PERCEPTION OF THE DURATION OF RAMPED AND DAMPED SOUNDS WITH RAISED COSINE RAMPS

Massimo Grassi[1] and Christopher J. Darwin[2]
[1]Università degli Studi di Padova, Dipartimento di Psicologia Generale, Padova, Italy
[2]Experimental Psychology, University of Sussex, Brighton, United Kingdom
E-mail: grassi@psy.unipd.it

Abstract

Schlauch et al. (2001) found that ramped tones (gradual attack and abrupt decay) of 10, 25, 50, 100, 200 ms were perceived as shorter than damped tones (abrupt attack and gradual decay) of the same physical duration. The size of this temporal asymmetry, in percent, depends on the overall duration of the sound: the longer the sound the smaller the asymmetry. In our first experiment listeners were asked to adjust the duration of a constant level sound in order to match the duration of either a similar sound, a ramped sound or a damped sound of 250, 500, 750 or 1000 ms of duration. Ramped and damped sounds were both modulated in amplitude with a raised cosine function. The results, extended to longer durations, confirmed what was previously found by Schlauch et al.: the duration of the damped sounds was systematically underestimated compared to the duration of the ramped sounds. Furthermore, the duration of ramped sounds was slightly underestimated compared to the duration of the steady sounds. In a second experiment, the same phenomena were investigated with a reaction time task. Subjects were asked to react to either the onset or the offset of the target sounds used in the first experiment so that the perceived duration could be calculated as the difference between two reaction times. Results demonstrated that, the delay in the reaction times to the onset of the periodic ramped sounds could explain their underestimation, but the anticipation of the offset of a damped sounds was too small to explain their underestimation. The results of the current research do not support the predictions of the AIM model (Patterson & Irino, 1998). According to this model ramped and damped sounds longer than 200 ms should be perceived as similar in their overall duration.

Schlauch, Ries and DiGiovanni (2001) found that the perceptual duration evoked by ramped and damped sounds is different. With both a magnitude estimation task and a matching task, sounds of 10, 25, 50, 100, 200 ms of duration that increased in level (ramped sounds) were judged to be longer than the same sounds reversed in time (damped sounds). This temporal asymmetry was also related to timbre: the difference between ramped and damped sound was larger with a sine wave than with broadband noise. Moreover, the difference in perceived duration between ramped and damped sounds decreases as the overall duration of the sounds is increased. According to the predictions of the first two stages of the AIM model (Patterson & Irino, 1998; Patterson *et al.*, 1995) the perceptual asymmetry between ramped and damped sounds should disappear for longer sounds since the suppression created by these two stages is limited to the first few milliseconds of the acoustic event, so it cannot have a large effect on the auditory image of long sounds. In this model the sound is filtered using an auditory filterbank, which converts the sound wave into a simulation of the basilar membrane motion (BMM). In a second stage of the model a bank of haircell simulators 'transduces' the BMM into a simulation of the Neural Activity Pattern (NAP) that the sound would produce in the auditory nerve. Both stages of the model act by suppressing across time the amplitude of the input sound. The most effective part of these suppressions are active within ~50 ms of the sound's onset. For this reason sounds longer than 50 ms are progressively less effected by the suppression. The aim of the current research is to see whether the temporal asymmetry found by Schlauch et al. disappears at longer durations as suggested by the AIM model.

Experiment 1: matching task

Target sounds could be either monotonically increasing, decreasing or steady in level. Four durations and four timbres were used for the target sounds, so each listener had to match a total of sixteen different targets. Target durations were 250, 500, 750, 1000 ms. Target timbres were: a sine wave at 1 kHz; a synthetic vowel with F0 at 100 Hz and formants at 450, 1450, 2450 Hz; a complex tone with the first ten harmonics and F0 at 500 Hz and a white noise 20-22050 Hz. Ramped sounds increased in amplitude with a raised cosine amplitude over their duration, their offset was a 12.5 ms raised cosine ramp for all the target durations used. The characteristics of the damped sounds' envelopes were the same as for the ramped sound but reversed in time. Ramped and damped sounds had no steady-state portion. Targets steady in level were turned on and off with two raised cosine ramps of 7.5 ms. The adjustable stimulus was a sound steady in level with 7.5 ms onset and offset raised cosine ramps. The timbre of the adjustable sound was always the same as that of the target. Peak level for all sounds was 65 dB SPL

Eleven listeners participated individually in the experiment in a single-skin sound-attenuated booth. The experiment was divided into sixteen experimental blocks. In each experimental block listeners matched thirty target sounds of the same duration and timbre: ten ramped sounds, ten damped sounds and ten sounds steady in level. Blocks were presented to the subjects in a nested randomised order. The order of the blocks was randomised across subjects. On each trial subjects heard two sounds consecutively: a target sound followed after a 500 ms pause by a sound adjustable in duration. Listeners could match the duration of the target sound by moving a roller ball. At the beginning of each trial the duration of the adjustable sound was chosen randomly from a range of durations varying from ±80% of the duration of the target sound to match.

Results and Discussion

The average duration of the adjustable sound was calculated for each target. This value was converted to the percent of underestimation of the adjustable sound with respect to the matched target. The results are presented averaged across subjects in figure 1. The three envelopes were perceived differently in their duration: $F(2, 20)=98.56$ $p<.0001$. Listeners were very accurate at matching the durations of steady targets: the adjustment error is within 4% of the target's duration for all timbres and all durations. The underestimation of damped sounds was in the range 35% to 27%. By contrast, the durations of ramped targets were slightly underestimated: this underestimation was in the range 3% to 10% for periodic sounds such sine waves, harmonic tones and vowels. The underestimation of ramped white noises was larger and was in the range 16% to 19%. This difference between ramped periodic sounds and ramped white noises was significant: $F(6, 60)=9.602$, $p<.0001$. The percentage underestimation of either ramped tones or ramped noises compared with steady sounds is constant across all target durations: respectively, $F(3, 30)=2.03$ $p>.05$ and $F(3, 30)=0.33$ $p>.05$. However, the underestimation of damped sounds compared to steady sounds seems to

decrease for long durations: the longer the damped target, the smaller the underestimation, $F(3, 30)=7.03$ $p=.001$.
Ramped sounds are slightly underestimated by a constant percentage of their duration. This result could be due to a fixed threshold phenomenon: listeners could only begin to perceive the ramped sound when it reaches threshold and so the beginning would not contribute to the estimated duration. In fact, sounds used in the experiment cover a dynamic range of more than 65 dB, so they begin well below 0 dB HL. In contrast, the percentage underestimation of the duration of damped sounds decreases for longer sounds but the fixed threshold could also contribute to the underestimation. If this threshold exists it could be detected by a reaction time task: reaction times to the onset of the ramped target sounds should be slower than those to the beginning of the damped and the steady sounds; reaction times to the offset of the damped target sounds subjects should be faster than those to the end of the ramped and the steady sounds. Furthermore, the magnitudes of both the anticipation and the delay can be compared to the underestimation obtained by ramped and damped sounds with the matching task.

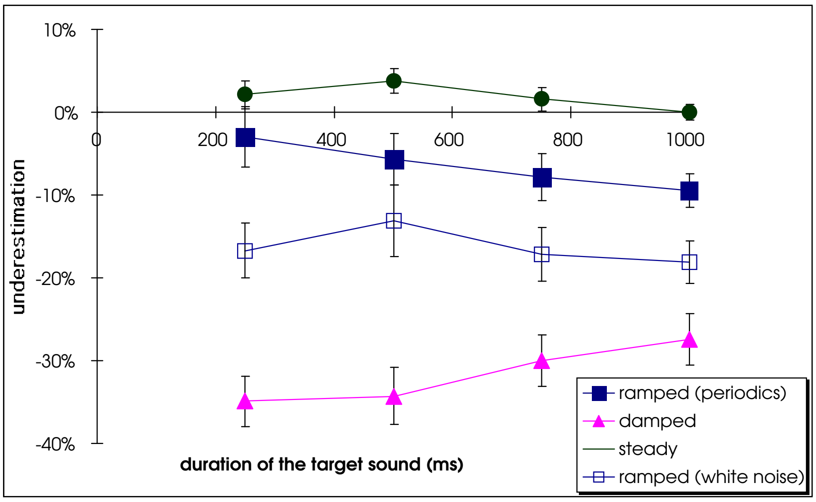

Figure 1: percentage underestimation of ramped, damped and steady sounds at function of the duration of the target.

Experiment 2: reaction time task

Ramped damped and steady sounds of 250, 500, 750, 1000 ms with the same envelopes as those of the first experiment were used as target stimuli. The target could be either the complex tone with the first ten harmonics and F0 at 500 Hz or white noise 20-22050 Hz. No other periodic sound was used since no difference was found within the periodic timbres in the first experiment.

Nine listeners from the first experiment participated individually in the experiment in a single-skin sound-attenuated booth. During each trial listeners heard an alert sound, a 30 ms 2 kHz sine wave, followed after a stimulus onset asynchrony interval (SOA) of either 1000, 1250, 1500, 1750 or 2000 ms, by the target sound. The target sound could be either a ramped sound, a damped sound or a sound steady in level. The subject's task was to press a key as soon as they perceived the onset of the target sound. Every combination of SOA and target was repeated five times during the experimental session. The procedure for the offset reaction time was identical to the onset reaction time but an extra target sound was added. In order to avoid possible anticipation in the reaction to damped sounds a fake damped sound was created. This sound was obtained by adding a low level steady sound to a damped sound. The maximum amplitude of the fake-damped was the same as that of the other targets. In this experiment the subject's task was to press a key as soon as they perceived the offset of the target sound. The same procedures were performed for the reaction time on the onset and offset of white noises. The order of the four sessions was randomised across subjects.

Results and Discussion

In the onset task, subjects were as fast in the reaction to the steady sounds as to the damped sounds. The reaction times to the onset of ramped sounds was delayed and the delay was a function of the duration of the ramped sound: the longer the ramped sound the longer the delay, $F(24, 6)=84.995$ $p<.0001$. Subjects' reactions to ramped sounds occurred after the 11-16% of the overall duration of the sound. In the offset task listeners were as fast in the reaction to the steady sounds as to the ramped sounds. Reaction times to the offset of damped sounds was anticipated: $F(8, 2)=35.603$ $p<.0001$. Subjects' reactions to damped sounds occurred 8-27% before the end of the sound. The difference found between periodic sounds and noises in the previous experiment is not found here: listeners reacted as promptly to tones as to noises both in the onset condition, $F(1, 9)=.003$ $p>.05$, and in the offset condition, $F(1, 9)=.014$ $p>.05$.

The results of the reaction time experiments show that a delayed perception of the onset is involved in the underestimation of ramped sounds and an anticipated perception of the offset is involved in the underestimation of damped sounds.

Comparison of the results of the two experiments

If a fixed threshold exists the durations of ramped and damped sounds should be equally underestimated in both experiments. For this reason, the results obtained with the two procedures were compared. The differences (in milliseconds) between the results obtained by the steady and the ramped sound either in the adjustment task or in the onset reaction times were calculated for each condition and each subject of the second experiment. These differences were calculated also between steady and damped sounds either for the adjustment or for the offset reaction times. Both differences were converted into percent of underestimation of the ramped-damped sound compared to the steady sound. The underestimation of ramped tones is similar in both tasks indicating that the underestimation of ramped tones is due to a perceptual threshold: $F(1, 9)=2.92$ $p>.05$. Ramped white noises, however, are slightly more underestimated with the adjustment task: $F(1, 9)=6.06$ $p=.03$. Furthermore, the underestimation of either damped tones, $F(1, 9)=25.9$ $p=.0007$, or damped white noises was much larger with the adjustment task, $F(1, 9)=20.95$ $p=.001$.

	Periodic sounds							
	Ramped sounds				*Damped sounds*			
Duration	*250*	*500*	*750*	*1000*	*250*	*500*	*750*	*1000*
Reaction time	16%	13%	11%	11%	27%	22%	13%	8%
Matching	5%	9%	9%	9%	36%	37%	31%	27%

	White noises							
	Ramped sounds				*Damped sounds*			
Duration	*250*	*500*	*750*	*1000*	*250*	*500*	*750*	*1000*
Reaction time	13%	14%	12%	12%	28%	21%	11%	8%
Matching	17%	16%	19%	18%	35%	34%	33%	29%

Table 1: percentages of underestimation of ramped and damped sounds compared to the steady sound with the matching task and the reaction time task.

Conclusions

Our experiments have demonstrated three things: firstly, subjects are accurate at matching durations of similar acoustics events; secondly, sounds that increase with a raised cosine amplitude over their duration are slightly underestimated in their physical duration: this underestimation is likely due to a perceptual fixed threshold; thirdly, the same sound reversed in time is substantially underestimated compared to its physical duration: this underestimation cannot entirely be explained by a perceptual threshold. Furthermore, as Schlauch et al. (2001) have shown, the percentage underestimation is related to the overall duration of the damped sound: the longer the sound the smaller the underestimation. According to Schlauch there are two possible explanations for this temporal asymmetry. On one side the abrupt offset at high level of the ramped sound results in a persistence of perception (forward masking) that is considered in judgements of perceived duration. On the other side, listeners may ignore a portion of the decay of the damped sound because they consider it an "echo". The first hypothesis does not find support in our data since ramped sounds are slightly, but consistently, underestimated in duration. However, this could be due either to the range of durations or the kind of ramp (raise cosine) used here. Instead, it seems that the hearing system is not able to track efficiently a sound that monotonically decreases in level. Stecker and Hafter (2000) found an asymmetry by comparing ramped and damped sounds in their perceived loudness. Neuhoff (1998) found an asymmetry in the perception of the change in loudness of such sounds. Patterson (1994) and Irino and Patterson (1996) found, by playing strings of very short ramped or damped sounds, a difference in the timbre evoked by the two strings. The results obtained in the current experiments do not suit the predictions of the AIM model by Patterson and Irino (1998) since the effective duration of the suppression of both filters is too brief and is therefore not effective over the range of duration used.

References

Irino, T., Patterson R. D. (1996). Temporal asymmetry in the auditory system. Journal of the Acoustical Society of America. 99(4), 2316-2331.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. Nature. 395, 123-124.

Patterson, R. D. (1994). The sound of a sinusoid: spectral modes. Journal of the Acoustical Society of America. 96(3), 1409-1418.

Patterson, R. D., Allerhand, M., Giguere, C. (1995). Time domain modelling of peripheral auditory processing: A modular architecture and a software platform. Journal of the Acoustical Society of America. 98(4), 1890-1894.

Patterson, R. D., Irino, T. (1998). Modelling temporal asymmetry in the auditory system. Journal of the Acoustical Society of America. 104(5), 2967-2979.

Schlauch, R. S., Ries, D. T., DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. Journal of the Acoustical Society of America. 109(6), 2880-2887.

Stecker, G. C., Hafter, E. R. (2000). An effect of temporal asymmetry in loudness. Journal of the Acoustical Society of America. 107(6), 3358-3368.