Synchronizing Gestures with Friction Sounds: work in progress

Etienne Thoret[1,3,4], Mitsuko Aramaki[1,3,4], Christophe Bourdin[2,3], Lionel Bringoux[2,3], Richard Kronland-Martinet[1,3,4], and Sølvi Ystad[1,3,4]

[1] CNRS - LMA - UPR 7051
[2] CNRS - ISM - UMR 7287
[3] Aix Marseille Univ.
[4] Ecole Centrale Marseille
thoret@lma.cnrs-mrs.fr

Abstract. This paper presents a work in progress dealing with the sensorimotor relation between auditory perception and graphical movements. An experiment where subjects were asked to synchronize their gestures with synthetic friction sounds is presented. A first qualitative analysis enabled to evaluate the influence of different intrinsic sound parameters on the characteristics of the synchronized gesture. This preliminary experiment provides a formal framework for a wider study which aims to evaluate the relation between audition, vision and gestures.

Keywords: Auditory perception, sensorimotor loop, friction sounds, gestures

1 Introduction

When we are interacting with the material world, we often think that only the visual modality is engaged to guide our actions, like for instance when we are walking, reaching a glass or drawing on a paper. The use of other modalities, like the proprioceptive system or the audition is more subtle and not as well conscious as vision which seems to occupy most of our attention. Nevertheless, several studies have shown our ability to recognize a geometric shape only by touching it [14], or to recognize events only from the sounds they produce [6, 7]. For instance, we are able to recognize and re-enact characteristic walking patterns from listening to the produced footsteps [21]. Other studies have focused on the auditory perception of biological movements in the case of graphical gestures. These studies revealed that we are able to identify a gesture, and to a certain extent the shape, drawn by a human only from the friction sound generated by the pen mine rubbing the paper [15]. Finally, the relations between

In this article, a work in progress about the direct relationship between the auditory system and the graphical motor competency is presented. Our aim is to investigate how audition can guide a graphical gesture in a synchronization task. The interaction between vision or proprioception and movements has already been studied. In different seminal studies, Viviani and colleagues have largely investigated such relations. For vision, they showed that we are more accurate when we follow a spotlight target which respects the dynamics of biological movements, also called the 1/3 power law. This law links the velocity of a movement to the curvature of its trajectory [17, 10]. The blindfolded manual reproduction of a kinesthetic stimulus has also been studied in another experiment [20] that likewise revealed that subjects based the reproduction of the kinesthetic stimulus on the velocity of the target.

From the auditory point of view, the manual production associated to the timbre variations of an acoustical stimulus has not been studied in the same way. Some studies investigated the relation between a sound and a gesture in specific situations. In particular, the case of musical gesture has been widely studied [12, 11, 8]. Such relations were for instance investigated in an experiment where subjects were asked to imitate a musical excerpt in three dimensions with their hands [4]. This study revealed that the subjects synchronized gesture parameters – mainly the velocity and the acceleration – on specific sound descriptors, mainly the pitch and the loudness. In another study, Caramiaux et al. [3] asked subjects to imitate environmental sounds with their hands, in order to distinguish different behaviors according to the *causality* of the imitated sound event.

In this article, we aimed at investigating the relation between sound and graphical movements with calibrated auditory stimuli which clearly evoked motions. In the case of vision, it is easy to create a calibrated stimulus which evokes a movement and to control its velocity using for instance a moving spot light. In the case of auditory stimuli, the problem is less simple. As mentioned before, the friction sound produced by a mine pen rubbing a paper clearly evokes a movement and has already been used to investigate the gesture evoked by such friction sounds [15]. Here, we used the same kind of sound to create calibrated auditory stimuli as acoustical targets evoking a specific movement. Rather than using recorded frictions sounds of someone drawing, we will use synthesized ones. Synthesis indeed enables to precisely control the friction sound produced by an object rubbing a surface by isolating one aspect for instance the kinematic linked to the movements that produced the sound.

This paper therefore presents an experiment where subjects were asked to synchronize their gestures on different synthesized friction sounds. They were asked to draw ellipses or circles according to the sound, and to translate the evocations of the sounds in their graphical productions. In particular, we investigated the influence of different mapping strategies between the gesture velocity and the sound parameters on the motor production task. The paper is organized in two parts, the synthesis process is firstly presented, and in a second time the experiment is presented and briefly discussed according to a first qualitative analysis.

2 Synthesis of Friction Sounds

In this study, timbre variations of friction sounds will be used as the acoustical target evoking movements. Such sounds naturally evoke movements according to their timbre variations without any spatialization processes. To synthesize such variations, we used a phenomenological model that has been proposed by Gaver [6] and improved by Van den Doel [16]. This model supposed that the sound produced by the friction between a plectrum and a rough surface results from a series of impacts on a resonator. The quicker the plectrum rubs the surface, the higher the number of impacts, and therefore the higher the pitch of the sound. From a signal point of view, the surface can be modeled by a noise[5]. And finally, the friction sound produced by the motion of the plectrum on the surface can be synthesized by low pass filtering the noise with a cutoff frequency proportional to the velocity of the plectrum. The modeling of the resonator is done by filtering the low pass filtered noise with a resonant filter bank tuned to the modal characteristics of the rubbed object [1].

Mapping Strategy. The synthesis process necessitates the determination of a mapping between gesture (velocity) and sound (cutoff frequency of the lowpass filter) parameters. The relationship between these parameters can be illustrated by the following example. We consider a simple surface, with regularly spaced asperities separated from a distance d_0, as presented in figure 1. If the surface is rubbed at a velocity v_0, the pitch of the produced sound will be proportional to v_0 and inversely proportional to d_0. If the surface is now rubbed at a velocity that by two times greater, the pitch of the friction sound produced will be doubled. In the same way, if the roughness d_0 is divided by two, the pitch will be doubled. The cutoff frequency of the law pass filter is linked to the velocity of the

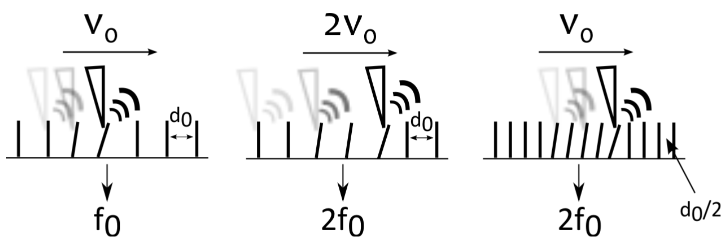

Fig. 1. A simple example of the phenomenological model of friction sounds. The three cases presented illustrate the ambiguity between velocity and roughness in the produced friction sound.

plectrum by a proportionality coefficient α: $f_c(t) = \alpha v_T(t)$. As explained before, this coefficient is linked to the surface roughness, and thus the cutoff frequency of the low pass filter can be seen from two angles as it encompasses two physical effects: the roughness of the surface and the relative velocity of the pen that interacts with the paper. Finally, when we listen to such a friction sound, there

[5] The more classical model of roughness for a surface is the fractal one, whose spectrum is defined by $S(\omega) = \frac{1}{\omega^\beta}$. When β is null, the noise is white, when β equals 1 the noise is pink. The higher the β the smoother the modeled surface is [16].

is an ambiguity between about the conveyed information. When increasing α, do we imagine that the surface because rougher, or do we imagine that the rubbing is twice as faster?

3 Experiment

The goal of the experiment was to evaluate the characteristics of the evocation induced by a friction sound using a synchronization task between a graphical gesture and a friction sound in open loop (i.e. the subjects were blindfolded during the task). It was effectuated in different acoustical conditions corresponding to different mappings between the velocity and the cutoff frequency to evaluate their influence on the produced graphical movements.

3.1 Methods

Subjects and Apparatus. Twelve participants took part in the experiment: two women and nine men. The average age was 24.17 years (SD=2.55). All the participants were right handed. None of the subjects were familiar with the topic of the study before the test. The subjects were blindfolded in front of a desk. The sounds were played through Sennheiser HD-650 headphones. The graphical gestures were collected through a Wacom Intuos 5 graphic tablet at a time rate of 133 Hz and with a spatial precision of 5.10^{-3} mm.

Geometric Shapes. Two geometric shapes were used. An ellipse of eccentricity 0.9 and semi-major axis of 9.05 cm, and a circle with a radius of 6.36 cm (i.e. an ellipse with null eccentricity and a semi major axis of 3.18 cm). The perimeters equal to 43.86 cm for the ellipse and 40 cm for the circle.

Velocity Profiles. Both shapes were defined according to the Lissajous motion:

$$\begin{cases} x(t) = a\cos\left(\frac{2\pi}{T}t\right) \\ y(t) = b\sin\left(\frac{2\pi}{T}t\right) \end{cases} \tag{1}$$

where a and b are respectively the semi-major and semi-minor axis of the ellipse (equal in the case of a circle). The chosen period T was 1.8 seconds, and 19 periods were generated. Thus, the durations of the stimuli were all equal to 34.2 seconds. Such a configuration of an ellipse (a fortiori a circle) implies that the motion follows the 1/3 power law (i.e. a biological motion). It is important to note that in the case of the circle, the tangential velocity is constant over the entire trajectory. In the case of the ellipse selected for the experiment, it varies between 13.88 $cm.s^{-1}$ and 31.66 $cm.s^{-1}$. In the case of the circle, it's constant and equals 22.7 $cm.s^{-1}$.

Acoustical Stimuli. Synthesized friction sounds were generated with the phenomenological model previously presented, and from the velocity profiles defined in the previous paragraph. The role of the mapping coefficient α is here evaluated, we arbitrarily chose 6 different values: 5, 10, 20, 50, 100 and 300 Hz.s.m^{-1}. These values provide friction sounds with very different timbres and influences mainly brightness of the friction sound. The higher the α the brighter the sound. The table 1 presents the minimal and maximal values of the cutoff frequency induced by these different values of α. Finally, 12 stimuli were generated: 2 (shapes) x 6 (mappings).

	Ellipse		**Circle**
α	f_{min}	f_{max}	f
5	69	159	111
10	138	318	223
20	276	635	445
50	690	1588	1135
100	1380	3176	2227
300	4140	9528	6681

Table 1. Minimal and maximal values of the low pass filter cutoff frequency in Hertz. In the case of the circle, the cutoff frequency is constant during all the stimuli as the velocity is constant.

Task and Procedure. The task consisted in drawing a shape – a circle or an ellipse – while being guided by the friction sound played through the headphones. The subjects were asked to synchronize their movement the sound variations in the counterclockwise direction during the 34.2 seconds of the friction sound. To investigate the direct relationship between the auditory modality and the evoked gesture, subjects were blindfolded throughout the duration of the experiment (also called in open loop) and encouraged to translate the evocation of the timbre variations in their production. It was explicitly asked to lift the elbow to make sure that the joint used during the movement involved both the shoulder and the elbow, and not the handle as during a handwriting task. The test was preceded by a training phase during which subjects would train on an example of such a circle and an ellipse. It was explicitly stated to the subjects that these two shapes were present. During the training phase, subjects could also adjust the height of the seat. Finally, each subject performed 36 trials: 2 (shapes) x 6 (mappings) x 3 (repetitions), which were randomized across the trials and the subjects. The subjects were not aware about the size of the shapes nor the orientation. Each performance lasted about 45 seconds and the subjects were encouraged to make a break when they felt the need.

Data Analysis. Data analysis focuses on two characteristics of the graphic productions: geometric characteristics and geometrico-temporal relationships. A data preprocessing is performed prior to the study of these two characteristics in order to overcome from the digital artifacts that appears when calculating the first and second derivatives of the data. The methods for calculating the different descriptors have been established in various articles of Viviani and colleagues [19, 20, 13]. Recorded data correspond to the coordinates of the stylus position on the tablet over time.

Pre-processing. A smoothing of the data is performed by low pass filtering the data at 10 Hz. Moreover, since subjects were blindfolded during the task, their graphic productions are not spatially accurate during the entire trace, and a low frequency deviation appears. In order to avoid the deflection of the center of gravity over time, a high-pass filtering is performed at 0.5 Hz on the recorded coordinates.

The recordings lasted 34.2 seconds, only 12 periods were selected, which corresponds to 21.6 seconds. Since it took some time for the subjects to produce a regular and synchronous movement with the sound (in the case of ellipses), the first six periods were excluded from the analysis, which corresponds to the first 10.8 seconds.

Geometric Characteristics:

- *The average eccentricty.* Eccentricity of an ellipse is defined with the following formula: $e = \sqrt{\frac{a^2-b^2}{a^2}}$, where a and b are respectively the semi major and minor axis of the ellipse. To determine the average eccentricity of the drawn shape, we used a method proposed by Viviani [19]. We considered each recording as a group of pointlike unitary masses and we computed the inertial tensor. It is well known in classical mechanics that the inertial tensor of a two-dimensional structure can be modeled by an ellipse characteristics are linked to the eigenvalues of the inertial tensor. The precise method is not described here, but we refer to the Viviani studies and to the Goldstein book for more details [9].

- *The average perimeter.* This is calculated from the recorded trace on the twelve analyzed periods.

Geometrico-Temporal Characteristics:

- *Accordance with the 1/3-power law.* An interesting descriptor of recorded gestures is the coherence with the 1/3 power law. This law is characteristic of biological movements which links the tangential velocity to the curvature of a trajectory, and is defined as: $v_t(s) = KC(s)^{-\beta}$ with β close to 1/3 in adults [10, 17, 19]. K is known as the *velocity gain factor* and mainly corresponds to the average velocity. In other words, this means that when we draw a shape, we accelerate in the more straight parts and slow down in the more curved ones. To evaluate the accordance of the produced gesture with the 1/3 power law, the exponent β is determined by linear regression between the quantities $\log_{10}(v_t)$ and $\log_{10}(C)$. The law can then be written according to the following linear relation: $\log_{10}(v_t) = \log_{10}(K) + \beta \log_{10}(C)$ available where coefficient r^2 and the significance of the correlation also are presented.

- *The mean velocity.* This value is calculated from the total length of the trace and the durations of the recordings.

3.2 Results

Table 2 presents the values of the geometric and geometrico-temporal descriptors averaged across subjects. Qualitative observations of such results enable to make some remarks. Concerning the geometric characteristics, the eccentricities of the drawn shapes seem to be consistent for each shape (circle and ellipse) regardless the different mapping inducing timbre variations in the perceived sounds. The eccentricity averaged across mappings is smaller for the drawn circles ($.62_{\pm .09}$) than for the drawn ellipses ($.90_{\pm .04}$). While the perimeters and the mean velocities of the drawn ellipses are also consistent through the different mappings, the perimeters and mean velocities of the drawn circles differ with the mapping: the lower value, the smaller the drawn shape (and therefore, the lower the mean velocity). Concerning the accordance with the 1/3 power law, the linear regressions for the β estimation are all significant. For the circles, the fitted values of the exponent are lower for low α values and range from .15 to .28. For the ellipse, the fitted values are consistent through the different values range from .34 to .39 and equal in average to $.36_{\pm .05}$.

To summarize, we can conclude that the mapping strategies did not much influence the motor production of ellipses. By contrast, a noticeable influence of the mapping is highlighted for the production of circles. We also found that all the descriptor values are smaller for the circle than for the ellipse. This is not surprising for the eccentricities which are intrinsically lower for drawn circles than for ellipses. However, the observed differences for the perimeters, the mean velocities, and the exponent values of the power law are interesting observations which will be discussed in the following section. An ongoing work is being currently made to assess these qualitative conclusions with statistical analysis.

3.3 Discussion

The goal of the experiment was to evaluate the influence of the mapping on the evoked movement and more precisely, to evaluate the extent to which the manipulation of the mapping coefficient α – informing either on the roughness of the rubbed surface or on the velocity of the evoked movement – modifies the movement produced by synchronization. In the case of drawn ellipses, the investigated descriptors did not differ between mappings, meaning that the subjects were guided by the temporal variations of timbre rather than the intrinsic sound

	α	e	$\bar{P}$(cm)	$\bar{v}$(cm.s^{-1})	$\bar{\beta}_{reg}$	$\bar{r}^2$
Circle	**5**	.61 $\pm$.10	11.36 $\pm$2.65	6.31 $\pm$1.48	.15 $\pm$.05	.21 $\pm$.07
	10	.56 $\pm$.11	12.77 $\pm$2.33	7.10 $\pm$1.30	.17 $\pm$.06	.22 $\pm$.08
	20	.58 $\pm$.10	13.38 $\pm$2.43	7.43 $\pm$1.35	.19 $\pm$.05	.28 $\pm$.11
	50	.64 $\pm$.10	15.63 $\pm$3.32	8.69 $\pm$1.84	.25 $\pm$.08	.37 $\pm$.16
	100	.66 $\pm$.05	16.66 $\pm$3.40	9.26 $\pm$1.89	.26 $\pm$.06	.40 $\pm$.11
	300	.64 $\pm$.07	15.82 $\pm$1.90	8.79 $\pm$1.05	.28 $\pm$.04	.47 $\pm$.09
Ellipse	**5**	.93 $\pm$.02	23.03 $\pm$2.48	12.80 $\pm$1.38	.38 $\pm$.04	.78 $\pm$.06
	10	.91 $\pm$.06	21.83 $\pm$4.11	12.13 $\pm$2.29	.38 $\pm$.06	.77 $\pm$.10
	20	.94 $\pm$.02	22.79 $\pm$2.70	12.67 $\pm$1.50	.39 $\pm$.05	.80 $\pm$.04
	50	.91 $\pm$.06	22.78 $\pm$2.77	12.66 $\pm$1.54	.36 $\pm$.06	.75 $\pm$.11
	100	.85 $\pm$.02	21.85 $\pm$3.41	12.14 $\pm$1.89	.34 $\pm$.04	.67 $\pm$.04
	300	.85 $\pm$.04	23.80 $\pm$4.00	13.23 $\pm$2.23	.34 $\pm$.05	.70 $\pm$.06

Table 2. Results of the experiment according to the geometric shape and the mapping parameters α – Mean and SD. All the correlations for the linear regression of β_{reg} were significant at 95% of confidence for all the subjects and mappings.

texture to synchronize their gestures with the sounds. Nevertheless, in the case of drawn circles, for which friction sounds contained no timbre variations and had a uniform sound texture, the geometric descriptors differed according to the different mappings: the lower the α values, the lower the mean velocities and the shorter the perimeters. This result suggests that when no temporal timbre variations are contained in the sound, subjects use the brightness to adjust their gesture. Finally, these results revealed that in a synchronization task, the temporal variations of timbre contained in the ellipses dominated over the absolute information given by the sound texture. The results for the circles revealed that the different mappings evoke different velocities: the higher the α values, the higher the speed of the gesture. Another interesting point to highlight is the average eccentricity value of .9 observed for the drawn ellipses. This result confirms classical results about preferred coordination patterns observed in different studies [2, 5]. It shows that the mapping and the synchronization task do not influence such motor attractors.

Finally, the geometrico-temporal characteristics of the drawn ellipses were in accordance with the 1/3 power law, which is a classical observation for biological movements. For the drawn circles, the exponent values are lower than 1/3 but not close to 0. This can be explained by the curvature of the drawn circles which are not exactly constant ($e = .62_{\pm.09}$), and therefore variations of velocities are made during the drawing. For the circle, differences on the exponent appear according to the mapping, the lower the value, the lower the exponent. This means that lower values induce a relatively constant movement while higher values may induce more variations in the produced movement.

4 Conclusion and Perspectives

This experiment enables to conclude that the mapping – expressed by the proportionality coefficient α – between the sound parameter and the velocity of a rubbing plectrum modifies the produced synchronized movements only if there are no slower timbre variations as in the case of the ellipses. It enables to conclude that, from a perceptual point of view, the α coefficient is more related to the roughness of the rubbed surface than to the velocity of the gesture. Further analysis should be conducted to completely assess the qualitative observations presented here. In addition, the analysis of asynchrony between the gestures and the acoustical sound parameters might reveal whether there are mappings that provided better synchronizations. An analysis by subject should also enable to evaluate whether intrinsic personal strategies have been used.

This preliminary study tackles the problem of the sensorimotor relation between an auditory input and a produced movement which has never been done before in such a formal approach. The obtained results will also help us to choose the adequate mapping strategy for friction sound synthesis model in further experiments. Such a framework should enables to evaluate the influence of the sound on the produced gesture in a multimodal context. Potential multisensory conflicts are currently being evaluated in different experimental conditions.

5 Acknowledgments

This work was funded by the French National Research Agency (ANR) under the MetaSon: Métaphores Sonores (Sound Metaphors) project (ANR-10-CORD-0003) in the CONTINT 2010 framework.

References

1. Aramaki, M., Gondre, C., Kronland-Martinet, R., Voinier, T., & Ystad, S.: Thinking the sounds: an intuitive control of an impact sound synthesizer. *Proceedings of ICAD 09 – 15th International Conference on Auditory Display* (2009)
2. Athenes, S., Sallagoïty, I., Zanone, P.G., & Albaret, J.M.: Evaluating the coordination dynamics of handwriting, *Human movement science*, 23(5),621–641(2004)
3. Caramiaux, B., Susini, P., Bianco, T., Bevilacqua, F., Houix, O., Schnell, N., & Misdariis, N.: Gestural embodiment of environmental sounds: an experimental study. *Proceedings of the International Conference on New Interfaces for Musical Expression. (NIME 2011) A.R. Jensenius et al., (Eds)* (2011)
4. Caramiaux, B., Bevilacqua, F., & Schnell, N.: Towards a gesture-sound cross-modal analysis. *Gesture in Embodied Communication and Human-Computer Interaction, Springer Heidelberg, Berlin*, 158–170 (2010)
5. Danna, J., Athènes, S., &, Zanone, P. G.: Coordination dynamics of elliptic shape drawing: Effects of orientation and eccentricity. *Human movement science*, 30(4), 698–710 (2011)
6. Gaver, W. W.: What in the world do we hear?: an ecological approach to auditory event perception. *Ecological psychology*, 5(1) 1–29 (1993)

7. Gaver, W. W.: How do we hear in the world? Explorations in ecological acoustics. *Journal of Ecological Psychology*, 5(4), 285–313 (1993)
8. Godøy, R.I., Leman, M.: Musical gestures: Sound, movement, and meaning. Taylor & Francis (2010)
9. Goldstein, H.: Classical mechanics (1962)
10. Lacquaniti, F., Terzuolo, C. A., & Viviani, P.: The law relating kinematic and figural aspects of drawing movements, *Acta Psychologica*, 54, 115–130 (1983)
11. Leman, M.: Embodied music cognition and mediation technology, Mit Press (2007)
12. Rasamimanana, N.: Geste instrumental du violoniste en situation de jeu: analyse et modélisation, *Ph.D Thesis – Université Pierre et Marie Curie – Paris VI* (2008)
13. Stucchi, N., Viviani, P.: Cerebral dominance and asynchrony between bimanual two-dimensional movements.,*Journal of Experimental Psychology: Human Perception and Performance*, 19(6) (1993)
14. Theurel, A., Frileux, S., Hatwell, Y., & Gentaz, E.: The haptic recognition of geometrical shapes in congenitally blind and blindfolded adolescents: is there a haptic prototype effect?,*PloS one*,7(6),e40251 (2012)
15. Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J. L., & Ystad, S.: Reenacting sensorimotor features of drawing movements from friction sounds. *In Aramaki, M., Barthet, M., Kronland-Martinet, R.,& Ystad, S. (Eds.): CMMR 2013. LNCS volume number 7900, Springer Heidelberg, Berlin* (2013)
16. Van Den Doel, K., Kry, P.G. & Pai, D.K.: 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)
17. Viviani, P., & Terzuolo, C.: Trajectory determines movement dynamics. *Neuroscience*, 7, 2, 431–437 (1982)
18. Viviani, P., Campadelli, P. & Mounoud, Pierre.: Visuo-manual pursuit tracking of human two-dimensional movements. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 62 (1987)
19. Viviani, P., & Schneider, R.: A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 198–218 (1991)
20. Viviani, P., Baud-Bovy, G., & Redolfi, M.: Perceiving and tracking kinesthetic stimuli: Further evidence of motor–perceptual interactions, *Journal of Experimental Psychology: Human Perception and Performance*, 23(4), 1232–1252 (1997)
21. Young, W., Rodger, M., & Craig, C. Perceiving and re-enacting spatio-temporal characteristics of walking sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 464–76 (2013)