RECOGNISING THE SIZE OF OBJECTS FROM SOUNDS WITH MANIPULATED ACOUSTICAL PARAMETERS

Massimo Grassi[a,b]
[a] Dipartimento di Psicologia Generale, via Venezia 8, 35131, Padova Italy
[b] Dipartimento di Scienze Filosofiche e Storico Sociali, via Antonini 8, 33100, Udine, Italy
emial: grassi@psy.unipd.it

ABSTRACT

Listeners were asked to categorise the recordings of a ball dropped on a plate according to its size in a 4AFC task. One control condition and four experimental conditions were investigated. In the control condition subjects were categorising the size of the ball by listening to the original recording. In the experimental conditions listeners were performing the same task either with sounds without the bounces, with sounds equated for RMS power, with sound low-pass filtered, or with sounds high-pass filtered. Each manipulation impaired subjects' performance compared to the control condition: without bounces and low-pass filtered sounds lowered by a constant proportion the performance; high pass filtered sounds and sounds equated for RMS power, instead, were affecting the subjective size of the balls.

When listening to everyday sounds we can extract useful information about the source events that are producing them. Recent research has shown that listeners can recover the length or the shape of an object by hearing the sound it produces in an impact (Carello, Anderson & Kunckler-Peck, 1998; Kunckler-Peck & Turvey, 2000). These results are surprising: hearing is always been considered a dynamic sense (time dependent), thus, uninformative about metrical properties of an object such the length or the shape. Gibson (1963) referred to these properties as 'useful dimension of sensitivity'. In fact, the ability of extracting useful information about objects from their sound could be evolutionary adaptive: the more information we can get about the objects in the environment the more we can adaptively adjust our behaviour. Furthermore, Carello *et al.* (1998) showed that subjects' performances are directly correlated with the physical properties of the source event rather than the acoustical properties of the stimulating wave pressure. Somehow, listeners seem to posses an accurate representation of the sound source (Gaver, 1993a; Gaver, 1993b), consequently, sounds are categorised according to the physical event that is producing them. However, at least some acoustical information need to be coded, in association with visual feedback, in order to achieve such representations. For this reason, the experimental manipulation of recorded ecological sounds could help in identifying which are the most important acoustic features when extracting metrical properties from sounds.

If we listen to the sound produced by a small ball falling on a plate we can easily guess its size even without looking at it. Furthermore if we imagine a big ball and a small ball falling on a same plate from the same height we can predict that the acoustic patterns produced will be different either for temporal structure, level or timbre. Which ball will produce more bounces? The sound of which ball will be highest in level? The timbre of which ball will be more brilliant? Overall: what are the main acoustic clues in order to identify the size of a ball dropped on a plate? According to the distinction set forth by Warren and Verbrugge (1984) everyday sounds are characterisable in two ways: according to their elementary properties, such level and spectral pattern, etc., or according to their *high order structure*, such the temporal structure of the signal (e.g. the number of bounces, the overall duration, etc.). Obviously, larger balls excite a more intense acoustic signal either in peak, in overall level or spectrum level. The temporal structure of the acoustic event would be quite different as well: smaller balls produce more bounces after the first impact than larger balls. Furthermore, bounces of small balls are quite spread in time while bounces of the larger balls are closer in time. For this reason the overall duration of the events is slightly different. Subjectively, larger balls sound louder than smaller balls, furthermore their sound is ‘flat’, on the contrary, the sound of smaller ball is ‘sharp’. Larger balls produce distinctively less bounces, etc..

The aim of the current research is to investigate systematically the effect of either the high order structure, or the elementary properties of sounds such amplitude and spectral content in categorising the size of balls from their sounds. Four experimental conditions and one control condition will be investigated: listeners will perform a size categorisation task either with unmodified sounds, with sounds without bounces, with sounds equated for RMS power, with sounds low-pass filtered or with sounds high-pass filtered. We expect: an impairment in the performance for all experimental conditions compared to the control condition. Furthermore, compared to the control condition, we expect a linear decrement in the performance for experimental conditions that are manipulating aspects of signals that are not necessary in order to obtain a good identification of the size of the object. On the contrary, we expect a non linear impairment for the experimental conditions that are manipulating acoustical parameters necessary for size recognition.

METHOD

Stimuli and apparatus

Stimuli were obtained by recording the sound of a wooden ball (pine) dropped on a crate plate. Balls could weight either 0.33, 1.2, 2.7 or 5.2 grams and plates had a diameter of either 160, 185 or 215 mm. Balls were dropped manually from a wooden frame in the centre of the plate from an eight of 300 mm. Sounds were recorded with a Tascam DA-P1 DAT and a Sennheiser MKH 20 P48 microphone in a quiet room. Recordings were then converted into sound files at 44.1 kHz sample rate and 16 bits resolution. Five conditions were produced, one control condition with the original recordings and four experimental conditions: (1) without bounces: all the bounces after the first impact were deleted; (2) sounds equated for RMS

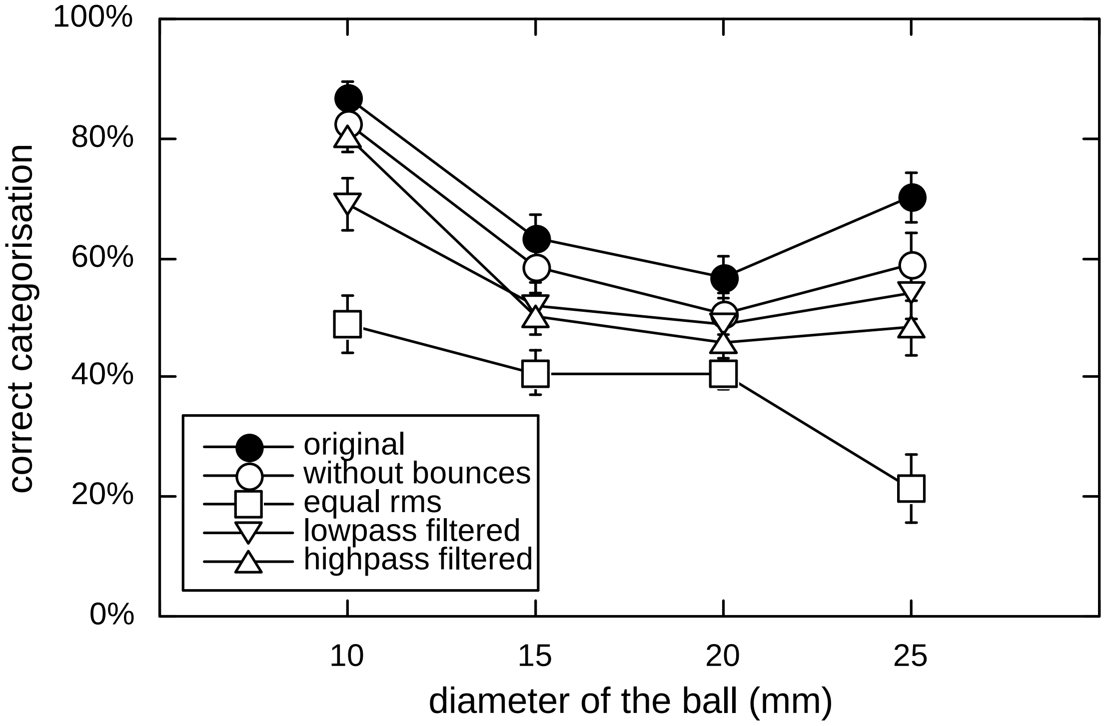

Figure 1: performance as a function of the size of the ball.

power; (3) sounds low-pass filtered with a Butterworth filter of the 4th order and cut-off frequency at 5 kHz; (4) sounds high-pass filtered with a Butterworth filter of the 4th order and cut-off frequency at 5 kHz. Manipulation (2), (3) and (4) were performed on the sounds without bounces. Filtered sounds were equalised to their original RMS power after the filtering. During the experiment sounds were played monophonically through a Yamaha sound card with a Asus computer into a Sennheiser HD 414 headphone.

Procedure

Six-teen undergraduates participated individually in the experiment on voluntary basis. They all reported having normal hearing. The experiment was divided in two sessions of three conditions each. Each session begun always with the control condition followed by two experimental conditions. Within the next day, subject performed the second session where a second control condition and two remaining experimental conditions were ran. The task consisted in categorising the dimension of the ball in a 4AFC task. Subjects were informed that the sounds were obtained with balls of four different size.

RESULTS

Subject answers were converted into percentages of correct categorisation. A 4×3×6 (ball-size, plate-diameter, block) ANOVA was conducted on the percentages of correct categorisation. A contrast between the two repeated control conditions did not show a significant effect: subjects did not improve their performance by repeating the control condition: $F(1, 15)=.356$ $p>.05$. For this reason, results obtained in those two blocks were averaged and a new 4×3×5 ANOVA was performed on the data. The size of the plate did not show any effect on the performance as well: subjects' categorisation was independent from the plate where the balls were dropped $F(2, 30)=.65$ $p>.05$.

The performance in all experimental condition was worse than in the control condition: $F(4, 60)=36.03$ $p<.0001$. In particular, the manipulation of the high order structure (without bounce condition) impaired by a constant proportion the performance in the categorisation: $F(3, 45)=.55$ $p<.05$. Also the performance with low-pass filtered sounds decreased by a constant proportion respect to the results in the control condition: $F(3, 45)=1.42$ $p>.05$. However, the pattern of categorisation changed largely for the sounds equated for RMS power $F(3, 45)=7.98$ $p=.0002$ and slightly for sound high-pass filtered $F(3, 45)=2.56$ $p=.06$ (although the result is not significant it is close to the significance). Under the experimental conditions investigated results suggest that level is the main acoustic clue in judging the subjective size of a ball.

CONCLUSIONS

Listeners showed good skills in categorising the original sound: the average performance was ~70%. Furthermore this ability on recognising the size of the ball was independent from the surface where the ball was impacting: the size of the plate did not affect the performance in the control condition. This result is interesting since could indicate a constancy phenomenon for object's size in hearing perception. In fact, every impact sound is the result of an interaction between an impacting object and an impacted object, therefore, each change in the impacted object results in a different sound produced by the interaction. In the without bounce condition performance was slightly poorer than in the control condition and the highest within the experimental conditions. Moreover, the pattern of responses was similar to the pattern of the control condition. Likely, the bounces following the first impact and their temporal distribution are necessary for the identification of the physical event per se: by listening to the temporal structure of the bounces people can distinguish between a ball dropped on a plate and, for example, a fork dropped on a plate. Warren and Verbrugge (1984) demonstrated that the temporal structure is responsible for a correct identification in comparing breaking and bouncing sounds. In the rest of the experimental conditions performance was largely impaired. In the low-pass filter condition the correct identifications were few. However the pattern of results is still similar to the control condition. Likely, this manipulation was affecting the perceptual quality of the sounds, without changing their identity. Low-pass filter stimulations are quite common in everyday life: when, for example, we listen to something trough a wall the acoustic pattern is low-pass filtered. Human beings are quite use to hear naturally low-passed sounds. For this reason this manipulation did not change the response pattern compared to the control condition. Performance on either the

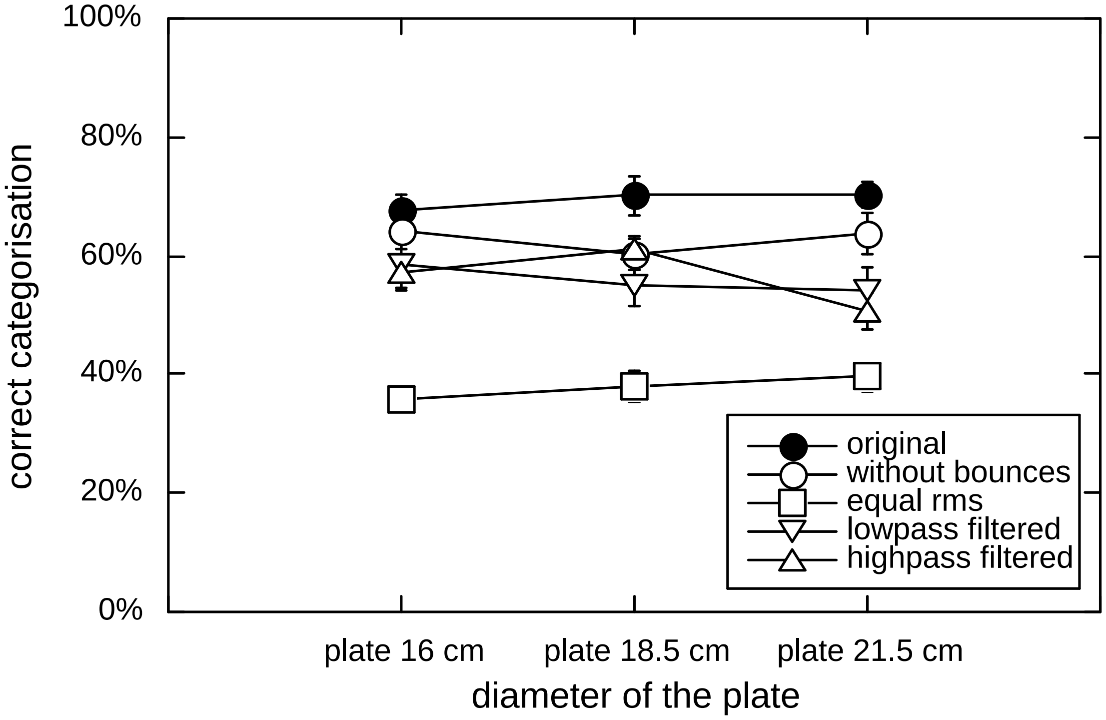

Figure 2: performance as a function of the diameter of the plate.

equal-RMS condition or the high-pass filter condition was, on average, the poorest. Sounds equated for RMS power were sounding similar each other to the subjects. An analysis of the response per se on this condition show that subjects were strongly biased towards the two middle categories: the 15 mm ball and the 20 mm ball. Likely, a task with a categorisation within fewer categories would improve the performance when categories are highly confused each other. The performance with high pass filtered sounds was biased as well: subjects were identifying all sounds as being produced by the smallest ball. The result is not surprising since smallest ball was producing the brighter sounds. In this condition we recorded the largest number of answer ‘10 mm’.

In conclusion, it is still unclear how subject can recognise the size of an object from its sound. Probably listeners select the more evident cues within the set of acoustical cues available. They could, for example, act by selecting in the first place a broad index: ‘this ball is producing many bounces and, therefore, is smaller than the previous one’, then, they can select some acoustical parameter to integrate their impression ‘this ball sounds louder than the previous one’ or ‘this ball sounds brighter than the previous one’. Level and spectral content are likely the main acoustical indicator for size although, a priori, the first is unreliable (a small ball dropped either close to the hear or two metres far is always perceived as a small ball despite the large difference in level of the two signals). According to the context either one or the other can achieve the role of crucial cue.

ACKNOWLEDGEMENT

This work has been supported by the European Commission under contract IST-2000-25287 ("SOb – The Sounding Object": www.soundobject.org).

REFERENCES

Carello, C., Anderson, K. L. & Kunckler-Peck, A. J. (1998). Perception of object length by sound. Psychological Science. 9, 211-214.

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 (1963). The useful dimensions of sensitivity. American Psychologist. 18, 1-15.

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.

Lakatos, S., McAdams & S., Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. Perception & Psychophysics. 59, 1180-1190.

Kunckler-Peck, A. J., Turvey, M. T., (2000). Hearing shape. Journal of Experimental Psychology: Human Perception and Performance. 26, 279-294.

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.