Chapter 12

A grand gesture: Vocal and corporeal control in melody, rhythm, and emotion

Iain Morley

Fitch's paper represents a timely enquiry into the rhythmic dimension of music origins, an aspect which has to date received far less attention than the tonal elements of musical behaviours. In particular, investigations into the relationships between musical and linguistic capabilities have often focused most on those elements which share the trait of vocal production—meaning, in the case of music, the production and perception of pitch and contour, and the structures within them. The research cited by Fitch in this paper and his own related ideas represent welcome additions to such investigations. It is increasingly evident that the rhythmic, gestural, expressive, and tonal production and perception systems are intimately related and interdependent, and any future considerations of music origins that do not consider all of these dimensions as integral to the emergence of the capacity will be incomplete.

There is far more of importance in Fitch's paper than it is possible to comment on here, so I must be selective and draw out just a few aspects to which I can immediately respond and which, in my opinion, raise particular queries for further elaboration. I then will respond in detail to one key point of interest raised in Fitch's paper—the relationship(s) between rhythm, melody, and corporeal motor control.

Fitch's division of 'rhythmic capability' into three components and discussion of the evidence for their presence in other higher primates is important, and the suggestion that it is the capacity for entrainment that is particular to humans is very interesting, and points to a key direction for future research. Fitch makes a distinction between **beat entrainment** (e.g. tapping one's foot to a beat) and **pattern generation** (*complex* motor output such as dancing or instrumental playing), with pattern generation relying on complex cross-modal integration, with information from the auditory cortex having to be successfully propagated to the motor cortex. Whilst there is clearly a distinction in complexity, surely this is a quantitative rather than qualitative difference between 'beat entrainment' and 'pattern generation', i.e. a difference in the number of different modes over which the entrainment is occurring?

When it comes to the separation of entrainment into two component activities, a 'synchronization component' and a 'pattern output component', it would seem to be worth also exploring the extent to which these components would need to have a two-way relationship. In other words the motor output must surely provide proprioceptive and auditory cues which then form both a part of the acoustic signal and a cue by which to tailor the relationship between the exterior stimulus, the listener's internal pulse and their motor pulse.

Fitch makes the important point that whilst the forms of the pattern output can be greatly divergent, the actual requirement of entraining one's motor control to an extracted pulse is equivalent whether for the purposes of playing another rhythm-producing instrument, clapping, or dancing (and the human body is in both latter cases itself a rhythm-producing instrument).

I wonder, however, whether he over-stresses the distinction between instrumental playing and dance, in the discussion of the generation of complex motor output. He says that 'extraction of a beat, and entrainment to that beat in a pure motor fashion (for example, in dancing) does not assure the ability to *musically* entrain, for one can be a dancer (motor → motor) but not an instrumentalist (motor → acoustic)'.

He highlights the necessity in instrumental playing of paying careful attention to the instrument's acoustic output and of '. . . closing the motor/auditory loop . . . adding an additional level of complexity over the (already challenging) dancing problem'. But surely dancing also involves some auditory feedback in moderating the motor act? Dancing must involve profound proprioceptive and other sensory feedback in moderating and maintaining it, as does instrumental activity. Admittedly this is not tonal but it is multi-modal nevertheless and may not be demonstrably less complex. Perhaps a distinction could be drawn between the use/role of the auditory loop in the two circumstances, rather than the implication that it is not used at all in the dance act.

Fitch makes the point that we can separate the perceptual and motor aspects of the process of 'rhythmic ability' because we can 'hear the beat, internally, without manifesting it via any body movements of one's own'. It will be very interesting to see in future research into the possible roles of mirror neuron activity in rhythmic and musical perception whether one actually *can* genuinely 'hear the beat internally' without the brain generating neurological activity which is equivalent to manifesting it via body movements. In a sense, 'hearing the beat internally' would thus rely on inducing the neurological corollary of its manifestation via actual body movements. I wouldn't be in the slightest bit surprised if it transpires that the brain can only extract rhythms to be heard internally through the generation of neurological activity that replicates its physical manifestation, in much the same way as has been evidenced to occur in mirror neurons for various other complex motor skills.

One aspect of auditory processing that is of particular interest in Fitch's paper is the apparent similarity between the structure of rhythmic and harmonic information processing, which 'suggests, perhaps surprisingly, that similar pattern-extraction mechanisms may apply in the domains of rhythm and harmony, despite gross differences in the musical surface analysed, providing support for the idea of broad cognitive/aesthetic principles applicable across musical domains'. He goes on to say that 'the discovery of isomorphisms in the rhythmic and harmonic domains raises fascinating questions about the nature of human perception and aesthetic cognition that cannot be answered at present. . . . Although the isomorphism is a fact, its causal basis remains a fascinating mystery, a mystery whose resolution may yield deep insights into (at least) musical cognition if explored experimentally at the perceptual, cognitive, and neural levels'.

It is possible that the following discussion, of research relating to relationships between vocal, corporeal, and rhythmic control, may shed some light on this isomorphism. Although more concerned with the production than the perception of vocalization and rhythm, it becomes evident that the production and perception are themselves intimately related.

The inter-relationship of vocal control and corporeal control in melody, rhythm, and emotion

Manual movement—and broader *corporeal gesture*—is a fundamental component of much musical behaviour and of vocal behaviour; it is also a key element of expression of emotion. Gesture and vocalization appear to be, in some respects, if not entirely, contiguous and interdependent systems, with neurological associations apparently existing between the fine control of orofacial and laryngeal musculature, and that of manual gesture. Further, investigation of the relationship between manual control and vocalization may help to shed light on the apparent strong association between rhythmic and melodic behaviours in music, and the structural parallel that Fitch highlights.

The involvement, often involuntarily, of manual gestures in vocal expression will be familiar to most people. This situation has led to a body of research examining the interrelationship of these functions (e.g. McNeill, 1992, 2000), and hypotheses that manual gesture and vocal control are functionally linked and share important neurological and/or evolutionary foundations. Some authors have suggested that manual gesture formed the foundation for syntactic elements of language (e.g. Hewes, 1973, 1992; Corballis, 1992; Armstrong, Stokoe, & Wilcox, 1994; Stokoe, 2000), whilst others have carried out experimental observation of interrelationships between manual and vocal control (e.g. Petitto & Marentette, 1991; Locke, Bekken, McMinn-Larson, & Wein, 1995; Feyereisen, 1997; Mayberry & Jaques, 2000; Locke, 2000).

Vocal content and manual gesture

Research by Mayberry and Jaques (2000) regarding gesture in normal speech and stuttered speech casts considerable light on the interrelationship between vocalization and gesture. They found that when disfluency occurs in stuttered speech it is accompanied by fewer gestures than fluent speech—in fact, at the time of stuttering, gesture ceases entirely. This is in contrast to disfluency in normal speech, in which case at a pause gesturing tends to increase along with vocalizations like 'uhm' and 'uhr'. These vocalizations actually maintain the elements of prosody in normal speech disfluencies, and gesture, likewise, is unaffected. In stuttering, on the other hand, the whole of speech prosody and rhythm cease, as does gesture.

So, in normal speech disfluency there is no disruption of the *motor* and *prosodic* elements of speech—these are maintained in ums and ahs, non-lexical vocalizations—the disfluency originates in a lexically-induced pause as one seeks the correct word or phrase, which does not affect gesture. This is not the case in stuttered disfluency. In the words of Mayberry and Jaques (2000, p.208):

> . . . gesture and speech are an integrated system in language production. When speech stumbles and stops as a result of stuttering, the hand always waits for speech so that the

> meanings being expressed by the hand in gesture always coincide with the meanings being expressed by the mouth in speech, even when the gesture must wait for a very long time. Gesture and speech are clearly not autonomous.

Mayberry and Jaques (2000, p. 209) observe that:

> the fact that the temporal concordance between gesture and speech execution is always maintained throughout stuttered and fluent speech suggests that the complex neuromotor patterns of gesture and speech are coordinated and integrated prior to their production in extemporaneous expression.

In fact, the concordance between speech and gesture does not appear to be instigated by the *lexical* components of the speech, but by *cyclical motor control*. Franz, Zelaznik, and Smith (1992) found that when subjects were required to produce repetitive movements with the finger, forearm, and jaw, and the repetition of a syllable, there was significant correlation within subjects of the cycle duration of each task; i.e. each subject reached a 'default' timing cycle for the repetitive muscular action, irrespective of which task was being performed. Franz et al. conclude that common timing processes are involved not only in movements of the limbs, but also in speech and non-speech movements of oral structures, and suggest this indicates a governing cognitive 'muscular timing module' responsible for instigating *all rhythmic cyclical muscular activity*.

This accords well with the findings of Alcock, Passingham, and Vargha-Khadem (2000) that the capacity to perform rhythms, both manually and verbally, forms an important foundation of oral/praxic ability—it seems that the capacity to perform planned sequences of complex muscular movements of rhythmic behaviour, both orally and manually, predate oral praxic abilities. Franz et al. (1992) also found that simultaneously produced finger, arm and oral movements concur whilst carrying out repetitive non-linguistic motor repetition, which Mayberry and Jaques (2000) suggest could be the mechanism by which gesture–speech coexpression occurs. The concordance of gesture and speech control appears to be instigated on a motor-control basis rather than a lexical basis (although the content of an utterance can clearly have some influence on the nature of the gesture). As Mayberry and Jaques observe, the harmonized complex motor patterns of the gesture and speech system must ultimately—in speech, at least—subsequently be integrated with the output of the conceptual linguistic systems.

The question still remains of what cycles in vocal control are being integrated with the gestural cycles. Mayberry and Jaques suggest from their research, and from that of McClave (1994) and Nobe (1996) that it is the prosodic patterns of speech that contain the oscillatory cycles of muscular control. McClave found that it is with the nuclei of tone groups that gesture co-occurs, rather than with syllable stress as was commonly thought, and Mayberry and Jaques' (2000) observations of the onset of gesture and vocalization in normal speakers and stutterers accord with those of McClave. The evidence suggests that the interrelationship between manual gesture and vocalization is a deep one, prelinguistic, and relates complex rhythmic motor production with complexity in tone.

Developmental findings confirm a deep association

From the earliest stage, infants' babbling and gesture seem to be interrelated (Masataka, 2000; Mayberry and Jaques, 2000). Locke et al. (1995) observed a strong association between the onset of babbling behaviour in infants and their exhibiting of lateralization of motor control.

Trevarthen (1999) found that even congenitally blind infants make manual gestures in time with the rhythm of parental vocalizations (which, incidentally, are predominantly tonal with no linguistic meaning to the infant). It seems the perception, as well as production, of vocalization can be linked with gesture. The earliest gestures made by babbling infants are not iconic; they are rhythmic and emotionally determined (Trevarthen, 1999; Falk, 2004a, 2004b) movements that accompany the vocalizations, but do not add meaning or symbolism to them.

The implication of all this research is that the cross-modal coordination of gesture and vocalization doesn't require a central representation or linguistic input initially; such integration occurs whether utterances are linguistic or non-linguistic. This relationship is innate, and relates not only to the production but also the perception of vocalization. It seems that there is a cognitive rhythmic motor coordinator which instigates such muscular sequences irrespective of the musculature (corporeal or orofacial) that is used, and that the complex patterns of gesture (finger, hand, arm, shoulder and joint musculature) and vocalization (orofacial, laryngeal, and respiratory musculature) are coordinated (Mayberry & Jaques, 2000). The cycles of vocalization that are integrated with the gestural cycles are prosodic, tonal ones; in the case of speech, as opposed to non-linguistic vocalization, linguistic meaning and narrative sentence structure are integrated into the gesture–speech system subsequently, before their physical manifestation.

Rhythm, corporeal movement, and emotion

So how does this bear on a relationship between rhythm, melody, and emotion? Whilst rhythm and melody are apparently processed in neurologically specialized areas of the brain which are somewhat independent of each other (for discussion of this see, for example, Morley, 2002, 2003; Peretz & Morais, 1989; Peretz & Zatorre, 2005), there is also clearly some important integration of these systems, especially in the case of the interrelationship of rhythmic physiological movement, vocalization, and emotional state. The expression and perception of emotion clearly happens over multiple media, in particular tonal contour in vocalization (vocal gesture) and posture, gesture, and 'body-language' (corporeal gesture). The evidence discussed above for the integration of vocal control and corporeal control provides several clues as to reasons for the interrelation of vocal-melodic, rhythmic, and corporeal activity; but the relationship between the instigation of vocalization, corporeal muscular control, and emotional expression extends beyond manual gesture accompanying vocalization.

Not only is the production of tonal vocalization related to rhythmic-motor coordination, but also to facial expression and the communication of emotion. Orofacial musculature has a fundamental role in shaping the upper vocal tract and modulating the output of the vocal chords, in all vocalization, linguistic or otherwise. It is also responsible for the control of facial expression. It is thus perhaps unsurprising (though highly significant) that there is a correlation between given facial expressions and characteristics of vocalizations made at the same time. Facial expressions of disgust, happiness, sadness, and anger are not learned and culturally determined, and the capacity to produce and recognize them seems to be innate and universal (e.g. Ekman & Friesen, 1971; Ekman, 1980; Carlson, 1994). In humans, Tartter (1980) showed that vocalizations made whilst smiling were of a higher mean frequency than those made with neutral expression due to the effect of smiling on altering the shape of the upper vocal tract; smiling increases second formant frequency (Tartter & Braun, 1994). Tartter and Braun also showed that vocalizations made whilst frowning had lower formant frequency and longer vowel duration than speech with a neutral expression; listeners were able to discriminate speech made whilst frowning, smiling, and with neutral expression from each other with no visual input. Furthermore, they were able to do so in both vocalized and whispered speech, suggesting that the same effects of facial expression on elements of affect of vocalization would be applicable to vocalizations even in the absence great vocal chord versatility (which accords well with Morton's (1977) findings regarding vocalizations in other animals). Given the universality of certain fundamental facial expressions and the correspondence between these and characteristics of vocalizations, we can also expect characteristics of particular emotional vocalizations to be universal and innate too.

This correlation between facial expression and vocal quality also apparently has an ancient provenance, being shared by our nearest evolutionary relatives. Chimpanzees also frequently couple particular vocalizations with particular emotional facial expressions, the vocalizations being moderated by alterations of the size and shape of the mouth, and thus the resonating upper vocal tract (Falk, 2004a, 2004b). Amongst bonobos, utterances always occur along with facial expressions, gestures, and tactile communication (Bermejo and Omedes, 1999; Falk, 2004a, 2004b), which is true for human infant-directed speech too (Falk, 2004a, 2004b). Humans also use facial affect and vocal affect to inform judgement about the affective content of each other; they seem to be interdependent systems in both production and perception (DeGelder and Vroomen, 2000), with the former having had a significant impact on the development of the nature of affective communication in the latter.

So facial expression and emotional content of vocalization are closely related, with particular orofacial musculature configurations resulting in particular vocal qualities, and the production of certain vocal tones and contours being dependent upon the instigation of certain facial expressions. Furthermore, core emotional facial expressions seem to be human universals (e.g. Ekman and Friesen, 1971; Ekman, 1980).

What is especially interesting is that feedback from facial expression actually *affects* our emotional state. Levenson, Ekman, and Friesen (1990) found that asking subjects to carry out various facial movements to create the expressions of fear, anger, surprise, disgust, happiness, and sadness caused distinct changes to the activity of the autonomic nervous system, such as changes in heart rate, galvanic skin response (skin conductivity due to perspiration), and temperature. The subjects were not informed as to the facial expression being replicated, they were simply told to carry out a sequence of muscular movements which resulted in the formation of a given expression, for example, to raise their eyebrows, then pull them together, whilst tightening the skin under their eyes and stretching their lips horizontally (producing an expression of fear). This particular expression resulted in increased heart rate and reduced skin temperature, whilst a facial expression of anger increased both heart rate and skin temperature, and a happy expression decreased heart rate whilst leaving skin temperature unaffected (Carlson, 1994).

Furthermore, humans have an innate tendency to imitate the facial expression of others with whom they are interacting. Even at the age of 36 hours, infants tend to imitate facial expressions that they see (Field et al., 1982); clearly this is an innate, and not a socially conditioned, tendency, and results in some 'contagion' of the emotion being expressed. Wild, Erb, and Bartels (2001) found that even when exposed to images of another's facial expression (showing sadness, disgust, happiness, anger, surprise, fear, or pleasure) for as little as 500 m/s, subjects reported corresponding changes in emotional state. Carlson (1994, p. 351) suggests that:

> . . . imitation provides one of the channels by which organisms communicate their emotions. For example, if we see someone looking sad, we tend to assume a sad expression ourselves. The feedback from our own expression helps put us in the other person's place and makes us more likely to respond with solace or assistance.

Kraut and Johnston (1979) found that people in situations that were likely to make them smile were more likely to do so in the company of other people than when they were on their own, and the considerable implications for selective advantages of the ability to empathize with conspecifics have also been elaborated (Schmidt and Cohn, 2001; Sloboda and Juslin, 2001).

The implications of all of this for the present discussion are threefold:

First, the production of particular vocal tones normally relies upon adopting certain facial expressions, and certain facial expressions correlate with certain emotional states.

Secondly, the production of a particular facial expression whilst producing a particular vocalization will result in some degree of feedback which actually affects emotional state.

Thirdly, there is a natural inclination to mimic such expressions and to feel such associated emotions; i.e. such physiological-emotional feedback may occur not only during production, but also during *perception* of such a stimulus.

But the non-verbal corporeal communication of affective state is not limited to facial expression, but incorporates the *whole body and posture*; i.e. vocal gesture, facial expression, and corporeal gesture (posture and body language) are all interdependent. The work of authors such as Trevarthen (1999) indicates that these above findings of feedback and 'contagion' of emotion, and their implications, also apply equally to whole-body expressions of affective state. This may go a considerable way towards explaining physical and emotional response to rhythmic and dance stimuli, both proprioceptively and visually.

As Mitchell and Gallaher (2001) observe, music and dance are historically interdependent developments, with many common features: they are both temporally organized, and described in terms of rhythm, tempo, beat, pace, movement, choppiness, or fluidity, for example. In fact, one rarely exists without some form of the other. Each of these properties can be observed and experienced across numerous modalities, in vision, audition, and kinaesthesis (the proprioceptive feedback associated with voluntary body-movement).

Particular pieces of music can elicit consistent emotional responses in listeners, and movements that often occur in response to music, which are kinaesthetically experienced, can also be a part of emotional experience.

Body posture and emotional state are strongly inter-related; our posture and movements can express a great deal about our emotional state, intentionally and unintentionally, and others' body-posture and movements thus provide important cues as to their emotional state. As well as being able to observe such cues, we can empathically experience something of their emotional state in mirroring them with our own bodies.

Musicality and rhythmic movement, Trevarthen (1999) points out, involve deliberate control and sequencing of this system. This can result in a self-directed feedback from movement into emotional state and, importantly, feedback and interaction between individuals, in terms of synchrony of movement and of emotional state.

Trevarthen (1999) has explored a functional system comprising the brain stem, basal ganglia and the limbic structures of the emotional motor system, which he calls the intrinsic motor formation (IMF). This is responsible for integrating functions of attention, learning, and physiological expression of emotion, including the synchronization and coordination of sequences of movements; it seems to be the case that *there is a close interrelationship between the emotional-controlling elements of the limbic system and the areas responsible for the co-ordination of motor sequences and posture*. This system is active automatically in all types of interpersonal interaction, but deliberate use of it is made in musical activities.

Other research that points in this direction, with specific reference to music and dance, has also been carried out. Krumhansl and Schenck (1997) found that subjects who listened to a piece of music and different subjects who observed a dance to that piece of music, showed great concordance both within and between the two conditions as to their interpretation of the timing of tension and emotion in each, and the emotion that they experienced at given points. That the dancer interpreted the music in such a way that crescendos in the music corresponded with high movements, high notes with leaps and staccato and legato sequences with matching physical movements perhaps renders that congruence in interpretation between the two less surprising, given the literality, or *iconicity*, of the interpretation; it nevertheless remains interesting that such physical movements are the (apparently) obvious way of physically representing such auditory phenomena, and that interpretation of them as such occurs. The main implication of the findings relevant to the current discussion is that both media equally represent tension, release and particular emotions, underlining the cross-modality of such affective expression and interpretation. The affective content is apparently interpreted equivalently in visual, auditory, and kinaesthetic media.

As with facial expression, this is not restricted to perception, but is also a two-way process. Mitchell and Gallagher (2001) review a considerable body of research which illustrates that 'music prompts kinesthetic (motor) responses in both children and adults that often match some aspect of the music' (Mitchell and Gallaher, 2001, p. 66), and that 'kinaesthetically experienced movements that sometimes occur in response to music can also be a normal part of emotions' (ibid. p. 66/67). This is a phenomenon with which most people are familiar already, of course, which is not to diminish the value of such research, but to add to it. That this coordination of emotional communication and of rhythmic behaviour is innate is evidenced by the aforementioned studies showing that very young infants can accompany emotive-tonal prosodic infant-directed speech, or in other circumstances singing, with rhythmic body movements which coordinate with the rhythm of the vocalization (Trevarthen, 1999). This is an innate capacity, as it can occur even in congenitally blind infants, who cannot possibly be imitating the movement of the parent.

Summary

- Orofacial expression of emotion (orofacial gesture) and corporeal expression of emotion (corporeal gesture/body language) are related and interdependent.
- The production of vocal sounds relies on the creation of orofacial expressions which correlate with emotional facial expressions; these can induce emotional response in both the vocalizer and perceivers, through voluntary or involuntary imitation and kinaesthetic feedback.
- There is an interdependence between complex orofacial vocal control and rhythmic corporeal gesture control—the production of complex sequences of vocal sounds relies on the same rhythmic-motor coordinating system as does complex manual/upper body gesture and the two frequently occur together.
- The production and perception of music (melody, rhythm, and dance) makes use of all of these systems—indeed, it *has* to.

Clearly the above conclusions are derived from a diverse and disparate body of research, and there is plentiful scope for further research specifically relating to the interrelationships above as they manifest themselves in musical production and perception. But it is possible that the connections highlighted go some way to elaborating some of the relationships that have previously been considered intriguing but mysterious.

It is clear that we are in a very exciting time regarding progress in our understanding of the relationships between the apparently great diversity of capabilities that comprise musical activity. It is increasingly evident that whilst diverse, these capabilities are also interdependent across a number of dimensions, and their congruence in musical activities may not be due to a convergence in their use but to genuine contiguity—interdependencies in their evolutionary heritage. Papers such as Fitch's presented in this volume are making important contributions to crystallizing our understanding and formulating how we might progress in furthering our understanding of this fascinating and fundamental human behaviour.

References

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, *75*, 34–46.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1994). Signs of the origins of syntax. *Current Anthropology*, *35*, 349–68.

Bermejo, M., & Omedes, A. (1999). Preliminary vocal repertoire and vocal communication of wild bonobos (*Pan paniscus*) at Lilungo (Democratic Republic of Congo). *Folio Primatologica*, *70*, 328–57.

Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition*, *44*, 197–226.

DeGelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*, 289–311.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–29.

Falk, D. (2004a). Prelinguistic evolution in hominin mothers and babies: For cryin' out loud! *Behavioural and Brain Sciences*, *27*, 461–62.

Falk, D. (2004b). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*, 491–503.

Feyereisen, P. (1997). The competition between gesture and speech production in dual-task paradigms. *Journal of Memory and Language*, *36*, 13–33.

Field, T., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression in neonates. *Science*, *218*, 179–81.

Franz E. A., Zelaznik, H. N., & Smith, A. (1992). Evidence of common timing processes in the control of manual, orofacial, and speech movements. *Journal of Motor Behavior*, *24*, 281–87.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, *14*, 5–24.

Hewes, G. W. (1992). Primate communication and the gestural origin of language. *Current Anthropology*, *33*, 65–84.

Kraut, R. E., & Johnston, R. (1979). Social and emotional messages of smiling: an ethological approach. *Journal of Personality and Social Psychology*, *37*, 1539–53.

Krumhansl C. L., & Schenk, D.L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae*, *1*, 63–85.

Levenson, R. W., Ekman, P., & Friesen, V. W. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*, 363–84.

Locke, J. L., Bekken, K. E., McMinn-Larson, L., & Wein, D. (1995). Emergent control of manual and vocal motor activity in relation to the development of speech. *Brain and Language*, *51*, 498–508.

Locke, J. L. (2000). Movement patterns in spoken language. *Science*, *288*, 449–51.

Masataka, N. (2000). The role of modality and input in the earliest stages of language acquisition: studies of Japanese sign language. In C. Chamberlain, J. Morford, & R. I. Mayberry (Eds.), *Language Acquisition by Eye* (pp. 3–24). Mahwah, NJ: Erlbaum.

Mayberry, R. I., & Jaques, J. (2000). Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In D. McNeill (Ed.), *Language and Gesture* (pp. 199–214). Cambridge: Cambridge University Press.

McClave, E. (1994). Gestural beats: the rhythm hypothesis. *Journal of Psycholinguistic Research*, *23*, 45–66.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. London: University of Chicago Press.

McNeill, D. (2000). *Language and Gesture*. Cambridge: Cambridge University Press.

Michell, R. W., & Gallaher, M. C. (2001). Embodying music: matching music and dance in memory. *Music Perception*, *19*, 65–85.

Morley, I. (2002). Evolution of the physiological and neurological capacities for music. *Cambridge Archaeological Journal*, *12*, 195–216.

Morley, I. (2003). *The Evolutionary Origins and Archaeology of Music: An Investigation into the Prehistory of the Human Musical Capacities and Behaviours*. Ph.D. thesis, Cambridge University, Cambridge. Available online at http://www.dar.cam.ac.uk/dcrr.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, *111*, 855–69.

Nobe, S. (1996). *Cognitive rhythms, gestures, and acoustic aspects of speech*. Ph.D. thesis, University of Chicago. [Cited in Mayberry and Jaques, 2000.]

Peretz, I., & Morais, J. (1989). Music and modularity. *Contemporary Music Review*, *4*, 279–93.

Peretz, I., & Zatorre, R. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89–114.

Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode–evidence for the ontogeny of language. *Science*, *251*, 1493–96.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: evolutionary questions in facial expression research. *Yearbook of Physical Anthropology*, *44*, 3–24.

Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P. N. Juslin, & J. A. Sloboda (Eds.), *Music and Emotion: Theory and Research* (pp. 71–104). Oxford: Oxford University Press.

Stokoe, W. C. (2000). Gesture to sign (language). In D. McNeill (Ed.), *Language and Gesture* (pp. 388–99). Cambridge: Cambridge University Press.

Tartter, V. C. (1980). Happy talk: perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics*, *27*, 24–27.

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America*, *96*, 2101–07.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae Scientiae*, *Special Issue*, 155–215.

Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research*, *102*, 109–24.