Acta Psychologica 143 (2013) 96–104

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Looking at the world with your ears: How do we get the size of an object from its sound?

Massimo Grassi [a,*], Massimiliano Pastore [b], Guillaume Lemaitre [c]

[a] Dipartimento di Psicologia Generale, Università di Padova, Via Venezia 8, 35131 Padova, Italy
[b] Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Via Venezia 8, 35131 Padova, Italy
[c] Università Iuav di Venezia, Dorsoduro 2206, 30123 Venezia, Italy

ARTICLE INFO

Article history:
Received 30 October 2012
Received in revised form 7 January 2013
Accepted 20 February 2013
Available online 28 March 2013

PsycINFO classification:
2326

Keywords:
Auditory cognition
Ecological acoustics
Size estimation

ABSTRACT

Identifying the properties of on-going events by the sound they produce is crucial for our interaction with the environment when visual information is not available. Here, we investigated the ability of listeners to estimate the size of an object (a ball) dropped on a plate with ecological listening conditions (balls were dropped in real time) and response methods (listeners estimate ball-size by drawing a disk). Previous studies had shown that listeners can veridically estimate the size of objects by the sound they produce, but it is yet unclear which acoustical index listeners use to produce their estimates. In particular, it is unclear whether listeners listen to amplitude (related to loudness) or frequency (related to the sound's brightness) domain cue to produce their estimates. In the current study, in order to understand which cue is used by the listener to recover the size of the object, we manipulated the sound source event in such a way that frequency and amplitude cues provided contrasting size-information (balls were dropped from various heights). Results showed that listeners' estimations were accurate regardless of the experimental manipulations performed in the experiments. In addition, results suggest that listeners were likely integrating frequency and amplitude acoustical cues in order to produce their estimate and although these cues were often providing contrasting size-information.

1. Introduction

The goal of auditory scientists is to understand how the sound waves are coded in/by the auditory system. For this reason, researchers are more interested in the proximal stimulus (i.e., the sound wave that reaches the eardrum) rather than the distal stimulus (e.g., the physical event that produces that sound wave). However, experimental studies have shown that naïve listeners spontaneously identify, categorize, and describe the distal rather than the proximal stimulus (Houix, Lemaitre, Misdariis, & Susini, 2012, Lemaitre, Dessein, Aura, & Susini, 2011; Lemaitre, Houix, Misdariis, & Susini, 2010), and a growing number of investigations on the perception of the distal stimulus have been reported in the recent past. This class of investigations, sometimes labeled as "ecological acoustics" (EA from now on in the text), studies whether and how it is possible to understand the characteristics of the distal stimulus by listening to the sound (i.e., the proximal stimulus) that this stimulus produces. For example, here the experimenter dropped a ball on a plate in one side of a panel-divided room. Successively, he asked the listener, sitting in the other side of the room, to estimate how large the ball was only on the basis of the sound produced by the ball–plate impact. Note that in the current study the listener was provided with no foregoing information about the apparatus producing the sound.

EA investigations have shown that listeners can recover a variety of sound source properties, although performance may vary depending on each attended particular piece of information. In general, listeners identify interactions between objects with great accuracy (Lemaitre & Heller, 2012). For example, they can distinguish almost perfectly whether a bottle dropped on the floor is breaking or bouncing (Warren & Verbrugge, 1984) or fill a vessel to the brim without overflowing just by listening to the resulting sound (Cabe & Pittenger, 2000). When sound source events are recognized (even implicitly), audition drives vision in resolving ambiguous visual motion patterns (Grassi & Casco, 2009, 2010) or a person's hand in grasping objects (Castiello, Giordano, Begliomini, Ansuini, & Grassi, 2010). However, in other cases, listeners' responses seem driven by stereotypical, potentially misleading, associations (Ballas, 1993; Giordano, McDonnell, & McAdams, 2010; Gygi, Kidd, & Watson, 2004, 2007; Repp, 1987). For instance, Repp (1987) showed that listeners systematically associated slow, low, and loud handclaps to male clappers even though handclaps produced by male hand clappers do not always follow this pattern. Giordano and McAdams (2006) showed that subjects are often unable to distinguish the material (metal and glass) of struck objects, but instead systematically associate small objects to glass and large objects to metal, regardless of their actual material.

* Corresponding author. Tel.: +39 0498277494.
E-mail address: massimo.grassi@unipd.it (M. Grassi).

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.02.005

Impacts are the most-studied category of sound events. Impact sounds result from the single or repeated contact between two (or more) solid objects (Gaver, 1993a, 1993b). By listening to impact sounds, listeners can identify the gender and the gait of a walker (Li, Logan, & Pastore, 1991; Pastore, Flint, Gaston, & Solomon, 2008) and, by the same token, walkers can identify the material they are walking upon (Giordano et al., 2012). Listeners can report to some extent the material of the two objects involved in an impact (Giordano & McAdams, 2006; Tucker & Brown, 2003) or their hardness (Freed, 1990; Giordano, Rocchesso and McAdams, 2010). But listeners are also able to provide metrical estimations of the objects involved in an impact such as the length of a rod dropped on the floor (Carello, Anderson, & Kunkler-Peck, 1998), the size of a ball dropped on plates of various dimensions (Grassi, 2005), or the shape of a plate struck by a pendulum (Kunkler-Peck & Turvey, 2000). Noticeably, in many of these experiments the listener is provided with few-to-none foregoing information about the sound source event.

The above studies report a remarkable finding: the listeners' response is veridical (i.e., accurate) in comparison to the distal stimulus. This finding is surprising. The listener is listening to only *one* sound. This sound, however, is the result of the contact between at least *two* objects. In the majority of impact sounds, these objects do not contribute equally to the resulting sound. A prototypical example is that of the drum and the drumstick: it is the drum that sounds, not the drumstick. Therefore, in many impacts, one object can be classified as the "sounding object" (SO) whereas the other can be classified as the "non-Sounding Object" (nSO). When the experimenter asks questions about the nSO, the veridicity of the listener's response is impaired (Giordano, Rocchesso, et al., 2010) and can be influenced by unrevealed changes in the SO (Grassi, 2005). For example, Giordano, Rocchesso, et al. (2010) found that listeners judging the hardness of the nSO needed twenty-five experimental sessions to reach a certain (mediocre) level of accuracy. On the contrary, listeners estimating the hardness of the SO reached a higher accuracy after only five sessions. Grassi (2005) found that listeners could provide veridical estimates of the size of a ball (nSO) dropped on a plate (SO). However, when the plate was changed without informing the listener, the estimations also changed: the same ball was estimated to be smaller or larger according to the size of the plate that it was dropped upon. In summary, these results suggest that listeners can give veridical estimate of both the size SO and the size nSO. However, estimations of the latter are more difficult as well as susceptible of being influenced by unrevealed changes in the SO.

The current study reports four experiments in which listeners estimated the size of a ball (nSO) being dropped on a single plate (SO) from various heights. In the previous study investigating this same sound source event, Grassi (2005, but see also Grassi, 2002) hypothesized that listeners attended to the sound's loudness to produce their estimates: the louder the sound, the larger the ball. Loudness addresses mainly to the intensity content of a sound, i.e., its amplitude content. It is often represented by the power of the sound averaged across its duration, i.e., the RMS power. However, that same study, as well as other studies (e.g., Carello et al., 1998; Giordano, Rocchesso, et al., 2010; Kunkler-Peck & Turvey, 2000), suggested an alternative hypothesis. Listeners could have attended to the sound's timbre and, in particular, to the sound's brightness to produce their estimates: the duller the sound the larger the ball. Brightness is mainly related to the frequency content of a sound and it is often represented by the spectral centroid (Grey & Gordon, 1978; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). In Grassi (2005) acoustical indices such as sound power and spectral centroid co-vary coherently so that it was impossible to disentangle which of the two acoustical cues was used by listeners to estimate size. However, if falling height is manipulated sound power and spectral centroid vary one independently from the other (see next paragraph and Appendix A for a detailed explanation) thus enabling to understand whether listeners exploit one or the other cue for estimating size. In the next section, we develop a simplified account of the physics of a falling ball and its theoretical consequences on the resulting sound. This will highlight how the acoustical cues vary as a function of the size of the balls and the height of the fall.

2. Acoustical cues potentially available to estimate the size and the falling height of a ball dropped on a plate

The sound produced by a plate impacted by a ball can be modeled by a sum of exponentially decaying sine waves (Wildes & Richards, 1988, chap. 25). In the current study, because plate did not change across experiments, the RMS power in dB of the first impact (P_{dB}) depends only on the logarithm of the diameter and on the logarithm of the falling height, as expressed by Eq. (1) (the complete demonstration is shown in Appendix A).

$$P_{\mathrm{dB}} = 30\log_{10}(d) + 10\log_{10}(h) + k \tag{1}$$

In this equation, d is the ball's diameter, h is the height of the fall and k (here and in the following equations) represents an undetermined constant. Because sound power in dB is a good index of a sound's loudness (Boullet, 2006; Stevens, 1955) the listeners of Grassi (2005) could estimate veridically the size of the ball simply by scaling the loudness of the impacts (balls were dropped from a fixed height). However, because here the falling height was changed within each experiment, loudness alone is not anymore an effective cue to estimate size: listeners using only loudness would overestimate ball size when falling height increases and underestimate it when height decreases. Note that Eq. (1) provides a precise prediction of the participants' estimate (assuming their estimation be only based on power): doubling the ball's diameter results in a size estimate three times larger than doubling the height of the fall.

Together with sound power, frequency domain cues also vary as a function of the size of the ball and the falling height. For example, the heavier the ball, the longer the time of contact between the ball and the plate (Avanzini, 2001). A longer contact time dampens vibrations of the plate whose periods are shorter than the time of contact itself. As a consequence, the sound is bright for small balls and dull for large ones. In addition, given a ball of mass m, the higher the fall the higher the energy of the event. Extra-energy can set into vibration the high frequency modes of vibrations of the plate, which require more energy than low frequency ones to be set into motion. As a consequence, the sound of a given ball is slightly brighter when it falls from a higher height. As we anticipated, the perceived brightness of a sound can be estimated by the spectral centroid that is the sum of the frequency components of a sound's spectrum weighted by their relative amplitude.

A change in the size of the balls and in the height of the fall affects also temporal cues and these cues are also potentially available to the listener. For instance, the time between the two first impacts (B2B) is directly related to the falling height (see Appendix A for a complete demonstration):

$$\log_{10}(\mathrm{B2B}) = \frac{1}{2}\log_{10}(h) + k \tag{2}$$

Note that, because of this, listeners can potentially separate the contribution of the falling height to the loudness of the impact, and veridically estimate the size of the ball independently from the falling height. Combining Eqs. (1) and (2) results in an index given by Eq. (3).

$$30\log_{10}(d) = P_{\mathrm{dB}} - 20\log_{10}(\mathrm{B2B}) + k. \tag{3}$$

Such a combined index could be considered as an invariant index of the size of the balls. However, this is true only with a simplified account of the ball's dynamics. In fact, the time between impacts also partially depends on the size of the balls so that the relation between loudness, size, and height differs in practice from this simple account (for a detailed description of the dynamics of a ball's bounce see Cross, 1999).

In summary, as the height of the fall increases, the sound of a given ball becomes louder (which is consistent with a "larger" ball) but also "brighter" (which is consistent with a "smaller" ball). Furthermore, bounces become more spaced in time (which is consistent with a "smaller" ball). If listeners attend to loudness to estimate the size of the ball they should not be able to estimate the size of the ball independently from the falling height. However, results can be different if listeners attend to the sound's brightness or if they combine several cues together (e.g., combining loudness and B2B). In the next section, we record and analyze the sounds of the balls using the apparatus of behavioral experiments. The analysis will enable to assess whether the above theoretical predictions are empirically corroborated.

3. Sound source event and acoustical event

The balls were seven solid wooden balls (pine) of 1, 1.5, 2, 2.5, 3, 4, and 5 cm in diameter, weighing respectively 0.35, 1.1, 2.9, 4.9, 8.6, 22.2, and 44.5 g (density, ~647 kg/m^3). The plate was a baked clay plate of 18.5 cm in diameter. The balls were dropped from 3 cm, 6 cm, and 12 cm. This height range was sufficiently large so that all balls produced at least one audible rebound, but also sufficiently narrow so that balls did not bounce out of the plate when bouncing. The balls were dropped in real time during the behavioral experiments. However, a large number of recordings were preliminarily made to investigate the acoustic cues potentially available to the listener. The recordings were made with a Sennheiser MKH 40 P48 microphone and a portable Tascam DA-P1 DAT recorder. Recordings had a 48-kHz sample-rate and a 24-bit resolution. A minimum of 30 exemplars was recorded for each combination of size and height. The ball was dropped manually onto the middle of the plate by the experimenter and it was allowed to bounce freely until its motion ended. The ball was always held in the same way before the drop to minimize variations. The plate was set on top of a foam block (40 by 50 cm by 4 cm) placed on a table. Therefore, the plate could vibrate freely after each impact, without transmitting the vibration to the table. The foam block just slightly damped the vibration of the plate, because the plate's reverse was not completely flat but finished with a circular edge 0.3 cm high and a 12.5 cm diameter (see Fig. 1).

Three acoustical parameters were extracted from the recordings: the power of the first impact averaged across its duration (the first 50 ms of the sound after the sound's amplitude peak) expressed in dB, the logarithm of the time between the first and the second bounce (i.e., B2B), and the logarithm of the spectral centroid of the first impact (calculated on the first 50 ms after the impact). An analogous procedure was used to record and analyze the stimuli of Experiment 4 (see later for a detailed description of this experiment). The acoustical values of these stimuli are reported in Appendix B.

Fig. 2 represents the RMS power in dB, the time between bounces, the spectral centroid, and the index described in Eq. (3), averaged across the different recordings for each diameter of the ball and each falling height. The figure shows that power fits well a linear model based on the logarithm of the size of the ball and of the falling height. Similarly, the time between bounces fits a linear model that mainly depends on the falling height, but that also changes as a function of the size of the ball (therefore structurally similar to that given by Eqs. (1) and (2)). The bottom left panel of the figure represents the spectral centroid. The spectral centroid is mainly affected by the size of the ball, although the predicted effect of falling height can be observed, making this cue another potential cue that subjects could use to estimate the size of the ball.

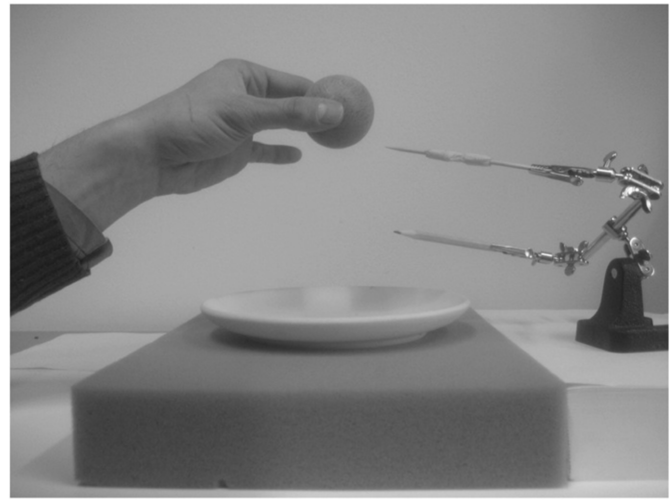


Fig. 1. Example of the apparatus. The photo depicts the apparatus used in Experiment 2. On the right side there are two reference points that were used by the experimenter to drop the ball.

We revised Eq. (3) (i.e., the model that integrates sound power and B2B) by estimating the parameters of the linear model for each cue, which lead to Eq. (4).

$$30\log_{10}(d) = P_{\mathrm{dB}} - 37\log_{10}(\mathrm{B2B}) + k$$
$$30\log_{10}(d) = P_{\mathrm{dB}} - 37\log_{10}(\mathrm{B2B}) \quad (4)$$

The weighting of the time between impacts is larger than predicted by the model. The lower right panel of the figure represents the empirical statistics of this index. For all but the largest ball, this index is a good predictor of the diameter of the ball, which does not depend on the falling height. Listeners using this index would therefore reach a better estimation of the size of the ball than if they would be only using the loudness of the first impact or the time between bounces.

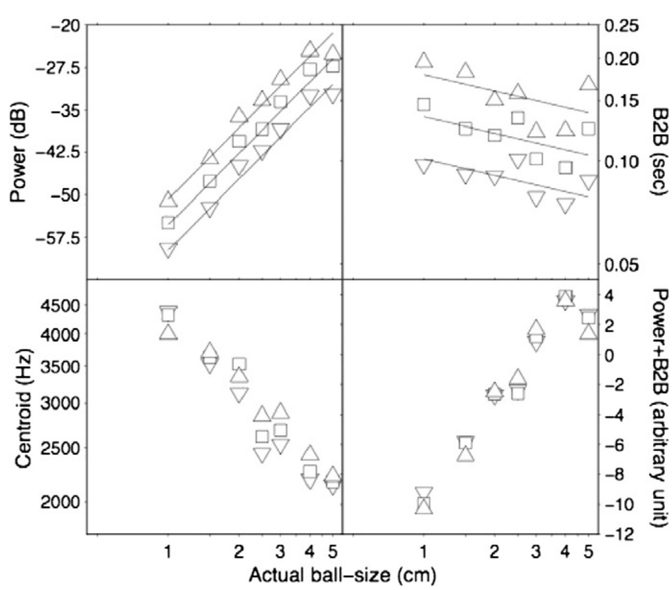


Fig. 2. Upper panels: RMS level and B2B as a function of the ball diameter and the falling height. The straight lines represent the linear model. Lower right panel: A potential index representative of the size of the ball, as a function of the size of the ball and the falling height. This index combines the sound power of the first impact and B2B, and is defined by Eq. (4). Lower left panel: Spectral centroid as a function of the ball diameter and the falling height. In all graphs, pointing down triangles represent h = 3 cm, squares represent h = 6 cm and pointing up triangles represent h = 12 cm. In all panes, standard errors are too small to be seen.

Overall, the analysis of the recordings corroborates the theoretical predictions, although, of course, empirical acoustical cues are characterized by a certain variance. In Experiments 1 and 2, we investigated the effect of a change in the ball's size and a change in the height of the fall on the size estimate produced by the listener. In these experiments, balls were dropped from either two (Experiment 1) or three (Experiment 2) heights. If listeners use loudness to estimate the size of the ball we expect estimations to increase as a function of the height of the fall. In contrast, if they use the sound's brightness we expect estimations to increase only as a function of the size of the ball.

4. Experiment 1

In Experiment 1 listeners were asked to estimate the size of the balls by listening to the sound they produce when impacting upon a plate. The balls could be dropped from two different heights. Listeners received no foregoing information about the sound source event.

4.1. Method

Fourteen listeners (4 females) participated in the experiment. They all reported normal hearing. The apparatus described in the previous section was used for the experiment. However, balls could be dropped from either 6 cm or 12 cm. Here, as well as in the successive experiments, the entire apparatus and the experimenter were 2 m away from the listener and hidden from the listener's view by a wooden 2.0 by 2.0 m removable, black panel.

Before the experiment, the experimenter dropped a ball, randomly selected from the ball set. Then, he asked the listener if s/he could tell the shape of the object that had just been dropped. The experiment then began. The experiment consisted of fifty-six trials resulting from presenting four times each of the seven balls at each of the two falling heights. Trials were presented in random order. In each trial, the same ball was dropped three times from the same height and each time was allowed to bounce freely until its motion ended. This triplex presentation of the same stimulus was adopted to provide the listener a sufficient exposition with the stimulus (the sound produced by the first two/three bounces is rather fleeting) and to allow the listener to accomplish his/her task in real time, i.e., while listening to the sound. The listener's task was to create and adjust a disk on a computer screen as large as the ball they thought was been dropped. At the beginning of each trial the computer screen was blank. A custom software allowed disks to be drawn ranging from ~0.05 cm up to ~30 cm of diameter. The listener controlled the disk-drawing by means of the keyboard. At the end of the experiment, and in all subsequent experiments, listeners were asked questions about the event. For example, listeners were asked to report how many different heights were used in the experiment.

4.2. Results and discussion

All the listeners replied to the first question that a spherical object had been dropped. The mean size estimates were calculated for each ball-size and each height and separately for each listener. The resulting values were log10 transformed and subjected to a 2 (height of the fall) by 7 (size of the ball) two-way analysis of variance. The size estimate increased as a function of the mass of the ball, $F(6, 78) = 383.69$, $p < .0001$, and as a function of the height of the fall, $F(1, 13) = 8.84$, $p = .011$. The interaction was not significant $F(6, 78) = 1.64$, $p > .05$. The effect size revealed that the influence of the ball's mass on the estimation was larger than the influence of the height of the fall: respectively, $\eta_p^2 = .967$ and $\eta_p^2 = .405$. In order to measure this difference in effect size, we calculated the increment in the subjective estimates resulting from a doubling of the size of the ball or from a doubling of the height of the fall: doubling the size of the ball had an effect nine times larger than doubling the height on the subject's estimate. Furthermore, we assessed how accurate was the listeners' scaling of the balls' sizes (in comparison to actual sizes) with two linear regressions run on log10-estimates, separately for the two heights of the fall. In both cases fit was very high (h = 6 cm, $F(1, 5) = 716.28$, $p < .0001$, $R^2 = .993$, slope = 1.28, intercept = −0.32; h = 12 cm, $F(1, 6) = 754.45$, $p < .0001$, $R^2 = .993$, slope = 1.29, intercept = −0.28). The listeners' estimations are shown in Fig. 3 (left) as a function of the log10 diameter of the ball. Finally, subjects thought that balls were dropped from more than one height.

In the current experiment, listeners were able to scale the size of the ball coherently with the actual balls' size although they showed a general underestimation of the actual balls' size. Listeners, however, were influenced by the height of the fall and judged the ball larger when the ball was dropped from a higher height. It is possible that the height of the fall did not affect much the estimations because listeners thought that balls were dropped from more than one height and weighted the estimations according to this belief. An increase of size estimations with an increase of falling height is consistent with the interpretation that subjects used loudness to estimate the size of the ball. To study how precisely does falling height influence size estimation, the next section reports the results of a similar experiment using several heights.

5. Experiment 2

In the previous experiment, listeners were able to scale the size of the balls coherently with the actual balls' size. However, listeners' estimations were affected by the height of the fall: the same ball was judged slightly larger (or slightly smaller) if it was dropped from a higher (or smaller) height. The current experiment extends the findings of Experiment 1 and further tests the effect of the height of the fall on the size estimate produced by the listener. Here, balls could be dropped either from 3, 6, or 12 cm.

5.1. Method

Fifteen new listeners (4 females) participated in the experiment. They all reported normal hearing. The apparatus and procedure were identical to Experiment 1 except that here balls were dropped also from a height of 3 cm and that here each stimulus (single combination ball-size dropping height) was presented three times to the listener totalling sixty-three trials.

5.2. Results and discussion

All the listeners replied to the first question that a spherical object had been dropped. The mean size estimates were calculated separately

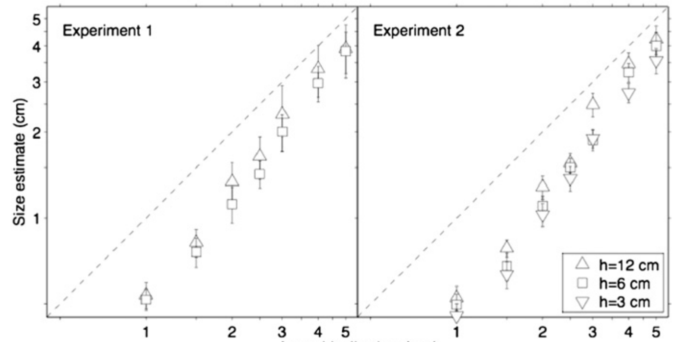


Fig. 3. Subjective size as a function of the actual size of the ball. The results of Experiment 1 are represented on the left panel. The results of Experiment 2 are represented on the right panel. Vertical bars are ±1 standard error of the mean.

for each listener, each ball and each height of the fall. The resulting values were log10 transformed and subjected to a 3 (height of the fall) by 7 (size of the ball) two-way analysis of variance. The size estimate increased as a function of the mass of the ball, $F(6, 84) = 283.83$, $p < .0001$, and as a function of the height of the fall, $F(2, 28) = 15.83$, $p < .0001$. The interaction was not significant $F(6, 66) = 1.64$, $p > .05$. The effect size revealed that the influence of the ball's mass in the estimation was larger than the influence of the height of the fall: respectively, $\eta_p^2 = .943$ and $\eta_p^2 = .531$. Doubling the size of the ball had an effect ten times larger than doubling of the height on the subject's estimate. The linear regressions showed that listeners' estimations were highly correlated to the actual balls' size (respectively, h = 3 cm, $F(1, 5) = 624.07$, $p < .0001$, $R^2 = .992$, slope = 1.34, intercept = −0.38; h = 6 cm, $F(1, 5) = 401.57$, $p < .0001$, $R^2 = .988$, slope = 1.36, intercept = −0.35; h = 12 cm, $F(1, 6) = 412.27$, $p < .0001$, $R^2 = .988$, slope = 1.36, intercept = −0.30). The listeners' estimations are shown in Fig. 3 (right) as a function of the log10 diameter of the ball. As in Experiment 1, listeners thought that balls were dropped from more than one height.

In the current experiment, as well as in Experiment 1, listeners scaled accurately the size of the ball, i.e., size estimations increased linearly (in the log–log domain) with the actual size of the balls, even if the slope and the intercept of the regressions showed that subjects had a tendency to underestimate the smallest sizes. Listeners were, however, affected by the height of the fall and judged the ball larger when it was dropped from a higher height. The results of the current experiment corroborate the findings of Experiment 1, namely, that the size estimate was affected by the height of the fall, therefore, that listeners attended to loudness to estimate the size of the ball. However, the results of the current and the previous experiment did not clarify the actual effect of the height of the fall on the size estimation. Listeners' responses were affected by the manipulation of this factor. However, the effect of the manipulation was unclear: results did not clarify whether the height of the fall affected the estimations in a fixed way (i.e., identical regardless the height value) or in a proportional way (i.e., proportional to the values of the height) because both explanations were compatible with the results. Experiment 3 was designed to answer this question.

6. Experiment 3

In Experiment 3 we investigated the way the height of the fall affects the size estimate produced by the listener. The results of Experiments 1 and 2 did not clarify the real effect and suggest a double interpretation. On the one side, it is possible that the increment in height returns a fixed increment in the listener's estimation regardless the absolute value of the height of the fall. On the other, it is possible that the height of the fall modulates proportionally the size estimate produced by the listener. In the current experiment the listeners performed the estimations in two sessions. In the first session, balls were dropped from either 3 or 12 cm, in the second session, from either 6 or 12 cm (or vice versa). If the height of the fall has a fixed effect on the estimation we expect the increment in the size estimation due to the height of the fall to be independent of the experimental session (i.e., similar in the two sessions). On the contrary, if the height of the fall modulates proportionally the size estimate, we expect the estimates of the first session to reveal a greater effect of the height than the estimates of the second session.

6.1. Method

Fourteen new listeners (4 males) participated in the experiment. They all reported normal hearing. The apparatus was identical to that used in the previous experiments, although in the current experiment a different procedure was used. The listeners performed the experiment in two sessions and the two sessions were performed in two consecutive days. In the first session the listener listened to the balls dropped from a height of either 3 cm or 12 cm (or 6 cm, 12 cm). In the second session the listener listened to the balls dropped from either 6 cm or 12 cm (or 3 cm, 12 cm). The order of the sessions was counterbalanced across listeners. At the end of the second session the listener was asked questions about the experiment such as in previous experiments. In particular, the listener was asked whether s/he had noted any difference between the first and the second session.

6.2. Results

All the listeners replied to the first question that a spherical object had been dropped. The mean size estimates recorded in the first and in the second sessions were calculated separately for each listener and sound source event. Successively, for each listener we calculated the difference between the size estimate for the 12-cm height drop and the 3 (or 6) cm height drop. These differences were subjects to a 2 (difference in height) by 7 (mass of the ball) two-way analysis of variance. The listeners' estimates were more different when they evaluated the stimuli of the 3–12 cm session than when they evaluated the stimuli of the 6–12 cm session: $F(1, 13) = 7.86$, $p = .015$. The factor mass of the ball was not significant, $F(6, 58) = 1.97$, $p > .05$. Finally, the interaction was not significant showing that the differences measured in the two sessions were roughly constant, i.e., independent from the size of the ball: $F(6, 68) = 0.29$. Interestingly, the majority of listeners (9 out of 14) asserted that the two sessions were identical.

The results of the current experiment corroborate the findings of Experiments 1 and 2, namely, that the estimate is affected by the height of the fall, therefore, that listeners are likely to use a loudness-related cue such as sound power to estimate the size of the ball. In addition, the results of the current experiment explain the effect of a change in the height of the fall on the size estimate produced by the listener: the size estimate is proportionally affected by a change in the height of the fall.

All together, the results of the current and the previous experiments suggest that listeners were not able to separate completely the contribution of the ball size and the falling height, as would be the case if they compensated loudness by the time between bounces. A radical hypothesis is therefore that subjects used only a loudness cue (such as power) to estimate size. To investigate such an idea, Experiment 4 used sounds with an equal energy, thus equal sound power (thus similar loudness) but produced by balls of different sizes.

7. Experiment 4

In all previous experiments listeners' estimations were larger when the balls were dropped from a higher height. This result shows, therefore, that listeners were likely to use a loudness cue such as power to produce their estimates. In the current experiment, we canceled the information carried by amplitude domain cues (i.e., those related to loudness) to understand the role of these cues in the estimation of the size of the ball. The three small balls (or three of the large balls) were dropped each from a different height so that sound events were identical for energy, thus, identical for the power of the resulting sound (see Appendix A). If listeners use power only to estimate the size of the ball they should be unable to make any distinction in the size estimate of the three smallest (or largest) balls. On the contrary, if listeners use more than one acoustical cue to produce the estimate they should be able to scale coherently the size of the ball within each energy group.

7.1. Method

Twelve new listeners (1 male) participated in the experiment. They all reported normal hearing. The apparatus for the current experiment was identical to the previous experiments however, in the current experiment the largest ball was not used. In the current experiment, each ball was dropped from a different height. The three smallest balls (i.e., 1, 1.5, and 2-cm Ø) were dropped from, respectively 20, 6.4, and 2.4 cm. Therefore, the corresponding events were all characterized by an energy of ~68 erg (68×10^{-7} J). The three largest balls (i.e., 2.5, 3, and 4 cm Ø) were dropped from, respectively 14.3, 8.1, and 3.2 cm. Therefore, the corresponding events were characterized by an energy of ~680 erg (i.e., 680×10^{-7} J). Each event was presented five times to the listener in random order. The listeners responded by drawing the disks as in previous experiments. At the end of the current experiment listeners were asked questions about the sound source event.

7.2. Results

Listeners' size estimations were averaged separately for each listener and each event and were subjected to a 2 (levels of energy) by three (size of the ball) two-way analysis of variance. The size estimate was large for the high energy events and small than for the low energy events, $F(1, 11) = 161.02$, $p < .0001$. Moreover, within each energy level, the estimates were different: $F(2, 22) = 70.85$, $p < .0001$. The listeners' estimates are represented in Fig. 4 as a function of the actual balls' size. In the current experiment, all listeners thought that balls were dropped from more than one height.

The results of the current experiment showed that listeners could scale the size of the three balls of each energy group, even without size information provided by amplitude domain cues. Therefore, results suggested that listeners used other cues to estimate size (e.g., the sound's brightness represented by the spectral centroid). However, results also showed that the size difference between the 2-cm ball and the 2.5-cm ball was larger than in the previous experiments. In the current experiment the difference in size between these two balls was more than 1 cm, whereas the same difference was much smaller in previous experiments (i.e., 0.3 cm). The large difference observed here might be due to the large difference in sound power produced by these two balls in the current experiment. The 2-cm ball was dropped from 2.4 cm, whereas the 2.5-cm ball was dropped from 14.3 cm. The large difference in heights resulted also in a large difference in the sound power produced by the two balls and, consequently, in a large difference in the subjective size of these two balls.

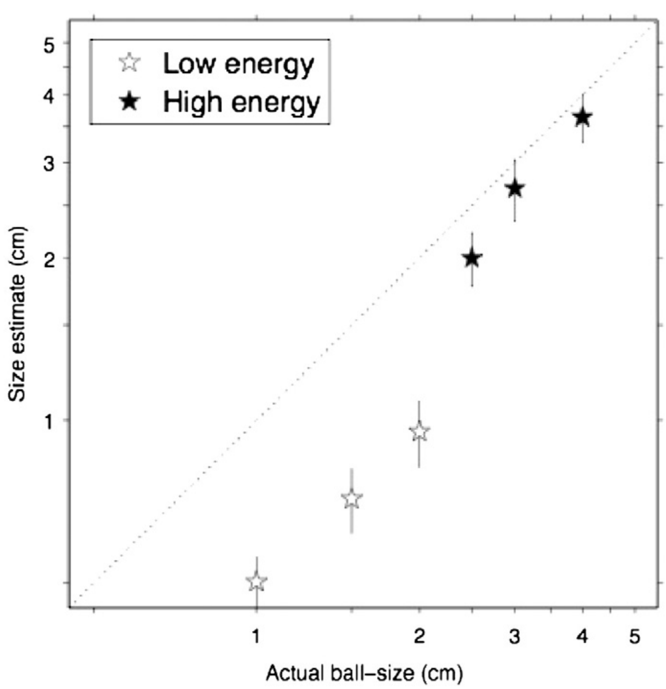


Fig. 4. Experiment 4. Size estimates as a function of the size of the ball. The diagonal dashed line shows the perfect estimate. Vertical bars represent ± 1 standard error of the mean.

In summary, the results of the current experiment suggested that listeners could estimate the size of the balls even when the information carried by the sound power was not available, thus suggesting that other cues were exploited in the size estimation. However, results also suggested that listeners listened to amplitude domain cues (i.e., those related to loudness) when estimating size. In summary, listeners integrated more than one cue together (and although these cues are contrasting) to estimate the size of an object from its sounds.

8. From sound the physical and subjective size

In the current section, we analyzed statistically (1) the relationship between the actual ball's size and the acoustical predictors and (2) the relationship between acoustical predictors and subjective size to understand, respectively, whether the acoustical predictors could predict the size of the ball and whether subjective estimates could be predicted by one (or more) of these predictors.

8.1. The relationship between ball's size and acoustical predictors

We performed a multiple linear regression between the log10-ball's size and the three acoustical predictors, i.e., sound power, B2B and spectral centroid. These analyses were not conducted for Experiment 3 because the sound source event of this experiment was identical to that of Experiment 2. Overall, acoustical indices predicted well the log10-size of the ball with R^2 ranging from 0.97 (Experiment 2) and 0.98 (Experiments 1 and 4). Residuals standard errors were ~0.03 log10 units (thus, ~1 cm) for all experiments. In synthesis, acoustical predictors enabled to scale veridically the balls' sizes and to estimate their sizes with an absolute error ranging within 1 cm. Successively, we also estimated the effect size of each predictor in order to understand which of the three acoustical indices was the best in fitting the balls' sizes. The sound's power was the best predictor of balls' size in all experiments ($0.57 < \eta_p^2 < 0.66$), the spectral centroid was as good as power for Experiments 1 and 2 (respectively, $\eta_p^2 = 0.65$ and $\eta_p^2 = 0.64$) whereas it was a poorer predictor for the data of Experiment 4 ($\eta_p^2 = 0.20$). In contrast, B2B was a relatively poor predictor across all experiments: $0.11 < \eta_p^2 < 0.36$.

8.2. The relationship between acoustical predictors and subjective size

We analyzed the relationship between acoustical predictors and subjective size with a linear model including all acoustical predictors. Such a model predicted well the data of all experiments: respectively, Experiment 1, $R^2 = .66$, Experiment 2, $R^2 = .74$, and Experiment 4, $R^2 = .69$. We also analyzed how much variance explained each of the acoustical indices by estimating each predictor's effect size. The sound's power was the best predictor of subjective size in all experiments ($.11 < \eta_p^2 < .18$). In contrast, spectral centroid and B2B were both poor predictors of the subjects' estimations (all $\eta_p^2 > .06$). In order to assess whether there was any individual difference in the relationship between subjective size and acoustical predictors we used the linear mixed model approach (Pinheiro & Bates, 2000), adding the subject as a random factor together with each singular estimate of the stimuli. Although individual estimates differed in absolute value, the distribution of the estimates across the predictors was similar for all subjects. In detail, we observed two, although small, sources of variation: the between-subjects intercepts and slopes. In Experiment 1, the standard deviation of the subjects' intercepts was 0.16 log 10 cm (i.e., 1.45 cm) and that of the subjects' slopes was 0.02 (i.e., .05 cm). In Experiment 2, the standard deviation of subjects' intercepts was 0.09

(i.e., 1.23 cm), and that of subjects' slopes was 0.02 (i.e., 1.05 cm). Finally, in Experiment 4 the standard deviation of the subjects' intercepts was 0.13 (i.e., 1.35 cm), and that of the subjects' slopes was 0.02 (i.e., 1.05 cm).

9. General discussion

In the current study we investigated how listeners can provide a good metrical estimation of the size of an object by just listening to the sound that this object produces when impacting a second object. In doing this, the current study extends the findings of a previous study that investigated the estimation of the size of a ball when dropped upon a plate (Grassi, 2005). There, listeners were able to scale coherently the size of the ball: listeners' estimations were veridical (i.e., close to the actual size of the ball) and were distributed similarly to the actual balls' sizes. In addition, some of the acoustical characteristics of the sound were found to be informative about the actual size of the ball and could be exploited by the listener to estimate size. However, in that study, the size-informative cues co-varied coherently and it was therefore impossible to understand which cue listeners actually listened to in order to provide a size estimate of the ball. The rationale moving the current study was to manipulate the sound source event in such a way that those acoustical parameters could be changed independently one from the other.

In the current study listeners were able to scale coherently the size of the ball in all experiments. In all experiments the ball was estimated larger when the ball's diameter was larger but also when the height of the fall was higher, especially in Experiments 1, 2 and 3. The effect of the height of the fall on the listener's estimate was revealed by Experiment 3. This experiment showed that the height of the fall has a proportional effect on the listener's estimate, i.e., the higher is the height the larger the estimate. In sum, results suggest that listeners use amplitude-related acoustical cues (such as power) to estimate the size of the ball. These indices, in fact, are the only ones predicting an increment in the size estimate both as a function of ball's size and as a function of falling height. Results, however, show also that the contribution of the height of the fall to the listener's estimate is nine/ten times smaller than that of the actual ball's size, i.e., three times smaller than that we could expect if size estimation was dependent only on sound power. This result is also stressed by the statistical effect size that, in Experiments 1 and 2, was larger for the “ball-size factor” than for the “height factor”.

However, results also show that amplitude domain cues are not the only acoustical indices that subjects use. In Experiment 4, subjects were able to scale well the three balls within the two energy levels used in the experiment, i.e., without any information by acoustical cues such as power. This result suggests that power alone is not sufficient to explain the listener's response and that listeners use at least some other acoustical cue to produce the size estimate, for example, a frequency domain cue that here was represented by the spectral centroid. Note that we believe that subjects exploited frequency cues (and not temporal cues) to estimate size for two main reasons. Firstly, frequency cues are much better predictors of the actual size than temporal predictors (see previous section). Secondly, literature suggests that frequency (and not temporal) cues are used by listeners to estimate size from sound (e.g., Carello et al., 1998; Giordano, Rocchesso, et al., 2010; Grassi, 2005; Kunkler-Peck & Turvey, 2000). There are, of course, several alternative explanations for the small (or lack of) effect of falling height. For example, it is possible that listeners understood that the falling height was manipulated (as suggested by the subjective reports at the end of the experiment) and attenuated the “weight” of sound power in the estimate, although weighing was not optimal.

The results of the current experiment show that listeners can provide estimations about an object, the nSO that does not contribute as much to the resulting sound. It is maybe for this reason that the perception of the characteristics of this object can be influenced by unrevealed changes in the characteristics of the SO (e.g., Grassi, 2005) or by unrevealed changes in the sound source event such as in the current experiments. Along this line, we might speculate whether the results of the current experiment support direct perception or not (e.g., Carello et al., 1998; Kunkler-Peck & Turvey, 2000). The acoustic array generated by a sound source event is uniquely determined by that sound source event. Therefore, theoretically, in order to recover the size of the ball from its sound the listener needs only to apply the inverse transfer function that links the acoustic waveform to the sound source event (Grassi, 2005). Under these circumstances, the listener can recover directly the size of the ball rather than using indirect size estimations that involve processes such as memory or reasoning. The direct perception interpretation implies that the sound contains an acoustical invariant, i.e., an acoustical feature that uniquely identifies the physical characteristic that the listener is asked to judge regardless of the changes across the experimental conditions. In the current study, we found at least one invariant that subjects could use (see Eqs. (3) and (4)). However, in all experiments subjects produced estimates that revealed that they were not listening to this (or other) invariant. Nonetheless, even in the current study, the best predictor of the listeners' performance was the size of the balls (and not any of the acoustical cues). In other words, the behavioral data can be better described in terms of the distal stimulus rather than the proximal stimulus. By the same token, we cannot exclude that subjects are producing “reasonable” size estimates that are modulated by the contingent acoustical characteristics of the sound they are listening to in a given trial. In synthesis, the answer to the direct/indirect question is, unfortunately, mixed.

10. Conclusions

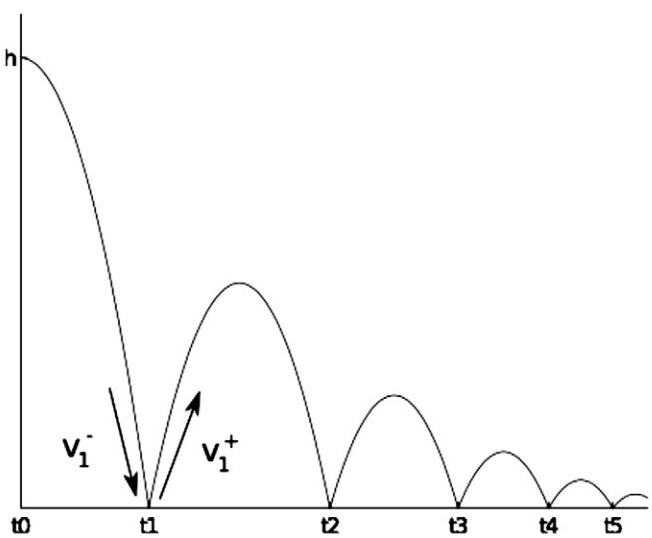
The results of the current study, together with the antecedents, suggest two main acoustical cues for recovering the size of an object from its sound. The first is necessarily a frequency domain cue such as the spectral centroid (e.g., Kunkler-Peck & Turvey, 2000). The second is an amplitude domain cue such as the sound's power (Grassi, 2005). The first cue is invariant in many circumstances (thus reliable), therefore, can be regarded as “globally invariant”. The second, in contrast, cannot. For example, it changes as a function of the distance dividing the listener and the event. Therefore, it has to be regarded as “locally invariant”: it is invariant only within limited contexts such as in the current study. The current results suggest another remark: listeners might attend to amplitude and frequency domain cues although in certain circumstances these cues provide contrasting size-information.

The results of current and previous studies suggest that the listener can tell some of the characteristics of the objects of the world just by listening to the sound that these objects produce impacting a second object. The auditory system has always been regarded as an alerting sense, i.e., a sense that can alert when something is occurring out of our visual field. But many different studies in ecological acoustics corroborate our everyday experience: audition can be used to explore the world as well as vision (e.g., Carello et al., 1998; Giordano et al., 2012; Kunkler-Peck & Turvey, 2000; Li et al., 1991; Pastore et al., 2008). Of course the visual capability of getting size information about objects is much higher. Nevertheless, we would like to stress a different aspect of the auditory ability in estimating objects' size. Humans have limited perceptual capabilities, and one way of overcoming this drawback is to maximize existing capabilities. The hearing system does not need to provide us with perfect estimates of the size of an object, but rather, with functional estimates of the size of the "event" we are listening to (e.g., the event "ball dropped on a plate"). The auditory estimate of an event's size may be crucial when objects are occluded from vision. In addition, large and voluminous events may be potentially more dangerous than smaller events. If we are listening to a low-frequency sound, we are likely encountering an event that is generated by a large object. If we are listening to a high-intensity sound, we are encountering an event that is close to us, or generated by a large object. A high intensity sound can be generated also by a high-energy event such as in the current experiments. In all the above cases, an overestimation of the event's size is advantageous for the perceiver, especially in the case the event is dangerous.

Acknowledgements

The authors would like to thank Elisa Soso and Pavle Buric for helping with the data collection and Judith Stevenson for critically reading the manuscript.

Appendix A



The initial potential energy of a ball dropped on a plate is $E_0 = mgh$ where m is the mass of the ball, g is the acceleration of gravity (thus a constant) and h is the height of the fall. After each impact of the ball on the plate, a portion of this initial energy is transferred to the plate and subsequently converted into acoustic wave; another portion of the energy is transferred to the ball as velocity, causing it rebound; finally, some energy is dissipated as friction, deformation of the ball, heat, etc. Therefore, the energy transferred to the plate at each impact n is $E_n = \alpha^{(n-1)}mgh$, where α is the coefficient of energy transferred back to the ball (coefficient of restitution). As a first approximation, α can be considered as constant for a given falling object and underlying surface.

Because the energy is the capacity for doing work, power is the rate of doing work over time, the sound power P is proportional to the energy of the event, and, by the same token, the power of the first impact, expressed in dB is proportional to the logarithm of E:

$$P \propto E_0$$
$$P_{\text{dB}} \propto 10\log_{10}E_0 \propto 10\log_{10}mgh \propto 10\log_{10}m + 10\log_{10}h + k.$$

The mass of the ball is equal to the volume of the sphere multiplied by the mass density ρ:

$$m = \rho\frac{4}{3}\pi r^3 = kd^3.$$

Therefore, the loudness of the first impact is directly proportional to the logarithm of the ball diameter and the falling height, as expressed in Eq. (1).

$$P_{\text{dB}} \propto 30\log_{10}d + 10\log_{10}h + k \tag{1}$$

After the first impact, the ball bounces up the plate with a kinetic energy:

$$E_1 = \alpha\text{mgh} = \frac{1}{2}mv_1^2,$$

where v_1^+ is the upward velocity of the ball bouncing back. Applying Newton's second law, the equation of motion for ball is therefore:

$$\text{m}x(t) = \text{mg}$$
$$x(t) = -\text{gt} + v_1^- = -\text{gt} + \sqrt{2\alpha\text{gh}}$$
$$x(t) = -\frac{1}{2}g\,t^2 + \sqrt{2\alpha\text{gh}}\text{t}.$$

Determining the time t_2 of the second impact implies solving $x(t) = 0$. This results in: $t_2 = t_1 + 2\sqrt{2\alpha^h/_g}$. The time between the two first impacts is therefore only proportional to the square root of the falling height.

$$\text{B2B} = t_2 - t_1 = k\sqrt{h}$$
$$\log_{10}(\text{B2B}) = \frac{1}{2}\log_{10}h + k \tag{2}$$

Appendix B

In the table below we report the average acoustical values that were extracted from the recordings of the stimuli used in Experiment 4.

Ball size (cm)	RMS power (dB)	Centroid (Hz)	B2B (s)
1	−32.2	5848.7	0.251
1.5	−33.7	4127.2	0.141
2	−34.0	2893.3	0.084
2.5	−22.9	2934.7	0.175
3	−20.5	2317.3	0.121
4	−22.3	2133.1	0.077

References

Avanzini, F. (2001). Computational issues in physically-based sound models. Unpublished doctoral dissertation, Università di Padova, Padova.
Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 250–267.
Boullet, I. (2006). The loudness of impulsive sounds: Perception, measures and models. Unpublished PhD thesis, Université de la Mediterrannée-Aix-Marseille. Available at http://tel.archives-ouvertes.fr/tel-00009870/en/
Cabe, P. A., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 313–324.
Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science*, *9*, 211–214.
Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., & Grassi, M. (2010). When ears drive hands: The influence of contact sound on reaching to grasp. *PLoS One*, *5*, e12240.
Cross, R. (1999). The bounce of a ball. *American Journal of Physics*, *67*, 222–227.
Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, *87*, 311–322.
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.
Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *The Journal of the Acoustical Society of America*, *119*, 1171–1181.
Giordano, B. L., McDonnell, J., & McAdams, S. (2010). Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds. *Brain and Cognition*, *73*, 7–19.
Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 462–476.
Giordano, B. L., Visell, Y., Yun Yao, H., Hayward, V., Cooperstock, J. R., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *The Journal of the Acoustical Society of America*, *131*, 4002–4012.
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.
Grassi, M. (2005). Do we hear size or sound: Balls dropped on plates. *Perception & Psychophysics*, *67*, 274–284.
Grassi, M., & Casco, C. (2009). Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 235–243.
Grassi, M., & Casco, C. (2010). Audiovisual bounce-inducing effect: When sound congruence affects grouping in vision. *Attention, Perception, & Psychophysics*, *72*, 378–386.
Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, *63*, 1493–1500.
Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral–temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America*, *115*, 1252–1265.
Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics*, *69*, 839–855.
Houix, O., Lemaitre, G., Misdariis, N., & Susini, P. (2012). A lexical analysis of environmental sound categories. *Journal of Experimental Psychology. Applied*, *18*, 52–80.
Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 279–294.
Lemaitre, G., Dessein, A., Aura, K., & Susini, P. (2011). Vocal imitations and the identification of sound events. *Ecological Psychology*, *23*, 267–307.
Lemaitre, G., & Heller, L. M. (2012). Auditory perception of material is fragile, while action is strikingly robust. *The Journal of the Acoustical Society of America*, *131*, 1337–1348.
Lemaitre, G., Houix, O., Misdariis, N., & Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology. Applied*, *16*, 16–32.
Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America*, *90*, 3036–3049.
McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*, 177–192.
Pastore, R. E., Flint, J. D., Gaston, J. R., & Solomon, M. J. (2008). Auditory event perception: The source–perception loop for posture in human gait. *Perception & Psychophysics*, *70*, 13–29.
Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* NY: Springer.
Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *The Journal of the Acoustical Society of America*, *81*, 1100–1109.
Stevens, S. S. (1955). The measurement of loudness. *The Journal of the Acoustical Society of America*, *27*, 815–829.
Tucker, S., & Brown, G. J. (2003). Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds. *Proceedings of the 114th Convention of the Audio Engineering Society. March, 2003, Amsterdam, the Netherlands*.
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.
Wildes, R. P., & Richards, W. A. (1988). Recovering material properties from sound. In W. A. Richards (Ed.), *Natural computation* (pp. 356–363). Cambridge, MA: The MIT Press.