

ASPECTS OF CONTROL AND PERCEPTION

Jan Tro

Department of Telecommunications
Acoustics Group
The Norwegian University of Science and Technology (NTNU)
tro@tele.ntnu.no

SUMMARY

This paper deals with the problem of artistic control in the performance situation and its effect on the perceived and experienced musical sound. This is not a presentation of controllers and control parameters, but a discussion in a wider framework where the description of variables and quantitative measures are emphasized.

Models of human interaction in a live acoustical environment are mentioned and some results from MIDI performance analyses are presented. Perceptual issues concerning musical sound variables, visualization of sound attributes (time-varying models), analyses of musical sound, and subjective testing are discussed in the last part.

Keywords: Control Models, Performology, Music Measurements, Perception.

1. INTRODUCTION

Issues concerning the artistic production of music and musical sounds include a "macro" and a "micro" chain; the first one involving people, equipment, intentions, environments and expectations, while the latter one is a more or less well defined structure concerning the artist's ability to control the performance within the established musical framework. Education, experience, training, rehearsals and creativity will all affect the basic situation where the composer or performer has to express the final artistic control.

However, methods and procedures for testing or evaluating the composer's and the performer's control ability regarding intentions, artistic creativity and improvisation have so far been given very limited attention.

Aspects of controlling digital audio effect was discussed at the DAFx99 [13], emphasizing real-time solutions including standard controllers (keyboard, mouse, buttons, faders) and

more exotic ones (data glove, radio baton) using hyperinstruments and motion detection.

This paper will **not** present and discuss sound controllers and control parameters. Instead we will be looking for description of control models and description of variables and quantitative measures in former and present research studies.

The presentation will focus on

- sound production
- control models and measurements
- perceptual issues

2. SOUND PRODUCTION.

A simple *macro* chain example is shown in **Figure 1** where the transmission of musical intention and information from

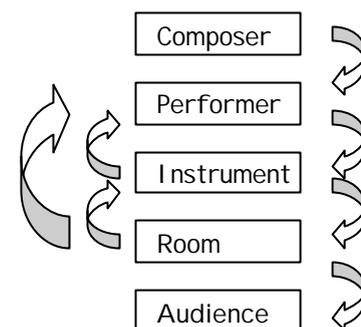


Figure 1. The Performance Feedback Model (PFM)

the composer to the listener is highly influenced by the uncontrolled feedback loops of tactile, acoustical and mental factors.

The problematic *micro* structure may be illustrated by a single performing singer. The mechanisms that affect the vocal message include psychology of the singer, musical context, rules of singing, cognitive aspects, physiology and muscle control, the sound-producing apparatus (lung, vocal folds and tract), bone and airborne sound feedback, acoustical environment and electro acoustic equipment. Similar factors may apply for the real time computer music performer.

Important questions (not completely discussed here) are:

- What kind of control do the performer demand?
- What kind of control is actually in use?
- How do musicians use controllers?
- How much control do musicians really have?

3. CONTROL MODELS AND MEASUREMENTS

There are lot of textbooks on MIDI control questions and similar data protocols. However, a more general approach to music control aspects is hard to find. Some resent conference papers and books on psychology and human behavior may, however, serve as important references ([6][12][8][1]).

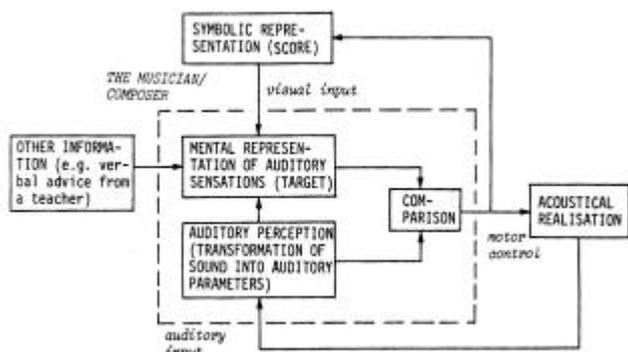


Figure 2. Cybernetic model of music realization [6][8].

The **Figure 2** shows a cybernetic model as an alternative to the *Performance Feedback Model* in **Figure 1**. Here the motor control is a result of the combination of visual and auditory inputs. Terhardt [8] states that this process comprises psychophysical items such as

- the technique of symbolic representation,
- visual and non-auditory perception,
- auditory perception,
- learning and memorizing,
- evolution and application of theoretical concepts,
- motor control of the musical instrument,
- the physics of the musical instrument,

- room acoustics in the auditory feedback.

Manual control and motor skills are comprehensively discussed in [1] pp. 386. It was found that the time required moving the hand or stylus from a starting point to a target obeys the basic principles of speed-accuracy trade-off. Faster movements terminate less accurately in a target, while targets of small area (requiring increased accuracy) are reached with slower movements.

Fitts ([1] p. 387) investigated the relationship among three variables of time, accuracy and distance. The test subjects had to move a stylus as rapidly as possible from the start to the target area. It was found that when movement amplitude (distance A) and target width (W) were varied, their joint effects were summarized by the simple equation that has become known as the Fitts's Law:

$$MT = a + b \left[\log_2 \left(\frac{2A}{W} \right) \right]$$

where *a* and *b* are constants. This equation describes the speed-accuracy trade-off in movement, e.g. movement time and accuracy (target width W) are reciprocally related. The quantity $\log_2(2A/W)$ is called the *index of difficulty (ID)*. This index forms one possible objective procedure that may be useful for the measurements of performer success.

We may define control variables of increasing complexity, such as *gain control* and *time delay* (1. order), *velocity* (2. order) and *acceleration* (3. order). A line of controllers for *pitch* (keyboard), *vibrato* (modulation wheel) and *spectrum manipulation* (modulation index control) may form a comparable hierarchy of complexity in sound response.

The performer's response or action, however, may be split into two different portions. **Figure 3** (from [1]) shows the possible distinction between resources underlying perception (necessary for the feedback loops) and resources underlying the selection of actions (decision of what to do).

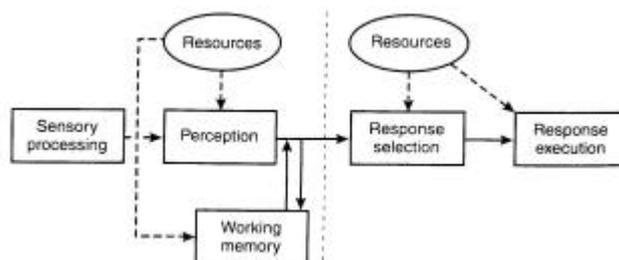


Figure 3. Distinction between resources for perception and action [1].

In the field of *quantitative performology* the performer action is measured either by some kind of automatic event recorder or documented by a thorough description of the performer control possibilities, intentions, deviations and variations. With a reliable and calibrated MIDI setup it is easy to measure basic control parameters such as keyboard timing and dynamics. Even if this is quite simple performer action, it still may indicate and document intriguing processes.

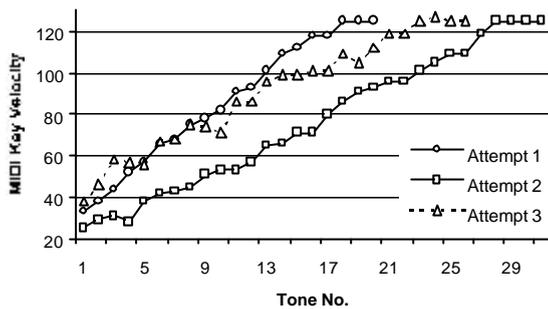


Figure 4. *Crescendo performance with one key [15].*

The ability of performing a slowly increasing crescendo by tapping one key repeatedly on the piano keyboard gives us an indication of the highest possible dynamical precision and sound level resolution in piano performance. On average three semi-professional pianists performed this crescendo with in total 9 dynamic steps.

Figure 4 (from [15]) shows an example of such a crescendo played by an extremely skilled Californian performer obtaining more than 20 dynamic steps using a Yamaha Disklavier Grand (recorded at the Center for Music Experiments, UCSD). The dynamic steps vary from 1 to 10, measured in MIDI Key Velocity steps. The average step size was close to 3. The recording and performing procedure was as follows:

The performer was asked to play one crescendo from very soft to very loud by tapping the middle c on the keyboard repeatedly as many time as necessary. This recording is marked *Attempt 1*. The performer was immediately asked to do one more performance, with the incredible *Attempt 2* as the result. Even if the performer was not expecting another try we made one more recording (*Attempt 3*).

The different results of the 3 consecutive attempts may be explained by one psychological and physiological *rehearsal effect* (from *Attempt 1* to a better result in *Attempt 2*), and one *fatigue effect* (from a superb performance in *Attempt 2* to a poorer controlled performance in *Attempt 3*).

A second professional piano performer obtained a similar effect.

Concerning the precision and reliability of the MIDI measuring methodology, some data are reported in [16]. We made a comparison of two anechoic DAT recordings; one original test performance (A) and one repeated MIDI controlled playback (B). The average time difference of tone attacks was 5 ms (std. 5 ms, max. dev. 12 ms). **Figure 5** shows RMS values (dB) for single tones. Level calculations coincide very well for the two recordings with an average difference of .2 dB (std. .4 dB).

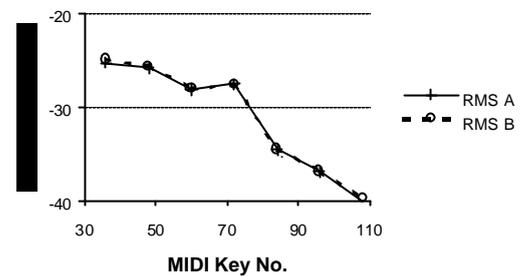


Figure 5. *Comparison of single tone sound levels in two recordings [16].*

These MIDI analysis examples give us an idea of how the performer struggles to control his instrument and, not least, how successful he normally is.

We accept technology as a part of or even the reason for the development of new tools, often with increased control abilities.

The transformation from hardware sound mixers to software processing units, however, is a good (or bad) example of how technology contributes to decreased control ability. Some software programs still include the mouse-cursor operated knobs or sliders as a substitute for the manually hand operated gears. Luckily some of the manufacturers have developed their own tailor-made controller in order to avoid the mouse-cursor dilemma.

It is not too difficult to find master reports and dissertations dealing with performers interaction and environmental influences (ensemble precision [11], room [10], delayed feedback [9] etc.) and there are books and reports on instrument directivity and well-documented equipment for sound distribution systems as well, both hardware and software.

However, the distribution of timbre intentions in a dynamic time-frequency space is a lot more complicated issue and cannot be defined in a single question.

Important questions (not completely discussed here) are:

- How to define control parameters?
- How to measure the parameters?
- How to measure the degree of control?
- What is the link between intended and experienced control?

4. PERCEPTUAL ISSUES.

Definition of musical sound variables, visualization of sound attributes, analyses of musical sound, and subjective testing and evaluation are basic questions in this paragraph.

The situation of listening to music may involve aspects similar to the microstructure of a performance, i.e. influenced by the environment, expectations, your hearing mechanism, multi dimensional stimuli, musical background and training, mental situation etc.

One definition of *timbre* is given by ANSI, the American National Standards Institute [17]: "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar."

This definition is difficult to use constructively as it merely tells us what *timbre* is not.

Timbre is often regarded in terms of a multi-dimensional *timbral space*. The dimensions of this timbral space represent different perceptual parameters of sound.

Timbre will normally refer to a holistic view of the sensation of a tone. It will include neither the basic pitch of the tone (if it exists), nor the loudness of the tone as a whole. It will, however, include pitch variations during the note (slurs, glissando, vibrato etc.) and the loudness envelope. Evaluated this way, the subjective attribute *timbre* corresponds to the objective terms *time-varying fundamental frequency*, *amplitude envelope* and *time-varying spectrum*. The identity of the subjective components of timbre is to a large extent regarded as unknown.

Although subjective parameters such as *sharpness* and *roughness* (Zwicker & Fastl [19]) have been proposed, researchers have reached no general consensus of how timbre should be broken down into smaller components.

Splitting timbre into time-varying *spectrum*, *pitch* and *loudness* functions is used extensively in many commercial sound synthesizers. An oscillator provides the pitch and a broadband spectrum, which is modified by a filter with time-varying parameters. The result is then run into a time-variant attenuator, which provides the loudness contour.

The fact that all these parameters are time-variant is an important observation.

Zwicker and Fastl [19] note that amplitude modulation of a sound is perceived as fluctuation at low frequencies (maximum fluctuation is perceived at 4Hz), and starts to be perceived as roughness around 15 Hz. The ear does not perceive a modulation with a frequency significantly higher than this as a modulation, but rather as a change in the spectrum. This is valid for both amplitude and frequency modulation.

Time-variant *spectral centroid* is one possible method of describing and visualizing spectral changes. The spectral centroid is a measure of the mean value of the spectrum, analogous to the center of mass for a physical object.

Referring to the DAFx Sound Working Group Specifications the centroid is computed as:

$$C = \frac{F_s}{2N} \cdot \frac{\sum_{i=0}^N i \cdot f(i)}{\sum_{i=0}^N f(i)}$$

where F_s is the sampling rate, $2N$ is the FFT size, and $f(i)$ is the magnitude of the Fourier transform of the input signal for the frequency bin i . The spectral centroid is often mentioned in connected with the subjective term *brightness* or *sharpness*.

This time-varying centroid is a direct extension of the steady-state concept. For evolving sounds, the movement of the centroid will probably be more interesting than an average value.

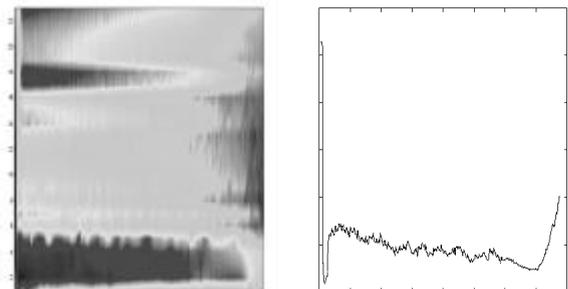


Figure 6. *Spectrogram and time-varying spectral centroid (left: time-Bark axes; right: time-frequency axes).*

A study of spectrogram representation combined with the time-varying spectral centroid measurement is reported in [4] (see **Figure 6**).

Spectrographical methods for analyzing dense sound mixtures and contemporary music have been reported [7][12]. Hettergott [7] presents a qualitative analysis of modern music as an attempt to broaden the methodology for this specific music genre where the pitch, the melodic and the chord sensation are typically not emphasized.

Ellis [12] develops a new model for a computational auditory scene analysis (CASA). Compared to former data-driven models this new approach is prediction-driven. The central operating principle is that the analysis proceeds by making a prediction of the observed cues expected in the next time slice based on the current state. This is then compared to the actual information arriving from the front-end. These two are reconciled by modifying the internal state, and the process continues.

Other parameters describing the spectrum have been proposed, like the *tristimulus* value (Pollard & Jansson [5], and *roughness* (Terhardt [20]).

The *tristimulus* parameter depends on the relative strengths of the fundamental (N_1), a group consisting of the second, third and fourth partials (N_2^4), and a group consisting of the fifth partial and all partials above (N_5^n). This parameter is time-variant, and is presented as a trajectory in a triangular graph. The total loudness, N , of the sound may be expressed as

$$N = N_1 + N_2^4 + N_5^n$$

This parameter is useful for visualizing the spectrum differences between instruments, between tones from one instrument and between portions of one evolving tone.

Roughness is a perceptual parameter related to amplitude fluctuations within a critical band. These fluctuations occur when a sound is subject to amplitude or frequency modulation. This parameter was first noted by Terhardt [20] and later discussed in [19].

Another interesting parameter is the *inharmonic* of the partials. Rossignol, Rodet, Soumagne, Collette and Depalle [18] define *inharmonic* for each partial as being given by:

$$h_n = \frac{|f_n - n \cdot f_0|}{n \cdot f_0}, n \in [2, N]$$

where f_n is the frequency of the partial, n is the number of the partial, and f_0 is the frequency of the fundamental. The total *inharmonic* for a complex tone is given as the sum of the *inharmonic* of the partials:

$$H(i) = \sum_{n=2}^N h_n$$

These formulas can be used to gain an objective measurement of *inharmonic*. However, the relation between this objective measure and the subjective sensation is intuitively not easy to establish.

Musically spoken *loudness* may be the most underestimated perceptive attribute as it refers to both the listening level as well as the fluctuation of musical dynamics. Some unexpected results have been reported.

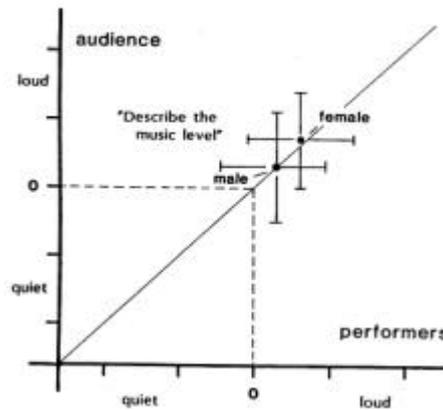


Figure 7. Evaluation of musical "strength" [14].

In a music listening study [14], we evaluating experienced long-term sound level in a group of musicians (162 performers) and listeners (38) after having attended the first performance of a 20 minutes piece of music. Analyzing the results from the question "Describe the over-all music sound level", we found a significant difference in the evaluation between male and female as shown in **Figure 7**. On the question "Describe the over-all dynamic range in the music", we found a significant difference between the audience and the performers.

5. SOME CONCLUDING REMARKS

This presentation tries to address the control question in a broad sense within an academic tradition. It seems to be necessary to include knowledge about classical and contemporary music related control aspects as

- composer control
- instrumentalist control (acoustical instruments as sound source)
- orchestral control (ensemble control)
- computer music control (real/not-real time signal synthesis and processing)

- artists performance control (real time)
- recording/transmission/distribution control (protocols, procedures, methods, quality, legal aspects)
- interaction and transmission of intention among performers and with the audience (psychological and philosophical aspects).

The main problem here is the lack of reliable data as the basic input for computer models. Even if stochastic data distributions may form acceptable first order estimations, we have to remember that human artistic behavior does not fit into a stochastic framework.

As a final remark, the question about music listening level should be taken very seriously due to the risk of hearing damage. In a music listening study [15] the correlation analysis among the four factors "preference", "level", "dynamics" and "rhythmics" showed that the *preference* was higher correlated to *dynamics* compared to *level*. This may be due to the acceptance of music dynamics as a more significant musical attribute compared to the sound level.

This can be interpreted in such a way that we have to emphasize the difference between soft and loud sound portions (dynamics) in order to enhance the standard music listening conditions.

6. ACKNOWLEDGEMENT

This report is a part of the ongoing Music Technology Research Project at the Acoustics Group, NTNU. Discussions and comments from laboratory colleagues Professor Ulf Kristiansen and Professor Peter Svensson have been highly appreciated. Karl Helmer Torvmark has written the centroid analysis routines, performed subjective listening tests and contributed to the present paragraph on perceptual issues.

7. REFERENCES

- [1] Wickens, C. D. & Hollands, J. G., "Engineering Psychology and Human Performance", ISBN 0-321-04711-7, Prentice-Hall, 3rd ed., 1999.
- [2] Arakawa, K. et al., "The Context Effect on Loudness in Listening to Music", *Journal of Music Perception and Cognition*, vol. 2, pp. 18-26 (1996) (in Japanese)
- [3] Kuwano, S. et al., "Impression of Smoothness of a Sound Stream in Relation to Legato in Musical Performance", *Perception and Psychophysics*, 56 (2), pp. 173-182 (1994).
- [4] Torvmark, K.H., "Presentation and Evaluation of Timbral Microstructures", M.Sc. Thesis, Norwegian University of Science and Technology, Trondheim, December 1999.
- [5] Pollard, H.F. & Jansson, E.V., "A Tristimulus Method for the Specification of Musical Timbre", *Acustica*, vol. 51, pp. 162-171, 1982.
- [6] Laine, P., "The Cybernetic Perspective to Music Algorithms - The Control Feedback in Cognitive Modelling", *Proceedings of the 5th International Conference on Music, Perception and Cognition (ICMPC)*, pp 165-170, Seoul, August 1998.
- [7] Hettergott, A., "Aspects in the Spectrographical Analysis of Modern Music", Poster, 4th Int. Conference on Music, Perception and Cognition (ICMPC), Montreal, August 1996.
- [8] Terhardt, E., "Impact of Computers on Music", in Clynes, M. (ed.), "Music, Mind and Brain", ISBN 0-306-40908-9, Plenum Press, New York, 1982.
- [9] Willey, R.K., "The Relationship Between Tempo and Delay and its Effects on Musical Performance", Ph.D. thesis, University of California, San Diego, 1990.
- [10] Bolzinger, S., "Contribution a l'etude de la retroaction dans la pratique musicale par l'analyse de l'influence des variations d'acoustique de la salle sur le jeu du pianiste", Dr. thesis, Université de la Méditerranée, Marseille, 1995.
- [11] Opperud, Chr., "Presisjon i musikkampspill", Music Technology Project, Norwegian Institute of Technology, Acoustics, Trondheim, 1989.
- [12] Ellis, D.P.W., "Prediction-driven Computational Auditory Scene Analysis for Dense Sound Mixtures", ESCA Workshop, Keele, July 1996.
- [13] Todoroff, T., "Controlling Digital Audio Effects", Proc. of 2nd COST G6 Workshop DAFx99, NTNU, Trondheim, December 9-11, 1999.
- [14] Tro, J., "Aspects of Long Term Music Listening", Proc. of 6th FASE Symposium, Sopron, 2-6 September, 1986.
- [15] Tro, J., "Aspects of Music Listening: Musical Dynamics", *Proceedings of NAM94*, Aarhus, 6-8 June, 1994.
- [16] Tro, J., "Data Reliability And Reproducibility In Music Performance Measurements", WESTPRAC VII Conference, Kumamoto 3-5 October, 2000.
- [17] ANSI, "USA Standard Acoustical Terminology (Including Mechanical Shock and Vibration)", S1.1-1960 (R1976). New York: American National Standards Institute, 1960.
- [18] Rossignol, S., Rodet, X., Soumagne, J., Collette, J.-L., Depalle, P., "Feature extraction and temporal segmentation of acoustic signals" [Online]. Paris : ICRAM. Available from: <http://www.ircam.fr/equipes/analyse-synthese/rossigno/icmc98/article6.html> [Accessed 19.11.1999]
- [19] Zwicker, E., Fastl, H., "Psychoacoustics : Facts and Models". Second edition. Berlin : Springer-Verlag, 1999.
- [20] Terhardt, E., "On the Perception of Periodic Sound fluctuations (Roughness)". *Acustica* 30, pp. 201-213, 1974.