (from 1 to 16 s), slopes (from around 1 dB/s to 15 dB/s) and dynamics (from 5 to 30 dB), but that the magnitude of the effect seems to depend on these parameters. In particular, we observed that the asymmetry seems to increase with the duration (Exp. 1) and the dynamics (Exp. 2) of the stimuli. The changes in magnitude of the asymmetry might actually reflect particular properties of the mechanisms underlying global loudness processing. We also observed that the asymmetry decreased with the mean intensity of the stimuli in certain cases. This latter finding is consistent with the intensity-region effects revealed in Chapters 1 and 2, but its causes still remain unclear.

In the modeling section of this chapter, we showed that the indicators most often used to predict the global loudness of time-varying sounds, i.e. STL_{max} and LTL_{max}, which correspond to the maximum of the short-term loudness and long-term loudness given by Glasberg Moore's model, respectively, could not directly account for the *integration mechanism*. Nevertheless, LTL_{max} would work better than STL_{max} to predict the increase of global loudness with duration (as obtained in Exp. 1) given that it relies on an additional temporal integration stage. To address the relevance of the presumed *integration mechanism*, we introduced another indicator that directly accounts for it. This indicator, called STL_{int}, is based on a temporal integration of the short-term loudness (STL) predicted by Glasberg & Moore's model. More precisely, STL_{int} is calculated as the average of the STL pattern contained in a rectangular temporal window located around the loudest portion of the sound. We showed in particular that it could then account for the rate of increase in global loudness judgments as a function of stimuli duration observed in Exp. 1, i.e. when the "stretching" was made at constant dynamics. It seems to reflect the *integration mechanism* to some extent.

However, we did not find any indicator yet capable of reproducing the asymmetries obtained at all durations. Although in Chapter 1, we showed that LTL_{max} could account for a portion of the asymmetries for 2-s ramps (1.3 dB over 4 dB), the present study shows this indicator would wrongly predict that the asymmetry decreases when the ramps are "stretched" in time like in Exp. 1.

All in all, the results of this study indicate that global loudness is more complex than what is typically assumed in current loudness models. Although the implication of the *integration mechanism* remains to be confirmed by further studies, it is interesting to note that it might reflect a common mechanism underlying global loudness processing of time-varying sounds in general, i.e. not only limited to the present case of rising and falling sounds. Besides, the asymmetries between rising and falling profiles seem to depend on the stimuli parameters in a complex way, as illustrated in Appendix D, which are not yet well understood. Note that all these results have been found with 1-kHz tones, and that it is not yet possible to affirm that the temporal mechanisms presently highlighted apply to stimuli with different spectral structures, as for example, broadband noises.

Chapter 4

Temporal loudness weights of rising and falling tones

Emmanuel Ponsot, Patrick Susini, Guillaume Saint Pierre, and Sabine Meunier. Temporal loudness weights for sounds with increasing and decreasing intensity profiles. *The Journal of the Acoustical Society of America*, 134(4):EL321– EL326, 2013.

Preface

The previous chapter (Chapter 3) explored how the overall parameters defining the temporal profiles of rising and falling-intensity stimuli (e.g., slope, duration, dynamics) shape their global loudness judgments. We found that the slope of the ramp is of primary importance, both for rising and falling tones. We indeed suggested that an *integration mechanism* could be at the basis of the observed effects (i.e., the increase of global loudness observed with a time-stretching at equal dynamics in Exp. 1 and the effect of the slope directly found in Exp. 2). However, asymmetries in global loudness between rising and falling tones were found at all ramp durations and remained unexplained by this *integration mechanism*. In the present paper, we addressed whether these asymmetries could be due to the fact that rising and falling sounds are weighted differently over time. Even for short ramp durations, rising sounds (which contain most of their energy in their final portion) might be judged louder than falling sounds (which contain most of their energy at the beginning) because the end of rising tones is more heavily weighted than the beginning of falling tones.

In the present chapter, we addressed this question by measuring the temporal weighting patterns underlying global loudness evaluations of rising and falling tones, respectively. We used the so-called "molecular psychophysics" (also known as psychophysical reverse-correlation; e.g., see Berg 1989; Ahumada and Lovell 1971) to quantify the respective contributions of the different temporal portions of these profiles into global loudness judgments. A "molecular" psychophysical experiment was thus conducted, which consisted of a global loudness judgment task including rising and falling tones with random level perturbations (varying from trial-to-trial) imposed on their mean level profile. Flat-intensity stimuli were also used to allow a comparison of our results with previous studies (e.g. Pedersen and Ellermeier 2008). These level-fluctuating stimuli with either flat, rising or falling profiles were presented in different blocks and listeners' task was to classify them as "soft" or "loud" with respect to their global loudness. Assuming that loudness judgments were based on a linear combination of the levels of each temporal portion of the stimuli (as it was done in previous studies employing this methodology; e.g., see Pedersen and Ellermeier 2008; Oberfeld and Plank 2011), the "perceptual weights" accorded to each temporal portion were derived from their responses using simple logistic regressions.

This work was published in *The Journal of the Acoustical Society of America*.

Ponsot, E., Susini, P., Saint Pierre, G., Meunier, S. (2013). Temporal loudness weights for sounds with increasing and decreasing intensity profiles. *The Journal of the Acoustical Society of America*, 134(4), EL321-EL326.

Additional analyses have been conducted after the publication of this work and are discussed together with the main findings of the paper in the partial synthesis section of this chapter. Appendix E presents a quantitative analysis of the results collected in the complementary experiment of the paper (see Section 4.3.3). In this appendix, the size of the asymmetry (in dB) was estimated for the stimuli employed in the experiment (i.e. time-segmented ramps) as a function of the presentation order of these ramps in the pair (rising-falling or falling-rising). In Appendix F, a second-order analysis of the main reverse-correlation experiment of the paper is conducted. Using identification tools for nonlinear systems, a computational model that attempts to simulate listeners' behavior during the task is also presented.

Abstract: Using molecular psychophysics, temporal loudness weights were measured for 2-s, 1-kHz tones with flat, increasing and decreasing time-intensity profiles. While primacy and recency effects were observed for flat profile stimuli, the so-called "level dominance" effect was observed for both increasing and decreasing profile stimuli, fully determining their temporal weights. The weighs obtained for these profiles were basically zero for all but the most intense parts of these sounds. This supports the view that the "level dominance" effect is prominent with intensity-varying sounds and that it persists over time since temporal weights are not affected by the direction of intensity change.

4.1 Introduction

Various studies have shown that the loudness of a sound could be affected by its temporal intensity profile. In particular, it has been found that sounds increasing in level over time – also referred to as up-ramps – are perceived as louder (Susini et al., 2007) or changing more in loudness (Neuhoff 1998; Olsen et al. 2010) than sounds decreasing over time – also referred to as down-ramps. Though several hypotheses have been proposed, the causes of these findings, involving perceptual phenomena specifically related to time-varying sounds, are still not clearly identified (e.g., Pastore and Flint 2011). It has been shown that the assessment of the global loudness of a time-varying sound is heavily guided by the maximum stimulus level and by its position towards the end of the sound (Susini et al. 2002; Susini et al. 2007). These results led us to the idea that the loudness difference observed between up-ramps and down-ramps stems from a dissimilar temporal loudness weighting between the two types of sounds.

So far, temporal weights of loudness have been assessed mainly for sounds with flat intensity profiles (e.g. Pedersen and Ellermeier 2008; Oberfeld and Plank 2011; Oberfeld et al. 2012). Using experimental designs based on molecular psychophysics (Berg, 1989), these studies show that the beginning and, to a lesser extent, the end of flat-intensity sounds receive the greatest temporal weights of loudness. In studies with non-steady profile sounds, a so-called "level dominance" effect was observed, the initial high-level components receiving greater attention from the listeners (Lutfi and Jesteadt 2006; Oberfeld and Plank 2011).

In the present experiment, the molecular approach was applied to the study of temporal weighting of loudness for sounds increasing and decreasing in intensity. We hypothesized that the high-level components would be more heavily weighted in the case of increasing intensity stimuli, as both "level dominance" and recency effects apply to them. A discrepancy in temporal weighting could thus help to explain why a loudness difference is observed between the two types of sounds.

4.2 Materials and method

4.2.1 Subjects

Eight volunteer participants (5 women and 3 men; age 21-29 years) took part in the experiment. None reported having hearing problems. They gave their informed written consent prior to the experiment and were paid for their participation. The participants were naïve with respect to the hypotheses under test.

4.2.2 Stimuli

Stimuli were 1-kHz pure tones lasting 2 s. They were all made of 16 consecutive 125-ms stationary segments with levels drawn independently from normal-truncated distributions (SD = 2 dB, restricted to Mean ± 5 dB). The mean values (in decibels) of the distributions were chosen in order to follow one of the three profiles under study. While increasing and decreasing profiles were examined for the particular purpose of this study, the flat profile was studied mainly to serve as reference in the following analyses and in order to compare our results with those of previous studies.

The means of the distributions from which the segment levels of the flat-level profile stimuli were drawn were all equal to 80 dB SPL, resulting in level-fluctuating sounds, as shown in the left panel of Figure 4.1. Regarding the increasing intensity stimuli, the means of the 16 distributions were chosen so that they created a globally increasing profile ranging from 65 to 80 dB SPL (see Fig. 4.1, middle panel). This profile was temporally reversed to create the [80-65 dB SPL] decreasing intensity stimuli (Fig. 4.1, right panel). Linear rise and fall times of 10 ms were imposed on the beginning and end of all stimuli. Inter-segment level variations were smoothed using 10-ms half periods of sinusoidal functions to avoid spectral splatter that occurs when there are abrupt changes in the sound intensity of pure tones. A fixed level increment of 0.9 dB was subtracted or added to each segment to create respectively low and high versions of these stimuli for each profile (see Figure 4.1). This value was chosen according to a preliminary experiment

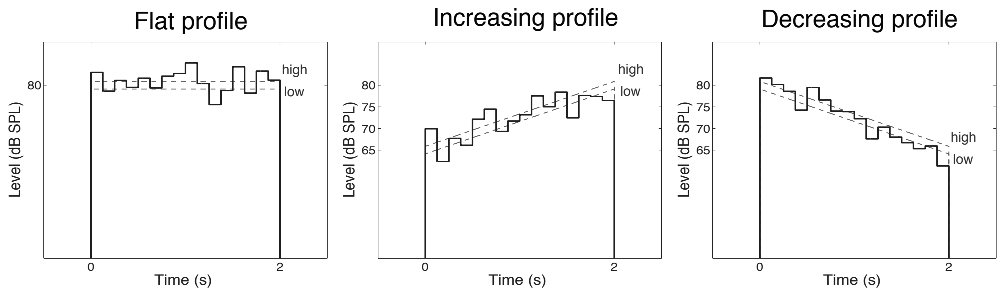

Figure 4.1: The stimuli used in the experiment consisted of sixteen 125-ms steady segments assembled end-to-end to follow one of the three mean profiles under study: flat [80 dB SPL], increasing [65-80 dB SPL] and decreasing [80-65 dB SPL]. For each stimulus, the segments were "randomly" varying in level (for details, see Sec. 4.2.2) either around a low or a high version of the intensity profile (± 0.9 dB around the mean profile, represented by the dashed gray-lines).

that indicated a 1.8 dB difference in mean level between the low and high versions of the stimuli (i.e., mean profile ± 0.9 dB) resulting in a mean discrimination score of 65 % for the three profiles.

4.2.3 Apparatus

The stimuli were generated at a sampling rate of 44.1 kHz with 16-bit resolution using Matlab. Sounds were converted using a RME Fireface 800 soundcard, amplified using a Lake People G-95 Phoneamp amplifier and presented diotically through headphones (Beyerdynamic DT 270 PRO). Levels were calibrated using a Brüel & Kjær 2238 Mediator sound-level meter placed at a distance of 4 cm from the right (left) earphone. Each participant was tested in a double-walled IAC sound-insulated booth.

4.2.4 Procedure

A standard 1I, 2AFC procedure was employed. The experiment was divided into six different sessions scheduled on different days. Each session was made up of 100 training trials followed by three blocks of 200 trials. Each profile (i.e., flat, increasing and decreasing) was presented in a particular block. In each trial, a sound randomly chosen to be a low or a high version of the corresponding profile was played to the participant. The participants had to determine whether the stimulus seemed to be soft or loud, on the basis of preceding stimuli listened during the block. It was an intensity discrimination task in which subjects were explicitly asked to consider the global loudness of the stimulus (corresponding to the judgment of the loudness over the duration of the sound) when making their judgment. In addition, they were instructed not to consider the loudness of the stimuli presented in previous blocks (i.e. other profiles) when making their judgment. All answers were entered directly into a Matlab interface. Participants became familiar with the experimental procedure by completing a training session at the beginning of the experiment. The blocks were presented in random order to each subject in each session. As we were interested in spontaneous and natural strategies, participants did not receive trial-by-trial feedback. They were only informed of their score (percentage of correct identifications of low / high incoming distributions) at the end of each block to ensure an optimum focus on the task in hand. Each session lasted approximately 1 hour.

4.2.5 Data analysis

We opted for a logistic regression analysis for weights estimation and employed the same decision model as in previous studies (Pedersen and Ellermeier 2008; Oberfeld and Plank 2011). Regressions were conducted using R (R Core Team, 2015) separately for each listener, each intensity profile and each distribution (low / high). Responses given in each trial served as the dependent variable. Soft responses were coded as 0 and loud responses coded as 1. The levels of each segment were taken as the 16 independent variables.

4.3 Results

The temporal weights, defined as the beta coefficients given by logistic regressions, were normalized individually for each profile and distribution. First, a within three-factor (segment by profile by distribution) repeated-measures analysis of variance (RM ANOVA) was performed on the full data set. Since a significant main effect of the profile was found $[F(2,14) = 6.26; p = .01]$ but no significant interaction between profile and distribution $[F(2,14) = 0.38; p = .69]$, temporal weights of the two distributions were pooled together afterwards.

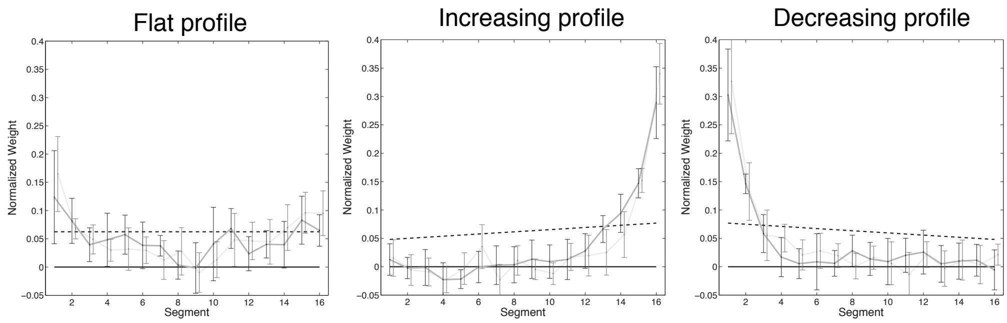

Figure 4.2: Normalized temporal weights obtained for each profile and each distribution as a function of the segment position (1-16). The low distribution corresponds to the light grey line and the high distribution to the dark grey line. Results are averaged over participants (N=8). The dashed-lines show the theoretical temporal weights from an "ideal observer" that would give equal attention to all the segments, and whose sensitivity would only be set by the near-miss effect. Error bars show the 95 % confidence intervals (regarding the population of participants) for each segment and distribution.

4.3.1 Temporal weights for the flat profile

The left panel of Figure 4.2 shows the mean temporal weights obtained for the flat profile as well as the temporal weights from an "ideal observer" (Berg, 1989), where the same weight would apply to each segment (represented by the dashed lines). The results show that the first and last segments received the highest weights from the listeners, resulting in a u-shaped temporal pattern. The middle-part of the profile was virtually ignored, as observed in the left panel of Figure 4.2 where the zero weight fell within the 95 % confidence intervals associated to most of these segments. Two-tailed Wilcoxon signed rank tests were conducted to compare the weights of segments 1 and 16 respectively with the mean weight of the middle-part segments (segments 6-11). Both primacy (W=34; p = 0.01) and recency (W=33; p = 0.02) effects were found to be significant.

4.3.2 Temporal weights for the increasing and decreasing profiles

The mean temporal weights obtained for the increasing and decreasing profiles are plotted, respectively, in the middle and right panels of Figure 4.2. For the increasing profile, only the weights of the last three segments differed from zero, which supports the assumption made in the introduction that the last portion of an increasing sound – corresponding to both its maximum and end-level – would receive the greatest attention in the task. Similarly, for the decreasing profile, all but the first three segments received a zero weighting. This latter result also supports the hypothesis that the highest weights would have been assigned to the beginning of the down-ramps, that is, the loudest portion. A common effect of saliency for the two profiles was obtained, where the most intense segments received exclusive attention from the listeners. This effect was also pointed out in other studies (Lutfi and Jesteadt 2006; Pedersen and Ellermeier 2008; Oberfeld and Plank 2011) that were interested in temporal weighting of non-steady profiles, and it was defined, as mentioned above, as the "level dominance" effect.

The weights from the "ideal observer" were also computed (using data from Rabinowitz et al. 1976) in order to take into account the fact that the sensitivity for level differences of pure tones is slightly better at high than at low levels; an effect known as the near-miss to Weber's law. Ideal weights were calculated so the sensitivity at 80 dB SPL (compared to 65 dB SPL) was increased by a factor around 1.6, and that the sum of the 16 weights should be equal to 1 (with normalization). As it can be seen in the middle and right panels of Figure 4.2, the near-miss effect can only marginally account for the trend obtained for the weighting profiles.

As the main motivation for this study was to account for the loudness difference between up-and down-ramps, their perceptual weighting patterns were compared. Our hypothesis was that the weights for the most intense components would be greater for the up-ramp than for the down-ramp. Thus, the weights of the last segments of the up-ramp were compared with the weights of the first segments of the down-ramp (simply by reversing their order to statistically compare them with up-ramp weights) using a within two-factor (profile by segment) RM ANOVA. As a result, the Segment × Profile interaction was not found to be statistically significant [$F(15, 210) = 0.61; p = 0.87$]. Thus, contrary to the assumption made in the introduction, it cannot be inferred from the results of the present experiment that high-level components are weighted more heavily for up-ramps than for down-ramps.

4.3.3 Loudness difference between up-ramps and down-ramps

In an additional experiment, we checked that the lack of asymmetry found between the two weighting patterns was not due to a failure to produce a loudness difference between the two types of time-segmented ramps. A new group of nine subjects naïve to the hypothesis under test (5 men and 4 women; age 20-29 years) took part in the complementary experiment, which was scheduled in one session lasting approximately 30 minutes. The apparatus used was the same as for the main experiment.

Participants compared level-fluctuating up-ramps and down-ramps in terms of global loudness in a 2I, 2AFC procedure. They had to indicate which sound of each pair they perceived as louder. They were presented with 300 trials (organized in six blocks of 50 trials). Half were composed of an increasing stimulus followed by its temporally reversed version, while the other half of the trials were presented in inverted order. In each pair, the reference ramp (an up-ramp or a down-ramp) was generated from a level-fluctuating [65-80 dB SPL] intensity profile ($SD = 2$ dB, restricted to Mean ± 5 dB) as in the main experiment. The comparison ramp was its temporally reversed version, which was presented at five different levels relative to the reference ramp ([60-75], [63-78], [65-80], [67-82], [70-85] dB SPL]) in order to introduce some variances into the loudness comparisons. The goal of the experiment was to check that more than 50 % of up-ramps would be perceived as louder than down-ramps, while the two types of profiles were presented with the same counterbalanced set of SPLs (in each session, up-ramps were higher than down-ramps in SPL on 150 / 300 trials). As a result, each of the nine participants answered more often that the up-ramp was louder (individual percentages: 66.3, 65.7, 67.3, 74.3, 65.7, 56.3, 58.3, 80.7 and 71.0 %; Mean = 67.3, SD = 7.5). These proportions were significantly higher than 50 %, as shown by a one-sample one-tailed Wilcoxon signed rank test (W=45; $p < 0.01$). Therefore, this result demonstrates that the asymmetry in global loudness between increasing and decreasing intensity stimuli still exists with level-fluctuating ramps.

4.4 Discussion and conclusion

In this study, we examined the perceptual weights for pure tones with three different intensity profiles as a function of time: flat [80 dB SPL], increasing [65-80 dB SPL] and decreasing [80-65 dB SPL]. On the one hand, clear primacy and recency effects were found for flat level profile sounds. Although the recency effect was greater in our experiment than in previous studies, most likely due to the fact that the sounds used were longer, these results broadly support the trends observed concerning this type of stimulus (Pedersen and Ellermeier 2008; Oberfeld and Plank 2011). On the other hand, this study used a molecular approach to examine in particular whether the temporal weights of high-level components of increasing sounds could be greater than those of decreasing sounds. Instead, a common effect of saliency – the "level dominance" effect – was found. The segments with the highest levels received greater attention from the listeners, but with no significant difference between increasing and decreasing intensity sounds. Thus, the present study shows that the difference in loudness between increasing and decreasing sounds, which still exists with level-fluctuating ramps as confirmed by the results of the additional experiment, does not arise from the fact that the high-level portion is more heavily weighted when presented at the end of the sound. Nevertheless, since the temporal weights were similar whether the loudest portion of the 2-s time-varying tones was located at the beginning or at the end – corresponding, respectively, to the down-ramp and the up-ramp in the experiment –, our results uphold those from another study (Turner and Berg, 2007), indicating that the "level dominance" effect seems to override primacy and recency effects and persist over several seconds. Further experiments using non-stationary sounds with various durations are required to determine precisely the duration of the effect.

Acknowledgments

We are especially grateful to Daniel Oberfeld for very helpful discussion and comments when not only preparing but also analyzing this experiment. We also like to thank three reviewers who provided fruitful comments on an earlier version of this manuscript. This work was supported by the project LoudNat funded by the French National Agency (ANR).

Partial synthesis

The main goal of this chapter was to investigate whether the global loudness asymmetry between rising and falling tones could be due to dissimilar temporal weighting processes. Using "molecular psychophysics", we measured the temporal weighting strategies of a group of listeners evaluating the global loudness of level-fluctuating sounds with flat, increasing or decreasing intensity-profiles.

First, we observed, as previous studies on this topic, that even for sounds with "flat" level profiles, observers did not weight similarly the different temporal portions of the stimuli. Rather, they assigned significantly greater weights to the first and the last segments as compared to the segments in the middle. These results support previous studies showing that "primacy" and "recency" effects are present in the overall evaluation of loudness of stimuli with flat level-profiles (e.g. Pedersen and Ellermeier 2008). For stimuli with increasing and decreasing level-profiles, neither "primacy" nor "recency" effects were observed; observers only paid attention to the loudest segments. This result, which had already been observed in psychophysical discrimination tasks where the level of the stimuli components was varied, is supposed to reflect the presence of another process known as "level-dominance" (e.g., Lutfi and Jesteadt 2006), which strongly constrains participants' attentional weighting toward the loudest elements of the stimuli. We found similar (symmetric) temporal weighting patterns for the two profiles. Our hypothesis that the asymmetry would result from the fact that the end of rising stimuli is weighted more heavily than the beginning of falling tones could thus not be supported. Because the asymmetry had so far only been established with continuous dynamic stimuli, we conducted a complementary experiment (see

4.3.3) to demonstrate that loudness asymmetries[1] are not specific to ramps changing continuously in loudness but also occur between "time-segmented" ramps as those employed in the molecular experiment (i.e., ramps composed of sixteen 125-ms segments of constant loudness, thus leading to fast but discrete variations of loudness).

How symmetrical weighting patterns can be compatible with global loudness asymmetries? A limitation of the "molecular" methodology is that the measured temporal weights only indicate the *relative* contributions of the different temporal portions (i.e. the segments) constituting the stimuli; they do not indicate any *absolute* value of the overall weight accorded to the stimuli because the weights of each profile are normalized separately (by the sum of their absolute value). As a result, we can only compare the trends of the weighting patterns but not their absolute values. It is thus possible that a similar (i.e., symmetrical) temporal weighting of loudness underpins global loudness evaluation of rising and falling sounds but that the overall "gain" accorded to rising stimuli is greater than the "gain" accorded to falling stimuli[2].

Another aspect that was not addressed in the paper is concerned with the complexity of the integration process. The analysis undertaken in the paper investigated whether a difference in the *linear* combination of individual stimuli components (i.e. the level of their temporal portions) in listeners'

[1]In Appendix E, psychometric functions from were derived from listeners' responses to estimate the size of the asymmetry (in dB). A significant effect of the presentation order of the ramps was found (the asymmetry was 3.4 dB in the falling-rising order whereas it was 1.3 dB in the rising-falling order) which was particularly consistent with the pair order effect observed with continuous *dynamic* stimuli in the second experiment of Chapter 2, a comparable, loudness matching experiment.

[2]One could think about using a two-interval paradigm to solve this issue. If pairs are made with one rising stimulus and one falling stimulus (in the one-interval level-discrimination task employed here, rising and falling stimuli were presented in separate blocks), their weighting patterns could be normalized together and then compared in absolute terms. One major difficulty, however, would be to equalize in loudness rising and falling stimuli *a priori* so that listeners alternate in their responses (rising louder / falling louder) from one trial to another depending on the introduced random level perturbations on the profiles.

judgments could explain the asymmetry between rising and falling tones. In Appendix F, a *nonlinear* analysis of the data collected for increasing and decreasing-intensity stimuli up to its second-order is proposed. First and second-order kernels of the system (i.e., listener's mind) involved in the task were derived. Second-order kernels revealed the presence of significant non-linearities in the integration process, which could indeed not be observed by the linear analysis undertaken in the paper. More interestingly, our data seem to indicate differences between the nonlinear characteristics underlying global loudness judgments of rising and falling stimuli. Indeed, non-zero components could be observed on the last segments of the second-order kernels for falling sounds; the symmetrical effect was not observable for rising sounds. However, it could not be verified whether this peculiarity indeed reflected the global loudness asymmetry between both profiles (probably due to the too small amount of data). It would be particularly interesting to further investigate this aspect. In Appendix F, we also show that a LNL (Korenberg) modeling structure can be inferred from an analysis of both first- and second-order kernels of the system and that this model quite well reflects the main characteristics of listeners' judgment process in the task. Considering small physiological plausible circuits to model global loudness processing could help identify potential underlying mechanisms by filling the gap between psychophysical studies on the one hand and physiological and neuroscientific studies on the other hand, which explore perceptual asymmetries using very different tools.

General Discussion

In this thesis, the big question of how listeners evaluate the *global loudness* of natural sounds was addressed by specifically focusing on *diotic*, controlled stimuli made of linear rising or falling level-profiles, i.e. with constant spectral characteristics. Such basic stimuli, which approximate the dynamics encountered in natural sounds, already activate neural circuits and create perceptual outcomes that are not yet well understood. A psychophysical investigation of the temporal integration processes involved in these time-varying diotic conditions thus constitutes a crucial first step towards a full understanding of the global loudness coding and processing of natural sounds. Binaural integration and the merging of both temporal and binaural integration processes are other aspects of the whole processing that were not considered in this thesis but that are addressed in the LoudNat project[3]. The different studies conducted in this thesis shared the same goals: to characterize and attempt to identify the perceptual processes and mechanisms underlying the *asymmetry* observed between global loudness judgments of diotically-presented rising vs. falling sounds.

This manuscript was organized in four chapters, each examining the characteristics and the factors governing the *asymmetry* from a particular perspective. Chapter 1 explored the influence of methodological and contextual factors on the asymmetry. Chapter 2 focused on the influence of two physical attributes of the sound: the spectral structure and the intensity-region. Chapter 3 investigated the relationships between the temporal profile characteristics of rising and falling sounds (slope, duration, dynamics) and their

[3]As mentioned in the introduction, this thesis was part of the LoudNat project, supported by the French National Research Agency (ANR-11-BS09-016-01).

global loudness judgments and resulting asymmetries. Chapter 4 employed "molecular" psychophysics to determine the underlying temporal weighting of loudness for these profiles. The present general discussion provides a synthesis of this work and attempts to bridge the findings of the different studies. First, a synthetic description of the results is made, allowing to characterize the asymmetry as a function of the experimental method, the context of presentation and the physical attributes of the stimuli. Second, we propose a discussion based on the potential mechanisms that might be involved in the asymmetry in different ways. In the third section, the findings of this thesis are synthesized at a higher level in the context of "global loudness processing of time-varying sounds" (i.e. not only in the case of rising and falling sounds). The fourth section discusses the extent to which current loudness models can account for the obtained results and presents some issues and perspectives for these models. In the fifth and final section, various directions are proposed for future research on global loudness processing and coding of time-varying stimuli. Some of the ideas presented in this section are tentative and would indeed need further development.

5.1 Characterization of the global loudness asymmetry between rising and falling sounds

5.1.1 Apparent robustness and constancy of the phenomenon

The purpose of the study presented in Chapter 1 was to investigate the robustness of the asymmetry to different measurement methods in different contexts of presentation. The asymmetry in global loudness between 1-kHz rising and falling ramps, lasting 2 seconds and made of 15-dB dynamics, were derived in various experimental configurations. Overall, very similar asymmetries were found between both profiles whether they were inferred from magnitude estimation or from loudness-matching tasks. Moreover, the context of presentation had, on average, no influence on the magnitude of the asymmetry. We found that the mean size of the asymmetry was about 3-4 dB. In Chapters 2 and 3 where the same 2-s, 1-kHz, 15-dB varying stimuli were used, similar asymmetries to the one obtained in Chapter 1 were found, on average over the different experimental conditions (measurement methods, contexts of stimuli presentation and intensity-regions): 3.3 dB in Chapter 2 and 2.8 dB in Chapter 3 (see Appendix D)[4].

Overall, it can be said that the asymmetry in global loudness is not due to a specific context or method, and is on average, around 3 to 4 dB for 1-kHz, 2-s ramps varying over 15-dB dynamics. The mechanism(s) that give rise to this phenomenon thus appear to be robust and stable across different contexts and conditions. This result constitutes one major outcome of this thesis and contrasts with *loudness change* judgments of rising and falling sounds that were explored in previous studies, which were found to be particularly sensitive to the definition of the percept itself (i.e., retrospective or continuous *loudness change*) and to the measurement method employed to measure it (for example, whether direct or indirect measurement procedures was employed; see Susini et al. 2010).

[4]Because "time-segmented" and not continuous ramps were used in Chapter 4, the asymmetries derived in this chapter are not considered here.

5.1.2 Influences of the physical attributes of the sounds

5.1.2.1 Intensity-region

A significant effect of the intensity-region of the stimuli has been observed in the different chapters of this thesis: the asymmetry was overall slightly smaller for the ramps presented in high intensity-regions (i.e. with high levels) compared to those presented in lower intensity-regions. However, the threshold between the low and the high intensity-regions at which this reduction occurred was seemingly dependent on the experimental context. In the magnitude estimation experiment of Chapter 1 (Exp. 1), we found that the asymmetry obtained for [75-90 dB SPL] ramps was smaller than the one measured for ramps presented in lower intensity-regions. In the paired-comparison experiment of the same chapter (Exp. 2), this reduction was found for [65-80 dB SPL] ramps. In the paired-comparison experiment of Chapter 2 (Exp. 2), it occurred for [70-85 dB SPL] ramps as compared to the other unique intensity-region tested ([50-65 dB SPL]). In the magnitude estimation experiment of Chapter 2 (Exp. 1) this reduction was revealed by a significant direction × intensity-region interactions but no particular intensity-region threshold was determined. In the magnitude estimation experiments conducted in Chapter 3, the reduction occurred for [75-90 dB SPL] ramps in Exp. 1 (Group A) and for ramps varying at 5 dB/s and that had a maximum level of 90 dB SPL in Exp. 2. However, we found no significant effect of the intensity-region or the maximum level of the ramps neither in Exp. 1 (Group B) nor for ramps varying at 2.5 dB/s in Exp. 2.

These results can also be qualified by considering local context effects such as the influence of previous stimuli direction in direct evaluation tasks or pair order effects in paired-comparison judgments. Indeed, we found different experiments in which the intensity-region effect was conditioned by such local contexts: the asymmetry was significantly reduced in high intensity-regions but only when rising preceded falling stimuli. In the “mixed context” group of the magnitude estimation experiment of Chapter 1, a significant local context effect was found: asymmetries were smaller in high intensity-regions only when the falling sounds were preceded by rising sounds. In the paired-comparison experiments of Chapter 2 (Exp. 2 and Appendix B), the intensity-region effect was found only when the pairs were composed of a rising followed by a falling stimulus. A consistent order effect was also observed in Chapter 4 for time-segmented ramps covering the [65-80 dB SPL] region (see Appendix E)[5]. Because we obtained similar results with white noises and narrow-band noises (Exp. 2 of Chapter 2 and Appendix B), it even suggests that this local context effect is not specific to tones. We did not attempt to characterize the magnitude of this reduction *per se* but our results indicate that it lays around 1-2 dB (see Figure 2.2 in Chapter 2 and see Appendix E). However, this contextual effect does not seem to be a *necessary* condition for the reduction to occur, because significant direction $\times$ intensity-region interactions were also obtained in Chapter 1 (cf. Exp. 1, "separate context") and in Chapter 2 (cf. Exp. 1), where rising and falling ramps were presented in separate blocks.

All in all, while the asymmetry is robust *on average*, its magnitude seems to depend on the intensity-region of the stimuli, conditioned in certain cases by the direction of the preceding stimulus.

5.1.2.2 Spectral structure

The influence of the sound spectral structure on the asymmetry was addressed in Chapter 2. In Exp. 1, where a magnitude estimation procedure was used, similar and significantly positive asymmetries were found with pure tones ranging from 125Hz to 8kHz, whereas no significant asymmetries were obtained with broadband noises (both white noises and pink noises). In Exp. 2, where a loudness-matching procedure was employed, the asymmetries were significantly greater for 1-kHz tones compared to white noises, but they were nevertheless significant in many conditions and were, on average, equal to 2 dB. The asymmetry is thus always smaller with broadband noises than with tones, but this difference seems to depend on the measurement method.

[5] The paired-comparison experiments of Chapter 1 (Exp. 2) could not reveal such local context effects because the asymmetries were averaged over the two presentation orders to compensate Time-Order Errors (e.g., Hellström 1979). Besides, the local context influence was not addressed in Chapter 3.

5.1.2.3 Temporal profile characteristics

In Chapter 3, we measured the asymmetry between rising and falling 1-kHz tones presented with various durations, slopes and dynamics. We found that the asymmetry occurred in many conditions but that its size was dependent on these temporal profile characteristics in a complex manner (as illustrated in Appendix D). For instance, we observed that the asymmetry increased with the duration of the ramps when their dynamics was fixed (see Exp. 1, Group A); this increase was significant between 2 and 6 seconds. Furthermore, the second experiment of this chapter showed a significant increase of the asymmetry with the dynamics of the ramps. At equal durations, slopes and maximum levels, ramps with smaller dynamics led to smaller asymmetries (see also Fig. D.1 in Appendix D). In the most extreme case, when the ramps were presented with dynamics of 5 dB, the asymmetry was almost reduced to zero (cf. Fig D.1). All together, these results indicate that the asymmetry is not a simple offset between judgments of rising vs. falling sounds, and that its dependency on the ramp profile characteristics (slope, duration, dynamics) is in appearance very complex (as highlighted in Appendix D), because different mechanisms are likely to interact (see Chapter 3). Therefore, the asymmetries have to be interpreted in the context of the global loudness processing scheme taken as a whole (see section 5.3 below).

5.2 Potential mechanisms underlying the asymmetry

5.2.1 Temporal weighting of loudness

In Chapter 4, we examined whether the asymmetry between rising and falling-intensity stimuli could be due to the fact that their temporal portions do not receive the same "weighting" in the overall evaluation. This question was directly addressed using psychophysical reverse correlation. We found symmetrical temporal weighting patterns for rising and falling sounds, with weights exclusively assigned to the loudest portions, similarly for both profiles. Therefore, this led us to conclude that the asymmetry is not due to a dissimilar temporal weighting. This finding eliminates the possibility that the asymmetry (between 2-s, 1-kHz, 15-dB ramps) is due to the fact that listeners include the last portions of decreasing stimuli (i.e. the softest portions) in their judgments. However, as noted in the partial synthesis of this chapter, because of the normalization of the temporal weights inherent to the regression analysis, the possibility that the asymmetry would be due to a greater overall "gain" given to the loudest portion of rising compared to falling stimuli cannot be ruled out.

5.2.2 Reduction at high intensity-regions and local context emphasis

Our results suggest that high-level contextual-dependent processes rather than low-level peripheral mechanisms are involved in the reduction of the asymmetry at highest intensity-regions. The fact that the intensity-region threshold at which this reduction occurs depends on the measurement method and on the experimental context (the different experiments of this thesis indeed yield different values, as discussed in section 5.1.2.1 above), and the fact that this reduction occurs only sometimes in specific contexts (when rising ramps precede falling ramps), go against peripheral explanations. It is interesting to note that similar order effects were reported in the literature on shorter asymmetrical stimuli. With 250-ms stimuli presented at 80 dB SPL (peak level), Stecker and Hafter (2000) found that the asymmetry was cancelled when rising sounds preceded falling sounds; they assumed that this effect could be explained by a "decay suppression" mechanism (see Chapter 1 for details). Ries et al. (2008) also found smaller asymmetries in the rising-falling order with 500-ms stimuli presented at peak levels of 70 dB SPL. However, contrary to Stecker and Hafter, the authors ruled out the "decay suppression" mechanism because it was not consistent with the perceived duration asymmetries they also measured. We believe our results are also inconsistent with a "decay suppression" mechanism. First, because this mechanism relies on a reverberation interpretation and thus loses its meaning with 2-s stimuli varying slowly in loudness as those employed in the present thesis. Second, because the effect only occurs at high intensities[6], something that can hardly be explained in the light of the "decay suppression" mechanism solely. In the end, our results rather indicate that another high-level process should be sought, compatible with the presently observed intensity-region $\times$ order interaction.

5.2.3 Decay of the loudness trace of falling stimuli

Many studies argued that "memory[7] processes" might be involved in overall loudness judgments of time-varying sounds because the salient events contained in a sound sequence of several seconds (such as loudness peaks) strongly influenced overall judgments, but this influence decreased as a function of the duration between the salient events and the end of the sequence (Kuwano and Namba 1985; Gottschling 1999; Hellbrück 2000; Kuwano et al.

[6]This interaction between the order of presentation and the intensity-region of the stimuli could not observed in previous studies because only one single intensity-region was tested.

[7]The use of the term "memory" is perhaps not ideal in that context because it was never directly demonstrated that memory mechanisms were recruited per se, e.g. by using measurement procedures characterizing change in sensitivity (*d'*) as a function of time. In studies conducted on overall evaluation of time-varying painful stimuli or annoyance judgments of time-varying sounds, they prefer using the terms "peak-end" rule (Kahneman et al. 1993; Ariely and Carmon 2000) to account for the fact that both the peak and the end appear to be the primary factors guiding overall judgment; a "heuristic" that also gives the peak a greater weight when it corresponds to the end of the sequence.

2003; Susini et al. 2002; Susini et al. 2007). In Chapter 3, we hypothesized that such an effect could be involved in global loudness evaluations of falling sounds; we called it the *decay mechanism*. Although we were not able to clearly assess its presence (i.e. significantly), we observed significant differences concerning the influence of duration on global loudness of rising and falling profiles in both experiments, whether the ramps were stretched in time at constant dynamics (Exp. 1) or at constant slope (Exp. 2). In particular, the results of the first experiment (Group A) showed a slight decrease of the estimates of falling tones between 9 and 12 s at all intensity-regions tested. Future studies offering greater statistical power could help determine whether this decrease might become significant in other contexts and at longer durations. Besides, note that an asymmetry between rising and falling tones was also found for 1-s (see Exp. 1) and for 2-s ramps (see Exp. 2), a result that cannot likely be accounted for by a decay mechanism[8].

Therefore, although the presence of some memory decay in the global loudness processing of falling sounds cannot be ruled out and might even contribute to increase the asymmetry rising vs. falling at long durations, it does not explain why an asymmetry is observed between ramps of short durations (e.g., at 1 or 2 s).

5.2.4 A specific circuit for rising *tones*?

In Chapter 2, we found that the global loudness asymmetry was modulated by the spectral structure of the stimuli and more specifically, that the asymmetry was reduced with broadband noises as compared to tonal stimuli. A possible interpretation of this effect has been proposed by Neuhoff (1998) in

[8]One may argue that a sensory decay process might be at the basis of the presentation order effect that we found, i.e. the smaller asymmetry in the rising-falling order as compared to the falling-rising order. Indeed, in the falling-rising order, listeners need to maintain the sensory trace of the beginning of the falling-intensity tone, whereas in the rising-falling order, they can directly compare the two loudest portions that are only separated by a small inter-stimulus interval. However, a question would remain: why would this effect apply only to stimuli presented in high intensity-regions (given that no order effect was obtained for lower intensity-regions)? This would imply that loudness memory coding is intensity-dependent, something that has, to the best of our knowledge, never been reported so far.

the context of *loudness change* asymmetries: noise and tonal stimuli would not have the same behavioral "relevance" to the listener and consequently would not lead to the same asymmetry. In other words, the size of the asymmetry would be mediated by the power of 'looming-like' rendering, since noise stimuli are less perceived as moving objects than tonal stimuli (e.g., Bach et al. 2015). This evolutionary interpretation has been supported by a number of behavioral and neuroscientific studies conducted on humans and primates (such as rhesus monkeys) showing that rising tones and rising noises do not lead to the same cortical activities and do not recruit the same brain circuitry to be processed (e.g., Ghazanfar et al. 2002; Maier et al. 2004; Maier and Ghazanfar 2007; Maier et al. 2008). However, in a recent study conducted in collaboration with Brice Bathellier and his team, we found the opposite: greater asymmetries with white noises than with 8-kHz complex harmonic sounds in mice, both at behavioral and at cortical levels (Deneux et al., submitted). Does this indicate that in mice, noise stimuli are more informative about potential approaching dangers than tones? The advantage of tonal stimuli over non-tonal stimuli cannot yet be taken as a general property of the mammalian auditory system.

The difference in brain coding of white noise and tones with rising and falling-intensity profiles is still a burning issue in the neuroscience domain (e.g., Bach et al. 2015). We believe it would be interesting to further examine the respective contributions of peripheral and high-level processes in this apparent tonal influence. Indeed, concerning global loudness asymmetries more specifically, it cannot yet be ruled out that the spectral structure effect observed in the present study also stems, at least in part, from the periphery. Although we showed similar effects with narrowband noises recruiting either one critical band or several critical bands (see Partial synthesis, Chapter 2), an experiment involving both complex tones with harmonic and non-harmonic (i.e. unresolved) spectral structures would be needed to clearly disentangle the differences of spread of excitation along the basilar membrane with the role of high-level cortical stages.

5.2.5 Non-sensory factors

In order to examine the role of peripheral and high-level mechanisms in psychophysics, the influence of non-sensory factors in observers' judgments (such as response biases, the adoption of specific evaluation strategies, etc.) has to be considered. For example, when looking at perceptual asymmetries or illusions, it is important to know whether the measured effects result from real perceptual differences or from shifts of decision criterion. According to the Signal Detection Theory (Macmillan and Creelman, 2004), this conceptual distinction is made to reflect the fact that a shift of the psychometric function can either come from a shift of mean internal perceptual representations (represented by Gaussian distributions) or from a shift of the criterion value (for an example of this issue in visual perception, see Carrasco et al. 2004, Schneider and Komlos 2008, and also Georgeson 2012). Although we did not conduct any experiment that specifically addressed this decision criterion aspect, we believe that our results provide converging evidence to argue that global loudness asymmetries are most likely related to a "real" perceptual difference rather than a shift of the decision criterion. First, because the average magnitude of the effect is robust to various measurement procedures and contexts, it would be very unlikely that the same criterion be used across very different kinds of conditions (for example, when the sounds are evaluated using numbers (Exp. 1 of Chapter 1) or compared against constant-intensity tones (Exp. 2B of Chapter 1)). Similarly, the results of Chapter 3 could hardly be explained by a simple decision criterion shift, given the complex modulations of the asymmetry as a function of the ramp profile characteristics. Other studies investigating physiological/objective behavioral responses to rising and falling tones, showing asymmetries induced by "looming" tones but not by "looming" noises, also provide support to the view that perceptual, i.e. not decisional, mechanisms are involved (e.g., Maier and Ghazanfar 2007; Leo et al. 2011).

However, an aspect that would nevertheless deserve to be further addressed in future studies concerns the large interindividual differences observed in the size of this asymmetry, as shown in Appendix A. Although the magnitude of the asymmetry is on average constant across different conditions, it strongly varies across listeners. The measurements of the asymmetry obtained with different listeners in the experiments of Chapter 1 followed an almost normal distribution with a standard deviation of about 4 dB (see Appendix A), which is about the same size as the effect itself. Investigating this issue could help go beyond the *apparent* robustness of the asymmetry on average, which does not hold true at an individual level. In particular, it remains to be determined whether intraindividual inconsistencies explain a part of this overall variability, e.g. by repeating the measurement several times with the same listeners in the same conditions[9], or whether they are entirely due to interindividual differences of positioning of the decision criterion. Finally, the respective contributions of low-level sensory differences (e.g., peripheral) and high-level processes such as decisional factors in this variability need to be addressed. In a comparable context, Florentine (2014) stated: “we should treasure [these] individual differences because they are highly likely to reveal important mechanisms that contribute to loudness”.

[9]Although this intraindividual variability was considered in some way in Appendix C, where the asymmetries measured for tones and noises were compared with the same listeners, it remains to be specifically explored using the same stimuli presented in the same context.

5.3 Global loudness processing of time-varying sounds

Throughout the different experiments of this thesis, which were specifically designed to better understand the mechanisms underlying global loudness *asymmetries* between rising and falling stimuli, we also learnt about global loudness processing of time-varying sounds at a more general level.

When the sounds are varying around a given mean level (their overall level profile is said to be "flat", like in Chapter 4), there are evidences that the underlying temporal weighting processes employed by listeners to judge their global loudness is not uniform. The first events and, to a lesser extent, the last events receive stronger weights compared to the others; "primacy" and "recency" effects have been proposed to account for these findings, respectively. While previous studies demonstrated these effects with broadband or narrowband noises lasting about 1 s (e.g., Pedersen and Ellermeier 2008), the results of this thesis show that it still holds true for tones (with 2-s, 1-kHz tones; see Chapter 4) and for noises of longer durations (with 3-s white noises; see the paper presented in Appendix G). However, the respective contributions of these "primacy" and "recency" effects seem to depend on the stimuli characteristics, such as their overall duration and both the size and duration of the introduced level perturbations. These aspects have to be further explored, as for instance, the "primacy" effects which seem to be downsized when the duration increases (see Appendix G).

When the overall level profile of a time-varying sound is not flat, as it is the case with the rising and falling stimuli specifically considered in this thesis, the picture is substantially more complex. First, we demonstrated in Chapter 4 (using 2-s ramps varying over 15-dB dynamics) that the "primacy" and "recency" effects disappeared and that global loudness judgments were then dominated by the loudest elements of the stimuli. This finding has been called the "level-dominance" effect (Lutfi and Jesteadt, 2006). Both Chapters 1 and 3 showed that global loudness judgments of sounds with increasing or decreasing level profiles are guided by their maximum level. We gained further insight into this "level-dominance" effect thanks to the manipulation of the ramp profile characteristics realized in Chapter 3. In particular, we confirmed in Exp. 2B of Chapter 1 and in Exp. 2 of Chapter 3 that global loudness judgments of such dynamic stimuli are always slightly lower than those of constant-level sounds presented at their maximum level (Susini et al., 2010), and we found that the slope of the ramp played an important role in this difference (see Exp. 1 and 2 in Chapter 3). We thus proposed that an *integration mechanism* of the loudest portion of these sounds could be at the basis of the effects of these two physical parameters, i.e. the maximum level and the slope of the ramps. Note that such an integration mechanism could be consistent with the temporal weighting patterns observed in Chapter 4 that appeared under the form of smooth exponential temporal functions located around the loudest portions of the stimuli. But the story does not end here: the integration process might also involve nonlinear components, as suggested by the second-order analyses undertaken in Appendix F. Indeed, not only the loudest levels of the different temporal portions of these level-fluctuating sounds are taken into account in global loudness judgments: both interactions and squared-terms of the levels of these portions appear to be involved. These nonlinear components of global loudness processing constitute a very interesting perspective for future research since they had never been addressed so far. Last but not least, as shown all along this manuscript, judgments of rising and falling sounds are substantially different. Rising sounds are perceived as louder than falling sounds and the magnitude of this "asymmetry" depends on many characteristics of the stimuli such as their slope, their dynamics, their duration and their spectral structure. The results obtained in Chapter 3 also tend to suggest that there might be a "decay mechanism" involved in global loudness processing where the value of the loudness integrated over the loudest portion(s) of the sound would decrease as a function of its position toward the end. This latter aspect need to be confirmed.

All together, these results show that several not yet fully understood mechanisms are involved in the global loudness processing of time-varying sounds. The effects highlighted throughout this thesis can definitely not be explained by considering loudness as a *static* concept, i.e. by considering time-varying stimuli as a sequence of small independent stationary parts[10]. In the next paragraph, the extent to which these results can be accounted for by current loudness models and by current indicators of global loudness is discussed.

[10]The same could be said (i) for continuous loudness perception and the associated "decruitment" effect found for falling-intensity sounds (see Olsen 2014) and (ii) for the effect of loudness adaptation induced by intermittent constant-intensity tones (e.g., Botte et al. 1982).

5.4 Loudness models for time-varying sounds

5.4.1 Predictions of global loudness asymmetries

Throughout this manuscript, we examined the extent to which current loudness models of time-varying sounds could account for the asymmetries derived between global loudness estimates of rising and falling stimuli. In particular, we tested Glasberg and Moore's "TVL" model (Glasberg and Moore, 2002) and, as suggested by the authors, two indicators that are assumed to reflect global loudness: STL_{max} and LTL_{max}, which correspond to the maximum values of short-term loudness and long-term loudness patterns provided as outputs, respectively. Overall, we showed in Chapters 1, 2 and 3 that both STL_{max} and LTL_{max} substantially underestimate the magnitude of the asymmetry measured between 1-kHz, 2-s, 15-dB rising and falling ramp: an asymmetry of about 0.5 dB is predicted by STL_{max} and around 1.3 dB by LTL_{max}. These values are consistent with our finding that rising tones are perceived as louder than falling tones (i.e. the asymmetries are positive) but the magnitude of the effect (around 3-4 dB) is definitely underestimated. In addition, Chapter 2 showed that TVL could not account for the spectral structure effect: it gave similar predictions for noises and tones. The most challenging results to be accounted for by the model were probably those obtained in Chapter 3, where the profile characteristics of the ramps were manipulated. First, we found that the trends of global loudness judgments, i.e. the significant increase with duration obtained with the "time-stretching" at constant dynamics both for rising and falling tones (i.e. explained by the so-called integration mechanism) could not be predicted neither by STL_{max}, nor by LTL_{max}. The increase of the asymmetry between these profiles as a function of the duration could also not be predicted by these indicators. In order to account for the *integration mechanism*, we introduced a new indicator called STL_{int}, which is based on an average of the maximum of STL contained in a fixed rectangular temporal window (see Chapter 3 for details). We found that the increase with duration could then better be predicted, at least qualitatively. However, the asymmetry between rising and falling tones remained unexplained by STL_{int}.

All together, these results show there is no simple way to adapt the current indicators derived from Glasberg and Moore's model 2002 so they can account for the effects observed throughout this thesis. Our results do not support the view that the long-term loudness output of the model directly reflects "relatively high-level cortical processes and involves memory"[11] (Moore, 2014). More particularly, we found that the maximum value of the long-term loudness pattern could not account for various effects obtained with rising and falling tones (e.g., asymmetries, increase of global loudness when the ramps are stretched in time at equal dynamics). One thus needs to further investigate the mechanisms and the computational code characterizing the transformation of the short-term loudness pattern (presumed to reflect the perceived loudness at any instant) into a single global loudness estimate.

The analyses undertaken in Chapter 3, 4 and in Appendix F also show that listeners' global loudness judgments of rising and falling stimuli take place on a certain temporal integration window (centered on the loudness peak of the stimuli) but also involve nonlinear mechanisms, which again suggests that simply taking the maximum of STL or LTL is not sufficient to account for the complexity of the processes underlying the transformation of momentary loudness values into global loudness. Finally, the fact that Glasberg and Moore's model fails to predict the asymmetries and the modulations of these asymmetries depending on stimuli parameters (intensity-region, slope, spectral structure) indicates that these effects are not reflected by the mechanisms presently implemented in the model. This outcome contrasts with what was found at shorter timescales where it was demonstrated that Glasberg and Moore's model could on average account for the magnitude of the asymmetries measured by Stecker and Hafter (2000) between 250-ms asymmetrical ramped and damped stimuli (Rennies et al. 2010; Moore 2013).

Throughout the different chapters of this thesis, model testing was mostly addressed with Glasberg and Moore's. It is very likely that the same conclusions would have been reached with the output of Challuper and Fastl's

[11] As for example, when listeners "judge the overall loudness of a relatively long segment, such as a sentence or musical phrase" (Moore, 2014)

model for time-varying sounds, which shares a very common structure – this model was actually examined in Chapter 1 and the same conclusions as with Glasberg & Moore's were obtained. Other available peripheral auditory models such as AIM (Patterson et al., 1995) or Zilany's model (Zilany et al., 2009) were not considered in this thesis because their outputs do not reflect loudness per se. Lentz and Shen (2011) demonstrated that Zilany's model, which takes into account neural adaptation of the auditory periphery, could not predict asymmetries between short duration 25-ms asymmetric ramps, and took this as an indirect support that the central (i.e. not peripheral) auditory system was involved. This aspect remains to be addressed with the much longer stimuli employed in the present study.

5.4.2 Challenges for future loudness models

We believe that current loudness models cannot predict the asymmetry because they do not yet involve the presumed high-level mechanisms that underpin global loudness processing of such intensity dynamics. Once these mechanisms will be fully identified and described, it remains to see how they can be implemented in future loudness models. We propose that future models might constitute extensions of current models. They could use STL as input and add complementary modeling stages to account for the different, presumably nonlinear integration mechanisms underlying global loudness processing, as synthesized in the previous section. LNL cascade models might constitute good candidates to this aim. Such an approach would be based on the assumption that the mechanisms considered in Glasberg and Moore's model and the mechanisms reflected in these later processing stages do not interact (i.e. that the efferent mechanisms could be neglected).

As noted by Brian Moore (2014), none[12] of the current models yet consider the high-level processes that may influence loudness in natural conditions. Even when dealing only with auditory effects (i.e., not audio-visual integration processes), there are many well-known effects that cannot be explained by low-level mechanisms. The most famous is probably the loudness

[12] De Coensel et al. (2009) proposed a model that attempts to address similar "high-level" issues in the context of overall annoyance evaluation.

constancy effect revealed by Zahorik & Wightman (Zahorik and Wightman, 2001), who showed that the loudness of a sound source whose distance from the observer was varied did not change much (although the sound pressure level at the ears indeed decreased with distance). Another result is concerned with the perceived loudness of speech stimuli, which was found to depend on the vocal effort of the talker, inferred by listeners using – in part – spectral cues (e.g., Allen 1971). The binaural loudness constancy effect (Epstein and Florentine, 2009), reflecting the fact that we perceive a talker in a room as having the same loudness whether we listen with one or two ears (a result which cannot be explained by the "traditional" binaural integration theory) is another type of high-level effect that cannot be accounted for by current auditory models. Although all the aforementioned effects taken from the literature arise with complex natural sounds, or at least presented in natural listening conditions[13], this thesis show that complex processes already occur with very basic diotic stimuli (identical 1-kHz tones at the two ears) that simply vary in level slowly over a few seconds.

The non-uniform weighting patterns underlying global loudness judgments of time-varying sounds with overall flat level profiles (e.g., Pedersen and Ellermeier 2008; Oberfeld and Plank 2011; Oberfeld et al. 2012; Ponsot et al. 2013; Ponsot et al. submitted) will also need to be considered in future loudness models.

Lastly, it should be addressed how current models might be able to take into account the fact that high interindividual differences in global loudness judgments are found in many situations, as with regard to the asymmetries highlighted in the present work (see Appendix A) or concerning temporal weighting strategies (see Appendix G). Indeed, current models are only based on the average characteristics of a relatively large group of listeners (Moore, 2014). If one wants to predict the loudness of a particular listener, a model must consider the physical characteristics of that listener (e.g., his HRTFs) but also his physiological, perceptual and cognitive peculiarities.

[13]With respect to these effects, Brian Moore (2014) noted that "these factors may be important in everyday listening situations, where our sensory systems seem designed to estimate the properties of auditory objects or sound sources, rather than to estimate the properties of the signals reaching the ears or the eyes."

For example, Hots et al. (2014b) raised this issue concerning the time constant reflecting temporal integration. The best way to reach this flexibility and accuracy is to build models where the constants reflect sensory or cognitive "parameters". Identifying the parameters of the model that account for different individual strategies can potentially bring important information about the functioning scheme of the underlying processes (Wackermann, 2014). With respect to global loudness evaluation, it would be interesting to propose a model whose parameters can be adjusted to account for the fact that listeners may use indeed different strategies.

To conclude, from the different results collected throughout this thesis as well as results from various previous studies, it can be said that loudness of time-varying sounds is *still* a challenge for current models. The present thesis shows that Glasberg and Moore's model cannot directly account for the global loudness of tones with increasing and decreasing levels. Other studies revealed discrepancies between model predictions and experimental data: for natural time-varying technical signals such as the sound produced by a gear wheel or an helicopter, these errors are indeed larger than 5 dB (Rennies et al., 2013b), sounds with subcritical bandwidths cannot well be predicted neither (Hots et al., 2013) and finally, the errors of the models in predicting simple amplitude modulated sounds with fast modulation frequencies (i.e. $fm > 32$ Hz) are found to be as high as 4 dB (Rennies et al., 2010)).

5.5 Main research perspectives derived from this thesis

5.5.1 Identifying the processes highlighted in this thesis

A very promising avenue for future research is to determine which of the results obtained in this thesis are specific to (global) loudness and which rather reflect high-level mechanisms involved in general in the processing of time-varying sensory information. Because Glasberg and Moore's model does not accurately predict the results obtained in this work, we believe that some of the effects presently highlighted might indeed reflect higher-level mechanism than those typically considered in loudness models.

For example, one could investigate whether asymmetries between rising and falling patterns are found in overall judgments of other sensory modalities (as suggested in Ariely and Zauberman 2000; Ariely and Carmon 2000) and which characteristics of these patterns modulate the magnitude of the asymmetries. In the visual domain, for example, it could be examined whether overall brightness judgments of rising and falling ramps of luminance follow the same trends as loudness judgments of intensity-ramps in the auditory domain, and for instance, whether an integration mechanism as the one presumably observed in Chapter 3 is involved. Such an approach was adopted in Teghtsoonian et al. (2000) where auditory decruitment induced by falling-intensity ramps was compared with visual decruitment induced by circles decreasing in size.

Another aspect that could be considered is related to the loudness after-effects induced by sequences of rising or falling-intensity ramps (e.g., Reinhardt-Rutland 1980; Reinhardt-Rutland and Anstis 1982; Reinhardt-Rutland 2004). When a series of contiguous 1-s falling (rising) intensity ramps is presented to a listener during an adaptation period of several dozens of seconds, a subsequent test tone presented at constant intensity leads to the perceptual illusion that it is growing louder (softer). Moreover, the absolute size of this perceptual illusion is different between series made of rising or falling inducers. It would be interesting to explore whether asymmetries in loudness after-effects induced by series of rising (falling) ramps and asymmetries in global loudness induced by single rising (falling) ramps are governed by the same mechanism(s) and are based on the same neural code.

5.5.2 Assessing the role of memory for falling tones

In Chapter 3, we found significant interactions between the direction of the ramps (rising / falling) and the duration of the stretching of their profiles, which could suggest the presence of a decay mechanism (see section 5.2.3 above) involved for falling sounds. An interesting perspective would be to design a specific experiment in which it would be possible to emphasize this interaction and more particularly what happens for falling tones with a better precision and power than with the magnitude estimation procedure employed in Chapter 3.

One possibility would be to investigate whether a trace memory decay of the beginning of the falling tones is at the basis of these effects. The most common approach is to use paired-comparison tasks and vary the inter-stimulus interval (ISI). Memory for loudness was mostly examined with simple stationary tones so far, by measuring the decay of their loudness traces in memory (e.g., Clément et al. 1999; Botte et al. 1992). Clément et al. (1999) investigated the decay function of sensitivity (*d'*) in a loudness discrimination task (using a roving-level paradigm) as a function of the inter-stimulus interval (ISI). The loss of sensitivity of rising and falling stimuli as a function of the ISI could be investigated using the same procedure as in Clément et al. (1999)[14]. However, in this case, the hypotheses would not be as straightforward as in Clément et al. because the memory of the peak of the falling tone would be disrupted by its subsequent decreasing portion whereas the memory of the peak of the rising tone would be disrupted by the increasing portion of the second comparison tone, as suggested by the results of Botte et

[14] A pilot experiment was only conducted with one subject using exactly the same procedure as Clément et al. (1999) but replacing the constant tones with 2-s, 15-dB, 1-kHz rising or falling tones. Preliminary results (not presented in this manuscript) did however not show any significant differences in sensitivity decay traces (ISIs ranging from 0.5 to 6 s) between rising and falling stimuli, respectively.

al. (1992). Investigating the role of memory mechanisms in global loudness processing of rising and falling-intensity tones thus appear to be significantly more complex than in the case of constant-intensity tones.

5.5.3 Disentangling decision criterion effects

In their study, Schneider and Komlos (2008) proposed an interesting and persuasive approach to tease apart the respective influence of perceptual and criterion effects in visual contrast perception. They showed that the contrast shift found in a visual task, where subjects performed a comparative judgment ("which target has higher contrast?"), entirely disappeared with another group of subjects performing equality judgment ("are the two targets equal in contrast?"), where judgment was now free of response bias. This demonstrated that the effect observed in the comparative judgment task was caused by a decisional bias and not related to any perceptual difference. It would be interesting to conduct a similar experiment to study the asymmetry between rising and falling sounds. Indeed, if the asymmetry found in a comparative judgment task also occurs in an equality judgment task (for example, asking "are the two sounds equal in global loudness?"), this would demonstrate that the effect has only perceptual (sensory) origins; otherwise, it would suggest that decisional biases (i.e. criterion effects) are involved. Other methodological approaches that help distinguish low-level sensory effects and criterion effect (i.e. response bias) can be found in the literature (e.g., the case of cross-modal enhancement of perceived visual intensity in Odgaard et al. 2003, Odgaard et al. 2004).

Another approach to assess the influence of non-sensory factors would be to manipulate the experimental instructions in order to corrupt spontaneous evaluation strategies, potentially biased. Such an approach has been used to assess the influence of non-sensory mechanisms on the asymmetry in perceived duration between short (50 and 500 ms) asymmetric ramps (see DiGiovanni and Schlauch 2007). To test the "echo-suppression" hypothesis proposed by Stecker and Hafter (2000), that listeners discard the late decay portion of damped sounds by associating it to the reverberation of the first portion, DiGiovanni and Schauch compared asymmetries between listeners who were instructed either to 'include all aspects of the sound' their judgments and listeners who did not receive any particular instructions. Their results showed that asymmetries were significantly reduced for the first group of subjects. Authors concluded that non-sensory factors were involved in the asymmetries. A direct transposition of their procedure in our context would be to ask subjects to ignore the direction of intensity change (increasing or decreasing) and to see whether the asymmetry would be the same with these new instructions.

5.5.4 Achieving a better understanding of temporal integration processes

A future challenge in the study of temporal integration processes underlying global loudness evaluation will be to determine the extent to which the results obtained from "molecular psychophysics" might be interpreted in the context of more « traditional » loudness estimation tasks. In this context, we recently conducted a study on temporal weighting of loudness comparing temporal weighting patterns in a level-discrimination task versus a magnitude estimation task. Although subjects evaluated the global loudness of the same stimuli in the two tasks, we found different temporal weighting patterns between the two tasks. In this study (which has been submitted for publication and is presented in Appendix G), we proposed that the measured temporal weighting patterns might be affected by the strategy employed by the listeners in the magnitude estimation task to minimize the cost of their evaluation process and their response time. Future studies have to be conducted to better understand which temporal integration processes are deployed as a function of the task and the context of loudness evaluation, in order to disentangle task-specific processes and "general" processes of global loudness evaluation.

A second aspect that will undoubtedly need to be considered is concerned with the role and the contributions of nonlinear components (i.e. squared-level components and interaction terms) in the temporal integration processes underlying global loudness evaluation, as we started to investigate in Appendix F.

5.5.5 Investigating global loudness processing with complex natural stimuli

This thesis investigated global loudness of synthetic auditory stimuli with very basic spectral structures (tones or noises) that linearly vary in level over time. Environmental sound events exhibit more complex time-varying characteristics, both spectral and intensity characteristics varying and even co-varying in more complex ways than what was considered here. In the present work, the effects of spectral and temporal characteristics were examined independently (Spectral in Chapter 2; Temporal in Chapters 3 & 4). How temporal and spectral cues interact when these characteristics co-vary remains to be determined. A direct perspective of the present work would be to examine whether the temporal mechanisms involved in global loudness processing of rising and falling sounds are independent of their spectral structure. For example, are the results of Chapter 3 specific to tones or can they also be found for noise stimuli? Besides, the level profiles of the stimuli considered in the present work consisted of simple monotonic linear increase or decrease, whereas real approaching sound sources have nonlinear level profiles. Are nonlinear level variations processed similarly as linear level variations? This issue has been raised in Wilkie (Wilkie, 2015). Another issue concerns the influence of the type of listening on the present findings. This thesis only addressed the processes underlying global loudness evaluation of diotic stimuli (i.e. the same signals at both ears). What happens with binaural listening in more complex natural environments where strong interaural differences occur? The aspect of binaural integration of stationary signals in strong dichotic conditions has been addressed in Michaël Vannier's Thesis (Vannier, 2015), which was also part of the LoudNat project. The merging of both binaural and temporal integration in the case of binaural non-stationary signals has to be considered.

Finally, it would be interesting to investigate how the temporal processes presently highlighted apply to long sequences made of several auditory events (e.g., circulation noise) or stimuli made of several increasing plus decreasing portions, like in previous studies (Susini et al. 2002; Susini et al. 2007). How the processes observed at an "intermediate" time scale (a few seconds) as in the present thesis (a few seconds) are integrated at longer time scales (a few minutes)? This issue is currently examined in the framework of the Chair UMPC PSA/Renault "MoUVie". The ultimate goal is to provide an overall modeling structure capable of describing global loudness processing of time-varying sounds at any time scale.

5.5.6 Determining the neural bases of auditory asymmetries

Investigating the neural coding of increasing and decreasing-intensity sounds is a very promising research perspective. In this context, a work has been initiated in collaboration with Brice Bathellier and his team during this PhD to investigate this aspect in mice. By combining behavioral tasks and two-photon calcium imaging techniques, it is investigated how mice react to rising and falling-intensity sounds and which brain processes are involved. Strikingly, asymmetries in the cortical population responses as well as in the behavioral measures are found, with both higher cortical activity and faster reactions associated to rising-intensity sounds. A paper based on these findings has recently been submitted for publication (see Deneux et al. submitted). Investigating the neural circuits and the neural computational code underlying these cortical nonlinearities in mice will certainly help to gain a better insight into the high-level processes engaged in perceptual asymmetries as the ones investigated in this thesis.

Appendix A

Interindividual variability in the size of the asymmetry

In Chapter 1, we found that the loudness asymmetry between rising and falling tones was particularly robust to various contexts of presentation and measurement methods. Moreover, it was reported that the mean asymmetry was about 4 dB. The goal of this appendix is to gain further insight into the variability of the asymmetry collected across conditions and participants.

In Figure A.1, the asymmetries are presented for each intensity-region and for each subject in the four experimental configurations of Chapter 1 (Exp. 1, "mixed" and "separate" groups; Exp. 2A, Exp. 2B). In Exp. 1, the asymmetries were computed from magnitude estimates given to rising and falling tones presented in intensity-regions from [45-60 dB SPL] up to [75-90 dB SPL] in two contexts of stimuli presentation. One group of participants evaluated the stimuli in a "mixed context" and the other group in a "separate context". In Exp. 2A and Exp. 2B, asymmetries were measured using direct and indirect loudness-matching tasks respectively, for intensity-regions ranging from [45-60 dB SPL] up to [65-80 dB SPL] (see Chapter 1, section 1.3.1.4 for details about the experimental procedures).

From the red squares plotted in these panels, it can easily be observed that participants had, on average, very different asymmetries. These dif-

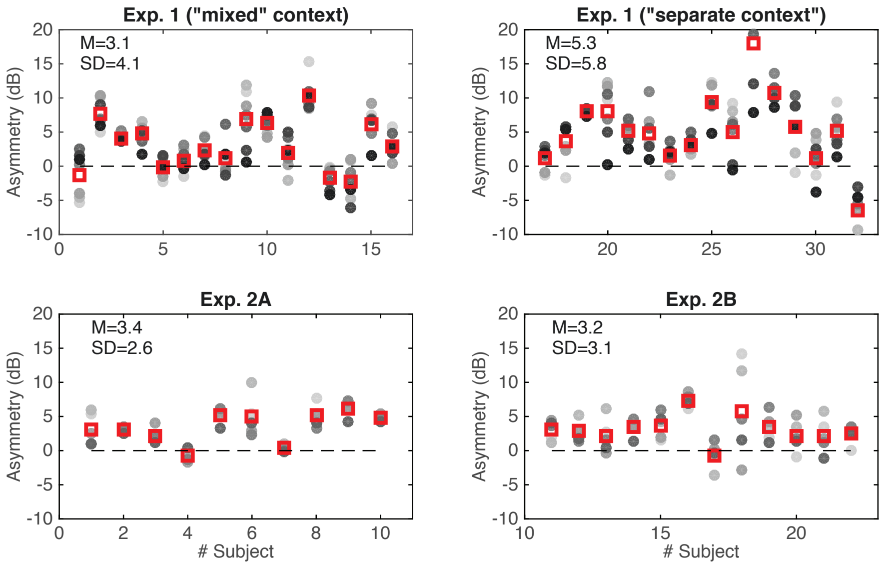

Figure A.1: The four panels show individual asymmetries obtained in each experiment of the study reported in Chapter 1, for each intensity-region (low intensity regions with dots of light grey colors; high intensity regions with dots of dark grey colors) and averaged over the different intensity-regions (red squares). Means and overall standard deviations of these asymmetries (over subjects and intensity-regions) are indicated inside each panel. Two "values" of subject 27 obtained in Exp. 1 ("separate context") being greater than 20 dB (as it can be seen on the top-right histogram presented in Figure A.2) are not visible in the top-right panel.

ferences appear to be more spread out in the two panels at the top (corresponding to magnitude estimation tasks), as compared to the two panels in the bottom (corresponding to loudness-matching tasks). Computing the standard deviations over these mean individual asymmetries (averaged over the different intensity-regions for each listener, i.e. variability of red squares in Fig. A.1) supported these observations: Exp. 1 "mixed context", SD = 3.7; Exp. 1 "separate context", SD = 5.2; Exp. 2A, SD = 2.3; Exp. 2B, SD = 2.0 (these values are those reported in Chapter 1 but not those that appear in Fig. A.1, which were calculated over subjects and intensity-regions). Many different listeners' behaviors are observed; some present clear positive asymmetries (e.g., S12 in Exp. 1, "mixed context"), others display zero or even negative asymmetries (e.g., S32 in Exp. 1, "separate context"). This finding is particularly interesting because, on an individual basis, the rule proposed in the article of Chapter 1 stating that the "mean asymmetry is equal to about 4 dB" does not apply to each listener. This variability can also be seen when considering individual results in some specific intensity-region. It is clear from Figure A.1 that whereas certain listeners had very different asymmetries across the different intensity-regions, some did not (or to a much lesser extent), such that the size of this variability across intensity-regions was very different among listeners. The fact that participants had significantly smaller asymmetries in highest intensity-regions compared to others as revealed by the overall analyses of Exp. 2A and Exp. 2B cannot be seen on an individual basis (see black dots, bottom panels, Fig. A.1).

From these analyses, we can conclude that the size of the asymmetry depends on the listener such that the "4dB rule" is only true on average. It is also found that the variability of these values is about twice as small in a loudness-matching task compared to a magnitude estimation task. However, it is not possible to determine whether this effect is obtained because one method is more accurate than the other or because listeners behave differently in the two tasks.

We also plotted the distribution of the asymmetries in the four experimental configurations of this study by pooling all the values together (across listeners and intensity-regions). The histograms obtained for each configuration are presented in Figure A.2. Figure A.3 also shows the "Grand" average histogram of the 334 estimates collected of this study, i.e. when all the asymmetries collected in Chapter 1 are aggregated. These two figures support the previous observations of a large variability of the asymmetry across listeners and intensity-regions. The distributions are significantly positive (t-tests yielded p-values $< .001$) but always contain zero (see the vertical black lines on Fig. A.2 and Fig. A.3). These data do not differ from normal family distributions, as indicated by Lilliefors tests (Conover, 1980) [Exp 1 "mixed context": K-S = .084, $p = .177$; Exp. 1 "separate context": K-S = .074, $p = .137$; Exp. 2A: K-S = .125, $p = .412$; Exp. 2B: K-S = .093, $p = .217$].

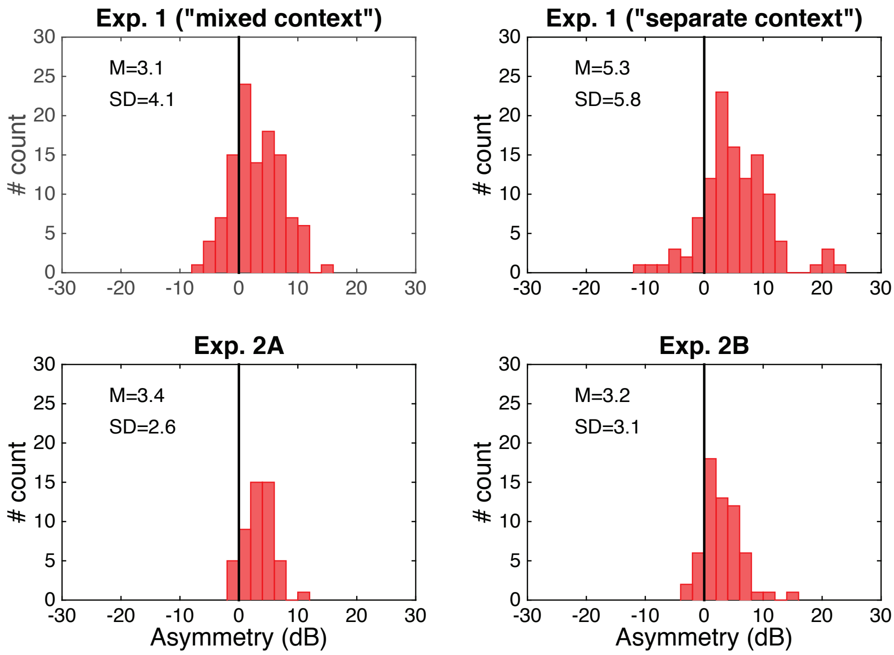

Figure A.2: Histogram of the asymmetries in the four experimental configurations of Chapter 1. Means and overall standard deviations of the data (over subjects and intensity-regions) are indicated in each panel.

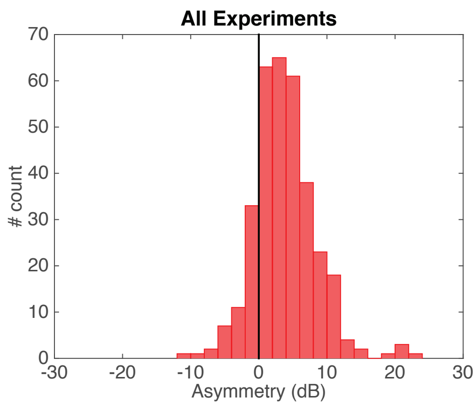

Figure A.3: Mean histogram of the asymmetries when the values obtained in the four experimental conditions of Chapter 1 are aggregated.

All in all, these results reveal large interindividual differences concerning the size of the asymmetry between rising and falling tones, and indicate that normal family distributions, whatever the measurement method was, well represent the distributions formed by individual data.

Appendix B

Asymmetries with narrow-band noises

In this appendix, we present data of an additional experiment conducted to specifically examine whether the differences in spectral bandwidth between tones and noises could explain the spectral structure effect obtained in the second experiment of Chapter 2.

In the second experiment of Chapter 2, we observed that overall, the asymmetries between stimuli made of 1-kHz tones or white noises followed similar trends as a function of the condition (intensity-region, pair order), but that they differed by a constant offset of about 1-2 dB. These results suggested that the influence of the spectral structure could thus simply be related to the bandwidth of the stimuli, such that asymmetries measured for sounds with narrow-band stimuli would also follow similar trends but would fall between these two extremes as a function of their spectral bandwidth (i.e., pure tone with an infinitely small bandwidth or noise with an infinitely large bandwidth). In order to test this hypothesis, we conducted a following experiment with a loudness-matching task similar to that of Exp. 2 (Chapter 2; see Section 2.3), where we measured asymmetries for narrow-band noises centered on 1 kHz (the frequency used in Exp. 2), which had either a 100-Hz bandwidth or a 400-Hz bandwidth (i.e. recruiting either one or several critical bands, respectively; Glasberg and Moore, 1990). This complementary experiment was conducted right after Exp. 2, i.e. in the same experimental session (with some of the subjects who already took part in Exp. 2). For this second part of the experiment, we did not make preliminary measurements to evaluate equal-loudness regions (compared to the regions where 1-kHz were presented) as we did in Exp. 2, but rather simply tested intensity-regions 5 dB lower than those used for 1-kHz tones in Exp. 2, identically for all listeners[1]. We hypothesized that (1) the same trends as those observed in Exp. 2 would be found (i.e. smaller asymmetries in the rising-falling configuration for ramps presented in the 80-dB SPL region) and (2) that asymmetries for 100-Hz bandwidth noises would be overall higher than asymmetries for 400-Hz bandwidth noises. Twelve listeners (4 men and 8 women) who already participated in Exp. 2 took part in this experiment. The design of the stimuli and the apparatus were the same as in Exp. 2 (see Section 2.3.1 for details).

The asymmetries obtained in this experiment are plotted in Figure B.1 in a similar manner as in Chapter 2, i.e. as a function of the maximum level of the stimuli intensity-region (60 or 80 dB SPL) and the presentation order (falling-rising on the left panel, rising-falling on the right panel). On average, asymmetries were positive[2], indicating that falling tones were matched higher in level than rising tones in order to be judged equally in loudness, but the size of this effect appeared to depend on the condition. A repeated-measure ANOVA was conducted with R (R Core Team, 2015) using a 2 $\times$ 2 $\times$ 2 (Spectrum, Intensity-Region, Order) factorial design to examine the results. Overall, the effect of the noise bandwidth, which showed slightly

[1] We assumed that narrow-bandwidths noises presented 5dB lower than 1-kHz tones will allow a comparison of the present measurements with those collected in the *equal-loudness* condition of Exp. 2 where white noises ramps were on average presented 5.8 dB and 6.1 dB lower than the [50-65 dB SPL] and [70-85 dB SPL] 1-kHz tones, respectively, as inferred from the preliminary loudness equalization experiment (theoretically, the 100-Hz bandwidth noise is equal in loudness to the 1-kHz tone and the 400-Hz bandwidth noise is just slightly louder than the tone due to spectral loudness summation). Since we found very comparable results before and after loudness equalization, this point should have however only a minor effect on the present results.

[2] This is true at least at the "observation level"; however, this complementary experiment comprised too few subjects to reveal whether the asymmetries were significantly positive in specific conditions.

greater asymmetries for the 400-Hz bandwidth noises compared to the 100-Hz bandwidth noises in most of the conditions, was not significant ($p > .05$). In the rising-falling order, asymmetries were smaller in the high compared to the low intensity-region whereas in the falling-rising order, the opposite was observed. These observations were supported by the statistical analysis which revealed a strong and highly significant Intensity-Region × Order interaction [$F(1,11) = 19.75, p = .001, \eta_p^2 = .642$]. No other significant effect or interaction was obtained.

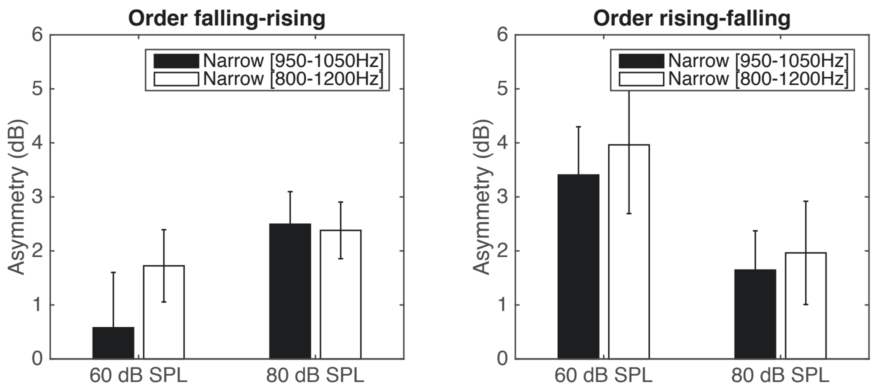

Figure B.1: Loudness asymmetries (in dB) derived from the loudness matches between rising and falling ramps, obtained for narrow-band noises centered on 1-kHz with two different bandwidths (100 Hz, i.e. [950-1050Hz] noises or 400 Hz, i.e. [800-1200Hz]). They are presented as a function of the pair order (rising-falling or falling-rising, as indicated in the title of each plot) and the maximum of the intensity-region at which the ramps were presented. Error bars show SEM.

In the rising-falling order, the asymmetries were on average 3 dB, which is very similar to what was obtained in Exp. 2 for 1-kHz tones (2.9 dB on average, see Figure 2.2, right panel). In the falling-rising order, the picture was this time slightly different to the one observed in Figure 2.2 given the fact that asymmetries in the falling-rising order appeared smaller in the low compared to the high intensity-region. To gain better insight this interaction, we conducted two additional ANOVAs (as *post-hoc* tests) to examine the effect of the intensity-region in each presentation order separately. The p-value threshold for significance was corrected using Bonferonni at the α level of .05 (.05/2 = .025). In the rising-falling order analysis, the effect of the intensity-region was still strong and highly significant [$F(1, 11) = 29.53, p < .001, \eta_p^2 = .729$]. However, in the falling-rising analysis, the effect of the intensity-region was not significant ($p > .05$). This indicates that the Intensity-Region $\times$ Order interaction revealed in the main analysis primarily resulted from the effect of the intensity-region in the rising-falling configuration, as it was the case in Exp 2 of Chapter 2.

Our first hypothesis, namely, that similar trends as those described in Exp. 2 would be found, was overall supported. However, this complementary experiment did unfortunately not provide any support to the "spectral bandwidth" hypothesis. Lastly, the present results still indicate that the asymmetry is reduced in high intensity-regions when the ramps are compared in pairs in the rising-falling order, as it was observed in Exp. 2.

Appendix C

Correlation analyses of the asymmetries measured in Chapter 2

Additional analyses and plots of the results obtained in the Exp. 2 of Chapter 2 based on correlation analyses are proposed here. The approach is similar to what was adopted to explore the relationships between simple and induced loudness adaptation (Canévet et al., 1985)). The present analyses should only be taken as tentative further insights into the properties of the underlying mechanisms governing the asymmetries.

First, it is important to remind that the main finding of Exp. 2 was that asymmetries with white noise ramps were significantly lower than asymmetries with 1-kHz tones in every condition. However, as described in Appendix A using the results of the study presented in Chapter 1, large interindividual differences were observed concerning the size of the asymmetry. However, our goal here was to make use of this large variability to investigate the correlations with respect to the size of the asymmetries measured for the same listeners measured in different conditions. We addressed two specific questions: Can correlations be found (1) between asymmetries measured with white noises and 1-kHz tones and (2) between asymmetries measured in one order of presentation compared to the other? Significant correlations might indicate a common underlying mechanism process involved for the two conditions considered (Canévet et al., 1985).

To this aim, we analyzed the individual values of the asymmetries collected throughout the experiment by computing linear correlations between asymmetries measured with the same listeners in different conditions.

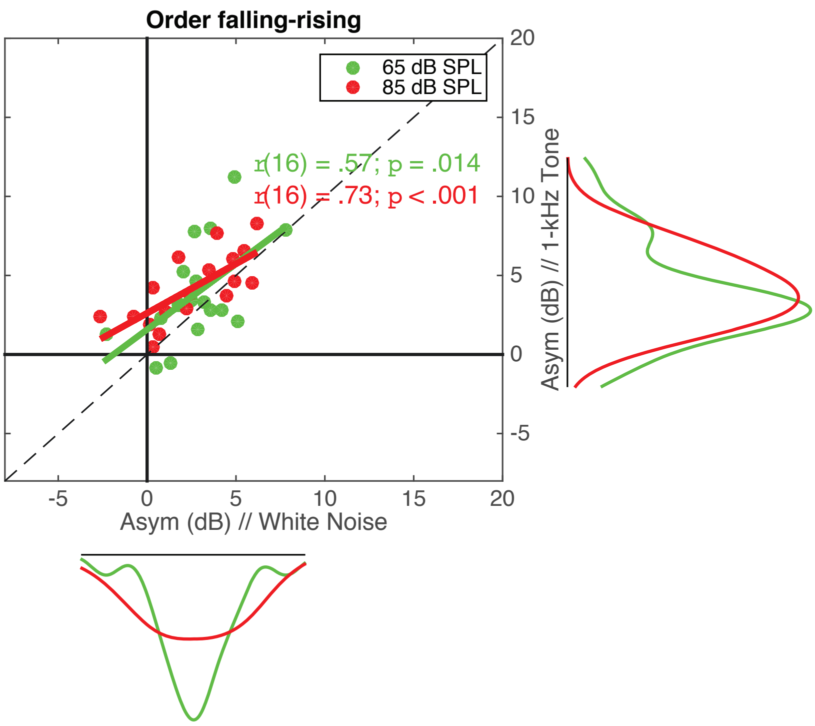

Figure C.1: Individual asymmetries obtained with 1-kHz tones against asymmetries obtained with white noise when the ramps were presented in the falling-rising order, plotted for each intensity-region (green symbols were used for the [50-65 dB SPL] ramps; red symbols for the [70-85 dB SPL] ramps). The underlying distributions estimated with kernel densities for each intensity-region are plotted on each axis. Best linear fits (in a least-squares sense) of the data are shown; corresponding correlation values and p-values are reported.

First, we analyzed the potential correlations between the asymmetries obtained with white noise and the asymmetries obtained with 1-kHz tones, either when the ramps were presented in the falling-rising order (Figure C.1) or in the rising-falling order (Figure C.2), for both intensity-regions employed in the experiment ([50-65 dB SPL] and [70-85 dB SPL]). Note that the effect of the spectral content reported in Chapter 2 is visible on Figures C.1 and C.2 because most of the points are above the main diagonal, which is consistent with greater asymmetries for 1-kHz tones compared to white noises. The results of the correlations and their associated p-values are reported in Figures C.1 and C.2. Significant correlations were obtained between the asymmetries for white noises and those for 1-kHz tones in every condition (intensity-region, pair order).

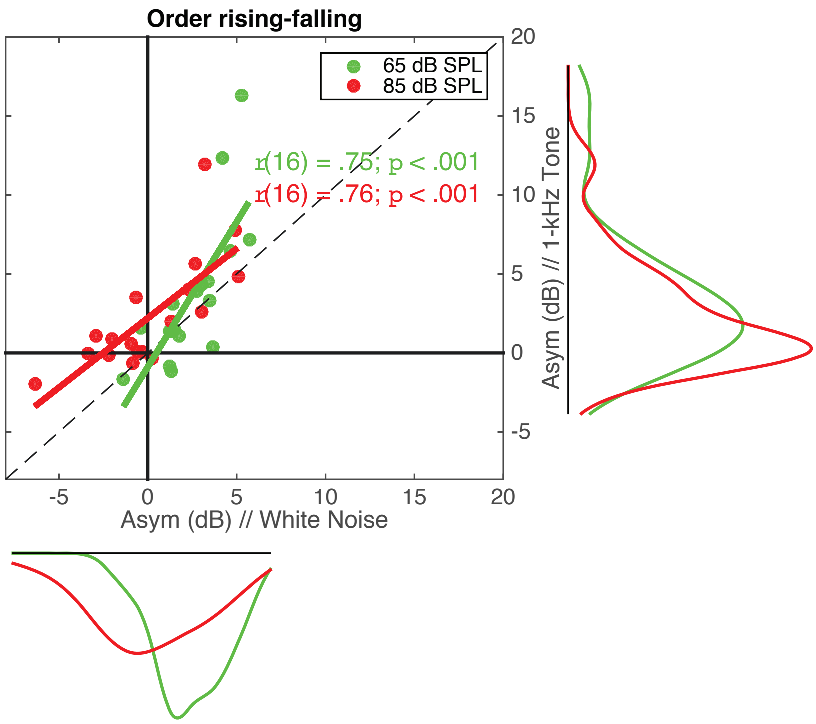

Figure C.2: Individual asymmetries obtained with 1-kHz tones against asymmetries obtained with white noise when the ramps were presented in the rising-falling order. Same format as Figure C.1.

Second, with respect to the second question presently addressed, we ana-

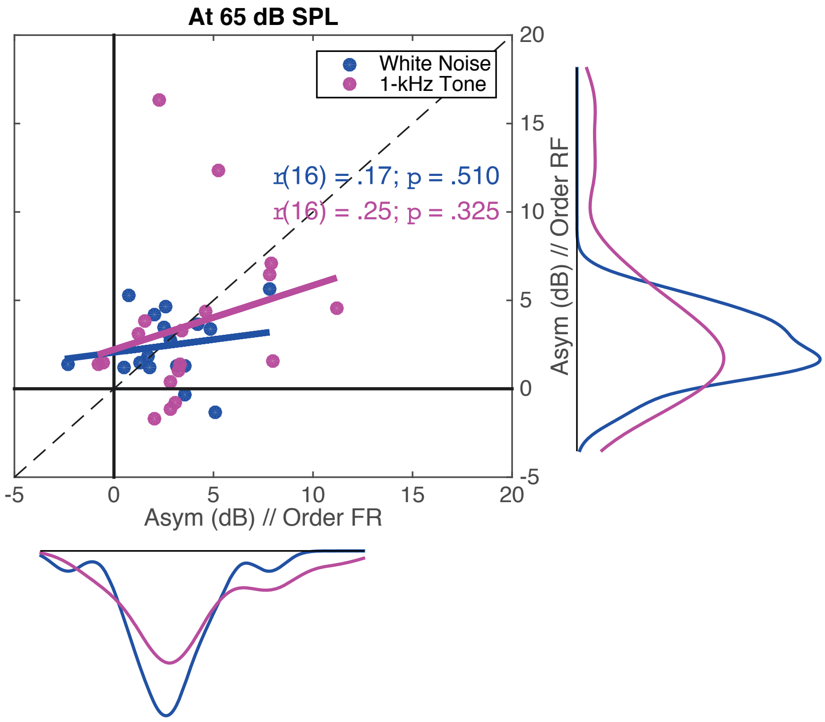

Figure C.3: Individual asymmetries with [50-65 dB SPL] ramps made of white noises (in blue) or 1-kHz tones (in magenta). Asymmetries for ramps presented in the rising-falling order (RF) are plotted against asymmetries obtained with the same subjects when the ramps were presented in the opposite order, i.e. falling-rising order (FR). The underlying distributions estimated with kernel densities for each type of spectral content (white noise, 1-kHz) are plotted on each axis. Best linear fits (in a least-squares sense) of the data are shown; corresponding correlation values and *p*-values are reported.

lyzed the correlations between the two orders of presentation for each spectral content (white noise, 1-kHz tone) in each intensity-region. Data, linear fits and correlation results are presented in Figures C.3 and C.4 (Figure C.3, [50-65 dB SPL] region; Figure C.4, [70-85 dB SPL] region). The same outcomes were found for both spectral contents: no significant correlations in the [50-65 dB SPL] region but significant correlations in the [70-85 dB SPL] region. This is an interesting finding to be related to the significant intensity-region

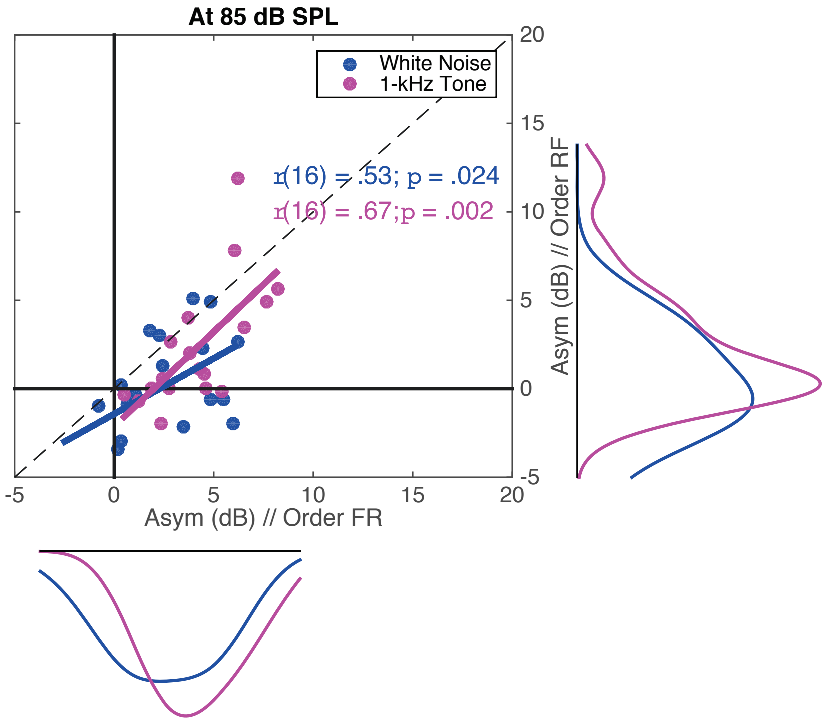

Figure C.4: Individual asymmetries and correlations obtained for [70-85 dB SPL] ramps between the two orders of presentation. Same format as Figure C.1.

× order interaction obtained in the experiment. While for the ramps presented in the [50-65 dB SPL] intensity-region the presentation order has no clear effect on the size of the asymmetry, it was significantly smaller in the rising-falling configuration compared to the falling-rising configuration when the ramps were presented in the [70-85 dB SPL] region (see Figure 2.2 in Chapter 2). The fact that significant correlations are obtained in this latter case means that listeners' asymmetries in the rising- falling order can be predicted by applying a simple linear transformation with a negative intercept to the asymmetries measured in the opposite order (or vice versa), but solely for ramps presented in high-intensity regions. At low intensity-regions, there was no significant correlation between the asymmetries in the two orders.

All in all, regarding the two issues raised in this appendix, the present analyses suggest that:

(1) It cannot be discarded that a similar mechanism would be responsible for the asymmetries observed for sounds having different spectral contents. Such a mechanism would thus be more or less “activated” depending on the spectral structure (e.g., the *tonalness*) of the stimuli. Although the correlations observed in this study were significant, they were not very high. If this “prediction error” is related to the measurement error for each spectrum, a ‘finer-grain’ measure of the asymmetry for each listener and spectral condition should increase the correlation coefficient. If not, then it could be possible that the mechanisms treating 1-kHz tones and white noises are in fact two distinct mechanisms, which only share certain similar properties, those reflected by the variance explained by the correlation coefficients only.

(2) The asymmetries of high-intensity ramps measured in pairwise comparisons were correlated between the FR order and the RF order. However, this was not found for ramps presented in lower intensity-regions, where the correlations became non-significant. We have no hypothesis to explain why the underlying processes would be similar in high intensity-regions and distinct in low intensity-regions. As a result, we do not take this result as evidence that the same mechanism is recruited for both presentation orders.

Of course, the two aforementioned statements remain to be confirmed because they solely rely on correlational analyses using the data of the second experiment of Chapter 2. It could be interesting to further investigate the former outcome and to understand how these correlations (suggesting common underlying mechanisms) can be reconciled with the view that a specific neural network is dedicated to rising tonal stimuli (e.g., Ghazanfar et al. 2002).

Appendix D

Size of the asymmetries in Chapter 3

This appendix presents global loudness asymmetries between rising and falling tones obtained in each experiment of Chapter 3. Since global loudness judgments were collected in magnitude estimation experiments, the size (in decibels) of these asymmetries was computed using the technique described in the general introduction (see Fig. 4) and detailed in Chapter 1 (we used the slope of individual loudness functions measured prior to these experiments). These asymmetries are presented in the different conditions as a function of the corresponding ramp characteristics in Figure D.1 (slope, duration dynamics, maximum level) in order to illustrate the complexity underlying asymmetries dependencies on temporal profile characteristics of ramps discussed in Chapter 3.

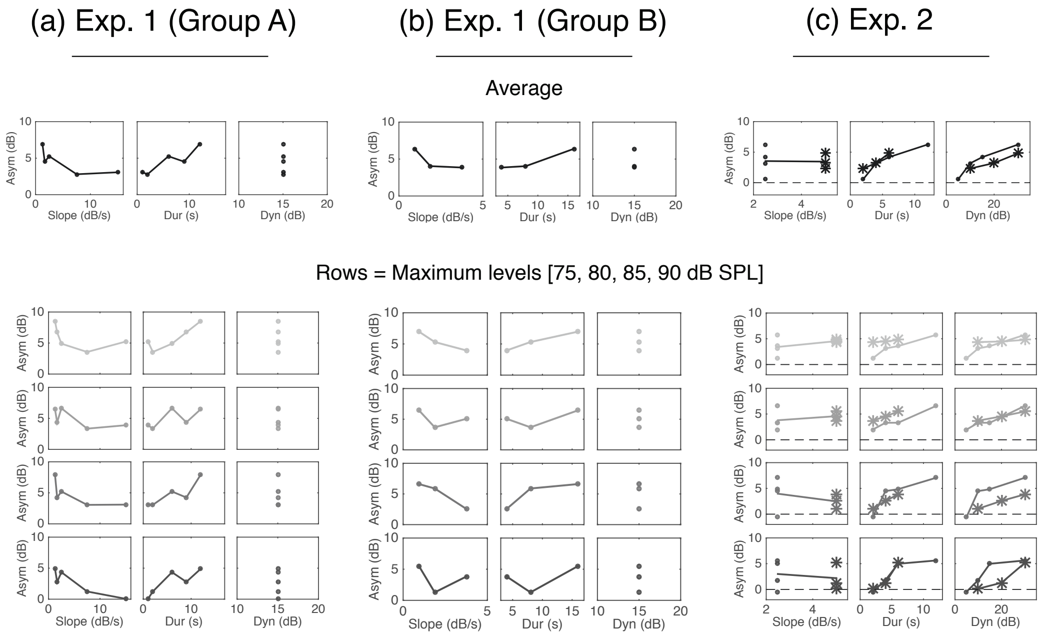

Figure D.1: Mean asymmetries (in dB) derived from the estimates obtained in each experiment of Chapter 3, calculated using the ratio of loudness estimates given to rising and falling tones and the slope of individual loudness function (for details about the procedure, see Chapter 1). Asymmetries are presented as a function of the different parameters values (slope, duration, dynamics) either on average over the different levels on the first row or for each maximum level in lower rows (from upper to lower rows: 75, 80, 85 and 90 dB SPL; light grey colors corresponding to lower levels and dark grey and dark colors to higher levels). In panel (c), the ramps varying with a slope of 5dB/s are referred to using stars symbols while dots are employed for the ramps varying at 2.5 dB/s. A positive value of x dB indicates that the rising sound was perceived x dB louder than its time-reversed falling version.

Appendix E

Psychometric functions derived from the complementary experiment in Chapter 4

This appendix presents additional analyses of the complementary experiment of Chapter 4. In this experiment, 9 participants compared level-fluctuating rising and falling ramps in terms of global loudness in a one-interval binary choice paradigm (for details, see Chapter 4). Listeners' task was to indicate which sound of the pair was louder. Half of the trials were composed of a rising stimulus followed by a falling stimulus, while the other half were composed of a falling stimulus followed by a rising stimulus. In each pair and on each trial, the reference ramp (rising or a falling) was generated from a level-fluctuating [65–80 dB SPL] intensity profile. The comparison ramp was its temporally reversed version (i.e. made of the same random fluctuations drawn for this trial) but it was presented at a different level: the same level decrement/increment selected among five possible values (-5, -2, 0, +2 or +5 dB) was subtracted/added to all the segments.

The primary goal of this experiment was to demonstrate that there were global loudness asymmetries even with time-segmented stimuli; the article thus only reported the results under the form of average percentage of responses 'rising tones perceived louder than falling tones'. However, with the present experimental configuration, psychometric functions[1] can also be computed to derive a quantitative measure of the asymmetry in dB. Individual as well as averaged psychometric functions were thus computed in the two orders of stimuli presentation. These functions are presented in Figure E.1.

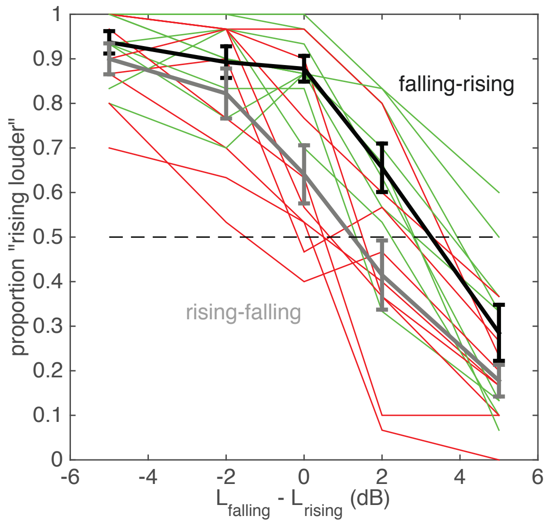

Figure E.1: Proportions of trials where the rising tone was reported as louder than the falling tone, plotted as a function of the mean level difference between the two ramps (i.e., the introduced level decrement/increment, which was equal to -5, -2, 0, +2 or +5 dB). Individual data are shown: green lines in the falling-rising order, red lines in the rising-falling order. Black and grey lines show averaged data in falling-rising and rising-falling orders, respectively. Error-bars correspond to SEM.

As it can be seen, these curves overall show that the subjects were 'biased' towards answering more often that the rising tone was louder than the

[1]Each participant was presented with 30 trials in each condition (order × comparison levels), which provides fair estimates of different points on the underlying psychometric functions.

falling tone, and this bias appeared to be larger in the falling-rising order compared to the rising-falling order. Sigmoid functions were fitted to the set of individual data using MATLAB to estimate the mean psychometric functions in each order (see Figure E.2).

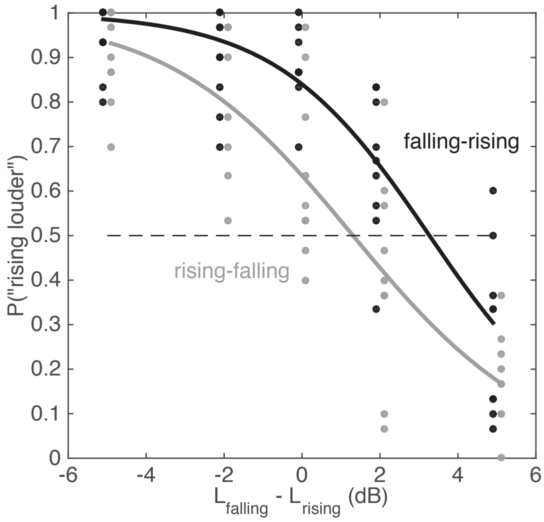

Figure E.2: Sigmoid functions fitted to the overall data (i.e. all listeners) collected in each presentation order (order rising-falling in grey; order falling-rising in black). Grey (black) dots also show individual proportions in the rising-falling (falling-rising) order. Same plotting conventions as in Figure E.1.

The average PSE, corresponding to the 0.5 value on each psychometric function, was derived for each presentation order to conduct statistical tests. Note that these measured PSEs directly reflect the asymmetry between rising and falling tones as defined everywhere in this thesis; a positive PSE of x dB indicates that a falling ramp needed to be, on average, x dB higher than a rising ramp to be perceived at equal-loudness. The asymmetries were significantly positive both in the falling-rising order [$t(8) = 6.746, p < .001$] and in the rising-falling order [$t(8) = 2.733, p = .026$], as shown by one sample t-tests. A paired-sample t-test also indicated that the asymmetry was significantly greater [$t(8) = 3.007, p = .017$] in the falling-rising order (M = 3.44 dB, SD = 1.53) compared to the rising-falling order (M = 1.28 dB, *SD* = 1.40). The present results are in particularly good agreement with those obtained in the second experiment of Chapter 2 where similar conditions as those considered here were employed. In Chapter 2, the stimuli were 2-s, 1-kHz tones varying over 15-dB regions matched in loudness in a 2I, 2AFC design using an interleaved adaptive procedure; however, these ramps were continuous (i.e. not time-segmented as in the present case) and the reference level was 5 dB higher, i.e. [70–85 dB SPL]. Despite these minor experimental differences, the asymmetry was also significantly greater in the falling-rising order (M = 4.2 dB, *SD* = 2.2) compared to the rising-falling order (M = 2.3 dB, *SD* = 3.5 dB), as indicated by a paired-sample *t*-test [$t(17) = 3.006, p = .008$].

Finally, the results reported in this Appendix for time-segmented ramps fully agreed the those of Chapter 2 for continuous ramps: rising ramps are perceived as louder than falling ramps but this asymmetry is significantly smaller when the ramps are matched in the rising-falling order compared to the falling-rising order.

Appendix F

Modeling global loudness processing of time-varying sounds: A second-order reverse correlation analysis[1]

This appendix presents an ongoing work containing preliminary results, analyses and model testing subject to further modifications.

[1]This work has been done in collaboration and under the supervision of Peter Neri.

Abstract: In this report, a computational model of the processes engaged in global loudness evaluation of basic time-varying sounds is proposed. The model is based on data previously obtained in an auditory psychophysical task where 1-kHz tones randomly varying in level over increasing or decreasing profiles were classified regarding their global loudness [Ponsot et al., J. Acoust. Soc. Am. **134**(4), EL321–EL326 (2013)]. First and second-order kernels of the system were computed using reverse correlation analysis, relating the overall loudness judgments to the level fluctuations introduced over the temporal profile of the stimuli. From the established structural properties of nonlinear systems, a LNL model (Korenberg cascade) was identified as the simplest model capable of reproducing most of listeners' behavioral characteristics in the task. Though tentative, we believe these results open the path to new and interesting research perspectives to investigate the role of nonlinear components in global loudness judgments that might be involved in judgment asymmetries observed between tones increasing and decreasing in level.

F.1 Introduction

Recent psychophysical studies have demonstrated the existence of a robust and consistent global loudness asymmetry between tones increasing vs. decreasing in level: Increasing-level tones are overall judged as being louder than their time-reversed, decreasing-level versions (Ponsot et al. 2013, 2015). Because current loudness models are not able to account for such an asymmetry, we believe that a mechanism not yet identified is presumably concerned (see Ponsot et al. 2015).

In the paper presented in Chapter 4 (i.e., Ponsot et al. 2013), we addressed the question of whether this asymmetry could be explained by a dissimilar temporal weighting of momentary loudness. A psychophysical experiment employing a reverse correlation technique (also called "molecular psychophysics" (Green, 1964)) was conducted to extract the temporal weighting of loudness underlying global loudness judgments of tones increasing and decreasing in level. A one-interval binary choice paradigm was used. The stimuli were 2-s increasing and decreasing stimuli time-segmented into sixteen 125-ms portions of constant intensity, and some "noise" was introduced on their level-profiles under the form of random level perturbations on each segment (see Figure F.1, panel (a) in the present document). Listeners' task, on each trial, was to classify the stimulus (increasing or decreasing) as "soft" or "loud" according to its global loudness as well as those of stimuli presented in previous trials. Because the responses varied as a function of the random level perturbations, the weights of each temporal portion of the stimuli in listeners' global loudness judgments could be derived using reverse correlation analysis. In this analysis, we assumed that listeners 'responses were linearly related to the level of each segment of the stimuli, as it was done in previous studies exploring temporal weighting of loudness (Pedersen and Ellermeier 2008; Rennies and Verhey 2009; Oberfeld and Plank 2011; Oberfeld et al. 2012). We found that observers exclusively weighted the loudest portion of decreasing and increasing stimuli (i.e., the beginning or the end, respectively) in a very similar fashion, such that symmetrical temporal weighting patterns were observed (see Ponsot et al. 2013). Therefore, we concluded that the asymmetry could not be explained by any dissimilar temporal weighting of loudness. These results thus suggested that the loudness asymmetries come from either (i) a larger "gain" placed on the loudest portion of increasing compared to decreasing stimuli ("gain" hypothesis[2]) or (ii) are due to nonlinear mechanisms not reflected in the linear analysis that was performed ("nonlinear weighting" hypothesis).

In the present work, the "nonlinear weighting" hypothesis is examined using a second-order reverse correlation analysis of the data collected with increasing and decreasing-level stimuli in Ponsot et al. (2013). Indeed, based on the results presented in Chapter 3 of this thesis, we came to the idea that nonlinear characteristics might reveal important information concerning the global loudness processing of this type of time-varying sounds. In the present document, nonlinearity was only addressed at the second order. Second-order characteristics have already been shown to be of primary importance to identify the mechanisms of very basic psychophysical tasks (as for example those involved in the visual detection of a bright bar embedded within spatio-temporal noise of bars varying in brightness; see Neri and Heeger 2002).

In the first part of this document, first and second-order kernels computed from the psychophysical data taken from Ponsot et al. (2013) are presented. The question of whether any learning occurred along the experiment was also addressed by splitting these data into two parts. Finally, using system identification tools, a computational model is derived from the psychophysical kernels. This model aims to simulate the circuit architecture underlying global loudness processing. Results and discussions of this report should be regarded as a preliminary attempt to address the question of how global loudness is processed using tools of nonlinear systems analysis.

[2]It is not possible to test the "gain" hypothesis using this set of data because the weighting patterns were obtained separately for increasing and decreasing stimuli (a one-interval paradigm was employed) and the weights were then normalized separately. One perspective to test this hypothesis would be to make subjects compare time-segmented increasing and decreasing stimuli against each other in a two-interval paradigm so the weights corresponding to each profile could be compared on the same scale.

F.2 Inferring observers' behavior from both linear and nonlinear kernels

As noted in the introduction, previous studies addressing the temporal weighting of loudness with "molecular psychophysics" were restricted to linear analyses, i.e. a first-order characterization of the underlying processes. To the best of our knowledge, there was only one attempt to address nonlinear characteristics potentially involved in global loudness judgment of time-varying sounds (see Pedersen 2006, pp. 44-53). Here, we propose an extension of the temporal weighting analysis up to its second-order, that is, by considering potential nonlinear components involved in the process. Most of the work presented in this document is based on results and properties of nonlinear systems presented in Westwick and Kearney (2003). Psychophysical kernel computations, analysis and modeling have been previously described in the literature (for details, one should refer to Neri and Heeger 2002; Neri 2004, 2010; Murray 2011).

F.2.1 First-order kernels

By pooling together the results of the eight listeners who took part in the experiment, a total of 14400 trials were available to derive the kernels (i.e., 7200 trials for each profile, increasing and decreasing, respectively). The first-order kernels were computed using the technique described in Neri and Heeger (2002), which was first introduced in auditory psychophysics (Ahumada 1967; Ahumada and Lovell 1971). The stimuli drawn from "high" distributions (see Fig. F.1, panel (a)) were considered as target present stimuli (we were expecting the subjects to report them as "loud") and, similarly, those derived from the "low" distributions were taken as target absent stimuli (we were expecting the subjects to report them as "soft"). The formula used to derive the first-order kernel was:

$$\mathbf{Kernel_1}(t) = \frac{1}{n_{loud/high}} \sum_{i=1}^{n_{loud/high}} \mathbf{s}_{loud/high}(t) - \frac{1}{n_{soft/high}} \sum_{i=1}^{n_{soft/high}} \mathbf{s}_{soft/high}(t)$$
$$+ \frac{1}{n_{loud/low}} \sum_{i=1}^{n_{loud/low}} \mathbf{s}_{loud/low}(t) - \frac{1}{n_{soft/low}} \sum_{i=1}^{n_{soft/low}} \mathbf{s}_{soft/low}(t)$$

where for example, $\mathbf{s}_{loud/high}(t)$ is a vector made of the 16 segment levels of a stimulus drawn from the "high" distribution that has been classified as "loud". All these vectors are averaged and then added or subtracted depending on their category (i.e., whether they were classified as "soft" or "loud"). The first-order kernels derived from these analyses are presented in Figure F.1 (panel (b)). As noted in Neri (2009), the terms of first-order kernels indicate how the different temporal portions of the stimuli are weighted. Thus, not surprisingly, they were virtually identical to the temporal weighting patterns obtained in Ponsot et al. (2013) using simple logistic regressions, and display highly symmetrical structures between the two profiles. From these kernels, it is only possible to conclude, as it was done in our paper (Ponsot et al., 2013), that participants give almost exclusive attention to the 3-5 loudest segments of the stimuli and clearly ignored (near-zero weights) what occurred on other segments, similarly for both profiles.

F.2.2 Second-order kernels

Then, second-order kernels of the system were computed using the covariance analysis technique presented in Neri and Heeger (2002)(for details, see Neri 2010). The formula used to derive the second-order kernel was:

$$\mathbf{Kernel_2}(t_1, t_2) = cov(\mathbf{S}_{loud/high}) - cov(\mathbf{S}_{soft/high})$$
$$+ cov(\mathbf{S}_{loud/low}) - cov(\mathbf{S}_{soft/low})$$

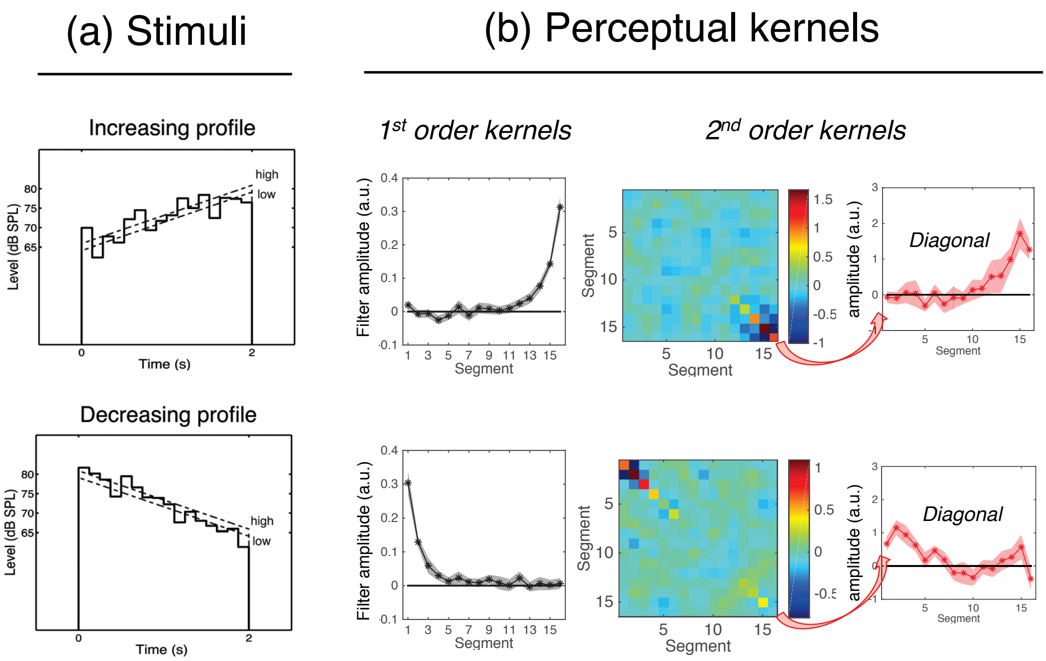

Figure F.1: Increasing and decreasing-level profile stimuli used in the psychophysical experiment are presented in panel (a). Mean first-order and second-order kernels derived from nonlinear reverse correlation analyses are shown in panel (b). Second-order kernels are wiener-filtered with MATLAB for a visualization purpose (using a [2 $\times$ 2] neighborhood matrix for the filtering). A focus is made on the diagonals of second-order kernels, which seem to display specific structures for increasing compared to decreasing profiles. Shaded areas show SEM.

where the covariance of the matrices made of the four different stimuli classes were computed, respectively. For example, $\mathbf{S}_{loud/high}$ is the [$n_{loud/high}$ $\times$ 16] matrix made of all the $\mathbf{s}_{loud/high}(t)$ vectors (the stimuli drawn from the "high" distribution classified as "loud') that have been concatenated vertically. The outputs of these covariance computations (e.g., $cov(\mathbf{S}_{loud/high})$) are two-dimensional 16 $\times$ 16 matrices, which in the end are added or subtracted from each other depending on their respective category (as for the derivation of the first-order kernels). The resulting second-order kernels are presented in Figure F.1 (panel (b)). Second-order kernels indicate how the modulations in different temporal portions interact with themselves and each other when the listener reported that the sound was "loud" (Neri, 2009). In other words, second-order kernels provide an image of the relative weights of the different interactions terms between the segment levels in the judgments. For instance, the term (i, j) indicates the relative weight of the interaction term $L_i \times L_j$ (i.e. between the levels of segments i and j) in the decision. The diagonal regions of the second-order kernels are also plotted to highlight potential differences between the two types of sounds. Nonlinear components involved in the judgment process are thus characterized by the nonzero terms. Specific structures within bottom-right and top-left corners of the kernels were obtained for increasing and decreasing stimuli, respectively. These results are particularly interesting because they demonstrate the presence of nonlinear components in the judgment process, as it can be seen both from second-order modulations of the loudest segments themselves (the positive terms of the diagonals of the kernel) and from the interactions between first and second neighboring segments (revealed by the negative flanks composing the kernel structure in these regions).

The structure observed for increasing stimuli also appears to be more spread out than the structure observed for decreasing stimuli, suggesting that nonlinear terms are of greater importance in that region for increasing stimuli. Furthermore, a small "bump" can be noticed on the last diagonal terms of the kernel obtained for the decreasing-profile stimuli. A specific look at the diagonal plots supports the observations on the overall kernels. We performed some statistical tests on raw kernel estimates to test for the significance of these observations. First, a repeated-measure ANOVA conducted to compare the terms of the diagonal of the increasing profile to the temporally inverted terms of the decreasing profile failed to show any significant difference between both profiles ($p > .05$). Concerning the presence of the 'bump' on the diagonal of the second-order kernel of the decreasing profile, a post-hoc t-test indicates for example that the diagonal term of segment 15 is significantly higher than the term of segment 8 [$t(7) = 2.578; p = .037$]; however, this significance does not survive Bonferroni correction.

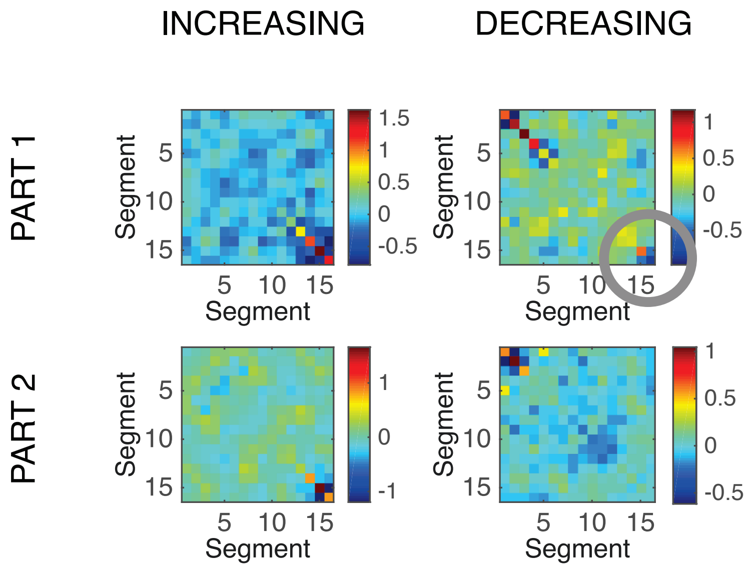

Figure F.2: Second-order kernels obtained after 2 × 2 wiener filtering for each profile (increasing or decreasing) in each part of the experiment, PART1 (i.e., sessions 2 & 3) or PART 2 (i.e., sessions 4, 5 & 6). For the decreasing profile, the 'bump' on the last diagonal terms (indicated by a grey circle) can be observed in PART 1 but not in PART 2.

F.2.3 Learning observed in second-order kernels

We investigated whether any learning could be observed throughout the experiment. The results collected in the first session were set aside because some participants did not complete this session, which served as training. The whole set of data was thus arbitrarily splitted between two parts: PART 1 comprising the data collected in sessions 2 & 3 and PART 2 comprising the data collected in the remaining sessions, namely sessions 4, 5 & 6. The first and second-order kernels of each profile were computed for each part separately, as described above. No differences were observed between the first-order kernels of PART 1 and PART 2. However, second-order kernels revealed slight but interesting differences between the two parts of the experiment. These kernels are presented in Figure F.2. Two differences can be noticed for the decreasing profile: (1) the structure around the diagonal of the kernel last longer in PART 1 compared to PART 2 and (2) the small 'bump' created by the last diagonal terms are visible in PART 1 but not in PART 2. For the increasing profile, one can also notice that the structure appears to be more spread out in PART 1 compared to PART 2. These differences can also be seen in Figure F.3 where the diagonal terms have been plotted for each profile in each part. For the decreasing profile, the 'bump' at the end of the diagonal is clear in PART 1 but not in PART 2. However, the differences between the two parts for the increasing profile are less clear.

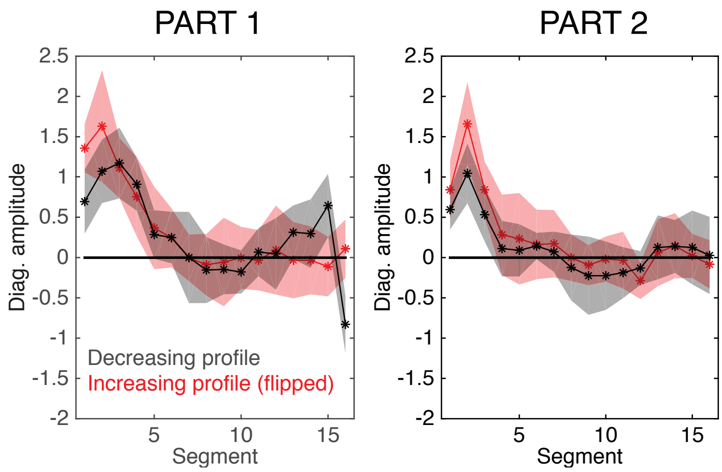

Figure F.3: Diagonal terms of second-order kernels obtained after 2 × 2 wiener filtering for each profile, increasing (in red, which has been flipped for a visualization purpose) or decreasing (in black). The left panel shows the results for PART1 (i.e., sessions 2 & 3) and the right panel shows the results for PART 2 (i.e., sessions 4, 5 & 6). It seems that the difference between the diagonal structures was present in PART 1 and then disappeared in PART 2. Shaded areas show SEM.

All these observations are intriguing because they suggest that the listeners might have "updated" their strategies during the experiment, and that the peculiarities observed in the mean kernel structures in Figure F.1 would actually reflect a specific perceptual processing occurring in PART 1 but not in PART 2. Unfortunately, all this discussion is limited to the observational level because the effects were not supported by statistical analyses, most likely because there was not enough data collected. Therefore, in what follows, we only considered the data collected in the whole experiment.

F.2.4 System identification through second-order kernel properties

Both first and second-order kernels provide important information about the functional organization of a system. If the order of a system is not higher than two, the first and second-order kernels as derived from the present analyses correspond to its first and second-order Voltera kernels, respectively (Westwick and Kearney, 2003). However, if the order of the system is higher than two, the kernel estimates as computed here may be slightly distorted (Neri, 2010). Therefore, we will make the assumption that the system presently investigated does not extend beyond second-order (as it is typically the case when conducting this type of analysis). Second-order kernels are particularly useful because they reflect properties of the underlying system and then can serve for its identification. We will not provide any mathematical formulation here; the reader is invited to refer to Westwick and Kearney's book for further details (book pages where these properties are presented will be specified).

It is easy to show from kernel properties that, among simplest cascade models that could be used to describe the process underlying the present psychophysical data, neither the Wiener model, which combines a dynamic linear function with a static nonlinearity (i.e., a LN structure) nor the Hammerstein model, which consists of a static nonlinearity followed by a dynamic linear function (NL structure) can be used. With the kernel derivation method employed here, the second-order kernel would be flat[3] (i.e., no mod-

[3]Note that this is different from the property of Wiener structures typically reported,

ulations) for a Wiener model (Neri, 2010) and there would be modulations only on the diagonal of the second order-kernel in the case of a Hammerstein model (Neri 2010; Westwick and Kearney 2003, pp. 78-80). Because of the negative flanks present around the diagonal of second-order kernels here, the present process cannot be accurately described by one of these two structures. The next more complex structure is the "sandwich" model, i.e. a LNL structure also called Wiener-Hammerstein model or Korenberg cascade (Westwick and Kearney 2003, pp. 78-83). A necessary condition to test for the LNL structure is that a sum of nonzero slices of the second-order kernel should be proportional to the first-order kernel (see p. 83). With the present data, plotted in Figure F.4, it is difficult to provide support to this condition; the traces made of the nonzero slices are very noisy. However, we believe that the LNL cannot be ruled out as easily as Wiener and Hammerstein models. An LNL structure is therefore examined in what follows to see whether it could account for the present dataset.

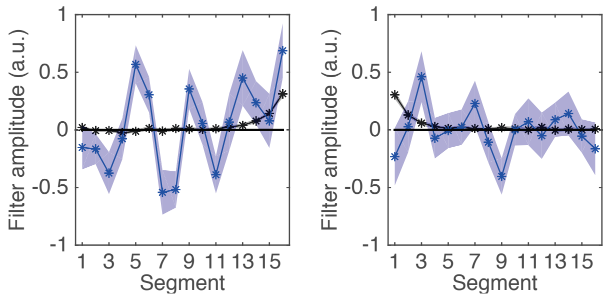

Figure F.4: Testing the condition for an LNL structure for the increasing and the decreasing-level stimuli, presented on the left and the right panel, respectively. The first-order kernels (in black) are superimposed with the sum of nonzero lines of second-order kernels (in blue); sums of slices 13-16 for the increasing profile, sums of slices 1-3 for the decreasing profile. The amplitude of each curve is arbitrary.

that any slice of their second order kernels is proportional to their first-order kernels (Westwick and Kearney 2003, pp. 75-77). Indeed, this property is found with the "standard" kernel derivation method but does hold true with the covariance derivation method.

F.3 Modeling

F.3.1 Potential modeling structure

A Korenberg model (LNL structure) is presented below, which attempts to reproduce most of the properties observed in both first-and second-order kernels of the psychophysical data described above (at least on a qualitative level). As a first step, the functions that compose this model were determined "manually" (i.e., no proper fitting procedure was adopted). The first dynamic linear function was taken as an impulse response with a negative overshoot, as visually inferred from the shape of the first line of the second-order kernels. The expression of this first dynamic linear filter was:

$$f_1(t) = t(e^{\frac{-2t}{5}} - \frac{1}{2}e^{\frac{-3t}{10}})$$

.

We used a simple half-wave rectifier with a zero threshold to model the static nonlinear function. After a few tests, we observed that, in order to produce kernels with the desired structures, the late dynamic linear function had to be a non-flat weighting function with greater weights at the beginning. The late dynamic linear function was thus chosen as a simple decreasing exponential function. After this weighting function, the components of the output vector were summed together and converted into binary responses using a simple threshold (see Fig. F.5). Note that a source of internal noise can be added so the transduction step function (which transforms a continuous variable into a binary response) becomes smoother.

F.3.2 Computations

To test such a cascade model, it is necessary to evaluate the output of thousands of trials in order to obtain stable kernels. Computations with this model were made over 100 replications of 10k trials. No response bias was considered in the model; the threshold of the late classifier of the continuous value into a binary value (corresponding to soft / loud answers) was taken

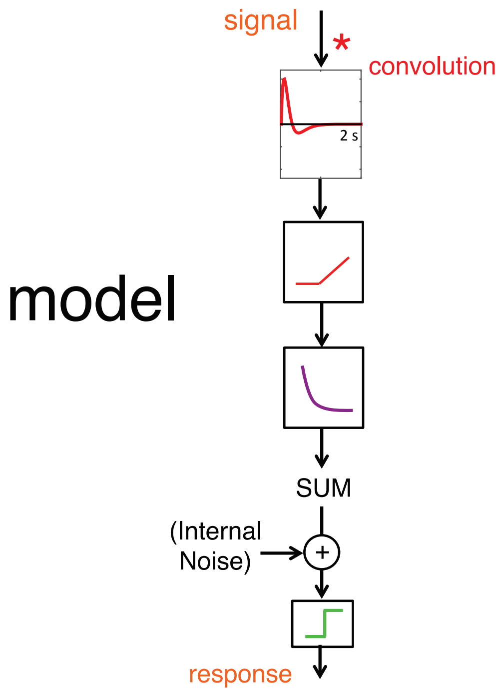

Figure F.5: Scheme of the overall LNL modeling structure used in the present report.

as the median value of each replication of 10k trials. We tested this structure with no internal noise (Internal Noise was set to zero in the model, see Figure F.5). The SNR (i.e. signal-to-noise ratio) of the stimuli (defined as the ratio between the segment *SD* and the segment mean level) was twice the standard deviation of the distribution used in the experiment in order to match human performance (in terms of *d'*). With this value of SNR, we obtained *d'* = 0.81 for both profiles, which is on average very similar to what was obtained in the experiment (*d'* = 0.86 for the decreasing profile and *d'* = 0.79 for the increasing profile). The kernels obtained with this model are presented in Figure F.6.

On average, this model reproduces most of the properties observed in the perceptual kernels. The first-order kernels of the model are almost identical to the first-order psychophysical kernels. The second-order kernel of

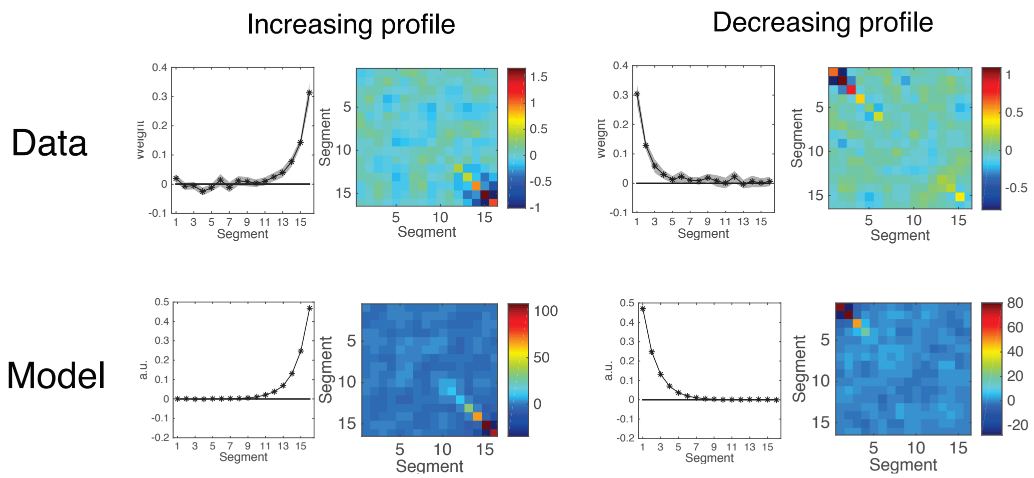

Figure F.6: First row: first-order and second-order psychophysical kernels. Second row: predicted kernels using the LNL structure presented in this report. Amplitude of each kernel is in arbitrary units. Computations were made over 100 replications of 10k trials. The mean of these computations is presented in the kernels, and the SEM between the replications is shown for the first-order kernels (second row). As before, second-order kernels are wiener-filtered for visualization purpose.

the model predicts a structure concentrated on the last diagonal terms and their segment neighbors for the increasing profile, and the mirror/flipped version for the decreasing profile. These structures show positive terms on the diagonal and negative flanks around. This is one property observed in the psychophysical kernels that could not have been obtained with LN or NL models.

We also tested whether asymmetries in global loudness judgments could arise with this simple modeling structure. To test this, we used signals without any noise on their level profile (SNR=0), i.e. linear increasing and decreasing ramps and fed them to the model. We calculated an index referred to as the " normalized asymmetry", taken as:

$$Asymmetry_{norm} = \frac{mean(output_{inc}) - mean(output_{dec})}{mean(output_{dec})}$$

where the outputs produced by the increasing or the decreasing profile before the threshold (converting the continuous output in a binary value) is applied (see Figure F.5) are compared. With the present model and set of parameters, we found that the normalized asymmetry was negative, with a mean value around -0.15. This means that the outputs produced by the increasing sound was on average 15% lower than the outputs produced by the increasing sounds. Preliminary tests indicate that the size of the asymmetry depends on both the late weighting function and the mean (i.e. DC component) of the first linear function to which the stimulus profile is convolved with. Further work has to be undertaken to properly determine which element(s) of the model give rise to this asymmetry between both profiles.

F.3.3 Limitations of the model

Three limitations of the present model can be noticed:

(1) The present model does not reproduce the little 'bump' in the second-order kernel of the decreasing stimuli. Except in kernel regions where the structure is situated, the terms of the second-order kernels characterizing the proposed LNL model are almost equal to zero everywhere. Besides, note that the same fixed structure is used throughout the computations, meaning that we did not address any potential learning effect here.

(2) Non-robustness of second-order kernels. We observed that changing SNR had no effect on the kernels. However, we found that the presence of a late internal noise or an observer's response bias removed the negative flanks in the second-order kernels.

(3) The predicted asymmetry between the outputs of the increasing and the decreasing profile is negative. This is not consistent with the fact that listeners typically report increasing-intensity sounds as louder than decreasing-intensity sounds (e.g. Ponsot et al. 2015). We did not find any LNL modeling structure capable of reproducing the kernel structures on the one hand and leading to a positive asymmetry between increasing and decreasing stimuli on the other hand.

F.4 General discussion

In this document, a second-order analysis and modeling of the data obtained for increasing and decreasing stimuli in Ponsot et al. 2013 is presented. It is shown that nonlinear components are in fact involved in the global loudness processing of increasing and decreasing-intensity sounds, a result that could not be observed in the linear analysis conducted in the paper (Ponsot et al., 2013) and that had never been specifically addressed in previous studies on global loudness. In particular, we show that listeners' global loudness judgments of increasing- and decreasing-intensity sounds are not only based on the level of loudest segments. Their judgments also result of positive contributions of the squared components (i.e. L_i^2) as well as negative contributions caused interactions between just contiguous segments ($L_i \times L_{i+1}$) and second-neighbors segments ($L_i \times L_{i+2}$) in these regions (namely, in the final portion of increasing stimuli and the starting portion of decreasing stimuli, respectively). This led to specific structures in derived second-order kernels, which were located where the loudest regions are (see Fig. F.1). The contribution and the importance of these nonlinear components in the whole process, as compared to the linear components, have yet to be determined, for example, by quantifying the gain of including nonlinear components in a model.

Furthermore, a little 'bump' was observed over the final terms of the second-order for the decreasing profile but not for the increasing profile. Separate analyses of the results collected during the first vs. the last experimental sessions revealed a possible learning effect, and perhaps that the final nonlinear 'bump' specific for decreasing stimuli would primarily occur in the first part of the experiment. The meaning and significance of this nonlinear 'bump' should be specifically addressed in future studies, because it suggests a difference in the nonlinear components underlying process of increasing and decreasing sound occurring only during the earliest experimental sessions. A LNL modeling was also presented as a potential circuitry reflecting the process of global loudness evaluation of increasing and decreasing tones. Though a few limitations have been pointed out (for example, the present model did not account for the 'bump' at the end of the second-order structure of decreasing stimuli nor for the dynamic evolution of the filters throughout the experiment (i.e., learning)), this model was able to reproduce most of the properties visible in the perceptual kernels. The model predicts asymmetries in global loudness between increasing and decreasing sounds, but it remains to be determined which element(s) of the model is (are) responsible for the asymmetry and why the present model only gives rise to negative asymmetries (decreasing predicted as louder than increasing sounds).

The reverse correlation technique employed here has been widely applied in previous studies to derive the behavior of sensory neurons or the processes engaged in various psychophysical discrimination/identification tasks, reflecting low-level sensory coding through the use of stimuli typically shorter than a few hundred milliseconds (e.g., Eggermont 1993; Ringach and Shapley 2004). This technique has also been applied at longer time-scales, but to a lesser extent (for examples of the literature, see p. 5 in Neri 2010). When dealing with long stimulus sequences like it is the case here, e.g., a few seconds, the noise introduced over the stimulus profile becomes perceptually distinguishable; a critical issue already formulated by Neri (2010): "it is unclear whether the critical parameter for the dynamics of these phenomena is time per se or rather the occurrence of a cognitive event". Global loudness processing of time-varying stimuli lasting a few-seconds falls into this category. The analyses and the modeling structure described in this report should be seen as preliminary results and as very first tentative of a nonlinear analysis and modeling of global loudness processing, which indeed takes place at high level stages of auditory processing. We believe they provide exciting perspectives to investigate global loudness processing using tools specifically dedicated to nonlinear systems. For example, it could be particularly interesting to take advantage of the large interindividual variability observed in global loudness asymmetries (see Appendix A) to investigate whether the magnitude of the asymmetry measured for a given listener can be predicted by second-order kernel properties from that specific listener.

Appendix G

Temporal weighting of loudness in two different loudness-judgment tasks[1]

Emmanuel Ponsot, Patrick Susini and Daniel Oberfeld (submitted).Temporal weighting of loudness: Comparison between two different loudness-judgment tasks.

[1]This work has been done in collaboration with Daniel Oberfeld. The article presented in this chapter has been submitted under this form.

Abstract:Psychophysical studies on loudness have so far examined the temporal weighting of loudness solely in level-discrimination tasks. Typically, listeners were asked to discriminate hundreds of level-fluctuating sounds regarding their *global loudness*. Temporal weights, i.e. the importance of each temporal portion of the stimuli for the loudness judgment, were then estimated from listeners' responses. Consistent non-uniform "u-shaped" temporal weighting patterns were observed, with greater weights assigned to the first and the last temporal portions of the stimuli, revealing significant primacy and recency effects, respectively. In this study, the question was addressed whether the same weighting pattern could be found in a traditional loudness estimation task. Temporal loudness weights were compared between a level-discrimination task (LD) and an absolute magnitude estimation task (AME). Stimuli were 3-s broadband noises consisting of 250-ms segments randomly varying in level. Listeners were asked to evaluate the global loudness of the stimuli by classifying them as "loud" or "soft" (LD), or by assigning a number representing their loudness (AME). Results showed non-uniform temporal weighting in both tasks, but also significant differences between the two tasks. An explanation based on the difference in complexity between the evaluation processes underlying each task is proposed.

G.1 Introduction

Psychophysical reverse-correlation, (Ahumada and Lovell 1971; Beard and Ahumada 1998; Ahumada 2002), also termed perceptual weight analysis (Berg, 1989), has been shown to provide a unique framework for identifying decision strategies underlying various perceptual evaluations (see Dai and Micheyl 2010; Murray 2011). In auditory research, this method has been successfully employed to examine the spectral weighting of individual components of complex sounds (e.g. Leibold et al. 2007; Leibold et al. 2009; Jesteadt et al. 2014), the temporal weighting of loudness for time-varying sounds (e.g. Pedersen and Ellermeier 2008; Oberfeld and Plank 2011; Ponsot et al. 2013), or spectro-temporal weights (Oberfeld et al., 2012). For the study of temporal loudness weighting, time-varying sounds composed of several temporal portions varying in level randomly and independently from trial-to-trial are typically judged in terms of *global loudness* (i.e., the loudness of the sound as a whole) over several hundred trials. Assuming that listeners' overall loudness judgments are based on a linear combination of each segment level, the relative weight of each portion can then be estimated, for example using multiple logistic regression (Oberfeld, 2009). The weights thus obtained indicate how strongly the global loudness is impacted when the level of a temporal portion of the sound is changed. For stimuli with a flat level profile (i.e., all temporal portions of the sound are drawn from distributions having the same mean level), these studies consistently showed that the first 100-300 ms receive a higher weight than later portions of the stimulus (i.e., a primacy effect) (Ellermeier and Schrödl 2000; Oberfeld 2008; Pedersen and Ellermeier 2008; Dittrich and Oberfeld 2009; Rennies and Verhey 2009; Oberfeld and Plank 2011; Oberfeld et al. 2012). This means that, for example, a 1 dB increase in the level of the first 100 ms of the sound causes a stronger increase in global loudness than a 1 dB increase in the level of the final 100 ms. Some studies also found a small recency effect (Pedersen and Ellermeier 2008; Ponsot et al. 2013), but this effects appeared to be weaker than the primacy effect (Dittrich and Oberfeld 2009; Oberfeld and Plank 2011). The observed non-uniform temporal weights are an important outcome for research in loudness, because primacy and recency effects are incompatible with the uniform temporal weighting assumed in current indicators of loudness such as L_{Aeq} or N_5. For example, a 1 dB increase in the level of the first 100 ms of a sound would result in the same increase in L_{Aeq} as a 1 dB increase in any other temporal portion of the signal if the level and spectral content is equal for the two temporal portions. In this context, Oberfeld and Plank (2011) demonstrated that adding temporal weights to current loudness indicators provides significantly better predictions of the psychophysical data.

Previous studies on the temporal weighting of loudness used sample (or level) discrimination tasks (Berg and Robinson, 1987). In these tasks, the levels of the different temporal portions of the sound (temporal segments) are drawn from random distributions. In case a one-interval task is used, there are two level distributions, one with a higher mean level (e.g., 61 dB SPL), and one with a lower mean level (e.g., 59 dB SPL) (e.g., Dittrich and Oberfeld 2009). On each trial, the segment levels of the stimulus are all commonly drawn either from the “high” or from the “low” distribution (selected randomly). Listeners’ task is to decide whether the presented sound was rather “loud” (i.e., originated from the higher level distribution) or rather “soft” (i.e., originated from the lower level distribution). Thus, it corresponds to a one-interval, two-alternative forced-choice level-discrimination task (i.e., an absolute identification task; Braida and Durlach 1972). This is an objective task (i.e., responses can be classified as being correct or incorrect), so that the accuracy or the sensitivity can be computed. If each temporal segment provides the same amount of information concerning the correct response (which was typically the case in previous experiments), then the “ideal” strategy (maximizing the accuracy) in this task would be to apply identical weights to all temporal portions of a sound (Berg, 1989). However, as discussed above, listeners’ weighting strategies are found to be significantly different from this flat “ideal weighting strategy”. The use of sample discrimination tasks with two-interval paradigms is also very common (e.g., Rennies and Verhey 2009; Oberfeld and Plank 2011), and results in very similar temporal weights to those described for one-interval tasks (Oberfeld and Plank, 2011).

These findings raise the question: How can the performance in the level-discrimination task be reconciled with loudness judgments that are typically obtained using more “traditional” loudness judgment tasks, such as magnitude estimation (e.g., Stevens 1956)? That is, to which extent do the weighting patterns observed in level-discrimination tasks reflect the temporal weighting underlying a loudness evaluation in, for example, a magnitude estimation task? The aim of the present study was to address these issues by investigating whether the pattern of temporal weights estimated in a level discrimination (LD) task could also be found in an absolute magnitude estimation (AME) task (Hellman and Zwislocki, 1963).

Our hypothesis regarding this research question was that the temporal weighting pattern observed in the AME task should be similar to the pattern observed in the LD task. Thus, we expected a pronounced primacy effect and probably a weaker recency effect. Indeed, the internal decision variable used by observers in LD tasks as in AME tasks should be based on global loudness in both cases. As discussed by Oberfeld and Plank 2011, in the sample discrimination tasks, the listeners are typically instructed to classify each sound as being either “soft” or “loud”, that is, to evaluate the global loudness of each sound with respect to the loudness of the previous sounds presented in a given block. Thus, it seems reasonable to assume that the subjective quality or sensory continuum on which listeners base their decisions in sample discrimination tasks is loudness (cf. Green and Swets 1966; Durlach and Braida 1969).

In the present study, two experiments were conducted. In within-subjects designs (to ensure good statistical power in the presence of interindividual differences), temporal loudness weights were measured and compared between AME and LD tasks. A number of experimental constraints were considered to accurately compare the weighting patterns between the two tasks, which deserve to be mentioned here. A first concern was related to the choice of the experimental procedures and the presentation paradigms. Since the AME task is a procedure where the stimuli are presented and evaluated one by one, we opted for a LD task based on a one-interval paradigm. A second experimental concern was that the stimuli had to induce sufficient variability in the magnitude estimates to estimate the weights. For this reason, the segment levels were drawn from normal distributions with comparably large standard deviations ($SD = 5$ dB), while previous studies used SDs between 2 and 3 dB. Finally, to minimize undesired psychoacoustical interaction and transient effects between consecutive segments caused by these large level modulations (such as forward masking caused by a loud segment preceding a much softer segment), the segments were rather long (250-ms), and cos2 on- and off ramps were used to smooth intersegment amplitude variations (for further details, see the procedure section below). Thus, the stimuli presented in this study were comparably long noises (3 s) fluctuating slowly (4 Hz) in level. In comparison, previous studies presented shorter sounds (< 1.5 s total duration) with shorter temporal segments (typically 100 ms) and smaller level fluctuations (e.g., *SDs* of 2 dB).

In the first experiment, the stimuli presented in the two tasks were varying around different mean levels. In the LD task, the segment levels were drawn from random distributions with means at 63.5 dB SPL ("low" distribution) and 66.5 dB SPL ("high" distribution) while in the AME task, random distributions with means at 54, 61, 68 and 75 dB SPL were used. We introduced this large range of levels (21 dB) in the AME task to ensure that subjects could easily judge the loudness of the stimuli by using different numbers. The second experiment was conducted to explore whether the results obtained in the first experiment were related, at least in part, to the larger range of stimuli levels used in the AME task compared to the LD task. To answer this question, in the second experiment, the stimuli presented in the AME task varied around the same mean levels (63.5 dB SPL and 66.5 dB SPL) as in the LD task.

G.2 Experiment 1

G.2.1 Materials and method

G.2.1.1 Participants

Seven subjects (4 women, 3 men; age 23-31 years) participated voluntarily. All reported normal hearing. They gave their informed written consent according to the Declaration of Helsinki prior to the experiment and were paid for their participation. The participants were naïve with respect to the hypotheses under test.

G.2.1.2 Stimuli

The stimuli were white (broadband) noises lasting 3 s. Static (constant-intensity) noises were used in the first session of the AME task, presented diotically with levels of 54, 61, 68 or 75 dB SPL. Otherwise, all the stimuli were level-fluctuating noises made of 12 consecutive 250-ms stationary noise segments with levels drawn independently from normal-truncated distributions (SD = 5 dB, restricted to Mean ± 2.5 SD). As mentioned above, the modulation depth of these noises was chosen to be sufficiently large to produce variability in participants' judgments, so that the temporal weights could be estimated. While 50-msec linear ramps were imposed on the amplitude envelopes at the onset and the offset of the stimuli, inter-segment level variations were smoothed using 100-ms cos^2 on and off ramps to avoid unwanted abrupt changes of sound intensity or temporal loudness masking effects. On each trial, the levels of all segments were randomly and independently drawn from the same normal distribution. In the AME task, the mean of the distribution was 54, 61, 68 or 75 dB SPL, presented with equal probability. In the LD task, the distribution means were 63.5 or 66.5 dB SPL, selected with equal probability.

G.2.1.3 Apparatus

The stimuli were generated at a sampling rate of 44.1 kHz with 16-bit resolution using Matlab. Sounds were converted using a RME Fireface 800 soundcard and presented diotically through headphones (Sennheiser HD 250 Linear II). Sound level was calibrated using a Brüel & Kjær artificial ear (Type 4153, IEC318). Participants were tested individually in a double-walled IAC sound-insulated booth at Ircam.

G.2.1.4 Procedure

The experiment was divided into two parts, which consisted of two different psychophysical tasks, as described below. The participants performed the two parts one after the other; their order of presentation was counterbalanced between participants. Each part involved several sessions scheduled on different days; Part LD refers to the level discrimination (LD) task that consisted of three 1-h sessions, Part AME refers to the Absolute Magnitude Estimation (AME) task, which was divided into five 1-h sessions.

In Part LD, a 1I, 2AFC procedure was employed. On each trial, a sound was presented with the segment levels drawn either from the "high" or the "low" distribution (see above), randomly chosen with *a priori* equal probability. The participant decided whether it had been presented a "loud" or a "soft" type of stimulus. Listeners were explicitly asked to consider the *global loudness* of the stimuli when making their judgment, corresponding to the judgment of the loudness over the entire duration of the sound (Pedersen and Ellermeier 2008; Ponsot et al. 2013). Each session comprised five blocks of 90 trials each. The answers collected during the first session of this part, which served as a training session, were removed from the analysis. Thus, a total of 900 trials were collected per listener in this task. The participants did not receive any trial-by-trial feedback, but the percentage of correct identifications of the "low" / "high" distributions (defined as corresponding to "soft" / "loud" responses, respectively) were displayed on the computer screen at the end of each block to ensure sustained attention on the task (Ponsot et al., 2013).

In Part AME, an absolute magnitude estimation (AME) procedure was used. No standard / modulus corresponding to a certain number was provided. The task was simply to give a number best representing the *global loudness* of each sound, regardless to the numbers assigned to previous stimuli (Hellman, 1976). There was no training session in this part. In the first session of the AME part, the static broadband noises were presented to the participants during 160 trials (40 repetitions of each stimulus), in random order. The participants produced magnitude estimates of loudness using their own scale. Level-fluctuating noises were then presented in the four sessions that followed. In these sessions, each trial consisted of a level-fluctuating noise drawn from one of the four defined mean levels (see above). Each "level-fluctuating" session comprised 250 trials, divided into three blocks (90-80-80 trials; the 10 first estimates of the 1rst block served as a training and were removed from the analysis). Thus, a total of 960 "level-fluctuating" trials were collected per listener in the AME task (240 trials per mean level and listener).

G.2.1.5 Fitting loudness functions

The magnitude estimates provided by each participant during the AME task were fitted with simple loudness functions. First, the levels (in dB SPL) of the stimuli (both static and level-fluctuating noises) presented in this part were converted into equivalent pressure units (in Pascals) using the following formula:

$$\mathrm{p}_{eq} = p_0 10^{L_{eq}/20},$$

where $p_0 = 20$ μPa and L_{eq} denotes the energy-equivalent sound pressure level of the stimulus (in dB).

The magnitude estimates of the listeners were then fitted individually using a power function, which represents one of the most simple loudness functions (cf. Suzuki and Takeshima 2004; Oberfeld et al. 2012):

$$\mathrm{E} = \mathrm{kp}_{eq}^{\alpha},$$

where E is the magnitude estimate (i.e. number) produced by the participant, and p_{eq} is the equivalent pressure of the stimulus. The constants k

and α were estimated by non-linear regressions for each listener, for static and level-fluctuating noises separately. Thus, a total of 14 regressions were conducted. Individual loudness exponents estimated by the regressions conducted on level-fluctuating noises estimates were used afterwards in the decision models to estimate temporal weights in both the AME and the LD tasks (see below).

G.2.1.6 Decisions models

Multiple regression analyses were used to estimate the temporal weights in the two tasks. In previous studies, the predictors used in the models were simply based on the sound pressure level of the temporal segment levels (e.g. Oberfeld and Plank 2011). However, because of the large variations imposed on the segment levels of the stimuli in the present study, we decided to use predictors based on *loudness* in order to match more accurately the human perceptual intensity scale. The loudness of each segment N_i was estimated using the individual loudness exponents α obtained from the regressions conducted on the estimates attributed to level-fluctuating noises in the AME task (the individual values of α were taken from the fitted level-fluctuating noises loudness functions are reported in Table G.1):

$$\mathrm{N}_i = p_i^{\alpha},$$

where p_i is the equivalent pressure (in Pa) of the i^{th} segment, computed according to Eq. G.2.1.5. Thus, 12 predictors, corresponding to the loudness values of the 12 segments, were used in the regression models to estimate the weights in the two tasks. In the AME task, the dependent variable (DV) was the magnitude estimate (number) given by the participants to evaluate global loudness. In the LD task, the DV was the binary response ("soft" or "loud") entered by the participants to evaluate global loudness.

In the AME task, estimates were assumed to be a linear combination of the loudness values of the 12 segments. Segment levels L_i (in dB) were first converted into pressure units p_i (in Pa) using the equation G.2.1.5. Second, using equation G.2.1.6, we estimated their equivalent loudness levels N_i. Therefore, the estimate (E) could be expressed as the linear combination of each segment loudness N_i:

$$\text{E (N)} = \sum_{i=1}^{12} w_i N_i + c,$$

where N is the vector of segment loudness values, N_i is the loudness of segment i, w_i is the perceptual weight assigned to segment i, and c is a constant.

The decision model chosen to account for the LD task was similar to previous studies on temporal weighting of loudness (e.g. Oberfeld and Plank 2011), except that the loudness values rather than the sound pressure levels of the segments were used as predictors. The loudness of each segment (N_i) was estimated using equation G.2.1.6, i.e., using the exponents inferred from the estimates given to fluctuating-noises in the AME task. The decision variable D was also assumed to be a linear combination of the loudness of the 12 segments,

$$\text{D (N)} = \sum_{i=1}^{12} w_i N_i + c,$$

where N is the vector of segment loudness values, N_i corresponds to the loudness of the i^{th} segment, the w_i are the perceptual weights, and c is a constant. The model assumes that a listener responds that the noise presented on a given trial was loud rather than soft if $D(N) > 0$, and that

$$\text{p("loud")} = \frac{e^{D(N)}}{1 + e^{D(N)}},$$

which corresponds to a logistic regression model (McCullagh and Nelder, 1989).

These two decision models are strictly identical except that the dependent variable is continuous in the first model (AME task) and binary in the second model (LD task). Multiple logistic regression was thus used to estimate the temporal weights in the LD task (Pedersen and Ellermeier 2008; Oberfeld and Plank 2011; Ponsot et al. 2013), and multiple linear regression was used in the AME task. Regressions were conducted separately for each listener and task. For each task, the twelve regression coefficients w_i were taken as the twelve temporal weights. The weights were normalized individually so the sum of their absolute values was 1 (Kortekaas et al., 2003). Statistical analyses were conducted with R (R Core Team, 2015).

G.2.2 Results

G.2.2.1 Loudness functions

Magnitude estimates were obtained for both static and level-fluctuating sounds in the AME task of Exp. 1 for each listener. As an example, the results obtained for one subject (S7) are reported in Figure G.1. Overall, subjects were well able to produce numbers not only reflecting the mean level of the stimuli but also reflecting level-modulations that were introduced: four overlapping scatterplots could be observed for each subject. Power functions were used to fit the magnitude estimates of the static noises and the level-fluctuating noises separately (see above). These fits correspond to the black line and the grey line as plotted in Fig. G.1 for S7, respectively. The parameters of each individual power function, estimated using non-linear least square fits, are reported in Table G.1. Except for S5, who deliberately changed his scale between the first (static noises) and the following sessions in the AME task (level-fluctuating noises) (indeed, the subject told the experimenter he wanted to use another scale with greater numbers – as it can be seen from the change of intercept reflected by the parameter k in Table G.1), very similar loudness function parameters (k and α) were obtained for static and level-fluctuating noises. The goodness of fit of these non-linear regressions, evaluated in terms of the proportion of variance accounted for (R^2), was reasonably high. It was only marginally significantly higher for static than for level-fluctuating noises [$t(6) = 2.286, p = .062$]. Loudness exponents α (see Table G.1) were in line with the values reported in the literature for white noise stimuli (e.g. Canévet et al. 2003;Teghtsoo-

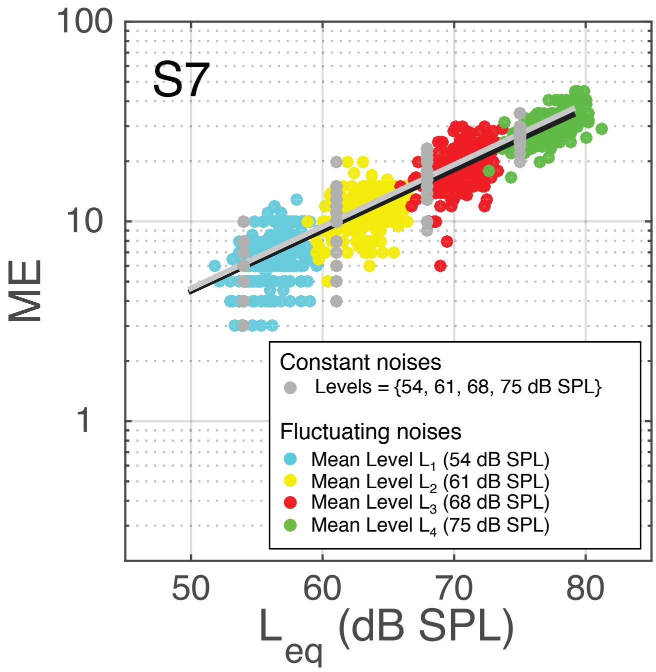

Figure G.1: Raw magnitude estimates produced by one subject (S7) in the AME task of Exp. 1. Subjects evaluated the global loudness of static noises (grey dots) and level-fluctuating noises (colored dots) at different mean levels. Estimates are plotted on a log-scaled y-axis as a function of the L_{eq} of the stimuli. Separate power functions were fitted to the MEs for constant noises estimates (grey line) and for level-varying noises estimates (black line). Parameters of the loudness functions are reported in Table G.1 for the different subjects.

nian et al. 2005). The exponents were significantly greater for static noises than for level-fluctuating noises [$t(6) = 3.911, p = .008$]. Finally, because the subjects were free to use their own response scale in the AME procedure, comparably large inter-individual differences were found regarding the multiplicative constant k (see the values reported in Table G.1), compatible with the literature (Hellman and Meiselman, 1988).

Experiment 1

		S1	S2	S3	S4	S5	S6	S7	**Average**	*SD*
Constant noises	k	31.73	21.48	206.27	16.64	2.48	24.39	106.29	**58.47**	73.33
	α	0.53	0.45	0.57	0.34	0.28	0.72	0.62	**0.50**	0.16
	R^2	.78	.75	.90	.80	.87	.87	.88	**.83**	.06
Fluctuating noises	k	19.2	10.16	122.71	14.81	9.89	10.34	97.55	**40.67**	48.12
	α	0.36	0.29	0.45	0.33	0.2	0.56	0.61	**0.39**	0.15
	R^2	.66	.78	.86	.64	.61	.77	.90	**.74**	.11

Table G.1: Parameters of the individual loudness functions (α, k) estimated by fitting non-linear power functions to the numbers assigned to both constant and level-fluctuating noises in the AME task of Exp. 1. Columns represent listeners (S1 to S7). For each fitted model, ordinary-R^2 obtained by the regression is indicated as a measure of goodness-of-fit.

G.2.2.2 Temporal weighting patterns

As explained above, the exponent of the loudness functions of fluctuating noises estimates obtained for each subject was fed into the regression models to estimate the temporal weights in each task. The temporal weights were then obtained for the 7 subjects in the two tasks. The averaged temporal weights obtained in the two tasks are presented in Figure G.2. As noted above, an “ideal observer” would apply uniform weights to the 12 segments (at least in the LD task), because each temporal segment element provides the same amount of information concerning the “correct” response (Berg, 1989). The data presented in Fig. G.2 revealed significant deviations from this uniform weighting pattern, as indicated by the confidence intervals of the weights that do not contain the horizontal black line corresponding to uniform weighting. Moreover, the weighting patterns obtained in the two tasks were rather different. In the AME task, a clear primacy effect but no recency effect was observed whereas in the LD task a small recency and a weaker primacy effect could be noticed.

The normalized weights were analyzed with a repeated-measures analysis of variance (rmANOVA), with the within-subjects factors segment and task, using a univariate approach with Huynh-Feldt (1976) correction for the

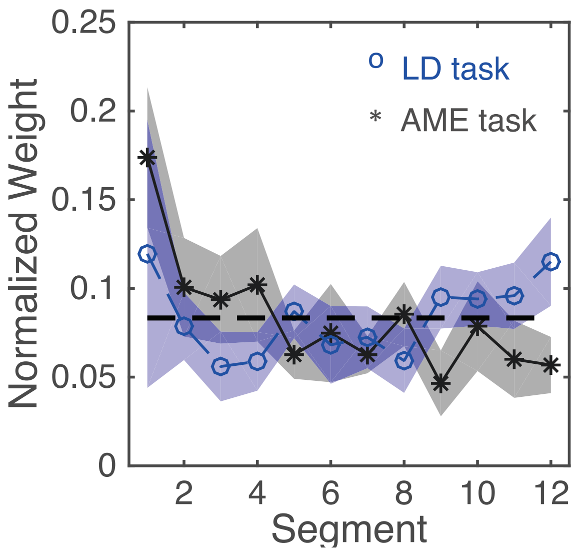

Figure G.2: Mean normalized temporal weights for both tasks of Exp. 1, presented as a function of the segment position (1-12). Blue dots symbols are used for the level discrimination (LD) task and grey asterisks are used for the absolute magnitude estimation (AME) task. Shaded areas correspond to 95 % confidence intervals. The horizontal dotted line represents the uniform temporal weighting pattern of an "ideal observer" in the LD task.

degrees of freedom. Unless otherwise specified, all the tests were two-tailed and used a probability level of .05 to test for significance. Effect sizes are reported using partial eta-squared, η_p^2. The effect of segment was significant $[F(11, 66) = 3.509, p = .032, \eta_p^2 = .369, \tilde{\epsilon} = .30]$, showing that the segments did not receive a uniform weighting. A significant segment × task interaction was found $[F(11, 66) = 5.014, p < .001, \eta_p^2 = .455, \tilde{\epsilon} = 1.0]$, supporting the view that a different weighting was applied in the two tasks. As it can be observed in Figure G.2, larger weights were applied to the first segments of the stimuli in the AME task, while moderately larger weights were applied to the first segment and to the final three segments in the LD task. Due to the weights normalization, there was of course no significant effect of task (F = 1.0). Additional ANOVAs with segment as a within-subjects factor were performed for each task separately. For the AME task, a significant effect of segment confirmed the presence of non-uniform weights [$F(11, 66) = 6.781, p < .001, \eta_p^2 = .531, \tilde{\epsilon} = .83$]. For the LD task, the effect of segment was not significant [$F(11, 66) = 2.018, p = .149, \eta_p^2 = .252, \tilde{\epsilon} = .27$]. One may notice the unexpected higher weight obtained for the fifth segment compared to other segments situated in the middle section of the stimulus. We have no explanation for this result.

Deviations from the flat weighting pattern were quantified on an individual basis by computing the coefficient of variation CV of the 12 weights (SD/M), for each listener and each task (see Oberfeld and Plank 2011). The mean CV was lower in the LD task (M = 0.45, SD = 0.27) than in the AME task (M = 0.54, SD = 0.16), indicating slightly more uniform weights in the LD task, but the difference was not significant [$t(6) = 1.058, p = .331$]. One subject (S3) showed particularly strong primacy effects in the two tasks, and higher CVs than the remaining subjects.

G.2.2.3 Predictive power of the decision models

The goodness of fit of the decision models was evaluated using the area under the ROC curve (AUC) for the models of the LD task, which is an index of the predictive power of the logistic regression model (Dittrich and Oberfeld, 2009). The proportion of variance (R^2) accounted for was used in the AME task. Fair to good model predictions were found for the two tasks: for the LD task, AUC ranged from .81 to .93 (M = .87, SD = .036). For the AME task, R^2 ranged from .61 to .90 (M = .74, SD = .11).

G.2.2.4 Increased predictive power by including temporal weights

The gain of using estimated temporal weights to predict loudness was evaluated in the two tasks. This was done by comparing different models to predict the present results: Restricted models containing only L_{eq} as predictor were compared with full models containing L_{eq} plus the twelve temporal weights as predictors (see Dittrich and Oberfeld 2009; Oberfeld and Plank 2011). The results are reported in Table G.2.

In the LD task, the AUC of the full model was significantly higher than the AUC of the restricted model, [$t(6) = 2.607, p = .040$]. Although this

Experiment 1

Goodness-of-fit index (Task)	Model	S1	S2	S3	S4	S5	S6	S7	**Average**	SD
AUC (LD Task)	restricted	.881	.847	.841	.878	.882	.877	.922	**.876**	.027
	full	.892	.869	.903	.884	.890	.889	.937	**.895**	.022
	p-value	.003	.001	$< .001$	.335	.071	.001	.001		
R^2 (AME Task)	restricted	.652	.777	.851	.638	.608	.761	.895	**.740**	0.111
	full	.665	.782	.869	.641	.614	.767	.901	**.749**	.113
	p-value	.001	.063	$< .001$	.756	.268	.013	.001		

Table G.2: Goodness of fit of full models (containing as predictors the L_{eq} and the 12 temporal weights) and restricted models (containing only the L_{eq} as predictor) used to predict individual loudness judgments obtained in each task of Exp. 1. Different indexes were employed to compare the models in the two tasks: R^2 for the AME task, AUC for the LD task. Outputs of likelihood-ratio tests (AUC) and F-tests (R^2) are reported to indicate significant improvements of the full model containing temporal weights as compared to the restricted model in each case. Columns represent listeners (S1 to S7).

difference in AUC is small, Cohen's d_z (Cohen, 1988) measuring the effect size indicates a large effect size ($d_z = 0.985$). Individual likelihood-ratio tests were conducted to compare the goodness-of-fit of the full and the restricted models (see Oberfeld and Plank 2011). A significantly better goodness-of-fit ($p < .05$) was obtained with the full model for all subjects but S4 and S5, who correspond to those having the more flat weighting patterns (i.e., smaller CVs).

In the AME task, the R^2 of the full model was significantly higher than the R^2 of the restricted model [$t(6) = 4.098, p = .006$]. The effect size was large, $d_z = 1.549$. F-tests on individual data showed that a significantly better goodness-of-fit was obtained with the full model for subjects S1, S3, S6 and S7 ($p < .05$).

Overall, these results indicate a significant gain of using the temporal weights to predict judgments both in the LD task and in the AME task. However, the AUC and R^2 only increased by small factors (2% and 1%

respectively). The likely reason for the small improvement by including temporal weights in the AME task is the large variation in overall loudness due to the variation of mean level from 54 to 75 dB SPL. Indeed, this 21 dB variation in level, which is completely captured by the L_{eq}, accounts for the greatest part of the variance of the magnitude estimates. Even with the rather large SD of the level perturbations, the variation caused by the level fluctuations around the mean segment level is of course weaker than the effect of the mean level. To examine the "real" improvement of considering temporal weights in the models for the AME task (by setting aside this large variation in mean level), additional weight analyses were conducted separately per mean level in the AME task.

G.2.2.5 Additional analyses per mean level in the AME task

We conducted separate regressions for each mean level with the results obtained in the AME task. The mean temporal weighting patterns thus obtained for each level are presented in Figure G.3.

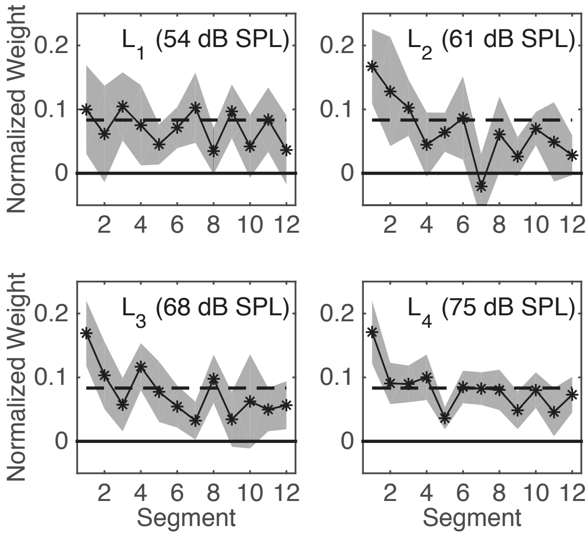

Figure G.3: Normalized averaged temporal weights for the AME task of Exp. 1, from the additional analyses conducted at each mean level separately.

A statistical analysis of these patterns indicates a significant influence of segment [$F(11, 66) = 4.900, p < .001, \eta_p^2 = .450, \tilde{\epsilon} = .97$]. The level ×

segment interaction was not significant $[F(33, 198) = 1.522, p = .071, \eta_p^2 = .203, \tilde{\epsilon} = .71]$. These results support what was observed in the overall analysis. The segments did not receive a uniform weighting, but rather a primacy effect was observed. Fig. G.3 shows that this weighting process depended descriptively (although not significantly) on the mean level: the weighting patterns estimated at 61, 68 and 75 dB SPL exhibit primacy effects, while a flat pattern was observed at 54 dB SPL, i.e. the lowest mean level employed in the AME task. We then compared the goodness of fit of the full and restricted models based on these weights obtained for each mean level. The proportions of variance accounted for by the models are reported at each mean level in Table G.3. Of course, the mean R^2 values are this time much lower than in the overall analysis presented above, but the data show large increases in R^2 when adding the temporal weights to the decision models (107%, 196%, 57% and 31% for each the mean levels in ascending order).

Level	Model	R^2				full vs. rest
						(nb of cases out of 7 where $p < .05$)
		Mean	*SD*	min	max	
L1 (54 dB SPL)	restricted	**.068**	.041	.013	.119	
	full	**.139**	.071	.067	.226	3
L2 (61 dB SPL)	restricted	**.043**	.030 .	011	.095	
	full	**.127**	.055	.056	.211	3
L3 (68 dB SPL)	restricted	**.138**	.052	.082	.207	
	full	**.217**	.074	.120	.331	5
L4 (75 dB SPL)	restricted	**.217**	.103	.065	.363	
	full	**.285**	.126	.117	.472	3

Table G.3: Comparison between full and restricted models at each mean level in the AME task of Exp. 1. The full models contained temporal weights inferred from the separate analyses of the estimates obtained at each mean level (54, 61, 68 and 75 dB SPL), in addition to the L_{eq}. The last column indicates the number of subjects for which the full model explained a significantly higher proportion of the variance (R^2) than the restricted model.

G.2.3 Discussion

In Experiment 1, the temporal weights of loudness were measured for the same participants in two psychophysical tasks, a level-discrimination task (LD) and an Absolute Magnitude Estimation task (AME). The loudness exponents estimated for each observer in the AME task where employed in the decision models to estimate the temporal loudness weights. Especially for the AME task, where the mean sound pressure level varied across a range of 21 dB, this procedure based on individual loudness values is in our view superior to the analyses based on sound pressure level used in most previous studies.

The data from Experiment 1 show that listeners assigned significantly non-uniform temporal weighting patterns in both tasks. This is an interesting finding because the sound duration (3 s) was considerably longer than in previous studies. Contrary to our hypothesis, there was a significant difference between the weighting patterns in the LD and the AME task. On average, participants assigned higher weights to the first three segments compared to later segments in the AME task (i.e. a primacy effect). In the LD task, we observed both a primacy effect and a recency effect. How could these different temporal weighting strategies be explained? One potential explanation is based on the difference between the stimuli presented in the two tasks. As explained above, we deliberately presented a much larger variation in mean level in the AME task than in the LD task, to ensure that subjects could easily do the task. The means of the four level distributions were separated by 7 dB, which is more than the SD of 5 dB, and covered a range of 21 dB. In contrast, in the LD task the means of the two level distributions differed by only 3 dB. For this reason, one could argue that the participants were influenced by the range of level variations of the stimuli. To test the hypothesis that the different weighting patterns can be attributed to the different range of segment levels presented in the two tasks, we conducted a second experiment presenting exactly the same stimuli in the AME as in the LD task.

G.3 Experiment 2

G.3.1 Materials and method

G.3.1.1 Participants

This experiment was conducted on a new group of seven subjects (4 women, 3 men; age 20-31 years), who participated voluntarily and were naïve with respect to the hypotheses under test. All participants reported normal hearing. They gave their informed written consent according to the Declaration of Helsinki prior to the experiment and were paid for their participation.

G.3.1.2 Stimuli and apparatus

All stimuli presented in this experiment were constructed exactly as those of the LD task of Exp. 1 (see above), both for the AME and the LD task. Thus, the segment levels of the stimuli were always drawn from normal distributions with means equal to 63.5 dB SPL ("low" distribution) or 66.5 dB SPL ("high" distribution). The apparatus were the same as in Exp. 1.

G.3.1.3 Procedure

The experiment comprised six 1-h sessions scheduled on different days. One session consisted of four 90-trials blocks, where two blocks were assigned to the AME task and the two remaining blocks to the LD task. The blocks were presented such that participants alternated between the two tasks (e.g., block1=LD, block2=AME, block3=LD and block4=AME). The type of the first block was also alternated between sessions. On each trial, a sound was presented with the segment levels drawn either from the "low" or the "high" distribution (see above), randomly chosen with a priori equal probability. The procedures were the same as in Exp. 1: in the LD blocks, the participant had to decide whether the sound was "soft" or "loud" while in the AME blocks, they had to give a number best representing the *global loudness* of each sound. Before the experiment, subjects were specifically informed that the variation in loudness between the stimuli would not be very large, so that in the AME task, they should select as many different numbers as possible to accurately capture the small variation in loudness from trial to trial. In order to have similar experimental conditions between the two tasks, the participants did receive feedback neither on a trial-by-trial nor on a block-by-block basis. The results of the first session, which served as a training session, were removed from the analysis. Thus, a total of 900 trials were collected per listener and per task. At the end of the last session, constant broadband noises were presented to the participants in an AME task, which consisted of 40 trials (10 repetitions of each level) to measure their loudness function with stationary stimuli. In this final part, participants were asked to use the same scale as they used during previous sessions.

G.3.1.4 Fitting loudness functions

The same fitting procedure as in Exp. 1 was used to fit their loudness functions, both for constant and fluctuating noises.

G.3.1.5 Decisions models

The decision models were strictly identical to those of Exp. 1.

G.3.2 Results

G.3.2.1 Loudness functions

Overall, the data show that the subjects had no problems to do the AME task and produced many different numbers to evaluate the global loudness of fluctuating noises with very similar mean levels. As an example, the magnitude estimates given by one participant (S8) in the AME task (for both fluctuating and constant noises) are plotted in Figure G.4. For every participant, the parameters of the loudness functions fitted to his estimates reported in Table G.4.

Loudness exponents α ranged between 0.30 and 0.80 for both static and level-fluctuating noises (see Table G.4). The slope was not significantly different between static and level-fluctuating noises [$t(6) = 0.909, p = .399$]. On average, these exponents are very close to those measured in the first experiment, with a mean value close to 0.5. As in Exp. 1, the values

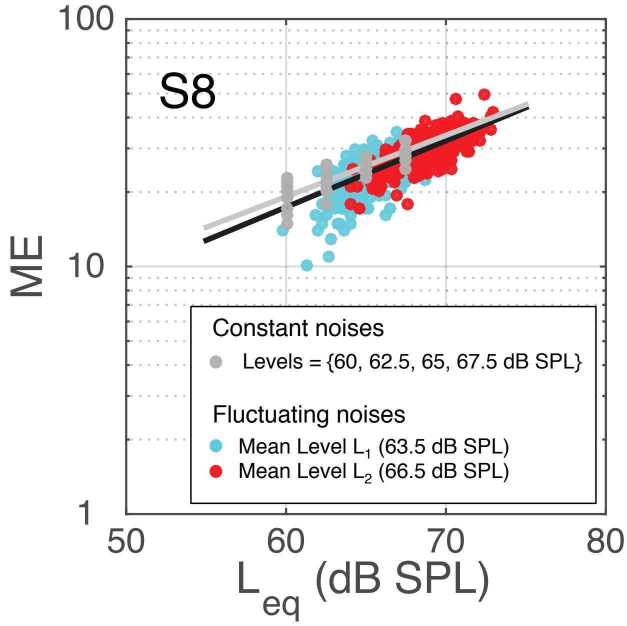

Figure G.4: Raw magnitude estimates produced by one subject (S8) in the AME task of Exp. 2. Same format as Fig. G.1.

Experiment 2

		S8	S9	S10	S11	S12	S13	S14	**Average**	*SD*
Constant noises	k	133.11	100.34	375.5	35.42	21.18	89.85	21.28	**110.95**	124.4
	α	0.50	0.30	0.60	0.57	0.45	0.63	0.43	**0.49**	0.11
	R^2	.74	.81	.66	.47	.73	.77	.18	**.62**	.22
Fluctuating noises	k	141.76	105.45	668.15	23.48	20.72	72.54	21.67	**150.54**	233.00
	α	0.54	0.35	0.80	0.48	0.47	0.58	0.48	**0.52**	0.14
	R^2	.65	.49	.43	.31	.42	.41	.15	**.40**	.16

Table G.4: Exp. 2, same format as Table G.1.

of the parameter k were very different between participants (who were allowed to use their own scale). The proportion of variance accounted for R^2 was significantly higher for static noises than for level-fluctuating noises $[t(6) = 4.523, p = .004]$.

G.3.2.2 Temporal weighting patterns

The averaged weighting patterns obtained for the two tasks, shown in Figure G.5, were similar to what was observed in the first experiment, with a clear primacy effect in the AME task while in the LD task, slightly higher weights were assigned to both the first and the last segments compared to the segments in the middle. A rmANOVA with the within-subjects factors segment and task showed no significant effect of segment [$F(11, 66) = 2.418, p = .11, \eta_p^2 = .287, \tilde{\epsilon} = .22$]. However, the significant segment $\times$ task interaction [$F(11, 66) = 4.480, p = .026, \eta_p^2 = .427, \tilde{\epsilon} = .22$] indicated that a different weighting was applied in the two tasks.

On average, the temporal weighting patterns were rather flat: as in Exp. 1, small CVs were measured both in the AME task (M = 0.40, $SD = 0.17$) and in the LD task (M = 0.65, $SD = 0.53$). There was no significant difference between the two tasks [$t(6) = 1.144, p = .296$]. One subject (S9) adopted a very specific strategy in the LD task by exclusively considering the three last segments of the stimuli to judge global loudness (for S9 in the LD task, CV= 1.792). This weighting pattern differed strongly from the weights assigned by the remaining participants, as indicated by the large confidence interval obtained for the weight on the last segment in the LD task (see Fig. G.5). After the end of the experiment, this participant told the experimenter that he had consciously and deliberately used this particular strategy in the LD task, although he was aware that the task was to judge the global loudness of the sounds similarly in the two tasks. However, he had no clear explanation why he adopted this strategy. The averaged weighting patterns when this participant was excluded are presented in Figure G.6, showing weighting patterns even more similar to the weights obtained in Experiment 1. In order to check whether the difference between the weighting patterns in the two tasks mainly relied on listener S9, we performed a second rmANOVA without this participant. The analysis still provided a significant segment $\times$ task interaction [$F(11, 55) = 5.269, p = .001, \eta_p^2 = .43, \tilde{\epsilon} = .45$]. Also, the effect of segment was now significant [$F(11, 55) = 3.770, p = .012, \eta_p^2 = .513, \tilde{\epsilon} = .58$].

Additional rmANOVAs conducted on each task separately without the data of S9 confirmed the presence of non-uniform temporal weights, with

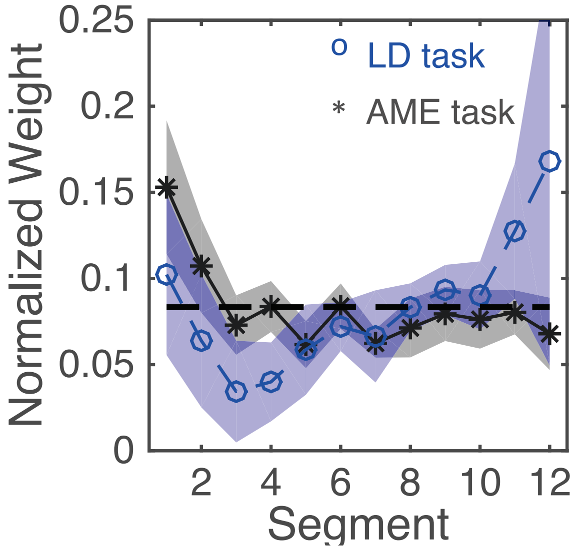

Figure G.5: Mean normalized temporal weights for both tasks of Exp. 2. Same format as Fig. G.2.

a significant effect of segment in the AME task $[F(11, 55) = 7.80, p < .001, \eta_p^2 = .61, \tilde{\epsilon} = .40]$, while in the LD task the effect just failed to reach significance $[F(11, 55) = 2.530, p = .051, \eta_p^2 = .34, \tilde{\epsilon} = .485]$.

G.3.2.3 Predictive power of the decision models

Fair to good model predictions were found for the LD task, except for Subject S14. In the LD task, AUC ranged from .69 (S14) to .92 (M= .85, SD= .08). In the AME task, R2 was quite low on average (M = .40, SD = .15); it ranged from .14 (S14) to .64. These values are comparable to what we observed for Exp. 1 in the separate analyses of the AME data per mean level.

G.3.2.4 Increased predictive power by including temporal weights

Full and restricted models were compared as in the first experiment to evaluate the benefit of adding temporal weights to a loudness model (for details about this model, see Exp. 1). The results are reported in Table G.5.

In the LD task, the AUC of the full model was not significantly higher

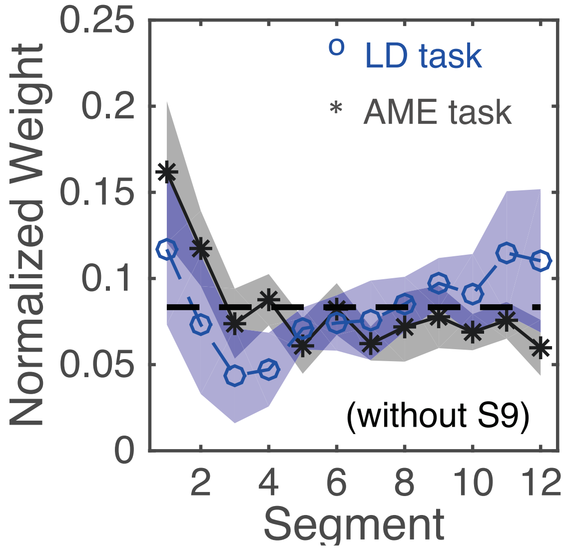

Figure G.6: Mean normalized temporal weights for both tasks of Exp. 2 when subject S9 is excluded, who differed from the remaining participants in exclusively considering the three last segments of the stimuli in the LD task. Same format as Fig. G.2.

Experiment 2

Goodness-of-fit index (Task)	**Model**	S8	S9	S10	S11	S12	S13	S14	**Average**	*SD*
AUC (LD Task)	restricted	.935	.749	.897	.876	.854	.870	.687	**.838**	.089
	full	.939	.885	.911	.891	.862	.911	.697	**.871**	.081
	p-value	.279	$< .001$	.001	.001	.023	$< .001$	.627		
R^2 (AME Task)	restricted	.645	.486	.423	.308	.415	.404	.141	**.403**	.155
	full	.655	.503	.474	.333	.440	.417	.172	**.428**	.150
	p-value	.018	.005	.001	.001	.001	.077	.002		

Table G.5: Exp. 2, same format as Table G.2.

than the AUC of the restricted model, $[t(6) = 1.839, p = .115]$. Cohen's d_z indicates a moderate effect size ($d_z = 0.695$), lower than what was measured in the first experiment ($d_z = 0.985$). However, likelihood-ratio tests indicated a significantly better goodness-of-fit ($p < .05$) with the full model compared to the restricted model for all subjects but S8 and S14. In the AME task, the R^2 of the full model ranged from .17 to .66 (M = .43, $SD = .15$) and was significantly higher than the R^2 of the restricted model [$t(6) = 4.673, p = .003$]. The effect size was large, $d_z = 1.7662$, similar to what was observed in the first experiment. A significantly better goodness-of-fit was obtained with the full model for all subjects but S13. Overall, this model comparison indicates a large and significant gain of adding temporal weights to predict loudness in the AME task but a smaller gain for the LD task.

G.3.3 Discussion

On average, the shape of the temporal weighting pattern in the AME task was strikingly similar to the pattern observed in the first experiment, with a significant primacy effect but no recency effect. In the LD task, there was a small primacy effect, but also a small recency effect, even if S9 was excluded from the analysis. The temporal weighting patterns differed significantly between the two tasks, similar to what was found for the first experiment. Thus, the specific "damped" trend of the temporal weighting pattern for the AME task (i.e., with primacy) was also observed in Exp. 2, where the range of levels was considerably smaller. As for Exp. 1, our results show that considering temporal weights in a model can yield significant improvements in the prediction of the loudness of level-fluctuating sounds.

G.4 General discussion and conclusion

The present paper presents the results of two experiments employing psychophysical reverse-correlation to compare the temporal weighting of loudness between a traditional absolute magnitude estimation task (AME) and a level-discrimination task (LD).

In the LD task, we observed "u-shaped" weighting patterns in both experiments, with marginally higher weights assigned to the first and final temporal segments of the sounds than to the temporal segments situated in the middle. Thus, the weights showed both small primacy and recency effects. This result is slightly different to what has been reported in previous studies for level-discrimination tasks, where much stronger primacy effects were usually observed (sometimes, weaker recency effects as well) (e.g. Pedersen and Ellermeier 2008; Oberfeld and Plank 2011; Oberfeld et al. 2012). In part, these differences are likely due to stimulus dissimilarities. Probably most important, in the present study, the stimuli were considerably longer (3 seconds) than in previous experiments on temporal weights, which mostly presented sounds of 1 second or less in duration. Thus, our data show that listeners judging the global loudness of longer sounds also assign non-uniform temporal weights. Second, the temporal weights inferred from sample discrimination tasks could in fact reflect effects related to memory processes (Dittrich and Oberfeld 2009; Oberfeld and Plank 2011). There is some evidence for weaker serial position effects in memory at slow presentation rates (Wickelgren 1970; Dosher 1999). Such an effect might have contributed to the weaker primacy effects observed in the present study, because the stimuli were varying in level at a slower rate (4 Hz) than in previous studies (≈10 Hz). Another stimulus difference that we should mention here is the variance of the random distributions from which the segment levels were drawn, which was substantially larger than in previous studies ($SD = 5$ dB vs. 2 dB). It is not unlikely that the higher “modulation depth” resulted in a percept that was qualitatively different from the “flat-profile” stimuli with comparably small level variations presented in earlier studies. It is not obvious, however, why a different percept should result in different temporal weights.

Furthermore, in the AME tasks of both experiments, we found evidence for significant primacy effects, i.e. the first few temporal segments of the sounds received greater weights compared to the following segments. However, there was no recency effect. The second experiment allowed us to rule out that the much larger level range presented in the first experiment in the AME task, compared to the LD task, was at the origins of the temporal weighting dissimilarities between the two tasks. Taken together, the data of both experiments show that the temporal weighting patterns obtained in the AME task were statistically different from those obtained in the LD task, which clearly contradicts the initial hypothesis proposed in the introduction.

Thus, the present study suggests a difference in the temporal weighting processes that depends on whether listeners are asked to do a binary loudness classification (LD task) or to evaluate the global loudness of time-varying sounds by assigning a number (AME task). Although the causes of the dissimilar weighting strategies presently observed are not obvious, one direction could be proposed to explain these results. This explanation relies on potential attentional differences and differences in the allocation of processing resources in the two tasks. Indeed, the AME task requires a more complex evaluation process as compared to the LD task: selecting a number from a potentially infinite set of numbers, or selecting one of only two possible responses. Thus, subjects probably need to invest higher effort in the AME task, which might lead to a more “economic” type of listening as compared to the other task. In addition, the fact that subjects repeated the tasks over hundreds of trials might have emphasized the need to adopt a strategy minimizing both the cost of the evaluation process and their response time in order to handle the task more easily. One way to achieve this in the AME task would be to start the process of selecting a number corresponding to the loudness of the sound soon after the sound has started rather than to wait until the end of the sound. The assignment of attention to the temporal dynamics of the sound would then be different between the two tasks, resulting in different temporal weighting patterns. However, we have no direct experimental evidence to support this argument.

As regard to loudness coding and evaluation more generally, the present results confirm the view that time-varying stimuli are not weighted uniformly and that the first portions play a greater role in global loudness judgments (primacy effect), as it was found in previous studies. Our data also indicate that for sounds considerably longer than 1 s, recency emerges. It would be interesting to explore even longer sounds, and one could speculate that with for example a 10-s sound the primacy effect would be further reduced, while the recency effect would increase (e.g. Susini et al. 2002). The amount of interindividual variation in the temporal weights observed in our experiments was somewhat higher than in most previous studies measuring temporal weights for shorter sounds with durations up to 1s, where typically very consistent primacy effects were found. The benefit of using individual temporal weights to predict loudness judgments was found to differ between tasks and listeners, but in the majority of cases it yielded significant improvements, which emphasizes the importance of considering non-uniform temporal weighting processes that underpin global loudness evaluation.

The main finding of the present study was the dissimilarity in temporal loudness weighting strategies found between global loudness judged in a sample discrimination task and in a magnitude estimation task. This effect was found even when the range of level variation did not differ between the two tasks. As discussed above, we however argue that this result does not invalidate temporal weighting processes typically inferred from sample discrimination tasks, but rather suggests that magnitude estimations repeated over hundreds of trials might lead people to adopt specific listening strategies in order to handle the required task in an economic fashion. Further work on this issue could help to gain better insight into the mechanisms underlying global loudness evaluation. To understand why and how listeners deploy specific temporal weighting strategies depending on the psychophysical task is certainly a promising direction for future research on loudness. In particular, it would be interesting to examine the proposed hypothesis that the dissimilar temporal weighting between different psychophysical tasks can be attributed to the amount of attention / processing resources distributed between the process of listening and forming a "sensory" representation of its global loudness, and the process of selecting a response (e.g., selecting a number in the AME task). Further studies are required before a precise implementation of a typical "universal" non-uniform temporal weighting function in future loudness models. In particular, it remains to be determined which weighting strategy closely reflects actual decisional processes used by listeners in more realistic (non-laboratory) loudness evaluation situations. The precise shape of the weighting function, its dependence on stimulus parameters (e.g. number of segments, variance of each segment) and its robustness to different stimulus configurations and psychophysical tasks still have to be addressed in more detail.

Acknowledgments

This work was supported by the project LoudNat funded by the French National Research Agency (ANR-11-BS09-016-01).

Bibliography

ISO 532B. Method for calculating loudness level. 1975.

Albert Jil Ahumada. *Detection of tones masked by noise;, a comparison of human observers with digital-computer-simulated energy detectors of varying bandwidths.* PhD thesis, University of California, Los Angeles, 1967.

Albert Jil Ahumada. Classification image weights and internal noise level estimation. *Journal of Vision*, 2(1):8, 2002.

Albert Jil Ahumada and John Lovell. Stimulus features in signal detection. *The Journal of the Acoustical Society of America*, 49(6B):1751–1756, 1971.

George D Allen. Acoustic level and vocal effort as cues for the loudness of speech. *The Journal of the Acoustical Society of America*, 49(6B):1831–1841, 1971.

ANSI. Psychoacoustical terminology. *New York: American National Standards Institute*, 1973.

Anne-Kristin Anweiler and Jesko L Verhey. Spectral loudness summation for short and long signals as a function of level. *The Journal of the Acoustical Society of America*, 119(5):2919–2928, 2006.

Yoav Arieh and Lawrence E Marks. Measurement of loudness, part ii: Context effects. In *Loudness*, pages 57–87. Springer, 2011.

Dan Ariely. Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making*, 11(1):19–45, 1998.

Dan Ariely and Ziv Carmon. Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making*, 13: 191–201, 2000.

Dan Ariely and Gal Zauberman. On the making of an experience: the effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making*, 13:219–232, 2000.

Dominik R Bach, Hartmut Schächinger, John G Neuhoff, Fabrizio Esposito, Francesco Di Salle, Christoph Lehmann, Marcus Herdener, Klaus Scheffler, and Erich Seifritz. Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex*, 18(1):145–150, 2008.

Dominik R Bach, John G Neuhoff, Walter Perrig, and Erich Seifritz. Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*, 74(1):28–33, 2009.

Dominik R Bach, Nicholas Furl, Gareth Barnes, and Raymond J Dolan. Sustained magnetic responses in temporal cortex reflect instantaneous significance of approaching and receding sounds. *PloS one*, 10(7):e0134060, 2015.

John C Baird. *Sensation and judgment: Complementarity theory of psychophysics*. Psychology Press, 2014.

Roger Bakeman. Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–384, 2005.

Bettina L Beard and Albert Jil Ahumada. Technique to extract relevant image features for visual tasks. In *Human Vision and Electronic Imaging III*, volume 3299, pages 79–85, 1998.

Bruce G Berg. Analysis of weights in multiple observation tasks. *The Journal of the Acoustical Society of America*, 86(5):1743–1746, 1989.

Bruce G Berg and DE Robinson. Multiple observations and internal noise. *The Journal of the Acoustical Society of America*, 81(S1):S33–S33, 1987.

JE Berliner, NI Durlach, and LD Braida. Intensity perception. vii. further data on roving-level discrimination and the resolution and bias edge effects. *The Journal of the Acoustical Society of America*, 61(6):1577–1585, 1977.

Marie-Claire Botte, Georges Canévet, and Bertram Scharf. Loudness adaptation induced by an intermittent tone. *The Journal of the Acoustical Society of America*, 72(3):727–739, 1982.

Marie-Claire Botte, C Baruch, and S Monikheim. Memory for loudness: the role of loudness contour. In *Auditory Physiology and Perception: Proceedings of the 9th International Symposium on Hearing Held in Carcens, France, on 9-14 June, 1991*, page 305. Pergamon, 1992.

Isabelle Boullet. *La sonie des sons impulsionnels: perception, mesures et modèles.* PhD thesis, Université de la Méditerranée-Aix-Marseille II, 2005.

Louis D Braida and Nathaniel I Durlach. Intensity perception. ii. resolution in one-interval paradigms. *The Journal of the Acoustical Society of America*, 51 (2B):483–502, 1972.

Louis D Braida, JS Lim, JE Berliner, Nathaniel I Durlach, WM Rabinowitz, and SR Purks. Intensity perception. xiii. perceptual anchor model of context-coding. *The Journal of the Acoustical Society of America*, 76(3):722–731, 1984.

Albert S Bregman. *Auditory scene analysis: The perceptual organization of sound.* MIT press, 1994.

ITU-R (Draft BS). Algorithms to measure audio programme loudness and true-peak level geneva. *Switzerland, April*, 2006.

Søren Buus, Hannes Müsch, and Mary Florentine. On loudness at threshold. *Acoustical Society of America Journal*, 104:399–410, 1998.

Georges Canévet and Bertram Scharf. The loudness of sounds that increase and decrease continuously in level. *The Journal of the Acoustical Society of America*, 88(5):2136–2142, 1990.

Georges Canévet, Bertram Scharf, and Marie-Claire Botte. Simple and induced loudness adaptation. *International Journal of Audiology*, 24(6):430–436, 1985.

Georges Canévet, Rhona P Hellman, and Bertram Scharf. Group estimation of loudness in sound fields. *Acta Acustica united with Acustica*, 60(4):277–282, 1986.

Georges Canévet, Bertram Scharf, Robert S Schlauch, Martha Teghtsoonian, and Robert Teghtsoonian. Perception of changes in loudness. *Nature*, 398(6729): 673–673, 1999.

Georges Canévet, Robert Teghtsoonian, and Martha Teghtsoonian. A comparison of loudness change in signals that continuously rise or fall in amplitude. *Acta Acustica united with Acustica*, 89(2):339–345, 2003.

Céline Cappe, Gregor Thut, Vincenzo Romei, and Micah M Murray. Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, 47 (4):1045–1052, 2009.

Céline Cappe, Antonia Thelen, Vincenzo Romei, Gregor Thut, and Micah M Murray. Looming signals reveal synergistic principles of multisensory integration. *The Journal of Neuroscience*, 32(4):1171–1182, 2012.

Marisa Carrasco, Sam Ling, and Sarah Read. Attention alters appearance. *Nature neuroscience*, 7(3):308–313, 2004.

Josef Chalupper and Hugo Fastl. Dynamic loudness model (dlm) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica*, 88(3):378–386, 2002.

Taishih Chi, Powen Ru, and Shihab A Shamma. Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, 118(2):887–906, 2005.

Janine C Clarey, Pascal Barone, and Thomas J Imig. Physiology of thalamus and cortex. In *The mammalian auditory pathway: Neurophysiology*, pages 245–251. Springer, 1992.

Sylvain Clément, Laurent Demany, and Catherine Semal. Memory for pitch versus memory for loudness. *The Journal of the Acoustical Society of America*, 106 (5):2805–2811, 1999.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Inc, 1988.

WJ Conover. *Practical nonparametric statistics*. Wiley New York, 1980.

David V Cross. Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, 14(3):547–552, 1973.

Huanping Dai and Christophe Micheyl. Psychophysical reverse correlation with multiple response alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4):976, 2010.

Christophe d'Alessandro and Michèle Castellengo. The pitch of short-duration vibrato tones. *The Journal of the Acoustical Society of America*, 95(3):1617–1630, 1994.

Bert De Coensel, Dick Botteldooren, Tom De Muer, Birgitta Berglund, Mats E Nilsson, and Peter Lercher. A model for the perception of environmental sound based on notice-events. *The Journal of the Acoustical Society of America*, 126 (2):656–665, 2009.

Thomas Deneux, Louise Francois, Suncana Sikiric, Emmanuel Ponsot, and Brice Bathellier. Cortical population nonlinearities reflect asymmetric auditory perception in mice. In *Cosyne 2015 Proceedings*, number 169, 2015.

Thomas Deneux, Alexandre Kempf, Aurélie Daret, Emmanuel Ponsot, and Brice Bathellier. Cortical bases of temporal asymmetries in auditory perception. submitted.

Jeffrey J DiGiovanni and Robert S Schlauch. Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecological Psychology*, 19(3):239–264, 2007.

Kerstin Dittrich and Daniel Oberfeld. A comparison of the temporal weighting of annoyance and loudness. *The Journal of the Acoustical Society of America*, 126 (6):3168–3178, 2009.

Barbara Anne Dosher. Item interference and time delays in working memory: Immediate serial recall. *International Journal of Psychology*, 34(5-6):276–284, 1999.

Nathaniel I Durlach and Louis D Braida. Intensity perception. i. preliminary theory of intensity resolution. *The Journal of the Acoustical Society of America*, 46(2B):372–383, 1969.

R EBU-Recommendation. Loudness normalisation and permitted maximum level of audio signals. 2011.

Jos J Eggermont. Wiener and volterra analyses applied to the auditory system. *Hearing research*, 66(2):177–201, 1993.

Wolfgang Ellermeier and S Schrödl. Temporal weights in loudness summation. In *Fechner Day 2000 Proceedings of the 16th Annual Meeting of the International Society for Psychophysics. Strasbourg: Université Louis Pasteur*, pages 169–173, 2000.

Michael Epstein. An introduction to induced loudness reduction. *The Journal of the Acoustical Society of America*, 122(3):EL74–EL80, 2007.

Michael Epstein. Context effects in loudness. In *Proceedings of Meetings on Acoustics*, volume 19, page 050004. Acoustical Society of America, 2013.

Michael Epstein and Mary Florentine. Loudness of brief tones measured by magnitude estimation and loudness matching. *Journal of the Acoustical Society of America*, 119(4):1943–1945, 2006.

Michael Epstein and Mary Florentine. Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear and hearing*, 30(2): 234–237, 2009.

Harvey Fletcher and Wilden A Munson. Loudness, its definition, measurement and calculation*. *Bell System Technical Journal*, 12(4):377–430, 1933.

Mary Florentine. *Loudness*. Springer, 2011.

Mary Florentine. A quarter-century's perspective on a psychoacoustical approach to loudness. In *Perspectives on Auditory Research*, pages 125–147. Springer, 2014.

Mary Florentine, Torben Poulsen, et al. Temporal integration of loudness as a function of level. *The Journal of the Acoustical Society of America*, 99(3):1633–1644, 1996.

Mark Georgeson. Sensory, perceptual and response biases: the criterion concept in perception. *Journal of Vision*, 12(9):1392–1392, 2012.

George A Gescheider and Barbara A Hughson. Stimulus context and absolute magnitude estimation: A study of individual differences. *Perception & psychophysics*, 50(1):45–57, 1991.

Asif A Ghazanfar, John G Neuhoff, and Nikos K Logothetis. Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences*, 99(24):15755–15757, 2002.

Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1-2):103, 1990.

Brian R Glasberg and Brian CJ Moore. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5):331–342, 2002.

Brian R Glasberg and Brian CJ Moore. The loudness of sounds whose spectra differ at the two ears. *The Journal of the Acoustical Society of America*, 127 (4):2433–2440, 2010.

Hedwig Gockel, Brian CJ Moore, and Robert P Carlyon. Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. *The Journal of the Acoustical Society of America*, 109(2):701–712, 2001.

Michael S Gordon, Frank A Russo, and Ewen MacDonald. Spectral information for detection of acoustic time to arrival. *Attention, Perception, & Psychophysics*, 75(4):738–750, 2013.

Gerd Gottschling. On the relations of instantaneous and overall loudness. *Acta Acustica united with Acustica*, 85(3):427–429, 1999.

Massimo Grassi and Christopher J Darwin. The subjective duration of ramped and damped sounds. *Perception & Psychophysics;*, 68(8):1382–1392, 2006.

Massimo Grassi and Andrea Pavan. The subjective duration of audiovisual looming and receding stimuli. *Attention, Perception, & Psychophysics*, 74(6):1321–1333, 2012.

David M Green. Consistency of auditory detection judgments. *Psychological Review*, 71(5):392, 1964.

DM Green and JA Swets. *Signal detection theory and psychophysics*. Wiley & Sons, Inc., New York, 1966.

Nicolas Grimault, Stephen Mcadams, and Jont B Allen. Auditory scene analysis: a prerequisite for loudness perception. In *Hearing–From Sensory Processing to Perception*, pages 295–302. Springer, 2007.

Giso Grimm, Volker Hohmann, and Jesko L Verhey. Loudness of fluctuating sounds. *Acta Acustica united with Acustica*, 88(3):359–368, 2002.

Deborah A Hall and David R Moore. Auditory neuroscience: The salience of looming sounds. *Current Biology*, 13(3):R91–R93, 2003.

Roelof Hamberg and Huib de Ridder. Time-varying image quality: Modeling the relation between instantaneous and overall quality. *SMPTE journal*, 108(11): 802–811, 1999.

Jürgen Hellbrück. Memory effects in loudness scaling of traffic noise. how overall loudness of short-term and long-term sounds depends on memory. *The Journal of the Acoustical society of Japan (E)*, 21(6):329–332, 2000.

Rhona P Hellman. Growth of loudness at 1000 and 3000 hz. *The Journal of the Acoustical Society of America*, 60(3):672–679, 1976.

Rhona P Hellman. Loudness, annoyance, and noisiness produced by single-tone-noise complexes. *The Journal of the Acoustical Society of America*, 72(1):62–73, 1982.

Rhona P Hellman and Carol H Meiselman. Prediction of individual loudness exponents from cross-modality matching. *Journal of Speech, Language, and Hearing Research*, 31(4):605–615, 1988.

Rhona P Hellman and J Zwislocki. Monaural loudness function at 1000 cps and interaural summation. *The Journal of the Acoustical Society of America*, 35(6): 856–865, 1963.

Rhona P Hellman, Andrzej Miśkiewicz, and Bertram Scharf. Loudness adaptation and excitation patterns: Effects of frequency and level. *The Journal of the Acoustical Society of America*, 101(4):2176–2185, 1997.

Åke Hellström. Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance*, 5(3):460, 1979.

Åke Hellström. Comparison is not just subtraction: Effects of time-and space-order on subjective stimulus difference. *Perception & Psychophysics*, 65(7): 1161–1177, 2003.

Cristy Ho, Charles Spence, and Rob Gray. Looming auditory and vibrotactile collision warnings for safe driving. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, pages 551–557, 2013.

J Hots, J Rennies, and JL Verhey. Loudness of subcritical sounds as a function of bandwidth, center frequency, and level. *The Journal of the Acoustical Society of America*, 135(3):1313–1320, 2014a.

Jan Hots, Jan Rennies, and Jesko L Verhey. Loudness of sounds with a subcritical bandwidth: A challenge to current loudness models? *The Journal of the Acoustical Society of America*, 134(4):EL334–EL339, 2013.

Jan Hots, Jan Rennies, and Jesko L Verhey. Modeling temporal integration of loudness. *Acta Acustica united with Acustica*, 100(1):184–187, 2014b.

Huynh Huynh and Leonard S Feldt. Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational and Behavioral Statistics*, 1(1):69–82, 1976.

Toshio Irino and Roy D Patterson. Temporal asymmetry in the auditory system. *The Journal of the Acoustical Society of America*, 99(4):2316–2331, 1996.

Walt Jesteadt. An adaptive procedure for subjective judgments. *Attention, Perception, & Psychophysics*, 28(1):85–88, 1980.

Walt Jesteadt and Lori J Leibold. Loudness in the laboratory, part i: Steady-state sounds. In *Loudness*, pages 109–144. Springer, 2011.

Walt Jesteadt, Daniel L Valente, Suyash N Joshi, and Kendra K Schmid. Perceptual weights for loudness judgments of six-tone complexes. *The Journal of the Acoustical Society of America*, 136(2):728–735, 2014.

Tomasz Kaczmarek and Anna Preis. Annoyance of time-varying road-traffic noise. *Archives of Acoustics*, 35(3):383–393, 2010.

Daniel Kahneman, Barbara L Fredrickson, Charles A Schreiber, and Donald A Redelmeier. When more pain is preferred to less: Adding a better end. *Psychological science*, 4(6):401–405, 1993.

Dennis H Klatt. Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, 67(3):971–995, 1980.

Reinier Kortekaas, Søren Buus, and Mary Florentine. Perceptual weights in auditory level discrimination. *The Journal of the Acoustical Society of America*, 113(6):3306–3322, 2003.

Sonoko Kuwano and Seichiro Namba. Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychological research*, 47(1):27–37, 1985.

Sonoko Kuwano and Seiichiro Namba. Loudness in the laboratory, part ii: Non-steady-state sounds. In *Loudness*, pages 145–168. Springer, 2011.

Sonoko Kuwano, Seiichiro Namba, Tohru Kato, and Jürgen Hellbrück. Memory of the loudness of sounds in relation to overall impression. *Acoustical Science and Technology*, 24(4):194–196, 2003.

Dave RM Langers, Pim van Dijk, Esther S Schoenmaker, and Walter H Backes. fmri activation in relation to sound intensity and loudness. *Neuroimage*, 35(2): 709–718, 2007.

Stefan Launer. Loudness perception in listeners with sensorineural hearing impairment. *Unpublished Ph. D. thesis, Oldenburg University, Oldenburg, Germany*, 1995.

Lori J Leibold, Hongyang Tan, Samar Khaddam, and Walt Jesteadt. Contributions of individual components to the overall loudness of a multitone complexa). *The Journal of the Acoustical Society of America*, 121(5):2822–2831, 2007.

Lori J Leibold, Hongyang Tan, and Walt Jesteadt. Spectral weights for sample discrimination as a function of overall levela). *The Journal of the Acoustical Society of America*, 125(1):339–346, 2009.

Jennifer J Lentz and Yi Shen. Investigating temporal asymmetry using masking period patterns and models of peripheral auditory processing. *The Journal of the Acoustical Society of America*, 129(5):3194–3205, 2011.

Fabrizio Leo, Vincenzo Romei, Elliot Freeman, Elisabetta Ladavas, and Jon Driver. Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Experimental brain research*, 213(2-3):193–201, 2011.

H Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2B):467–477, 1971.

Estella H Liu, Eduardo Mercado III, and Barbara A Church. Multidimensional processing of dynamic sounds: more than meets the ear. *Attention, Perception, & Psychophysics*, 73(8):2624–2638, 2011.

Gregory R Lockhead. Psychophysical scaling: Judgments of attributes or objects? *Behavioral and Brain Sciences*, 15(03):543–558, 1992.

Thomas Lu, Li Liang, and Xiaoqin Wang. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature neuroscience*, 4(11):1131–1138, 2001.

Robert A Lutfi and Walt Jesteadt. Molecular analysis of the effect of relative tone level on multitone pattern discrimination. *The Journal of the Acoustical Society of America*, 120(6):3853–3860, 2006.

Robert A Lutfi, Ching-Ju Liu, and Christophe Stoelinga. Level dominance in sound source identification. *The Journal of the Acoustical Society of America*, 124(6):3784–3792, 2008.

Neil A Macmillan and C Douglas Creelman. *Detection theory: A user's guide*. Psychology press, 2004.

Joost X Maier and Asif A Ghazanfar. Looming biases in monkey auditory cortex. *The Journal of Neuroscience*, 27(15):4093–4100, 2007.

Joost X Maier, John G Neuhoff, Nikos K Logothetis, and Asif A Ghazanfar. Multisensory integration of looming signals by rhesus monkeys. *Neuron*, 43(2): 177–181, 2004.

Joost X Maier, Chandramouli Chandrasekaran, and Asif A Ghazanfar. Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology*, 18(13):963–968, 2008.

Lawrence E Marks. Magnitude estimation and sensory matching. *Perception & Psychophysics*, 43(6):511–525, 1988.

Lawrence E Marks. The slippery context effect in psychophysics: Intensive, extensive, and qualitative continua. *Perception & Psychophysics*, 51(2):187–198, 1992.

Lawrence E Marks. Contextual processing of multidimensional and unidimensional auditory stimuli. *Journal of experimental psychology: Human perception and performance*, 19(2):227, 1993.

Lawrence E Marks and Mary Florentine. Measurement of loudness, part i: Methods, problems, and pitfalls. In *Loudness*, pages 17–56. Springer, 2011.

Jeremy Marozeau. Models of loudness. In *Loudness*, pages 261–284. Springer, 2011.

Peter McCullagh and John A Nelder. *Generalized linear models*, volume 37. CRC press, 1989.

Josh H McDermott. The cocktail party problem. *Current Biology*, 19(22):R1024–R1027, 2009.

Sabine Meunier, Patrick Susini, Régis Trapeau, and Jacques Chatron. Global loudness of ramped and damped sounds. In *20th International Congress on Acoustics*, 2010.

Sabine Meunier, Michaël Vannier, Jacques Chatron, and Patrick Susini. Asymmetry in perceived duration between up-ramp and down-ramp sounds as a function of duration. *The Journal of the Acoustical Society of America*, 136(2):EL166–EL172, 2014.

Brian CJ Moore. Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otology & neurotology*, 24 (2):243–254, 2003.

Brian CJ Moore. Revisiting the loudness of sounds with asymmetric attack and decay. *The Journal of the Acoustical Society of America*, 134(5):4195–4195, 2013.

Brian CJ Moore. Development and current status of the "cambridge" loudness models. *Trends in hearing*, 18, 2014.

Brian CJ Moore and Brian R Glasberg. Modeling binaural loudness. *The Journal of the Acoustical Society of America*, 121(3):1604–1612, 2007.

Brian CJ Moore, Brian R Glasberg, and Thomas Baer. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–240, 1997.

Richard F Murray. Classification images: A review. *Journal of Vision*, 11(5):2, 2011.

Yoshitaka Nakajima, Sonoko Kuwano, and Seiichiro Namba. The effect of temporal patterns of sound energy on the loudness of intensity increment sounds. *Psychological research*, 45(2):157–175, 1983.

Peter Neri. Estimation of nonlinear psychophysical kernels. *Journal of Vision*, 4 (2):2, 2004.

Peter Neri. Nonlinear characterization of a simple process in human vision. *Journal of Vision*, 9(12):1, 2009.

Peter Neri. Stochastic characterization of small-scale algorithms for human sensory processing. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4): 045118, 2010.

Peter Neri and David J Heeger. Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature neuroscience*, 5(8):812–816, 2002.

Veronika Neuert, Daniel Pressnitzer, Roy D Patterson, and Ian M Winter. The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids. *Hearing research*, 159(1):36–52, 2001.

John G Neuhoff. Perceptual bias for rising tones. *Nature*, 395(6698):123–124, 1998.

John G Neuhoff. An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, 13(2):87–110, 2001.

Bärbel Nieder, Søren Buus, Mary Florentine, and Bertram Scharf. Interactions between test-and inducer-tone durations in induced loudness reduction. *The Journal of the Acoustical Society of America*, 114(5):2846–2855, 2003.

Daniel Oberfeld. Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both? *The Journal of the Acoustical Society of America*, 121(4):2137–2148, 2007.

Daniel Oberfeld. Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale*, 62(1):24, 2008.

Daniel Oberfeld. The temporal weighting of the loudness of time-varying sounds reflects both sensory and cognitive processes. *Proceedings of Fechner Day*, 25 (1):261–266, 2009.

Daniel Oberfeld and Tina Plank. The temporal weighting of loudness: Effects of the level profile. *Attention, Perception, & Psychophysics*, 73(1):189–208, 2011.

Daniel Oberfeld, Wiebke Heeren, Jan Rennies, and Jesko Verhey. Spectro-temporal weighting of loudness. *PloS One*, 2012.

Eric C Odgaard, Yoav Arieh, and Lawrence E Marks. Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Perception & Psychophysics*, 65(1):123–132, 2003.

Eric C Odgaard, Yoav Arieh, and Lawrence E Marks. Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2):127–132, 2004.

Kirk N Olsen. Intensity dynamics and loudness change: A review of methods and perceptual processes. *Acoustics Australia*, 42(3):159, 2014.

Kirk N Olsen and Catherine J Stevens. Perceptual overestimation of rising intensity: Is stimulus continuity necessary? *Perception*, 39(5):695, 2010.

Kirk N Olsen, Catherine J Stevens, and Julien Tardieu. Loudness change in response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6):1631, 2010.

Kirk N Olsen, Catherine J Stevens, Roger T Dean, and Freya Bailes. Continuous loudness response to acoustic intensity dynamics in melodies: Effects of melodic contour, tempo, and tonality. *Acta psychologica*, 149:117–128, 2014.

Richard E Pastore and Jesse Flint. Magnitude judgments of loudness change for discrete, dynamic, and hybrid stimuli. *Attention, Perception, & Psychophysics*, 73(3):886–907, 2011.

Geoffrey R Patching, Mats P Englund, and Åke Hellström. Time-and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 38 (4):915, 2012.

Roy D Patterson. The sound of a sinusoid: Time-interval models. *The Journal of the Acoustical Society of America*, 96(3):1419–1428, 1994.

Roy D Patterson and Toshio Irino. Modeling temporal asymmetry in the auditory system. *The Journal of the Acoustical Society of America*, 104(5):2967–2979, 1998.

Roy D Patterson, Mike H Allerhand, and Christian Giguere. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, 98(4):1890–1894, 1995.

Benjamin Pedersen. *Auditory temporal resolution and integration-stages of analyzing time-varying sounds*. PhD thesis, Aalborg University, 2006.

Benjamin Pedersen and Wolfgang Ellermeier. Temporal weights in the level discrimination of time-varying sounds. *The Journal of the Acoustical Society of America*, 123(2):963–972, 2008.

Dennis P Phillips, Susan E Hall, and Susan E Boehnke. Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing research*, 167(1):192–205, 2002.

Emmanuel Ponsot. Sonie des sons non stationnaires : processus temporels perceptifs dans l'évaluation des sons crescendo et decrescendo. Rapport de stage de fin d'études, Ecole Centrale de Lyon, 2012.

Emmanuel Ponsot, Patrick Susini, Guillaume Saint Pierre, and Sabine Meunier. Temporal loudness weights for sounds with increasing and decreasing intensity profiles. *The Journal of the Acoustical Society of America*, 134(4):EL321–EL326, 2013.

Emmanuel Ponsot, Patrick Susini, and Sabine Meunier. A robust asymmetry in loudness between rising-and falling-intensity tones. *Attention, Perception & Psychophysics*, 77(3):907–920, 2015.

Emmanuel Ponsot, Sabine Meunier, Abbes Kacem, Jacques Chatron, and Patrick Susini. Are rising sounds always louder? influences of spectral structure and intensity-region on loudness sensitivity to intensity-change direction. *Acta Acustica United with Acustica*, in press.

Emmanuel Ponsot, Patrick Susini, and Oberfeld Daniel. Temporal weighting of loudness: Comparison between two different loudness-judgment tasks. -, submitted.

Torben Poulsen. Loudness of tone pulses in a free field. *The Journal of the Acoustical Society of America*, 69(6):1786–1790, 1981.

EC Poulton. Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, 91(3): 673, 1982.

EC Poulton and PR Freeman. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 66(1):1, 1966.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `http://www.R-project.org/`.

William M Rabinowitz, JS Lim, Louis D Braida, and Nathaniel I Durlach. Intensity preception. vi. summary of recent data on deviations from weber's law for 1000-hz tone pulses. *The Journal of the Acoustical Society of America*, 59(6): 1506–1509, 1976.

Anthony H Reinhardt-Rutland. Two auditory aftereffects and their dependency on carrier frequency. *Perception & psychophysics*, 28(6):569–571, 1980.

Anthony H Reinhardt-Rutland. Perceptual asymmetries associated with changing-loudness aftereffects. *Perception & psychophysics*, 66(6):963–969, 2004.

Anthony H Reinhardt-Rutland and Stuart Anstis. Auditory adaptation to gradual rise or fall in intensity of a tone. *Perception & psychophysics*, 31(1):63–67, 1982.

Jan Rennies and Jesko L Verhey. Temporal weighting in loudness of broadband and narrowband signals. *The Journal of the Acoustical Society of America*, 126 (3):951–954, 2009.

Jan Rennies, Jesko L Verhey, and Hugo Fastl. Comparison of loudness models for time-varying sounds. *Acta Acustica united with Acustica*, 96(2):383–396, 2010.

Jan Rennies, Inga Holube, and Jesko L Verhey. Loudness of speech and speech-like signals. *Acta Acustica United with Acustica*, 99(2):268–282, 2013a.

Jan Rennies, Jesko L Verhey, Jens E Appell, and Birger Kollmeier. Loudness of complex time-varying sounds? a challenge for current loudness models. In *Proceedings of Meetings on Acoustics*, volume 19, page 050189. Acoustical Society of America, 2013b.

Virginia M Richards, Yi Shen, and Charles Chubb. Level dominance for the detection of changes in level distribution in sound streams. *The Journal of the Acoustical Society of America*, 134(2):EL237–EL243, 2013.

Dennis T Ries, Robert S Schlauch, and Jeffrey J DiGiovanni. The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *The Journal of the Acoustical Society of America*, 124(6):3772–3783, 2008.

Dario Ringach and Robert Shapley. Reverse correlation in neurophysiology. *Cognitive Science*, 28(2):147–166, 2004.

Vincenzo Romei, Micah M Murray, Céline Cappe, and Gregor Thut. Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current biology*, 19(21):1799–1805, 2009.

Lawrence D Rosenblum, A Paige Wuestefeld, and Helena M Saldana. Auditory looming perception: Influences on anticipatory judgments. *Perception*, 22:1467–1467, 1993.

ANSI S-3.4. Procedure for the computation of loudness of steady sounds. *Acoustical Society of America*, 2007.

Roberta Santoro, Michelle Moerel, Federico De Martino, Rainer Goebel, Kamil Ugurbil, Essa Yacoub, and Elia Formisano. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS computational biology*, 10(1), 2014.

Bertram Scharf. Dichotic summation of loudness. *The Journal of the Acoustical Society of America*, 45(5):1193–1205, 1969.

Bertram Scharf. Loudness. *Handbook of perception*, 4:187–242, 1978.

Robert S Schlauch. A cognitive influence on the loudness of tones that change continuously in level. *The Journal of the Acoustical Society of America*, 92(2): 758–765, 1992.

Robert S Schlauch, Dennis T Ries, and Jeffrey J DiGiovanni. Duration discrimination and subjective duration for ramped and damped sounds. *The Journal of the Acoustical Society of America*, 109(6):2880–2887, 2001.

Josef Schlittenlacher, Takeo Hashimoto, Hugo Fastl, Seiichiro Namba, Sonoko Kuwano, and Shigeko Hatano. Loudness of pink noise and stationary technical sounds. In *Proceedings of Internoise*. Citeseer, 2011.

Josef Schlittenlacher, Wolfgang Ellermeier, and Takeo Hashimoto. Spectral loudness summation: Shortcomings of current standards. *The Journal of the Acoustical Society of America*, 137(1):EL26–EL31, 2015.

Keith A Schneider and Marcell Komlos. Attention biases decisions but does not alter appearance. *Journal of Vision*, 8(15):3, 2008.

Charles A Schreiber and Daniel Kahneman. Determinants of the remembered utility of aversive sounds. *Journal of experimental psychology. General*, 129(1): 27–42, 2000.

Christoph E Schreiner and Brian J Malone. Representation of loudness in the auditory cortex. *The Human Auditory System: Fundamental Organization and Clinical Disorders*, 129:73, 2015.

Michael Schutz and Jonathan M Vaisberg. Surveying the temporal structure of sounds used in music perception. *Music Perception: An Interdisciplinary Journal*, 31(3):288–296, 2014.

Erich Seifritz, John G Neuhoff, Deniz Bilecen, Klaus Scheffler, Henrietta Mustovic, Hartmut Schächinger, Raffaele Elefante, and Francesco Di Salle. Neural processing of auditory looming in the human brain. *Current Biology*, 12(24): 2147–2151, 2002.

Ville Pekka Sivonen and Wolfgang Ellermeier. Binaural loudness for artificial-head measurements in directional sound fields. *Journal of the Audio Engineering Society*, 56(6):452–461, 2008.

Ville Pekka Sivonen and Wolfgang Ellermeier. Binaural loudness. In *Loudness*, pages 169–197. Springer, 2011.

Robert Lee Smith and JJ Zwislocki. Short-term adaptation and incremental responses of single auditory-nerve fibers. *Biological cybernetics*, 17(3):169–182, 1975.

G Christopher Stecker and Ervin R Hafter. An effect of temporal asymmetry on loudness. *The Journal of the Acoustical Society of America*, 107(6):3358–3368, 2000.

Stanley S Stevens. The measurement of loudness. *The Journal of the Acoustical Society of America*, 27(5):815–829, 1955.

Stanley S Stevens. The direct estimation of sensory magnitudes: Loudness. *The American journal of psychology*, pages 1–25, 1956.

Patrick Susini, Stephen McAdams, and Bennett K Smith. Global and continuous loudness estimation of time-varying levels. *Acta Acustica united with Acustica*, 88(4):536–548, 2002.

Patrick Susini, Stephen McAdams, and Bennett K Smith. Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta acustica united with acustica*, 93(4):623–631, 2007.

Patrick Susini, Sabine Meunier, Régis Trapeau, and Jacques Chatron. End level bias on direct loudness ratings of increasing sounds. *The Journal of the Acoustical Society of America*, 128(4):EL163–EL168, 2010.

Clare AM Sutherland, Gregor Thut, and Vincenzo Romei. Hearing brighter: Changing in-depth visual perception through looming sounds. *Cognition*, 132 (3):312–323, 2014.

Yoiti Suzuki and Hisashi Takeshima. Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, 116(2):918–933, 2004.

Alejandro Tabas, Emili Balaguer-Ballester, Daniel Pressnitzer, Anita Siebert, and André Rupp. Hierarchical processing of temporal asymmetry in human auditory cortex. In *Pattern Recognition in Neuroimaging, 2014 International Workshop on*, pages 1–4. IEEE, 2014.

Ana Tajadura-Jiménez, Aleksander Väljamäe, Erkin Asutay, and Daniel Västfjäll. Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion*, 10(2):216, 2010.

Robert Teghtsoonian, Martha Teghtsoonian, and Georges Canévet. The perception of waning signals: Decruitment in loudness and perceived size. *Perception & psychophysics*, 62(3):637–646, 2000.

Robert Teghtsoonian, Martha Teghtsoonian, and Georges Canévet. Sweep-induced acceleration in loudness change and the "bias for rising intensities". *Perception & Psychophysics*, 67(4):699–712, 2005.

Frédéric E Theunissen and Julie E Elie. Neural processing of natural sounds. *Nature Reviews Neuroscience*, 15(6):355–366, 2014.

Régis Trapeau. Etude perceptive d'une évolution temporelle. Master ic2a – spécialité art, sciences, technologies, Université de Grenoble, 2009.

Matthew D Turner and Bruce G Berg. Temporal limits of level dominance in a sample discrimination task. *The Journal of the Acoustical Society of America*, 121(4):1848–1851, 2007.

Sascha Tyll, Björn Bonath, Mircea Ariel Schoenfeld, Hans-Jochen Heinze, Frank W Ohl, and Tömme Noesselt. Neural basis of multisensory looming signals. *Neuroimage*, 65:13–22, 2013.

Guillaume T Vallet, David I Shore, and Michael Schutz. Exploring the role of the amplitude envelope in duration estimation. *Perception*, 43(7):616, 2014.

Michaël Vannier. *Sonie de champs acoustiques stationnaires en situation d'écoute dichotique*. PhD thesis, Insa-Lyon, 2015.

Daniel Västfjäll. The "end effect" in retrospective sound quality evaluation. *Acoustical Science and Technology*, 25(2):170–172, 2004.

Jesko L Verhey and Birger Kollmeier. Spectral loudness summation as a function of duration. *The Journal of the Acoustical Society of America*, 111(3):1349–1358, 2002.

Jirı Wackermann. Universality versus individuality: Place for inter-individual differences? *Fechner Day 2014*, page 67, 2014.

J Wang, L Qin, S Chimoto, S Tazunoki, and Y Sato. Response characteristics of primary auditory cortex neurons underlying perceptual asymmetry of ramped and damped sounds. *Neuroscience*, 256:309–321, 2014.

David T Westwick and Robert E Kearney. *Identification of nonlinear physiological systems*, volume 7. John Wiley & Sons, 2003.

Wayne A Wickelgren. Time, interference, and rate of presentation in short-term recognition memory for items. *Journal of Mathematical Psychology*, 7(2):219–235, 1970.

Sonia Wilkie. *The Effect of Audio Cues and Sound Source Stimuli on Looming Perception*. PhD thesis, School of Electronic Engineering and Computer Science Queen Mary, University of London, 2015.

Pavel Zahorik and Frederic L Wightman. Loudness constancy with varying sound source distance. *Nature neuroscience*, 4(1):78–83, 2001.

Chaoying Zhang and Fan-Gang Zeng. Loudness of dynamic stimuli in acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 102(5):2925–2934, 1997.

Muhammad SA Zilany, Ian C Bruce, Paul C Nelson, and Laurel H Carney. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America*, 126(5):2390–2412, 2009.

Eberhard Zwicker. Über psychologische und methodische grundlagen der lautheit. *Acta Acustica united with Acustica*, 8(Supplement 1):237–258, 1958.

Eberhard Zwicker. Procedure for calculating loudness of temporally variable sounds. *The Journal of the Acoustical Society of America*, 62(3):675–682, 1977.

Eberhard Zwicker. Dependence of post-masking on masker duration and its relation to temporal effects in loudness. *The Journal of the Acoustical Society of America*, 75(1):219–223, 1984.

Eberhard Zwicker, Georg Flottorp, and Stanley Smith Stevens. Critical band width in loudness summation. *The Journal of the Acoustical Society of America*, 29(5):548–557, 1957.

Eberhard Zwicker, Hugo Fastl, and H Frater. *Psychoacoustics, Facts and Models,*, volume 22. Springer (Series of Information Sciences), Berlin, 2nd updated edition, 1999.