MIMICRY OF TONE PRODUCTION: RESULTS FROM A PILOT EXPERIMENT

Tommaso Bianco
IRCAM, CNRS - UMR STMS, Paris
tommaso.bianco@ircam.fr

Marcelo M. Wanderley
Idmil, McGill University, Montreal
marcelo.wanderley@mcgill.ca

Frederic Bevilacqua
IRCAM, CNRS - UMR STMS, Paris
frederic.bevilacqua@ircam.fr

ABSTRACT

In this paper we present the description and the first results of a pilot experiment in which participants were requested to mimic the production of sonic elements trough different control modalities. Results show different degrees of dependence of the control temporal profiles with the dynamic level and temporal ordering of the stimuli. The protocol and methodology here advanced may turn useful for ameliorating existing mapping strategies for gesture based interactive media, with particular emphasis to adaptive control of physics-based models for sound synthesis.

1. INTRODUCTION

The decoupling between control input and acoustical output in computer music instruments opened a problematic that still remains central for research in the musical domain: the combination between ease of use and effectiveness, two fundamental components of system's usability [1]. If we consider the case of control for virtual acoustical instruments based on physical modeling, this combination can become even more baffling. The user may indeed be compelled to control an instrument without the interaction feedback, and with a physical interface that bases on a control modality far different from the real case.

The connection between user's intention and sonic outcome has been previously tackled at different levels: at a physical and physiological level for the choice of the machine transducers [2] [3], at gestural level for the definition of tasks and evaluation of performance [4] or evaluation of similarities [5], and at a cognitive level, for the understanding of mental coding of musical experience though motor-mimetic imagery [6] [7]. If aiming at developing intelligent machines for music production, it is evident that these layers must be considered in conjunction. To achieve maximal effectiveness, the mapping model should therefore adapt to the idiosyncrasies of the physical interface and personal abilities of the user. In a similar situation, the natural skills developed through everyday activity could guarantee the user an initial level of expertise, even to people who usually don't have the opportunity to make music [8].

In this direction we conducted a pilot experiment where the user had to mimic the production of simple musical elements. A similar concept appeared in previous literature, under the terms of motor-mimetic sketching, sound-tracing [6], sound "gestureification" [9].

Analogously to those experiments, here as well subjects were demanded to transfer a mental imagery of a perceived acoustical event onto human movement. However, the concurrency of the following aspects tells apart the present account from previous experiments:

- the user's intentions were to be communicated trough a specific task and *control modality* [1]
- the mimicry activity was performed trough a control modality different from the one that produced the sound stimulus (as happens on the contrary for air playing activity [6])
- the user had to rely only on primary feedback [2], i.e. visual, tactile and proprioceptive cues
- the musical stimulus was limited to simple tones, and the user was explicitly asked to address only to sound intensity; this reduction was motivated by the attempt to limit influences of cognitive and cultural aspects raised by melodic and rhythmic developments.
- stimuli perception and control action were tasks sequentially separated

2. METHOD

This experience is intended to explore gestural control for sound production. In particular, we focus on position and velocity control for different sound dynamics, and on gesture coarticulation, a term that defines the "process whereby the properties of a segment are altered due to the influences exerted on it by neighboring segments" [11], referred to with the term *articulation* in the musical domain [12]. In outline, a group of participants was asked to listen to trumpet tones and to mimic their production acting with a device. In the following we describe in detail the components and protocol of the experiment.

[1] for *control modality* we refer to the modality of physical interaction that allows the user to communicate her intentions through a device as in [10]

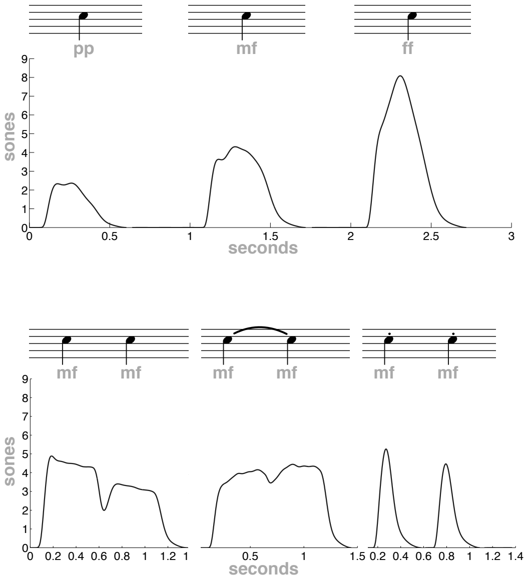

Figure 1. Score notation and loudness profiles computed through Zwicker method [13], for isolated (up) and consecutive (down) notes

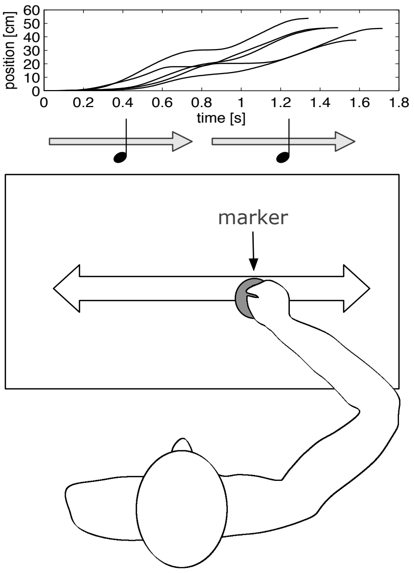

Figure 2. Sketch of the material and configuration setup for the experiment. At top, the displacement over time of the marker, for the mimicry of production of two consecutive tones trough control in velocity

2.1 Stimuli

The acoustic stimuli presented to the subjects consisted of a set of tones produced by a real trumpet performer and recorded during a previous experiment [14]. The audio was presented to the subjects through a headphones set, and the rate of audio amplification was adjusted for each subject in order to assure a neat but comfortable listening. Score notation and loudness profiles of tones are presented in Fig. 1 Isolated notes were presented in triplets of increasing or decreasing loudness, forming a listening stream of ca 2.5s, whereas couple of articulated notes were presented singularly for each condition, with a maximal duration of ca 1.2s.

2.2 Participants

Five subjects took part to the experiment: four males and one female, aged between 25 and 30 years. All subjects were student members or collaborators of the Idmil laboratory, thus involved in research in the musical domain. Part of them were also trained musicians in piano, trumpet and violin performance.

2.3 Material and apparatus

The subjects were seated on a chair in front of a table, and were asked to move a marker leaning on the table by performing planar movement along a line. A sketch of the task setup is shown in Fig. 2 The marker was a sensor of the Polhemus Liberty interface, which tracked its position in time at 120 Hz by means of a custom made driver software developed at Idmil laboratory. The marker was cloth-covered in order to reduce the mechanical resistance of friction of the contact with the table surface. The participants were left free to position their arm and body with respect to the material setup, with the only precaution of finding the most comfortable solution in order to avoid obstruction or limitation of the movement during the performance of the tasks.

2.4 Protocol

The protocol consisted of two stages, each of which involved a different control modality. During the first stage, participants were presented with the series of notes and were asked to mimic the production of the sounds in loudness by relating instantaneous position of the marker with instantaneous value of loudness. Before performing the trials, they were instructed to choose along the line on the table two extreme positions distant 50cm, representatives of zero loudness and of maximal loudness perceived. After listening to each stimulus, subjects had to move the marker from the zero loudness point, across the loudness scale, and back to the origin point, so as to reproduce the loudness profile of the sounds. In the second stage, subjects were asked to relate loudness of sounds with instantaneous velocity of the marker. In this configuration, still position of the marker represented the silence, while an arbitrary maximal linear velocity the maximal loudness perceived. At the beginning of each experiment, the testers were let to familiarize with the task in order to determine the most comfortable movement amplitudes and speeds. Participants were explicitly told to reproduce the loudness profile as accurately as possible, bot in amplitude and temporal variations, and to discard all other auditory attributes. When presented with couples of joined notes, control though velocity was explicitly asked to be performed on the same movement direction, whereas direction was left arbitrary for the tasks with isolated notes.

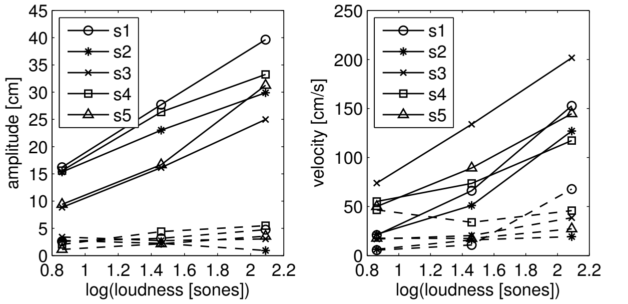

Figure 3. Semi-logarithmic plot of ratio between loudness and maxima in position profiles (left) and velocity profiles (right): mean (solid lines) and standard deviations (dashed line)

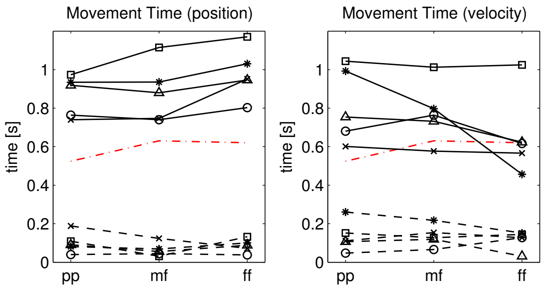

Figure 4. Temporal duration of stimuli and participants' control profiles: : mean (solid lines) and standard deviations (dashed line)

2.5 Analysis

Loudness profiles of the acoustic stimuli were computed with Zwicker method, which makes use of several psychoacoustical principles in order to give an estimate of the average person's impression of the sound intensity for temporally variable sounds [13]. The profiles computed are displayed in Fig. 1. It is worth pointing out that the values of loudness in sones are dependent on an estimation of the real sound pressure level presented to the subjects. These were indeed supposed to have been exposed to an average sound pressure level of 69 db, which lies in the decibel scale as the average comfortable volume for a quiet laboratory setting. The role of possible error in the normalization was investigated and we found that even an error of 20dB would have caused only minor relative difference in our present study.

Each movement trial was manually trimmed from the recordings and examined for amplitude and temporal variations. In a second step, each segment has been resized and rendered in a functional form by means of the FDA Matlab toolbox. Functional conversion with smoothing penalty allowed to represent each movement profile as a continuous and derivable function, and to control smoothness degree of higher order derivatives (which presented irregular spikes due to errors in the sensing and streaming of the capturing system). Control with the generalized cross-validation criterion assured a good compromise between the smoothing effect and the fit to the original curves [15]. In order to have more representative curves for each condition, trial records for the same condition (up to 5) were subsequently aligned by means of landmarks registration. Roughly, this process allows to align features by estimating a strictly increasing nonlinear transformation of time that takes all the times of a given feature into a common value [15].

Indicative measures of shape similarity between normalized curves have been computed by means of the mean square error method described in Ch.8.5 of [15]. This method assured the consideration of an eventual temporal deformation introduced by the registration process. These measures have been computed for the single isolated notes, the single notes embedded in sequence of two notes, and for the entire sequence comprising two notes.

3. RESULTS

The subjects revealed a systematic behavior in the duration and maximal amplitude of control profiles. Fig. 3 displays the maximum values of stimuli loudness versus maximum values of control profiles for the three dynamics conditions. The plot evidences the increase in movement amplitude and peak velocity with loudness, as explicitly requested in the task. For both control modalities, amplitudes generally scale linearly with the logarithm of loudness in sones (a measure unit directly proportional to loudness). The light deviation from a linear trend in the semilog plot for some subjects correlates with singular values in standard deviation or in movement durations. In the case of control by position, higher amplitude for subject 5 in the louder note goes with a greater duration if supposing that he maintained the same amount in velocity. In the case of control by velocity, subjects exhibiting light divergence from linear trend present higher standard deviation or erroneous duration estimation (Fig. 4).

Standard deviation revealed higher for control through velocity, indicating a possible higher difficulty in delivering consistency among trials.

The mean durations of movements together with the ground truth of stimuli events durations for each condition is shown in Fig. 4. Globally the participants tended to overestimate the duration of the event. Comparison of the two plots reveals that this overestimation is dominant for position control. For louder tones, the two modalities reveal a divergent behavior: an increase in duration for control though position and a decrease for control though velocity.

Temporal profiles of stimuli loudness, position, and velocity control for a representative participant are reported in Fig. 5. In plots on second and third columns, the three superposed curves resume the mean of five registered trials for each dynamic condition. Position curves and derivatives conform previous results of human point to point reaching movements, characterized by amplitude invariant symmetric bell-shaped velocity [16]. On the contrary, amplitude invariance does not behold for velocity profiles, whose shape varies to a greater extent between cases. All participants except one manifested a behavior similar to that in figure, transiting from triangular or trapezoidal shape into a bell shape when movement peak velocity exceeds 60cm/s. Subject 3, who was the only participant without

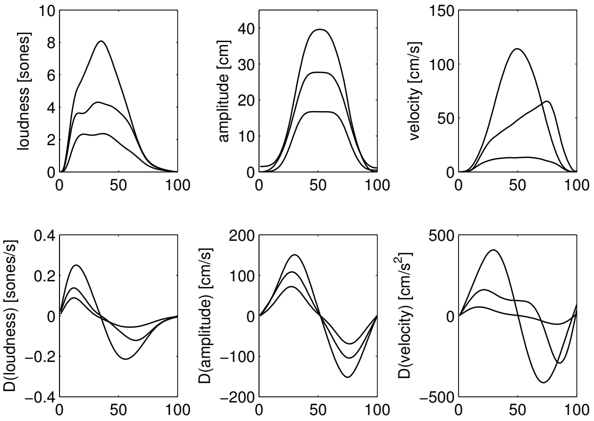

Figure 5. Extract of control profiles for one subject in isolated tones task, for *pp*, *mf*, and *ff* dynamics

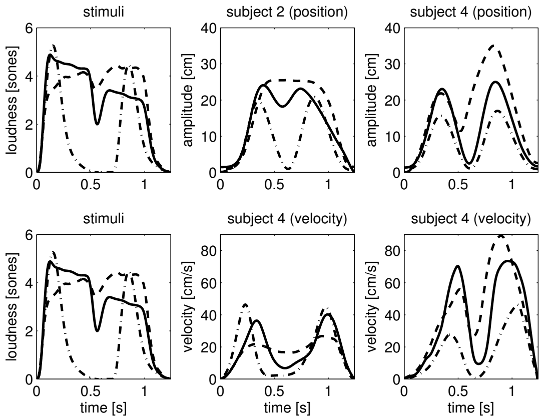

Figure 6. Extract of control profiles for two subjects in consecutive tones task for *non legato* (solid), *legato* (dashed), and *staccato* (dash-dot) articulation. The up row reports profiles in control trough position, the down row profiles in control trough velocity. In first column stimuli loudness, equal for position and velocity tasks

professional musical training, singularly presented similar profiles at all movement amplitudes.

Control profiles for coarticulated notes for two subjects are displayed in Fig 6. The curves reveal distinct method between subjects for the performance of the task, yet preserving the relation between conditions visible in the stimuli profiles. Articulatory degree between the tones reflects in the transient parts between the two strokes, with the level of descend correctly marking the distinction between the three coarticulation strategies. Subject 2 manifested the extreme behavior for position control, performing no descend for the *legato* condition. Subject 4, who was musically trained in violin performance, showed the closest match in profile shapes between the two control modalities.

Indicative values of similarity between curves for each condition are given in Fig. 7, in which higher values represent higher discrepancy between stimuli profile and gesture profile. Results show that control trough position in general performs better then control with velocity in simulating the profile of the stimulus. Moreover, performance required in the task decreases for consecutive notes, both when considering the entire couple segment and when the single embedded note (segmented at minima).

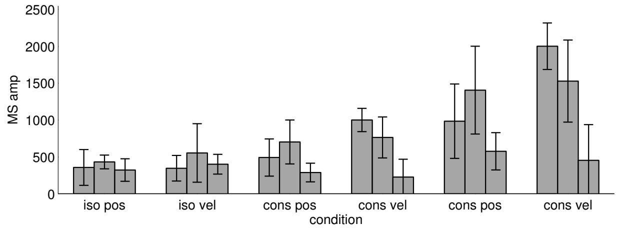

Figure 7. Values of joint amplitude and phase variability as computed by formulas 8.3 and 8.5 in [15]. Higher values represent higher discrepancy between stimulus profile and gesture profile. Values are computed for single note in isolated *pp*, *mf*, and *ff* condition, single note (average) embedded in consecutive *nl*, *l*, and *s* condition, and on entire record for consecutive *nl*, *l*, and *s* condition

4. DISCUSSION

The goal of the present study was to investigate the role of musical dynamics and articulation on human motor control for the mimicry of sound production. Two control modalities were examined, position and velocity.

Both modalities revealed to overestimate the duration of the stimuli, and to scale almost linearly with the logarithm of the loudness perceived. We advance the hypothesis that this scaling could be the outcome of two causes. On a performance level, faster and wider movements could have been automatically reduced by the participant because of arm biomechanical limitations. On a cognitive level, sound loudness could have been mentally associated to movement kinematics - as explicitly demanded in the task - in conjunction with movement effort. Effort on joint torques and muscular activation, which has been shown to scale in amplitude and duration with the movement speed [17] [18] [19], may have participated in reducing the performance of the tasks for wider/faster movements.

The two modalities deviated to some extent in variability. If we consider variability in-between trials as an indicator of difficulty, Fig. 3 and Fig. 4 suggest that control trough velocity could be a harder task than control by position. A substantial reliance on visual over kinesthetic feedback for position control could be the cause for a more consistence performance. Support of this hypothesis is given by the fact that participants with developed acuteness for arm velocity profiles (typical of violin players, alike subject 4) delivered comparable results between the two modalities.

The two control modalities differentiate also in terms of profile shape (Fig. 5). Position control manifested over all dynamics conditions a bell-shaped velocity profile typical of point to point movements, whose velocity profile has been largely proved to be invariant to duration, distance, and peak velocity [17] [21] [16] [22]. Control in velocity, on the contrary, did not show the same invariant property. By comparing our profiles (Fig. 5) with theoretical accounts that addressed the modeling of movement in terms of effort minimization [23] [24], we can advance the hypothesis that a change in the motor control strategy adopted in the mimicry took place along with the change in stimuli dynamics.

To our knowledge, no study on human pointing movements proposed requirements similar to our experiment. Literature on pointing movement tasks usually refers to Fitt's law for a quantitative description between movement time and amplitude. However, Fitt's model grounds on the condition of self-paced movements, that is movements in which the execution time behaves as a by-product of the speed-accuracy trade off - participants are required to move as fast and as accurate as possible. This model however has proved to fail in describing tasks where subjects are required to move at a specified time [20]. In this experiment the movement task is based both on temporal and on spatial constraints. Participants had to assure maximal accuracy both in the end point positions (intensity levels) and in the movement timing (duration and temporal profile). In the musical domain, similar conditions form the requisites for the performance of the violin, for which bowing techniques have been investigated in terms of motor control strategies in [25].

The comparison between profiles for isolated and consecutive notes revealed that the sequential ordering of gesture units do not resolve in simple temporal sequencing, but entails a structural change to its constituents which affects the all sequence. Velocity profiles for velocity based control, indeed, converted to bell-shaped in all participants even for peak velocities under 60 cm/s, contrary to the case of isolated notes, thus revealing that in terms of motor control, a different strategy may be in use.

Whether the mimicry of two notes should be considered as the repetition of two discrete tasks or as a per-se unary rhythmic task is an open question, which still puzzles general motor control research [26]. Brain imaging studies revealed that the performance of rhythmic movements involve different brain areas then when performing discrete movements [27], giving support to the idea that rhythmic movements cannot be considered as the concatenation of discrete movements. In the present experiment, stimuli quantity was limited to two elements in order to prevent the emergence of frequency or pace effects. However, change in profiles and different values of similarity between gestures and stimuli (Fig. 7) lead us to suppose that to some degree a change in behavior took place.

5. CONCLUSION AND FUTURE DIRECTION

In this paper we have presented the description and results of a pilot experiment in which participants mimicked with different control modalities the production of sound stimuli. In the broader context of human-computer interaction, our experiment sets, to put it as Buxton [28], on *pragmatism*, in that it considers the interdependence of transducer or control modality with the visual and kinesthetic skills engaged in the interaction. Idiosyncrasies between modalities and between users, both in values and temporal profiles, indicate that a mapping strategy capable to adapt to the control channel and to the user natural skills could accelerate the "process whereby novices begin to perform like experts" [28]. On a higher level, recognition of coarticulation effects may help in extracting semantic cues on the embedding of gesture units in human-computer phrasal dialogue.

For a future case study, we envision to improve some aspects of the experiment. First, the substitution of headphones with loudspeakers will help in the monitoring of the effective acoustical intensity delivered to the subjects. Secondly, we foresee to use trumpet tones synthetically produced by a physical modeling synthesis software [2] as acoustical stimuli. Synthesized material will permit to avoid dissimilarity between tones (see Fig. 1) caused by inconsistency in the trumpet player performance, consequently excluding the presence of uncontrolled external factors that could interfere in the mimicry task. We are currently working on the trumpet model in order to augment tongue and airflow interaction for a more realistic simulation of articulation techniques.

6. AKNOWLEDGEMENTS

This work was partially supported by Short-Term Scientific Mission program of the COST Action IC0601 on Sonic Interaction Design

7. REFERENCES

[1] B. Shackel, "Human factors and usability," pp. 27–41, 1990.

[2] R. Vertegaal, T. Ungvary, and M. Kieslinger, "Towards a musician's cockpit: Transducers, feedback and musical function," in *Proceedings of the International Computer Music Conference*, pp. 308–310, 1996.

[3] M. M. Wanderley, J. Viollet, F. Isart, and X. Rodet, "On the choice of transducer technologies for specific musical functions," in *Proceedings of the International Computer Music Conference*, 2000.

[4] M. M. Wanderley and N. Orio, "Evaluation of input devices for musical expression: Borrowing tools from hci," *Comput. Music J.*, vol. 26, no. 3, pp. 62–76, 2002.

[5] B. Caramiaux, F. Bevilacqua, and N. Schnell, "Towards a gesture-sound cross-modal analysis." Lecture Notes in Computer Science. Springer-Verlag (to appear), 2008.

[6] R. I. Gody, E. Haga, and A. R. Jensenius, "Playing air instruments: Mimicry of sound-producing gestures by novices and experts," in *Proceedings of the 6th International Gesture Workshop* (J.-F. K. Sylvie Gibet, Nicolas Courty, ed.), 2006.

[7] M. Leman, *Embodied Music Cognition and Mediation Technology*. Cambridge: MIT, 2008.

[2] http://forumnet.ircam.fr/701.html

[8] A. Boulanger, “Expressive gesture controller for an individual with quadriplegia,” in *Proceedings of the 10th Ubicomp 2008 Adjunct Programs*, pp. 113–116, 2008.

[9] N. Schnell, “Collaboration on sound gesturefication.” SID STSM Report, 2008.

[10] Q. Wang, S. Harada, T. Hsieh, and A. Paepcke, “Visual interface and control modality: An experiment about fast photo browsing on mobile devices,” in *Proceedings of Human Computer Interaction INTERACT*, Lecture Notes in Computer Science, Springer Berlin, 2005.

[11] R. Hammarberg, “The metaphysics of coarticulation,” *Journal of Phonetics*, vol. 4, pp. 353–363, 1976.

[12] R. J. Jackson, *Performance Practice: A Dictionary-Guide For Musicians*. New York: Routledge, 2005.

[13] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag New York, Inc., 2006.

[14] T. Bianco, V. Freour, N. Rasamimanana, F. Bevilacqua, and R. Caussé, “On gestural variation and coarticulation effects in sound control,” *Lecture Notes in Computer Science [to appear]*.

[15] J. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB*. Springer, 2009.

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

[17] C. Gielen, K. van den Oosten, and F. Pull ter Gunne, “Relation between emg activation patterns and kinematic properties of aimed arm movements,” *Journal of Motor Behavior*, vol. 17, no. 4, pp. 421–42, 1985.

[18] Hollerback and T. Flash, “Dynamic interactions between limb segments during planar arm movement,” *Biological Cybernetics*, vol. 44, no. 1, pp. 67–77, 1982.

[19] Flanders and Herrmann, “Two components of muscle activation: scaling with the speed of arm movement,” *Journal of Neurophysiology*, vol. 67, no. 4, pp. 931–943, 1992.

[20] R. A. Schmidt, H. N. Zelaznik, B. Hawkins, J. S. Frank, and J. T. Quinn, “Motor-output variability: A theory for the accuracy of rapid motor acts,” *Psychological Review*, vol. 86, pp. 415–451, 1979.

[21] C. G. Atkeson and J. M. Hollerback, “Kinematic features of unrestrained arm movements,” *Journal of Neuroscience*, vol. 5, pp. 2318–2330, 1985.

[22] J. Soechting and F. Lacquaniti, “Invariant characteristics of a pointing movement in man,” *Journal of Neuroscience*, vol. 1, pp. 710–720, 1981.

[23] W. Nelson, “Physical principles for economies of skilled movements,” *Biological Cybernetics*, vol. 46, no. 2, pp. 135–147, 1983.

[24] N. Hogan, “An organizing principle for a class of voluntary movements,” *Journal of Neuroscience*, vol. 4, no. 11, pp. 2745–2754, 1984.

[25] N. Rasamimanana and F. Bevilacqua, “Effort-based analysis of bowing movements: evidence of anticipation effects,” *The Journal of New Music Research*, vol. 37, no. 4, pp. 339 – 351, 2008.

[26] N. Hogan and D. Sternad, “On rhythmic and discrete movements: reflections, definitions and implications for motor control,” *Experimental Brain Research*, vol. 181, no. 1, pp. 13–30, 2007.

[27] S. Schaal, D. Sternad, R. Osu, and M. Kawato, “Rhythmic arm movement is not discrete,” *Nature Neuroscience*, vol. 7, no. 10, pp. 1136–1143, 2004.

[28] W. Buxton, “Chunking and phrasing and the design of human-computer dialogues,” pp. 475–480, 1986.