TOWARD A NEW MODEL FOR SOUND CONTROL

Roberto Bresin, Anders Friberg, Sofia Dahl

Department of Speech, Music and Hearing
Royal Institute of Technology, KTH
{roberto.bresin, anders.friberg, sofia.dahl}@speech.kth.se

ABSTRACT

The control of sound synthesis is a well-known problem. This is particularly true if the sounds are generated with physical modeling techniques that typically need specification of numerous control parameters. In the present work outcomes from studies on automatic music performance are used for tackling this problem.

1. INTRODUCTION

In the recent past, sound synthesis techniques have achieved remarkable results in reproducing real sounds, like those of musical instruments. Unfortunately most of these techniques focus only on the "perfect" synthesis of isolated sounds. For example, the concatenation of these synthesized sounds in computer-controlled expressive performances often leads to unpleasant effects and artifacts. In order to overcome these problems Dannenberg and Derenyi [1] proposed a performance system that generates functions for the control of instruments, which is based on spectral interpolation synthesis.

Sound control can be more straightforward if sounds are generated with physics-based techniques that give access to control parameters directly connected to sound source characteristics, as size, elasticity, mass and shape. In this way sound models that respond to physical gestures can be developed. This is exactly what happens in music performance when the player acts with her/his body on the mechanics of the instrument, thus changing its acoustical behavior. It is therefore interesting to look at music performance research in order to identify the relevant gestures in sound control. In the following paragraphs outcomes from studies on automatic music performance are considered.

2. CONTROL MODELS

The relations between music performance and body motion have been investigated in the recent past. Musicians use their body in a variety of ways to produce sound. Pianists use shoulders, arms, hands, and feet; trumpet players make great use of their lungs and lips; singers put into actions their glottis, breathing system, phonatory system and use expressive body postures to render their interpretation. Dahl [2] recently studied movements and timing of percussionists when playing an interleaved accent in drumming. The movement analysis showed that drummers prepared for the accented stroke by raising the drumstick up to a greater height, thus arriving at the striking point with greater velocity. In another study drummers showed a tendency for privileging auditory to tactile feedback [3]. These and other observations of percussionists' behavior are under further investigation for the modeling of a control model for physics-based sound models of percussion instruments. This control model could be extended to the control of impact sound models where the human action is used to manipulate the sound source.

The research on music performance at KTH, conducted over a period of almost three decades, has resulted in about thirty so-called performance rules. These rules, implemented in a program called Director Musices [4], allow reproduction and simulation of different aspects of the expressive rendering of a music score. It has been demonstrated that rules can be combined and set up in such a way that emotionally different renderings of the same piece of music can be obtained [5]. The results from experiments with emotion rendering showed that in music performance, emotional coloring corresponds to an enhancement of the musical structure. A parallel can be drawn with hyper- and hypo-articulation in speech; the quality and quantity of vowels and consonants vary with the speaker's emotional state or the intended emotional communication [6]. Yet, the structure of phrases and the meaning of the speech remain unchanged. In particular, the rendering of emotions in both music and speech can be achieved, and recognized, by controlling only a few acoustic cues [5][7]. This is done in a stereotype and/or cartoonized way that finds its visual correspondent in email emoticons. Therefore cartoon sounds can be produced not only by simplifying physics-based models, but also by controlling their parameters in appropriate ways.

3. WALKING AND RUNNING

As a first sound control case we chose that of locomotion sounds. In particular we considered walking and running sounds. In a previous study Li and coworkers [8] demonstrated the human ability to perceive source characteristics of a natural auditory event. They ask subjects to classify the gender of a human walker. Subjects could correctly classify walking sounds as being produced by men or women. Subjects showed also ability in identifying the gender in walking sequences even with both male and female walkers wearing the same male's shoes.

3.1. Sounds

In their study Li and coworkers considered walking sounds on a hardwood stage in a art theater. From various analyses applied on the walking sounds, they found a relationship between auditory

events and acoustic structure. In particular male walking sounds were characterized by "(1) low spectral mean and mode (2) high values of skewness, kurtosis, and low-frequency slope, and (3) low to small high-frequency energy". On the other hand female walking sounds were characterized by (1) high spectral mean and mode, and significant high spectral energy".

In the present study we considered sounds of walking and running footstep sequences produced by a male subject running on gravel. The choice was motivated by the assumption that (1) an isolated footstep sound produced by a walker on a hard surface would be perceived as unnatural, i.e. mechanical, (2) the sound of an isolated footstep on gravel would still sound natural because of its more noisy and rich spectrum (Figures 1, 2, 3 and 4).

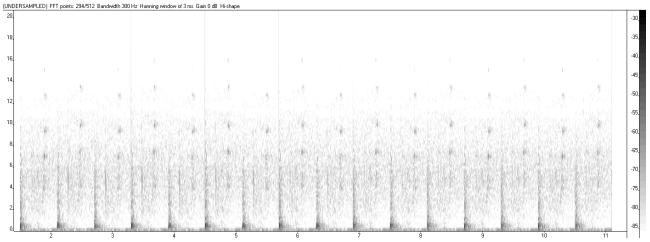

Figure 1. *Spectrogram of a walking sound on concrete floor (16 footsteps).*

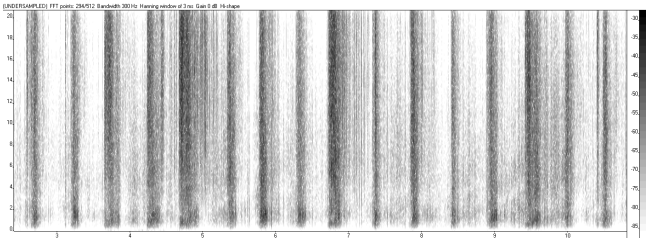

Figure 2. *Spectrogram of walking sound on gravel (16 footsteps).*

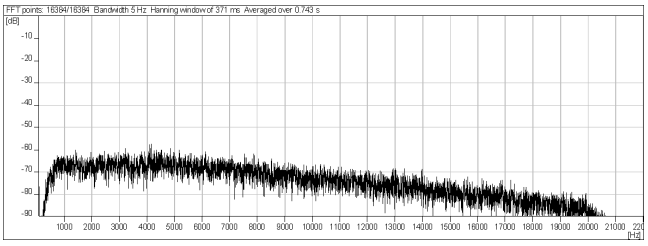

Figure 3. *Long time average spectrogrum (LTAS) of a walking sound on gravel (40 footsteps).*

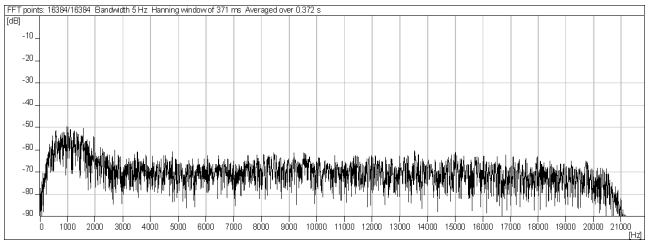

Figure 4. *Long time average spectrum (LTAS) of a running sound on gravel (60 footsteps).*

3.2. Timing

When walking, a double support phase is created when both feet are on the ground at the same time, thus there is a step overlap time. This is shown also in the spectrogram of three footsteps of Figure 5; there is no silent interval between to adjacent steps. This phenomenon is similar to *legato* articulation. Figure 6 plots the key-overlap time (KOT) and the double support phase duration (T_{dsu}) as a function of the inter-onset interval (IOI) and of half of the stride cycle duration ($T_c/2$), respectively. The great inter-subject variation in both walking and *legato* playing, along with biomechanical differences, made quantitative matching impossible. Nevertheless, the tendency to overlap is clearly common to piano playing and walking. Also common is the increase of the overlap with increasing IOI and increasing ($T_c/2$), respectively.

Both jumping and running contain a flight phase, during which neither foot has contact with the ground. This has also a visual representation in the spectrogram of three footsteps of Figure 7; there is a clear silent interval between two adjacent steps. This is somewhat similar to *staccato* articulation in piano performance. In Figure 8 the flight time (T_{air}), and key-detach time (KDT) are plotted as a function of half of stride cycle duration (Tc/2) and of IOI. The plots for T_{air} correspond to typical step frequency in running. The plots for KDT represent *mezzostaccato* and *staccato* performed with different expressive intentions as reported by Bresin and Battel [9]. The similarities suggest that it would be worthwhile to explore the perception of *legato* and *staccato* in formal listening experiments

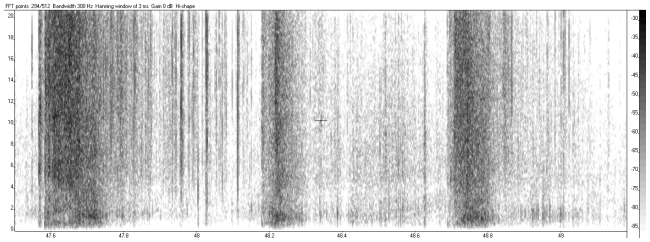

Figure 5. *Spectrogram of three steps extracted from the walking sound on gravel of Figure 1.*

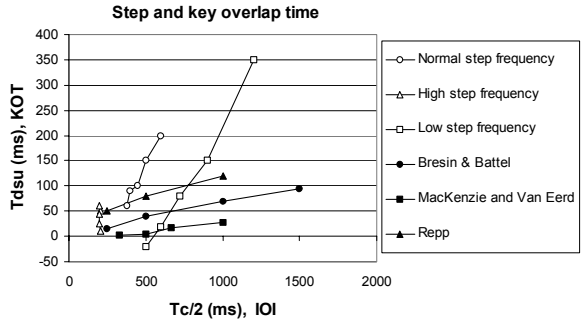

Figure 6. *The double support phase (T_{dsu}, filled symbols) and the key overlap time (KOT, open symbols) plotted as function of half of stride cycle duration (Tc/2) and of IOI. The plots for T_{dsu} correspond to walking at step frequency as reported by Nilsson and Thorstensson [10][11]. The KOT curves are the same as in Figure 1, reproducing data reported by Repp [12], Bresin and Battel[9], MacKenzie and Van Eerd [13].*

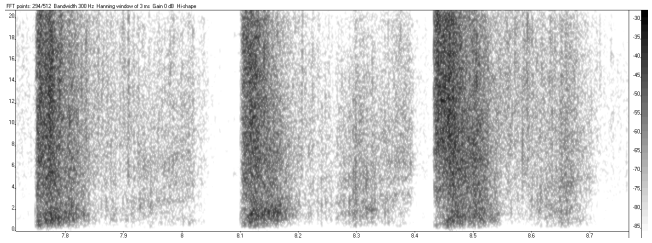

Figure 7. *Spectrogram of three steps extracted from the running sound on gravel of Figure 2.*

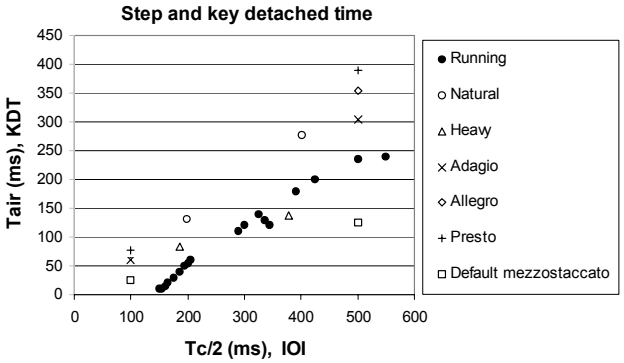

Figure 8. *The time when both feet are in the air (T_{air}, filled symbols) and the key detached time (KDT, open symbols) plotted as function of half of stride cycle duration (Tc/2) and of IOI. The plots for T_{air} correspond to normal frequency steps in running [10][11]. The KDT for mezzostaccato (KDR = 25%) is defined in the Oxford Concise Dictionary of Music [14]. The values for the other KDTs are reported in works by Bresin and Battel [9] and Bresin and Widmer [15].*

4. CONTROLLING THE SOUNDS OF WALKING AND RUNNING

Among the performance rules developed at KTH there are rules acting on a short timescale (micro-level rules), and rules acting on a long timescale (macro-level rules) [16]. Examples of the first class of rules include the "Score Legato Articulation" rule, which realizes the acoustical overlap between adjacent notes marked *legato* in the score, and the "Score Staccato Articulation" rule, which renders notes marked *staccato* in the score [17]. A macro-level rule is the "Final Ritard" rule that realizes the final *ritardando* typical in Baroque music [18]. Relations between these three rules and body motion have been found. In particular Friberg and Sundberg demonstrated how their model of final *ritardando* was derived from measurements of stopping runners, and in the previous paragraph we pointed out analogies in timing between walking and *legato*, running and *staccato*. In both cases human locomotion is related to time control in music performance.

Friberg et al. [19] recently investigated the common association of music with motion in a direct way. Measurements of the ground reaction force by the foot during different gaits were transferred to sound by using the vertical force curve as sound level envelopes for tones played at different tempi. The results from the three listening tests were consistent and indicated that each tone (corresponding to a particular gait) could clearly be categorized in terms of motion.

These analogies between locomotion and music performance open to a challenging field for the design of new control models for artificial walking sound patterns, and in general for sound control models based on locomotion. In particular a model for humanized walking and one for stopping runners were implemented in two pd patches. Both patches control the timing of the sound of one step on gravel.

The control model for humanized walking was used for controlling the timing of the sound of one step of a person walking on gravel. As for the automatic expressive performance of a music score, two performance rules were used to control the timing of a sequence of steps. The two rules were the "Score Legato Articulation" rule and the "Phrase Arch" rule. The first rule, as mentioned above, presents similarities with walking. The "Phrase Arch" rule is used in music performance for the rendering of *accelerandi* and *rallentandi*. This rule is modeled according to velocity changes in hand movements between two fixed points on a plane [20]. When it was used in listening tests of automatic music performances, the time changes caused by applying this rule were classified as "resembling human gestures" [7]. The "Phrase Arch" rule was then considered to be interesting for use in controlling tempo changes in walking patterns and combined with the "Score Legato Articulation" rule. In figure 9 the tempo curve and the overlap time curve for walking sounds, produced by the model, is presented.

The control model for stopping runners was implemented with a direct application of the "Final Ritard" rule to the control of tempo changes on the sound of running on gravel.

In the following we describe a pilot experiment where the control models presented here are tested for the first time.

5. PILOT EXPERIMENT: LISTENING TO WALKING AND RUNNING SOUNDS

A listening test comparing step sound sequences without control, and sequences rendered by the control models presented here, was conducted. We wanted to find out if (1) listeners could discriminate between walking and running sounds in general and (2) if they were able to correctly classify the different types of motion produced by the control models.

5.1. Stimuli

Eight sound stimuli were used. They were 4 walking sounds and 4 running sounds.

The walking sounds were the following; (1) a sequence of footsteps of a man walking on gravel indicated with W_SEQ in the following, (2) 1 footstep sound extracted from stimuli (1) W_1STEP, (3) a sequence of footsteps obtained by looping the same footstep sound W_NOM, (4) a sequence of footsteps obtained applying the "Phrase arch" and the "Score legato articulation" rules W_LP.

The running sounds were; (5) a sequence of footsteps of a man running on gravel R_SEQ; (6) 1 footstep sound extracted from stimuli (5) R_1STEP, (3) a sequence of footsteps obtained by looping the same footstep sound R_PD, obtained with a pd patch, (4) a sequence of footsteps obtained applying the "Final Ritard" rule R_RIT.

5.2. Subjects and procedure

The subjects were four, 2 females and 2 males. Their average age was 28. The subjects all worked at the Speech Music Hearing Department of KTH, Stockholm.

Subjects listened to the examples individually over Sennheiser HD433 adjusted to a comfortable level. Each subject was instructed to identify for each example (1) if it was a walking or running sound and (2) if the sound was human or mechanical. The responses were automatically recorded by means of the Visor software system, specially designed for listening tests [21]. Listeners could listen as many times as needed to the each sound stimuli.

5.3. Results and discussion

We conducted a simple preliminary statistical analysis of the results. In Figure 10 average subjects' choices are plotted. The Y axe represents the scale from Walking, value 0, to Running, value 1000, with 500 corresponding to a neutral choice. It emerges that all walking and running sounds were correctly classified as walking and running respectively. This including the footstep sequences generated with the control models proposed above. This means that listeners have in average no problem to recognize this kind of stimuli. It is however interesting to notice how the stimuli corresponding to a single footstep were classified with less precision than the other sounds in the same class (walking or running). In particular one of the subjects classified the stimulus W_1STEP as a running sound.

The R_1STEP was classified as mechanical, although it is a real sound. This could be the reason why the sequences of footstep sounds produced by the control models were all classified as mechanical, since these sequences loop the same footstep.

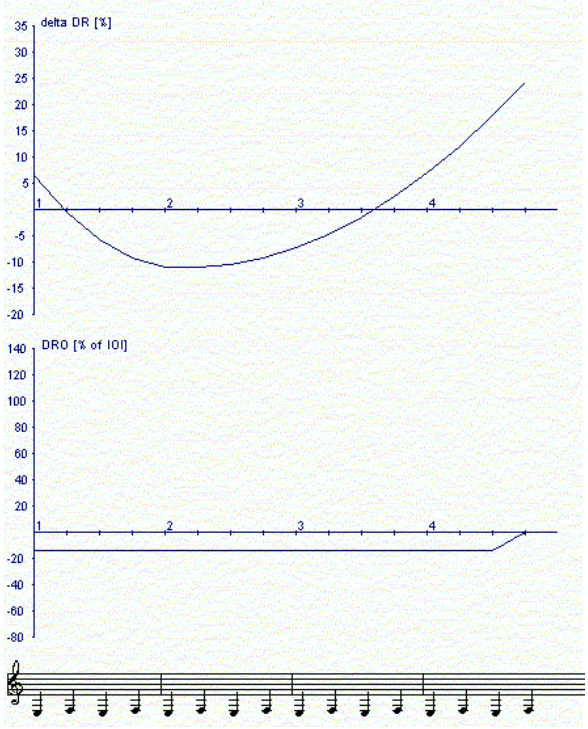

Figure 9. *Tempo curve and overlap time curve used for controlling a walking sound.*

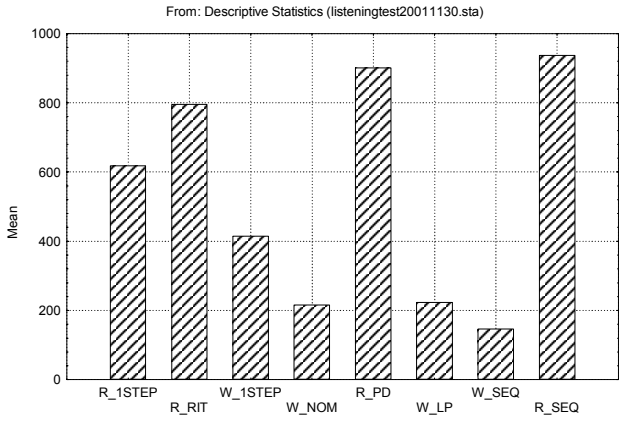

Figure 10. *Mean classification values, with pooled subjects, for the scale Running (1000) - Walking (0)*

6. CONCLUSIONS

In this paper we proposed a new model for controlling physics-based sound models. This control model is rule-based and it is derived from analogies between music performance and body motion that have been observed in previous works of the KTH research group. We used walking and running sounds that in the future will be substituted by physics-based models of sounds, such as a model of scratching sounds.

Even though this is still a preliminary work, a listening test was conducted. The main result was that subjects correctly classified different types of motion produced by the models.

The proposed rule-based approach for sound control is the first step toward the design of more general control models that respond to physical gestures. At present, studies on the expressive gestures of percussionists and disk jockeys are under development. New rules could be designed to develop control models. These models would produce a natural expressive variation of the control parameters in accordance with the dynamics of the gestures.

7. LINKS

The Sounding Object project home page:
http://www.soundobject.org

The art of music performance:
http://www.speech.kth.se/music/performance

Sound control models and example sounds:
http://www.speech.kth.se/music/performance/controlmodels

Articulation rules and example sounds:
http://www.speech.kth.se/music/performance/articulation

The Director Musices program:
http://www.speech.kth.se/music/performance/download/

8. ACKNOWLEDGMENTS

This work was supported by the European Union (SOb - The Sounding Object project – no. IST-2000-25287; http://www.soundobject.org), and by the Bank of Sweden Tercentenary Foundation (FEEL-ME - Feedback Learning in Musical Expression - Dnr 2000 - 5193:01; http://www.psyk.uu.se/forskning/projekt.html#pj).

9. REFERENCES

[1] Dannenberg, R. and Derenyi, I. (1998). Combining Instrument and Performance Models for high-quality Music Synthesis. *Journal of New Music Research,* 27(3):211-238

[2] Dahl, S. “The Playing of an Accent. Preliminary Observations from Temporal and Kinematic Analysis of Percussionists”. *Journal of New Music Research*, Vol. 29, No 3, pp. 225-234, 2000

[3] Dahl, S., Bresin, R. “Is the player more influenced by the auditory than the tactile feedback from the instrument?”, In *Proceedings of DAFx01*

[4] Friberg, A, Colombo, V, Frydén, L and Sundberg, J “Generating Musical Performances with Director Musices”, *Computer Music Journal*, vol. 24, no. 3, pp. 23-29, 2000

[5] Bresin, R and Friberg, A (2000) “Emotional coloring of computer controlled music performance”, *Computer Music Journal*, Vol. 24, No. 4, 44-62

[6] Lindblom, B. “Explaining phonetic variation: a sketch of the H&H theory”, In Hardcastle & Marchal (Eds.) *Speech production and speech modeling*, Dordrecht: Kluwer, pp. 403-439, 1990

[7] Juslin, N.P., Friberg, A., and Bresin, R. “Toward a computational model of expression in music performance. The GERM model”, *Musicae Scientiae.* in press, 2001

[8] Li, X., Logan, R.J., and Pastore, R.E. “Perception of acoustic source charachteristics: Walking sounds”, *J. Acoust. Soc. Amer.*, vol. 90, no. 6, pp 3036-3049, 1991

[9] Bresin, R. and Battel, G.U. “Articulation strategies in expressive piano performance. Analysis of legato, staccato, and repeated notes in performances of the Andante movement of Mozart's sonata in G major (K 545)”. *Journal of New Music Research*, vol. 29, no 3, pp. 211-224, 2000

[10] Nilsson, J., & Thorstensson, A. “Adaptability in frequency and amplitude of leg movements during human locomotion at different speeds”, *Acta Physiol Scand*, 129, 107-114, 1987

[11] Nilsson, J., & Thorstensson, A. “Ground reaction forces at different speeds of human walking and running”, *Acta Physiol Scand*, 136, 217-227, 1989

[12] Repp, B. “Acoustics, perception, and production of legato articulation on a computer-controlled grand piano”, *J. Acoust. Soc. Amer.*, vol. 102, no. 3), pp. 1878-1890, 1997

[13] MacKenzie, C.L., & Van Erde, D.L. “Rhythmic Precision in the Performance of Piano Scales: Motor Psychophysics and Motor Programming"”, In M. Jeannerod (Ed*.) Proceedings of the Thirteenth international Symposium on Attention and Performance*, Hillsdale: Lawrence Erlbaum Associates, Inc., Publishers, pp. 375-408, 1990

[14] Kennedy, M. *Oxford Concise Dictionary of Music*. Oxford: Oxford University Press, 1996

[15] Bresin, R., and Widmer, G. “Production of staccato articulation in Mozart sonatas played on a grand piano. Preliminary results”, *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, Vol. 2000/4, pp. 1-6, 2000

[16] Friberg, A. “Generative Rules for Music Performance: A Formal Description of a Rule System”, *Computer Music Journal*, vol. 15, no. 2, pp. 56-71, 1991

[17] Bresin, R. “Articulation Rules For Automatic Music Performance”, In *Proceedings of the International Computer Music Conference* 2001

[18] Friberg, A. and Sundberg, J. “Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners”, *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp 1469-1484, 1999

[19] Friberg, A, Sundberg, J and Frydén, L “Music from Motion: Sound Level Envelopes of Tones Expressing Human Locomotion”, *Journal of New Music Research*, vol. 29, no 3, pp. 199-210, 2000

[20] Flash, T. and Hogan, N. “The coordination of arm movements: an experimentally confirmed mathematical model”, *The Journal of Neuroscience*, vol. 5, no. 7, pp. 1688-1703, 1985

[21] Granqvist, S. “Enhancements to the Visual Analogue Scale, VAS, for listening tests”, *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, 1996, no. 4, pp. 61-65, 1996