

Acknowledgements	3
Prologue.....	5
References.....	6
Subjective duration of ramped and damped sounds	7
Riassunto.....	7
Abstract	8
Introduction.....	10
Experiment one	15
Method	15
Results.....	17
Discussion.....	19
Experiment two	20
Method	20
Results.....	21
Discussion.....	24
Experiment three	25
Method	25
Results.....	27
Discussion.....	28
Comparison of experiments 2 and 3 and discussion.....	29
Experiment four	31
Method	31
Results.....	33
Comparison of the results of the two procedures.....	37
Discussion.....	38
Experiment five	40
Method	41
Results.....	44
Discussion.....	47
General discussion.....	48
References.....	54

From sound to size: balls dropped on plates	57
Riassunto.....	57
Abstract	58
Introduction.....	60
Experiment one	63
Method	63
Results and discussion.....	64
Method	66
Results and discussion.....	67
Experiment three	69
Method	69
Results and discussion.....	70
Analysis of the physical and acoustical event.....	72
Analysis of experiment one	74
Analysis of experiment two.....	74
Analysis of experiment three.....	75
Experiment four	76
Method	76
Results and discussion and comparison with results of previous experiments	76
General discussion.....	79
References	85
Appendix.....	87

Acknowledgements

For the first research I would like to thank prof. Chris J. Darwin (University of Sussex). All the experiments performed have been conducted while I was visiting scientist in the Hearing Laboratory of the University of Sussex. The fifth experiment has been performed while I was research fellow under a Marie Curie contract.

For the second research I would like to than prof. Davide Rocchesso (Università di Verona). Thank to Davide we have been involved in a very fructuous project (“The Sounding Object”). All the experiments of the second research have been supported by the European Commission under contract IST-2000-25287 (“SOB – The Sounding Object”: www.soundobject.org).

Prologue

The progress of any research area goes along with the progress of the technological tools used for the investigations. Psychoacoustics is no exception. After the second world war the development of vacuum tube electronics lead to a noticeable progress in the investigations of auditory phenomena. By means of such equipments the experimenter could produce accurately a number of tones or noises of precise duration, level spectral distribution, etc. Nonetheless, while on the one side such a new equipment permitted the investigation of sounds in a highly controlled fashion it also directed the research towards the investigation of those auditory events that were easy to produce with such a technology. Consequently, classical investigations in psychoacoustics dealt with sounds or noises constant in level and in spectral distribution. In synthesis, the research pointed its effort towards those phenomena that were easily obtained with the equipment.

However, our everyday environment is far from being full of sounds constant in frequency and level. On the contrary, the majority of sounds we listen to is spectrally complex and, moreover, not constant in frequency and level. As a consequence, the traditional approach to auditory perception was only partially informative about how our hearing system deals with everyday sounds. We were in fact facing a paradox: we knew much about how the hearing system codes sounds of fixed frequencies and levels but such sounds could be found only in the laboratory.

The study of the perception of everyday sounds is fundamental. In fact, with no doubts our hearing system has evolved to deal with such sounds rather than the unrealistic stimulations of the laboratory. In the recent past the development of the digital technology has lead to a considerable increment in the possibility of the psychoacoustical research. Nowadays, the experimenter can easily obtain sounds of any level of complexity. It is not by chance that during the last congress of the Acoustical Society of America (Nashville, May 2003) an entire scientific session was dedicated to the discussion of researches in ecological psychoacoustics. The goal of this new topic is to study the perception of everyday sounds by means of the rigorous approach of the traditional psychoacoustics. A strong importance is given to both the stimulation itself and the goal of the research. The stimulation has to reflect aspects of everyday sounds as the goal of the research is to understand how the hearing system has evolved to deal with such sounds.

The ecological psychoacoustics investigates the perception of sounds by means of stimulations that reflect some aspects of everyday sound or everyday sounds themselves. In the

current dissertation I present two examples of researches in ecological psychoacoustics. In the first, I studied the perception of sounds monotonically increasing or decreasing in sound pressure level. In the second, I studied the perception of the size of an object from the sound it produces impacting a second object.

References

Neuhoff, J. G. (in press). Ecological psychoacoustics: Introduction and history. In J. G. Neuhoff (Ed.) *Ecological Psychoacoustics*. New York: Academic Press.

Subjective duration of ramped and damped sounds

Riassunto

La durata soggettiva di un suono non dipende esclusivamente dalla sua durata fisica ma anche dal modo in cui la sua intensità cambia nel tempo. In una recente ricerca Schaluch, Ries e DiGiovanni (2001) hanno dimostrato come brevi suoni (da 10 a 200 ms) crescenti monotonicamente in livello di pressione sonora (suoni crescenti) siano percepiti come più lunghi degli stessi suoni rovesciati nel tempo (suoni calanti). Ma i suoni crescenti non sono solamente percepiti come più lunghi dei suoni calanti, essi vengono percepiti anche come aventi un'intensità soggettiva maggiore (Stecker e Hafter, 2000) e un cambiamento d'intensità soggettivo maggiore (Neuhoff, 1998, 2001) dei suoni calanti.

Patterson e Irino (1998) hanno proposto un modello del sistema uditivo periferico che spiega la differenza percettiva di tali suoni per durate brevi (> 50 ms). Per durate superiori, sia Stecker e Hafter (2000) che Neuhoff (1998, 2001) hanno proposto un'ipotesi a spiegazione della differenza soggettiva dei suoni crescenti e calanti. Secondo Stecker e Hafter (2000) gli ascoltatori percepirebbero come un eco la debole coda del suono calante, quindi, essi non includerebbero tale porzione nel computo dell'intensità sonora del suono stesso. Per contro, secondo Neuhoff (1998, 2001), il percetto generato dai suoni crescenti sarebbe accresciuto rispetto a quello generato dai suoni calanti.

Entrambe le ipotesi quindi predicono una differenza nella durata soggettiva dei suoni crescenti e calanti per durate maggiori ai 200 ms ma si differenziano nel trovare la causa di tale differenza: secondo la prima ipotesi la differenza avrebbe origini cognitive; secondo la seconda ipotesi la differenza avrebbe origini sensoriali. Scopo dei miei esperimenti era duplice: (I) investigare se i suoni crescenti fossero percepiti come più lunghi dei suoni calanti con suoni di durata superiore ai 200 ms; (II) investigare quale delle due ipotesi potesse spiegare una eventuale differenza.

Negli esperimenti ho misurato le durata soggettiva dei suoni crescenti e calanti (durate di 250, 500, 750, 1000 ms). I risultati hanno dimostrato che sia i suoni crescenti che quelli calanti sono percepiti come più brevi rispetto a suoni costanti in livello di pressione sonora nel tempo (suoni costanti). La durata dei suoni costanti è stata stimata veridicamente, la durata dei crescenti è stata sottostimata di circa il 10% mentre la durata dei calanti è stata sottostimata di circa il 35-40%.

Successivi esperimenti hanno dimostrato come l'inizio dei suoni crescenti e la fine dei suoni calanti siano sotto soglia, troppo poco intensi per poter essere percepiti. Tuttavia, sia la soglia percettiva agente all'inizio dei suoni crescenti che quella agente alla fine dei suoni calanti potevano spiegare sottostime del 10-15% della durata del suono. In conclusione, i risultati degli esperimenti dimostrano che (I) i suoni crescenti sono percepiti come più lunghi dei suoni calanti. Ciò nonostante (II) i risultati degli esperimenti non ho fornito alcuna evidenza a supporto di un origine sensoriale di tale differenza.

Abstract

The perceived duration of a sound does not simply depend on its actual duration. It also depends on the way that its amplitude changes over time. Schlauch, Ries, and DiGiovanni (2001) showed that the subjective durations of ramped sounds of 10 to 200 ms of duration (slow attack and fast decay) were much longer than the subjective durations of the same sounds reversed in time (damped sounds). However, other experimental evidences showed that, with durations exceeding 200 ms, ramped sounds were perceived as louder than damped sounds (Stecker and Hafter, 2000) and ramped sounds were perceived as more changing in loudness than damped sounds (Neuhoff, 1998, 2001).

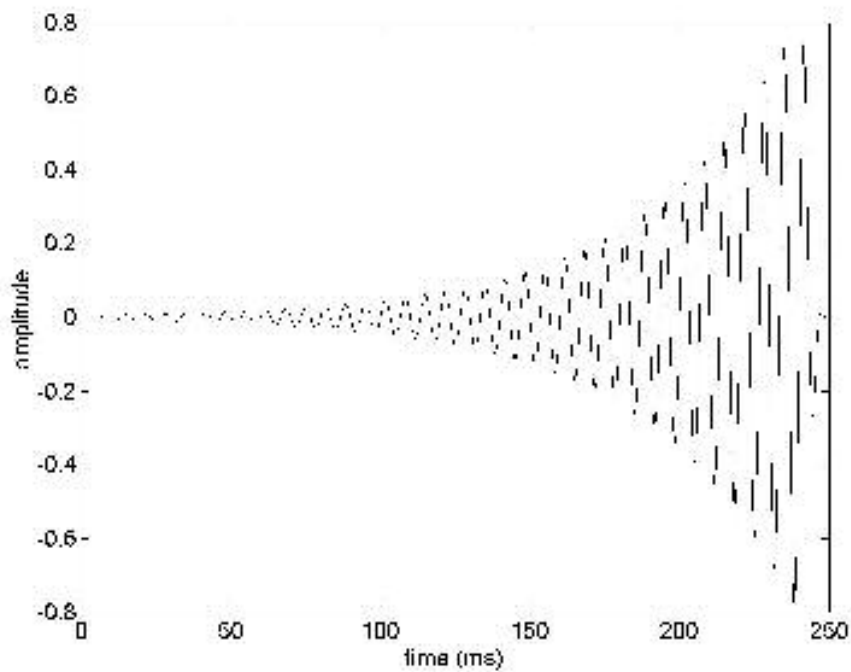
For brief durations (< 50 ms) a model of the auditory system (Patterson & Irino 1998) can explain the difference in the perception of the two sounds. For durations exceeding 200 ms both Stecker and Hafter (2000) and Neuhoff (1998, 2001) proposed hypotheses for explaining the different percepts evoked by the two sound. According to Stecker and Hafter (2000) listeners judge the quiet tail of a damped sound as the result of an echo and for this reason they do not include such portion of the sound in the computation of its overall loudness. On the contrary, according to Neuhoff (1998, 2001), the difference rises as the two sounds are coded differently in the auditory system.

The aim of the current experiments was to: (I) test whether ramped sound remained longer than damped sounds for duration exceeding 200 ms; (II) test which of the two hypotheses could explain the difference in subjective duration between the two sounds. I measured the subjective duration of ramped and damped sounds in comparison to a sound steady in level across time (steady sound) and I found that: the duration of steady sounds was estimated veridically, and both ramped and damped sounds were underestimated in duration. Ramped sounds were only slightly

underestimated (10%) while damped sounds were much more underestimated (40%). In other experiments, I found that the beginning of ramped sounds and the end of damped sounds were sub-threshold, therefore, too quiet to be perceived by the listener. However, both the threshold acting at the onset of ramped sounds and the threshold acting at the offset of damped sounds could explain underestimations up to a maximum of 10% of the duration of the sound. In conclusion, (I) ramped sounds are perceived as longer than damped sounds for duration exceeding 200 ms. However, (II) the experiments failed to find any evidence supporting a difference in the sensory coding of the two sounds.

Introduction

The perceived duration of a sound does not simply depend on its actual duration. It also depends on the way that its amplitude changes over time. A sound that starts at a low level and gradually increases in amplitude throughout its duration will be heard as about 30% longer than the same sound reversed in time so that it starts at a high level and then decreases in amplitude. Schlauch, Ries and DiGiovanni (2001) found such an asymmetry in subjective duration between ramped (increasing in level) and damped (decreasing) sounds ranging from 10 to 200 ms. In their first experiment the subjective duration of ramped and damped sounds was measured using the magnitude estimation procedure. In a second experiment, the subjective duration was measured with a matching procedure. Both experiments gave similar results. Ramped sounds were always perceived longer than damped sounds having identical physical duration. Furthermore, the difference in subjective duration between ramped and damped sounds was larger when measured with tones than with broad band noises.



Waveform of a 150 Hz ramped sound of 250 ms. The waveform of a damped sound is the same waveform reversed in time.

Other perceptual asymmetries in the perception of sounds that monotonically increase or decrease in amplitude over time have been reported. The decruitment phenomenon (Canévet, 1986, Canévet & Scharf, 1990; Schlauch, 1992; Teghtsoonian, Teghtsoonian, & Canévet, 2000; Canévet, Teghtsoonian, & Teghtsoonian, 2003) involves the loudness of sounds that increase or decrease continuously in intensity over long durations (usually from 10-20 seconds up to two or three minutes). Loudness estimates for sounds that increase and decrease in level are usually similar to each other except at low levels where the loudness of damped sounds decreases quite abruptly. This decruitment phenomenon is usually interpreted as the result of auditory adaptation.

Asymmetries have also been found for other psychological dimensions (*i.e.* overall loudness, change in loudness, timbre) and for sounds of much shorter durations. Intensity changes in ramped sounds are more salient than those in damped. Small (1977) used temporally reversed sound to show that listeners more easily detected whether a 2000-ms tone was increasing than decreasing in level. In a similar vein, Stecker and Hafter (2000) found that a 250-ms tone that has a slow increase and a fast decrease in amplitude is overall louder than the same sound when reversed in time. Furthermore, this difference is even greater when the fast-slow tone precedes the slow-fast one.

Further evidence for the salience of increases in level comes from the studies of Neuhoff and his colleagues. Neuhoff (1998) showed that the subjective change in loudness of a 1.8 s sound that increases in level over time was greater than the subjective change in loudness of the same sound reversed in time. Although this asymmetry was evident for tones (either sine waves, complex tones or speech like sounds) it did not occur with broadband noises. Furthermore, Neuhoff found the asymmetry between rising and falling tones in two other experiments (Neuhoff, 2001; Seifritz, Neuhoff, Bilecen, Scheffler, Mustovic, Schächinger, Elefante, & Di Salle, 2002). Recently, Canévet et al. (2003) failed to replicate Neuhoff's results. They asked a group of listeners to evaluate the initial and the final loudness of a 1.8 s sound rising (or falling) in intensity. Results showed no difference in the change in loudness of rising tones versus falling tones. However, in the experiments performed by Neuhoff (1998, 2001) listeners were asked to judge the change in loudness directly, not the initial and final loudness of the sound.

Asymmetries between increases and decreases in intensity have also been found for much shorter tones. Patterson (1994a, 1994b) demonstrated that the timbre of twenty short (25 ms each) concatenated ramped tones is different from the timbre of the same sound reversed in time. In particular, while the sequence of ramped tones is perceived as a rhythmic component together with a sinusoidal component, in the string of damped sounds only the sinusoidal component can be

heard. In two papers Akeroyd and Patterson (1995) and Irino and Patterson (1996) showed that the perceptual difference between ramped and damped sounds extends also to noise carriers: the percept from ramped noises contains a continuous hiss that the percept from damped noises lacks. However, the asymmetry measured with white noises was smaller than with tones (Irino & Patterson, 1996). In all these experiments the perceptual difference between ramped and damped sounds was also as function of the half-life¹ of the amplitude modulator: its rate of change in dB/s. For very short half-lives (>2 ms) the perceptual quality of ramped and damped sounds was similar, both of them evoked a click-like sensation. For half-lives ranging from 2 up to 16-32 ms the difference was maximal. Finally, for half-lives longer than 32 ms both sounds were perceived as possessing the continuous plus rhythmical quality.

Asymmetries in the perception of ramped versus damped sounds of widely-different duration are thus well-established in humans. Moreover, similar, asymmetries have also been found with animals both with behavioural tasks (Ghazanfar, Neuhoff, & Logothetis, 2002) and with physiological recordings at various points of the auditory pathway (Fay, Chronopoulos, & Patterson, 1996; Winter, & Patterson, 2000; Neuert, Pressnitzer, Patterson, & Winter, 2001, Lu, Liang & Wang, 2001a; Lu, Liang & Wang, 2001b).

Although all experiments described above asked listeners to evaluate different aspects of the sound stimulus, the results obtained are strikingly similar. Firstly, almost all the experiments reported asymmetries in the perception of ramped and damped sounds; Canévet et al. (2003) reported the only result where no asymmetry was found. Secondly, the asymmetry is clearly present with tones but less clear with white noise (Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Schlauch & al. 2001) or null (Neuhoff, 1998; Neuhoff, 2001).

There are currently three different approaches for explaining the asymmetry between ramped and damped sounds. For short-duration sounds, Patterson and his colleagues proposed an auditory model (AIM) (Patterson, Allerhand & Giguere, 1995; Patterson & Irino, 1998) in which the sound is first filtered by an auditory filterbank, which converts the sound wave into a simulation of the basilar membrane motion. In a second stage of the model a bank of haircell simulators transduces the motion into a simulation of the neural activity pattern that the sound would produce in the auditory nerve. Both stages of the model suppress across time the amplitude of the input sound. The most effective suppression occurs within ~50 ms of the sound's onset. For this reason sounds longer than 50 ms are progressively less affected by the suppression. According to the predictions of the first two stages of the model (Patterson & al., 1995; Patterson & Irino, 1998) the

¹ The half-life of a sound is the time required to reach the half of its initial amplitude.

perceptual asymmetry between ramped and damped sounds should disappear for durations longer than ~50 ms. The first two stages of the AIM model can predict many of the results obtained by Schlauch et al. (2001). However, the data collected by these authors also show that the perceptual asymmetry persists at longer durations. Even a damped sound of 200 ms needs to be 1.5 times longer than a ramped sound in order to have the same subjective duration. Patterson's model can thus explain well the asymmetry for very brief durations, however, it does not explain why the perceptual asymmetry persists for sounds much longer than a few milliseconds as in the experiments of Neuhoff and colleagues (Neuhoff, 1998, Neuhoff 2001, Seifritz, & al. 2002), Stecker and Hafter (2000) and the longer-duration sounds used by Schlauch et al. (2001).

According to Stecker and Hafter (2000), the different loudnesses evoked by ramped and damped sounds is the result of listeners ignoring the ending part of damped sounds. They argue that natural sounds are generally characterised by a fast attack and a decay that may be prolonged by environmental reverberation. Listeners may interpret the ending part of a damped sound as due to reverberation rather than being intrinsic to the sound source and consequently ignore it in the computation of subjective loudness. The echo hypothesis also predicts that ramped sounds will be perceived as longer in duration than damped sounds since, if listeners ignore the ending part of damped sounds they will be perceived as shorter than ramped sounds of the same physical duration.

On the other hand, Neuhoff (1998, 2001) proposed that the perceptual asymmetry is due to an evolutionarily advantageous bias for tones increasing in intensity. Such tones have the same amplitude change as ones steady sound sources which are approaching the listener. On the contrary, tones decreasing in intensity would indicate a sound source going away from the listener. An overestimation of the change in loudness of increasing sounds could facilitate reactions to the approaching sources. The existence of the asymmetry for tones and not for noises would indicate that the asymmetry is selective for biological motion versus non biological motion, therefore, selective for the motion of a single coherent sound source versus multiple incoherent sound sources (Neuhoff, 2001). According to Neuhoff, tones are normally generated by biological sources while noises, although they can also be produced by biological sources, are usually the result of multiple sound sources sounding together. Furthermore, Neuhoff found that the listening of rising and falling sounds activate regions of the cortex usually associated with the perception of horizontal and vertical auditory motion (Seifritz, Neuhoff, Bilecen, Scheffler, Mustovic, Schächinger, Elefante, & Di Salle, 2002). According to Neuhoff (1998, 2001) listeners are more sensitive to approaching sound sources than receding sound sources. If this higher sensitivity would correspond to listeners

overtracking the change in level of rising sounds compared to falling sound we might expect that sounds rising in intensity are also perceived longer in duration than sounds falling in intensity. Furthermore, Neuhoff (2001) speculates that such different sensitivity may result from the two sounds being coded differently in the neural auditory pathway.

In summary, Stecker and Hafter's (2000) hypothesis attributes the asymmetry to listeners unconsciously ignoring the ending part of the damped sound, while Neuhoff's (1998, 2001) hypothesis, by postulating a bias to the ramped sound, considers that the asymmetry may originate from evolutionarily advantageous, although unspecified, sensory processing. Both of the explanations proposed by Stecker and Hafter (2000) and Neuhoff (1998, 2001) can account for the results of experimental studies using sounds longer than those used in the experiments performed by Patterson and his colleagues (Patterson, 1994; Akeroid & Patterson, 1995; Irino & Patterson, 1996). The aims of the current research are to investigate (I) whether the asymmetry in subjective duration reported by Schlauch et al. (2001) persists for durations longer than 200 ms and (II) whether the hypotheses proposed by Stecker and Hafter (2000) and Neuhoff (1998, 2001) may explain the asymmetry in the subjective duration between ramped and damped sounds.

Experiment one

In the following experiment the subjective duration of ramped and damped sounds was measured by means of an adjustable sound that was constant in amplitude across time (steady sound): listeners manipulated the duration of the steady sound in order to match the duration of one of three target sounds: ramped, damped or steady. The adjustable sound could either precede or follow the target. This procedure enabled us to measure the subjective duration of ramped and damped sounds in comparison with a control condition (steady vs. steady sound).

The goal of this first experiment was twofold. First, to measure the asymmetry in perceived duration for ramped and damped sounds for a wider range of durations than that used by Schlauch et al. (2001). My informal observations had shown that the subjective duration of ramped sounds remained longer than that of damped sounds for durations exceeding 200 ms. Second, to control for and thereby clarify the size of the time order error in which the relative duration of a stimulus is influenced by its position within a sequence (see Allan, 1979 for a review).

Method

Listeners

Eight paid listeners participated in the experiment. None of them reported a hearing loss or other difficulties with their hearing. Four listeners were naïve, and four were PhD students or post doctoral fellows in hearing perception.

Stimuli

Stimuli were ramped sounds, damped sounds and steady sounds of 125, 250, 500, 750, 1000 ms of duration. Steady sounds were gated on and off with two raised cosine ramps of 7.5 ms of duration but were otherwise constant in level. Ramped sounds were obtained by modulating the amplitude of a tone with a raised cosine function:

$$a(t) = \frac{1}{2} \left(1 - \cos \left(\pi \left(\frac{t}{T} \right) \right) \right)$$

where $a(t)$ is the amplitude of the sound as a function of the time t , and T is the overall duration of the sound ($T = 125, 250, 500, 750, 1000$ ms). The offset of ramped sounds was amplitude modulated with a 15-ms raised cosine ramp. Damped sounds were obtained by reversing in time the ramped sounds. The timbre of all sounds was a complex tone with the first ten harmonics of a 500-Hz fundamental all at the same amplitude.

Apparatus

Sounds were synthesised in real time at 44.1 kHz and 16 bits resolution using a custom software running on a Macintosh Power PC that also controlled the experiment. The output of the sound card was passed through Tucker-Davis attenuators and presented through Sennheiser HD414 headphones in a single-skin sound-attenuated IAC booth.

Procedure

During the experiment the listeners' task was to match the duration of a steady adjustable sound to that of a fixed duration ramped, damped, or steady target sound that either preceded or followed it by 500 ms. The experiment was divided into ten experimental blocks. Five blocks consisted of trials where the adjustable sound preceded the target, and five where it followed it. During each block, listeners matched fifteen target sounds of the same duration presented in random order: five ramped, five damped and five steady. Different blocks were taken in a random order with each block lasting about 15 minutes. Peak level for all targets and adjustable sounds was 65 dB SPL. During a trial the listener adjusted the duration of the target sound by moving a roller ball. The listener could hear the result of the adjustment by pressing the roller-ball's button. As the listener moved the roller-ball, a screen cursor also moved which was recentered after each button press so that the listeners could not base their adjustment on the cursor's absolute position across adjustments. Moving the cursor by half a screen led to a change in the adjustable sound of about 1/8 of the target sound's duration. At the beginning of each trial, the duration of the adjustable sound was chosen randomly from a range of $\pm 80\%$ of the duration of the target sound. This range was also the maximum permitted adjustment during each experimental trial. If the adjustment was outside the permitted range, an alert sound was played and the duration of the adjustable sound was reset randomly. At the beginning of the experiment, listeners made one or two adjustments in the presence of the experimenter to familiarise them with the task. The results of these adjustments were not analysed. After the experiment many listeners reported that the task was difficult for the shortest target sounds (125 ms).

Results

The average duration of the adjustable tone (in milliseconds) was calculated for each target tone duration and sequence position for each listener. These values were converted into the difference between matched duration and target duration as a percentage of target duration (D%) according to the following equation:

$$D\% = \frac{d_{\text{vrt}} - d_{\text{t}}}{d_{\text{t}}}$$

where d_{vrt} is the matched duration (in ms) of a given target sound and d_{t} is the actual duration of that target sound (see figure 1). Positive values of D% indicate overestimation of the duration of the target sound, negative values indicate underestimation and a null value a correct match.

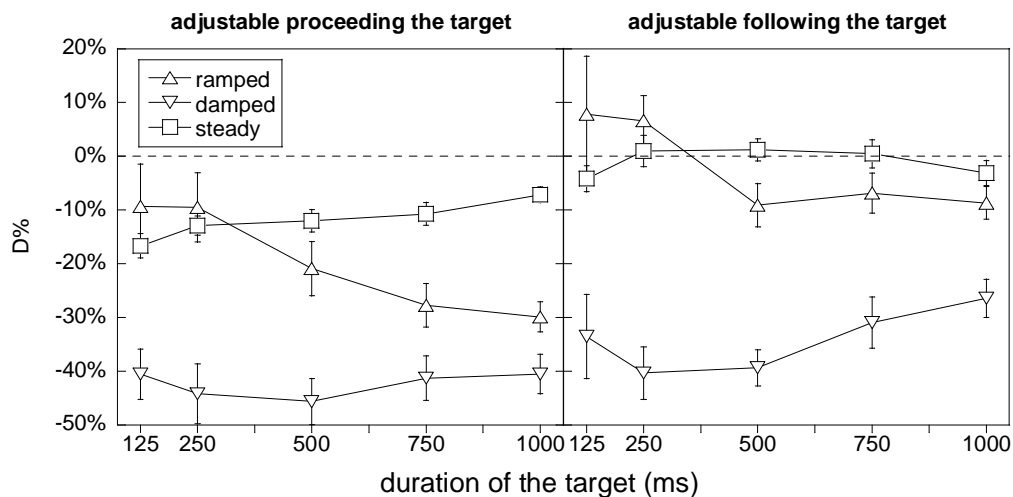


Figure 1: D% as a function of the overall duration of the target sound. Results obtained with the adjustable sound preceding the target sound are presented in the left graph. The horizontal dashed line represent perfect match. Vertical bars represent ± 1 standard error of the mean.

A first analysis was conducted exclusively on the results collected with the control condition (steady vs. steady sound) in order to investigate which target-adjustable sequence induced less time order error in its estimation. A 2 (position of the steady target) \times 5 (duration of the steady target) ANOVA was conducted on D% recorded with steady targets. Listeners underestimated slightly the steady target when the adjustable sound was preceding it, but gave veridical judgements when the adjustable sound followed the steady target: $F(1, 7)=24.80, p=.002$. The overall duration of the steady target did not influence the extent of underestimation, $F(4, 28)=3.68, p>.05$, nor were

listeners' estimations different for the two steady target-adjustable sequences across the five durations investigated: $F(4, 28)=1.82, p>.05$. The D% measured for the steady target preceding the adjustable and steady target following the adjustable were averaged separately across the five target durations. A one sample t-test (D% =0 null hypothesis) showed that the duration of the steady targets was significantly underestimated when the steady target was following the adjustable sound: $t(7)=-8.59, p<.0001$. A second t-test revealed that when the steady target was preceding the adjustable sound listeners neither overestimated nor underestimated the duration of the steady target: $t(7)=-.613, p>.05$.

A second ANOVA was performed in order to see whether the three target sounds (ramped, damped and steady) differed in their matched durations. A 2 (position of the target) \times 3 (shape of the envelope) \times 5 (duration of the target) ANOVA was performed on D% measured with all target sounds. Listeners estimated the three target sounds as having different durations depending on their envelopes: $F(2, 14)=42.23, p<.0001$. On average, steady sounds were slightly underestimated in duration, ramped sounds were moderately underestimated in duration and damped sounds were consistently underestimated in duration. Furthermore, the difference between ramped, damped and steady sounds D% depended on whether the target preceded or followed the adjustable sound: $F(2, 14)=7.75, p=.005$. On average, the D% of the three targets were markedly different from each other when the target followed the adjustable sound with the steady targets being underestimated by 12%, ramped targets being underestimated by 19% and damped targets being underestimated by 43%. On the contrary, when the target preceded the adjustable sound, the D% measured with ramped, damped and steady targets assumed more close values: 2%, 2% and 35% the underestimations measured for, respectively, steady, ramped and damped sounds.

For this reason I analysed in detail whether the order of the target adjustable sequence influenced the difference in subjective duration between ramped and damped sounds. I performed a 2 (order adjustable-target) \times 2 (ramped or damped target) \times 5 (target duration) exclusively with the D% collected with ramped and damped sounds. Ramped sounds were always judged longer than damped sounds, however, the difference in subjective duration between the two stimuli was even greater when the target preceded the adjustable sound: $F(1, 7)=22.63, p=.002$. In fact, when the target preceded the adjustable sound, the difference between the D% measured with ramped and damped sounds ranged from 42% to 17%. On the contrary, when the target sound followed the adjustable sound, the difference between the D% measured with ramped and damped sounds ranged from 35% to 10%.

Discussion

The results collected with steady targets demonstrated that listeners were able to estimate veridically the steady targets' durations when the steady target preceded the adjustable sound (average D% -1%). When the steady target followed the adjustable sound, listeners underestimated the duration of the steady target (average D% -11%). Overall, ramped targets were matched as shorter than steady targets having the same physical duration and this result was evident both when the ramped target preceded and when it followed the adjustable sound. The matched duration of ramped targets only slightly exceeded the duration of steady targets for the two shortest durations investigated (125 and 250 ms). However, the variance in the adjustments performed with ramped targets for such durations was the highest and such a variability corroborates listeners' reported difficulty in estimating the duration of the shortest sounds. On the contrary, for the three longest duration (500, 750, 1000 ms) the matched duration of ramped targets was moderately shorter than that of steady targets. Damped targets were always substantially underestimated compared to ramped or steady targets. D% for damped targets ranged from -45% to -30% regardless of the sequence order of target and adjustable sound.

In conclusion listeners underestimated both the duration of ramped sounds (moderately) and the duration damped sounds (substantially). This result did not depend on order errors in duration matching since listeners were able to match veridically the duration of steady sounds. In addition, the asymmetry in subjective duration between ramped and damped sounds was not fixed, but depended somewhat on the sequence order of target and adjustable sounds.

Experiment two

The first experiment showed that although listeners were able to match veridically the duration of a steady sound they substantially underestimated the duration of damped sounds and moderately underestimated ramped sounds. The aim of the current experiment was to investigate whether the subjective duration of ramped and damped sounds depended on the timbre of the carrier. Schlauch et al. (2001) found that the asymmetry in subjective duration between ramped and damped sounds was large with tones but smaller (although consistently present) with noises. A reduction of the ramped-damped asymmetry with noises has also been found in studies investigating psychological continua other than duration (Irino & Patterson, 1996; Neuhoff, 1998; Neuhoff, 2001).

In experiment 2 I also aimed to replicate, with double the number of matches, the results of the first experiment from the conditions that had the adjustable sound following the target sound – the condition that minimised the time order error.

Method

Listeners

Eleven paid listeners, recruited from the University population, participated in the experiment. None of them reported a hearing loss or other difficulties with their hearing. All listeners were naïve and none of them had participated in the previous experiment.

Stimuli

Stimuli were ramped sounds, damped sounds and steady sounds of 250, 500, 750, 1000 ms duration. The shortest duration for the target sound used in the previous experiment (125 ms) was excluded from the current experiment as such a duration was within the range of durations investigated by Schlauch et al. (2001). The amplitude envelope shapes were identical to those used in Experiment 1. Four timbres served as carriers for target and matching sounds: (I) a complex tone with the first ten harmonics at same amplitude ($f_0 = 500$ Hz); (II) a 1 kHz sine wave; (III) a synthetic vowel ($f_0 = 100$ Hz) with formants at 450, 1450, 2450 Hz and bandwidths of, respectively 90, 110, 110 Hz, and (IV) a white noise.

Apparatus and procedure

The apparatus was identical to the apparatus used in the first experiment. The procedure was also similar except that all matches were performed with the target sound preceding the adjustable sound. The experiment was divided into 16 experimental blocks. During each block a trial consisted of a target sound followed after 500 ms by an adjustable sound, of identical timbre, whose duration listeners had to match to that of the target. The adjustable sound was steady in amplitude and gated on and off with two raised cosine ramps of 7.5 ms. Each experimental block contained ramped, damped and steady targets sharing the same timbre and duration. Within each experimental block listeners matched thirty target sounds presented in random order: ten ramped sounds, ten damped sounds, ten steady sounds. Listeners took the experimental blocks in a nested randomised order. Experimental blocks with targets of the same overall duration were taken consecutively to obtain more consistent adjustments (Wright, Buonomano, Mahncke and Merzenich, 1997) and the order in which the different durations were taken was randomised across listeners. Within these four series of four blocks sharing the same target duration, the timbre order was randomised. Each block lasted about 30 minutes. Peak level for all target and adjustable sounds was 65 dB SPL.

Results

	Mean				Standard error of the mean			
	250	500	750	1000	250	500	750	1000
sine	2.57%	3.36%	2.35%	0.31%	1.93%	1.95%	1.71%	0.93%
complex tone	0.76%	3.38%	1.35%	-0.68%	1.59%	1.53%	1.63%	1.25%
vowel	2.97%	4.59%	1.13%	0.54%	1.99%	1.39%	1.88%	1.04%
white noise	0.61%	3.13%	1.42%	-0.13%	1.60%	1.56%	1.81%	0.96%

Table 1: D% measured for steady targets as a function of the overall duration of the steady sound. The left part of the table shows means D%, the right part of the table shows the correspondent standard error of the mean.

The average duration of the adjustable sound in milliseconds was calculated for each target sound and for each listener. A preliminary 4 (timbres) \times 4 (durations) ANOVA was performed on D% collected with steady targets in order to test whether timbre and overall duration were affecting the subjective duration of steady targets (see table 1). Listeners were highly veridical in estimating the duration of the steady targets and such veridicity was neither affected by the timbre of the carrier, $F(3, 30)=1.05, p>.05$, nor by the overall duration of the sound, $F(3, 30)=2.17, p>.05$, nor by these two factors interacting, $F<1$. All D% collected with steady targets were averaged separately

for each listener. A one sample t-test ($D\% = 0$, null hypothesis) performed on these data showed that listeners' $D\%$ for steady targets were in fact no different from zero: $t(10)=1.72, p>.05$.

A second ANOVA 4 (timbre of the target) \times 3 (shape of the envelope) \times 4 (duration of the target) was performed on $D\%$ measured with all target sounds. Listeners perceived the three envelopes as having different durations, $F(2, 20)=97.99, p<.0001$. Furthermore, three contrasts highlighted that ramped sounds were perceived moderately shorter than steady sounds, $F(1, 10)=19.00, p=.001$, and that damped sounds were perceived much shorter than either steady sounds: $F(1, 10)=152.35, p<.0001$ or ramped sounds, $F(1, 10)=115.53, p<.0001$.

In order to understand in detail the effect of timbre and overall duration of the target on the subjective duration of ramped and damped sounds, matched durations of ramped and damped targets were converted into the difference between the matched duration of the ramped (or damped) target and the matched duration of the steady target as a percentage of target duration with the following formula:

$$DD\% = \frac{d_{vrdt} - d_{vst}}{d_{\phi t}}$$

where d_{vrdt} is the subjective duration in milliseconds of the ramped (or damped) target, d_{vst} is the subjective duration of the steady target and $d_{\phi t}$ is the actual duration of the target in milliseconds. Positive values of $DD\%$ indicate overestimation of the duration of the ramped (or damped) target compared to the matched duration of the steady sound, negative values indicate underestimation of the duration of the ramped (or damped) target compared to the matched duration of the steady target and a null value indicate identical matched duration for ramped (or damped) target and steady target. $DD\%$ provides a measure of the percentage of under/overestimation of the duration of the ramped (or damped) target relative to that of the steady target. This measure controls, for example, for any stimulus-order effects that are constant across envelopes.

$DD\%$ were calculated for each ramped and damped target and separately for each listener (see figure 2). A 4 (timbre of the target) \times 2 (ramped/damped envelope) \times 4 (duration of the target) ANOVA was performed on $DD\%$. The timbre of the carrier affected the subjective duration of the target sounds, $F(3, 30)=7.80, p=.001$, and this result was due to ramped white noises being more underestimated than either ramped sines, ramped complex tones or ramped vowels: $F(3, 30)=10.75, p<.0001$. In fact, the same analysis of variance performed with the exclusion of the results collected with white noise targets showed no difference between sine,

complex tone and vowel carriers ($F < 1$). On average while periodic ramped targets were underestimated of $\sim 10\%$, ramped white noises were underestimated by $\sim 15\text{-}20\%$. On the contrary, all damped targets were similarly underestimated by $\sim 40\text{-}30\%$, whatever their timbre. Moreover, while ramped sounds seemed constantly underestimated across the four duration investigated the underestimation of damped sounds decreased as the overall duration of the sound increased: $F(3, 30) = 9.85, p < .0001$.

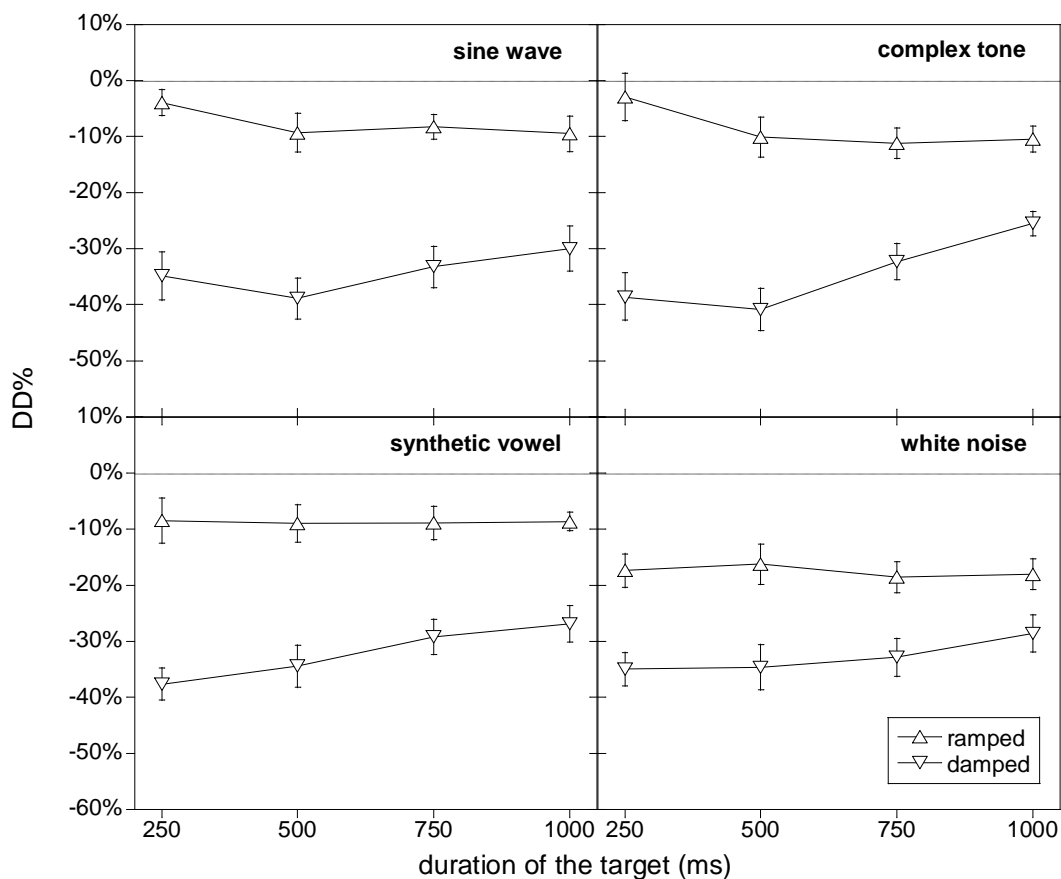


Figure 2: DD% as a function of the duration of the target sound. Upper left graph show results for sine carrier, upper right graph results for complex tone carrier, lower left graph show results for synthetic vowel carrier, lower right graph show results for white noises. The horizontal dashed line correspond to a DD% of zero. Vertical bars represent ± 1 standard error of the mean.

Two separated analysis of variance performed on, respectively, the DD% recorded with ramped sounds or with damped sounds highlighted such a difference. The first analysis reported that the underestimation of ramped sounds was in fact constant across the four durations investigated, $F(3, 30) = 1.45, p > .05$, independently from the timbre of the carrier: $F(9, 90) = 1.62$,

$p > .05$. On the contrary, the underestimation of the duration of damped sounds was large for short sounds, and smaller for long sounds, $F(3, 30) = 6.71, p = .001$.

Discussion

The results collected with the steady targets demonstrated that listeners were extremely accurate in evaluating the duration of such sounds. Despite the great accuracy demonstrated with steady targets both ramped and damped sounds were underestimated in duration. Ramped sounds were moderately underestimated in duration. Such underestimation was not dependent on the overall duration, but it was influenced by timbre. Ramped tones of any duration were underestimated by $\sim 10\%$ and ramped noises of any duration were underestimated by $\sim 15\text{-}20\%$. Damped sounds were substantially underestimated in duration both compared with steady sounds and with ramped sounds. The timbre of the carrier did not affect the perceived duration of damped sounds with damped sounds of any timbre being underestimated by $\sim 40\text{-}30\%$ of their duration. On the contrary, the overall duration of the damped sounds did have an influence on their subjective duration: short damped sounds were underestimated much more than long damped sounds.

In conclusion, when the timbre of the carrier was aperiodic I observed a reduction of the asymmetry in perceived duration between ramped and damped sounds. This reduction was due to a change in the perception of ramped noises rather than a change in the perception of damped noises. Furthermore, the asymmetry in subjective duration between ramped and damped sounds decreased as the overall duration of the sounds increased.

Experiment three

Experiment 2 showed that the durations of ramped sounds were consistently underestimated by a constant proportion. A parsimonious explanation of such a result is a sensory threshold at a fixed level. The modulation function (raised cosine) starts at zero and crosses a fixed threshold at a constant proportion of the total duration. The experimental result could thus be explained by listeners ignoring the initial fraction of the sound that has a sub-threshold amplitude. Although some detection threshold must exist, it is an experimental question whether it accounts for the duration-matching results.

A sensory threshold may also explain part of the underestimation measured for damped sounds. By the same argument, a constant proportion of the end of a damped sound would be at a sub-threshold level. However, in order to fit the results, such a threshold must be different from that hypothesised for the ramped sounds. Firstly, damped sounds were much more underestimated in duration than were ramped sounds, and so would need a higher threshold. But, more importantly, the proportional underestimation of damped sounds *decreased* as the duration of the sound increased. A simple threshold model would have to be modified to take account of the rate of change of level of the damped sound, with a lower threshold for long than for short sounds. Such a modification would be compatible with forward masking, though not with adaptation.

The aim of the current experiment was to test the existence of such sensory thresholds by means of a reaction time experiment. I assume that the reaction time to the onset of a ramped sound will be triggered when the sound becomes detectable: the lower the threshold, the shorter the reaction time. The experimental question is whether the increase in reaction time as a stimulus duration increases will be appropriate to explain the duration underestimation data. A similar argument can be made for a hypothesised fixed threshold for the offset of a damped sound.

Method

Nine paid listeners from the previous experiment participated in the experiment. By excluding the software the apparatus was identical to that of the first two experiments. The software was custom and permitted to gather reaction times with an accuracy of 16.6 ms. Ramped, damped and steady sounds of 250, 500, 750, 1000 ms, identical to the target sounds used in the experiment 2, were used as target stimuli. Two target timbres were used: complex tone and white

noise. No other periodic carrier was used since no difference was found between, sine waves, complex tones and synthetic vowels, in the previous experiment. Peak level for all sounds was 65 dB SPL.

Procedure

The experiment was divided in four blocks. In two blocks of trials listeners reacted to the onset of target tones or noises. In the remaining two blocks listeners reacted to offsets. Listeners took these four blocks in a randomised order, with each block preceded by a short practice session to familiarize listeners with the task and to warm-up listeners' motor reactions.

On each trial, the listener heard an alert sound (30-ms, 2-kHz sine wave) followed after a silent interval of either 1000, 1250, 1500, 1750 or 2000 ms, by a ramped, a damped, or a steady sound. The listeners' task was to press a keyboard key as soon as they heard the target sound. The response terminated the sound. Every combination of silent interval and target was randomly repeated five times during each block giving a total of 300 trials (5 silent intervals \times 3 target envelopes \times 4 target durations \times 5 repetitions). Each onset block lasted approximately 30 minutes. Listeners could choose to repeat a trial if they thought that they had made an anticipation or had unduely delayed their response. Onset reaction times were measured as the temporal difference between the response and the actual onset of the target sound.

The procedure for the offset reaction time was similar. The listeners' task was to press a keyboard key as soon as they heard the target stop and again the response terminated the sound. In order to avoid possible anticipations in the reaction to damped sounds, an extra target sound (fake-damped) was added to the target set. In this sound, a low level steady sound was added to the damped sound in order to produce a sound with an initial decrement in level but followed by a distinctly audible end. The amplitude peak of the fake-damped sound was identical to that of the other targets. In each block listeners reacted to a total of 400 sounds (5 silent intervals \times 3+1 target envelopes \times 4 target durations \times 5 repetitions) presented in random order. Each block lasted approximately 40 minutes. Offset reaction times were measured as the temporal difference between the response and the physical offset of the target sound. Data from the fake-damped targets were not analysed.

Results

Onset

Before the analysis was carried out I excluded outliers from the data set of each listener. Outliers were defined as reaction times smaller or greater than ± 2 standard deviations above (or below) the mean of the listener. With this method I removed $\sim 5\%$ of all observations. The remaining onset reaction times were averaged for each target separately for each listener (figure 3) and subjected to a 2 (timbres) \times 3 (envelopes) \times 4 (durations) ANOVA. Overall average onset response times to tones and noises were identical, $F < 1$, as were those to steady and damped targets: $F < 1$. On the contrary, the reaction to ramped sounds was delayed, $F(2, 16) = 741.23$, $p < .0001$, and such a delay increased as the duration of the ramped sound increased: $F(6, 48) = 85.01$, $p < .0001$.

Offset

Outliers were removed and a similar analysis was made for the offset reaction times. Overall average offset response times to tones and noises were identical, $F < 1$, and the latencies to the offsets of steady and ramped targets were similar $F(1, 8) = 2.77$, $p > .05$. On the contrary, reaction times to the offsets of damped sounds were anticipated, $F(2, 16) = 35.61$, $p < .0001$, but this anticipation did not increase as the overall duration of the damped sound increased: $F(6, 48) = 2.00$, $p > .05$.

Delays and anticipations recorded with onset and offset reaction times were transformed in DD%. For example, average reaction times to a 250 ms steady sound and ramped sound of 232 and 268 ms respectively give $DD\% = \frac{|268 - 232|}{250} = 14\%$ of the duration of the ramped sound. In other words, with a target duration of 250 ms, the reaction to the ramped sound occurred 14% later than the reaction to the steady sound implying that the first 14% of the ramped sound was not heard by the listener. The same calculation was performed for the offset reaction times to damped sounds (see figure 4).

A 2 (timbres) \times 2 (envelopes) \times 4 (durations) ANOVA was performed on such DD%. Overall, listeners ignored similar portions of tone targets and noise targets: $F(1, 8) = 1.23$, $p > .05$. For short durations, the portion below threshold measured on damped sounds was much larger than the portion below threshold measured on ramped sounds, on the contrary for long durations listeners neglected similar portions of the two sounds: $F(3, 24) = 27.11$, $p < .0001$. Overall, the threshold decreased as the overall duration of the sound increased: $F(3, 24) = 74.97$, $p < .0001$. Specifically, the threshold on ramped sounds decreased slightly with increasing overall duration:

$F(3, 24)=5.72, p=.004$, whereas the threshold on damped sounds decreased considerably: $F(3, 24)=55.01, p<.0001$.

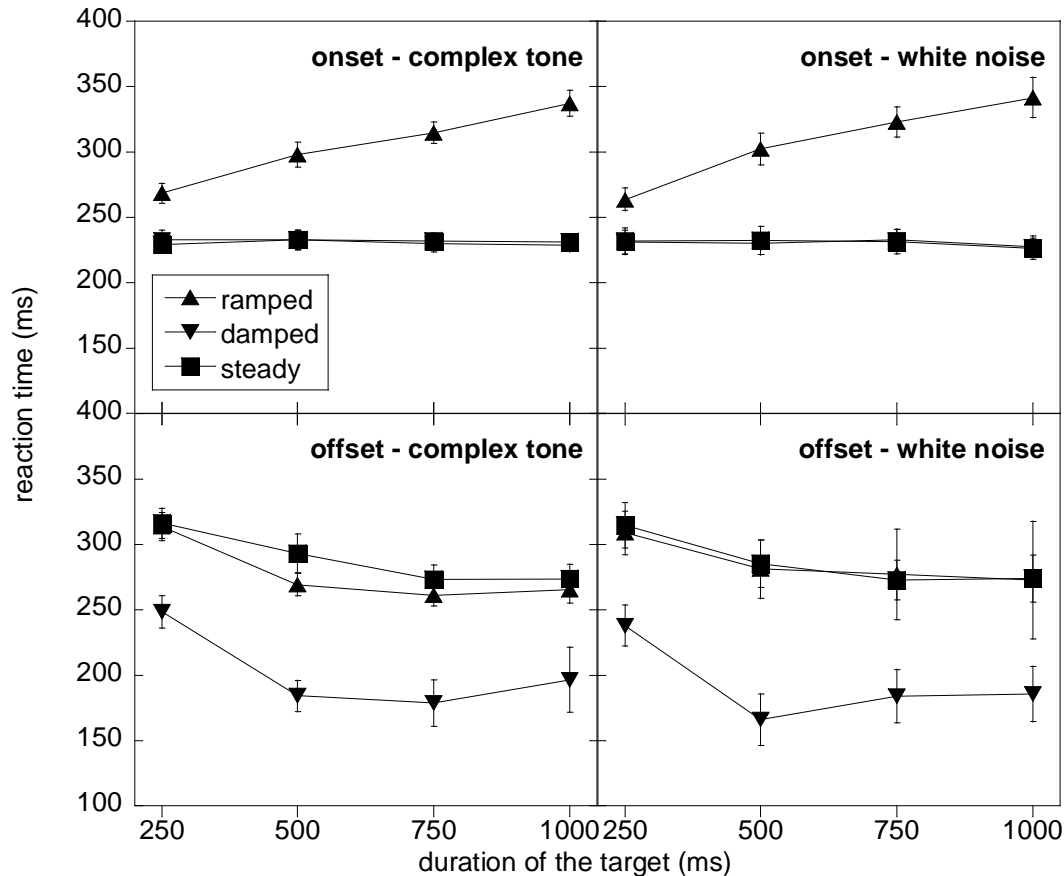


Figure 3: onset and offset reaction times as a function of the duration of the ramp. Upper graphs show results collected with the onset reaction times. Lower graphs show results collected with the offset reaction time. Vertical bars represent ± 1 standard error of the mean.

Discussion

Reaction times to the onsets of ramped sounds were longer than those to steady or damped sounds; reaction times to the offsets of the damped sounds were shorter than to the steady and the ramped sounds. For the ramped sound onsets the reaction time results of experiment 3 converted into DD% show that listeners ignored c. 10-15% of the sound, whatever its timbre, and that this ignored portion decreased slightly with increasing overall duration. For the damped sound offsets listeners ignored c. 25% of short damped sounds, decreasing to c. 10% for long ones.

Qualitatively these results are not surprising since the very beginning of the ramped sounds and the very end of the damped sounds were sub-threshold. The issue though is whether these results can quantitatively explain the matching data from experiments 1 and 2.

Comparison of experiments 2 and 3 and discussion

DD% scores from experiments 2 and 3 were compared. The DD%s measured with ramped tones, ramped, damped tones and damped noises were subjected to four 2 (experimental task) \times 4 (durations) ANOVA. These analyses used only the nine listeners that participated in both experiments, and were confined to the complex tones and white noises. Scores for the ramped sounds were quite similar in the two experiments. They did not differ significantly for the complex tones $F(1, 8)=1.07, p>.05$, and showed only a small difference for the noises: $F(1, 8) = 5.71, p=.044$. By contrast, for both types of damped sound, their underestimation from experiment 2 was much larger than the portion below threshold estimated with reaction time in experiment 3: $F(1, 8)=35.42, p<.0001$ for tones, $F(1, 8)=18.99, p=.002$ for noises.

The ramped-sound results generally support the hypothesis that the small underestimation of their duration is due to a fixed threshold, rendering the first part of the sound inaudible. The data from the complex tones conforms exactly to this hypothesis, but there is a small discrepancy for the noise stimuli. The underestimation of ramped white noises with the matching experiment is 5% larger than the ignored portion measured with the reaction time experiment. The stochastic nature of the noise stimuli could be responsible for this difference, if early isolated randomly large amplitudes were effective at triggering a reaction time but could be ignored in estimating the sound's duration.

Results collected for damped sounds with the two experiments differ substantially. The percentage underestimation is much smaller in the reaction time experiment than it is in the duration estimation experiment. A simple threshold can neither explain this difference, nor the change in both measures with overall duration.

In summary, although a simple threshold model can explain most of the data for ramped sounds it fails to explain both the size of the duration underestimation of damped sounds and its variation with overall duration.

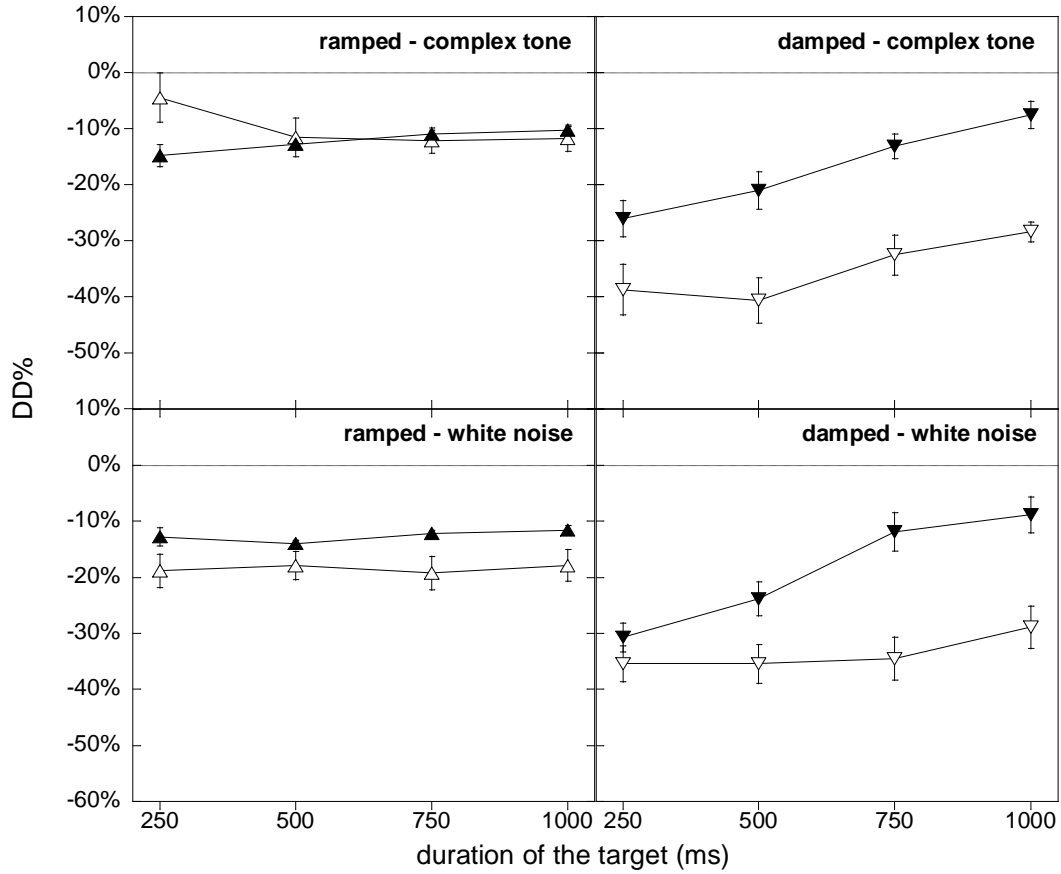


Figure 4: DD% measured with the matching task and the reaction time task in experiment 2 and 3 as a function of the duration of the target sound. Black symbols represent DD% measured with the reaction time and white symbols represent DD% measured with the matching task. In all graphs, the horizontal dashed line corresponds to DD% of zero. Vertical bars represent ± 1 standard error of the mean.

Experiment four

The results of experiment 3 suggested that the underestimation of ramped sounds could be well explained by a sensory threshold at a fixed amplitude, which occurred later for shallower ramps causing a proportional underestimation of duration. However, a fixed sensory threshold could explain only a small part of the larger underestimation of damped sounds.

The current experiment compares subjective durations and sensory thresholds of ramped and damped sounds modulated either with the raised cosine function or with an exponential function that gains ~ 43 dB over its time course. For the first $\sim 5\%$ of its time course, the rising exponential function is greater than the raised cosine function, but is lower for the remainder (see figure 5). As the fixed threshold acting on ramped sounds raised cosine modulated (CR) occurred after the 10% of their duration (~ 33 dB SPL) the sensory onset of ramped sounds exponentially modulated (ER) should be triggered later (see figure 5). As a consequence, I expect larger underestimation for ER compared to CR as listeners will perceive later the onset of ER sounds. By the same token I expect anticipated sensory offset with damped sounds exponentially modulated (ED) than with damped sounds raised cosine modulated (CD). However, as the sub-threshold portion explains only partially the large underestimation of damped sounds, I cannot predict whether an anticipated perception of the offset of damped sounds will correspond to a larger underestimation of their duration.

Method

Listeners

Ten paid listeners, recruited from the University population, participated in the experiment. None of them reported a hearing loss or other difficulties with their hearing. All listeners were naïve and none of them had participated in the previous experiments.

Stimuli

Target sounds were steady, CR, CD, ER, ED of four durations, 250, 500, 750, 1000 ms. The steady, the CR and the CD were identical to those used in experiment 2 and 3. The ERs were obtained amplitude modulating a steady carrier with the following function:

$$a(t) = \exp\left(5\left(\frac{t}{T} - 1\right)\right)$$

were $a(t)$ is the amplitude of the modulator at time t , and T is the overall duration of the sound ($T = 250, 500, 750, 1000$). This function decreases by $\sim 43\text{dB}$ across the sound's duration. The final 15 ms was given an additional raised cosine ramp to bring the overall amplitude smoothly to zero. ED targets were obtained by reversing in time the ER targets. The timbres of target sounds were a white noise and a complex tone (f_0 500 Hz) with the first ten harmonics at the same amplitude.

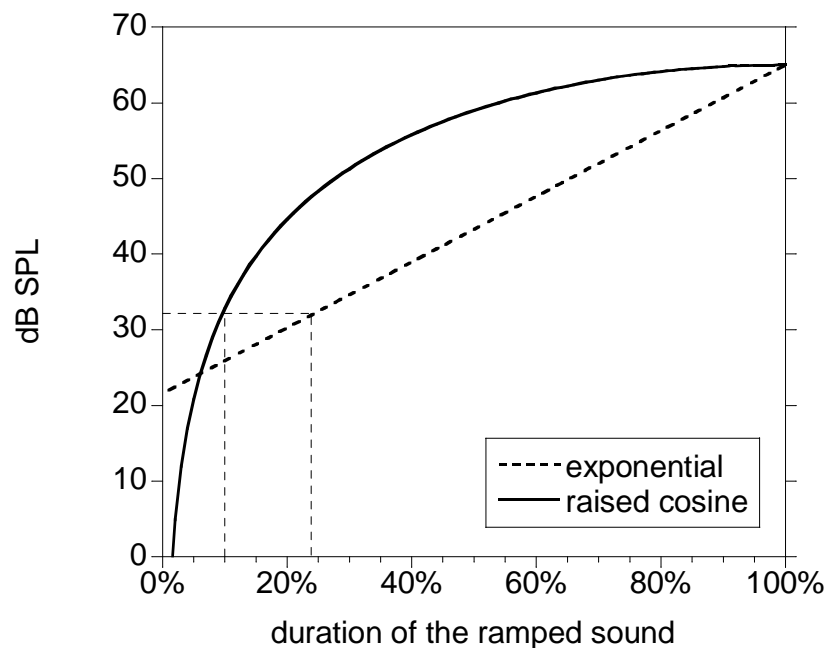


Figure 5: sound pressure level of the ramped sound as a function of the duration of the ramp expressed in percentages. The horizontal dashed line shows the approximate location of the sensory threshold measured experiment 3.

Apparatus and procedure

The apparatus was identical to that used in the first three experiments. The experiment consisted of two tasks, a matching task and a reaction time task, similar to those described in experiment 2 and 3. Listeners performed the two tasks consecutively and the order was balanced across listeners. The two tasks were performed within a maximum of one week one another. Peak level for all sounds used in the experiment was 65 dB SPL.

Matching task

The matching task was divided into eight blocks within which target duration and timbre were constant. Listeners took blocks of identical target duration consecutively with the order of the two timbres balanced across listeners. The order of the four block pairs was randomised across listeners. Each block consisted of fifty matches, ten for each target sound, performed with ER, ED, CR, CD and steady targets.

Reaction time task

The procedure was identical to that used in experiment 3. Listeners reacted separately to the onset of tones, the onset of noises, the offset of tones and the offset of noises. The procedure for onset and offset response times was identical to experiment 3. However, in each block the number of scored targets that listeners had to react to was five-hundred (5 silent intervals \times 5 target envelopes \times 4 target durations \times 5 repetitions). As in experiment 3, in each offset block the target set was extended with fake-CD and fake-ED targets. Fake damped targets were obtained by adding a low-level steady tone (or noise) either to ED or CD targets. Therefore, listeners reacted to the offset of a total of seven-hundred sounds (5 silent intervals \times 5+2 target envelopes \times 4 target durations \times 5 repetitions). The reactions to fake damped targets were excluded from the analysis.

Results

Matching task

The average duration of the adjustable sound was first calculated for steady targets separately for each listener and converted into D%. A 2 (timbres) \times 4 (durations) ANOVA was performed on these D%s. As in experiment 2, matches did not differ either with carrier timbre, $F(1, 9)=1.01, p>.05$, the duration of the sound, $F<1$ or the two factors interacting, $F<1$ and were not different from veridical $t(9)=-.61, p>.05$.

	Mean				Standard error of the mean			
	250	500	750	1000	250	500	750	1000
sine	-1.92%	-1.79%	-1.17%	-2.63%	1.93%	2.08%	2.79%	3.29%
white noise	0.93%	-0.04%	-1.59%	-0.69%	2.01%	2.00%	1.56%	2.40%

Table 2: D% measured for steady targets as a function of the overall duration of the steady sound. The left part of the table shows means D% and the right part of the table shows standard error of the mean.

A second 2 (timbres) \times 5 (targets) \times 4 (durations) ANOVA was performed with the D% measured for all target sounds. Results showed that, overall, different target envelopes were differently estimated in duration: $F(4, 36)=82.49, p<.0001$. Four contrasts showed that all amplitude

modulated targets were underestimated in duration compared to steady targets: ER, $F(1, 9)=13.76$, $p=.005$; CR, $F(1, 9)=23.72$, $p=.001$; ED, $F(1, 9)=148.68$, $p<.0001$; CD, $F(1, 9)=113.69$, $p<.0001$.

Results collected with ramped and damped targets were converted in DD% and a new 2 (timbres) \times 2 (ramped/damped targets) \times 2 (modulating function) \times 4 (durations) ANOVA was performed on such data (see figure 6). Noise targets were more underestimated than tone targets: $F(1, 9)=31.71$, $p=.0003$ and such a result was due to ramped noises being much more underestimated than ramped tones while damped noises were slightly more underestimated than damped tones: $F(1, 9)=7.26$, $p=.02$. Damped sounds were substantially underestimated compared to ramped sounds: $F(1, 9)=150.49$, $p<.0001$. The underestimation was generally greater when sounds were modulated with the exponential function than when sounds were modulated with the raised cosine function: $F(1, 9)=20.96$, $p=.001$. Such a decrement was evident for ED tones and noises and ER noises but not for ER tones: $F(1, 9)=6.18$, $p=.03$.

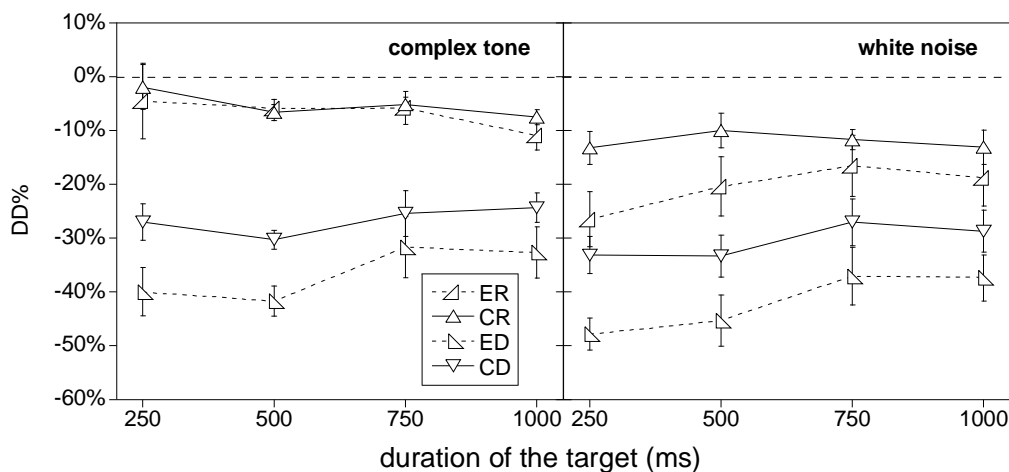


Figure 6: DD% as a function of the duration of the sound. On the left graph results collected with complex tones on the right results collected with white noises. Vertical bars represent ± 1 standard error of the mean.

Four separate 2 (modulating function) \times 4 (durations) ANOVAs were performed on DD%_s measured with ramped tones and noises and damped tones and noises in order to understand the role of duration and modulating function on their underestimation. Overall, the underestimation of ramped tones was constant across all durations investigated, $F<1$, and constant despite the function used for modulating the amplitude of the sound, $F<1$. The underestimation of CR noises was constant across durations, $F<1$. In particular, the underestimation was constant also for ER noises although there was a slight effect of duration: $F(3, 27)=2.65$, $p=.06$.

The underestimation of ED sounds decreased as the overall duration of the sound increased: $F(3, 27)=3.91, p=.01$ for tones, and $F(3, 27)=2.94, p=.05$ for noises. On the contrary, the underestimation of CD sounds was constant across the four durations, $F(3, 27)=1.75, p>.05$ for tones and $F(3, 27)=1.47, p>.05$ for noises.

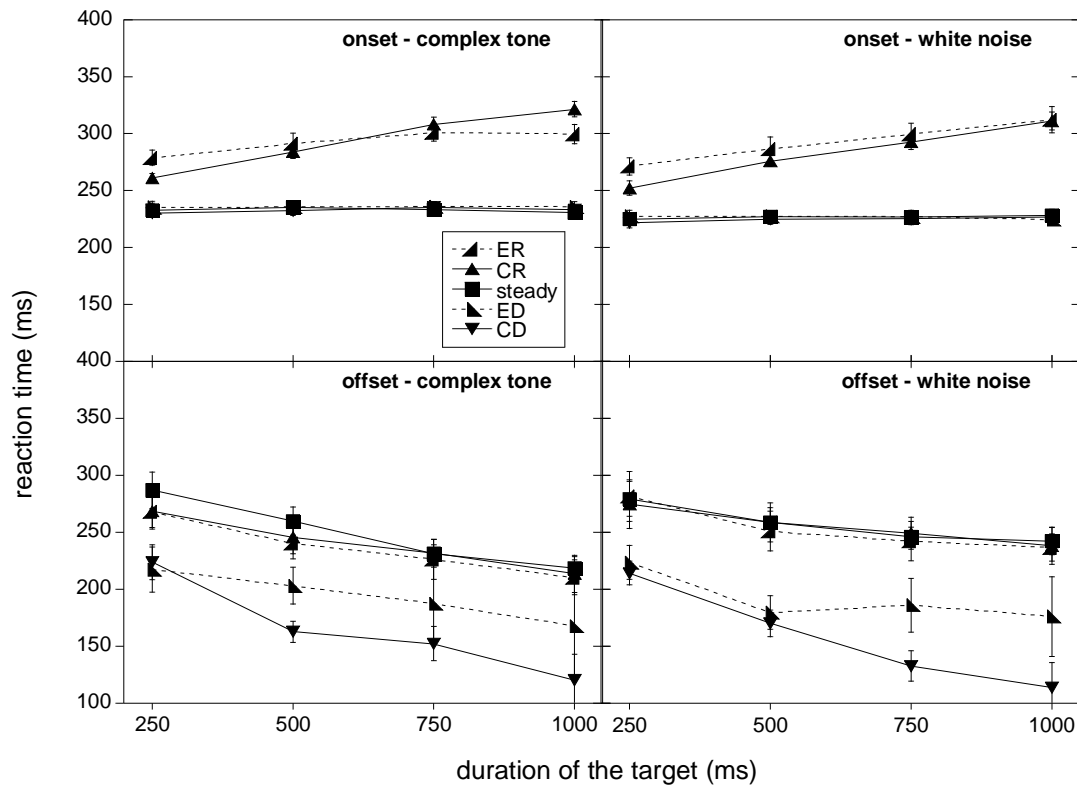


Figure 7: reaction time as a function of the duration of the target sound. Upper panels show results collected with the onset reaction time, lower panels show results collected with offset reaction time. Vertical bars represent the standard error of the mean.

Reaction time task

Before the analysis was carried out I excluded outliers with the same method used in experiment 3. Onset reaction times were collected and averaged for each singular target and separately for each listener. Resulting values were analysed by means of a 2 (timbres) \times 5 (targets) \times 4 (duration) ANOVA (figure 7). Onset response times to tones were on average about 7 ms longer than to noises $F(1, 9)=6.03, p=.03$. A contrast between reactions to steady and damped targets showed that listeners reacted with identical latencies to ED and CD: $F<1$ and $F(1, 9)=1.30, p>.05$ respectively. On the contrary, the reaction time to the onset of all ramped sounds was delayed, $F(4, 36)=193.32, p<.0001$, by an increasing amount as the duration of the ramped sound

increased: $F(12, 108)=42.46, p<.0001$. However, a new 2 (timbres) \times 2 (modulating function) \times 4 (duration) ANOVA performed only on reaction times to ramped targets revealed that latencies to ER increased more slowly as a function of duration than did latencies to CR ones: $F(3, 27)=19.48, p<.0001$.

Outliers were removed and offset reaction times were separately averaged for each target listener. On these data a 2 (timbres) \times 5 (envelopes) \times 4 (durations) ANOVA was performed (figure 7). Offset response times to tones and noises were identical, $F<1$. A contrast between response times to steady and ramped targets demonstrated that listeners reacted with similar latencies to the offsets of steady and ramped targets both when exponentially modulated, $F(1, 9)=1.35, p>.05$ and when raised-cosine modulated $F<1$. On the contrary, the reaction time to the offsets of damped sounds was anticipated, $F(4, 26)=18.26, p<.0001$, but the anticipation did not increase as the overall duration of the damped sound increased: $F(12, 108)=1.39, p>.05$. Furthermore, a contrast between the reaction times to ED and to CD showed that reactions to the first were faster: $F(1, 9)=7.04, p=.02$.

Onset reaction times to ramped targets and offset reaction times to damped targets were converted in DD% separately for each timbre, amplitude modulator, overall duration of the target and each listener. On these values a 2 (timbres) \times 2 (ramped or damped targets) \times 2 (modulating function) \times 4 (durations) ANOVA was conducted (see figure 8). The timbre of the target sound had no effect on DD% ($F<1$). The portion below threshold was larger for damped sounds than for ramped sound, $F(1, 9)=5.89, p=.03$, and such a difference was even larger for short sounds while smaller (or null) for long sounds: $F(3, 27)=5.39, p=.005$. The overall duration of the sounds affected the threshold with short sounds having larger ignored portions than long sounds: $F(3, 27)=29.65, p<.0001$. Finally, the modulating function affected the threshold in a different way for ramped and damped sounds. When reacting to ER listeners ignored larger portions of the sound than when reacting to CR. However, this outcome was reversed with damped sounds. When reacting to ED listeners ignored smaller portions of the sound than when reacting to CD: $F(1, 9)=25.33, p=.001$.

In order to understand in detail the relationship between the position of the threshold on the one hand and the overall duration of the sound and the modulating function on the other, two new 2 (modulating function) \times 4 (durations) ANOVAs were performed. Such analyses were run separately for ramped sounds and damped sounds averaging listeners' results collected with tones and noises. Listeners ignored larger portions of ER sounds than of CR ramped sounds: $F(1, 9)=13.86, p=.005$ and such a difference was particularly evident at short sound duration: $F(3,$

27)=42.53, $p<.0001$. Listeners ignored constant portions of CR ramped sounds, independently from the duration of the sound: $F(3, 27)=2.60$, $p=.07$, while the ignored portion of ER sounds decreased as the overall duration of the sound increased: $F(3, 27)=37.82$, $p<.0001$. When reacting to damped sounds listeners ignored smaller portions of the tail of ED sound rather than CD sound: $F(1, 9)=8.82$, $p=.01$ and such a portion was large for short sounds and smaller for long sounds: $F(3, 27)=16.35$, $p<.0001$. However, the difference in DD% between the two modulators decreased parallelly across the four durations investigated: $F<1$.

A final analysis was conducted in order to compare the sensory thresholds measured with ramped and damped sounds separately for the two modulating functions. Two 2 (timbres) \times 2 (ramped or damped targets) \times 4 (durations) ANOVAs were performed on DD%. The ignored portions measured with ER and ED sounds were overall identical $F<1$ and identical across the four durations investigated $F(3, 27)=2.13$, $p>.05$. On the contrary, listeners ignored larger portions of CR sounds than CD sounds: $F(1, 9)=33.85$, $p=.0003$. In detail, while listeners ignored $\sim 10\%$ of CR sounds across the four durations investigated the ignored portions of CD decreased as the overall duration of the sound increased passing from $\sim 25\%$ for short durations to $\sim 10\%$ for long durations: $F(3, 27)=7.72$, $p=.0007$.

Comparison of the results of the two procedures.

The DD% measured with the matching and the reaction time task were compared. Such a comparison was performed by means of eight 2 (task) \times 4 (durations) ANOVAs performed separately for each target sound. The ignored portion measured with the reaction time could explain all the underestimation of ER tones in the matching task, $F(1, 9)=1.84$, $p>.05$, while it explained the largest part (but not all) of the underestimation of the duration of CR tones in the same task: $F(1, 9)=6.67$, $p<.02$. Furthermore, the results of the reaction time could explain also both the underestimations measured with the matching for both ER and CR noises: respectively, $F(1, 9)=2.31$, $p>.05$, and $F<1$. On the contrary, the results obtained with the reaction time could not fully explain any of the large underestimations measured for damped sounds with the matching. Much different were the underestimations of either damped tones or noises exponentially modulated with the two tasks: respectively, $F(1, 9)=41.04$, $p<.0001$, and $F(1, 9)=62.99$, $p<.0001$. Although smaller, the same difference was evident for CD: $F(1, 9)=11.29$, $p=.008$ for tones, and $F(1, 9)=14.34$, $p=.004$ for noises.

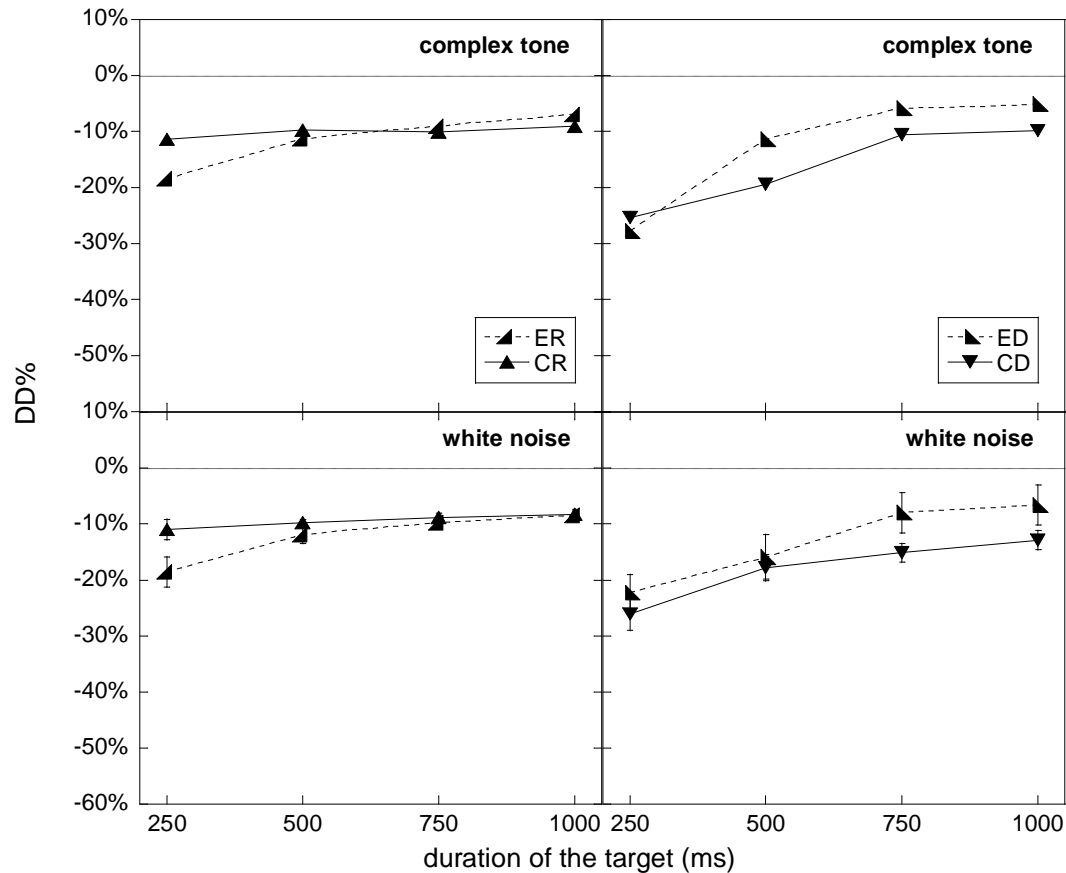


Figure 8: DD% as a function of the duration of the target sound. For sake of clarity data are plotted in four graphs. Left graphs show results collected with ramped sounds, right graphs show results collected with damped sounds. Vertical bars represent ± 1 standard error of the mean.

Discussion

I predicted that the exponential function would produce a forward shift in the position of the sensory onset of ramped sounds with a corresponding (i) reduction of their subjective duration; (ii) resulting from a further delay in their onset response. The results of the matching experiment partially confirmed such a prediction: exponentially modulated ramped noises were more underestimated in duration than raised cosine modulated ramped noises. However, all ramped tones were similarly underestimated in duration independently from the modulating function. In the reaction time task I found that the exponential modulator induced a longer delay in the perception of the onset of ramped sounds than did the raised cosine modulator.

Overall, the results collected with ramped sounds demonstrated that the small underestimation of their duration is mainly dependent on a threshold acting on their onset: the beginning of the sounds is too quiet to be perceived and for this reason listeners do not include such a portion in the computation of the subjective duration of the sound. According to my data such a threshold is either fixed or dependent on the rate of change in level of the sound: higher for a high rate of change in dB/s. A rate of change dependent threshold was particularly evident with ER sounds.

According to my predictions, an anticipation in the sensory offset of damped sounds should correspond to an anticipated perception of their offset and, likely, in a decrease in their subjective duration. Behavioural results showed larger underestimations for exponentially-modulated damped sounds than for raised-cosine-modulated damped sounds. However, listeners perceived the offset of exponentially modulated damped sounds later than the offset of raised cosine modulated damped sounds. Of course one explanation for such inconsistency may lie in the fact that the two functions behave quite differently across time. Therefore, it may be possible that the modulating function played a larger role than shifting the position of the offset mark backwards as I intended. In addition, while in experiment 2 the underestimation of damped raised cosine sounds was in the range of 30-40% of their duration the underestimation for the same sounds in the current experiment was much smaller and was in the range of 20-30% of their duration. In the current experiment only the damped sounds exponentially modulated were underestimated in a range between 30-40% of their duration.

Overall, damped sounds are largely underestimated in duration and their underestimation cannot be explained with models based on sensory thresholds as deduced from the reaction time experiments. In particular, the results collected with CR and CD sounds show that these two sounds are perceived as having different durations. The corresponding sensory thresholds are also different, however, such a difference is not large enough for explaining the difference in subjective duration. Results collected with ER and ED sounds are even more paradoxical. On average ER sounds are perceived as being 40% longer than ED sounds. However, the sensory thresholds measured with the two sounds are identical.

Experiment five

The results of Experiments 1, 2, 3 and 4 showed (I) that listeners slightly underestimated the duration of ramped sounds, (II) that such underestimation originate at the beginning of the ramped sounds and (III) that such underestimation can be explained with a sensory threshold acting in the region of their onset. In the same experiments I found that listeners (I) largely underestimated the duration of damped sounds, (II) that such underestimation originate at the end damped sounds, however, (III) such underestimation can be explained only partially with a sensory threshold acting in the region of their offset. The lack of sensory ground for explaining the asymmetry in subjective duration between ramped and damped sounds may be due to the results obtained with the reaction time task. Such a task could be inaccurate for measuring the sensory thresholds acting at the onset of ramped sounds and at the offset of damped sounds.

The aim of the current experiment was measure the sensory thresholds acting at the onset of ramped sounds and the offset of damped sounds with a psychoacoustical procedure more robust than the reaction time: an adaptive procedure. In the current experiment listeners performed such a task together with the two previous tasks. The idea at the basis of the adaptive procedure was the following. Let consider a damped sound like those used in the previous experiments. The ending part of this sound is too quiet to be perceived, therefore, a second damped sound whose envelope is truncated when the sound is no longer perceived should posses the same subjective duration of the first. *Mutatis mutandis*, let consider a ramped sound, according to the data, its beginning is too quiet to be perceived. As a consequence a second identical ramped sound starting at threshold level should posses the same subjective duration of the first. In the adaptive task listeners heard two damped (or ramped) sounds in succession, one shorter than the other. Their task was to indicate which of the two sounds was the shortest in duration. If the answer was correct the duration of the shortest sound was increased, and, consequently the difference between the two sounds was reduced. This process was iterated until listeners perceived the two sounds as having a similar durations.

In conclusion, the aim of the current experiment was to test whether the results collected with the adaptive procedure could provide any sensory ground for (I) the short subjective duration of damped sounds and, consequently, (II) the large asymmetry in subjective duration between ramped and damped sounds. In the current experiment the three tasks were performed only with

ramped and damped sounds exponentially modulated as with such sounds the difference in the results collected with the matching and the reaction time was maximal. Furthermore, as the time required for testing each listener was long (six hours), the three tasks were run only with tones.

Method

Listeners

Twelve paid listeners, recruited from the university student population, were tested individually. They had normal audiometric thresholds and none of them participated in the previous experiments.

Stimuli

Stimuli were ramped, damped and steady sounds of four durations: 250, 500, 750, 1000 ms. The modulating function of ramped sounds was the following:

$$a(t) = \exp\left[k\left(\frac{t}{T} - 1\right)\right]$$

where $a(t)$ is the amplitude of the waveform as a function of the time t , T is the overall duration of the sound ($T = 250, 500, 750, 1000$), k is the value that determines how rapidly the function increases in dB. For ramped sounds of the matching task, ramped sounds of the reaction time task, and standard ramped tone of the adaptive procedure k assumed the value of 5.18. Therefore, for any overall duration T , and for $t=0$, the level of the sound is -45 dB from the peak level. As discussed in the previous experiments the absolute change in dB of ramped sounds is independent from its overall duration but the rate which ramped sounds change in dB per time unit is different: 180 dB/s, 90 dB/s, 60 dB/s and 45 dB/s, respectively, for the four durations investigated.

With the adaptive task, my goal was to present to the listeners pairs of ramped sounds (one fixed in duration and one variable) whose rate of change was identical but whose overall duration was different. In order to obtain such result the value of k for the variable ramped sound had to be calculated trial by trial for any given overall duration d (with $d < T$) of the variable ramped sound. Consequently, given the four rates of change in dB/s of the ramped sound:

$$dB(d) = dB/s\left(\frac{d}{1000}\right)$$

where $dB(d)$ represents the decrement (in dB) in the given duration d according to the rate of change in dB/s, we can calculate the value of k for any duration d by keeping constant the rate of change in dB/s as follow:

$$k = -\ln\left(10^{\frac{-dB(d)}{20}}\right)$$

Damped targets were obtained reversing in time ramped targets. The beginning and the end of all sounds used in the experiment were gated on and off with two 10 ms raised cosine ramps. The timbre of all sounds used in the experiment was a complex tone with the first ten harmonics at the same amplitude ($f_0 = 500$ Hz). The tasks were performed only with tones to save listeners from a 12 hours long experiment.

Apparatus and procedure

All sounds were synthesised in real time at 44.1 kHz and 16 bits resolution using custom Matlab functions running on a Macintosh G4 that also controlled the experiment. The output of the sound card was passed through Tucker-Davis attenuators and presented through Sennheiser HD414 headphones in a double-skin sound-attenuated IAC booth. The experiment consisted in three tasks, a matching task, a reaction time task and an adaptive task. The order which listeners performed the tasks was balanced across listeners. The experiment was performed in six sessions of one hour each and all listeners completed the experiment within ten days. Peak level for all sound used in the experiment was 65 dB SPL.

Matching task

This task was divided in four blocks. Each block consisted in matches performed with one target duration. The four blocks were performed by listeners in a random order. Within each block listeners performed the same task of the second experiment, with a total of thirty matches for each block: ten with ramped sounds, ten with damped sounds, ten with steady sounds.

Reaction time task

Task was divided into two experimental blocks. Listeners performed one block reacting to the onset of the target sounds, one reacting to the offset of the target sounds. In the onset block, during each trial listeners heard an alert sound (30-ms, 2-kHz sine wave) followed after a silent interval of either 1000, 1250, 1500, 1750 or 2000 ms, by the target sound. Their task was to react as soon as they perceived the beginning of the target sound. Each combination target-silent interval was presented five times during each block in a random order, therefore, listeners reacted to a total

of three-hundred trials. Trials used for measuring the offset reaction times were identical, however, the trial set was extended with one-hundred new trials created with fake damped targets. Fake damped targets were obtained adding a low level steady tone (or noise) to damped targets. In the offset block listeners' task was to react as soon as they perceived the end of the target sound. Listeners took the two blocks in random order and this order was balanced across listeners. Results collected with fake-damped sounds were excluded from the statistical analysis.

Adaptive task

The adaptive task was divided into four blocks. In each block listeners run the adaptive procedure on the same target duration. Listeners took the four blocks in random order. Within each block, listeners performed the adaptive procedure with ramped, then with damped and finally with steady sound. Such order was balanced across listeners. Thresholds for sensitivity in difference of duration between target sound and variable sound were measured ten times for each target. A two interval forced choice, 2 down 1 up adaptive procedure, estimating the 70.7% of the psychometric function was used (Levitt, 1971). This procedure consists in trials articulated in a series of runs. During each run, listeners heard a pair of ramped sounds (or damped or steady), respectively, a target and a variable, separated by a silent interval of 500 ms. The order of this sequence was randomised run by run. After listening to the pair, listener's task was to indicate which of the two sounds was shortest in duration. The answer was considered positive if the listener responded correctly to two runs in a row. The answer was considered negative if the answer to a single run was wrong. When the answer was positive, in the successive run, the difference in duration between target and variable was reduced by increasing the duration of the variable sound by a certain amount (step). When the answer was negative the difference in duration between target and variable was increased by increasing the duration of the variable sound by the amount step. When the listener performed a certain number of reversals (positive run followed by a negative run or vice versa) the step size was reduced. Overall, the procedure used three step sizes (respectively, 1/10, 1/20, 1/50 of the duration of the target sound). The number of reversals in order to change step size were, respectively, 2, 2 and 8. Listeners received feedback after the response to each pair. At the beginning of each trial the duration of the variable sound was the half of the duration of the target. This value was chosen on the basis of a pilot experiment. The silent interval between the variable and the target sound was varied in duration within a range of 500 ± 25 ms. This random variation was introduced to prevent listeners judging the sound pair on the basis of the rhythm of

the pair. The threshold was calculated as arithmetic mean of the duration of the variable sound at the final eight reversals.

Results

Matching task

The average duration of the adjustable sound (in ms) was calculated for each target sound and separately for each listener. Matched durations measured with all targets were converted into D%. A first analysis was performed only on D% measured for steady targets (see table 3). Listeners slightly overestimated the duration of short steady targets and underestimated the duration of long steady targets: $F(3, 33)=4.50$, $p=.009$. On average, listeners overestimated the duration of steady targets ($\sim 3.5\%$) and such a overestimation was just slightly higher than zero (veridical match): $t(11)=2.70$, $p=.02$.

Mean (D%)				Standard error of the mean			
250	500	750	1000	250	500	750	1000
8.49%	6.20%	1.65%	-1.63%	2.76%	2.89%	1.51%	1.68%

Table 3: average D% measured for steady sounds as a function of the overall duration of the sound (left half of the table). The right half of the table show the correspondent standard errors of the mean.

Successively D% collected with all three targets were compared. On average listeners perceived the three target sounds as having different duration according to their envelopes: $F(2, 22)=31.65$, $p<.0001$. Two contrasts highlighted that ramped sounds were perceived as slightly shorter than steady sounds and damped sounds were perceived much shorter than steady sounds: respectively, $F(1, 11)=5.38$, $p=.04$ and $F(1, 11)=41.24$, $p<.0001$. A further contrast showed that damped sounds were perceived as much shorter than ramped sounds: $F(1, 11)=32.82$, $p<.0001$.

Results collected with ramped and damped sounds were successively converted in DD% in order to understand in detail the role of duration on the underestimation of ramped and damped sounds. Two separate one way ANOVA with the duration of the sound as a factor were performed on DD% measured for ramped and damped sounds. Both the underestimation of ramped sounds and the underestimation of damped sounds were constant across the four duration investigated: respectively, $F(3, 33)=2.86$, $p>.05$ and $F<1$.

Reaction time task

Reaction times to each singular target were calculated separately for the onset reaction times and the offset reaction times and for each listener (see figure 9). In the onset reaction time, listeners reacted promptly to the onset of steady and damped sounds, while the reaction to ramped

sounds was delayed, $F(2, 22)=111.58, p<.0001$, and such delay increased as the overall duration of the sound increased: $F(6, 66)=19.59, p<.0001$. In addition, a contrast revealed that listeners reacted with identical latencies to the onset of steady and damped sounds: $F(1, 11)=3.48, p>.05$.

In the offset reaction time, listeners' reaction was fastest with damped sounds while, in comparison, the reaction to the offset of steady and ramped sounds was slower: $F(2, 22)=10.08, p=.008$. Furthermore, a contrast revealed that listeners latencies to the offset of steady and ramped sounds were identical: $F(1, 11)=1.46, p>.05$. With all sounds the reaction decreased as the overall duration of the sound increased: $F(3, 33)=26.08, p<.0001$. Such effect was more evident for damped tones than for steady and ramped targets: $F(6, 66)=4.48, p=.001$.

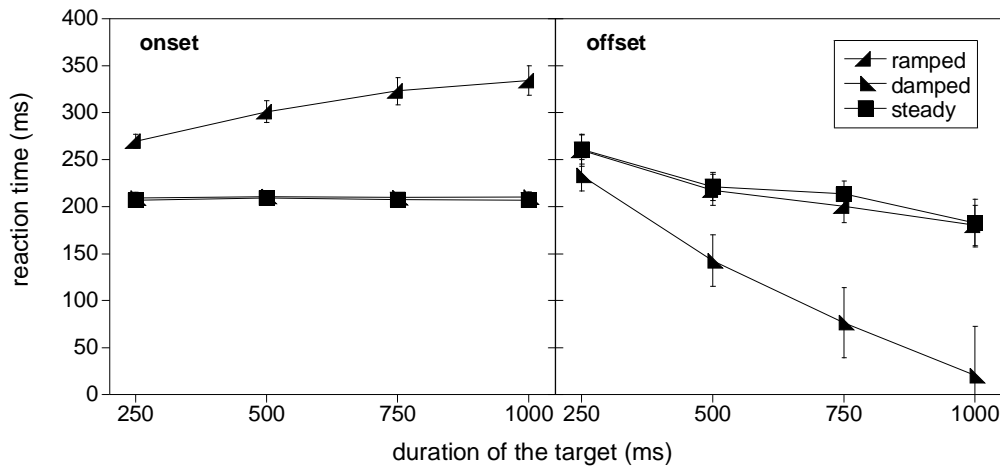


Figure 9: reaction time to the onset (left) and the offset (right) of the target sounds as a function of the overall duration of the target. Vertical bars show ± 1 standard error of the mean.

The reaction times to the onset of ramped sounds and offset of damped sounds were transformed into DD% (see figure 10). A 2 (envelopes) \times 4 (durations) ANOVA was performed on these data. The magnitude of the portions below threshold measured with ramped and damped sounds were identical, $F<1$. Two new analysis of variance revealed that (i) the portion below threshold measured with ramped sounds decreased as the overall duration of the sound increased: $F(3, 33)=38.13, p<.0001$. On the contrary, the second analysis revealed that listeners ignored a fixed portion of damped sounds ($\sim 15\%$): $F<1$.

Adaptive procedure

Thresholds were calculated for each target and each listener. These thresholds are reported in table 4.

	Mean (ms)				Standard error of the mean (ms)			
	250	500	750	1000	250	500	750	1000
Ramped	174.2	380.0	578.2	809.5	8.0	13.9	20.9	23.8
Damped	177.3	384.8	606.2	827.6	8.9	15.1	21.8	25.9
Steady	214.9	420.3	659.4	878.1	3.7	10.9	13.8	22.7

Table 4: thresholds for durations of the target sounds as a function of the duration of the target.

A 3 (envelopes) \times 4 (durations) ANOVA was performed on such thresholds. Thresholds were dependent on the envelope of the sound: $F(2, 22)=20.52$, $p<.0001$. Listeners had low thresholds for the discrimination of duration of steady sounds and higher thresholds for both ramped and damped sounds.

Thresholds measured with ramped and damped sounds were converted into DD% (see figure 10). On such DD% a 2 (envelopes) \times 4 (durations) ANOVA was performed. The thresholds calculated with ramped and damped sounds were similar, $F(1, 11)=2.45$, $p>.05$, and they both decreased as the duration of the sound increased: $F(3, 33)=7.76$, $p=.0005$.

Comparison of the results of the three procedures

The DD% measured with the three procedures were compared by means of a 3 (procedure) \times 2 (envelopes) \times 4 (durations) ANOVA. Behavioural data were dependent on the experimental task, $F(2, 22)=8.82$, $p=.001$. Furthermore, while the estimates collected with ramped and damped sounds were similar with the reaction time and similar with the adaptive procedure they were much different with the matching task: $F(2, 22)=18.98$, $p<.0001$.

Successively the statistical analysis was conducted separately for ramped and damped sounds in order to understand whether the underestimations measured with matching task could be explained by the results of the reaction time or the adaptive procedure. A first contrast revealed that the underestimation measured for ramped sounds in the matching task was comparable to the portion below threshold measured with the adaptive procedure, $F<1$, while the underestimation measured with the reaction time was larger: $F(1, 11)=10.71$, $p=.007$. The underestimation measured for damped sounds with the matching task could be explained neither with the results of the adaptive task, $F(1, 11)=26.92$, $p<.0001$ nor with the results of the reaction time task, $F(1, 11)=12.38$, $p=.005$.

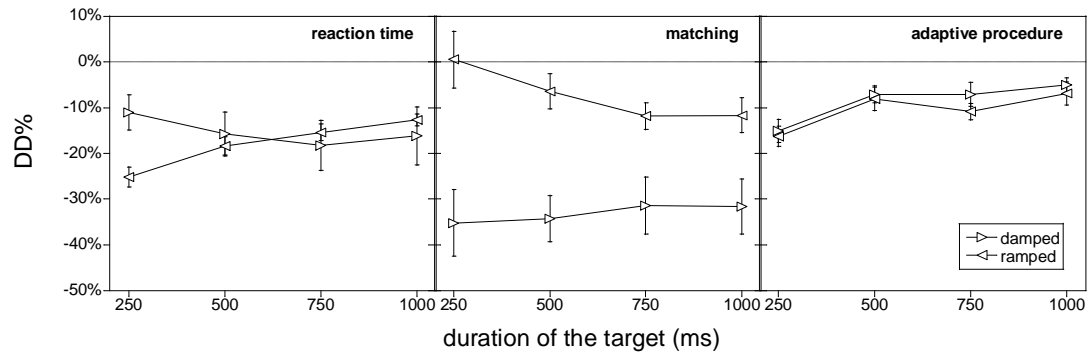


Figure 10: DD% obtained with the three tasks as a function of the duration of the target sound. On the left panel the results collected with the reaction time task, on the centre panel the result collected with the matching task and on the right panel the results collected with the adaptive procedure. Vertical bars represent ± 1 standard error of the mean.

Discussion

The results of the current experiment corroborated the findings of the previous experiments. In the matching task I measured a great asymmetry in subjective duration between ramped and damped sounds. Such an asymmetry was the result of ramped sounds being slightly underestimated in duration and damped sounds being largely underestimated in duration. The results of the reaction time task showed that the underestimation of ramped sounds could be explained by listeners ignoring the beginning of the ramped sounds. In the same task listeners were not able to perceive the end of damped sounds. However, the magnitude of the portions below threshold were identical for ramped and damped sounds.

The results of the adaptive task corroborated the findings of the reaction time task and showed that sensory thresholds are active in the region of the onset of ramped sounds and in the region of the offset of damped sounds, therefore, it demonstrated that the beginning of ramped sounds and the end of damped sounds are not perceived and not included in the computation of the subjective duration of ramped and damped sounds. Also with the adaptive procedure the magnitude of the portions below threshold were identical for ramped and damped sounds.

In conclusion, both with the reaction time and the adaptive procedure I failed to find a sensory explanation for the difference in subjective duration of ramped and damped sounds measured in the matching experiment.

General discussion

The results of current experiments demonstrate that listeners are able to estimate veridically the duration of a sound steady in level. Such a result was quite consistent across all matching experiments performed. Furthermore, results highlights that, with the steady sounds, the Weber ratio holds across all the durations investigated. The veridicity of listeners' matches with steady sounds was also independent from the timbre of the carrier. However, in one condition of experiment 1 the duration of the steady sounds was slightly underestimated. In that condition listeners matched the duration of a steady sound that succeeded in time the adjustable sound that listeners had to manipulate.

The accuracy of listeners in measuring the duration of short sounds is well known. Stevens (1957) gave a value close to the unity to the power function for the estimation of durations, claiming an almost linear relationship between the psychological time and the physical time. Moreover, the highest precision in the estimation of duration is found with auditory stimulation (Westheimer, 1999). Westheimer (1999) measured the just noticeable difference for a set of stimuli ranging from 150-1500 ms. Stimuli were presented in three modality: touch, vision, or hearing. The lowest JNDs were always recorded with the auditory stimulations. Finally, listeners' inaccuracy when the target sound proceeded the target (experiment 1) could reflect a time order error: steady targets were underestimated in duration because of their position in the target-adjustable sequence. The experimental literature on time perception often highlights this error (see Allan, 1979 for a review).

Despite the great accuracy in estimating the duration of the steady sounds listeners underestimated both the duration of ramped sounds (slightly) and the duration of damped sounds (largely). Ramped tones were almost constantly underestimated in duration. The subjective duration of ramped tones was, overall, ~10% shorter than the actual duration of the tone. This result was particularly consistent for the estimates collected with the three longest targets, while with the shortest target, results showed smaller underestimations. On the contrary, the subjective duration of ramped noises was ~5-10% shorter than the subjective duration of ramped tones but it was overall constant across all the durations investigated.

I hypothesized that the small underestimation of ramped sounds could be explained by a sensory threshold placed at a fixed level of the ascending ramp: the beginning of the ramped

sounds was too quiet to be perceived and for this reason listeners did not include such a quiet portion of the sound in the computation of its subjective duration. The results collected with the reaction time demonstrated the existence of such a sensory threshold. The beginning of ramped sounds was perceived with a certain delay compared to the beginning of both steady and damped sounds. Overall, the sub-threshold portions measured with the reaction time could explain all the underestimation of ramped tones and almost all the underestimation of ramped noises. However, thresholds for noises and tones were identical and, moreover, with both carriers the magnitude of the sub-threshold portion decreased as the overall duration of the sound increased. The results collected with the adaptive procedure corroborated these findings: the beginning of ramped tones was sub-threshold and the sub-threshold portion decreased as the overall duration of the tone increased.

The magnitude of the sub-threshold portion may indeed be influenced by the overall duration of the ramped sound, therefore, by its rate of change in dB/s. There is, in fact, a consistent number of researches demonstrating that the latency of the neural onset response to the onset of tones is a direct function of the rate of change in level at tone's onset (Heil & Irvine, 1997; Heil, 1997a, 1997b, 2003). Moreover, such a finding extends to noise carriers (Phillips, Hall, Guo, & Burkard, 2001). The results of the current experiment show that also a behavioural onset response is a direct function of the rate of change in level of the ramp that is used to modulate the amplitude of the sound. In addition, the mismatch between the results collected with ramped white noises in the matching task and in the reaction time task could rise from the stochastic nature of the noise stimuli: early isolated randomly large amplitudes were effective at triggering a reaction time but could be ignored in estimating the overall sound's duration.

The subjective duration of the damped sounds was much shorter than both the duration of steady and ramped sounds. Across all matching experiments listeners underestimated the duration of damped sounds by ~30-40% independently from the timbre of the carrier. Overall, such an underestimation decreased as the duration of the damped sound increased. With both the reaction time and the adaptive procedure I tested whether such large underestimation could depend on a sensory threshold, (I) positioned at the end of damped sounds, and (II) placed at a higher level than that operating at the onset of the ramped sounds. The results showed that a threshold exists, however, such a threshold could explain only underestimations much smaller those measured with the matching task. In particular, the results were different for short and long durations. Hence, the threshold acting at the end of damped sounds is dependent on the rate of change in level of the sounds as well as the threshold acting at the beginning of ramped sounds. In detail, thresholds

measured with short damped sounds explained underestimations up to a maximum of 30% of sound's duration. On the contrary, thresholds measured with long damped sounds could explain underestimation up to a maximum of 10% of sound's duration. However, none of the thresholds could explain entirely the large underestimation of damped sounds measured in the matching task.

In synthesis, I found that ramped tones are perceived as longer than damped tones. Furthermore, I found that the difference in subjective duration between ramped and damped tones is great for short tones (high rate of change in dB/s) and decreases as the duration of the tone increases. Furthermore, the same pattern of results has been observed with ramped and damped noises. However, the difference in subjective duration between ramped and damped noises is smaller than the difference in subjective duration between ramped and damped tones. Such a decrement results from a further underestimation of ramped noises when compared to both steady sounds and ramped tones. Overall, these results are in line with the results obtained by Schlauch et al. (2001). In fact, also Schlauch et al. (2001) found that ramped sounds were perceived as longer than damped sounds and that ramped noises were more underestimated in duration than ramped tones. In addition, I found that ramped tones are perceived as shorter than steady tones and that such an underestimation was due to a threshold acting at their onset while the underestimation of damped sounds could be only marginally explained by a threshold acting at their offset.

Overall, the results of all researches performed comparing the subjective responses to ramped and damped sounds seem to converge. Ramped sounds are perceived as longer (Schlauch & al., 2001 and the current experiments) as louder, (Stecker & Hafter, 2000) as more changing in loudness (Neuhoff, 1998, 2001) as having a stronger sound quality (Patterson, 1994a, 1994b; Akeroyd & Patterson, 1995; Irino & Patterson, 1996) than damped sounds. Furthermore, the difference in the subjective estimates of ramped and damped sounds decreases (Schlauch & al., 2001; Akeroyd & Patterson, 1995; Irino & Patterson, 1996) or disappears (Neuhoff, 1998, 2001) when the carrier is a noise rather than a tone. Stecker and Hafter (2000) failed to find such a reduction, however, in their experiment the comparison between ramped and damped tones and noises was performed between subjects and not within subjects as in the other studies cited above. Moreover, in the current experiments as well as in those performed by Schlauch et al. (2001), Stecker and Hafter (2000), and Patterson and colleagues (Patterson, 1994a, 1994b; Akeroyd & Patterson, 1995; Irino & Patterson, 1996) I found a reduction of the difference in the behavioural response to ramped and damped sounds when the modulating function was shallower rather than steep. Finally, in the current experiment (see experiment 1) as in the experiment performed by Stecker and Hafter (2000) I found that the difference in the percepts of ramped and damped

sounds may be influenced by the order of the stimuli in a sound sequence. In synthesis, there is the possibility that such a convergence of results arises as all researches were investigating different aspects of the same phenomenon. Moreover, the results of the current study show clearly the direction of such a perceptual asymmetry. According to my results the asymmetry originates from a diminished perception of damped sounds rather than an augmented perception of ramped sounds as stated by Neuhoff (1998). In fact, the results collected with the reaction time and the adaptive procedure could explain the estimates given to ramped sounds in the matching task. On the contrary, the same experiments could not explain the estimates collected with damped sounds in the matching task.

The results of the current experiments suggest that the asymmetry has its origins in the perception of damped sounds rather than ramped sounds. Furthermore, within the range of duration investigated, the short percept of damped sounds cannot be explained by analysing the perceptual response to its low level end. This result suggests that, within the range of durations investigated, the asymmetry in the perception of ramped and damped sounds cannot be explained entirely as a peripheral phenomenon. In fact, I found either faint evidences (with the shortest duration) or null evidence for a different perceptual coding of the onset of ramped sounds and the offset of damped sounds.

At the state of the art the asymmetry can be explained as a peripheral phenomenon only for much briefer sounds and for very high rate of change in dB/s. The neural response to ramped and damped sounds, like those used by Patterson and colleagues (Patterson, 1994a, 1994b; Akeroyd & Patterson, 1995; Irino & Patterson, 1996), is asymmetric along the auditory pathway (Lu, & al., 2001a, 2001b; Fay, & al., 1996; Pressnitzer, & al., 2000; Neuert, & al., 2001). There is only one evidence supporting a different neural response to the two sounds for longer durations and lower rates of change in dB/s. In a fMRI study with human listeners Seifritz et al. (2002) demonstrated that ramped more than damped sounds activated a distributed neural network subserving space recognition, auditory motion perception and attention. However, such a difference does not inform us about the response strength or response timing to ramped and damped sounds but only about the cortical areas that the two sounds activate. Moreover, the different cortical activation found by Seifritz et al. (2002) could originate as ramped sounds more than damped sounds where evoking the impression of the movement of a sounds source as reported by the listeners of the behavioural experiment performed in the same study. Interestingly, according to Wang, Lu and Liang (2003) the neural response to ramped and damped sounds becomes more and more asymmetrical as we reach the central sites of the auditory system.

The results of the current experiments show that damped sounds are perceived as shorter than ramped sounds and that ramped sounds could be perceived veridically if their time course was completely over-threshold. Therefore, my results seem to support the hypothesis purposed by Stecker and Hafter (2000) more than the hypothesis purposes by Neuhoff (1998, 2001). It may be possible that listeners interpreted the quiet tail of damped sounds as the result of an echo and for this reason they decided to ignore this portion of the sound in the estimation of its subjective duration. Nonetheless, the hypothesis suggested by Stecker and Hafter (2000) rises some problem. According to Stecker and Hafter (2000) listeners would segregate the attack portion from the decay portion as the attack contains information about the source event while the decay contains information about the environmental where the event took place. However, in everyday sounds the ratio between the duration of the attack portion and the duration of the decay portion may be influenced by a number of factors. For example, in impact sounds the duration of the attack portion may be influenced by the material and shape of the objects involved in the sound source event while the duration of the decay portion may be influenced by the acoustical characteristics of the environment. As a consequence, we can have sounds with many possible ratios between the duration of the attack and decay portion. For this reason, we should expect listeners estimations of damped sounds as quite spread across listeners as any listener could decide arbitrarily how much of the sound corresponds to its attack and how much corresponds to its decay. By the same token, as ramped sounds should be perceived as having no reverberant portion (Stecker and Hafter, 2000) we should expect the subjective estimations of ramped sounds as quite consistent across listeners. On the contrary, the results of the matching experiments show that, estimations of ramped and damped sounds were similarly consistent and similarly consistent across listeners.

In addition, according to Stecker and Hafter (2000), the asymmetry between ramped and damped sounds may originate as listeners switch from “literal mode” perception to “constancy mode” perception. At the beginning listeners act in the literal mode, thus, they perceive the two sounds as having similar loudnesses (or durations in this case). Successively they switch from literal to constancy mode and they interpret the damped sound as composed of two parts (attack plus reverberation) so they ignore the tail of damped sounds. According to Stecker and Hafter (2000) such a switch in the perceptual attitude of the listener would require some time to be actuated. Such explanation would suite for the difference in the results collected with damped sounds in the reaction time and in the matching task. In the reaction time listeners would perceive in literal mode as the answer is virtually given in real time. On the contrary, in the matching task listeners would perceive in constancy mode as the answer is given delayed in time. However, the same explanation

does not suit for the difference in the results collected with the adaptive procedure where the answer was delayed compared to the perception of the stimuli, thus, listeners should have acted in the constancy mode.

In conclusion, I believe that the speculations purposed by Stecker and Hafter (2000) and Neuhoff (1998, 2001) are plausible. Short damped sounds may be perceived as everyday sounds and composed of two portions: the attack and the reverberant portion. Furthermore, ramped and damped sounds are indeed the acoustic pattern produced by approaching and receding sound sources. Nonetheless, the asymmetry is much evident even with ramped and damped sounds that do not remind any of the two analogies because they are too short (Patterson, 1994a, 1994b; Akeroyd & Patterson, 1995; Irino & Patterson, 1996) or too long (Canévet, 1986, Canévet & Scharf, 1990; Schlauch, 1992; Teghtsoonian, Teghtsoonian, & Canévet, 2000; Canévet, Teghtsoonian, & Teghtsoonian, 2003). So, why these two sounds should be perceived so different? I believe that the explanation of the phenomenon lay at a higher level. There is definitely a strong difference between ramped and damped sounds. In fact, while the evolution across time of ramped sounds is unpredictable the evolution of damped sounds across time is highly predictable as the sound will dissolve, at some point, into silence. The peculiarity of the sense of hearing is that any stimulation occurs along a single dimension: time. The ongoing perceptual flow of information we listen to in everyday situations requires high capability of detecting events that are occurring in very narrow temporal spaces. It may be possible that top-down suppression of highly predictable stimulations would proof useful for the perceiver. In this way the perceiver can keep low the amount of information he/she is analysing and reset promptly the cognitive load for the incoming new stimulations. A fast top-down suppression of events that are highly predictable would enable a prompter response to new events. On the one hand the hearing system has the possibility of tracking sound events along all their time course. On the other it may decide to suppress those information that do not require further analysis and attention.

References

- Allan, L. G. (1979). The perception of time. Perception & Psychophysics, *26*, 340-354.
- Akeroyd, M. A., & Patterson, R. D. (1995). Discrimination of wide band noises modulated by a temporally asymmetric function. Journal of the Acoustical Society of America, *98*, 2466-2474.
- Canévet, G. (1986). Estimation de sonie pour de sons purs à variation monotone de niveau. Acustica, *62*, 256-264.
- Canévet, G., & Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. Journal of the Acoustical Society of America, *88*, 2136-2142.
- Canévet, G., Teghtsoonian, R., & Teghtsoonian, M. (2003). A comparison of loudness change in signals that continuously rise and fall in amplitude. Acta Acustica United with Acustica, *89*, 339-345.
- Fay, R. R., Chronopoulos, M., & Patterson, R. D. (1996). The sounds of a sinusoid: Perception and neural representations in the goldfish (*Carassius Auratus*). Auditory Neuroscience, *2*, 377-392.
- Ghazanfar, A.A., Neuhoff, J.G., & Logothetis, N.K. (2002). Auditory looming perception in rhesus monkeys. Proceedings of the National Academy of Sciences, *99*, 15755-15757.
- Heil, P. (1997a). Auditory onset response revisited. I. First-spike timing. Journal of Neurophysiology, *78*, 2616-2641.
- Heil, P. (1997b). Auditory onset response revisited. II. Response strength. Journal of Neurophysiology, *78*, 2642-2660.
- Heil, P. (2003). Coding of temporal onset envelope in the auditory system. Speech Communication, *41*, 123-134.
- Heil, P., & Irvine, D. R. F., (1997). First-spike timing of auditory nerve fibers and comparison with auditory cortex. Journal of Neurophysiology, *78*, 2438-2454.
- Irino, T., & Patterson R. D. (1996). Temporal asymmetry in the auditory system. Journal of the Acoustical Society of America, *99*, 2316-2331.
- Levitt, H., (1971). Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America, *49*, 467-477.

- Lorenzi, C., Gallégo, S., & Patterson, R. D. Discrimination of temporal asymmetry in cochlear implantees. Journal of the Acoustical Society of America, 102, 482-485.
- Lu, T., Liang, L., & Wang, X. (2001a). Neural representations of temporal asymmetric stimuli in the auditory cortex. Journal of Neurophysiology, 85, 2364-2380.
- Lu, T., Liang, L., & Wang, X. (2001b). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nature Neuroscience, 4, 1131-1138.
- Neuhoff, J. G. (1998). Perceptual bias for rising tones. Nature, 395, 123-124.
- Neuhoff, J. G. (2001) An adaptive bias in the perception of looming auditory motion. Ecological Psychology, 13, 87-110.
- Neuert V., Pressnitzer D., Patterson R. D., & Winter I. M. (2001). The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids, 159, 36-52.
- Patterson, R. D. (1994a). The sound of a sinusoid: spectral models. Journal of the Acoustical Society of America, 96, 1409-1418.
- Patterson, R. D. (1994b). The sound of a sinusoid: time-interval models. Journal of the Acoustical Society of America, 96, 1419-1428.
- 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, 1890-1894.
- Patterson, R. D., & Irino, T. (1998). Modelling temporal asymmetry in the auditory system. Journal of the Acoustical Society of America, 104, 2967-2979.
- Philips, D. P., Hall, S. E., Guo, Y., & Burkard, R. (2001). Sensitivity of unanesthetized chinchilla auditory system to noise burst onset and effects of carboplatin. Hearing Research, 155, 133-142.
- Pressnitzer, D., Winter, I. M., & Patterson, R. D. (2000). The responses of single units in the ventral cochlear nucleus of the guinea pig to damped and ramped sounds. Hearing Research, 149, 155-166.
- Schlauch, R. S. (1992). A cognitive influence on the loudness of tones that change continuously in level. Journal of the Acoustical Society of America, 92, 758-765.
- 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, 2880-2887.

- Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, D., Mustovic, H., Schächinger, H., Elefante, R., & Di Salle, F. (2002) Neural processing of auditory 'looming' in the human brain. Current Biology, 12, 2147-2151.
- Small, A. M. (1977). Loudness perception of signals monotonically changing sound pressure. Journal of the Acoustical Society of America, 61, 1293-1297.
- Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry in loudness. Journal of the Acoustical Society of America, 107, 3358-3368.
- Stevens, S. S. (1957). On the psychological law. Psychological Review, 64, 153-181.
- Teghtsoonian R., Teghtsoonian M., & Canevet, G. (2000). The perception of waning signals: Decruitment in loudness and perceived size. Perception & Psychophysics, 62, 637-646.
- Wang X., Lu T., & Liang L. (2003). Cortical processing of temporal modulations. Speech Communication, 41, 107-121.
- Westheimer G. (1999). Discrimination of short time intervals by the human observer. Experimental Brain Research, 129, 121-126.
- Wright B.A., Buonomano D.V., Mahncke H.W., & Merzenich M.M. (1997). Learning and generalisation of auditory temporal-interval discrimination in humans. Journal of Neuroscience, 17, 3956-3963.

From sound to size: balls dropped on plates

Riassunto

Alcune recenti ricerche hanno dimostrato che, dato un suono d'impatto, siamo in grado di stimare la grandezza dell'oggetto che lo ha generato (Lakatos, McAdams & Caussé, 1997; Carello, Anderson & Kunckler-Peck, 1998; Kunckler-Peck & Turvey, 2000). I risultati di tali ricerche sono sorprendenti per due motivi: l'orecchio è raggiunto da un solo suono che è il risultato del contatto tra due oggetti. Ciò nonostante gli ascoltatori sono in grado di fornire stime della grandezza fisica di un oggetto indipendentemente dalle caratteristiche dell'altro (segregazione). Inoltre, le stime di grandezza fornite dai soggetti sono veridiche, cioè, esse sono spiegate al meglio dalla grandezza fisica degli oggetti stessi piuttosto che dalle caratteristiche acustiche del suono che giunge all'orecchio (percezione diretta).

Scopo dei miei esperimenti duplice: (I) investigare se la segregazione avviene realmente, cioè se si riesca ad estrarre le proprietà di un oggetto indipendentemente da manipolazioni effettuate sul secondo oggetto. (II) Investigare in che misura si potesse parlare o meno di percezione diretta della grandezza dell'oggetto. In un primo esperimento gli ascoltatori hanno stimato la grandezza di sfere a partire dal suono che queste facevano impattando un piatto. In un secondo ed in un terzo esperimento gli ascoltatori hanno effettuato lo stesso compito ma le sfere potevano cadere, a loro insaputa, su due (esperimento 2) o su tre piatti (esperimento 3). In un quarto esperimento ho chiesto invece ai soggetti di stimare la grandezza delle sfere facendole loro vedere.

I risultati hanno dimostrato che gli ascoltatori erano in grado di fornire stime veridiche della grandezza delle sfere e che le loro stime di grandezza erano spiegate al meglio se confrontate con la grandezza reale delle sfere piuttosto che con le caratteristiche acustiche del suono. La funzione psicofisica che meglio descriveva la relazione tra la grandezza soggettiva uditiva e quella fisica era una funzione potenza con esponente pari a 1.5. Inoltre, i risultati hanno dimostrato che tanto più grande era il piatto su cui la sfera veniva lasciata cadere tanto più grande era la stima di grandezza associata alla sfera. Nel quarto esperimento le stime visive della grandezza delle sfere erano legate alla grandezza fisica delle stesse con una funzione potenza solo leggermente positivamente accelerata (esponente pari a 1.1).

Complessivamente i risultati hanno dimostrato che (I) gli ascoltatori non sono stati in grado di segregare la grandezza della sfera indipendentemente da quella del piatto su cui la sfera andava a cadere. (II) Per quanto concerne invece la capacità degli ascoltatori di percepire direttamente la grandezza delle sfere i risultati sono contrastanti. Le stime uditive erano spiegate al meglio dalla grandezza delle sfere piuttosto che dalle caratteristiche acustiche del suono. Ciò nonostante, secondo Gibson (1963) la grandezza è una dimensione fisica utile all'essere umano, quindi, indipendente dal senso che si utilizza per percepirla. Invece, le funzioni psicofisiche uditive e visive per la stima della grandezza delle sfere differivano significativamente.

Abstract

Recent researches demonstrated that listeners are able to evaluate the size of an object from the sound it produces when impacting upon a second object (Lakatos, McAdams & Causse, 1997; Carello, Anderson & Kunckler-Peck, 1998; Kunckler-Peck & Turvey, 2000). The results of the researches reported above are doubly surprising: the wave pressure reaching the ear is the result of the contact between two objects. Nonetheless, listeners are able to estimate the size of one object independently from the other object (segregation). The listeners are usually accurate in evaluating the size of the objects and the best predictors of listeners' performance can be found in matching listeners estimations with the actual size of the object rather than matching listeners' estimations with the acoustical characteristics of the resulting sound (direct perception).

The aim of the current experiments was (I) to test whether listeners could really estimate the size of an object independently from the size of the other object; (II) to study the extent of direct perception of size from impact sounds. In the first experiment listeners evaluated the size of spheres from the sound they produced when dropped upon a plate. In the second and third experiment listeners performed the same task, however, spheres could be dropped to two (experiment 2) or three (experiment 3) plates of different diameters. Listeners participated in the experiments with no foregoing information about any of the characteristics of the sound source event. In a fourth experiment observers evaluated the visual size of the balls used in the first three experiments.

Results demonstrated that listeners could evaluate veridically the size of the balls from the sounds they produced impacting upon the plate. Furthermore, listeners' estimations were better explained when compared to the actual size of the ball rather than the acoustical characteristics of

the sounds. Listeners' estimations could be described with a power function with an exponent of 1.5. Moreover, listeners' estimations were dependent from the size of the plate where the ball was dropped upon: the larger the plate the larger the subjective size of the ball. In the fourth experiment visual estimations of size were as a power function of the actual size of the ball (exponent of 1.1).

Overall, results demonstrated that (I) listeners were not able to segregate the properties of the ball from the properties of the plate. For what concerns direct perception results were ambiguous. On the one side listeners' estimations could be described at best by the actual size of the ball rather than the acoustical characteristics of the sound. On the other, according to Gibson (1963) size is a useful dimension of sensitivity, thus, independent from the sense that we use to estimate it. However, the exponents of the auditory and visual psychophysical functions differed greatly.

Introduction

Imagine the situation: you are in your living room talking to a friend about the size of an object that is in your bedroom. You decide to solve the problem by going into the bedroom to get the object. When you arrive in the bedroom your friend from the living room asks you to drop the object on the floor. After listening to the sound of the impact, he says: "it's more or less 50 millimeters wide!".

Recent research has, in fact, pointed out that listeners are able to extract from a sound information about the physical event that is producing it. Results obtained are doubly surprising. The wave pressure reaching the ear is the result of the physical interaction of at least two objects. Nonetheless, listeners are able to segregate from this single wave pressure the information concerning the sole object under investigation. Moreover, after listening to the sound, they are accurate in judging static properties of the physical event that generated it. For example, participants were accurate in judging physical properties such as the length of a rod dropped on the floor (Carello, Anderson & Kunckler-Peck, 1998), the ratio dimensions (height and width) and the shape of either struck bars or plates (Lakatos, McAdams & Caussé, 1997; Kunckler-Peck & Turvey, 2000). In other experiments, listeners could identify the hardness of mallets striking different cooking pans (Freed, 1990) and the gender of walkers from the sound of their footsteps (Li, Logan & Pastore, 1991). Contemporarily, many authors have studied the perception of environmental sounds by means of synthetic stimuli. Klatzky, Pai and Krotkov (2000) and Lutfi and Oh (1997) found that, respectively, decay and frequency play a major role in the categorization of the material type. In a successive study Lutfi (2001) found that these two acoustic parameters contribute also to the subjective sensation of hollowness.

In such experiments, participants describe directly the source and the physical behavior of the source, the distal stimulus, rather than distinguishing between the acoustic parameters of the sound that the source is producing: the proximal stimulus (Gaver, 1988; VanDerveer, 1979a; VanDerveer 1979b). Hence, the most reliable predictors of the performance are the physical properties of the distal stimulus (length, shape, material, etc.) involved in the sound producing event, rather than the acoustical indexes of the proximal stimulus such as sound pressure level, frequency content of the sound, etc. According to Gaver (1993a, 1993b) and Fowler (1990, 1991) we perceive sources, and the source properties have consequences for our behavior. In particular,

Gaver (1993a, 1993b) distinguishes between everyday listening and musical listening. While the first involves direct perception and identification (either correct or wrong) of the sound source event (the distal stimulus) the second does not involve a recognition of a sound source event but rather a simple listening to properties of the proximal stimulus, the acoustic wave.

However, the physical interaction between objects and its relation with human perception needs to be further analyzed. Impact sounds, for example, are the result of an interaction between, at least, two objects. The resulting air pressure wave is dependent on which of the objects is vibrating. Of course, impact sounds can be produced in a number of different ways where, either one object or the other object (or both), is put into vibration. While listening to a book dropped on a concrete floor the resulting air pressure wave is mainly produced by the vibration of the book itself; conversely, the same book dropped on a table will produce an air pressure wave that is the result of the vibration of the book and the vibration of the table. Furthermore, the strength of the impact can affect significantly the resulting acoustic pattern: if we drop a book on the floor from a height of ten centimeters or one-hundred centimeters the resulting acoustic waves will be different. A higher strength in the impact will correspond to a different solicitation of the so-called modes of vibration of the object (Morse, 1981) and, therefore, in a different acoustic pattern. When put into vibration each object vibrates in a certain limited number of ways, characteristic of that object. The modes of vibrations are the possible patterns of vibration of a given object. Therefore, the sound produced by any object is characterized by only certain frequencies, proper of that object, and consequently by a particular timbre.

None of the researches previously reported has clearly stressed the distinction between the two partners producing the impact and their contributions to the resulting acoustic wave. It is likely that the more one object contributes to the resulting sound the more information (*i.e.* mass, size, shape) reveals to the listener about itself. Therefore, if we drop a fork on the floor or into a steel pan we will hear two different acoustic stimulations: in the former mainly the fork is vibrating; in the latter both the fork and the pan are vibrating. As a consequence, the first sound will provide us with much information about the physical characteristics of the fork itself (mass, size, material, etc.) and not as much about the floor (its area, its thickness etc.). Alternatively, the second sound will provide us a series of physical information about both the fork and the pan. All investigations on the perception of objects' size by the sound it produces were focused on the most vibrating object: a rod dropped on the floor in the research by Carello et al. (1998), a plate struck by a pendulum in the research by Kunckler-Peck and Turvey (2000), a struck bar in the research by Lakatos et al. (1997).

The role played by each object within the impact is identified with names. The object that provides the energy for the vibration of the second object is called exciter. Instead, the object that receives the energy from the exciter is called resonator. The two names indicate which action is due to each specific object: the former put into vibration the latter. So far, we saw that researches has been concentrated on investigating the perception of the physical properties of the resonator (Lakatos & al. 1997; Carello & al. 1998; Kunckler-Peck & Turvey, 2000). Freed (1990) performed the only research investigating the exciter. The current experiments want to provide further evidences on the perception of exciter's properties from impact sounds. All experiments reported show studies about the perception of the less vibrating object within the impact. In the investigations, the exciter (a wooden ball) is dropped upon the resonator (a backed clay plate). Therefore, the aim of the current research was to investigate systematically whether a listener can extract the size of an object even when the object's vibrations are contributing only slightly to the resulting air pressure wave.

Experiment one

The aim of the first experiment was to investigate whether listeners could estimate the size of balls from the sound they produce when dropped onto a plate. Therefore, whether they could give a metrical estimation of the size of the object that vibrates less as a result of the physical interaction. The reason why the vibration of the ball is negligible compared to the vibration of the plate is mainly due to the geometry of the two objects and the effect of the geometry on an efficient pressure of the air. The vibrational pattern of a ball is characterized by stress waves: when the ball impacts the plate it begins to vibrate omni-directionally. As a consequence, the perturbation of the air surrounding the ball is only marginally efficient. On the contrary, when the plate is put into vibration by the ball its vibrational pattern is much more efficient because is characterized by bending waves: because of its geometrical shape the plate behaves like a dipole with the vibration moving much air along its two flat sides (Morse, 1981).

Method

Subjects

Ten undergraduates of the University of Padova volunteered for the experiment. They all reported having a normal hearing.

Apparatus

Seven solid wooden balls (pine) of 10, 15, 20, 25, 30, 40, 50 mm of diameter and weighing, respectively, 0.35, 1.1, 2.9, 4.9, 8.6, 22.2, 44.5 g (mass density, $\sim 647 \text{ Kg/m}^2$), and a baked clay plate of 215 mm of diameter were used for the experiment. Balls were dropped in the middle of the plate and from a height of 150 mm through a plastic pipe, placed vertically, closed on the lower end by a window that held the ball still. A lever allowed the experimenter to open the window and release the ball. The plate was placed on a table and it was divided from the table with a 500 by 500 mm foam block 40 mm thick. Therefore, after the impact, the plate could vibrate freely without transmitting the vibration to the table. The foam block damped just slightly the vibration of the plate since its reverse was not completely flat but finished with a 3 mm high and 125 mm diameter circular edge. Listeners sat facing the apparatus. The entire apparatus and the experimenter were

two meters away from the listener and they were occluded from listener's view by a 1.5 by 1.5 m frame covered with opaque paper.

Procedure

Before the session started, the experimenter dropped a ball, randomly selected from the ball set. Then, the experimenter asked the listener if he/she could tell the shape of the object that had just been dropped. The main experiment then began. On each trial the experimenter dropped the same ball three times following listener's request. In each presentation, the ball was allowed to bounce freely until its motion ended. At the same time, during these three successive presentations the listener had to create and adjust a circle on a computer screen as large as the ball they thought had just been dropped. In particular, listeners were asked to imagine the ball and produce a disc whose diameter was as large as the diameter of the ball they had imagined. Within the experimental session each trial (single ball-size dropped on the plate) was repeated five times in randomized order. Therefore, thirty-five trials were presented to the listener. At the beginning of each trial the computer screen was blank. A custom software allowed circles to be drawn ranging from ~ 0.5 mm up to ~ 300 mm of diameter. The drawing of the circle was controlled by the listener with two pairs of keyboard keys: one pair led to an increment-decrement of ~ 0.5 mm of diameter of the circle at each key press; the other permitted an increment decrement of ~ 15 mm of the diameter of the circle at each key press. At the end of the experiment, as in the successive ones, listeners were asked questions about the shape and material of the impacting object (the ball) and what surface, shape and material, it was dropped on. These questions were intended to determine what overall knowledge the listener had obtained. Finally, in this experiment, as in the successive ones, listeners received no information about either the size of the balls, the height from which the balls were dropped or the material of the balls and the plate.

Results and discussion

As far as the preliminary question is concerned, all listeners answered that a spherical object had been dropped. A first analysis was conducted on the average circle diameters drawn by listeners for each ball size. This analysis showed that the relation between subjective and actual diameters could be described at best with a power function: linear fit $r^2 = 963$, $F(1, 5) = 131.11$, $p < .0001$, power fit $r^2 = .985$, $F(1, 5) = 338.55$, $p < .0001$. Therefore, geometric means of the log transformation of the circle diameters drawn by listeners (in millimeters) were calculated for each stimulus. On these data a linear regression was performed. The linear function highly fitted the log-

log transformation of subjective estimations and actual sizes: $r^2 = .985$, $F(1, 5) = 339.61$, $p < .0001$ ($B = 1.48$). Overall, listeners underestimated the size of the ball and such underestimation decreased as the size of the ball increased (see Figure 1). Furthermore, the psychophysical functions of all listeners were positively accelerated with slopes ranging from 1.26 up to 1.71.

Eight listeners reported they heard metal balls, one reported hearing glass balls, and the last listener reported hearing stone balls. Seven listeners reported that balls were dropped on a metal plate and three reported that a ceramic plate had been used.

Listeners produced reasonable estimates of the size of the ball. In fact, although the diameter of the balls dropped was underestimated compared to their actual diameter, the magnitude of such underestimation was generally small (5 to 7 mm). Furthermore, listeners had no foregoing knowledge of any of the characteristics of the physical interaction, therefore such underestimation could be simply incidental. Finally, the psychophysical function describing listeners' estimations was positively accelerated with large error in the estimation of small balls and small errors in the estimations of large balls. The linear interpolation touched almost all the data points, however, listeners' estimations had a flexion at the largest ball size.

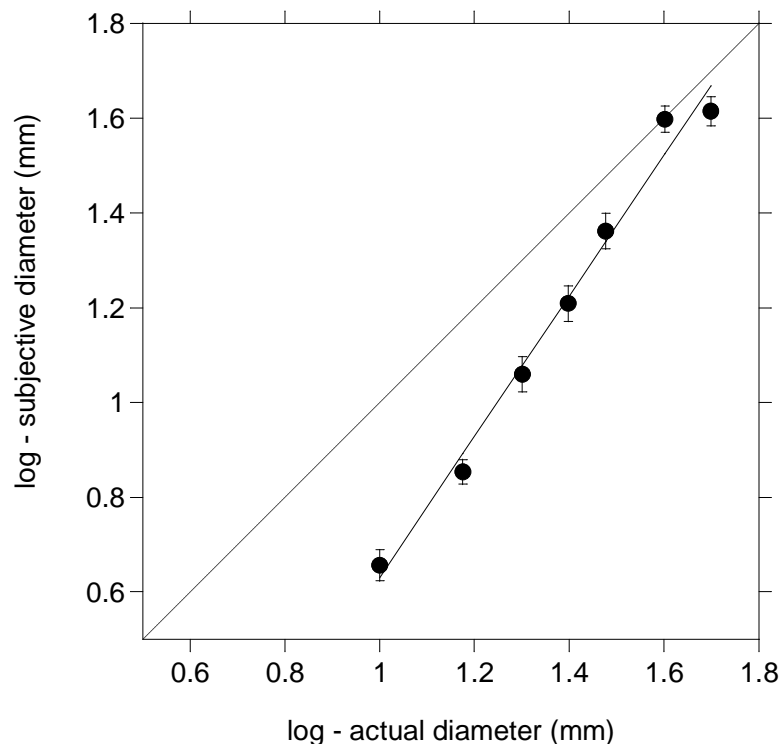


Figure 1: log-subjective diameter as a function of the log-actual diameter of the balls. The solid line represents the linear interpolation of the data. The diagonal dashed line represents perfect match. Vertical bars show ± 1 standard error of the mean.

Experiment two

The aim of the second experiment was to investigate whether listeners could still give a reliable metrical estimation of the size of the ball when listening to the sound of balls falling on two plates that had different diameters. In fact, all experiments investigating the perception of the size of an object from its sound have manipulated, within the experiment, only one of the two objects participating in the impact. However, impact sounds are the result of the contact between at least two objects: the resonator and the exciter. As a consequence it is still unknown whether listeners are able to extract the physical properties of one object independently from changes in the properties of the second object. If listeners perceive directly the size of the ball the subjective estimation of the diameter should be independent from the object where the ball is dropped upon. Therefore, a 50 mm diameter ball should be perceived as being ~50 mm large either when dropped on a table or, for example, on the floor.

Method

Subjects

Ten undergraduates of the University of Padova volunteered for the experiment. They all reported having a normal hearing. None of the subjects had participated in the previous experiment.

Apparatus and procedure

The procedure was the same as in experiment one. A second baked clay plate of 185 mm of diameter was used in addition to the 215 mm diameter plate from Experiment 1. This second plate had the same characteristics of the plate used in the previous experiment (material, shape etc.) and was only different in diameter. As in the first experiment, before the experimental session, listeners heard the sound of a single ball, then they were asked if they could recognize the shape of the impacting object from its sound. The procedure was the same as in the first experiment. However, the stimuli set (single combination ball-plate) was extended with the thirty-five new trials obtained dropping the balls also on the 185 mm diameter plate. This resulted in a total of seventy trials per experimental session. Each listener evaluated a random sequence of the stimuli. The duration of the experimental session was, approximately, 45 minutes.

Results and discussion

As far as the preliminary question is concerned, all listeners answered that a spherical object had been dropped. Geometric means of the log transformation of the circle diameters drawn by listeners (in millimeters) were calculated for each stimulus. Psychophysical functions for the estimations measured with the 215 and the 185 mm diameter plates were calculated on these data. A very high fit was obtained with two linear regressions performed with the log transformation of the actual diameter of the ball: $r^2 = .991$, $F(1, 5) = 567.30$, $p < .0001$ ($B = 1.59$) for the 215 mm diameter plate and $r^2 = .989$, $F(1, 5) = 458.49$, $p < .0001$ ($B = 1.47$) for the 185 mm diameter plate (see Figure 2).

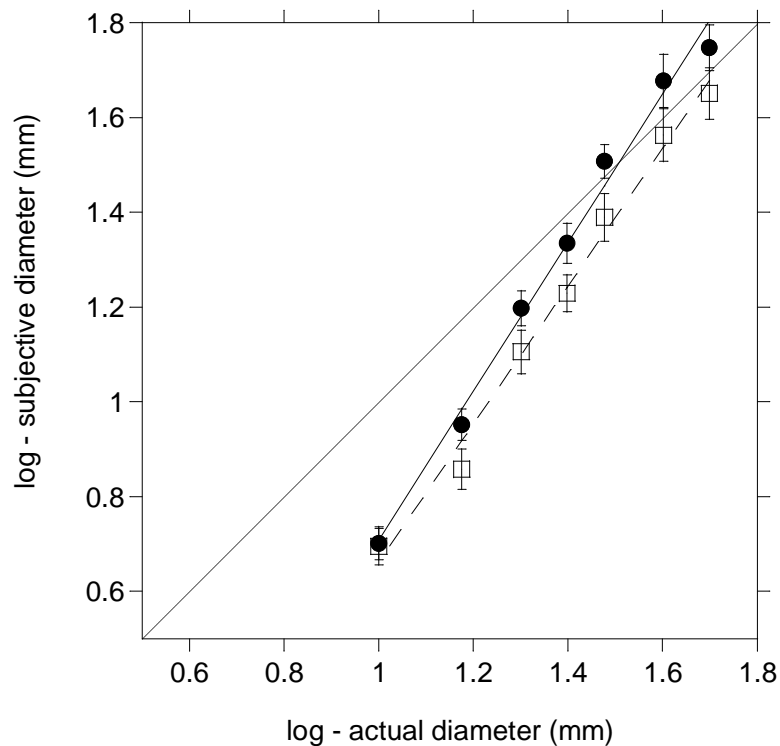


Figure 2: log subjective diameter as a function of the log actual diameter of the balls. Filled circles and empty squares represent, respectively, estimations collected with the 215 and the 185 mm diameter plate. The solid and the dashed line represent, respectively, the linear function fitting the estimations collected with the 215 and the 185 mm diameter plate. The diagonal dashed line represent perfect match. Vertical bars show ± 1 standard error of the mean.

An analysis of variance was performed on the data. The diameter of the balls dropped on the small plate was underestimated compared to the diameter of the same balls when dropped on the large plate: $F(1, 9) = 77.08$, $p < .0001$ (see Figure 2). In particular, the different sounds produced

by the two plates affected the perceived dimension of all balls in a similar fashion, $F(6, 54)=2.14$, $p>.05$: all balls increased their subjective size when dropped upon the largest plate.

Listeners reported that they heard either metal, ceramic, glass or wooden balls (one listener) falling on one or more plates made either of metal or ceramic. Some listener interpreted the different timbres of the two plates as due to the ball hitting the same plate in different points. The majority of the listeners reported that, within the experiment, either balls or plates could be made of more than one material.

Also in this experiment, listeners produced reasonable estimations of the size of the ball, however, the subjective diameter of the ball was dependent upon the plate on which the ball was dropped. This effect was evident for all balls. Compared to their actual diameter, the perceptual diameter of the balls was slightly underestimated when balls were dropped on the small plate and both underestimated and overestimated when balls were dropped on the large plate. As in the previous experiment the error of listeners' estimates was large for the smallest ball and small for the largest balls. Furthermore, the data of the current experiment showed a larger variability than the results of the first experiment (compare standard errors in Figure 1 and 2). In fact, while in the first experiment results showed a strong consistency across listeners, with all of them estimating similar perceptual sizes, in the second experiment listeners' estimations were more variable.

The slopes of the psychophysical functions estimated in the current and in the previous experiment are really similar with values oscillating around 1.5: the plate influences only the intercept of the regressions but not its slope. Furthermore, as in the previous experiment, listeners' estimations had a flexion at the largest ball size.

Experiment three

The aim of the third experiment was to investigate whether the difference in subjective size observed in the previous experiment was purely incidental or, on the contrary, was the result of the sound produced by the small plate being different. In the previous experiment listeners estimated the same ball larger when it was dropped upon the large plate and smaller when it was dropped upon the small plate. This result could be due to the fact that the sound produced by the ball when dropped upon the small plate was, overall, higher in frequency content. In the following experiment a third plate will be added. As in the previous experiment this plate will differ from the other two only for its diameter. In particular, the additional plate will be smaller than both plates previously used. If listeners' estimations are biased by the frequency content of the sound the estimations for the balls dropped onto this last plate should be lower than estimations for the other two plates.

Method

Subjects

Twelve undergraduates of the University of Padova volunteered for the experiment. They all reported having a normal hearing. None of the subjects had participated in the previous experiments.

Apparatus and procedure

The procedure was the same as in experiment two. A third baked clay plate of 165 mm of diameter was used in addition to the 215 mm diameter plate and the 185 mm diameter plate used in the second experiment. This third plate had the same characteristics of the plates used in experiment two (material, shape etc.) and was only different in diameter. As in the first and the second experiment, before the experimental session, listeners heard the sound of a single ball, then they were asked if they could recognize the shape of the impacting object from its sound. The procedure was similar to the procedure used for the second experiment. However, in the random stimuli sequence that listeners had to evaluate each stimulus (singular combination ball-plate) was repeated three times and not five as in the previous experiments. This resulted in a total of sixty-three trials (seven balls, three plates, three repetitions) per experimental session. The duration of the session was, approximately 40 minutes.

Results and discussion

As far as the preliminary question is concerned, all participants answered that a spherical object had been dropped. The geometric mean of the log transformation the circle diameters drawn by listeners (in mm) were calculated for each stimulus. Psychophysical functions for the estimations measured with the 215, the 185 and the 165 mm diameter plates were calculated on these data. Three linear regressions performed on the log transformation of the actual diameters of the balls highly fitted listeners' estimations: $r^2 = .985$, $F(1, 5) = 338.57$, $p < .0001$ ($B = 1.57$) for the subjective estimations measured with the 215 mm plate, $r^2 = .994$, $F(1, 5) = 841.12$, $p < .0001$ ($B = 1.54$) for the 185 mm diameter plate and $r^2 = .992$, $F(1, 5) = 677.44$, $p < .0001$ ($B = 1.56$) for the 165 mm diameter plate (see Figure 3).

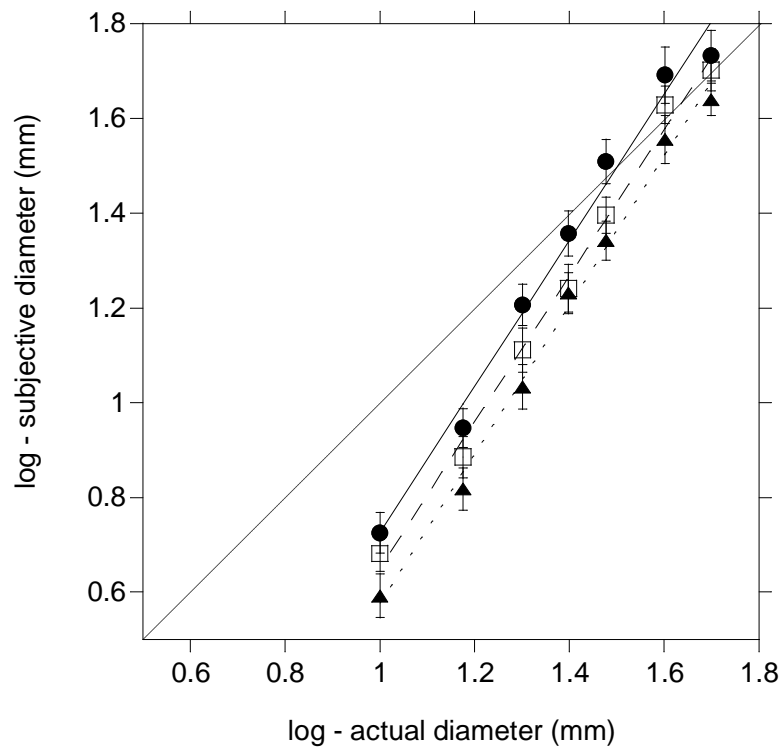


Figure 3: log subjective diameter as a function of the log actual diameter of the balls. Filled circles, empty squares and filled triangles represent, respectively, estimations collected with the 215, the 185 and the 165 mm diameter plate. The solid, dashed and dotted line represent, respectively, the linear function fitting the estimations collected with the 215, the 185 and the 165 mm diameter plate. The diagonal dashed line represent perfect match. Vertical bars show ± 1 standard error of the mean.

An analysis of variance was performed on the data. As in the previous experiment, the subjective estimate was dependent from the size of the plate: the bigger the plate the bigger the

estimate of the diameter of the ball, $F(2, 22)=33.95, p<.0001$ (see Figure 3). Furthermore, as in the previous experiment, the sounds produced by the three plates affected the subjective diameters of all balls in a similar fashion: $F<1$.

Listeners reported that they heard either metal, ceramic, glass or wooden balls falling on one (the largest part of the listeners) or more plates (two listeners) made either of metal or ceramic. As in Experiment 2 listeners interpreted the different timbres produced by the three plates as the result of using several materials for both the ball and the plate or as the result of the ball being dropped onto different points of a singular plate.

Also in this experiment the listeners provided reasonable estimations of the diameter of the ball. Compared to their actual diameter, the perceptual diameter of the balls was always underestimated when balls were dropped on the 165 mm diameter plate and both underestimated and overestimated when balls were dropped on 185 mm diameter plate or on the 215 mm diameter plate. Furthermore, as Experiment 1 and 2 the error in the estimation was maximal for the smallest balls and decreased as the size of the ball increased.

The results of the current experiment replicated and extended the findings of the second experiment. In fact, as in Experiment 2, the dimension of the plate had an effect on the subjective diameter of the ball: the larger the plate the larger the estimate of the diameter of the ball. As in the previous experiment the effect due to the plate diameter was additive: the change of the plate did not influence the slope of the psychophysical function but its intercept. Furthermore, the slope of the psychophysical functions estimated in the current experiment were very similar to those estimated in the first two experiments with exponents ranging around 1.5. Finally, as in Experiment 1 and 2 listeners' estimations had a flexion at the largest ball size.

Analysis of the physical and acoustical event

The physical interaction between the ball and the plate and its consequences on the resulting acoustic waveform can be described. On the one side, the physical event is characterized by a ball of mass m falling on the plate from height h , subject to gravity acceleration g . When the ball reaches the plate all the potential energy:

$$E = mgh$$

is converted into kinetic energy. The ball reaches the plate at a velocity v :

$$v = \sqrt{2gh}$$

and its kinetic energy (E_c) is equal to:

$$E_c = \frac{1}{2}mv^2$$

After the impact, part of this energy is transformed into acoustic energy while part is reconverted into kinetic energy (the ball bounces, the plate vibrates) and so on.

Overall, since the energy is the capacity of doing work, the power is the rate of doing work over time, and work, in the event under analysis, is the displacement of the air that is causing the sound, the perception of the size of the ball from its sound could be dependent on the ongoing air pressure wave reaching the ear (its power). In addition, other acoustical parameters, whose role may be important for estimating the size, are theoretically predictable. Firstly: the greater the mass impacting the plate the greater the oscillation of the plate. Consequently, the amplitude of the acoustic waveform will be greater and so will be the loudness of the resulting sound. Secondly: the greater the mass of the ball the longer the duration of the oscillation and, consequently, the duration of the resulting waveform. Although this is theoretically true, the subsequent impacts can damp the vibration of the plate and therefore shorten its oscillation. Thirdly: the greater the mass of the ball the longer the time of contact between the ball and the plate (Avanzini, 2001). Time of contact alters the frequency content of the sound produced by balls of different masses when dropped onto the plate. A long time of contact would damp vibrations whose periods are shorter than the time of contact itself. Consequently, the higher the mass impacting the plate the higher

would be the damping of the high frequency modes of vibration of the plate with a resulting attenuation of the high frequency components. As a result, the sound produced by a light ball impacting the plate is bright while the sound produced by a heavy ball impacting the plate is dull. This difference can be captured by the centroid. The centroid is the sum of the frequencies of a sound weighted by their relative amplitude. This acoustical index is related to the perceptual brightness of the sound (Grey & Gordon, 1978). Finally, the substitution of the plate with a smaller plate, as in Experiment 2 and 3, should affect mainly the acoustical event. In particular, a decrease in the plate diameter produces a sound higher in pitch as a result of all frequency components being shifted up by a certain ratio.

An increment in the mass of the ball impacting the plate affects not only the sound on its first impact but also the *high order structure* of the event (Warren & Verbrugge, 1984). The high order structure is defined as the temporal distribution of signal, and consequently, in the current event, as the temporal distribution of the bounces of the ball. In fact, in impacts where b is kept constant and the mass of the ball m is increased bounces will be less spread in time. The velocity v is independent from the mass of the ball m . Furthermore, the higher the mass m the longer the time of contact between the ball and the plate. As a consequence, the dissipation of kinetic energy for long time of contact will be greater than the dissipation for short time of contact. Therefore, after the impact, light balls will travel for longer times than heavy balls and, consequently, the distance between the first and the second bounce will be greater for lightest balls so will be the overall number of bounces. *Mutatis mutandis*, the analyses can be repeated identically for all the bounces following the first.

The goal of the following analysis was to test whether any of the acoustical characteristics of the sounds could fit listeners' performance. The sounds of the dropped ball were recorded with a Sennheiser MKH 40 P48 microphone into a portable Tascam DA-P1 DAT recorder at 44.1 kHz sample rate and 16 bits resolution. Twenty recordings were recorded for each ball/plate combination. The duration of the sound, its amplitude peak, its frequency centroid, its average root mean square (RMS) power and the time between the first and the second bounce (see appendix for details) were calculated from each recording and then averaged for each ball/plate combination. On the contrary, for the number of bounces the modal value was recorded. These values were used to compute forward multiple linear regressions on the log transformation of the subjective diameters collected in the first three experiments. This kind of multiple regression starts by choosing the predictor which explains the most variation in the dependent variable. Successively, if necessary, it chooses a second predictor which explains the most residual variation, and then recalculate

regression coefficients. The process continues until no variables significantly explain residual variation.

Analysis of experiment one

In the forward multiple linear regression the following predictors were included: the average RMS power of the sound, its amplitude peak, its centroid, its overall duration, the temporal distance between the first two impacts and the overall number of bounces. The power of the sound alone was a very good predictor of the performance: $r^2 = .978$, $F(1, 5) = 217.44$, $p < .0001$.

Analysis of experiment two

The average RMS power of the sound could predict well the estimations provided by the listeners for the 215 mm diameter plate: $r^2 = .989$, $F(1, 5) = 450.56$, $p < .0001$. On the contrary, the average duration was the best predictor for the estimations recorded for the 185 mm diameter plate: $r^2 = .977$, $F(1, 5) = 208.36$, $p < .0001$. However, by excluding this predictor, the power of the sound was the best predictor for listeners' performance: $r^2 = .973$, $F(1, 5) = 179.79$, $p < .0001$.

The results of the second experiment highlighted also that the perceptual size of the ball was dependent on the plate which the ball is dropped upon. However, the change in plate diameter should, theoretically, affect only the frequency content of the sound.

A new forward multiple linear regression was performed. This analysis was performed in order to understand whether any acoustical predictor could explain the difference between the perceptual diameters of the balls dropped onto the small plate compared to the perceptual diameters of the same balls when dropped onto the large plate. The difference between the log perceived diameters of the balls dropped on the 215 mm plate with the log perceived diameters of the same balls dropped on the 185 mm plate was calculated for each of the seven ball sizes used in the experiment. The same differences were calculated for each of the acoustical indexes. These differences were used, respectively, as a dependent variable and predictors for the new multiple forward regression. The difference between the centroids given by the two plates could explain the difference in the subjective estimates of the size of the balls dropped onto the two plates: $r^2 = .594$, $F(1, 5) = 9.76$, $p = .026$.

Analysis of experiment three

The average RMS power of the signal was the best predictor of the performance for the subjective diameters recorded with the 215 mm diameter plate, $r^2 = .985$, $F(1, 5) = 318.69$, $p < .0001$ and the results recorded for the 165 mm diameter plate, $r^2 = .981$, $F(1, 5)$, $p < .0001$. On the contrary, a model including together the overall duration of the sound and the centroid fitted well the results recorded for the 185 mm diameter plate, $r^2 = .996$, $F(2, 4) = 529.95$, $p < .0001$. However, by excluding this predictor, the power of the sound resulted the best predictor: $r^2 = .975$, $F(2, 4) = 197.74$, $p < .0001$.

As for the results of the second experiment, the analysis between the differences in the log subjective diameters measured with the three plates on the one side and the difference in the acoustical predictors on the other was conducted. None of the acoustical predictors could explain the different estimations measured for the three plates.

Experiment four

In the first three experiments listeners evaluated the size of the balls after listening to the sound they produced when impacting upon a plate. The purpose of the current experiment was instead to investigate how the actual diameter of the ball maps into the its visual estimate. In the current experiment a new group of participants evaluated the visual size of the balls used in the previous experiments by mean of the disc generating software. Therefore, it will be possible to compare the psychophysical functions obtained with the visual and hearing estimations.

Method

Subjects

Four teen undergraduates of the University of Padova with normal or corrected to normal vision volunteered for the experiment. None of the subjects had participated in the previous experiments.

Apparatus and procedure

Observers participated in the experiment individually. Each observer sat front of a computed screen placed 60 cm away from him/herself. The distance between the observer and the computer screen was kept constant by mean of a chin rest. On each trial the experimenter placed a ball on a pedestal that was placed on the left hand side of the observer. The ball was positioned on a point that was 75 cm away from the observer and 32 cm away from the point of the computer screen where the disc appeared. During each trial the task for the observer was to generate a disc as large as the ball placed on the pedestal. In particular, the observer was asked to create a circle whose diameter was as large as the diameter of the ball. Each observer evaluated a random sequence of twenty-one stimuli, therefore, each ball was visually measured three times by each observer.

Results and discussion and comparison with results of previous experiments

The geometric means of logarithmic transformation of circle diameters drawn by observers (in mm) were calculated separately for each ball copied and each observer (see Figure 4). Observers slightly underestimated the actual diameter of the ball and the value of this underestimation was,

approximately, 3.9 mm. A linear regression performed on the log transformation of the actual balls' diameters highly fitted the behavioral data: $r^2 = .999$, $F(1, 5) = 6982.06$, $p < .0001$ ($B = 1.13$). The slopes of the individual psychophysical functions ranged from .95 up to 1.26.

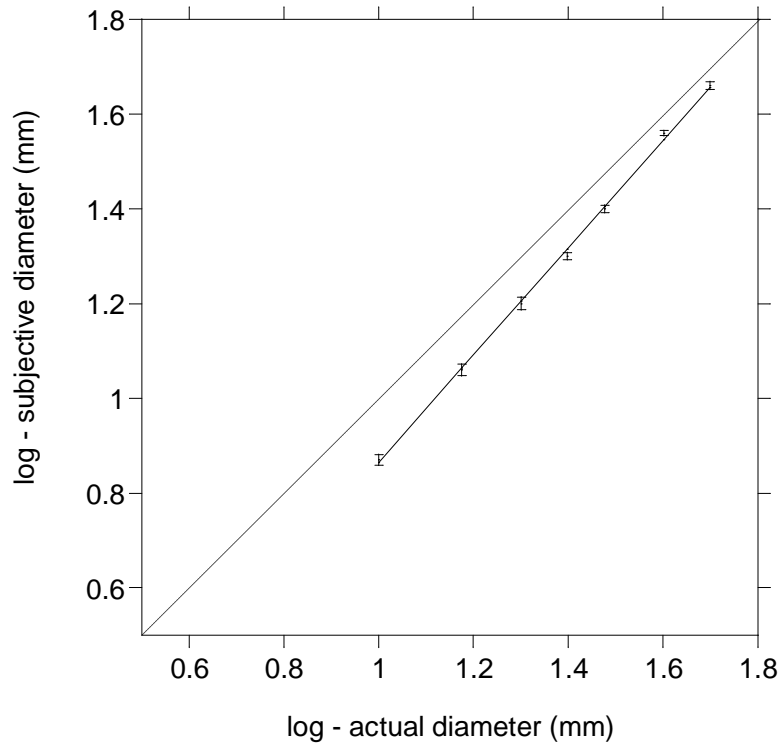


Figure 4: log subjective diameter as a function of the log actual diameter of the balls (visual task). The solid line represents the linear function interpolating the data. The diagonal dashed line represent perfect match. Vertical bars represent ± 1 standard error of the mean.

The slopes of the psychophysical functions calculated in Experiment 1, 2, 3 and in the current experiment were compared. Individual slopes were calculated for each subject of the four experiments. In the case of Experiment 2 and 3, as the plate did not influence the slope of the psychophysical functions the mean slope of listener's slopes was calculated. An one way analysis of variance with the experiment as between factor was performed on the slopes. Slopes measured in the four experiments differed, $F(3, 42) = 6.28$, $p = .001$ and such a difference was due to the slopes measured in the current experiment being much lower than the slopes measured in Experiment 1, 2 and 3. In fact, the same analysis of variance performed without the data of the current experiment was not significant: $F < 1$.

Overall, the visual subjective size of the balls was slightly underestimated compared to their actual sizes. The psychophysical function was just slightly accelerated with large errors in the

estimations of the smallest balls and small errors in the estimations of the largest balls. Furthermore, the slope of the visual psychophysical function was consistently lower than the slope of the hearing psychophysical function.

General discussion

The results of the first three experiments showed that listeners produced reasonable estimations of the actual size of the balls. Moreover, in all experiments, listeners underestimated the size of the smallest balls and either estimated veridically or slightly overestimated the size of the largest balls. Throughout the experiments, the results were very consistent with all the psychophysical functions positively accelerated and their exponents always ranging around 1.5.

However, in the second and the third experiment, listeners were influenced by the size of the plate: the same ball was estimated larger when dropped upon a large plate and smaller when dropped upon a small plate. The effect of the plate on the subjective estimates was additive: the plate did not influence the exponent of the psychophysical functions but the intercept of the linear regressions in log-log coordinates.

The comparison between subjective estimates and acoustical parameters showed that some acoustical index (*i.e.* RMS power) was almost as efficient as the actual size of the ball for fitting listeners' estimations. Moreover, at least for the results of the second experiment, a frequency domain index, the spectral centroid, could provide an explanation for the effect of the plate on the subjective estimates.

Finally, in the fourth experiment, observers provided a veridical visual copy of the balls by means of the disc generating software. In this experiment, the size of the ball was just marginally underestimated and the resulting psychophysical function was only slightly positively accelerated. Moreover, the exponent of the psychophysical function calculated in Experiment 4 was significantly lower than the exponents calculated in Experiment 1, 2 and 3.

The result of the first three experiments showed that listeners can provide reasonable estimations of the size of an object from its sound even when the vibrations of the object are minimal, thus, they are contributing only slightly to the resulting sound pressure wave. Freed (1990) already showed that listeners can extract information about the less vibrating object. In his research subjects were judging on a continuous arbitrary scale the hardness of a set of mallets striking cooking pans. However, the results of the current experiments seems more remarkable: firstly, listeners' answer was isomorphic to the stimulus, therefore, it was possible to compare directly listeners' estimations with the physical parameter under investigation. Secondly, Freed (1990) estimated the hardness of the mallets only on an ordinal scale.

Throughout the first three experiments, the results showed a steady psychophysical function. Such a function was a power function positively accelerated with an exponent slightly larger than 1.5. In addition, all psychophysical functions shared a second characteristic: they all had a flexion at the largest ball size. Such a flexion reflects flexions and reversals of almost all acoustical parameters at the largest balls size (see appendix). It may be possible that, with the largest ball and consequently with the greatest mass, the modal vibrations of the plate were saturated. Hence, the use of the largest ball did not lead to an appropriate increment (or decrement) in the acoustical parameters. Therefore, listeners' estimations mirrored the acoustical characteristics of the sound. By excluding the data collected with the largest ball all exponents would increase considerably: Experiment 1 from 1.48 to 1.57; Experiment 2 from 1.56 and 1.45 to 1.65 and 1.49, for the estimates collected with, respectively, the 215 and the 185 mm diameter plate; Experiment 3 from 1.54, 1.53 and 1.57 to 1.65, 1.58 and 1.63 for the estimates collected with, respectively, the 215, the 185 and the 165 mm diameter plate.

In Experiment 2 and 3 listeners perceived systematically larger (or smaller) the balls when dropped upon a large (or small) plate. As all plates shared the same shape and material an increment (or decrement) in their diameter resulted only in a different timbre of the sound as all frequency components were shifted up (or down) by a certain ratio (see values of the spectral centroid in appendix). Therefore, a high pitch and bright sound conveyed the impression that the sound was produced by a small ball and this result was supported by the centroid being the most powerful predictor for the different subjective estimates measured in Experiment 2. At the beginning of the last century many researchers dedicated themselves to the study of the tonal volume (Rich, 1919; Boring, 1926). According to the tonal volume hypothesis, high pitch tones are extensively associated with small objects while low pitch tones are extensively associated with large and voluminous objects (Perrot, Musicant & Schwethelm, 1980). Likely, the size of a ball dropped onto a small plate was underestimated compared to size of the same ball when dropped onto the large plate as the resulting sound was both higher in pitch and brighter.

The regressive analysis between the behavioral measures and the acoustical event showed that a change in the size of the ball was mainly affecting the amplitude content of the sound: its power. Grassi (2002) demonstrated that the manipulation of this acoustical parameter was the most effective in impairing the performance in a size-from-sound research. In his experiments listeners had to categorize the sounds of four balls dropped onto a plate according to their size in a four alternative forced choice task. Sounds could be either unmodified, or deprived of their bounces, or low-pass filtered, or high-pass filtered, or equated for RMS power. The worst performance was

recorded when sounds were equated for RMS power. Power may be the crucial acoustical clue linking the actual size of the exciter to the subjective size of the exciter. Nonetheless, according to listeners' reports after the experiments, not all of them based their estimates on the amplitude content of the sound. In fact, few listeners reported that, during the experiment, they were paying more attention to the frequency content of the sound in order to produce their estimates.

In studies on impact sounds both the perception of the properties of the resonator (Lakatos & al. 1997; Carello & al. 1998; Kunckler-Peck & Turvey, 2000) and the exciter (Freed 1990) have been investigated. In two of the researches investigating the resonator (Lakatos & al. 1997; Kunckler-Peck & Turvey, 2000) authors indicated the frequency domain as responsible for carrying information about the ratio between the height and the width of a stuck bar (Lakatos & al. 1997) and the information about the shape of a plate (Kunckler-Peck & Turvey, 2000). In fact, the vibrational pattern of any object is dependent on its physical dimension and shape. As a consequence, any manipulation of the metrical dimensions of the resonator corresponds to a different frequency content in the resulting sound.

In the experiments reported here I studied the perception of the exciter of the physical impact: the object that provides to the sounding object the necessary energy for the vibration. In an ideal interaction where the exciter does not vibrate, a variation of its impacting force corresponds mainly to a variation in the amplitude content of the sound produced by the resonator. Of course, in interactions between resonators and exciters, both objects vibrate but the second vibrates less than the first and, consequently, the statement above remains valid: a manipulation concerning the exciter results mainly in a variation of the amplitude domain.

In the introduction of this paper I mentioned the two problems raised by the researches using real impact sounds: the perceptual segregation of the informations concerning the two objects involved in the impact and the veridicality of listeners' estimations of the physical properties of the objects.

In impact sounds, the wave pressure reaching the ear is the result of the physical interaction of (at least) two objects. Nonetheless, in the current and in previous experiments (Lakatos & al. 1997; Carello & al. 1998; Kunckler-Peck & Turvey, 2000), listeners seemed able to segregate from this single wave pressure the information concerning the sole object they were asked to evaluate. Did listeners really segregate the informations concerning the plate from those concerning the ball? How accurate was the reconstruction of the sound producing event? In the current experiments, listeners easily understood that the sound was the result of a ball being dropped on a circular plate and both results can be explained. Balls are clearly distinguishable from

other objects because of the regularity of their bounces (Warren & Verbrugge, 1984). Furthermore, Kunckler-Peck and Turvey (2000) demonstrated that we can recognize the shape of a resonator (the plate) from its sound. Nonetheless, a certain confusion arose when listeners had to recognise the material of the two objects, especially the non-vibrating object. Across the first three experiments only few listeners reported to hear wooden balls, all the rest thought that balls could be made of metal, glass or ceramic. In addition, in the second and third experiment, only few listeners understood that the sounds were obtained with two (or three) plates. On the contrary, listeners thought either that the ball was hitting the plate in different points or that the balls could be made of different materials. Nonetheless, despite the largest part of listeners thought that only one plate was used in the experiment the results showed that they were indeed influenced in their estimations by the size of the plate. In conclusion, listeners were unable to segregate from the resulting sound the effects due to the size of the plate from the effects due to the size of the ball and the reconstruction of the sound producing event was only partially veridical.

In the current and in previous experiments (Lakatos & al. 1997; Carello & al. 1998; Kunckler-Peck & Turvey, 2000), after listening to the sound, listeners were usually accurate in judging static properties of the physical event that generated it. Furthermore, the best predictors for listeners' performance were usually found in the physical properties of the object itself rather than in the acoustical characteristics of the sound it produced. Did listeners provide a direct estimate of the size of the ball? In the current experiment, the simplest backward path for recovering the size of the ball would include: the recognition of the shape of exciter and resonator and the selection of that relevant acoustic feature (for example sound power) that conveys size information about the exciter (the invariant). At this stage, the listener need to apply the transfer function that links the mass of the ball to the power of the sound. The final step would require the calculation of the logarithm of mass of the ball, its transformation into the logarithm of diameter by multiplying its value for a constant. Such an output can be successively scaled according to some secondary properties of the event such as the material of the ball (its mass density), the height from which the balls are dropped or the distance of the sound source event. In the current experiments listeners provided reasonable estimations of the size of the balls and their actual sizes resulted the best single predictor for listeners' estimations. Therefore, at a first glance, results could reflect a direct perception of size. However, results showed psychophysical functions always positively accelerated demonstrating an evident error in the transfer function of the process. For example, if listeners thought that balls were made out of steel we should have expected a constant underestimation of the subjective size of the balls compared to their actual size as the mass density of steel is higher

than wood. On the contrary, none of the psychophysical functions measured in the experiments showed a constant error in the estimation. In fact, as many acoustical parameters could efficiently scale the subjective size of the ball, it may be also plausible that listeners used the disc generating software to scale one or more aspects of the auditory sensation rather than evaluating directly the size of the ball. Moreover, according to Gibson (1963), size is a useful dimension of sensitivity. Therefore, if size is important, whichever function links subjective size and actual size, we should expect the exponents of such a function to be constant, independent from the sense that we use to experience the stimulation. The current experiment failed to corroborate such prediction: the psychophysical functions of hearing subjective size and visual subjective size had very different exponents. As a corollary: such a difference was not the result of an unreliable estimation of the hearing psychophysical function; such a difference would have been even greater by excluding the largest ball size in the computation of the hearing psychophysical functions. In conclusion, it is not clear whether listeners provided a direct estimation of size or not, nonetheless, although size is a useful dimension of sensitivity, the relation between subjective size and actual size is different in hearing and vision.

How the ability of reconstructing the sound source event and the ability of evaluating the physical properties of the event are related one another? The research demonstrated that listeners are able to recognize an event among a class of possible events (VanDerveer, 1979a; VanDerveer 1979b; Warren & Verbrugge, 1984; Gaver, 1988). Furthermore, listeners are able to estimate veridically the physical properties of the event (Freed, 1990; Carello, & al., 1998; Lakatos, & al. 1997; Kunckler-Peck & Turvey, 2000). The results of the current experiment show that even when listeners are able to recognise partially or wrongly the sound producing event they can still estimate the physical properties of an object. Therefore, a veridical reconstruction of the sound source scene is not a necessary condition for a veridical estimation of the physical property of an object.

In conclusion, it may be trivial to stress that certain senses are specializes for certain tasks. In fact, I think that the hearing system need to provide us neither veridical information about the size of an object nor size estimates similar to visual estimates. On the contrary, the hearing system has to provide a functional estimate of the size of the object we are listening to, or even better, a functional estimate of the size of the event we are listening to (*i.e.*, in the current experiment, the event plate plus ball). In fact, a hearing estimate of size may be crucial when, for some reason, objects are occluded to vision. In addition, large and voluminous events may be potentially more dangerous than smaller events. As a consequence, an overestimation of the size of a large event, as in the Experiment 2 and 3, would be advantageous for the perceiver. If we are listening to a low

frequency sound we are likely facing an event that is generated by a large object. If we are listening to a high amplitude sound we are facing an event that is either close to us or generated by a large object or both. In addition, all psychophysical functions measured here were positively accelerated. Therefore, the subjective estimate of size increases much faster than its physical counterpart. As a consequence, our perceptual system seems to provides us with even greater margins of safety.

References

- Avanzini, F. (2001). Computational issues in physically-based sound models. Unpublished doctoral dissertation, Università di Padova, Padova.
- Boring, E. G. (1926). Auditory theory with special reference to intensity, volume, and localization. American Journal of Psychology, *37*, 157–188.
- Carello, C., Anderson, K. L., & Kunckler-Peck, A. J. (1998). Perception of object length by sound. Psychological Science, *9*, 211–214.
- Fowler, C. A. (1990). Sound-producing sourcing of perception: Ratenormalization and nonspeech perception. Journal of the Acoustical Society of America, *88*, 1236–1249.
- Fowler, C. A. (1991). Auditory perception is not special: We see the world, the fell the world, we hear the world. Journal of the Acoustical Society of America, *89*, 2910–2915.
- Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. Journal of the Acoustical Society of America, *87*, 311–322.
- Gaver, W. W. (1988). Everyday listening and auditory icons. Unpublished doctoral dissertation, University of California, San Diego.
- Gaver, W. W. (1993a). What in the world do we hear? An ecological approach to auditory event perception. Ecological Psychology, *5*, 1–29.
- Gaver, W. W. (1993b). How do we hear the world? Explanations in ecological acoustics. Ecological Psychology, *5*, 285–313.
- Gibson, J.J. (1963). The useful dimension of sensitivity. American Psychologist, *18*:115,
- Grassi, M. (2002). Recognising the size of objects from sounds with manipulated acoustical parameters. In J. A. Da Silva, E. H. Matsushima, & P. Ribeiro-Filho (Eds.), Fechner Day 2002: Proceedings of the International Society for Psychophysics (pp. 392-397). Rio de Janeiro, Brazil: International Society for Psychophysics.
- Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America, *63*, 1493–1500.
- Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. Presence: Teleoperators and Virtual Environments, *9*, 399-410.
- Kunckler-Peck, A. J., & Turvey, M. T. (2000). Hearing shapes. Journal of Experimental Psychology: Human perception & Performance, *26*, 279–294.

- Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. Perception & Psychophysics, *59*, 1180-1190.
- Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. Journal of the Acoustical Society of America, *90*, 3036–3049.
- Lutfi, R. A., & Oh, E. L. (1997). Auditory discrimination of material changes in a struck-clamped bar. Journal of the Acoustical Society of America, *102*, 3647–3656.
- Lutfi R. A. (2001). Auditory detection of hollowness. Journal of the Acoustical Society of America, *110*, 1110–1119.
- Morse, P. M. (1981). Vibration and sound (3rd ed.). New York: Acoustical Society of America.
- Perrot, D. R., Musicant, A., & Schwethelm, B. (1980). The expanding image effect: The concept of tonal volume revisited. Journal of Auditory Research, *20*, 43–55.
- Rich, G. J. (1919). A preliminary study on tonal volume. Journal of Experimental Psychology, *1*, 13–22.
- VanDerveer, N. J. (1979a, June). Acoustic information for event perception. Paper Presented at the Celebration In Honour of Eleanor J. Gibson, Cornell University, Ithaca, NY.
- VanDerveer, N. J. (1979b, June). Confusion errors in the identification of environmental sounds. Paper Presented at the Celebration In Honour of Eleanor J. Gibson, Cornell University, Ithaca, NY.
- Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. Journal of Experimental Psychology: Human Perception and Performance, *10*, 704-712.

Appendix

log10-actual diameter (mm)	Overall duration (ms)	Amplitude peak (dB)	Average RMS power (dB)	log10-Centroid (Hz)	Bounce to bounce (ms)	Number of bounces
Plate 215 mm						
1.00	765.1	-20.9	-51.8	3.82	213.0	4
1.18	782.1	-16.3	-48.2	3.75	195.0	3
1.30	716.1	-9.9	-42.9	3.71	170.8	3
1.40	707.5	-10.0	-42.1	3.65	167.0	3
1.48	671.7	-6.9	-39.3	3.64	137.1	2
1.60	664.5	-5.5	-36.9	3.63	111.9	2
1.70	903.8	-5.0	-34.4	3.60	195.0	3
Plate 185 mm						
1.00	825.9	-20.8	-54.1	3.84	204.2	3
1.18	876.7	-17.1	-49.1	3.78	198.1	3
1.30	961.9	-10.5	-44.5	3.76	175.6	3
1.40	966.9	-9.8	-43.6	3.73	175.3	3
1.48	1021.5	-7.0	-41.3	3.73	149.2	3
1.60	1123.7	-5.1	-38.5	3.73	146.6	2
1.70	1115.2	-6.6	-37.6	3.67	191.8	3
Plate 165 mm						
1.00	818.1	-20.2	-52.8	3.86	220.9	4
1.18	844.4	-17.5	-48.7	3.80	198.2	3
1.30	846.5	-12.6	-44.2	3.77	173.4	3
1.40	895.8	-10.6	-43.2	3.74	171.4	3
1.48	927.1	-8.5	-39.9	3.74	159.0	2
1.60	979.2	-5.3	-36.8	3.74	153.6	3
1.70	970.5	-6.7	-37.0	3.69	173.7	3

Average acoustical values extracted from the recordings of the stimuli used in Experiment 1, 2 and 3. The overall duration of the sound was calculated from the onset of the sound up to the point where the amplitude of the waveform reached level of the background noise amplitude. The average RMS power of the sound was calculated over all the event by mean of a 50 ms sliding window. The computation of the centroid was performed on all the event by averaging the centroids computed by mean of concatenated temporal windows of 50 ms. The calculation of the centroid included frequencies up to 22050 Hz. In the computation of the centroid we excluded all frequencies components whose amplitude was 30 dB below the highest frequency component. This permitted to exclude frequencies components of the background noise. Bounce to bounce is the time between the first and the second bounce.