Sonifying drawings: characterization of perceptual attributes of sounds produced by human gestures

Etienne Thoret, Mitsuko Aramaki, Richard Kronland-Martinet, Jean-Luc Velay, Solvi Ystad

ACOUSTICS 2012

Sonifying drawings: characterization of perceptual attributes of sounds produced by human gestures

E. Thoret[a], M. Aramaki[a], R. Kronland-Martinet[a], J.-L. Velay[b] and S. Ystad[a]

[a]Laboratoire de Mécanique et d'Acoustique, 31, Chemin Joseph Aiguier - 13402 Marseille Cedex 20

[b]Institut de neurosciences cognitives de la méditerranée, 31 Chemin Joseph Aiguier 13402 Marseille Cedex 20

thoret@lma.cnrs-mrs.fr

Friction sounds produced by the pencil of a person who is drawing on a paper are audible and could bring information about his gestures. Here we focused on the perceptual significance of the morphology of such sounds, and to what extent human gestures could be retrieved by sounds. Inspired by a study dealing with visual perception of a moving point light which exhibited a 2/3-power law relation between kinetics and shape curvature for the perception of a constant velocity, experiment was carried out with a synthesis model of friction sounds. Subjects were asked to adjust the power law exponent to obtain the most realistic sound. Results revealed an exponent close to 2/3 as found in vision and thus highlighted a similar power law in the auditory modality, providing an ecological way to determine velocity profiles from shapes and to generate sounds coherent with human gestures. Two association tests where subjects had to univocally associate friction sounds with different shapes were carried out. Sounds were recorded or synthesized from the velocity profile recorded during drawing sessions. Results showed that subjects were able to associate sounds with correct shapes and that the auditory characterization of the shape depended on the velocity profile.

1 Introduction

Ecological acoustics theory has for a long time dealt with the perception of everyday non speech sounds. In particular, many studies proposed different taxonomies to classify these sounds. The most suitable and widely used nowadays is probably the one introduced by Gaver in [3] where everyday sounds are classified in three main categories: aerodynamic sounds, liquid sounds and vibrating object sounds (i.e. scraping, rolling or impact). The main auditory properties governing the perception of such sound events have been called invariants by ecological psychologists in [7]. They are classified in two categories; i.e. the structural invariants related to the recognition of the resonant object, for instance the material or the shape and transformational invariants related to the recognition of the means with which the resonant object was excited (the nature of the action). For instance, transformational invariants related to bouncing and breaking events were studied in [10] according to behavioral considerations. The authors have compared the perception of such events with regard to the differences between temporal sequences of impacts of such sonic events without considering the material or the shape of the impacted object.

In this study, we are interested in identifying transformational invariants dealing with the action of scraping by a human. From a physical point of view, when somebody is drawing or writing on a rough surface, the interactions between the pencil and the surface produce friction sounds. We aimed at investigating whether such sounds convey perceptual information on the nature of the underlying gesture and furthermore, on the characteristics of the drawn shape. Based on previous studies conducted in visual and motor modalities, we assumed that the velocity profile is a relevant attribute as a transformational invariant of sound perception produced by human gestures.

The paper is organized as follows: a physically-based model of friction sounds is firstly presented. This model allows synthesizing realistic friction sounds from high-level control parameters related to the nature of the scraped surface (perceived material, size and shape) and to the dynamical characteristics of the action (velocity and pressure). In this study, we assumed that the nature of the surface and the pressure are constant to focus on the influence of the velocity profile on the generated sound. In a second part, three listening tests designed to investigate the assumptions previously described will be presented. Finally, some applications using the obtained results will be mentioned.

2 Friction Sounds

Friction sounds produced when somebody is drawing or writing are due to the interactions between the lead of the pencil and the paper. The physical model of such sounds is complex and takes into account many mechanical and tribological aspects of the interaction. In this study we investigated the perception of these phenomena by using a simplified and efficient synthesis model where the dynamical variations of the gesture are taken into account. A physically based model of friction sound is used for this purpose where the pressure is assumed to be constant and only the kinematics of the gesture is taken into account.

2.1 A Physically Based Synthesis Model

A phenomenological model of friction sounds has been introduced by Gaver in [3] and improved by Van den Doel in [9]. This model considers that friction sounds results from successive micro-impacts of a plectrum on the asperities of the surface, and is based on a source-filter scheme. From a physical point of view, the source contribution is related to the excitation pattern and the filter represents the resonance of the vibrating surface. The profile of the surface is modeled by a noise where the heights of the asperities are linked to the roughness of the surface. Figure 1 sums up the general scheme of this model. The source consists in reading a noise-

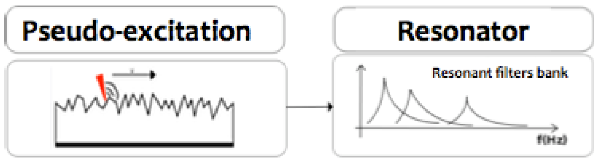

Figure 1: Physically Based Model of Friction

table with a velocity linked to the velocity of the plectrum. In practice, the noise is lowpass filtered with a cutoff frequency linked to the velocity of the plectrum. The resonator models the object which is rubbed. It is composed by a resonant filter bank with frequencies, gains and dampings of the filters linked to the modal response of the object.

2.2 Control

At low-level, the control parameters of the previous friction sound synthesis model are complex for non expert users. Easier ways to control the different parameters have already been studied and give perceptual controls on the material or on the roughness of the surface of the rubbed object:

- A way to control the resonant filter bank with regard to the perception of the material and the shape of the object is proposed in [1] and [2].

- The nature of the noise is controlled by the roughness of the surface. In [9], it is suggested to use a fractal noise which is a noise with a spectrum inversely proportional to a frequency power; $S(\omega) \propto \frac{1}{\omega^\beta}, \beta > 0$. From a physical point of view, the fractal noise was well adapted to model the asperities over different observation scales of the surface.The surface profile was smoother with increasing β-values.

In addition different dynamical parameters take part in the production of such sounds. Particularly the velocity and the pressure of the pencil on the paper are the main characteristics of the gesture which have audible effect on the produced sound. The velocity of the plectrum directly controls the reading rate if the noise table. This synthesis process gives an easy way to synthesis friction from a given velocity profile and to control the nature object which is rubbed[1]. The synthesis process with perceptual controls is summed up in figure 2.

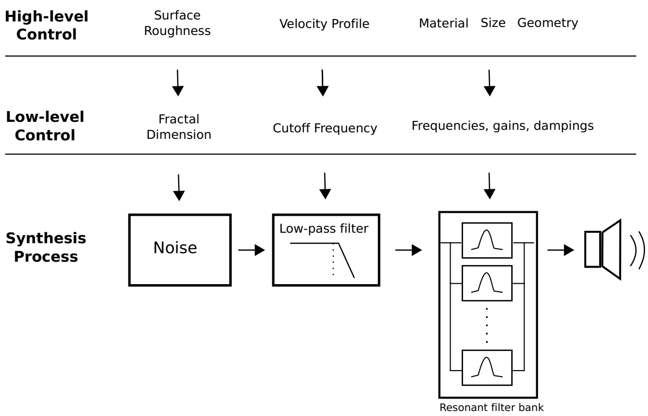

Figure 2: Sound Synthesis Model of Friction with High-Level Controls

3 Velocity Profiles of Gestures: a Transformational Invariant of Auditory Perception

In the following, we aimed at investigating the auditory perception of human gesture and the relevance of the velocity profile to induce the perception of human feature in the evoked motion. Moreover we wanted to find out if the auditory information transmitted by this velocity profile is sufficient to evoke a specific gesture. For this purpose, we used the friction sound synthesis model previously presented which enables to control independently each dynamical parameter of the gesture, and therefore to study its influence on the perception of the produced sound.

In a first experiment the relevance of a biological law linking the curvature of a shape to the velocity of the gesture which is drawing is investigated from the auditory point of view. This experiment aimed at highlighting that the velocity profile may follow specific characteristics to evoke a human gesture in the auditory perception of an action such as drawing.

In a second step, our ability to recover a static drawn shape from the generated friction sound is investigated in experiments 2 and 3. In particular, we hypothesized that the auditory information related to the produced gesture during the drawing is contained in the characteristics of the velocity profile. From a general point of view, this series of experiments aimed at supporting the assumption that the velocity profile is a relevant invariant for the perception of human gestures (we refer the reader to [6] for more details on methods, results and discussion).

3.1 Experiment 1

Many psychological studies have proved that biological movements are easily distinguishable from other movements in the visual system. In [5], an experiment where subjects had to indicate if they perceived a human walking only with point lights located at the main joints gave really convincing results in the sense of an integration of our biological movement in the visual perception system.

Different studies conducted by Viviani [6, 12] dealing with the motor competences in drawing tasks revealed that when somebody is drawing, the tangential velocity v_t of his or her gesture is linked to the curvature C of the drawn shape by the following relation called the 2/3-power law:

$$v_t(t) = KC(t)^{1-\beta}, \ \beta = 2/3 \tag{1}$$

where K is a constant linked to the mean velocity.

Viviani et *al.* also examined this law from the perceptual point of view. In [10], The authors highlighted that the perception of a shape traveled by a moving point light was closer to the actual shape when the point light kinematics followed the 2/3-power law. In [11], an experiment revealed that the perception of a constant velocity of a moving point light was constrainted by this biological law.

3.1.1 Method

Stimuli Friction sounds were synthesized with the synthesis model previously presented. The velocity profiles were generated with the power law in equation 1, with $K = 10\,\text{m.s}^{-1}$ and C the curvature of a pseudo random trajectory such as the one presented in figure 3. Since we did not aim at focusing on a specific shape, we decided to consider pseudo-random trajectories. The β value of the exponent was fixed by the subjects during the experiment.

Procedure Twenty one subjects took part in this experiment (9 females, 11 males, mean aged = 30.65 years). Subjects listened to the friction sound generated with the synthesis process and were asked to adjust the sound by two buttons ('<' and '>') until they assessed that the sound could have been produced by a human gesture. Subjects didn't know that they were adjusting the β-value of this law between 0 and 1.028 by step of 0.0416 acting on the two buttons. No visual stimuli were presented during the experiment. Each subject did six trials.

[1]Examples of synthesized friction sounds and stimuli of the following listening tests are available on the website: `http://www.lma.cnrs-mrs.fr/~kronland/SonifyingDrawings/`

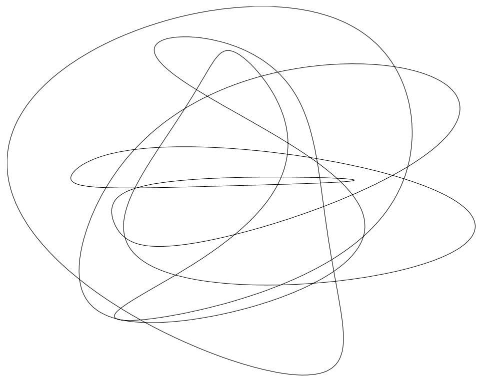

Figure 3: Example of Pseudo Random Trajectory

3.1.2 Results and Discussion

The β-values were averaged for each subject and then across all subjects. The mean β-value of the power law was found to be equal to 0.639 (SD=0.084). A one sample t-test with a population mean equal to 2/3 was conducted and no significant difference was found ($p = 0.14$). This means that the *auditory* β-value respects the theoretical β-value previously found for motor competences in drawing tasks and for the perception of constant velocity of a dynamical shape.

The results of this experiment argues in the sense that the velocity profile is an invariant of the auditory perception of our gestures. In particular, it reveals that a biological motion can be recovered in a friction sound if the velocity profile is modulated by the 2/3-power law.

3.2 Experiments 2 & 3

The first experiment focused on the possibility to recover a human gesture from auditory information. The good agreement of the results with the biological law encouraged us to go a step further in the investigation of auditory perception of velocity profiles of the gestures for specific shapes.

Then two following experiments were set up to investigate if different shapes could be discriminated from friction sounds produced when somebody is drawing. These listening tests were done with recordings and synthesized friction sounds generated from the previous model. These experiments aimed at highlighting the relevance of the velocity profile in the case auditory perception of sound produced by gestures. For that, the synthesized sounds were directly modulated from the velocity profiles. Besides, we aimed at bringing to light that the information, conveyed by these gestures' characteristics in the sound, is used by the auditory system to create a mental representation of the gesture underlying a friction sound.

3.2.1 Method

Subjects Twenty subjects took part in experiment 2 while eighteen took part in experiment 3.

Stimuli The stimuli were composed of recordings and synthesized friction sounds. Recordings were obtained from sounds produced by a writer who drawn two corpuses of shapes on a graphic tablet that enabled recordings of the velocity profiles for each shape. The first corpus, used in experiment 2, contained clearly distinct shapes from the auditory point of view. The second one, used in experiment 3, contained closer shapes. The two corpuses are presented in figure 4. The synthesized friction sounds were generated by using the previous synthesis model and the recorded velocity profiles.

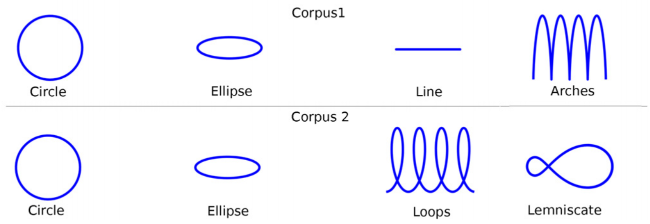

Figure 4: Corpuses of the Experiments 2 & 3

Procedure Each experiment was composed of eight trials. In each one, subjects had to univocally associate four friction sounds (recorded or synthesized) to the four visual shapes drawn. No time constraints were imposed.

3.2.2 Data Analysis

In each experiment, two confusion matrices were collected, one for the recorded sounds and one for the synthesized ones. These matrices represented the rates of association between sounds and shapes for each type of sound. The data analysis was conducted to answer two questions:

1. Were the friction sounds properly associated to the shapes?
2. Did synthesized and recorded sounds give similar results?

3.2.3 Results

1. For distinct shapes (experiment 2), all the sounds were associated to the correct shapes with rates of association higher than 80% for each shape for each type of sound. For closer shapes (experiment 3), each shape was associated with a rate significantly higher than statistical chance excepted the recorded *Loops*. For this corpus, significant confusions appeared between the *Loops* and the *Ellipse* both for synthesized and recorded sounds.
2. Both in experiment 1 and 2, the confusion matrices between synthesized and recorded sounds were significantly correlated. This means that recorded and synthesized sounds led to results that did not differ from each other.

3.2.4 Analysis of Velocity Profiles

Analysis of velocity profiles was further done to compare the results of the listening tests with an *objective* classification derived from the velocity profiles. Euclidean distances between shapes were computed from the velocity profiles that were firstly normalized and resampled on 512 points.

Statistical comparisons between the obtained distance matrix and the ones computed from the confusion matrices derived from perceptual results revealed that there were no significant differences between the perceptual and the *objective* classification.

3.2.5 Discussion

Based on the high rates of association obtained for most shapes, these experiments revealed that subjects were generally able to retrieve a drawn shape by processing the produced friction sound. Since the synthesized sounds were generated by controlling the parameter related to the velocity profile of the gesture, some features contained in the recorded sounds not reproduced in the synthesized sounds. Nevertheless, the lack of differences between results from recorded and synthesized sounds (for both experiments 2 and 3) allowed us to conclude that the velocity profile contains the necessary information to deduce the drawn shape from the sound. Moreover, the distance matrix computed from the euclidean distance between velocity profiles enabled to predict the perceptual classification obtained from the recorded sounds. To summarize, these results supported the hypothesis of velocity profile as a relevant transformational invariant in the auditory perception of drawn shapes by human gestures.

4 Sonification Perspectives

Results from the three experiments presented above allowed us to address different applications in the sound design domain. Based on these findings, we proposed a sonification strategy to generate a friction sound evoking the drawing by a human of a shape exclusively its geometrical characteristics. For that, the curvature profile of the shape was firstly determined from a starting point and a direction along the shape defined by the user. Then, the velocity profile was computed from the 2/3 power law and finally used as a dynamical parameter of the synthesis mode previously presented to generate the friction sound. In addition to a real-time control of the synthesis model for instance from gesture capture devices (graphic tablet, camera ...), this sonification strategy offers a new offline control based on the analysis of a given shape to enable the generation of a sound evoking drawings produced by a human.

5 Conclusion

The experiments presented in the article led to the suggestion that the velocity profile of a human gesture is a transformational invariant of the auditory perception. Experiment 1 highlighted that sounds may present specific features related to the velocity profiles that enable the recognition of human gestures. In particular, it revealed that the 2/3 power law previously found in visual and motor modalities, was also valid in the auditory modality, and that the velocity profile of human gestures is constrained by the geometrical curvature of the drawn shape. This aspect supported the motor theory of perception [4] which argues that the way we perceive stimuli from our everyday world is influenced by the way we interact with the mechanism which produced the stimuli. Moreover, results from experiments 2 and 3 revealed that sounds synthesized from the velocity profiles were correctly associated with the corresponding shape and consequently, enabled the evocation of a specific drawing gesture. From a cognitive point of view, we assumed that the mental association between sound and shape is based on the internal representation of the underlying gesture activated during the sound processing. This fundamental question linked to the auditory-motor relationship is currently investigated and may probably be addressed by considering an amodal representation of the human gesture.

Acknowledgments

This work is supported by the *MetaSon - Sonic Metaphors* Project – CONTINT 2010: ANR-10-CORD-010 - French National Research Agency (ANR).

References

[1] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad. Controlling the Perceived Material in an Impact Sound Synthesizer. IEEE Transaction On Speech and Audio Processing, 19(2) :301–314, February 2011.

[2] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier and S. Ystad. Thinking the sounds: an intuitive control of an impact sound synthesizer. Proceedings of ICAD 09 – 15th International Conference on Auditory Display, 2009.

[3] W. W. Gaver. What do we Hear in the World: an Ecological Approach to Auditory Event perception, Ecological Psychology, Vol. 10, No. 5, 704-712, 1984

[4] J. J. Gibson. The Senses Considered as Perceptual Systems . Boston: Houghton-Mifflin, 1966.

[5] G. Johansson. Visual perception of biological motion and a model for its analysis. Attention, Perception, & Psychophysics, 14(2) :201–211, 1973.

[6] F. Lacquaniti, C. A. Terzuolo, and P. Viviani. The law relating kinematic and figural aspects of drawing movements, Acta Psychologica, 54, 115–130, 1983.

[7] S. E. McAdams and E. E. Bigand. Thinking in sound: The cognitive psychology of human audition. Oxford Science Publication. Chapter 6, 1993.

[8] E. Thoret, M. Aramaki, R. Kronland-Martinet, J. L. Velay, S. Ystad. From Sound to Shape: auditory perception of the velocity profile of the gesture underlying a friction sound. *To be submitted in* Journal of Experimental Psychology: Human Perception and Performance.

[9] K. Van Den Doel, P. G. Kry and D. K. Pai. FoleyAutomatic : physically-based sound effects for interactive simulation and animation. Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 537–544, ACM, 2001.

[10] P. Viviani, and N. Stucchi. The effect of movement velocity on form perception: Geometric illusions in dynamic displays. Attention, Perception, & Psychophysics, 46(3) :266–274, 1989.

[11] P. Viviani, and N. Stucchi. Biological movements look uniform: Evidence of motor-perceptual interactions. Journal of Experimental Psychology: Human Perception and Performance, 18, 603–623, 1992.

[12] P. Viviani, and T. Flash. Minimum-jerk, two-thirds power law and isochrony: Converging approaches to movement planning. Journal of Experimental Psychology: Human Perception and Performance, 21, 32–53, 1995.

[13] W. H. Warren, and R. R. Verbrugge. Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics Journal of Experimental Psychology: Human Perception and Performance, 21, 32–53, 1995.