10ème **C**ongrès **F**rançais d'**A**coustique

Lyon, 12-16 Avril 2010

Les imitations vocales de sons environnementaux permettent-elles leur reconnaissance ? (Do vocal imitations enable the recognition of the imitated sounds?)

Guillaume Lemaitre[1,2], Arnaud Dessein[1], Karine Aura[3], Patrick Susini[1]

[1] STMS-Ircam-CNRS, Paris, France, {lemaitre, dessein, susini}@ircam.fr

[2] Carnegie Mellon University, Department of Psychology, Pittsburgh, USA, guillaum@andrew.cmu.edu

[3] Laboratoire Octogone-J. Lordat EA 4156, Université de Toulouse, France, aura@univ-tlse2.fr

Si nous faisons l'hypothèse qu'au cours d'une conversation courante, l'emploi d'imitations vocales par un locuteur permet à son interlocuteur d'identifier le son imité, il est alors fondé de penser que l'étude de ces imitations peut beaucoup nous apprendre concernant les mécanismes psychologiques d'identification des sons environnementaux. Nous présentons ici deux études qui investiguent cette question. Dans une première étude, nous avons demandé à des paires de participants d'écouter des séries de sons. Un des participants avait alors à décrire à l'autre un son sélectionné, et le second participant avait alors à deviner ce son. Les participants étaient absolument libres d'utiliser tous les moyens nécessaires pour décrire le son. Les résultats montrent que dans plus de 50% des cas, les participants ont spontanément utilisé des imitations vocales, et que l'emploi de ces imitations améliore les performances de reconnaissance. Dans une seconde expérience, nous avons enregistré des imitations vocales de sons environnementaux (bruits de cuisine). Ces sons présentaient l'intérêt d'avoir été étudiés dans un travail préalable, et leur catégorisation perceptive était donc connue. Nous avons alors demandé à des participants de classer ces imitations, en fonction de ce qu'ils croyaient être imité. Une analyse en clusters hiérarchique montre alors que les catégories d'imitations correspondent de manière remarquable aux catégories de sons imités. Ces catégories semblent définies par certaines caractéristiques acoustiques systématiques. Ce dernier point est confirmé, en montrant, à l'aide de techniques de classification automatique simples, que la catégorisation des imitations peut être prédite en se basant sur quelques descripteurs psychoacoustiques classiques. Ces résultats montrent donc que les imitations vocales contiennent l'information suffisante à la reconnaissance du son imité, et que cette information se caractérise simplement du point de vue acoustique. Ils laissent donc espérer que l'étude des imitations vocales de sons environnementaux va permettre de caractériser l'identification de ces sons.

1 Introduction

A question specific to the perception of environmental sounds is that of the acoustical information used by listeners to identify the source of the sounds, and to recover its properties. Usual experimental methods consist of categorization experiments and of analyses of the verbalizations produced by listeners required to describe sounds. Among the various linguistic devices used to described a sound, any informal observation would suggest that vocal imitations are commonly and spontaneously used during conversations. Imitations are a convenient device to emphasize and highlight the properties of a sound, thus allowing its recognition.Therefore, through the constraints of voice production and perception, vocal imitation can be thought of as mirroring the cognitive representations of sounds shared by speakers and listeners, thus enabling communication. As such, studying vocal imitations holds the promise of characterizing these representations and their relationships to the acoustical properties of the sounds. It is the goal of the experimental studies reported here with French speakers to examine the potentiality of such a promise.

1.1 Environmental sound perception

Human listeners are very good both at assessing the properties of an acoustical signal, and at identifying the properties of the mechanical event that have caused the sounds (see [11] for an overview). In other words, two different (even if they also may overlap) types of information are accessible to the listeners: the sound, and the event causing the sound.

1.2 Vocal imitations, onomatopoeias and sound symbolism

Onomatopoeias have probably been the most commonly studied types of vocal imitations. A very interesting definition is provided in [16]: "An onomatopoeia is a word that is *considered by convention* to be *acoustically similar* to the sound, or the sound produced by the *thing* to which it refers."

Three notions are important here: first onomatopoeias are words, the meaning of which is conventionally agreed among a given language: “meow”, “hiccup”, “honk” are instances of English words listed in dictionaries, and referring respectively to the sounds of a cat crying, an involuntary spasm of the respiratory organs, and a car horn. Second, this word refers to an object that makes a sound, or to the sound itself: “cuckoo” can refer to the sound made by a cuckoo, or to cuckoo itself. Third, the relationship between the word and the referent sound or sound event is based on some acoustical similarity (“its form is motivated by its content”, [19]). For instance, despite spelling differences, the English, French and Norwegian words for the cry of a rooster (“cock-a-doodle-doo”, “cocorico” and “kykkeliky”) display phonetic similarities clearly related to their referent sound. Onomatopoeias have therefore an interesting status for the study of language. Being conventional words, some authors have posited that they are a “bridge” between symbolic and sensorial representations [4], and to play an important role in the acquisition of language [7], as other forms of imitating behaviors do.

The sound symbolism of onomatopoeias has been studied for several Western languages [19, 21, 18, 2], or for more specific technical languages [14]. They show evidence of relationships between the phonetics of the onomatopeias (within the peculiarities of each language), and the acoustical properties of the referent sound. For instance, both [17] and [13] have submitted that there is evidence for a systematic mapping from certain categories of physical events onto strings of imitative phonemes, when examining English phonemic representations of inanimate sounds, as in comics. According to their studies, initial plosives are descriptive of sudden onsets and final nasals of sounds with a prolonged decay; and /i/-like vowels generally suggest brighter sounds than /u/-like vowels.

Japanese onomatopoeias have a specific status, and have been studied a lot. In fact, there exists in Japanese a large number of specific mimetic words: *giongo*, mimicking sounds; *giseigo*, mimicking voices; and *gitaigo*, mimicking manner, states of doing something. For the two former categories, a direct resemblance (iconic) between the word and the referent sound can be assumed, whereas for the latter category, the relation is more likely to be symbolic. [7] have experimentally shown that, when required to rate giongo on a set of semantic dimensions, English (with no proficiency in Japanese language) and Japanese listeners displayed consistent correlations, which was not the case for gitaigo. Particularly they agreed in the case of giongo for the scales related to the acoustical properties of the sounds, but not on the scales related to the aesthetic of the sound (i.e. scales referring to beauty, pleasantness, vulgarity, etc.). As it is the case for other languages, systematic relationships between the phonetical content of the words and the acoustical properties have been highlighted in Japanese [20]. However, the phonetic/acoustical relationships are not simply direct mappings. [21] showed for instance that stop consonants in English indicate the shortness and abruptness of the sound reffered to, provided they do not appear in words containing long vowels, diphthongs, prolonged consonants, or other restrictive elements.

In comparison to onomatopoeias, non-conventional vocal imitations have been rarely studied. Such imitations can be simply defined by dropping the first part of Pharies’ definition of onomatopoeias: a non-conventional imitation is a creative utterance intended to be acoustically similar to the sound, or the sound produced by the thing to which it refers. [9] showed that human-imitated animal sounds were well recognized by listeners, even better than the actual animal sounds [8], yet they did not have any problem to discriminate between the two categories [10]. This effect is probably close to that of sound effects used in movies and video games [5].

1.3 Scope and outlines of the study

Studying how speakers produce vocal imitations of environmental sounds is expected to provide insights into the properties of the sounds that are important for the identification of the event causing the sounds. In other words, we posit here that studying imitations will help understanding how these two types of information (the sound and the sound event properties) are tied together. The studies reported in this article examine some of the initial motivating assumptions.

2 Experimental study 1

The initial observation motivating this study was that, in an everyday conversation, people would spontaneously use vocal imitations or onomatopoeias when having to communicate a sound that they have heard, or that they have in mind. This section reports on an experimental study investigating this assumption.

2.1 Method

Participants Twelve participants (5 women and 7 men) volunteered for the experiment. They were aged from 26 to 45 years old. All reported having normal hearing and being French native speakers. They performed the experiment in couples. Three couples were made with people who already knew each other, and 3 couples with people who never met before.

Stimuli The stimuli were 30 monophonic sounds with a 16-bit resolution and a sampling rate of 44.1 kHz. They were divided into 3 groups of 10 sounds. The first two groups contained kitchen sounds, sampled from the corpus used in [11] and [6]. For these sounds, they measured the identification *confidence*, following [1]. This measure was used to select sounds in two groups: sounds with high confidence values, indicating that they were easy to identify (Group 1), and sounds with lower confidence values, indicating these were more difficult to identify (Group 2). The third group (Group 3) contained car horn sounds recorded and studied by [12]. These sounds have therefore the same kind of causing event: a driver (electrodynamical or pneumatic) loaded by a resonator (horn or metal plate). The only differences between these sounds are their pitch, and their timbre.

In each of theses groups, three target sounds were selected. These sounds were to be communicated by the participants during the experiment.

Procedure Two participants were invited in each session. They each had a different role (Participant 1 or 2) that was randomly attributed at the beginning of the session. The experiment was divided into nine series following the same procedure. Each series corresponded to one of the nine sounds to be communicated (three groups times three sounds). The procedure was the same for each series: Participant 1 was first isolated in a sound-attenuated booth and had to listen to every sound of the series. Then, the interface indicated her or him a target sound to communicate to Participant 2. Participant 1 heard this sound three times. Afterwards, she or he joined Participant 2 and had to communicate her or him the selected sound. The participants could freely talk, and were not specified how to communicate. Particularly, the possibility to use vocal imitations was not mentioned. Once the conversation finished, participant 2 was isolated in the sound booth. She or he had to listen to the ten sounds, and to select the sound that she or he believed was the one communicated by participant 1. The presentation order of the sounds in the interface was different for the two participants. The participants were told that the sound to be retrieved could be the same between different series.

2.2 Results

For each sound communicated, two indexes were collected: the presence or absence of vocal imitations or onomatopoeias during the conversation (wild and tame imitations were not distinguished here), and the correct or incorrect identification of the target sound. The former index was collected by the experimenters by a-posteriori analyzing and annotating the video recordings of the conversations. Three of the coauthors independently analyzed the videos. Their annotations were completely identical. The experiment had therefore a within-subjects design, with two independent variables (the sounds, and the category of the sound to be imitated), and two dependent variables: the presence of imitation, and the correct identification.

Presence of vocal imitation and onomatopoeias Vocal imitations and onomatopoeias were present in 59.3 % of the conversations. They therefore appear to be a common linguistic device spontaneously to communicate the sounds.

The presence of vocal imitation and onomatopoeias is not the same in the three groups of sounds. They occurred in 50% of the conversations for the easy-to-identify kitchen sounds (Group 1. high confidence scores), 72.2% of the difficult-to-identify kitchen sounds (Group 2, low confidence scores) and 55.6% for the car horn sounds (Group 3). Difficult-to-identify sounds therefore tend to have caused more vocal imitations and onomatopoeias, yet these differences are not statistically different (χ^2(1,N=18)=1.87, 0.115, 1.08, and p=0.17, 0.78 and 0.3 respectively, when contrasting Group1 vs. Group2, Group1 vs. Group 3, and Group 2 vs. Group3).

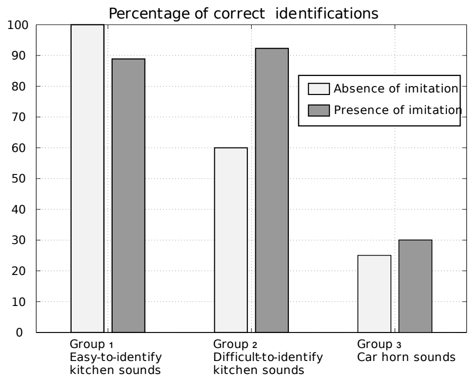

Figure 1: Percentages of correct identification in Experiment 1.

Correct identifications Communications of easy-to-identify kitchen sounds (Group 1) resulted in 94.4 % of correct identifications, and difficult-to-identify kitchen sounds (Group 2) in 77.8 % of correct identifications. For these two groups of sounds, identification of the communicated sound was good. Difficult-to-identify kitchen sounds (Group 2) resulted in slightly less correct identifications, but the difference is not statistically significant (χ^2(1,N=18)=1.12, p=0.29). Car horn sounds (Group 3) resulted in 27.8 % of correct identifications. As assumed the task was therefore more difficult for the car horn sounds, which were much more similar one to each other, and therefore difficult to differentiate verbally, than for the kitchen sounds (χ^2(1,N=18)=16.8 p<0.01 for Group 1 vs. Group 3, χ^2(1,N=18)=11.2, p<0.01 for Group 2 vs. Group 3).

Influence of the presence of imitations and onomatopoeias on the identification Though the experiment was not designed to systematically assess such a question[1], the collected indexes can be binned together so as to compare the percentage of correct identifications in the conversations that contained or did not contain vocal imitations or onomatopoeias. Figure 1 represents such a comparison. As noted above, easy-to-identify kitchen sounds (Group 1) resulted in a large number of correct identification, and car horn sounds (Group 3) to a fewer number of correct identifications. In both cases, the percentage of correct identifications does not depend on the presence or absence of imitations (Fischer's exact two-sided tests: p=0.50 and p=1.00). In the case of difficult-to-identify kitchen sounds (Group 2), there appears that the conversations including vocal imitations resulted in slightly more correct identifications than the conversations that did not, though is effect is not statistically significant (Fischer's exact two-sided tests: p=0.17).

[1] This issue resulted in some cells in the contingency tables with very few elements. Therefore Fisher's exact tests were preferred over Pearson's χ^2.

3 Experimental Study 2

Experimental Study 2 focused on how listeners categorize a set of vocal imitations of kitchen sounds. More precisely, we studied here only "wild imitations". Using a categorization task was motivated by the assumption that, if listeners are able to recover the sounds that are imitated, they should categorize the imitations in a similar fashion as they would categorize the imitated sounds. If kitchen sounds are categorized according to their causing events, so should be their imitations. The results of the categorization task were also used to fit a model that predicts the categories on the basis of some acoustical features of the imitations, thus providing some insights into the acoustical cues that are important for sound identification.

The stimuli used in Experimental Study 2 were based on vocal imitations of a set of kitchen sounds (the "referent sounds"). The selection of these sounds was made on the basis of the results of one of the categorization experiments reported in [6]. There were 12 referent sounds: 3 sounds were selected in category of sounds made by solid objects, and labeled S1 to S3; 3 sounds were selected in the category of liquid sounds (L1 to L3); 3 sounds were selected in the category of gases (G1 to G3); finally, 3 sounds were selected in the category of electrical appliances (E1 to E3).

3.1 Imitating the kitchen sounds

Twenty participants were hired (10 men and 10 women, aged from 18 to 50 years old). The participants were required to listen to the sounds. For each sound, they had to record three instances of an imitation of the sound. They were required to imitate the sounds "in such a way that another person could recognize it". They were instructed not to use any word or onomatopoeia.

A total of 720 (4 categories x 3 sounds x 20 participants x 3 trials) imitations were recorded. These recordings were edited and screened to remove those that were of poor quality, and those who included words or onomatopoeias. Eventually, only four categories of kitchen sounds (liquids-water, gas, solids-cutting and electrical), three sounds per category, and the imitations of six participants (three men and three women) were selected, making a total of 72 imitations.

3.2 Method

Participants Twenty participants (10 women and 10 men) volunteered for the experiment. They were aged from 18 to 50 years old.

Procedure The participants saw a white screen, on which red dots labelled from 1 to 72 were drawn, each dot corresponding to a sound. Participants were asked to move the dots to group together the sounds. They were allowed to form as many groups as they wished and to put as many sounds in each group as they desired. Participants were required to group together the vocal imitations "on the basis of what is imitated".

3.3 Analysis

The categorizations made by the participants were submitted to a hierarchical cluster analysis. The dendrogram of vocal imitations is represented in Figure 2. Exploring the clusters of vocal imitations highlights the principles that rule the organization of the vocal imitations, as well as the characteristics of the imitations potentially responsible for the clustering. Considering the dendrogram from the highest fusion level, the first branching-off distinguishes the imitations of gases from all the other sounds. The former imitations are clearly distinct from the others because of their breathy (unvoiced) character. The latter imitations are further divided into two clusters: on the left hand side, a cluster that includes a subcluster of mostly electrical sounds (characterized by the presence of a continuous steady pitch), and an hybrid subcluster that mostly includes imitations of liquid sounds (sound with rhythmic pitch). On the right hand side a cluster is further subdivided into a subcluster of imitations of solid sounds, and an hybrid cluster of liquid and solid sounds. These imitations have all in common to display a repetitive pattern. Thus, the initial branching-off of the dendogram results in four distinct and coherent clusters: the imitations of gases, electrical sounds, of some liquid sounds, and sounds of solid objects. The other imitations of liquids are rejected in a hybrid cluster, and mixed either with imitations of solids, or imitations of electrical sounds.

At the finest level of the hierarchy height clusters correspond to the four main categories of referent sounds:

- $\mathcal{G}$: imitations of gases
    - (1) $\mathcal{G}_1$ made of 6 imitations of the sound G_1;
    - (2) $\mathcal{G}_2$ made of 5 imitations of the sound G_2;
    - (3) $\mathcal{G}_3$ made of 5 imitations of the sound G_3;
- $\mathcal{E}$: imitations of electrical sounds
    - (4) $\mathcal{E}$ made of 12 imitations of E_1, E_2 and E_3;
- $\mathcal{L}$: imitations of sounds of liquids
    - (5) $\mathcal{L}'$ made of 6 imitations of L_1 and L_3;
    - (6) $\mathcal{L}_2$ made of 6 imitations of the sound L_2;
- $\mathcal{S}$: imitations of sounds of solid objects
    - (7) $\mathcal{S}''$ made of 8 imitations of S_1, S_2 and S_3;
    - (8) $\mathcal{S}'$ made of 5 imitations of S_1 and S_2.

One cluster ($\mathcal{X}$) is mathematically consistent, but includes imitations of different sounds.

Overall, 58 imitations (out of the 72) fall in the four clusters $\mathcal{G}$, $\mathcal{E}$, $\mathcal{L}$ and $\mathcal{S}$ made by grouping together the eight mathematically consistent clusters of imitations, and corresponding to the categories of referent sounds. Among these 58 imitations, only three are clustered in a cluster that does not correspond to the category of referent sounds. Therefore, if we consider the categories of referent sounds as an appropriate level of accuracy, 55 of the vocal imitations (76.4 %) were consistently classified. This indicates that, for a large majority of the imitations, listeners have been able to access to the category of the referent sound.

3.4 Acoustical properties of the imitations

The description of the clusters of imitations suggests that they might be characterized by a few distinctive

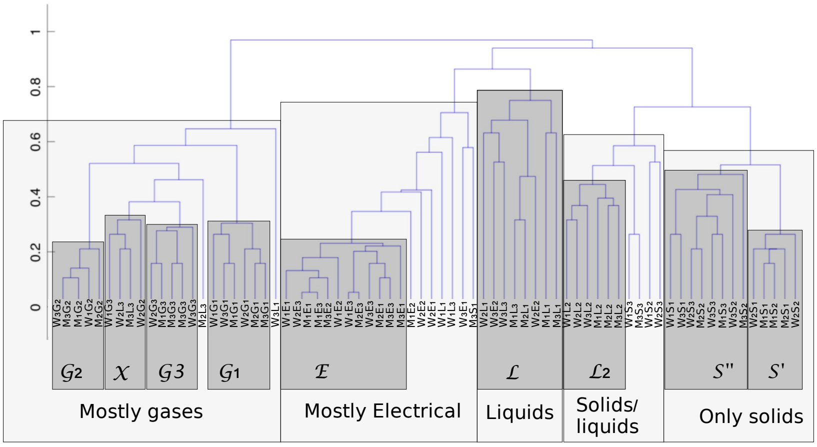

Figure 2: Dendrogram of the vocal imitations. The first letter of the index of each imitations describes the imitator (M stands for man, and W for woman), whereas the second letter describes the category of the imitated sound (L is for liquid, S for solid, M for machine and G for gases).

acoustical features. To uncover such features, the data were submitted to a binary decision tree analysis. The goal of such an analysis is to predict the classification of a set of clusters from a set of binary decision rules in a feature space. A binary decision tree is therefore the most simple model to fit the data, if we assume that the distinctions between the most consistent categories of imitations found in our data are clear-cut and depend only on a few acoustical properties of the imitations.

It is therefore important to choose categories of imitations that are compact and well distinct one from the others, and that correspond to a similar referent sound, or to sounds from a same category of referent sounds. We selected the 49 imitations in the 8 clusters $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$, $\mathcal{E}$, $\mathcal{L}_2$, $\mathcal{S}'$ and $\mathcal{S}''$ that fit these criteria. The acoustical features were computed with the IrcamDescriptor toolbox [15], and Yin algorithm to compute the fundamental frequency and the aperiodicity of the signal [3].

The binary decision was able to perfectly classify the 49 imitations in the 8 clusters, with the following rules:

- The *minimum aperiodicity* allowed one to distinguish the imitations of electrical appliances ($\mathcal{E}$) from all the other sounds.
- The *modulation amplitude* of the energy envelope then discriminated the imitations of gases ($\mathcal{G}$) from the imitations of sounds of liquids and solids ($\mathcal{R}$).
- The *temporal increase* of the energy envelope first discriminated between $\mathcal{G}_2$ on the one hand, and $\mathcal{G}_1$ and $\mathcal{G}_3$ on the other hand.
- The *standard deviation of the spectral centroid* distinguished $\mathcal{S}'$ from $\mathcal{L}_2$ and $\mathcal{S}''$.
- The difference between $\mathcal{G}_1$ and $\mathcal{G}_3$ was then simply captured by their *effective duration*.

4 General discussion

The study reported in this article started from the following observations. First, when required or willing to communicate a sound to a pair, a speaker very often makes use of vocal imitations to describe what she or he has in mind. Second, vocal imitations are an efficient linguistic device, because they enable the listener to recover what the speaker was intending to communicate.

We have reported the results of two experimental studies examining these ideas. The results of Experimental Study 1 clearly confirmed the intuition, that, despite the intimidating presence of the experimenter, people did spontaneously use vocal imitations and onomatopoeias to communicate a sound. Yet the experiment was not designed to statistically test such effects, the results nevertheless display some very interesting patterns: imitations were used more often to communicate sounds that were not easily namable, but at the same time not too difficult to vocalize. Furthermore the results suggested that a potential advantage of vocal imitations and onomatopoeias over verbal-only descriptions only occurred in this case: for very easily identifiable kitchen sounds, participants only had to name the sound event to allow perfect recognition. For the set of very similar car horn sounds, imitating the sound was helpless, probably because the timbre differences were too subtle to be vocally rendered.

Experimental Study 2 examined the second part of the assumption: that of the meaning conveyed by the vocal imitations. It showed that, when required to sort a set of “wild” vocal imitations of everyday sounds on the basis of what they thought was imitated, participants made categories of imitations that overall corresponded to the categories of the referent sounds. And most of the vocal imitations were grouped together in clusters corresponding to their referent sounds. The referent sounds had been chosen because they belong to contrasting categories of different sounding events. Therefore, these results suggest that the vocal imitations convey enough information for the listener to recover the causing events of the sounds. However, two other aspects are also worth mentioning. First, there were also imitations that were misclassified. Two explanations might be proposed: the sound events were not recognizable, and their imitations do not fit well in a general pattern where all the other sounds are organized with respect to their causing event; or these sounds could not be successfully imitated by the participants required to do so. Second, the fact that the resulting categorization could be perfectly predicted by a set of simple acoustical features (if we consider only the successful imitations) sheds an interesting light on these results: there is a perfect overlap between the acoustical similarity of the imitations and the causal similarity of referent sound events: for instance, all the imitations of the gases were unvoiced steady sounds. If we consider that the sounds produced by a same kind of mechanical event share a least a minimum set of common properties, these results would suggest that the imitations have been particularly successful in emphasizing these very features.

References

[1] James A. Ballas. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2):250–267, 1993.

[2] Susana M. Capitão Silva, Luis M. T. Jesus, and Mário A. L . Alves. The use of onomatopoeias to describe environmental sounds. In *Proceedings of the XXII Encontro Nacional de Associaçaõ Portuguesa de Linguistíca*, Lisbon, Portugal, 2007.

[3] A. de Cheveigné and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *Journal of Acoustical Society of America*, 111(4):1917–1930, 2001.

[4] Teruo Hashimoto, Nobuo Usui, Masato Taira, Izuru Nose, Tomoki Haji, and Shozo Kojima. The neural mechanism associated with the processing of onomatopoeic sounds. *Neuroimage*, 31:1762–1170, 2006.

[5] Laurie Heller and Lauren Wolf. When sound effects are better than the real thing. In *Proceedings of the 143^{rd} ASA meeting*, Pittsburgh, PA, 2002.

[6] Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis, Patrick Susini, and Isabel Urdapilleta. Categorization of environmental sounds, focus on sound event categories. 2010. Unpublished paper.

[7] Noriko Iwasaki, David P. Vinson, and Gabriella Vigliocco. What do English speakers know about *gera-gera* and *yota-yota*? A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese-language education around the globe*, 17:53–78, 2007.

[8] Norman J. Lass, Sandra K. Eastham, William C. Parrish, Kathleen A. Sherbick, and Dawn M. Ralph. Listener's identification of environnmental sounds. *Perceptual and Motor Skills*, 55:75–78, 1982.

[9] Norman J. Lass, Sandra K. Eastham, Tammie L. Wright, Audrey HR. Hinzman, Karen J. Mills, and Amy L. Hefferin. Listener's identification of human-imitated sounds. *Perceptual and Motor Skills*, 57:995–998, 1983.

[10] Norman J. Lass, Audrey HR. Hinzman, Sandra K. Eastham, Tammie L. Wright, Karen J. Mills, Bonita S. Bartlett, and Pamela A. Summers. Listener's discrimination of real and human-imitated sounds. *Perceptual and Motor Skills*, 58:453–454, 1984.

[11] Guillaume Lemaitre, Olivier Houix, Nicolas Misdariis, and Patrick Susini. Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: applied*, 2010. In press.

[12] Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg, Boris Letinturier, and Stephen McAdams. The sound quality of car horns: a psychoacoustical study of timbre. *Acta Acustica united with Acustica*, 93(3):457–468, 2007.

[13] Robert L. Oswalt. Inanimate imitatives. In L. Hinton, J. Nichols, and J. Ohala, editors, *Sound Symbolism*. Cambridge University Press, 1994.

[14] Aniruddh Patel and John Iversen. Acoustical and perceptual comparison of speech and drum sounds in the North India tabla tradition: an empirical study of sound symbolism. In *Proceedings of the 15^{th} International Congress of Phonetic Sciences*, Barcelona, Spain, 2003.

[15] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Cuidado projet report, Institut de Recherche et de Coordination Acoustique Musique (IRCAM), Paris, France, 2004.

[16] D. A. Pharies. *Sound symbolism in the Romance languages*. PhD thesis, University of Columbia, Berkeley, 1979.

[17] R. Rhodes. Aural images. In L. Hinton, J. Nichols, and J. Ohala, editors, *Sound Symbolism*. Cambridge University Press, 1994.

[18] Vincent Rioux. Methods for an objective and subjective description of starting transcients of some flue organ pipes integrating the view of the organ-builder. *Acustica united with Acta Acustica*, 86(4):634–641, 2000.

[19] Włodzimierz Sobkowiak. On the phonostatistics of English onomatopoeia. *Studia Anglica Posnaniensia*, 23:15–30, 1990.

[20] Masayuki Takada, Kazuhiko Tanaka, and Shin ichiro Iwamiya. Relationships between auditory impressions and onomatopoeic features for environmental sounds. *Acoustic Science and Technology*, 27(2):67–79, 2006.

[21] Rafał Żuchowski. Stops and other sound-symbolic devices expressing the relative length of referent sounds in onomatopoeia. *Studia Anglica Posnaniensia*, 33:475–485, 1998.