

Acoustics of speech and environmental Sounds

Susana M. Capitão Silva¹, Luis M. T. Jesus² and Mário A. L. Alves³

¹Secção Autónoma de Ciências da Saúde, Universidade de Aveiro and Direcção Regional da Educação do Norte, Ministério da Educação, Portugal

²Escola Superior de Saúde da Universidade de Aveiro and Instituto de Engenharia Electrónica e Telemática de Aveiro, Universidade de Aveiro, Portugal

³Hospital Senhora da Oliveira, Guimarães, Portugal

Abstract

In this study we present a preliminary acoustic analysis of environmental sound stimuli, based on Gaver's (1993) classification. Results showed similarities between sounds produced by objects with the same type of material and interaction. We also designed an experiment, where the subjects had to propose onomatopoeic representations for environmental sounds. The onomatopoeic representations used by the subjects shared common speech features (manner and place of articulation, and vowels used).

Introduction

In our everyday life we hear different kinds of sounds. The most common and to which we are exposed since birth are the environmental sounds. It is difficult to describe an environmental sound, but when we do, onomatopoeic representations are often used, and thereby speech sounds are used to report environmental sounds.

Environmental sounds acoustic features have been studied by various authors (Gygi, Kidd, and Watson 2004). According to Gaver (1993), musical listening and everyday listening are quite different processes: when listening to music we analyse features of the sound itself; when we try to perceive environmental sounds we listen to events and not sounds, we pay attention to what is generating the sound and not the emotional sensations or acoustic features conveyed by the sound. Therefore, research methods should consider the materials involved (solids, liquids or gases), and the type of interaction (solids: impacts, scrapping, rolling, deformation; liquids: drip, pour, splash, ripple; gases: explosion, gusts, wind).

Onomatopoeic representations of environmental sounds have been recently analysed by Takada, Tanaka and Iwamiya (2006), for a set of commercially available stimuli. The authors asked their subjects to describe the sounds according to quality rating scales and using onomatopoeic represen-

tations. Results showed similar acoustic properties in the stimuli expressed by onomatopoeic representations classified into the same clusters.

Environmental Sounds Stimuli

We selected 26 stimuli from audio recordings commonly used in clinical practice by Portuguese speech and language therapists. The sample contained a total of 42 environmental sounds that included sounds of animal vocalizations, sounds produced by the human body, sounds of nature, sounds of objects and sounds of means transportation. The selected stimuli were classified according to the type of interaction occurring between materials (Gaver 1993), which can be solid, liquid or gases. Most of the sounds consist of a single event, i.e., involved only one type of interaction between materials, but there were some stimuli that involve two types of interaction. There were 14 stimuli generated by solids, 6 by liquids and 6 by gases.

Method

Time waveforms and spectrograms of all stimuli were analysed using *Praat*. The following acoustic measures were extracted: duration, and F0, F1, F2 and F3 for periodic signals. Most of the sounds were noise signals, so multi-taper spectra (PSD Thomson estimates) were calculated with 11 ms windows left aligned to the start of the samples using *Matlab*, and peaks and broad peaks in the spectra were analysed.

Ten subjects (5 male and 5 female) with normal hearing abilities participated in this study. Subjects were asked to describe the selected stimuli using an onomatopoeic representation. Participants listened to the stimuli through Senheiser eH 1430 headphones. The stimuli were presented in the randomized order, and the subjects were able to listen to the sound stimuli for as many times as they felt necessary. Onomatopoeic representations produced by the subjects were recorded using a Philips SBC ME 400 unidirectional condenser microphone connected to a PC through a Sound Blaster Live! 24 bit soundcard.

The onomatopoeic representations were coded using 15 phonetic parameters: 5 places of articulation (bilabial, labio-dental, alveolar, palatal and velar), 6 manners of articulation (plosive, fricative, liquid), voiced, voiceless, nasal, and 4 vowel features (Group 1 - /ə, a, ε, e/; Group 2 - /i/; Group 3 - /u, o, ɔ/). These groups were based on the results presented by Takada, Tanaka and Iwamiya (2006), i.e., they were the most relevant features characterizing onomatopoeic representations.

Results

Temporal and spectral analysis of the selected sounds revealed some acoustic characteristics that allowed us to establish relations between speech and environmental sounds. The acoustic effects of source attributes proposed by Gaver (1993) were used as a reference, and new frequency and temporal domain characteristics of sounds generated by solids, liquids and gases with specific properties, were defined.

Duration was analysed according to the type of material producing the acoustic event. Results revealed that sounds produced by solids are shorter (specially the ones caused by a deformation) than sounds produced by liquids and gases (continuous sounds). The acoustic properties of sounds generated by an impact between solids are related to the acoustic properties of stops. In both cases we have an interaction between two solids that involves some sort of deformation. Aerodynamic sounds can be generated by an explosion (similar to the stops' plosion) or by more continuous sources (e.g. wind) which are similar to those used to produce fricatives. Liquid sounds, like those generated by water dripping, are produced by resonance cavities, with slight variations of pitch. When the interaction is a splash the sound is continuous like in fricatives, and when it is generated by waves or by water that is poured, those sounds share characteristics with laterals.

Sounds resulting from impacts of solids presented an acoustic aperiodic signal with high average amplitude which decayed over time. Most of the frequency components of the noise signal produced by a river were located above 1 kHz and by the wind below 1 kHz, as shown in Figure 1.

We observed that most onomatopoeic representations of solids used unvoiced alveolar stops and vowels from Group 3, as shown in Figure 2. An example of an onomatopoeic representation would be [tək tək].

Liquids were also predominantly represented by an alveolar place of articulation, but palatal consonants were also used. Voiceless fricatives are often used, indicating the absence of a periodic source.

Sounds produced by gases interactions were represented by alveolar and palatal places of articulation (mostly voiceless fricatives). Nasal consonants were used more often than for any other group. Vowels from Group 3 were also used. One example of onomatopoeia would be /fu:/.

Conclusions

This paper proposes novel ways of understanding how speech perception relates to environmental sounds perception, and presents a preliminary acoustic description of different categories of environmental sounds. Results showed that sounds produced by basic level events generated by similar materials or interactions shared common acoustic properties.

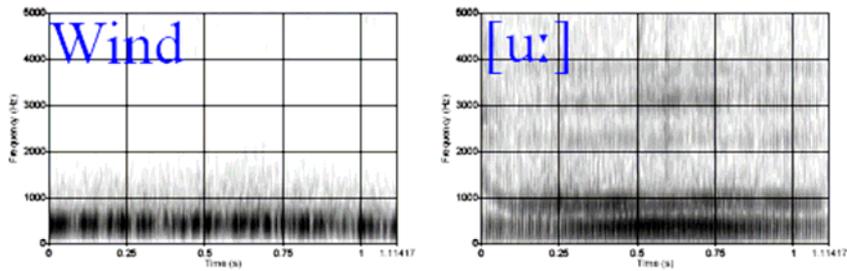


Figure 1. Spectrogram of stimulus “wind” and its onomatopoeia [u:].

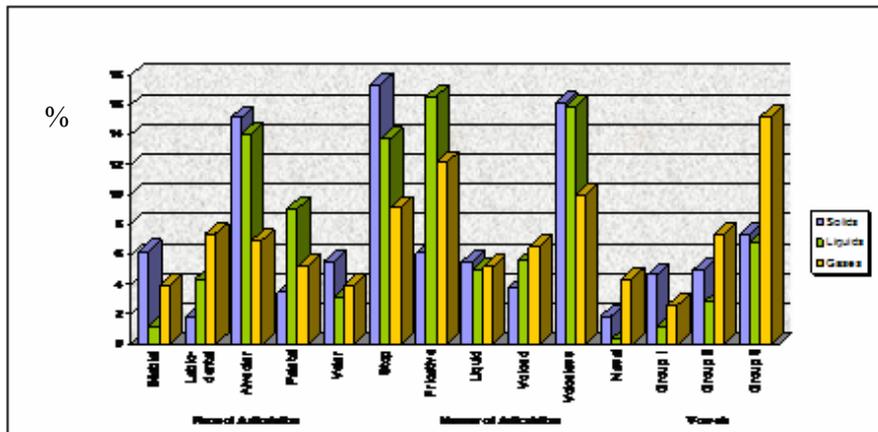


Figure 2. Phonetic features used to classify onomatopoeic representations of the different types of material: solids, liquids and gases.

References

- Gaver, W. 1993. What in the world do we hear?: an ecological approach to auditory event perception. *Ecological Psychology* 5(1), 1-29.
- Gygi, B., G. Kidd, C. Watson 2004. Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustic Society of America*, 115 (3), 1252-1265.
- Takada, M., K. Tanaka, S. Iwamiya 2006. Relationships between auditory impressions and onomatopoeic features for environmental sounds. *Acoustic Science Technology* 27(2), 67-79.