Published article: Windsor, WL 'Gestures in music-making: action, information and perception', in New Perspectives on Music and Gesture ed. Anthony Gritten and Elaine King (Farnham: Ashgate, 2011), pp. 45-66(21).

http://www.ashgate.com/default.aspx?page=637&lang=...

Chapter 2
Gestures in Music-making: Action, Information and Perception

W. Luke Windsor

This chapter will explore some issues related to the analysis, perception and production of gestures in music-making. Although it will touch on the most obvious modality through which gestures are viewed by audiences, the visual, it will take a more poietic (and perhaps neutral) approach to defining gestures themselves, rather than an esthesic approach (see Nattiez 1990 for an explanation of these distinctions): for the purposes of this chapter, gestures will be considered to be certain *movements made by musicians*. Such movements may complement or indeed express other kinds of 'gestures' (such as those a music analyst or performer might discover in a score) and may be perceived through a range of modalities and media.

The focus here will, therefore, be on the bodily gestures made by musicians, the nature of the traces these gestures leave on the environment, and how these traces might be picked up and interpreted by audiences. Throughout, the theoretical underpinnings of my arguments are drawn from perceptual psychology, in particular the work of the 'ecological' psychologist James Gibson (1966, 1979; see Heft 2001 for an attempt to integrate Gibson's work into a broader history of realist and pragmatic psychology and philosophy). This ecological approach has recently received interest from a small number of researchers of music who wish in different ways to integrate action and perception (e.g. Clarke 2005, 2007; Dibben 2001; Dibben & Windsor 2001; Windsor 1996, 2000), although they mainly focus on the perceptual side of this equation, with some attention paid to electroacoustic manipulation.

The Problem of Gesture in Performance

It has become commonplace to talk of gesture in music: that this volume exists at all is testament to the recent popularity of the term. For the purposes of this discussion a number of questions will be addressed, first in relation to the Western art music tradition of performance, then through an intentionally paradoxical turn to the performative nature of gesture in electroacoustic composition.

First, how are musical gestures perceived? Normally, gestures are considered to be visual signals, but in music it has become quite normal to talk about gestures being expressed through sound, sometimes independently of any visual complement. Both visual and auditory perception will therefore be considered. Second, what is a musical gesture made by a performer? Is it any movement they make (whether or not it results in sound), or only certain movements? This question necessitates some analysis of the kinds of movement performers make, and some reflection on conscious and unconscious intentionality. Third, how can gestures be audible? An everyday definition of a gesture is an expressive movement (like a shrug, wink or a hand-wave) that is seen: how are gestures heard, and what does it mean to say we hear a gesture? To address these questions requires attention to the actions performers make, the extent to which they can be perceived through different media, how they are perceived and what constraints limit their perception under different reception conditions.

Action in Music

Musicians engage in all kinds of actions, many of which result directly in the production of sound. The movements of our bodies can interact with objects and air columns to produce a wide variety of sounds. However, many of the movements made by musicians are not strictly necessary for producing sound. It is possible to play a woodwind instrument without making movements of the upper body which would be visible to a distant onlooker: the lungs need to be filled and emptied and the fingers need to operate the mechanism of the keys, but the visible swaying of a body, and resulting movement of a whole instrument is not a direct source of sound. Of course, such movements are made by musicians, and have been studied fairly extensively for the clarinet at least (e.g. Vines et al. 2006) and also for the piano (e.g. Davidson 1993, 1994, 1995). Some musicians are explicitly coached in making such movements, either early in their training (e.g. Dalcroze eurythmics) or when they become involved in opera or musical theatre. Davidson has looked at the way such movements can communicate meaning and expression in performance (Davidson 2001) and has also suggested that pianists have a repertoire of movements that may serve similar ends (Davidson 1995, 2002).

For the purposes of the ensuing discussion, musicians' movements can be categorized in two ways, both of which might be helpful in clarifying the gestural qualities of musical performance. First, actions can be categorized by their relative importance to sound production: they directly make sounds, they indirectly affect the making of sounds, or they supplement the making of sounds. In the first category would fall the movements of a pianist's fingers that are necessary to set the piano's hammers in motion, or the changes in embouchure made by a flautist changing octave, or the movements of a guitarist's hand to create vibrato. Note that although some of these movements do not initiate vibration, but modify it, all have a physical mapping from movement to acoustic consequence. In the second category would come cyclical movements such as foot tapping, head nodding or body sway, where these are phase-locked to the beat or tactus, or related to tempo change, dynamic shaping or the like. Similarly there are the punctual movements that accompany events of structural importance such as those discussed by Davidson (2002). Such movements play no necessary physical role in making the sound itself (indeed performers are often warned away from their excessive use as they can disturb the mechanics of performance) but they certainly accompany aspects of musical sound production in a potentially predictable manner and in many cases affect the sound that is produced, or at least seem to. Whether or not they have a truly causal relationship with aspects of musical performance is a question that will be returned to later in this chapter. Lastly, there are supplementary movements that appear to have no causal relationship with the sound (although this is an un-asked empirical question as yet) but certainly seem to play a huge role in the performance: the raised eyebrow of an opera singer to signal surprise, the closing of a pianist's eyes throughout a delicate passage.

The second manner of categorization is related to the first but does not beg questions of causality. Movements are either correlated to some acoustic parameter or not: such correlations can be determined empirically and do not require any appeal to cognitive psychology or philosophy. One can either measure movements and relate them to acoustic dimensions in more or less direct ways, or one can ask participants in experiments to make judgements about music and sound and see whether these judgements are correlated. What becomes important is whether sound and movement are potentially related, and whether such a relationship is detectable by an observer/listener. Focusing on this second type of categorization suits a more ecological approach to studying the perception of musical expression as it reflects a desire to discover what patterns of acoustic or visual structure determine our perceptions: movements (of objects) very closely specify the sound that results and it is argued that these predictable relationships are perceived according to a kind of ecological psychophysics (e.g. Freed 1990; Warren & Verbrugge 1984; Warren et al. 1987). In particular, temporal patterns of excitation are preserved in acoustic information and can be shown to have perceptual relevance (Warren & Verbrugge 1984). Similarly, the pressure and velocity with which a bow moves across a violin string (and the movements of the arm that holds the bow) has a physically predictable relationship with the resulting timbre, one that we can become sensitive to in a direct, rather than conventional, manner, and one that can be modelled in sound synthesis (e.g. Cadoz et al. 1984). In semiotic terms, such a relationship between what would be called expression and content is indexical, rather than symbolic, or, if one prefers Eco's terminology to that of Peirce, the sign function is 'motivated' rather than relying on a purely arbitrary and systemic basis (Eco 1979: 190–92).

Aural and Visual Perception of Performance

What then do observers perceive when they see and hear performance? The movements of performers create patterns in sound and light that are picked up by the visual and auditory systems of audience members. These patterns more or less clearly specify what the performer is doing. They do so in rather general terms in that certain patterns can only be produced by human organs: for example, the human voice is closely specified through a peculiar relationship between the vibration of the vocal chords and the resonances of the vocal cavities. Greater levels of specificity are achievable, however: the sounds as well as the visible movements of a pianist can specify the force with which the performer is depressing keys, the tempo of the music, even the degree of expressivity (Davidson 1993, 1994). Patterns that specify the magnitude and temporal structure of action reveal an enormous amount about what a performer is *doing* and are often cross-modal in that similar descriptions of structure can be found that specify sources in both visual and acoustic domains. This can be true of the patterns of a bouncing ball and their relationship with the ball's elasticity (Warren et al. 1987), or the patterns of movement that specify expressivity of a performance (Davidson 1993). If one is willing to take a radical ecological approach (i.e. Gibson 1979) then what is perceived is movement of a person (in interaction with an instrument), not sounds and light. Auditory and visual perception are processes of picking up information that specifies events and objects. In the case of music, the objects are people and instruments, the events are sets of movements that constitute musical performances (whether seen or heard, or both). This rather counter-intuitive view of musical performance may seem overly radical and seems to downplay a more typical focus on sonic descriptions of music. However, if we are to study movements, and that subset of movements we call gestures, then focusing on movement and how it is jointly perceived through sound and light rather than any metaphorically gestural quality of sound seems a sensible step. This is not to say that the 'structure as such' (Gibson 1966: 225) of sound is not of interest to psychologists or musicologists. However, if we wish to focus on actual rather than metaphorical gestures then sound plays the same role as light: it can tell us what movement another person makes and whether it was a gesture, or an accidental, insignificant or uncommunicative movement.

Other researchers may wish to focus on more metaphorical approaches to musical gestures, but here an attempt will be made to analyse visual and acoustic information and reveal how it specifies the kinds of movements and patterns of movements that are gestural in quality, whether more continuous in nature or more punctual. What do musicians do that is like the flowing continuous hand gestures that accompany speech and how do we perceive these gestures? What do they do that acts like a shrug, or a wink? Clearly the gestures that 'accompany' music may not play quite the same role as the non-verbal communicative movements that supplement other forms of communication nor may they be structured in the same way. However, they are potentially a primary manner in which an audience has direct contact with the performer: we cannot see what a performer thinks but we can hear and see them move, and this may provide us with useful information about their conception of the music they are playing, or at least allow us to form an interpretation of what we think this conception might be.

In order to exemplify the role that the perception of underlying movements may have in performance, two areas of musical activity will be considered. First, this chapter will examine expressive timing in music, along with its relationship to body movement. Some observations will be made about what performers do that we might consider gestural, how these gestures might be perceived by audiences and how gestural information can be recovered from the movements of performers for the purposes of analysis. Second, this chapter will turn from a focus on performers' gestures to compositional gestures, in an attempt to integrate two sides of musical production that are too easily portrayed as being at either end of a continuum between the practical and conceptual.

Gesture and Expressive Timing in Performance

A huge number of psychological studies attest to the expressive nature of musical performance, most of which focus at least in part on timing. These studies are largely concerned with the control of timing in performance (e.g. Shaffer 1981, 1984), what patterns of timing can tell us about cognitive representations of musical structure and vice-versa (e.g. Clarke 1988; Palmer 1989), and how timing might encode (or indeed communicate) mood or other interpretative intention (e.g. Juslin 2000). The concept of gesture is rarely used in such work, although possible relationships between patterns of expressive timing and movement have been explored in both theoretical and empirical studies (e.g. Friberg & Sundberg 1999; Shove & Repp 1995; Todd 1995). Of course, expressive timing in music is directly the result of patterns of movement that directly or indirectly create sound. Whether such movements are gestural in nature, and how they might be analysed if they are, will be explored below.

Most researchers on expressive timing agree that many patterns of rubato, durational accenting, fermata, voice asynchrony and articulation seem to be generated in a rule-like and predictable manner from an interpretation of musical structures (in particular metrical and phrase structure). One way of conceptualizing these rules is in terms of gesture: we move in a particular way to attempt to communicate something. The movements determined by the canonic demands of the score are supplemented by additional movements that may enhance the projection of certain musical elements, or even communicate elements that are to an extent independent of the canonic notation. Of particular interest might be the way in which such movements (whether visible or audible) work in parallel given the multi-modal nature of musical expression, and the notions of parallel, nested and hierarchical gestures will be addressed below through attention to empirical, computational and theoretical findings and methodological issues. First, however, a consideration of some definitions of and assumptions about gesture in performance and its relationship to movement are in order.

Gestures as Expressive Action in Performance

Outside of musicology, and in current definitions, gestures are considered to be body movements that are employed expressively to convey either thought or feeling. However, earlier usages are either broader than this (all bodily movements) or narrower (religious or oratorical postures or movements) and even include the historic usage 'to walk proudly' in verb form. Gesture has also become figuratively employed as will be familiar to all readers of this book: indeed the *Oxford English Dictionary* cites the compositional 'gestures' of new music as an example of this figurative usage (*Oxford English Dictionary*, n.d.). If one eschews figurative definitions then we might indeed come to the conclusion that gestures in musical performance are all expressive actions (or manners of inaction), whether visible or audible. Movements 'expressive of thought or feeling' might be thought to comprise all movements made by performers, although one might wish to distinguish between such expressive movements as called for by the score and those merely implied by it, or added by the performer. Such movements may be discrete, applying to a single time point in a musical performance, or across multiple time points. Just as the hand may be used to signal *stop* when the palm is displayed motionless, the same hand may describe a flowing gesture intended to signify rising emotional intensity. The movements of a performer might indicate a single event in a performance with a discrete movement such as depressing a key on the piano forcefully, or might gesture continuously when a series of key depressions vary in force.

In music, just as in other domains, gestures can occur in parallel. I might in response to a question both shake my head and shrug, or shake my head and frown. In the first instance I might be perceived to be negative but unsure, in the second, negative and displeased. In musical performance, because some very small movements and series of movements become audible the potential for parallel gestures becomes extremely rich. The combined audibility *and* visibility of gesture in musical performance creates a rich possibility for combining parallel gestures across or within modalities. Parallel gesturing may involve more than one modality: a performer might accompany a slowing of tempo with a waggle of the head, for example. It can also occur within a single modality: for example one might slow down and play more staccato. Within a single modality one might vary more than one expressive parameter, as in the case of getting louder and faster.

Moreover, multiple gestures can occur even where a single expressive parameter is involved: a continuous slowing of tempo can be accompanied by a pause of an agogic accent. To illustrate this, imagine playing a short melodic piece on a hand-cranked musical box. Given the same canonic structure, we can articulate at least two parallel strands of gestural timing, despite the constraints of the instrument. By turning the crank faster or slower, stopping or starting, or modulating the frequency with which the crank turns (following any kind of continuous or stepped function) one can produce a range of performances. Each of these performances is the result of performative actions or gestures that combine both sequentially and simultaneously. The tune might be played with a relatively static and fast tempo (roughly periodic rotation of the hand, high velocity), a slow tempo with much rubato (aperiodic rotation of the hand with continuously modulated velocity, average velocity low), a slow tempo with pauses (periodic rotation interrupted with zero velocity segments), or any combination of the above. The results reflect the additive combination of gestures. Interestingly it may be hard to separate out the contribution of each kind of gesture: a gradual deceleration in combination with a final cessation of movement could be perceived in combination as a stepped deceleration. Such problems will be returned to below when analysis techniques are explored.

Given that gestures may be parallel, it is worth noting that they can both complement and/or contradict one another. For example, a pause and a gradual ritardando can together signal the approach and arrival of a cadential figure, whereas an accelerando followed by a fermata has a very different flavour. Dynamic accents can mark individual onsets, but if they occur too close to the peak of crescendo, can be hard to discern as such (e.g. Clarke & Windsor 2000). Across modalities, movements can be intermodally related more or less strongly: tempo, along with the structure of bars and beats and other temporal markers may be correlated with head, torso or instrument positions and velocities and may contribute to our perception of form and its expression (Clarke & Davidson 1998; Vines et al. 2006; Windsor et al. 2003). For example, in Windsor et al. (2003) the head and shoulder movements of two pianists described a roughly oval continuous movement towards and away from the piano keyboard with a period of one or two bars. In contrast, a third pianist nodded his head at the beginning of each bar, the upper body and head remaining otherwise static.

In summary, the actions of performers can be viewed as thoroughly gestural in that they give character to a musical performance. They are available to audiences either through sound or light, or both, can be produced in parallel, can be more or less independent or complementary. In the following section the analysis of such gestures will be investigated, and in particular how one can separate out parallel gestures given only a single parameter, in this case expressive inter-onset timing.

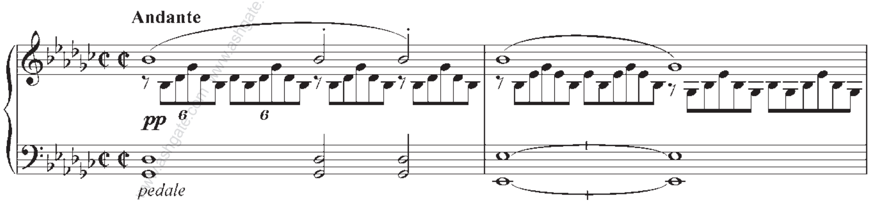
Retrieving Expressive Gestures from Performance

Although thus far this chapter has focused on the actions of performers, the discussion will now turn to the traces such actions leave on the environment. In particular, expressive variability in tempo, dynamics and timbre within performance is information that can be picked up by the listener that can more or less unequivocally specify the gestures that create it. The processes that lead to such direct perception of action will be considered later in this chapter: here the focus will be on the structure of expressive inter-onset timing and how it can be decomposed to reveal the kinds of parallel and multiple patterns discussed in the previous section. It will be assumed that the structured patterns of timing discussed here are the basis for perceiving performative gestures through sound.

Many researchers have worked to extract and measure time-varying properties from acoustic or mechanical signals collected from performances. Such measurements are often represented numerically and/or visually, and attempts are made to model the predictable aspects of such signals (for a review, see Windsor 2008). However, if the processes that create such signals are multiple, it may not always be obvious how to extract such different expressive components. Here, the problems of parallel sources of expressive timing will be exemplified through a re-analysis of some data and model-fitting originally carried out by Windsor and Clarke (1997), and then some solutions will be explored, in particular a decomposition technique developed by Windsor et al. (2006).

Windsor and Clarke (1997) assessed the fit between an algorithmic model of expressive inter-onset timing (and dynamics) developed by Todd (1992) and an expert performance of an extract from a Schubert's Impromptu in G♭ (see Example 2.1). In this study the researchers optimized the fit between model and human performance separately for inter-onset timing and dynamics to minimize error. The best fit achieved by the model for timing was around 43 per cent of the variability observed in the human performance, which was highly significant given the number of inter-onset intervals modelled. However, despite this close relationship between model and performance, the residuals (the remainder of the variance after model fitting) show considerable structure. The model produces roughly polynomial patterns of tempo change, curves that have been observed in many empirical studies of timing in piano performance. After these have been accounted for, the remaining variability in onset timing seems to suggest either additional agogic accents and micro-pauses at the beginnings and ends of phrases in addition to the behaviour predicted by Todd's model, which accelerates continuously towards the middle of phrase then decelerates towards the phrase boundary, although this was only given as a speculative interpretation. Is this a case of parallel sets of gestures being combined additively?

Example 2.1 Schubert, Impromptu in G♭, bars 1–2



Taking the first 24 note onsets (the first bar) of the piece only and applying the same model settings as in the original study gives a fit of around 53 per cent. Figure 2.1 shows the durations of each inter-onset interval (IOI) in the performance (nominally equal events in the score) plotted alongside the output of the model. The qualitative lack of fit here is striking, especially for the initial and terminal events. Leaving aside the small-scale structure, can one fit these two events from the performance better in quantitative terms? A simply linear model that assumes only the first and last events vary in duration from the others, but does not predict by how much, fits the performance data to a much greater extent, without even attempting to match the small-scale detail (see Figure 2.2).

Unfortunately, simply combining these two models to account for both continuous and discrete modulation of timing is neither easy nor terribly logical. The two models are quite different in their conception: one is highly constrained and based on a complex and thorough set of theoretical predications about expressive timing, whilst the other is a thoroughly naïve and under-theorized approximation.

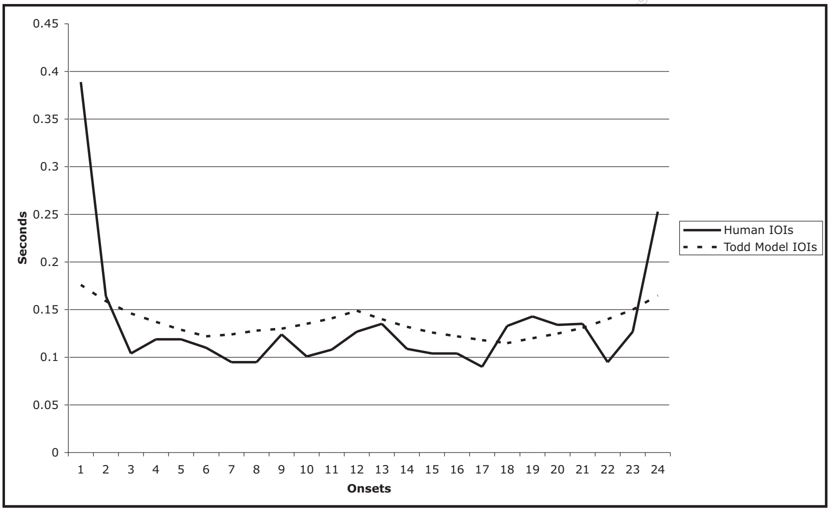


Figure 2.1 An overlay chart plotting the inter-onset timing of the first bar of Schubert's *Impromptu* in G♭ for a human performer and the optimized output of Todd (1992). Model fitting is as in Windsor and Clarke (1997)

The Todd model tests some clear predictions at least, whereas the better-fitting model simply says: the first and last notes in a phrase may be longer or shorter than the others. Moreover, the Todd model does not account for the residuals left behind by the other model, nor vice-versa. Although the simplistic model does capture some aspects of what Todd's model fails to capture, it does not do so comprehensively. This kind of ad-hoc, post-hoc attempt to fit data to a model is unsystematic and lacking in theoretical rigour. What is needed is an approach to expressive timing that *a priori* assumes that expression may have multiple sources that combine in an additive manner. Todd's model does assume that expression is additively combined across multiple hierarchical levels of phrase structure, but each level of timing determines each other level such that expressive timing follows a single underlying equation. Although this is an elegant idea, it is not

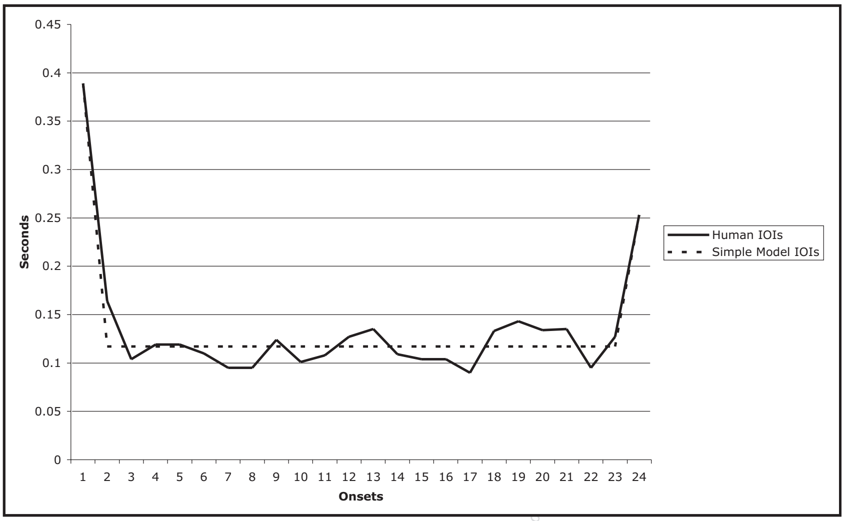


Figure 2.2 An overlay chart plotting the inter-onset timing of the first bar of Schubert's *Impromptu* in G♭ for a human performer (data from Windsor & Clarke 1997) and a simple linear model with an agogic accent at the first event and a final micro-pause

the only approach to combining expressive gestures. There are now a number of algorithmic approaches to combining different aspects of expressive performance, such as *Director Musices* (Friberg et al. 2000) and GERM (Juslin et al. 2002). Here, many rules interact with varying parameters to create a simulated performance that can be evaluated in relation to human performances or judgements. For example, optimizing these parameters to best fit with a set of real performances has been an approach taken by Sundberg et al. (2003) and Zanon and de Poli (2003a, 2003b).

An alternative to such rule-based approaches is to optimize the fit of a performance (or performances) to a model that has few assumptions about what form expressive gestures might take, only that they will be associated in a predictable fashion with structural features in the score. One such model is that of Windsor et al. (2006): here a hierarchical representation of musical structure is used to predict where expressive timing will occur, and a process of optimization finds the individual contributions of each pattern as well as the additively combined result, which should match a target performance closely, given that the musical structure chosen has some close relationship to that expressed through the performance. This process delivers an analysis of expressive timing that breaks down the performed timing into separate repeated gestures associated with different levels and types of structure: a single event may be associated with an accent and be lengthened, a phrase may accelerate, another phrase may accelerate then decelerate, and so on. Figure 2.3 shows an example of such an analysis of timing and Figure 2.4 the result of adding together each of these gestural layers alongside the original performance. The analysis reveals how different sources of expression combine, despite their different time-spans, shapes and structural types.

For example, the analysis shows that the largest phrasal division into one unit of 48 quavers, and two of 36, does not produce uniform timing strategy in the manner of models such as Todd (1992). The performer accelerates towards the end of the first phrase, then follows a romantic ebb and flow pattern across the next two phrases. The analysis also manages to capture the potentially distorting contribution of the fermata at the end of the second long phrase: the ritardando here is extremely deep, and in a more conventional analysis tends to conceal the fact that when it is removed the underlying pattern of rubato is actually quite repetitive. In Figure 2.4 it is much harder to see that there is a repeated pattern of timing, and a statistical analysis of the rubato would have similar problems without such decomposition. Note also the way in which two parallel patterns of timing interlock at the finest level of structure, one associated with metrical regularity, one with out-of-phase grouping structure. Finally, observe the way that this kind of analysis allows one to test whether continuous *and* discrete deviations from metronomic performance are working in parallel: the two leaps down from a grace note are associated with a slight delay that combines additively with all the other patterns to produce the combined model shown in Figure 2.4 (interestingly the other grace notes have no such effect on global timing).

Returning to the issue of gesture, how are these patterns of timing gestural? Each acceleration or deceleration, pause or agogic accent is generated by action, and as we will return to in the next section, the perceptual system is attuned to picking up information that specifies such actions. We do not perceive sound just for itself, but as a source of information about the various bodily gestures that create that sound. Whether each of the gestural layers in such an analysis presented above is perceptible, and whether some layers are perceptually (or motorically) distinct or not remains an empirical question, but decomposition of the expressive signal is a clear precursor to such questions.

The captions for Figures 2.3 and 2.4 (that appear overleaf) are as follows:

Figure 2.3 The decomposition of the expressive profiles: the x axis shows score position, the y axis the parameter values in quaver-note IOIs in seconds. The names of the structural units are displayed to the right of each profile. Score annotations as in Figure 2.4. Note that the x axis is warped to align with the musical notation, distorting the canonically regular shape of the profiles (from Windsor et al. 2006: 1188)

Figure 2.4 Observed and predicted IOIs plotted against a score annotated with phrase, bar and non-contiguous segments. L = leap; R = ritardando; C-R = chord-ritardando. Phrase segments are identified by their duration in score time measured in quavers. Note that the x axis is warped to align with the musical notation (from Windsor 2006: 1187)

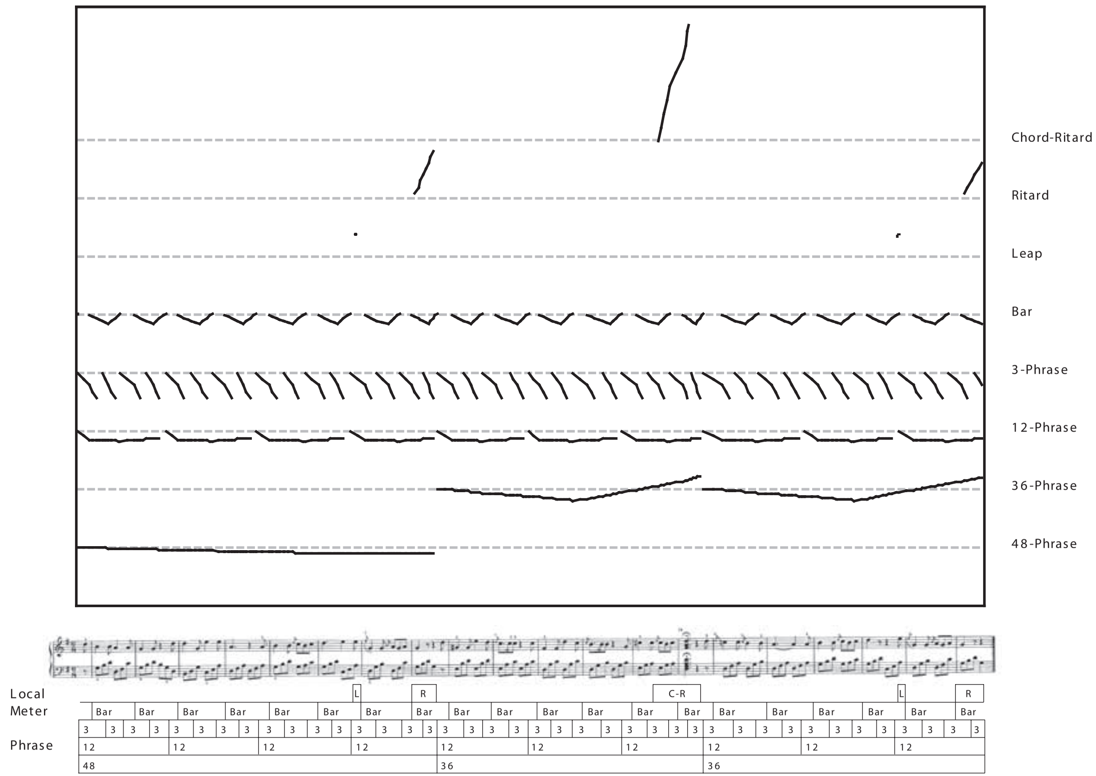

Chord-Ritard
Ritard
Leap
Bar
3-Phrase
12-Phrase
36-Phrase
48-Phrase
Local
Meter
Phrase
L
R
C-R
L
R
Bar
3
12
48
36
36

Figure 2.3

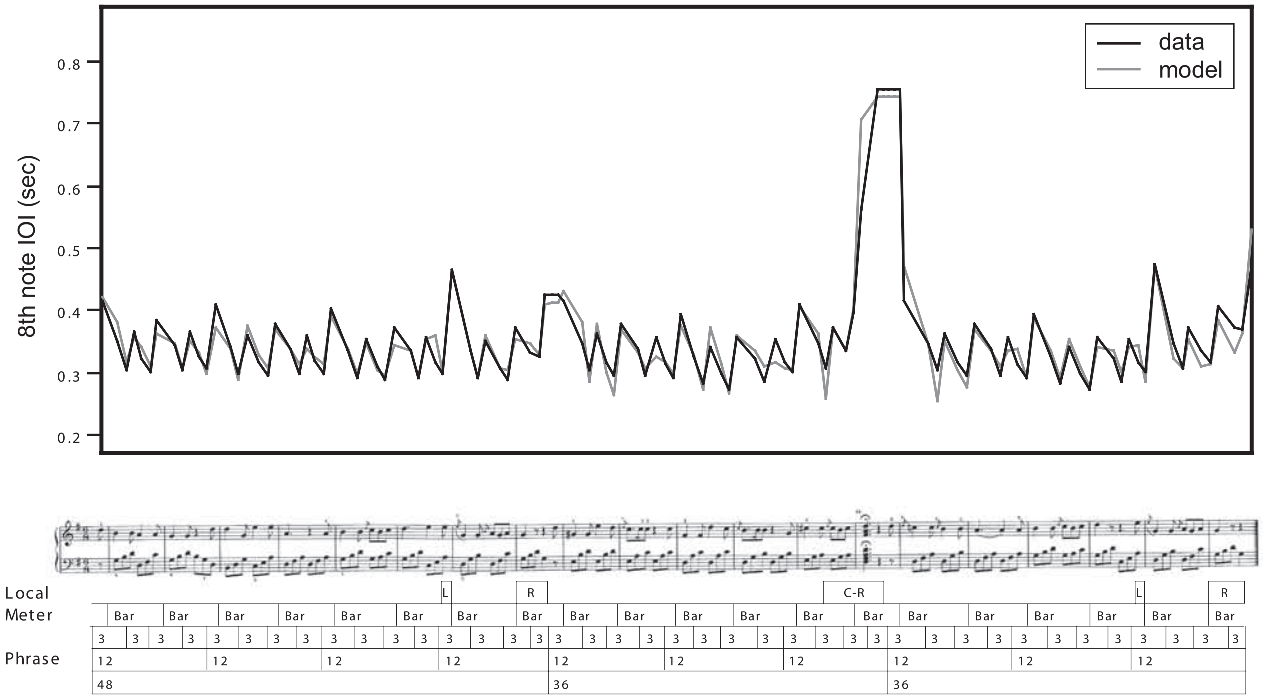

data
model
8th note IOI (sec)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Local
Meter
Phrase
L
R
C-R
Bar
3
12
48
36

Figure 2.4

Gesture in Acousmatic Composition

A performance, as has been demonstrated above, may be made up of a complex combination of movements, which can be considered gestural in that they give character to a piece of music. Acousmatic music, composed for and performed by loudspeakers, and as conceived by Pierre Schaeffer (e.g. Schaeffer 1966), rests upon a theoretical disjunction between sound and source, the idea of reduced listening, a disjunction called into question by many of his successors. Schaeffer's theoretical approach to sound in acousmatic composition seems diametrically opposed to the Gibsonian approach to everyday perception (Gibson 1966, 1979), where sound is simply information for events and objects (see also Gaver 1993). Such a disjunction, if carried over into perception would place acousmatic music in clear contrast to acoustic performances where the relationship between sound and source is visible.

However, given that perceptual processes are peculiarly able to pick up information that specifies the origin of sounds (Gaver 1993) it has been argued that although acousmatic experience is 'impoverished' in relation to everyday perception, we still default to source identification as a primary manner of making sense of what we hear, and that the tension between the ambiguous and incomplete information about sources in acousmatic presentation and our tendency to ascribe them is precisely that which leads to much of acousmatic music's aesthetic richness (Windsor 2000; see also Windsor 2004). In order to form a richer view of gesture from the perspective of action this section will look for compositional gestures (as well as information for inanimate objects and events) in acousmatic music, or the traces of these gestures in the acoustic signal.

Musician versus Listener

Regardless of whether music is presented acousmatically or not, the listener's experience is impoverished in relation to the musician's experience of the music as they are making it. The listener can hear, possibly see (and very possible smell) what a musician is doing, but has no access to the haptic and proprioceptive experiences of the musician. Beyond this, however, the listener can perceive what a musician does to a great extent, although acousmatic presentation certainly widens the gap between making and listening, as much of the information used by a listener to figure out what a musician is doing is undoubtedly visual. This impoverishment of experience, however, is where music becomes interesting. Gibson (1966: 303) argues that in cases where our perceptual field is limited we act to do something about this uncertainty:

> More typical of life than absence of stimulation, however, is the presence of stimulation with inadequate information – information that is conflicting, masked, equivocal, cut short, reduced, or even sometimes false... With conflicting or contradictory information the overall perceptual system alternates

> or compromises... ... but in lifelike situations a search for additional information begins, information that will reinforce one or the other alternative... If detection still fails, the system hunts more widely in space and longer in time.

Arguably, it is in this action that interpretation (as opposed to communication) begins when in a musical context (Windsor 2000). However, the listener is not entirely cut off from the actions of the musician: we can perceive events veridically even in degraded acousmatic, unimodal situations; indeed our ability to do so is well documented (Kendall 1991: 71):

> In everyday life, sound events arise from action, in fact, from the transfer of energy to a sounding object. The auditory system provides us with perceptual characterizations of the energy transfer and of the internal structure of the objects involved. Early in childhood one learns to recognize the occurrence of sound events and to relate them to physical events.

We can perceive the hardness of a mallet from the sound it makes on impact (Freed 1990) and the elasticity of a ball from the timing of its audible bounces (Warren et al. 1987); sonar operators can identify different kinds of ship and propeller from sound alone (Howard & Ballas 1983; Solomon 1958, 1959a, 1959b): such an ecological physics arguably underpins much of our everyday perceptual activity. Hence, it is not exactly a bold claim to suggest that the body's involvement in the production of many musical sounds is specified through predictable relationships between sound and action: such skills are the skills of the expert instrumental teacher who can diagnose deficits in technique through hearing a student. Although acousmatic presentation may make this harder, the traces of human activity are readily perceptible where such lawful relationships between sound and action are not obscured.

The 'Hand' of the Composer

Of course, a primary role of the acousmatic composer may be to obscure the actual origins of sounds, or deceive us about them. The acousmatic composer chooses sounds that specify events and objects more or less closely, sequences and aggregrates of sound objects that more or less resemble environmental 'lawfulness' (e.g. Howard & Ballas 1980): the more closely the real world is alluded to the easier it is for the listener to fall back on direct perception. Conversely, the less closely it is mimicked the more active and potentially imaginative our perception becomes, and paradoxically the 'hand' of the composer becomes audible. In more everyday listening situations we have access to the actions of musicians through the presence of acoustic correlates of action, but in many acousmatic situations the 'hand' of the composer is revealed by the absence of such correlates or the re-ordering or combination of sounds in such a way as to contradict their origins. We become aware of intentionality because the sounds we hear do not sound as if they were simply recorded and re-presented. Of course, many compositions also reveal the actions of their composers through the use of gestural interfaces, their own voices or instruments. What is interesting is that in both cases, the composer is perceivable through the traces he or she has left on the work, in the same way as a performer leaves an audible trail that identifies their contribution. An example where such different kinds of intentionality can be perceived is 'Toccata' from Francois Bayle's *camera oscura* (2000[1976]). This piece combines many allusions to the real, as well as the hyperreal and surreal. Bayle himself 'appears' as a *personnage sonore*: we hear the sounds of a real person interacting with the recorded environment, an environment that seems to sometimes obey everyday rules, and at other times its own internal logic.

To summarize this argument, we tend to hear the bodies of musicians as the primary 'cause' of sounds in instrumental music, whereas in acousmatic music our attention is shifted to more compositional causes. The acousmatic can act to conceal or distort the information we would normally use to identify the gestures made by musicians, but instead the gestures of the composer are highlighted through the traces such gestures leave behind. From an ecological standpoint such traces provide information that specifies the presence of human activity, and actions that communicate or signify character.

Gesture in Context: Action and Perception

In this final section an attempt will be made to integrate the preceding arguments. Gestures, it has been proposed in this context are actions, rather than abstract, figurative or metaphorical entities. Moreover, they are communicative or perceived as such. The central questions underlying this chapter are how musical gestures relate to physical actions by musicians, how such gestures can be recovered in analysis, and how their perception – and attempts to conceal their perception – affect our understanding of intentionality.

In phenomenological terms gestures are perceived through the traces they leave on the environment, whether immediately on their production or preserved over time, as in a recording. The ecological approach to perception adopted here gives a more precise understanding of how such traces betray their origins: patterns of sound and light lawfully specify events and objects including the details of their human origins: the trace of a gesture is the information that specifies a particular action. To conclude this chapter an attempt will be made to characterize the flow of information, perception and action that pertains when a single musician plays on their own, and when that musician is observed and heard by a single listener.

Action, Perception and Information in Performance

Figure 2.5 represents the flow of information that applies when a single musician (performer, improviser or composer) is at work. Actions here result in events that

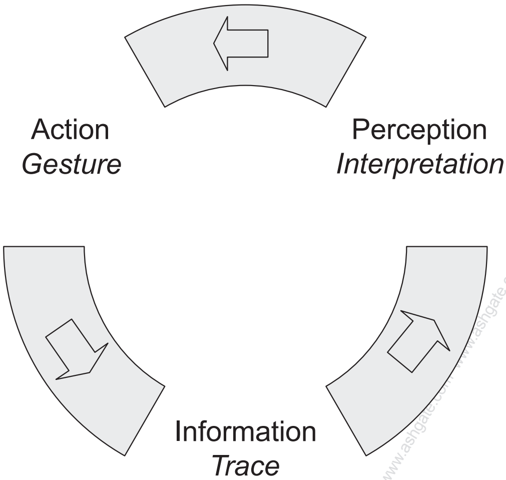


Figure 2.5 Diagram representing the flow of perception and action for a single musician

provide information for perception, which then guide further action. Actions here might be movements that create sound, or incidental movements that accompany sound production: the events they generate, whether sounding or not, are accessible to the musician because they provide information for the musician's perceptual systems to pick up. This information guides further actions, not only though providing feedback on the success of previous actions, but also guiding information further gathering movements.

Figure 2.6 shows what happens when a listener/observer breaks into this loop: the listener/observer acts to perceive by orienting their head towards the auditory and visual information that is being produced by the musician. The movements that are perceived through these sources of information more or less closely specify what the musician is doing, and the gestural character of this activity. What is perceived then generates further action on the part of the listener/observer, either immediate and exploratory, or deferred: one might focus on the hands of a pianist, or attend to a line in a polyphonic texture; or one might seek out an explanation of a detail from the pianist after the performance (or switch one's attention to the programme note). Of course, this situation becomes even more complex and interesting when the musician perceives the actions of listener/observer and modifies his or her actions accordingly.

The shared information about action available to the listener/observer might be thought of as being never as detailed as that available to the musician themselves, but in many ways the listener/observer can observe and listen to the body of the performer in much more detail and freedom, unconstrained by technical limitations and from a distant vantage point. The oddity of seeing and hearing oneself performing on video for the first time bears witness to the privileged viewpoint of the spectator.

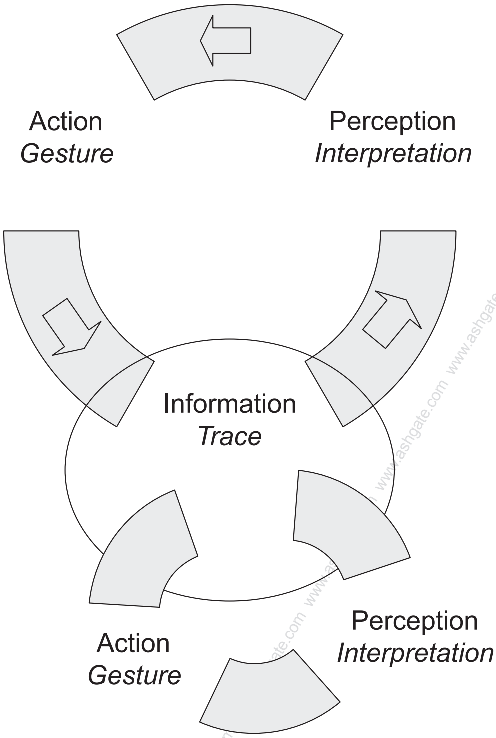


Figure 2.6 Diagram representing the flow of perception and action for a musician and listener/observer

Composers, however physically absent from a performance, still arguably leave a trace of their physical actions in the sounds that are produced, although this is far less specific, leaving much more to the listener/observer's interpretation.

Performance and Composition: Analysing Gesture

What this chapter has endeavoured to show is that it is possible to recover from musical signals traces that betray the gestural origins of sound. However, the empirical basis for many of the claims advanced here is relatively weak. It is one thing to show how a musical signal can be decomposed to reveal a nested hierarchy of temporal trajectories that originate in the gestures of the human body, quite another to detail the extent to which these are perceived. Similarly, although an attempt has been made to show how even in acousmatic composition the human presence of the composer can be detected, it is another thing to show the extent to which this occurs and identify the invariant properties of sound that specify such human causation.

However, the aim here has been to emphasize that any analysis of gesture in music has to consider the real actions of musicians and how these are perceived through the eyes and ears of the audience. Gestures are actions that musicians make, and the supreme virtue of music in this respect is that it can make audible gestures that are near invisible. Moreover, the potential for parallel strands of gesture within a performance also gives music a peculiar efficacy. Much work needs to be done to investigate the manner in which sound and light specify the actions of performers and how these give meaning to musical experience if we are to understand music and its gestural qualities.

References

Bayle, F. (1976/2000). *camera oscura*. Paris: Magison.

Cadoz, C., Florens, J-L. & Luciani, A. (1984). Responsive Input Devices and Sound Synthesis by Simulation of Instrumental Mechanisms: The CORDIS System. *Computer Music Journal* 8/3: 60–73.

Clarke, E. F. (1988). Generative Principles in Music Performance. In J. A. Sloboda (ed.), *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition* (pp. 1–26). Oxford: Clarendon Press.

— (2005). *Ways of Listening. An Ecological Approach to the Perception of Musical Meaning*. New York: Oxford University Press.

— (2007). The Impact of Recording on Listening. *Twentieth Century Music* 4/1: 47–70.

— & Davidson, J. W. (1998). The Body in Performance. In W. Thomas (ed.), *Composition, Performance, Reception: Studies in the Creative Process in Music* (pp. 74–92). Aldershot: Ashgate.

— & Windsor, W. L. (2000). Real and Simulated Expression: A Listening Study. *Music Perception* 17/3: 1–37.

Davidson, J. W. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. *Psychology of Music* 21: 103–13.

— (1994). What Type of Information is Conveyed in the Body Movements of Solo Musician Performers? *Journal of Human Movement Studies* 6: 279–301.

— (1995). What does the Visual Information Contained in Music Performances Offer the Observer? Some Preliminary Thoughts. In R. Steinberg (ed.), *The Music Machine: Psychophysiology and Psychopathology of the Sense of Music* (pp. 105–13). New York: Springer Verlag.

— (2001). The Role of the Body in the Production and Perception of Solo Vocal Performance: A Case Study of Annie Lennox. *Musicae Scientiae* 5/2: 235–56.

— (2002). Understanding the Expressive Movements of a Solo Pianist. *Musikpsychologie* 16: 9–31.

Dibben, N. (2001). What Do We Hear When We Hear Music? Music Perception and Musical Material. *Musicae Scientiae* 5/2: 161–94.

— & Windsor, W. L. (2001). Constructivism in Nicholas Cook's Introduction to Music: Tips for a 'New' Psychology of Music. *Musicae Scientiae* 2: 43–50.

Eco, U. (1979). *A Theory of Semiotics*. Bloomington: Indiana University Press.

Freed, D. J. (1990). Auditory Correlates of Perceived Mallet Hardness for a Set of Recorded Percussive Sound Events. *Journal of the Acoustical Society of America* 87: 311–22.

Friberg, A., Colombo, V., Frydén, L. & Sundberg, J. (2000). Generating Musical Performances with Director Musices. *Computer Music Journal* 24/3: 23–9.

— & Sundberg, J. (1999). Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners. *Journal of the Acoustical Society of America* 105/3: 1469–84.

Gaver, W. W. (1993). What in the World Do We Hear? An Ecological Approach to Auditory Event Perception. *Ecological Psychology* 5/1: 1–29.

Gesture. *Oxford English Dictionary Online* (http://dictionary.oed.com; accessed August 2008).

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. London: Unwin Bros.

— (1979). *The Ecological Approach to Visual Perception*. New Jersey: Lawrence Erlbaum.

Heft, H. (2001). *Ecological Psychology in Context: James Gibson, Roger Barker and the Legacy of William James's Radical Empiricism*. Mahwah, NJ: Lawrence Erlbaum.

Howard, J. H. & Ballas J. A. (1980). Syntactic and Semantic Factors in the Classification of Nonspeech Transient Patterns. *Perception and Psychophysics* 28: 431–9.

— & Ballas, J. A. (1983). Perception of Simulated Propeller Cavitation. *Human Factors* 25/6: 643–55.

Juslin, P. N. (2000). Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception. *Journal of Experimental Psychology: Human Perception and Performance* 26: 1797–813.

—, Friberg, A. & Bresin, R. (2002). Toward a Computational Model of Expression in Performance: The GERM model. *Musicae Scientiae*, Special Issue (2001–02): 63–122.

Kendall, G. S. (1991). Visualisation by Ear: Auditory Imagery for Scientific Visualisation and Virtual Reality. *Computer Music Journal* 15/4: 70–73.

Nattiez, J-J. (1990). *Music and Discourse: Toward a Semiology of Music* (trans. C. Abbate). Princeton: Princeton University Press.

Palmer, C. (1989). Mapping Musical Thought to Musical Performance. *Journal of Experimental Psychology: Human Perception and Performance* 15/12: 331–46.

Schaeffer, P. (1966). *Traité des Objets Musicaux*. Paris: Seuil.

Shaffer, L. H. (1981). Performances of Chopin, Bach and Bartok: Studies in Motor Programming. *Cognitive Psychology* 13: 326–76.

— (1984). Timing in Solo and Duet Piano Performances. *The Quarterly Journal of Experimental Psychology* 36A: 577–95.

Shove, P. & Repp, B. H. (1995). Musical Motion and Performance: Theoretical and Empirical Perspectives. In J. Rink (ed.), *The Practice of Performance: Studies in Musical Interpretation* (pp. 55–83). Cambridge: Cambridge University Press.

Solomon, L. N. (1958). Semantic Approach to the Perception of Complex Sounds. *Journal of the Acoustical Society of America* 30/3: 421–5.

— (1959a). Search for Physical Correlates to Psychological Dimensions of Sounds. *Journal of the Acoustical Society of America* 31/4: 492–7.

— (1959b). Semantic Reactions to Systematically Varied Sounds. *Journal of the Acoustical Society of America* 31/7: 986–90.

Sundberg, J., Friberg, A. & Bresin, R. (2003). Attempts to Reproduce a Pianist's Expressive Timing with Director Musices Performance Rules. *Journal of New Music Research* 32/3: 317–25.

Todd, N. P. (1992). The Dynamics of Dynamics: A Model of Musical Expression. *Journal of the Acoustical Society of America* 91: 3540–50.

— (1995). The Kinematics of Musical Expression. *Journal of the Acoustical Society of America* 97: 1940–49.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M. & Levitin, D. J. (2006). Cross-Modal Interactions in the Perception of Musical Performance. *Cognition* 101/1: 80–103.

Warren, W. H. & Verbrugge, R. R. (1984). Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics. *Journal of Experimental Psychology: Human Perception and Performance* 10/5: 704–12.

—, Kim, E. K. & Husney, R. (1987). The Way the Ball Bounces: Visual and Auditory Perception of Elasticity and the Bounce Pass. *Perception* 16: 309–36.

Windsor, W. L. (1996). Perception and Signification in Electroacoustic Music. In R. Monelle & C. Gray (eds), *Song and Signification: Studies in Music Semiotics* (pp. 64–74). Edinburgh: Edinburgh University Faculty of Music.

— (2000). Through and Around the Acousmatic: The Interpretation of Electroacoustic Sounds. In S. Emmerson (ed.), *Music, Electronic Media and Culture* (pp. 7–33). Aldershot: Ashgate.

— (2004). An Ecological Approach to Semiotics. *Journal for the Theory of Social Behaviour* 34/2: 179–98.

— (2008). Measurement and Models of Performance. In S. Hallam, I. Cross & M. Thaut (eds), *Oxford Handbook of Music Psychology* (pp. 323–31). Oxford: Oxford University Press.

— & Clarke, E. F. (1997). Expressive Timing and Dynamics in Real and Artificial Musical Performances: Using an Algorithm as an Analytical Tool. *Music Perception* 15/2: 127–52.

—, Ng, K., Davidson, J. W. & Utley, A. (2003). Investigating Musicians' Natural Upper Body Movements. Paper presented at the First International Conference on Music and Gesture, University of East Anglia, UK, 28–31 August.

—, Desain, P. W. M., Penel, A. & Borkent, M. (2006). A Structurally Guided Method for the Decomposition of Expression in Music Performance. *Journal of the Acoustical Society of America* 119/2: 1182–93.

Zanon, P. & de Poli, G. (2003a). Time-varying Estimation of Parameters in Rule Systems for Music Performance. *Journal of New Music Research* 32/3: 295–316.

— (2003b). Estimation of Parameters in Rule Systems for Expressive Rendering in Musical Performance. *Computer Music Journal* 27/1: 29–46.