



HAL
open science

Loudness Processing of Time-Varying Sounds: Recent advances in psychophysics and challenges for future research

Emmanuel Ponsot, Patrick Susini, Sabine Meunier

► **To cite this version:**

Emmanuel Ponsot, Patrick Susini, Sabine Meunier. Loudness Processing of Time-Varying Sounds: Recent advances in psychophysics and challenges for future research. InterNoise, Aug 2016, Hamburg, Germany. <hal-01429698>

HAL Id: hal-01429698

<https://hal.science/hal-01429698v1>

Submitted on 9 Jan 2017

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

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



HAL Authorization



Loudness Processing of Time-Varying Sounds: Recent advances in psychophysics and challenges for future research

Emmanuel PONSOT¹; Patrick SUSINI¹; Sabine MEUNIER²

¹ STMS lab (Ircam, CNRS, UPMC), 1 place Igor Stravinsky, 75004 Paris, France

² LMA, CNRS, UPR 7051, Aix-Marseille Univ, Centrale Marseille, impasse Nikola Tesla, CS 400006, 13 453
Marseille Cedex 13, France

ABSTRACT

Current loudness models and indicators of time-varying sounds such as the 95th percentile of the loudness distribution N_5 all compute global loudness from short-term loudness time-series by leaving the temporal dimension out. Putting it differently, these models all rely on the assumption that the global loudness of a time-varying sound does not depend on the shape of its loudness pattern, i.e. that it can be predicted by its loudness distribution only. This assumption has been challenged and invalidated by a number of recent studies investigating time-varying sounds with both flat (e.g., fluctuating stimuli) and non-flat (e.g., rising- or falling-intensity stimuli) overall intensity profiles. We will intend to review the main outcomes of these studies and illustrate the implication of various high-level processes (cognitive, memory) in overall loudness evaluations of time-varying sounds, not yet considered in current loudness models. Then, we will discuss how the association of “molar” and “molecular” (i.e. reverse-correlation) psychophysics could help characterize and identify the whole underlying perceptual machinery.

Keywords: Loudness, Time-Varying sounds, Psychophysical methods - I-INCE Classification of Subjects Number(s): 63.1 (See. <http://www.inceusa.org/links/Subj%20Class%20-%20Formatted.pdf>.)

1. INTRODUCTION

For sound sequences of several seconds varying through time, the loudness of a stimulus cannot be considered as a *static* perceptual variable. As soon as listeners have access to the dynamics of their sensory representations, i.e. they are able to judge their momentary perception at different times, one needs to distinguish momentary sensations from what can be experienced and evaluated at the end of a sound, here defined as *global loudness*. Global loudness corresponds to the “overall loudness” perceived after the end of the sound, when the whole momentary loudness pattern (i.e., the vector made of “momentary” loudness snapshots over time) is merged into one scalar value by the human brain. This retrospective *subjective* merging is an operation of interest for the industry and the media who need to assess as accurately as possible how loud a sequence of several seconds or minutes will be perceived (e.g., the sound produced by the passing-by of an airplane, or an advertisement to be broadcasted on the radio).

Psychoacoustical experiments investigating dynamic loudness perception showed that global loudness does not correspond to a simple average of momentary loudness, but is rather strongly influenced by the loudest events (e.g. 1, 2, 3). Current indicators of global loudness are based on this finding. In the media, the overall loudness of a program is taken as the integration of its momentary loudness values (predicted by simplified auditory models) that exceed a certain threshold (e.g., see the R-128 EBU-Recommendation in 4). Other indicators employed in the industry, such as N_5 or L_{Aeq} also rely on the assumption that global loudness of a time-varying sound does not depend on the shape of its loudness pattern, i.e. that it can be predicted by its loudness distribution only. Even in

¹ ponsot@ircam.fr

the context of psychoacoustical research, Zwicker (5) or Glasberg and Moore (6) suggested to use the peak of the “short-term loudness” or the “long-term loudness” time-series predicted by their models to estimate the global loudness of time-varying sounds. However, recent studies conducted even with very basic sounds have pointed out limitations to such assumption, showing for example that even with 1-kHz stimuli ramping up or down in level, global loudness is not simply based on the maximum of their short-term or long-term loudness (e.g. 7).

Why current models fail to predict global loudness of time-varying stimuli? How could this limit be overcome? The present paper attempts to provide answers to these questions. In Section 2, we discuss why, in our view, current loudness models cannot yield to accurate predictions of loudness for long time-varying sounds, as they were not designed for that purpose. In Section 3, we intend to expose the main outcomes of recent psychophysical studies on loudness of time-varying stimuli that shed light onto mechanisms not yet identified and thus not considered in current models neither. These mechanisms were exposed by the use of two broad classes of psychophysical approaches, known as “molar” and “molecular” methods (8). We explain why a wise combination of these approaches can be valuable for anyone interested in characterizing a whole complex psychophysical processing such as the one considered here.

2. CURRENT LOUDNESS MODELS WERE NOT CONCEIVED FOR GLOBAL LOUDNESS PREDICTIONS

Historically, psychoacoustics has mostly been dedicated to the study of fine timescale sensory and perceptual processing (e.g., temporal integration, backward and forward masking, detection of intensity increments/decrements), generally by conducting experiments involving very brief and simple stimuli (e.g., a few hundred milliseconds). This was the necessary step to reach a good characterization of the basic components of auditory processing, such as outer and middle ear filtering, inner ear coding, or more central stages of processing where the information incoming from both ears are merged together (e.g., binaural loudness summation). This led to quite sophisticated auditory models such as loudness models that, for the most advanced versions, are said to be suitable for temporally varying sounds (e.g., 5, 6). We could schematize this approach by saying that the study of loudness has mainly been focused on what occurs in the Stage A (see Fig. 1) of the perceptual processing. The experiments were often designed so as to minimize the effects of the higher level processes potentially present in the later stages (e.g., memory processes, heuristic rules), notably through the design of experiments employing short and simple stimuli. In these cases, loudness could simply be considered as a unitary variable (i.e., the concepts of momentary loudness and global loudness were confounded).

However, in certain cases of psychophysical judgments, our perceptual system operates on a larger timescale that inescapably involves the processes present in Stage B. This is the case of *global loudness* processing of time-varying sounds lasting several seconds: our perceptual system needs to integrate and to store what occurred during the whole stimulus presentation, thus potentially involving “higher-level” processing stages such as memory and decisional mechanisms. We believe that this is the main reason why current loudness models cannot yield to accurate predictions of global loudness for long time-varying stimuli: they were not conceived to reveal and thus to implement the processes occurring in Stage B. In order to overcome this limitation, future auditory models that seek to predict global perceptual variables would have to incorporate higher-level processing stages identified through experiments specifically designed for that purpose. In the next section, we present several recent studies on loudness that attempted to identify and characterize these latter processes by using two classes of psychophysical approaches. The stimuli employed in these studies were 1-kHz tones or noises varying in intensity over several seconds; they were either fluctuating around a given flat intensity profile or with rising / falling intensity profiles, the latter constituting a first-order approximation of intensity dynamics encountered in natural (musical or environmental) sounds.

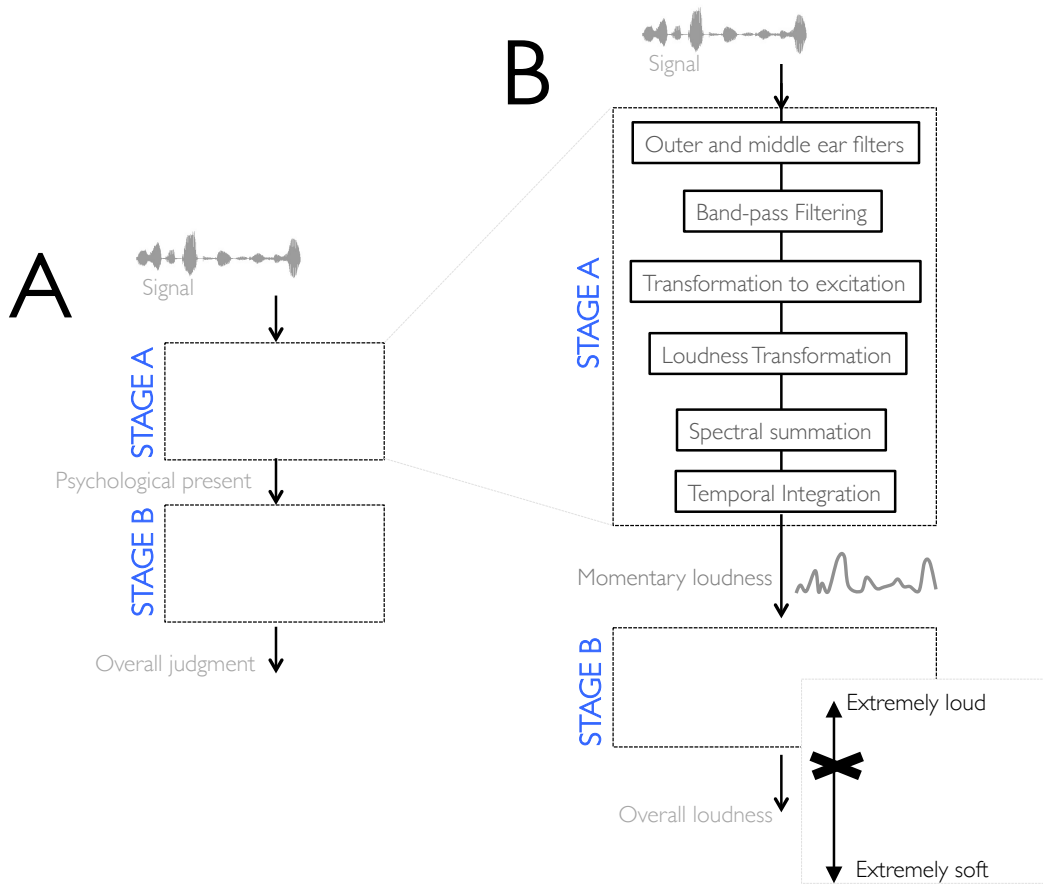


Figure 1 – (A) General schematic view of human perceptual processing operating at different timescales; early processes in Stage A provide continuous representation of the sensory input while latter processes of Stage B are solicited for integrating past information into an overall judgment. (B) This processing chain is rendered explicit in the case of global loudness processing. Since, the mechanisms in Stage B can potentially influence those involved in Stage A, this figure must be taken as a schematic illustration only.

3. RECENT EXPERIMENTAL STUDIES ON LOUDNESS OF TIME-VARYING SOUNDS: TWO CLASSES OF PSYCHOPHYSICAL APPROACHES

3.1 Generalities concerning molar and molecular approaches

The terms “molar” and “molecular” psychophysics were initially introduced by Green (8) to differentiate between two classes of psychophysical approaches. The main opposition could be formulated as follows (a tentative illustration is proposed in Figure 2).

In molar psychophysics, a limited set of stimuli is typically used at the input of the perceptual system. Each stimulus is presented many times until a significant number of responses are reached, so that performance metrics (e.g. means of responses) can be computed in each condition. In the end, only these metrics, based on an average of individual responses, are considered in the analyses. In molecular psychophysics, one main stimulus is used to create an infinity of very similar but always different stimuli, this by adding some noise on the physical dimension(s) of interest (in Fig. 2, this is illustrated for the intensity profile of a rising-intensity stimulus). Observers then evaluate these noisy stimuli in hundreds or thousands of trials. In the end, their responses are reverse-correlated with noise properties to identify the mechanisms that underpin the way they processed the stimuli. As compared to molar psychophysics, the analysis undertaken in molecular psychophysics (i.e. reverse-correlation) relies on the noise introduced in each stimulus and the observer’s response on each trial.

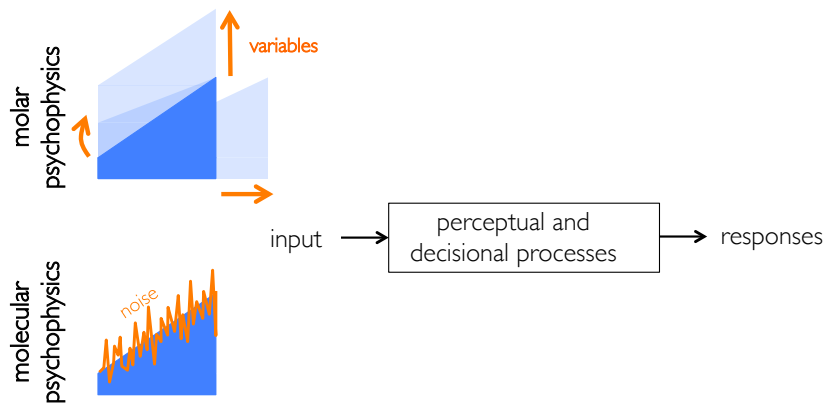


Figure 2 –The two main classes of psychophysical approaches illustrated in the case of the investigation of how humans process loudness of rising intensity-stimuli.

3.2 Main results from recent loudness studies using the two classes of approaches

3.2.1 Molar psychophysics for characterizing processes in a wide range

Molar psychophysics is probably the most direct tool for psychophysicians. Indeed, many questions can very often be addressed experimentally using a restricted number of conditions. A dependent variable is measured precisely in each condition (e.g., using adaptive tracking procedures) and an analysis of variance is conducted on the data to see whether the dependent variable measured in these conditions differ significantly or not. The main advantages of this approach are that very different conditions can be examined and that there is virtually no restriction in the design of such experiments. If one wants to assess the loudness difference between the loudness of a 1-second stimulus and the loudness of a 10-s stimulus (an order of magnitude apart), an appropriate experimental design can easily be proposed.

Molar psychophysics was thus adopted in a couple of recent psychophysical studies investigating global loudness processing of rising and falling-intensity tones with very different characteristics. Notably, specific temporal integration and memory processes were exposed by significant stretching of stimuli amplitude envelopes: durations of the ramp stimuli were varied from 1 to 16 seconds, the slope of the stimuli and their dynamics (range of intensity variation) were also manipulated in wide ranges (9). A molar approach was also used to investigate global loudness processing of rising and falling ramps at very different intensity-regions, apart from more than 20 dB (7, 10).

3.2.2 Molecular psychophysics for characterizing processes at a given operating point

Even though the molecular psychophysical approach exists for a long time and was widely used in the neuroscience and physiology community to characterize sensory systems, the method has only been recently (for a decade or so) transposed to the study of loudness processing. It appeared to be particularly relevant for investigating how humans integrate the different temporal portions of a time-varying stimulus, because the dimension of interest being the intensity profile, some noise can be introduced on it directly to examine temporal weighting strategies in the global loudness judgments produced by the listeners. Weighting strategies have thus been exposed with time-varying sounds using both “flat-level” profiles stimuli (e.g., fluctuating stimuli) and “non-flat” profiles stimuli (e.g., rising and falling-intensity stimuli).

When the sounds are varying around overall “flat” profiles, there are evidences that the underlying temporal weighting processes employed by listeners to judge global loudness are not uniform. The first events and, to a lesser extent, the last portions of the stimuli receive stronger weights compared to the others; “primacy” and “recency” effects have been proposed to account for these findings, respectively (e.g. 11). Most previous studies demonstrated these effects with broadband or narrowband noises lasting about 1 s; we recently showed that it was also true for noises of longer durations (with 3-s white noises fluctuating in level; see 12). When the overall level profile of a time-varying sound is not flat, other mechanisms seem to be engaged. Oberfeld and Plank (13) showed that when a gradual increase of intensity was imposed on the beginning of the stimuli, the primacy effect was transferred to the first next highest-level portions. Using stimuli with rising and

falling-intensity profiles, Ponsot et al. (14) found that “primacy” and “recency” effects disappeared and that the loudest portions of the stimuli then dominated global loudness judgments; a phenomenon referred to as “level-dominance” (15). Second-order analyses of these data were recently conducted, showing nonlinear components in the integration process of the loudest portion that had not been considered so far (see 16, Appendix F).

3.3 How combining molar and molecular methods could be helpful to characterize the whole system

On the one hand, molar psychophysics appears to be an indispensable tool for quantifying how the perceptual system responds to stimuli whose characteristics vary in a wide range. Temporal integration and memory mechanisms involved in the case of rising and falling-intensity sounds would have never been exposed without the use of stimuli of very different durations. On the other hand, level-dominance effects and the shape of the temporal integration window used by the listeners could directly be accessed to by the molecular psychophysical approach and nonlinear components of global loudness processing were identified through a molecular analysis too.

4. CONCLUSION AND PERSPECTIVES

In this paper, we present several recent psychophysical studies that focused on the identification of the mechanisms underlying global loudness processing of time-varying sounds lasting several seconds, which adopted either molar or molecular approaches. We show that specific aspects of this issue can more easily be tackled with one approach or the other, depending on the range and the precision of the characterization required. When one starts to investigate a complex psychophysical issue that operate at different timescales, we believe that it would worth identifying upstream the processes that can be exposed by molar or molecular approaches because the two do not make use of the same conditions and stimuli.

All the processes reported above invalidate the assumption that the global loudness of a time-varying sound does not depend on the shape of its momentary loudness pattern; thus, current models cannot account for these effects. Further attention is deserved to refine our understanding of the whole underlying perceptual machinery involved in global loudness processing. In particular, memory mechanisms involved in the integration process were highlighted but not quantified precisely yet. The respective contributions of “primacy” and “recency” effects exposed by molecular psychophysics seem to depend on the stimuli characteristics, such as their overall duration and both the size and duration of the introduced level perturbations. Such aspects have to be further explored, as for instance the fact that “primacy” effects appear to be downsized when the duration increases or the fact that the type of judgment used by listeners to evaluate global loudness alters the strength of these edge-emphasis weightings (12). Efforts should also be given to examine the role of nonlinear components in the whole integration process.

ACKNOWLEDGEMENTS

The studies conducted by the authors which are reported in this paper were realized in the framework of the *LoudNat* project funded by the French National Research Agency (ANR-11-BS09-016-01).

REFERENCES

1. Kuwano, S., & Namba, S. (1985). Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychological research*, 47(1), 27-37.
2. Susini, P., McAdams, S., & Smith, B. K. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acustica*, 88(4), 536-548.
3. Susini, P., McAdams, S., & Smith, B. K. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica United with Acustica*, 93(4), 623-631.
4. EBU-Recommendation, R. (2011). Loudness normalisation and permitted maximum level of audio signals.
5. Fastl, H., & Zwicker, E. (2006). *Psychoacoustics: facts and models* (Vol. 22). Springer Science & Business Media.
6. Glasberg, B. R., & Moore, B. C. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5), 331-342.
7. 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.
8. Green, D. M. (1964). Consistency of auditory detection judgments. *Psychological Review*, 71(5), 392.
9. Ponsot, E., Susini, P., & Meunier, S. (2014). The role of duration in global loudness evaluation of rising and falling-intensity sounds. *Fechner Day 2014*, 58.
10. Ponsot, E., Meunier, S., Kacem, A., Chatron, J., & Susini, P. (2015). Are rising sounds always louder? Influences of spectral structure and intensity-region on loudness sensitivity to intensity-change direction. *Acta Acustica united with Acustica*, 101(6), 1083-1093.
11. Pedersen, B., & Ellermeier, W. (2008). Temporal weights in the level discrimination of time-varying sounds. *The Journal of the Acoustical Society of America*, 123(2), 963-972.
12. Ponsot, E., Susini, P., & Oberfeld, D. (2016). Temporal weighting of loudness: Comparison between two different psychophysical tasks. *The Journal of the Acoustical Society of America*, 139(1), 406-417.
13. Oberfeld, D., & Plank, T. (2011). The temporal weighting of loudness: Effects of the level profile. *Attention, Perception, & Psychophysics*, 73(1), 189-208.
14. 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.
15. Lutfi, R. A., & Jesteadt, W. (2006). 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.
16. Ponsot, E. (2015). Global loudness processing of time-varying sounds (Doctoral dissertation, UPMC).