End level bias on direct loudness ratings of increasing sounds

Patrick Susini[a)]
Institut de Recherche et de Coordination Acoustique/Musique and Unité Mixte de Recherche 9912, Centre National de la Recherche Scientifique, 1 place Igor Stravinsky, 75004 Paris, France
patrick.susini@ircam.fr

Sabine Meunier, Régis Trapeau, and Jacques Chatron
Laboratoire de Mécanique et d'Acoustique, Unité Propre de Recherche 7051, Centre National de la Recherche Scientifique, 31 chemin Joseph-Aiguier, 13402 Marseille, France
meunier@lma.cnrs-mrs.fr, regis.trapeau@umontreal.ca, chatron@lma.cnrs-mrs.fr

Abstract: Three experiments on loudness of sounds with linearly increasing levels were performed: global loudness was measured using direct ratings, loudness change was measured using direct and indirect estimations. Results revealed differences between direct and indirect estimations of loudness change, indicating that the underlying perceptual phenomena are not the same. The effect of ramp size is small for the former and important for the latter. A similar trend was revealed between global loudness and direct estimations of loudness change according to the end level, suggesting they may have been confounded. Measures provided by direct estimations of loudness change are more participant-dependent.

PACS numbers: 43.66.Cb, 43.66.Lj [Q-JF]
Date Received: June 8, 2010 **Date Accepted:** August 5, 2010

1. Introduction

A significant number of studies have recently investigated the loudness of sounds that increase and decrease continuously in level (Neuhoff, 1998, 2001; Canévet *et al.*, 2003; Teghtsoonian *et al.*, 2005; Susini *et al.*, 2007; Olsen *et al.*, 2010; Olsen and Stevens, 2010). This question is important for the study of moving sources for example, as intensity change is the dominant motion cue for mediating the effects of approaching sound sources (Bach *et al.*, 2009). Results are different among the studies. On the one hand, an overestimation of loudness change for upramp sounds was revealed (Neuhoff, 1998, 2001; Teghtsoonian *et al.*, 2005), explained by Neuhoff (1998, 2001) as a survival advantage for detecting an approaching sound source. In these studies direct ratings were used, i.e., the loudness change reported was a single judgment. The overestimation for upramps has been found to be strongly dependent on end level of the ramp rather than on ramp size (Teghtsoonian *et al.*, 2005). On the other hand, no loudness change overestimation for upramp sounds was found (Canévet *et al.*, 2003; Susini *et al.*, 2007), but a strong effect of ramp size was observed—as it could have been expected when measuring loudness change (Canévet *et al.*, 2003). In these studies, indirect ratings were used, i.e., the loudness was measured at start and end level of the ramp, and loudness change was defined by the ratio of these two measurements. It was hypothesized by Susini *et al.* (2007) that upramps' direct ratings of loudness change could be confounded with those of global loudness—corresponding to the judgment on the total sound energy contained over the duration of the sound—and heavily influenced by end level. A complementary hypothesis is that the perceptual phenomena corresponding to loudness change measured respectively by the two procedures, direct and indirect ratings, are not the same. These hypotheses are based on different published

[a)]Author to whom correspondence should be addressed.

studies, which differ significantly in detail. Therefore, in order to examine these assumptions, direct and indirect ratings of loudness change, as well as direct ratings of global loudness were compared in a similar experimental setup in the present study. Neuhoff (2001), Olsen *et al.* (in press) and Olsen and Stevens (2010) used paired comparisons to compare loudness change of upramps and downramps. In the other studies cited in this introduction, a single stimulus was evaluated in one trial. As the aim of this study was to compare direct and indirect measurements and not the difference between upramps and downramps, only upramps were tested and stimuli were rated individually.

2. Experiments

2.1 Participants

Three groups of 16 participants ranging in age from 14 to 59 years, with a mean of 31, participated in the experiment. All listeners had thresholds of 15 dB HL or lower for octave frequencies from 125 Hz to 8 kHz. Each group performed the primary experiment (loudness function) and then rated the same upramps in one of the three experimental conditions described below. Participants were randomly assigned to each group.

2.2 Stimuli

Synthetic vowel (e) and 1-kHz pure tone were tested. The synthetic vowel sound was based on Klatt's (1980) algorithm with a fundamental frequency of 100 Hz and three formant frequencies (450, 1450 and 2450 Hz) as in Neuhoff's (1998) study. In order to compare our results with those from previous studies (Neuhoff, 1998; Teghtsoonian *et al.*, 2005), similar configurations of the control parameters were chosen.

2.3 Loudness functions

In a primary experiment, loudness functions for the steady-state synthetic vowel and 1-kHz pure tone were obtained in order to compare ratings for each group of participants. Each stimulus had a duration of 500 ms and had a rise and fall of 30 ms.

Loudness functions were measured by Absolute Magnitude Estimation (AME), with similar instructions to those proposed by Hellman (1982). The tone and the synthetic vowel were presented in two different stimuli sequences. Each sequence consisted of 10 steady state stimuli of different levels (from 45 to 90 dB SPL), randomly presented nine times, following the procedure used by Canévet *et al.* (2003) in order to reduce any assimilation effect (see Cross, 1973, for a review). Participants randomly began with the tone or the synthetic vowel sequence.

Loudness estimations (geometric mean magnitude estimates, $N=16$ for each group) follow a power function of sound pressure with an exponent of 0.39, 0.38 and 0.41 for the 1-kHz tone, and 0.40, 0.37, 0.36 for the vowel sound respectively for the three groups of participants. The curves obtained have the usual shapes. The function for the 1-kHz tone is represented in Fig. 1.

2.4 Upramps

Stimuli were the 1-kHz tone and the synthetic vowel sound ramped with a linear increasing level variation in decibels over a duration of 1.8 s as in Neuhoff's (1998) and Teghtsoonian *et al.*'s (2005) studies, with a rise and fall of 30 ms. There was no plateau at either the beginning or the end of the ramps. There were two ramp sizes: 15 and 30 dB. Seven regions of level changes were presented for the 15 dB ramp size: 45–60, 50–65, 55–70, 60–75, 65–80, 70–85, and 75–90 dB SPL; and four for the 30 dB ramp size: 45–75, 50–80, 55–85, and 60–90 dB SPL, for a total of 11 different ramps. Each ramp was presented twice. Therefore, for a given type of sound (1-kHz tone or synthetic vowel), 22 upramps were randomly presented to each participant. Participants randomly began with the tone or the synthetic vowel sequence.

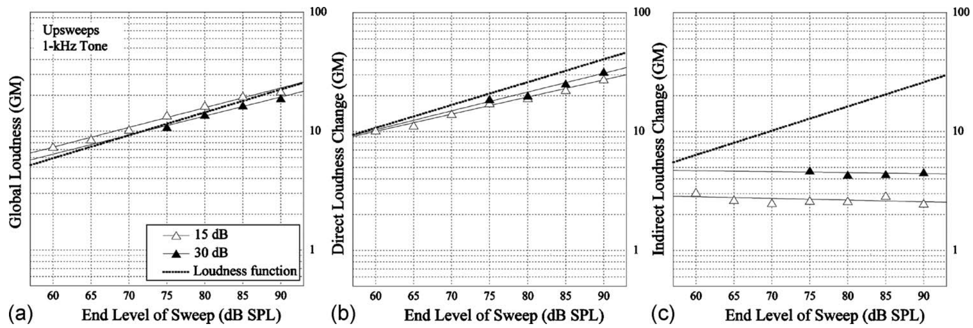

Fig. 1. Ratings for the 1-kHz upramp tones as a function of end level for 15 or 30 dB ramp sizes and a duration of 1.8 s. (a) Direct ratings of global loudness obtained for the first group of participants (left panel), (b) direct ratings of loudness change obtained for the second group of participants (middle panel), and (c) indirect ratings of loudness change for the third group of participants (right panel). The bold lines represent the magnitude estimation function for loudness of a 1-kHz tone, respectively for the three groups of participants (values of the exponent are: 0.39, 0.38 and 0.41).

These stimuli were presented to the three groups of participants. The only difference between the three experimental conditions was the instructions given to the groups.

2.4.1 Direct rating of global loudness

In the first experiment, direct ratings of global loudness were obtained for each upramp. The instructions given to the first group of participants were to assign a number to their global loudness impression for each ramp. Listeners had to answer the question: "How loud was this sound over its entire duration?"

2.4.2 Direct rating of loudness change

In the second experiment, direct ratings of loudness change were obtained. The instructions given to the second group of participants were to assign a number to the amount of loudness change for each ramp. Listeners were told not to make an overall loudness judgment and had to answer the question: "How much did this sound change in loudness?" as it was asked in Neuhoff's (1998) and Teghtsoonian *et al.*'s (2005) studies.

2.4.3 Indirect rating of loudness change

In the third experiment, loudness change was not measured directly, but it was evaluated as the ratio of separate loudness ratings of the start and end levels as in Canévet *et al.*'s (2003) study. Teghtsoonian *et al.* (2005) showed that when an estimation of the start level is asked just after the end of the sound, the loudness is overestimated compared to the loudness of a 10-s plateau played prior to the beginning of an upramp. Then, in this experiment, the loudness estimation of the start level was extracted from the loudness function obtained for the third group during the primarily experiment. The loudness of the end level was rated by participants of this group for each upramp by the method of Absolute Magnitude Estimation. Thus, listeners had to answer the question: "How loud was this sound at the very end?"

3. Results

The effects of the parameters ramp size and end level (region of loudness change), as well as type of sound (1-kHz tone and vowel sound), were evaluated in ANOVAs of log judgments. Three 15-dB ramps were omitted in each analysis (45–60, 50–65, and 55–70 dB), since there was no matching 30-dB with the same end level, which is how they are represented in Fig. 1. Measures obtained in the three experiments are compared with the loudness function of each individual group (bold lines in Fig. 1).

3.1 Direct rating of global loudness

Figure 1(a) presents ratings of global loudness for the two ramp sizes as a function of end level for the 1-kHz tone. The same tendency is observed for the vowel sound. The figure clearly shows that the global loudness increases with end level quite similarly for the two ramp sizes, and follows the loudness function for the 1-kHz tone. Figure 1(a) shows that global loudness of a 1-kHz upramp is similar to the loudness of a steady state sound (loudness function) presented at a level equivalent to the end level of the ramp. The analysis of variance reveals a significant major effect of end level [$F(3,45)=37.7$, $p<0.0001$, $\eta^2=0.08$ [1]], and no interaction between type of sound (tone or vowel) and end level, which means that the end level effect is independent of the type of sound. However, the slopes of the linear regression of the global loudness ratings and of the loudness function are slightly different, especially for the vowel (0.32 against 0.39 for the 1-kHz tone, 0.27 against 0.40 for the vowel).

The analysis also reveals a small, but significant difference [$F(1,15)=28.9$, $p<0.0001$, $\eta^2=0.01$] between ratings obtained for the two ramp sizes: global loudness ratings for the 15 dB ramp size are higher. Finally, the analysis reveals no significant interaction between ramp size and the other factors, implying that the difference between the two ramp sizes is valid independently of the type of sound and the end level.

3.2 Direct rating of loudness change

Figure 1(b) presents direct ratings of loudness change for the two ramp sizes as a function of end level for the 1-kHz tone. The same tendency is observed for the vowel sound. The curves show profiles quite similar to those found for the global loudness, revealing a small but significant effect of the ramp size [$F(1,15)=8.8$, $p<0.01$]–ratings of loudness change are higher for the 30 dB ramp size–and a greater influence of the end level [$F(3,45)=26.7$, $p<0.0001$, $\eta^2=0.02$], loudness change is judged larger when the end level increases as the ramp size is the same. The deviation between the slopes of the linear regression of the direct ratings of loudness change and the slopes of the loudness function obtained for the second group of participants (0.30 against 0.38 for the 1-kHz tone and 0.26 against 0.37 for the vowel) is greater than the one observed for the global loudness experience; slopes are less steep especially for the vowel. This latter result is confirmed by a small, but significant, interaction between type of sound and end level [$F(3,45)=4.3$, $p<0.01$]. Finally, there is a small interaction between type of sound and ramp size [$F(1,15)=4.8$, $p<0.05$], implying that the dynamic effect depends on the type of sound. The difference between 15 and 30-dB ramp sizes is lower for the 1-kHz tone than for the vowel.

3.3 Indirect rating of loudness change

Figure 1(c) presents indirect ratings of loudness change for the 1-kHz tone. The curves profiles are very different to those obtained for direct ratings of loudness change. Similar profiles are obtained for the vowel sound. Results reveal that indirect ratings of loudness change are much higher for the 30-dB ramp size. In addition, the regression lines for the 30 and 15 dB ramp sizes are parallel, and the value of the slopes is nearly zero, which means that indirect ratings for a given ramp size are similar, irrespective of end level. These results are corroborated by the analysis of variance revealing a significant effect of the ramp size parameter [$F(1,15)=236.4$, $p<0.0001$, $\eta^2=0.12$], no effect for the end level and no interactions. On average, loudness change for a 30 dB ramp size is 1.6 times greater than for a 15 dB ramp size. This ratio is similar to the one obtained by Canévet *et al.* (2003) between upramps of 45–75 and 60–75 dB.

4. Discussion and conclusion

The results of the present study showed that direct ratings of loudness change are dependent on the ramp size and on the end level of upramps with a big effect of the end level as in Teghtsoonian *et al.* (2005). Participants were asked to rate loudness change, and thus, similar ratings of loudness change for upramps with the same ramp size irrespective of end level were expected (e.g., [70–85] or [75–90] dB). Indirect ratings of loudness change are only dependent on the size of upramps. An effect of end level has also been found on the global loudness ratings. Both ratings, global loudness and direct loudness change follow the same trend according to the end level, the size of the effect being bigger in the global loudness task than in the direct loudness change task. Thus, the assumption made by Canévet *et al.* (1999) on Neuhoff studies on loudness change stating that: "although Neuhoff cautioned his subjects to avoid making a judgment of overall loudness, they may have done so," may be confirmed. However, both measures, loudness change and global loudness, are slightly affected by the size of the ramp. For two upramps with the same duration and ending at the same level, global loudness is higher for a ramp size of 15 dB compared to a ramp-size of 30 dB, which is consistent with the fact that there is more energy in the first ramp. Inversely, loudness change is higher for a ramp-size of 30 dB, which is coherent, considering that participants were asked to rate this physical variation of the signal.

A comparison of the total variance due to inter-participant differences between the three tasks (82% for global loudness, 96% for direct loudness change, 76% for indirect loudness change) revealed that direct ratings of loudness change were more participant-dependent than direct ratings of global loudness or indirect ratings of loudness change. Asking participants to rate loudness change directly appeared to be the least obvious task, and thus different strategies were involved. It looks like participants understood instructions when asked to rate loudness variations (although the effect of the ramp size was small), but it was difficult for some of them to avoid merging their judgments with a global loudness impression, as their judgments were also strongly influenced by the end loudness. For direct ratings of loudness change, the intra-individual coefficient of variation computed over the different 15 dB ramp sizes were between 0.02, for participant 1, and 0.71, for participant 11. This difference reveals that ratings provided by participant 11 for the different regions of level change showed much greater variation than the ratings provided by participant 1. For example, for the highest regions of level change [60–75], [65–80], [70–85] and [75–90], standardized ratings were 1.0, 1.0, 0.96, 1.0, and 0.4, 0.7, 0.8, 2.0 for participants 1 and 11, respectively. It appears that estimations of participant 1 were non-end level dependent, whereas estimations of participant 11 were. It seems clear that direct ratings of loudness change are very participant-dependent, perhaps because of the use of different strategies by the participants. Comparison of the mean coefficient of variation between direct ratings of loudness change (0.89) and of global loudness (0.50) reveals that the former is much more participant-dependent than the latter.

According to our results, indirect judgments of loudness change are consistent with the physical variations of the sounds. Nonetheless, with this indirect procedure, loudness change is inferred from discrete ratings of loudness; there is no evidence that this measure accurately reflects the listener's impression of loudness change. Neuhoff already raised this point in response to Canévet *et al.* (1999): "There can be inherent shortcomings in inferring characteristics of dynamic perception by extrapolating from static judgements."

Results of the three experiments performed using similar experimental setups show that the perceptual phenomena corresponding to loudness change measured by the two procedures, direct and indirect ratings, are not the same. Moreover, differences between participant's strategies have to be taken into account when measuring loudness change with a direct procedure.

References and links

[1]The rule of thumb adopted here to interpret values Eta-squared is: small=0.01; medium=0.06; large=0.14.

Bach, D. R., Neuhoff, J. G., Perrig, W., and Seifritz, E. (**2009**). "Looming sounds as warning signal: The function of motion cues," Int. J. Psychol. **74**, 28–33.

Canévet, G., Scharf, B., Schlauch, R. S., Teghtsoonian, M., and Teghtsoonian, R. (**1999**). "Perception of changes in loudness," Nature (London) **398**, 673–674.

Canévet, G., Teghtsoonian, R., and Teghtsoonian, M. (**2003**). "A comparison of loudness change in signals that continuously rise or fall in amplitude," Acta. Acust. Acust. **89**, 339–345.

Cross, D. V. (**1973**). "Sequential dependencies and regression in psychophysical judgments," Percept. Psychophys.

14, 547–552.
Hellman, R. P. (**1982**). "Loudness, annoyance, and noisiness produced by single-tone-noise complexes," J. Acoust. Soc. Am. **72**, 62–73.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthetizer," J. Acoust. Soc. Am. **67**, 971–995.
Neuhoff, J. (**2001**). "An adaptive bias in the perception of looming auditory motion," Ecological Psychol. **13**, 87–110.
Neuhoff, J. G. (**1998**). "Perceptual bias for rising tones," Nature (London) **395**, 123–124.
Olsen, K. N., and Stevens, C. J., **2010**, "Perceptual overestimation of rising intensity: Is stimulus continuity necessary?," Perception **39**, 695–705.
Olsen, K. N., Stevens, C. J., and Tardieu, J., "Loudness change in response to dynamic acoustic intensity," J. Exp. Psychol. In press.
Susini, P., McAdams, S., and Smith, B. (**2007**). "Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings," Acta. Acust. Acust. **93**, 623–631.
Teghtsoonian, R., Teghtsoonian, M., and Canévet, G. (**2005**). "Sweep-induced acceleration in loudness change and the 'bias for rising intensities'," Percept. Psychophys. **67**, 699–712.