PERCEPTION OF THE SIZE AND SHAPE OF RESONANT OBJECTS

Stefan Bleeck, Niamh O'Meara

Institute of Sound and Vibration Research, University of Southampton, United Kingdom

Corresponding author: Stefan Bleeck, Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, United Kingdom, e-mail: bleeck@soton.ac.uk, Tel.: 0044 2380 596682

Abstract

Background: We investigated the ability of naïve, untrained listeners to identify the physical parameters of 3D polystyrene objects from listening to single impulse sounds generated by an impact collision. We were specifically interested in the perception of object shape and object size and their interaction.

Material and methods: Twenty polystyrene objects of various shapes (spheres, hearts, cubes, eggs, rings, and cones) and sizes (between 64 cm^3 and 2278 cm^3) were used in three experiments investigating size and shape perception. In the first experiment, the task was to identify the 'odd one out' of three sounds originating from objects of different shape or size. In the second experiment the task was to identify the shape and size of an object just by listening to it. In the third experiment the task was to rate how similar two sounds were.

Results: Results show that listeners were able, to a degree, to identify the size and shape of objects without reference and in relation to each other. Multidimensional scaling suggests that shape (most salient) and size (second most salient) are the predominant perceptual dimensions.

Conclusions: We conclude that humans, to some degree and without training and without prior experience, have the ability to infer the physical properties of object size and shape by listening to single impulse sounds. Size and shape seem to be independent and are the most salient parameters.

Keywords: shape perception • size perception • auditory perception

---

PERCEPCIÓN DE LA FORMA Y DEL TAMAÑO DE LOS OBJETOS RESONANTES

Resumen

Introducción: En un grupo de oyentes no entrenados, se ha estudiado su capacidad de identificar los parámetros físicos de los objetos tridimensionales, hechos del poliestireno, en base a los impulsos sonoros que estos generaban durante los choques. Estamos particularmente interesados en la percepción de la forma y del tamaño de los objetos y de sus interacciones.

Materiales y métodos: Para la realización de tres experimentos para estudiar la percepción del tamaño y de la forma, se han utilizado veinte objetos hechos del poliestireno, de varias formas (esferas, corazones, cubos, huevos, anillos y piñas), y de varios tamaños (de 64 cm^3 a 2278 cm^3). En el primer experimento se debía identificar un sonido distintivo, de entre los tres sonidos provenientes de objetos de diferente forma y tamaño. En el segundo experimento se trataba de identificar la forma y el tamaño de un objeto escuchando los sonidos que éste emitía. En el tercer experimento, por fin, la tarea fue estudiar la similitud entre los sonidos.

Resultados: Los resultados demuestran que, hasta cierto punto, los oyentes han sido capaces de determinar el tamaño y la forma del objeto, ya sea solo o en comparación con otros objetos. La evaluación multidimensional sugiere que la forma (la más importante) y el tamaño (segundo factor más importante) son las características percibidas dominantes.

Conclusiones: Una persona sin previo entrenamiento ni experiencia dispone de ciertas habilidades de deducción de las propiedades físicas del tamaño y de la forma del objeto, escuchando el sonido de los impulsos individuales. El tamaño y la forma parecen ser parámetros independientes, pero al mismo tiempo los más importantes.

Palabras clave: percepción de la forma • percepción del tamaño • percepción auditiva

ВОСПРИЯТИЕ РАЗМЕРА И ВЕЛИЧИНЫ РЕЗОНИРУЮЩИХ ПРЕДМЕТОВ

Изложение

Вступление: Исследована способность группы неподготовленных слушателей к идентификации физических трехразмерных параметров предметов из стиропора путем издаваемых ими отдельных звуковых импульсов, генерируемых при столкновениях. Нас особенно интересовало восприятие формы и величины объектов, а такжеих взаимодействий.

Материалы и методы: Для проведения трех экспериментов, исследующих восприятие величины и формы использовали двадцать предметов из пенопласта разной формы (шары, сердца, кубики, яйца, кольца и шишки) и размера (от 64 см^3 до 2278 см^3). Заданием первого эксперимента было идентифицировать один отличительный звук среди трех, исходящих от предметов разной формы и величины. Заданием второго опыта было опознать форму и размер объекта, слушая издаваемые им звуки. А задание в третьем опыте – это оценка сходства звуков.

Результаты: Результаты показывают, что слушатели в определенной степени были в состоянии определить размер и форму предмета независимо или путем сравнения с другими предметами. Многоразмерное шкалирование предполагает, что форма (самая важная) и размер (второй в отношении важности) – это доминирующие воспринимаемые свойства.

Выводы: Неподготовленный, не имеющий предварительного опыта человек имеет определенные способности к умозаключениям о физических качествах величины и формы предмета, слушая звуки отдельных импульсов. Кажется, что размер и форма независимые, но самые важные параметры.

Ключевые слова: восприятие формы • восприятие размера • слуховое восприятие

PERCEPCJA ROZMIARU I WIELKOŚCI REZONUJĄCYCH OBIEKTÓW

Streszczenie

Wprowadzenie: Zbadano zdolność grupy nieprzeszkolonych słuchaczy do identyfikacji parametrów fizycznych trójwymiarowych obiektów ze styropianu poprzez wydawane przez nie pojedyncze impulsy dźwiękowe, generowane przy zderzeniach. Byliśmy szczególnie zainteresowani percepcją kształtu i wielkości obiektu oraz ich interakcji.

Materiały i metody: Do przeprowadzenia trzech doświadczeń badających percepcję wielkości i kształtu użyto dwudziestu obiektów ze styropianu o różnych kształtach (kule, serca, kostki, jajka, pierścienie i szyszki) i rozmiarach (od 64 cm^3 do 2278 cm^3). W pierwszym doświadczeniu, zadaniem było zidentyfikowanie jednego wyróżniającego się dźwięku spośród trzech pochodzących z obiektów o różnym kształcie i wielkości. W drugim doświadczeniu zadaniem było rozpoznanie kształtu i rozmiaru obiektu poprzez słuchanie wydawanych przez niego dźwięków. A zdaniem w trzecim doświadczenia była ocena podobieństwa pomiędzy dźwiękami.

Wyniki: Wyniki pokazują, że słuchacze do pewnego stopnia byli w stanie określić rozmiar i kształt przedmiotu niezależnie albo przez porównanie z innymi przedmiotami. Wielowymiarowe skalowanie sugeruje, że kształt (najważniejszy) i rozmiar (drugi pod względem ważności) są dominującymi postrzeganymi cechami.

Wnioski: Człowiek nieprzeszkolony i bez wcześniejszego doświadczenia posiada pewne umiejętności wnioskowania o właściwościach fizycznych wielkości i kształtu obiektu poprzez słuchanie dźwięku pojedynczych impulsów. Rozmiar i kształt wydają się być niezależnymi ale najważniejszymi parametrami.

Słowa kluczowe: percepcja kształtu • percepcja rozmiaru • percepcja słuchowa

Background

Humans possess a remarkable ability to differentiate and identify natural objects by listening to the sounds they generate. We are also, to a degree, able to infer some physical properties of objects by sounds. For example it is easy to hear the difference between two otherwise identical objects when one is made of glass and the other of wood. It is, however, not so obvious that we have access to other physical attributes like size or shape. However, in some cases, this ability is surprisingly precise. Related phenomena have been investigated with a wide variety of objects, materials, and sound events. Listeners can, for example, estimate the ratio between height and width of bars and plates that are hit by a mallet [1,2] or judge the length of a rod dropped on the floor [3,4]. It has been shown that the acoustic structure of a vibrating object generally contains information about the object's physical properties, and these can be utilized by listeners to extract, for example, shape, size, spatial dimensions, hollowness, or material type [5–8]. These results have led to the provocative hypothesis that at least in some cases it might be appropriate to describe auditory perception in terms of the physical properties of the sound source rather than in traditional terms of the physical properties of the sound (Grassi, 2005; [9]). In this study we contribute to this issue by analysing our ability to perceive the physical attributes of natural objects presented in an extreme way: as a single presentation of a single-impact impulse.

Participants in psychophysical studies are often asked to rate stimuli according to relatively basic perceptual features like pitch or loudness. This focus facilitates 'musical' or 'analytical' listening [10,11] in which listeners concentrate on a single aspect of a sound. However, in everyday situations, we usually do not concentrate on the individual physical properties of a sound, although we can often effortlessly attend to the objects and the events that are the source of the sound. When we hear a sound, we usually want to know what happened and what it means for us. For example, when we hear the sound of a liquid filling a glass, we might not be able to tell the pitch, but we might be able to tell that it was the last glass from a bottle. We argue here that experiments that utilize such 'everyday listening' in real situations contribute to our understanding of the general principles of hearing, principles that might be harder to appreciate when using artificial stimuli and concentrating on analytical listening. Such subjective listening modes are reviewed in Stoelinga, 2009 [12]. Stimuli that are interesting in this respect evoke an association between an object and an event.

Physical objects can generate many different sounds, depending on the way they are excited. Several ways of perceiving sounds of different physical origin have been investigated psychophysically: for example striking with a hammer [13], bouncing [14], breaking [15], and rolling [16]. Acoustic properties that are correlated with the ability to detect the physical properties of an object are, for example, the damping time or length of a sound [17]. For a review of the physical properties of sounding objects and their perception, see [9]. The perception of the physical properties of sounding objects is arguably important in auditory scene analysis. It is also well known that the size perception of a speaker plays a role in speech communication [18].

So far, research has investigated the size or the shape of simple sounding objects independently, leaving open the question how these perceptions interact. In this paper we investigate how good listeners are in detecting and discriminating the physical properties of sounding objects among a large selection of 6 different geometries and 4 different sizes which vary at the same time in the same experiment. To avoid the question of training, we only used untrained, naïve listeners that had never had exposure to the sounds.

The sound generating process can be separated conceptually into the sounding 'object' and the 'event' [10,11]. The object (not to be confused with an 'auditory object' [19] is the physical object that generates the sound. The 'event' is what happens to the object (striking, bouncing, rolling, breaking, etc.). Only the two together create an 'ecological' perception. Both seem to be independent; there is no described evidence of obvious interaction (an interaction would be, for example, when an object is perceived as a different object when the event changes). In this paper we investigate the ability to perceive object properties with different geometric features while keeping the event as constant as possible. For this purpose, we chose collision impacts as the event for two reasons: first, collision impacts allow the highest grade of practical repeatability and second, more importantly, impact sounds of simple objects are similar to the theoretical impulse responses. That means, assuming linear sound transmission, that an impact sound is the shortest sound that conveys complete information about the object. Impact sounds of natural objects often have short decay times – in the order of tens of milliseconds. Such impulsive impact sounds are frequently encountered in natural environments and are also related to communication sounds: a wide variety of animal vocal utterances are caused by repeated impulse-type stimulation of resonances. Patterson dubbed the term 'pulse resonance communication' [20] for this type of sound. Examples of repeated pulse resonance communication sounds include frog calls, bird songs, fish sounds, and the voiced components of human speech (repetition of identical pulse resonances only adds periodicity pitch information). Pulse resonances can therefore be considered a simplified model for communication sounds; they are the basic building blocks, the 'atoms', of voiced communication [20].

In evolutionary terms, sound producing mechanisms throughout the animal kingdom have probably evolved separately in an example of convergent evolution [20]. There are also plenty of examples of everyday, pulse resonance sounds from non-living sources: musical instruments, hand claps, footsteps. The world around us is full of pulse resonance sounds, and our auditory system is able to analyse these automatically and effortlessly. However, in order to investigate the ability to hear, discriminate, and identify these building blocks, it would be useful to have a model that is, on the one hand, realistic enough to have a high ecological validity, and on the other hand controllable in its physical properties. Impulsive impact sounds have these properties, and we chose polystyrene objects for our investigations because they are readily available in a wide variety of shapes and sizes. The impedance of the acoustic component of two colliding objects depends on the density and bulk modulus of each material. Both quantities differ significantly between polystyrene (density 20–23 kg/m^3, bulk modulus 3–3.5 GPa) and a striking metal ball (steel 10,700 kg/m^3, bulk modulus 160 GPa). The difference is so large that for all practical purposes the kinetic energy of the collision is dissipated only by the polystyrene object – the striking ball is inaudible.

The usefulness of this approach is that we can ask questions about very basic perceptional qualities. Assuming that pulse resonance sounds are ubiquitous in nature, we should understand the perception of these simplest sounds before we try to investigate more complex sounds. On a contextual level, researchers have attempted to understand the psychology of human sound perception of physical objects by considering how we integrate environmental information from complex sources. Giordano et al. [13] found that the more information we can exploit from an acoustic feature, the more perceptual weight it has.

Table 1. Specification of objects used in all three experiments.

	Object	Abbreviation	Weight (g)	Max dimension (cm)	Volume (cm^3)	Density (g/cm^3)
1	Sphere 1	s1	3.4	6.0	113	0.030
2	Sphere 2	s2	7.3	8.0	268	0.027
3	Sphere 3	s3	17.5	10.0	523	0.033
4	Sphere 4	s4	29	12.0	905	0.032
5	Egg 1	e1	3.3	5.6	102	0.032
6	Egg 2	e2	5.2	8.4	175	0.029
7	Egg 3	e3	7.7	10	233	0.033
8	Ring 1	r1	5.6	14.6	195	0.028
9	Ring 2	r2	11.5	16.7	429	0.026
10	Ring 3	r3	27.8	21.5	869	0.032
11	Ring 4	r4	51	24.3	1579	0.029
12	Cone 1	co1	7.6	15.3	262	0.029
13	Cone 2	co2	13.5	20.4	445	0.030
14	Cone 3	co3	26.3	26.1	1018	0.026
15	Cone 4	co4	75.4	31.3	2278	0.033
16	Heart 1	h1	2	5.5	71	0.028
17	Heart 2	h2	4.6	8.3	142	0.032
18	Heart 3	h3	8.4	11.2	251	0.033
19	Cube 1	cu1	1.9	4.0	64	0.029
20	Cube 2	cu2	6.4	6.0	216	0.029

Here we employ a descriptive approach using extremely simple, yet ecologically 'interesting' sounds. In order to later inform computational and speech processing models, we do not allow learning or give contextual information. We report three experiments that are novel in two important ways: first, we use a wide range of simple shapes, and secondly we also alter size within each shape family so that we are able to discern interactions. We asked three related research questions about perception ability: how good are naïve listeners at discriminating resonating objects in relation to each other? How good are listeners at absolute perception of either shape or size of sounding objects? Finally, what are the underlying dimensions of acoustic perception?

Methods

Overview

In all three experiments reported in this paper, untrained, naïve participants listened to a set of recorded stimuli via headphones while given different tasks. In the first experiment, the task was to identify the 'odd one out' of three sounds. In the second experiment the task was to identify the shape and size of an object just by listening to it. In the third experiment the task was to rate how similar two sounds are. None of the participants had any previous exposure to the sound of the particular polystyrene objects in the experiment. No feedback was given at any time during the experiments and so no useful learning could happen.

Method

Twenty homogeneous polystyrene objects of similar density were sourced from a craft shop (Craftmill Ltd, UK, www.craftmill.co.uk). The properties of the objects are described in Table 1. Homogeneity of objects was confirmed by visual inspection and the absence of holes or denser areas.

Stimuli were recorded through a free-field microphone in a sound proof room. Each polystyrene object was attached to a length of string and hung from the ceiling to allow free movement. A metal ball (weight 11.25 g, diameter 14 mm) was attached to a 163 cm string hung from the ceiling, 5 cm away from the centre of the object, at a specific height for each object (see Figure 1), creating a pendulum with which to strike the polystyrene object. A reference maker was used to ensure that the metal ball was released at the same distance for the polystyrene object each time. Using markers on each object, the pendulum hit the objects through a line containing its centre of gravity, and all efforts were made to ensure the impact position and force applied was as uniform as possible for

Figure 1. For each of the objects used, the dot indicates the point where collision occurred.

each object type. The microphone was placed 45 cm radial to the point of impact.

The recorded sounds were checked audiovisually for signal to noise ratio and repeatability. The differences between all presentations were generally small. Three recordings were selected for each object and used randomly in the experiments. The length of all signals was short (generally less than 30 ms, but longer for the ring) and the amplitude of each presentation was calibrated to 80dB (A). Presentation amplitude during the experiment was roved randomly within −5 to 0 dB. All sounds can be downloaded from the website of the *Journal of Hearing Science*.

Participants

The participants were otologically normal (self-reported) men and women between the ages of 18 and 35. Numbers of participants varied between experiments. Participants were unpaid volunteers, mostly students of the Audiology program at the University of Southampton and were different in all experiments. In Experiments 2 and 3, participants could see all 20 objects which were arranged in the experimental room, but they were not allowed to touch them. This was done to give participants a reasonable motivation of the background of the experiment without giving them any acoustical information. Participants received basic verbal and written instructions for the task. Each participant performed in only one experiment. All experiments were conducted under ethical approval of the Ethics Committee of the Institute of Sound and Vibration Research at the University of Southampton, UK.

Experiment 1: Relative identification of size or shape

Only a subset of objects was used in this experiment: two groups of objects of each shape, a 'small' group – co1; e2; s2; r1; h2 – and a 'large' group – co2; e3; s3; r2; h3. These were selected from the overall pool of objects in order to be roughly comparable in size among each group and different between groups. The rationale behind the experiment was that participants were presented with three stimuli, of which two came from one category (either size or shape) and the third one was the 'odd one out' (different in shape or size). The participant's task was to identify the sound of the object that was either different from the other two in size (Experiment 1_{size}) or shape (Experiment 1_{shape}). The question was whether and to what degree participants could identify the odd one out in this three alternative forced choice paradigm. The two paradigms are schematically shown in Figure 2. We expected that participants would be able to do both experiments, but would be better for size than for shape identification.

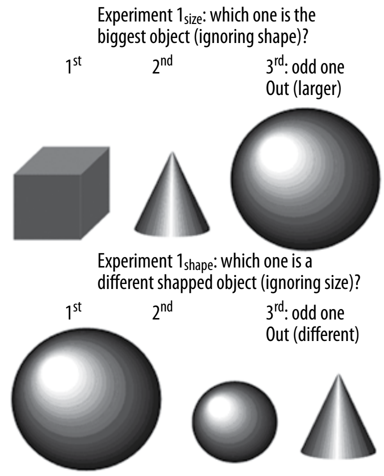

Figure 2. Paradigms for Experiment 1_{size} (top) and Experiment 1_{shape} (bottom). Explanation in text.

Experiment 1_{size} The goal of this experiment was to investigate whether participants could judge which of three sounds stemmed from the largest object. The question asked was 'which one of these three sounds comes from the biggest object?' All combinations of three different shapes were used (30 combinations), e.g. 'small cube', 'small cone', 'large sphere'. Here the 'large sphere' is the odd one out, because it is larger than the other two.

Experiment 1_{shape} The goal of this experiment was to see if participants could judge which of three sounds stemmed from the one with a different shape. The question asked was 'which one of these three sounds comes from an object that has a different shape from the other two?' All combinations of two different shapes were used (leading to 24 combinations), e.g. 'large sphere', 'small sphere', 'small cone'. Here the 'small cone' is the odd one out, because it is of a different shape to the other two.

The experiment took around 25 minutes to complete. Stimuli were presented in randomized sequences of three, with 500 ms intervals between them. Each sequence was presented twice (not successive) to check repeatability. Participants were allowed to replay the sequence once. Exactly 29 participants completed the experiment. Participants sat in front of a computer screen displaying a user interface with three unlabelled buttons. The buttons lit up successively during presentation of the three sounds and the buttons were then used by the participants to answer the respective question ('Which one is biggest?'; 'Which one is of different shape?'). No feedback was given at any stage.

Results

The question in these experiments was to determine how good participants were in comparing the shape and the size of an object in relation to other objects just by listening to them and identifying the 'odd-one-out'. The odd-one-out in the results below is referred to either as $target_{size}$ (Experiment 1_{size}) or $target_{shape}$ (Experiment 1_{shape}). A Wilcoxon signed rank test showed no significant performance

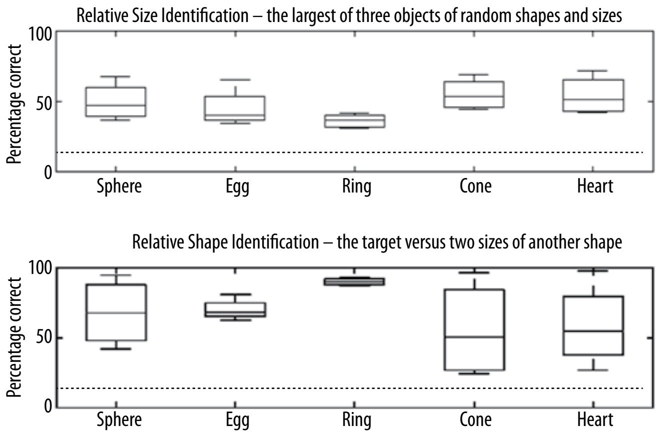

Figure 3. Percentage correct scores for Experiment 1_{Size} (top panel) and 1_{Shape} (bottom panel). In the boxplots, the horizontal line is the median; boxes span the interquartile range, and whiskers show either 1.5 times IQR or max (min) values, depending on which is smaller or larger. Chance level (33.3%) is indicated by dashed lines.

Table 2. Percentage correct scores for Experiment 1_{shape}. Columns show objects which appeared twice, while the rows show $target_{size}$. Chance level was 33%.

Target	Comparison Shape					*Average*
	Sphere	Egg	Ring	Cone	Heart	
Sphere	–	62.5	92.9	24.6	26.8	54.2
Egg	81.5	–	87.0	71.9	61.1	75.37
Ring	94.7	80.7	–	96.6	98.2	92.55
Cone	54.4	68.0	91.1	–	49.1	65.65
Heart	41.8	69.1	88.9	29.1	–	57.22

difference between the first and second run so data from both runs were pooled.

Experiment 1_{size}

The top panel of Figure 3 shows the results of Experiment 1_{size}. Boxes represent the distribution of the average correct responses. For example, the sphere was correctly identified as the larger object in around 50% of the cases when compared with 2 other smaller objects. It is not possible to show the raw results in a systematic table because each condition has three different shapes. All size targets were identified significantly above the chance level of 33.3% ($F(4)=7.12$, $p<0.01$), and the overall mean value for correct identification was 48.2%. The large cone (co2) was correctly identified most often. The large ring (r2) resulted in the lowest overall average score and had the smallest range.

Experiment 1_{shape}

The bottom panel of Figure 3 and Table 2 show the distribution of correctly identified $target_{shape}$, averaged over all comparison shapes. For example, the sphere was correctly identified as being different from the eggs in 62.5% of the cases and on average the sphere was identified correctly in 54.2% of all cases. The overall mean value for correct identification was 68.7%. All $target_{shape}$ were on average identified significantly above chance level ($F(4)=13.5$, $p<0.01$). The large ring (r2) produced the highest average score of 92.6%. Table 2 breaks the averages down into the percentage identification when compared to specific shape pairs: the rows show $target_{shape}$; the columns show which object in a trial appeared twice as its comparison. The scores for identifying the ring showed the smallest range, while the sphere had the largest range. Participants reported that the rings had a distinctive sound which made the shape easier to recognise. However, the top panel in Figure 3 shows that the ring scored the lowest when it was a $target_{size}$. This does not necessarily indicate that participants could not hear the size of the ring (see Experiment 2), but this might indicate that participants were confused by the fundamentally different sound quality of the ring when making a judgement about size.

Only three combinations (cones *vs.* target heart, hearts *vs.* target sphere, and cones *vs.* target sphere) were around chance level, indicating that the sounds of these objects are similar to each other. Interestingly, when comparing the converse results – i.e. hearts *vs.* target cone, spheres *vs.* target heart, and spheres *vs.* target cone – the identification was much better.

The spectra of two objects of the same shape are related to each other and modulated by size [21]. The parameters of the spectra that define shape, but not size, are shape invariant.

This implies that shape identification benefits from the shape invariants contained in the comparison pair, or that the different sizes of the comparison shape made the difference between the shapes more noticeable. The highest scores were from any combination involving rings, which implies again that rings are distinctly different in shape from all others.

We conclude that untrained listeners can identify the 'odd-one-out' of three different sounds, and, contrary to the initial hypothesis, they performed better for shape (68.7%) than for size (48.2%). Note, however, that this depends on the actual range of shapes and sizes of the presented objects, and might not generalize to other objects.

Experiment 2: Absolute identification

The research question in Experiment 2 involved absolute shape and/or size identification: could participants match sounds that they heard to the objects they saw in front of them? This experiment consisted of three parts, with participants being asked to identify which of the set of selected objects in front of them produced the sound that they heard. Participants were not permitted to handle the objects.

- Experiment 2A – out of a choice of 2 objects
  - $2A_{shape}$: similar size, but different shape
  - $2A_{size}$: same shape, but different size
- Experiment 2B – out of a choice of 3 objects
  - $2B_{shape}$: similar size, but different shape
  - $2B_{size}$: same shape, but different size
- Experiment 2C: out of a choice of all 20 objects of different shapes and sizes.

In experiments 2A and 2B, the experimenter placed the relevant number of objects (2 or 3) on a table in front of the participant. The closed headphones that participants were wearing prevented them from hearing any resulting sounds. The design of the experiments was an adapted alternative-forced-choice paradigm, with 2 or 3 choices depending on the experiment. The experiment was double blind: the objects were randomly chosen by the computer, the experimenter was informed which objects to select, without the knowledge of the participant. The experimenter had no knowledge about the correct order of the presented sounds. The sounds (2 for Experiment 2A; 3 for Experiment 2B) were presented to the participants and they had to decide which sound was produced by which object, by pointing at the objects in the same order that the sounds had appeared. The experimenter recorded the response by computer, again without feedback. In Experiment 2A (2 objects), combinations of all 20 objects were used; in Experiment 2B (3 objects) all shapes except the cubes were used, as only two cubes were available. Therefore, five different shapes were used in three sizes (s1, s2, s3, e1, e2, e3, r1, r2, r3, co1, co2, co3, h1, h2, and h3). In Experiment 2C, all 20 objects were placed in front of participants, and only one sound was played at a time, repeating every sound three times.

In total, 47 otologically normal (self-reported) participants (age between 18 and 35; 34 females) undertook Experiment 2, but each participant participated in only one part of the experiment, to avoid cross-learning effects. There were 14 participants who took part in Experiment 2A with two objects, 25 participants in Experiment 2B with three objects, and 8 participants in Experiment 2C with all 20 objects. Experiments took between 15 and 35 minutes to complete.

Results of Experiment 2

Experiment 2A: Choice of two objects

The question in Experiment 2A was to determine whether participants could identify the sounds made by two objects which they were seeing in front of them. In Experiment $2A_{shape}$, objects had different shapes but similar size, in Experiment $2A_{size}$ objects had different size but identical shape. In Experiment $2A_{shape}$ every shape was compared against 5 other shapes, and in Experiment $2A_{size}$ each object was compared against all the others in its family. Figure 4 shows the average results for Experiment $2A_{size}$. Table 3 shows the confusion matrix for Experiment $2A_{shape}$ (these values are also shown in Figure 5 with the results of Experiment $2B_{shape}$). Perhaps not surprisingly after the results of Experiment 1, in Experiment $2A_{shape}$ the ring was again the easiest to identify, with more than 97% accuracy (chance level =50%). The heart was the most difficult, with 71.4%; however, all shapes were identified significantly better than chance ($ps<0.01$). The greatest confusion, with 12.9%, was between cubes and hearts. Experiment $2A_{size}$ shows good size identification in all object families, except cubes and small spheres. The sizes of the rings were correctly identified most often.

Experiment 2B: Three objects

This experiment was an extension to Experiment 2A, this time with three objects. As before, two experiments were conducted: in Experiment $2B_{shape}$, 3 objects with similar size, but different shapes were used, and in Experiment $2B_{size}$, 3 objects with the same shape but different size were used. Figure 5 compares the results of Experiments $2A_{shape}$ and $2B_{shape}$, to indicate how identification performance decreased with the addition of another object. There is no data for the cubes in Experiment $2B_{shape}$, as there were only two cubes in total. Figure 6 shows the results of Experiment $2B_{size}$. In Experiment $2B_{shape}$ all shapes were identified significantly better then chance (33%) (t-tests, $p<0.05$). As in Experiment 2A, the ring was easiest to identify with 91.7% and the heart was hardest to identify with 54.2%. The ring was significantly easier to identify then all other shapes; no other combination was statistically significant (ANOVA with Games-Howell post hoc test). In Experiment $2B_{size}$, all sizes were on average identified better then chance (33%) apart from e2 and h2 (t-tests, $ps<0.05$).

Experiment 2C: Twenty objects

In Experiment 2C, a sound was played to the participants and, out of the 20 objects placed in front of them, they were asked to point to the one they believed they had heard. Again, no feedback was given. Participants scored a 'correct shape' point when they chose any size from the correct shape family. Table 4 shows the same results broken down for shape, averaged over all sizes. Participants found this experiment much more challenging than the experiments with two or three objects. Chance levels for this experiment are complicated due to the different number of objects in each group; the sphere, ring, and cone had a chance level of 20%, the egg and heart 15%, and the cube 10%. Only objects of the ring and cone families were identified significantly above chance level (t-tests, $p<0.05$).

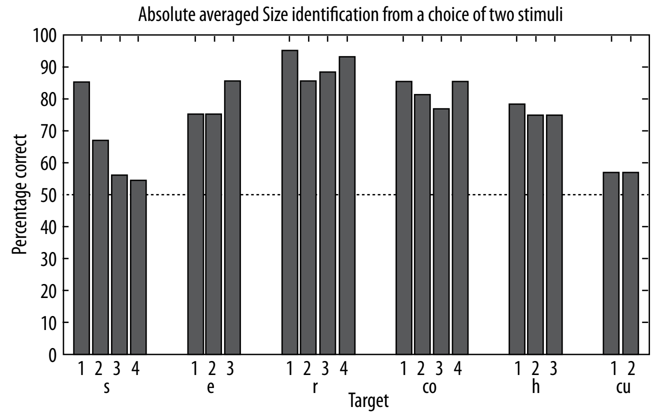

Figure 4. Percentage correct scores from Experiment $2A_{size}$, the absolute identification of an object's size in a 2AFC task. Each column is the average of that object's percentage averaged over all of its comparisons. Dashed line indicates 50% chance level.

Table 3. Percentage correct scores for Experiment $2A_{shape}$, displayed as a confusion matrix. Diagonal elements (dark grey) represent correct identification, off-diagonal elements represent confusions. Columns show the target shape to be identified; rows show the shape it was identified as.

		Shape presented					
		Sphere	**Egg**	**Ring**	**Cone**	**Heart**	**Cube**
Object identified	**Sphere**	**82.9**	7.1	0.0	4.3	4.3	1.4
	Egg	7.1	**81.4**	0.0	5.7	4.3	1.4
	Ring	0.0	0.0	**97.1**	1.4	1.4	0.0
	Cone	4.3	5.7	1.4	**77.1**	5.7	5.7
	Heart	4.3	4.3	1.4	5.7	**71.4**	12.9
	Cube	1.4	1.4	0.0	5.7	12.9	**78.6**

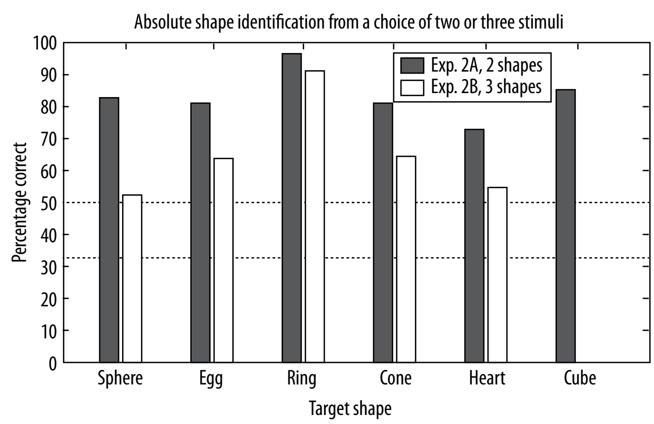

Figure 5. Percentage correct scores for Experiments $2A_{shape}$ and $2B_{shape}$, the absolute identification of an object's shape from two or three other objects. The dash-dot line represents chance level (50%) for the 2AFC, and the dashed line is at chance level (33%) for the 3AFC.

Experiment 3: Dimensions of perception

The purpose of the third experiment was to establish whether the subjective perception of sounds can be matched to measured objective dimensions. We hypothesised from the results of the previous experiments that the most important aspect of perception is determined by the shape of the object followed by its size, and that these aspects are independent of each other. This hypothesis was tested in Experiment 3 by measuring subjective dissimilarity between object pairs. We discarded the rings for this experiment due to the results from the previous experiments which suggested that their sound is much more distinct than the others. Also, its topology in comparison to the other shapes is much different: the ring has a hole. This leads to circular modes of excitation with much longer decay times: in some modes, waves can propagate around the ring for a long time, which gives the percept a much more tonal character.

The 16 objects used in this experiment were s1, s2, s3, s4, e1, e2, e3, co1, co2, co3, co4, h1, h2, h3, cu1, and cu2. Each object was compared to the others, resulting in 16×15=240 comparisons. All 240 pairs were presented

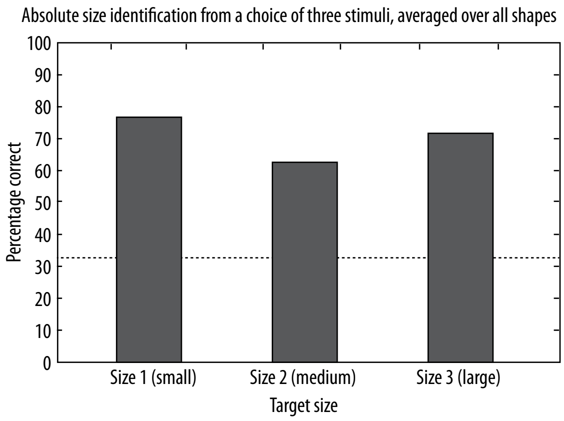

Figure 6. Percentage correct scores from Experiment $2B_{size}$, the absolute size identification from a 3AFC task. These results show the correct size identification percentage averaged over all object shapes. The dashed line represents the chance level of 33%.

twice to the participants in randomised order. Participants sat in front of a computer screen and judged the similarity of the sounds by adjusting a horizontal slider, which was ~10 cm long, representing a visual analogue scale. The slider was labelled 'very dissimilar' to the left and 'very similar' to the right. Two example sound pairs were played to the participants in advance to give them an idea of what the range of the sounds was. The two pairs were e1 and s1 (similar size and shape) and cu1 and co4 (dissimilar in size and shape). Exactly 47 participants took part in this experiment, and the experiment took around 20 minutes to finish. In general, participants reported that the task was difficult, but intuitively do-able without training. Informal feedback also indicated that participants could not easily put in words why two sounds were rated as similar or not.

Experiment 3: Results

This experiment was carried out to test the hypothesis that shape and size are two independent components of perception by asking participants to rate the similarity of all combinations of 20 different objects subjectively. Participants reported that the task was not easy because they had never been exposed to such sounds or task. This is reflected in considerable variation between the first and second runs of individual participants. These were correlated significantly, but only with a 'medium' correlation coefficient of r=0.31. However, when averaged over the whole group of participants, the first and second presentation of each pair were correlated with a coefficient of r=0.92. This indicates that even if the task was difficult for individual participants, as a group they responded in a highly repeatable way. Data of all participants was pooled and analysed as a triangular dissimilarity matrix using a classical metric, multidimensional scaling (MDS) (the 'cmdscale' function from the MATLAB (2012b) statistics toolbox). MDS works by finding an embedding from the objects into a subspace, while preserving the distances. Distances were calculated by the MATLAB function 'pdist' based on the Euclidian distance. Our hypothesis was that shape and size are two independent components of perception, and this translates into two properties of the MDS: first, when the distances are embedded in 2 dimensions, these two dimensions should project the shape and the size characteristics of the objects; secondly the first two eigenvalues of the transformation should be much larger than any further values.

The first two dimensions of the MDS algorithm obtained for all 16 objects in Experiment 3 are shown in Figure 7. The first 5 eigenvalues of the first dimensions as calculated by the MDS algorithm are 8.99, 2.89, 1.107, 0.79, and 0.53. The first two dimensions therefore have by far the highest eigenvalues and are thus the most prominent. Dimension 1 correlates strongly with the different shapes that were used in the experiment; dimension 2 correlates similarly strongly with the size. All object shape groups, with the exception of s2 and h3, are non-overlapping on dimension 1, or in other words, objects of the same shape are always 'next-door neighbours' in dimension 1. We therefore argue that this dimension is a good representation of an object's 'shape'. Dimension 2 shows that larger objects of one class are always arranged higher than smaller objects of the same class. There is therefore a strong relationship between relative size and dimension 2, and we therefore conclude – as above – that this dimension represents an object's relative 'size'. Note that the absolute size – the size of one shape class compared to another – is not necessarily preserved. For example, egg3 has more mass and a bigger volume than sphere2. However, dimension 2 is strongly correlated with the volume of the objects (from Table 2) with a coefficient of variation R^2=0.38 (p<0.01). Note that this relationship is even stronger (R^2=0.51, p<0.01) when

Table 4. Percentage correct scores for shape perception in Experiment 2C (20 objects), shown as a confusion matrix.

		Object family presented					
		Sphere	Egg	Ring	Cone	Heart	Cube
Object family identified	Sphere	**18.8**	33.3	9.4	15.6	37.5	16.7
	Egg	25.0	**12.5**	12.5	15.6	0.0	20.8
	Ring	3.1	12.5	**62.5**	6.3	18.8	8.3
	Cone	28.1	4.2	9.4	**40.6**	12.5	20.8
	Heart	15.6	20.8	0.0	12.5	**18.8**	20.8
	Cube	9.4	16.7	6.3	9.4	12.5	**12.5**

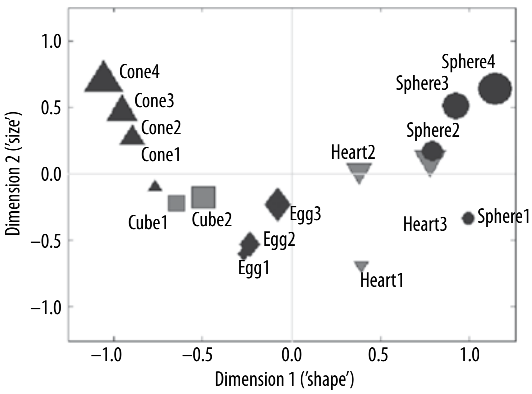

Figure 7. Results of Experiment 3 plotted using multidimensional scaling. Dimension 1 (shape) is shown on the *x*-axis and dimension 2 (size) is shown on the *y*-axis.

dimension 2 is correlated against the logarithm of the volume. The coefficient of determination, R^2, of the MDS was 0.61, indicating that 61% of the variance of the scaled data can be accounted for by the MDS procedure.

Discussion

The main new finding of this study is the ability of untrained, naïve participants to obtain information about both shape and size at the same time from impulse resonances of 3D polystyrene objects. Since in our experiments we were restricted to the sounds of the objects that were available to us, i.e. a fixed set of polystyrene objects, we cannot quantitatively compare the ability to hear shape and to hear size directly (for example in order to answer the question if we are better at size or at shape perception). Differently shaped or sized objects might well have produced higher or lower discrimination values for either. We can, however, conclude that both perception dimensions exist, that they seem to be independent, and that naïve people can discriminate them reliably.

Limitations

Due to practical limitations, it was not possible to have the same number of objects in each family or to have objects between families that had exactly the same weight. Since we were restricted to available objects, they were matched in groups of similar, but not identical size. This means that it was not possible to compare size discrimination precisely. This problem could be overcome in future experiments if objects were used that were tailor-made for the experiment.

In Experiment 1_{size}, an inconsistency went unnoticed. We asked participants for the 'larger' object. However, the word 'larger' might be understood differently by different participants, because it can either refer to volume, mass, or maximum dimension. However, in all but two cases both volume and maximum dimension was bigger in the 'larger' object. The exception is that the maximum dimension of c1 is larger than the maximum dimension of e3 and s3. These cases potentially introduce an error in the analysis.

Most participants in the experiments were audiology students (mostly young and mostly female) from the University of Southampton, UK. While they were not trained, they were all experienced ('expert') listeners in audiological experiments, and are thus not an ideal sample to represent the general population.

Ecological validity

The experiments in this study were designed with ecological validity in mind. In light of the research question (which was to learn about the perception of the physical attributes of natural objects), we tried to encourage listeners to use 'everyday listening', that is, to perceive the nature of the event producing the sound and not the perception of the acoustic properties of the signal [9]. Participants were therefore informed about the nature of the sounds that they were to hear, and they were able to see (but not touch) the objects in front of them. Ecological validity could be improved if the collisions happened physically in the same room as the participants (behind a screen), but practical limitations of time and space made this approach impossible. In any case, the striking mechanism's limitations meant it could not guarantee identical sound production every time, and the experiments would have taken too long. Therefore, stimuli in our experiments were presented bilaterally via headphones. It has shown in a comparable experiment with falling rods that both methods produce similar results [14], but it needs to be taken into account that a recording via a single microphone and presentation via headphones does not allow subjects to make use of a sound's normal spatial cues.

Informal feedback from the participants indicated that the setup did encourage 'everyday listening': participants often reported that the strategy they employed to solve the task was to visualise the objects that could have produced the sounds. However it is an open question of what quantitative impact the experimental design and use of everyday listening had on the results. Apart from the participants' verbal feedback, there is no way of confirming that the participants really did listen in 'everyday' mode.

No training took place at any point during the experiments, no feedback on performance was given, and no subject participated in more than one experiment. We can therefore rule out the effect of training on the results.

Dimensions of perception

Size and shape emerged in our experiments as two independent dimensions of perception. The results of Experiment 3 demonstrate that shape and size are the most prominent aspects of the perception of the sound that an object produces. A difference in shape was – for the used set of objects – more perceptible than a difference in size, but both are reliable features of perception.

Although the experiments were not specifically designed to investigate interaction between these dimensions, we found no evidence for it. We never observed a case where a change of shape would fool the listener to believe that it changed its size, or vice versa. Whatever acoustical properties were responsible for the participants' ability to determine the shape and size of an object, they seem to be different and independent.

Note that we cannot ultimately prove that the dimension 2 in the MDS experiment (Figure 7) represents size alone and is independent of dimension 1 (shape). However, the results from Experiment 1 – the observed correlation between dimension 2 and absolute size – and the absence of any systematic size – shape confusion leads us to conclude that the two dimensions are probably independent.

The analysis of the two perceptional 'dimensions' shape and size must, however, be guided by caution. While size can be parameterized by a single continuous dimension, there is no easy way to parameterize shape, which is in our experiments a nominal descriptor (either an object belongs to a distinct class or not). Apparent physical shape similarities (for example between an egg and a sphere) do not necessarily translate into similar perceptions. Instead, we found the closest perceptional similarity between spheres and hearts. It is not clear what properties of sound production led to the perceived similarity or dissimilarity. A future study could investigate the correlation between perceived similarity and acoustic sound properties.

Impact sounds

Impact sounds are a very simple model of animal communication: vowels as produced by humans (and other animal vocal communication sounds) are the result of repeated pulsatile excitation. In humans, the vocal folds vibrate rhythmically, and the complete vowel is the product of the interaction of the pulse and the frequency resonances that are produced in the vocal tract. Single impact sounds as used in this study might therefore be regarded as the 'atoms', the basic building blocks of pulsatile communication sounds. Patterson [20] hypothesized that the mammalian auditory system is evolutionary optimized to process single pulsatile sounds. There it was argued that repetition of identical pulsatile sounds, as in vowels, is utilised to improve the signal to noise ratio of single impulse sounds without adding more information, apart from the repetition rate (or pitch). The experiments presented here demonstrate that our auditory system is indeed able to process single impulse sounds with high precision, suggesting that such sounds can be seen as simplified models of more complex auditory communication sounds, because the only information that repetition adds is, like in a vowel, temporal pitch. In a future paper we plan to investigate how the number of repetitions of single impulse sounds affects our ability to detect and discriminate them.

Possible explanations

There are several possible explanations to describe our ability to hear shape and size. The first one is that it is learned. This would require that many generic sound patterns of objects that are spherical, conical, and so on, are learned during the development of the auditory system, probably with visual feedback. We would therefore be able to recognise certain aspects of the spectrum as being 'spherical', 'conical', etc. A similar argument could explain the ability to discriminate size: we have learned generically that objects which have a specific size have resonances at about a specific frequency (influenced by the material). The results of our experiments cannot confirm or reject this hypothesis, since all our subjects were adults who had had many years of acoustic experience. However, we note that our subjects had never heard the specific sounds that the objects in the experiments made. A learned ability would therefore require a degree of abstraction of an almost infinite variety of shape classes and families, combined with an almost infinite amount of variation in size.

A different, and possibly more intriguing, explanation is that we can somehow infer from the sound of an object how it is shaped (and potentially the other way round: we might be able to imagine from the sight of an object how it would sound). This explanation would not require memory, but an ability to somehow 'calculate' (or reconstruct) sounds in the brain. This 'auditory reconstruction' theory that we suggest here is related to the well-known 'motor theory of speech', the theory that speech perception is influenced by perceiving motoric gestures of the vocal tract. The motor system in this theory is recruited for perceiving speech; the general idea here is that the motor system's role is not only to produce speech sounds, but also to detect them. For a review of the motor theory see Galantucci et al. [22], and for criticism see Massaro & Chen [23]. We suggest here that there is an analogy between the motor theory and the reconstruction theory. In the motor theory, speech sounds are not so much perceived, but rather mirrored, by the motor system. In this way, speech is better described by the physical properties of production than perceptional parameters. This representation might aid perception. Similarly, in the reconstruction theory, impulse sounds would be described appropriately by the physical properties of the sound production process rather than the perceptional parameters. The neural mechanism that would solve such a task would need to be able to order objects for size and to reconstruct objects (to a degree) for shape.

We have shown in a previous paper [21] that the spectral cue is dominant for the auditory size discrimination of transient signals. The cues for size discrimination are similar in transient sounds and in speech sounds. We argued there that single impulse responses can be thought of as a simple model of voiced speech. We also presented a mathematical model for the analysis of transient signals based on the Mellin transform and the auditory image model (tAIM) and were able to predict with high confidence which of two signals from objects of the same shape is bigger. This model was developed on the basis of spherical shapes, but it also works for all the shapes reported here because the objects used in both papers are identical (see Figure 8 in [21]). This provides evidence that a relatively simple algorithm that evaluates relatively simple spectral and temporal features of a sound can differentiate for shape and order for size, which are the two major components of our auditory reconstruction theory.

Nevertheless, this theory is still untested. The fact that participants had no prior experience with the sound of these particular objects, and that the strategy they employed was often described informally as 'visualization' or 'reconstruction', is intriguing and speaks for the reconstruction theory, but clearly more work needs to be done to establish which theory (or a combination of theories) is correct. The 'learning' hypothesis could be tested directly by asking very young children, who are still relatively inexperienced acoustically, if they can perform the task.

Conclusions

- Participants, without prior experience, were able to infer the physical properties about the size and shape of polystyrene objects from their impact sounds.
- These abilities do not require specific training.
- Size and shape seem to be independent and are the most salient parameters.

Acknowledgements

Many thanks to Helen Hill, Elisabeth Clarke, and Natasha Porter for performing the experiments as part of their undergraduate project in the Audiology course at the University of Southampton.

References:

1. Kunkler-Peck AJ, Turvey MT. Hearing shape. J Exp Psychol Hum Percept Perform, 2000; 26: 279–94.
2. Lakatos S, McAdams S, Caussé R. The representation of auditory source characteristics: simple geometric form. Percept Psychophys, 1997; 59: 1180–90.
3. Carello C, Anderson KL, Kunkler-Peck AJ. Perception of object length by sound. Psychological Science, 1998; 9(3): 211–4.
4. Kirkwood BC. The influence of presentation method on auditory length perception. In: Cummins-Sebree (ed.), Studies in Perception and Action IX. Psychology Press, 2005.
5. Aramaki M, Besson M, Kronland-Martinet R, Ystad S. Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. In: Ystad S, Kronland-Martinet R, Jensen K (eds.), Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music (Vol. 5493, pp. 1–17). Springer: Berlin, Heidelberg. Retrieved from http: //dx.doi.org/10.1007/978-3-642-02518-1_1.
6. Lutfi RA, Stoelinga CN. Sensory constraints on auditory identification of the material and geometric properties of struck bars. J Acoust Soc Am, 2010; 127(1): 350–60.
7. Lutfi RA. Auditory detection of hollowness. J Acoust Soc Am, 2001; 110(2): 1010–9
8. McAdams S, Chaigne A, Roussarie V. The psychomechanics of simulated sound sources: material properties of impacted bars. J Acoust Soc Am, 2004; 115(3): 1306–20.
9. Carello C, Wagman J, Turvey MT. Acoustic specification of object properties. In: Anderson J, Anderson B. (eds.), Moving Image Theory: Ecological Considerations (pp. 79–104). Southern Illionois University Press, 2005.
10. Gaver WW. What in the world do we hear? An ecological approach to auditory event perception. Ecological Psychology, 1993; 5(1): 1–29.
11. Gaver WW. How do we hear in the world? Explorations in ecological acoustics. Ecological Psychology, 1993; 5(4): 285–313.
12. Stoelinga C. A Psychomechanical Study of Rolling Sounds. VDM Verlag, 2009.
13. Giordano BL, Rocchesso D, McAdams S. Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. J Exp Psychol Hum Percept Perform, 2010; 36: 462–76.
14. Kirkwood BC. Maintaining Realism in Auditory Length-perception Experiments. PhD Papers in Technology and Science (pp. 2–4). Aalborg University, 2005.
15. Warren WH, Verbrugge RR. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. J Exp Psychol Hum Percept Perform, 1984; 10: 704–12.
16. Houben MMJ, Kohlrausch A, Hermes DJ. Perception of the size and speed of rolling balls by sound. Speech Communication, 2004; 43(4): 331–45.
17. Giordano L, McAdams S. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. J Acoust Soc Am, 2006; 119(2): 1171–81.
18. Ives DT, Smith DRR, Patterson RD. Discrimination of speaker size from syllable phrases. J Acoust Soc Am, 2005; 118(6): 3816–22.
19. Bregman AS. Auditory Scene Analysis. MIT Press: Cambridge, 1994.
20. Patterson RD, Smith DRR, van Dinther R. Walters TC, van Dinther R. Size information in the production and perception of communication sounds. In: Yost WA, Popper AN, Fay RR (eds.), Auditory Perception of Sound Sources (pp. 43–75). New York: Springer Science + Business Media, 2008.
21. O'Meara N, Bleeck S. Size discrimination of transient sounds: perception and modelling. Journal of Hearing Science, 2013; 3(3): 32–44.
22. Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. Psychol Bull, 2006; 13(3): 361–77.
23. Massaro DW, Chen TH. The motor theory of speech perception revisited. Psychol Bull, 2008; 15: 453–7.