Modeling piano performance:
Physics and cognition of a virtual pianist

Richard Parncutt
Department of Psychology, Keele University
r.parncutt@keele.ac.uk, http:// www.keele.ac.uk/depts/ps/rpbiog.htm

In: T. Rikakis (Ed.), Procs Int. Computer Music Conf. (Thessaloníki/GK, 1997) (pp. 15-18). San Francisco/CA, 1997.

Abstract

The musical quality of computer-generated performances of piano scores might be improved by incorporating a physical model of the pianist's body as it interacts with the instrument. Issues considered include the various aspects of musical structure (as conceived by a virtual pianist) that interact with expression, and the expressive function and algorithmic determination of fingering (as used by a virtual pianist).

1 Introduction

Computer simulations of expressive piano performance are becoming increasingly sophisticated. But they do not yet produce consistently satisfying results. In this regard, musical expression is lagging behind timbral synthesis. Most instrumental timbres can now be convincingly synthesized by physical models (Smith, 1996).

What's wrong? Have researchers been neglecting vital aspects of the problem? Most recent computer-based attempts to create expressive musical performances (Clynes, 1985, 1986; Friberg, 1991; Honing, 1990; Mazzola, 1994; Sundberg, 1988; Todd, 1989) have focused on the role of musical structure (Clarke, 1988). But perhaps the relationship between structure and expression is only part of the challenge.

Some new and creative approaches have recently appeared. Clarke (1995) attempted to explain aspects of performance expression in semiotic terms, citing Shaffer's (1992) proposal that a musical performance is a kind of narrative, in which *events* are determined by the musical structure, and the characters or personalities of *protagonists* are created by the performer through expressive interpretation. Such arguments, while convincing at a qualitative level, do not immediately lend themselves to application in a computer-based performance system.

How about the physical means by which musicians create their interpretations – their bodies, brains, ears, lungs, lips, fingers? Could the modeling of piano performance, like timbral synthesis, benefit from the introduction of physical models?

The model of temporal integration and integrated energy flux proposed by Todd (1994) and developed by Clarke (1995) may be regarded as a physical model of expression based on the physiology of the auditory periphery. But the most important physical models of expression address music's apparent ability to imply movement (Shove & Repp, 1995). Todd's (1995) model of the kinematic implications of rhythm and phrasing may be regarded as a physical model of expression based on musical structure as conceived by the pianist. Similarly physical in their origins and implications are Truslit's implied motional forms (Repp, 1993), and Langer & Kopiez's (1995) and Large & Kolen's (1994) banks of coupled, resonating oscillators. I will not consider these alternative approaches in this paper. Instead, I will consider the possibility of modeling perhaps the most conspicuously "physical" element in the system: the body of the pianist as it interacts with the instrument.

The idea of a synthetic performer has a long history in computer music circles (e.g., Vercoe, 1984), but the idea of including the physical properties of the performer's body in such a model is rather newer. The musical output of Vercoe's (1988) virtual accompanist was determined entirely by the notated and acoustic structure of the music. Similarly, Jean-Claude Risset's piece *Duet for One Pianist* (Risset & Van Duyne, 1996) involved only MIDI interaction between a live performer and a computer.

Here, "virtual pianist" alludes to a hypothetical model that combines both physical and cognitive aspects. I will consider possible influences on expression both of the pianist's conception of the musical structure and of the pianist's body as it interacts with the instrument.

2 Structural Models

Let us first overview the various structural models of expressive performance that have been developed in recent years. Most have been biased toward certain aspects of structure, while neglecting others. The tendency for performers to speed up at the start of phrases and slow down at the end of phrases, sections, or whole pieces, and for timing patterns to reflect the hierarchical structure of phrases at different levels, has been studied in considerable quantitative detail (Repp, 1992; Sundberg & Verrillo, 1980; Todd, 1985), and was the primary focus of the model of Todd (1989). The frustratingly idiosyncratic model of Clynes (1985, 1986) was mainly concerned with modulation of timing and dynamics as a function of metrical position, again at different hierarchical levels. In contrast, Sundberg's (1988) model focuses on the musical surface (*leap articulation, harmonic* and *melodic charge*, etc.). A major recent advance in the structural modeling of expression (Widmer, 1995) has allowed a system of rules, similar to that of Sundberg and colleagues, to be derived directly and objectively by comparing performance data with structural analyses of the music based on the theories of Lerdahl & Jackendoff (1993) and of Narmour (1977).

Clearly, there is no easy way out. A general structural model of expression would need to consider, in a coherent fashion, a wide variety of different aspects musical structure, appropriately balancing surface and deeper levels (Mazzola & Zahorka, 1994).

A recent attempt at a unified model (Parncutt, 1997) is based on a broad definition of an *accent* as any relatively salient event, or any event that attracts the attention of a listener more than surrounding events (see also Parncutt, 1994). In both speech and music, listeners need to get a feel for the importance of individual events (syllables, notes) relative to each other if they are to correctly infer the underlying structure and meaning.

Musical accents may be classified following Lerdahl and Jackendoff (1983):

STRUCTURAL or between-category accents

• time	– grouping
	– metrical
• pitch	– melodic (contour)
	– harmonic
• loudness	– dynamic
• timbre	– instrument/orchestration

EXPRESSIVE or within-category accents

• time	– onset time (agogic)
	– duration (articulatory)
	– amplitude envelope
• pitch	– intonation
• loudness	– stress
• timbre	– coloration

Grouping accents occur at starts and ends of note groups at different levels, from phrases through sections to whole pieces. *Metrical* accents are similarly hierarchical, and may be identified and quantified both within and between beats, and within and between measures. *Melodic* accents may be divided into *turns* and *skips* (Drake & Palmer, 1993), of which turns are more important (Huron & Royal, 1996). Turns are peaks and valleys of the melodic contour (peaks being more important, Thomassen 1982). Skips are disjunct intervals between consecutive tones; the wider the interval preceding the tone, the stronger is its accent, and rising skips produce stronger accents than falling. *Harmonic* accents occur either "horizontally", at sudden changes of harmony, or "vertically", at points of acoustic dissonance. *Dynamic* accents are explicitly marked in the score. *Timbral* accents occur at changes of instrumentation or articulation.

The various kinds of *expressive* accent listed above represent the means that musicians have for bringing out structural accents. A performer may manipulate onset and offset times, amplitude envelopes, precise pitch (intonation), physical intensity (stress), and timbre. Of these, only stress, agogic accent, and articulatory accent are available to the pianist.

Based on the above taxonomy, expressive music performance may be regarded as a process whereby *expressive accents reinforce structural accents*. This broad concept is consistent with the variety of "good" interpretations of a single piece. A musical score typically includes many different kinds of structural accent, of various strengths or saliences (see Parncutt, 1997 for examples). During practice and performance, performers are constantly making decisions – largely intuitively and unconsciously – as to which accents should be emphasized, and how and to what extent. Current knowledge does not yet allow us to create a viable scientific model of the underlying cognitive processes

Consequently, computer-based simulations of musical performance, if they are to be musically convincing, really cannot avoid *involving the user* in the interpretative process. It appears for the moment to be impossible to arrive at a good interpretation solely by means of abstract and supposedly general principles. The user needs to be invited to choose which accents to emphasize, and by how much. The user should also be free to manipulate associated modulations of timing and dynamics. In the case of dynamics, stresses may be large or small, sudden or gradual. A "gradual stress" would involve a gradual increase in loudness for notes leading to the accent, and a gradual decrease afterwards. In the case of timing, agogic accents may involve a slowing of local tempo before the event and a speeding up after it; and articulatory accents may involve a temporary increase in *legato* for tones near a given event. The creation of user-friendly interfaces for such adjustments, including "pop-up", "nudgeable" local and global curves for modulation of timing and dynamics in the vicinity of specific accents, would be a straightforward programming exercise. After each change to the interpretation, the result would be synthesized, listened to, and appraised by the user, who would then make further changes. After many such iterations, the user would arrive at an interpretation of the piece that is not only musically acceptable (assuming that the user has sufficient musical skills, talent, and patience) but also unique and personal.

A relatively theory-free music interpretation system has recently been developed by Dalgarno (1997), with encouraging results. Dalgarno's system could be made more flexible by adding elements of the theory set out above, while always allowing the user the option of skipping the theory altogether and directly manipulating the surface parameters.

Learning to play an instrument invariably requires thousands of hours of practice (Sloboda & Howe, 1991), including the development of technique by seemingly endless repetition of scales, arpeggios, exercises, and studies. There is no compelling reason why some performers should not now take advantage of appropriate technology to reduce the time they spend on technical work to allow more time for interpretation. Moreover, computer-assisted musical interpretation would be attractive for musicians with physical disabilities that otherwise prevent them from reaching high levels of technical expertise (Dalgarno, 1997).

3 A Physical Model of Piano Fingering

Miller (1996) showed that fingering can affect the sound of a performance. Pianists played an unfamiliar piece by Beethoven using 3 different fingerings: their own; that of the composer; and a fingering devised by the experimenter. When presented with audio recordings, independent listeners consistently preferred the first performances to the second, and the second to the third.

The note-by-note mechanism by which fingering affects performance has not yet been systematically investigated. But interview data (Clarke et al, 1997) suggest that fingering may affect performance in specific ways that would be directly amenable to computer modeling. For example, *legato* (as measured by the degree of temporal overlap between successive tones, Repp, 1995) is more likely to be maintained *within* hand positions than *between* them (that is, at changes of hand position), and between strong fingers than between weak fingers. Notes played with the thumb tend to be played more loudly and held down for longer than notes played by other fingers. In fast passages, articulation tends to be clearer (the onset time of a note correlating more closely with the offset time of its predecessor, and key velocities being more consistent) when stronger fingers are used (*1*, *2*, *3* rather than *4*, *5*).

How could these ideas be applied to computer-based interpretation of piano music? First, one would need to decide on a fingering. Second, performance parameters could be modified to take account of fingering effects.

A suitable fingering could be decided for the piece, either directly by the user, or by a model. Better, these two approaches could be combined, the model generating a list of possible fingerings and the user deciding among them. Best of all, successive "performances" of the piece could be made subtly different by having the model prescribe a probability value for each fingering; the user could then adjust the values. Such changes in fingering would generate aleatoric variations in performance parameters, and may shed light on otherwise inexplicable "unintentional" or "random" fluctuations.

A model has recently been advanced for the prediction of fingerings of melodies (Parncutt, Sloboda, Clarke, Raekallio, & Desain, 1997). The fingerings used by keyboard players are determined by a range of ergonomic (anatomical/motor), cognitive, and music-interpretive constraints. The model attempts to encapsulate only the most important *ergonomic* constraints; it may thus be regarded as a kind of physical model, based on the physiology of the fingers and hands.

The model begins by generating all possible fingerings for a melodic fragment, limited only by maximum practical spans between finger-pairs. Many of the fingerings generated in this way seldom occur in piano performance. For example, the melodic fragment C4-E4-G4 may be fingered *121, 123, 124, 125, 131, 134, 135, 141, 145, 212, 213, 214, 215, 231, 234, 235, 241, 245, 312, 313, 314, 315, 341,* or *345.*

Next, the difficulty of each fingering is estimated using a rule system. Each is named after a technical difficulty that a pianist might seek to avoid; e.g., the "Weak-Finger Rule" accounts for the tendency to avoid the weaker fingers *4* and *5*, and the "Stretch Rule" accounts for the avoidance of unnecessary stretches between fingers. The other rules are called Small-Span, Large-Span, Position-Change-Count, Position-Change-Size, Three-Four-Five, Three-to-Four, Four-on-Black, Thumb-on-Black, Five-on-Black, and Thumb-Passing. The sum of all 12 rule contributions is the predicted difficulty or "cost" of a specific fingering. The relative importance of each rule may be adjusted by applying a weight (or linear coefficient) to it; different pianists and musical styles may require different relative weightings. For full details of the model see Parncutt et al. (1997) or Sloboda, Parncutt, Clarke, & Raekallio (1997).

Once fingerings have been assigned difficulty estimates, they are ranked in order of calculated difficulty. The fingering with the lowest difficulty is predicted to be used most often in performance. The fingerings that pianists actually use or recommend are expected to appear among the least difficult calculated fingerings.

4 A Physical Model of the Pianist's Body

Fingering, of course, is only one aspect of the physics of a virtual pianist. The sound of a piano performance depends also on *how* fingers are used to play notes – not to mention the role of the wrists, forearms, upper arms, back, and torso. A recent series of interviews clarified the complex relationship between fingering, interpretation, and the physical interaction between the pianist and the keyboard (Clarke et al., 1987).

A model of piano performance incorporating a virtual pianist could improve on Parncutt et al.'s (1997) model of fingering in two ways. First, it might do away with some of the existing 12 rules and instead estimate difficulty directly from a physical model of the pianist's hands and arms in their relationship with the keyboard. For example, certain fingerings make it physically easier to play consistently loudly, legato, or with "arm weight". In a second stage (closely related to the first), the model could take into account the physics of the body in modeling the execution of the notes. Both stages could take advantage of recent advances in related disciplines such as motor control and robotics.

A physical modeling approach could eventually account for all physical aspects of piano performance. A model including room acoustics could determine when and how the pedal is used, as investigated by Repp (1996). A model knowing about physical properties of the hands, including interactions between fingers and between hands (cf. Parncutt, Sloboda, & Clarke, 1997), could account for timing of arpeggiated chords (Repp, in press, a) and for the complex combination of expressive intentions and motor constraints that determine onset asynchronies in notated simultaneities (Palmer, 1996; Repp, in press, b). A physical model of the pianist's body could even inform the process by which pianists get used to the touch of unfamiliar pianos, as "information flows from the instrument to the performer, via sound, and also through a bi-directional information flow from the haptic senses (tactile, kinesthetic, force, etc.)" (Van den Berghe, De Moor, & Minten, 1995, p. 1).

Even the relationship between acoustic interpretation and gross bodily movements of the performer might be amenable to modeling. Observers can more reliably recognize the expressive intent of a pianist by *vision* than by hearing. (Davidson, 1993, 1994). Specific parts of the pianist's body, such as the hands and upper torso, are related to specific forms expression; for the pianist that Davidson studied, hand lifts occurred at rests or sustained notes, and sharp forward movements of the head occurred at cadence points. Larger movements were consistently associated with higher degrees of expressiveness. Modeling of body movements might explicate otherwise mysterious cyclic developments in tempo and dynamics occurring partially independently of the musical structure in pieces such as Chopin's Etude in E minor, for which an abundance of performance data already exist (Clarke, 1995; Sloboda, Lehmann, & Parncutt, 1997).

5 Discussion

At first sight, the various physical constraints on performance may appear as *hindrances* to the realization of a desired musical effect. Surely if a computer model knows about the performer's subliminal concept of the musical structure and intentions for conveying the structure to the audience, that should be sufficient to produce a convincing simulation of a real performance? Aren't physical constraints exactly what computer modeling can succeed in liberating us from?

These questions cannot be answered until convincing computer-generated realizations of musical scores become everyday reality. Meanwhile, some insight may be gained by considering the parallel case of *timbral synthesis*. Synthesized musical instrument sounds require a great deal of physical complexity to fool trained ears. If any audible component of the sound is missing, the result usually sounds musically worse than the original. Yet much of this complexity does not seem to have any particularly *musical* point to it. To take a simple example, for what intrinsically *musical* reason does a clarinet tone need a noisy onset? Apparently, what sounds *good* is often no more than what sounds *familiar* from good musical performances. In other words, the subjective quality of synthesized sound is determined by auditory association.

Much the same principle may be expected to hold in the case of simulated expressive piano performance. Just as the individual tones of the performance need to sound genuine in their acoustic complexity, so too, I would venture, should the timing and dynamics of the performance reflect the complexity of the physical processes that enable that performance to be realized by a real human being sitting at a real piano keyboard – and that includes the constraints that limit, often frustratingly, what real performers can do.

6 References

Clarke, EF 1988. "Generative Principles in Music Performance," in Sloboda, J (Ed.), *Generative Processes in Music.* Oxford .

Clarke, EF 1995. "Expression in Performance: Generativity, Perception, and Semiosis," in J Rink (Ed), *Practice of Performance.* Cambridge U P.

Clarke, EF, R Parncutt, M Raekallio & JA Sloboda 1997. "Talking Fingers: An Interview Study of Pianists' Views on Fingering," Musicae Scientiae **1**.

Clynes, M 1985. "Music Beyond the Score," Communication & Cognition **19,** 169–194.

Clynes, M 1986. "Generative Principles of Musical Thought," Communication & Cognition AI, **3**, 185–223.

Dalgarno, G 1997. "Creating an Expressive Performance Without Being Able to Play," Brit J Music Ed.

Davidson, JW 1993. "Visual Perception of Performance Manner in Movts of Solo Musicians," Psychol Mus **21**, 103–113.

Davidson, JW 1994. "Which Areas of a Pianist's Body Convey Information about Expressive Intention?," J Hum Mov Stud **6**, 279–301.

Drake, C & C Palmer 1993. "Accent Structures in Music Performance," Mus Percept **10**, 343-378.

Friberg, A 1991. "Generative Rules for Music Performance," Comp Mus J **15** (2), 56–71.

Honing, H 1990. "POCO: An Environment for Analysing, Modifying, and Generating Expression," Proc ICMC.

Huron, D & M Royal. 1996. "What is Melodic Accent?" Mus Percept **13**, 489–516.

Langner, J & R Kopiez 1995. "Entwurf einer neuen Methode der Performanceanalyse," Jahrbuch Musikpsychologie **12**, 9–27.

Large, EW & JF Kolen 1994, "Resonance and Perception of Musical Meter," Connect Sci **6**, 177–208.

Lerdahl, F & R Jackendoff 1983. *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Mazzola, G & O Zahorka 1994. "Tempo Curves Revisited," Comp Mus J **18** (1), 40–52.

Miller, MF 1996. "Piano Fingerings and Musical Expression," Poster, Soc Mus Theory, Baton Rouge LA.

Narmour, E 1977. *Beyond Schenkerism.* U Chicago P.

Palmer, C 1996. "On Assignment of Structure in Music Performance," Mus Percept **14**, 23–56.

Parncutt, R 1994. "Perceptual Model of Pulse Salience and Metrical Accent," Mus Percept **11**, 409–464.

Parncutt, R 1997. "Accents and Expression in Piano Performance," in W. Auhagen et al (Eds), *Systemische Musikwissenschaft.* Krefeld: Dohr.

Parncutt, R, JA Sloboda, EF Clarke, M Raekallio & P Desain 1997. "Ergonomic Model of Fingering," Mus Percept.

Parncutt, R, JA Sloboda & EF Clarke 1997. "Interdependence of right and left hands in sight-read, written, and rehearsed fingerings," Proc. Euro. Soc. Cog. Sci. Music, Uppsala.

Repp, BH 1992. "Diversity and Commonality in Music Performance," J Acoust Soc Am **92**, 2546–2568.

Repp, BH 1993. "Music as Motion: A Synopsis of Truslit (1938)", Psychol Mus **21**, 48–72.

Repp, BH 1995. "Acoustics, Perception, and Production of Legato Articulation," J Acoust Soc Am, **97**, 3872–3874.

Repp, BH 1996. "Pedal Timing and Tempo in Expressive Piano Performance," Psychol Mus **24**, 199–221.

Repp, BH in press, a. "Some Observations on Pianists' Timing of Arpeggiated Chords," Psychol Mus.

Repp, BH in press, b. "Patterns of Note Onset Asynchronies in Expressive Piano Performance," J Acoust Soc Am.

Risset, J-C, & S Van Duyne 1996. "Real-Time Performance Interaction with Computer-Controlled Acoustic Piano," Comp Mus J **20** (1), 62–75.

Shaffer, LH (1992). "How to Interpret Music," In MR Jones & S Holleran, *Cognitive Bases of Musical Communication*. Washington, DC: Am Psychol Assoc.

Shove, P, & BH Repp 1995. "Musical Motion and Performance" in J Rink (Ed), *Practice of Performance*. Cambridge U P.

Sloboda, JA & MJA Howe 1991. "Biographical Precursors of Musical Excellence," Psychol Mus **19**, 3–21.

Sloboda, JA, A Lehmann & R Parncutt (1997). "Perceiving Intended Emotion in Concert-Standard Performances of Chopin's Prelude No. 4 in E-Minor," Proc Euro Soc Cog Sci Mus, Uppsala, Sweden.

Sloboda, JA, EF Clarke, R Parncutt & M Raekallio 1997. "Determinants of Fingering Choice in Piano Sight-Reading," J Exp Psychol: Hum Perc Perf.

Smith, J O III 1996. "Physical Modeling Synthesis Update," Comp Mus J **20** (2), 44–56.

Sundberg, J 1988. "Computer Synthesis of Music Performance," in JA Sloboda (Ed), *Generative Processes in Music.* Oxford.

Sundberg, J & Verrillo, V 1980. "On Anatomy of Retard," J Acoust Soc Am **68**, 772–779.

Thomassen, J 1982. "Melodic Accent: Experiments and Tentative Model," Am J Psychol **12**, 546–560.

Todd, NPMcA 1989. "Computational Model of Rubato," Contemp Mus Rev **3**, 69–88.

Todd, NPMcA 1994. "Auditory 'Primal Sketch'," J New Mus Res **23** (1), 25–70.

Todd, NPMcA 1995. "Kinematics of Musical Expression," J Acoust Soc Am **97**, 1940–1949,

Van den Berghe, G, B De Moor & W Minten 1995. "Modelling Grand Piano Key Action," Comp Mus J **19** (2), 15–22.

Vercoe, B 1984. "Synthetic Performer in Context of Live Performance." Proc ICMC.

Vercoe, B 1988. "Connection Machine Tracking of Polyphonic Audio." Proc ICMC.

Widmer, G 1995. "Modeling Rational Basis for Musical Expression," Comp Mus J **19** (2), pp. 76–96.