journal of interdisciplinary music studies
spring/fall 2008, volume 2, issue 1&2, art. #0821211, pp. 173-186

Exploring Relationships between the Kinematics of a Singer's Body Movement and the Quality of Their Voice

Geoff Luck and Petri Toiviainen

University of Jyväskylä, Finland

Background in music psychology. Physical movement plays an important role in musical perception and production. It is generally agreed among professional singers and vocal teachers, for example, that there are relationships between the kinematics of a singer's body and the quality of their voice. Thus, we might expect to find relationships between quantifiable indicators of a singer's vocal performance and quantifiable features of their movements while they sing.

Background in computing, mathematics, and statistics. High-resolution motion capture systems have been used in several studies to investigate connections between music and movement (e.g., Palmer & Dalla Bella, 2004; Wanderley, Vines, Middleton, McKay, & Hatch, 2005; Luck & Toiviainen, 2006). The overall orientation of different body parts and the amo"unt of their movement can be estimated from the motion capture data, and these features can be subsequently modeled computationally.

Aims. To synthesize basic research on the singing voice, human movement, and quantification of audio and movement data in an exploration of relationships between bodily posture and singing quality.

Main contribution. Relationships between the spatial arrangement of the limbs and selected audio features of 15 singers' performances of *Tuulantei* by Oskar Merikanto were examined statistically. Results indicated that, while there were individual differences in the relationships observed, features relating to timbre seemed to be frequently associated with the lateral angles of the head and neck. The frontal angles of the upper body, and the frontal angle and rotation of the head, were also important.

Implications. Relationships between the kinematics of a singer's body and their vocal performance have been identified. The present study combines empirical methods of music psychology with sophisticated mathematical, statistical, and signal processing methods to produce formalized knowledge on singing that has application areas in music education.

Keywords: Embodied cognition, motion-capture, singing, computational analysis

•*Correspondence:* Geoff Luck, University of Jyväskylä, Jyväskylä, Finland; e-mail: luck@cc.jyu.fi

Introduction: Music and movement

If one takes the embodied view of human cognition (e.g., Varela, Thompson & Rosch, 1991; Port & van Gelder, 1995), that cognitive processes are governed by an organism's sensorimotor capacities, body, and environment, one can see that musical expression and bodily movement are inextricably connected. There is no music without movement, no musical expression without expressive movement. Similarly, when we hear music, we parse the elements of the music through, for example, body movement, such as foot-tapping or body-sway. At times, our comprehension of the actions responsible for producing music is undetected at a conscious level, but activation of so-called 'mirror neurons' in the brain (e.g., Rizzolatti, Fadiga, Gallese & Fogassi, 1996) reveal its presence nonetheless.

Physical movement plays an important role in musical interaction and communication. In an ensemble, for instance, musicians employ various physical gestures to facilitate synchronization with each other (e.g., Williamon & Davidson, 2002), while, in an orchestra, the conductor's gestures both help maintain synchronization between the musicians, and convey expressive qualities of the music. Movement also plays an important role in the communication of emotions and expressive ideas between musicians.

Research has shown that people can perceive the performance manner of musicians (Davidson, 1993) and the emotional characteristics of dancers (e.g., Dittrich, Troscianko, Lea & Morgan, 1996), and that there exist systematic relationships between expressive dance and music (Krumhansl & Schenk, 1997). From a movement-production point of view, research suggests that children are able to express emotional meaning in music through expressive body movement (Boone & Cunningham, 2001) while, from an embodied cognition perspective, the time-course of runners slowing down to a stop has been shown to closely match that of the final *ritardandi* at the end of classical music performances (Friberg & Sundberg, 1999).

However, most of the work on music and corporeality has to date been either theoretical (e.g., Todd, Lee & O'Boyle, 1999; Godøy, 2003; Leman & Camurri, 2006) or application-oriented (e.g., Wanderley & Depalle, 2004; Camurri, Mazzarino & Volpe, 2004), while fewer empirical investigations have been carried out. Moreover, of the empirical work that has been carried out, most is based on the use of video recordings. For example, Williamon (2000) examined the coordination of movements in duo piano performance, while Schmidt (2002) and Seddon (2005) observed the movements of performing jazz musicians. Due to their limited temporal resolution and two-dimensional image, video recordings are not optimal for studying movement. A more accurate and comprehensive investigation requires the use of a motion-capture system, which allows the movement to be captured at a high resolution, and in three-dimensions. High-resolution motion capture systems have been used in several studies to investigate connections between music and movement.

Wanderley, Vines, Middleton, McKay and Hatch (2005), for instance, carried out an exploratory study of clarinettists' ancillary gestures, while Palmer and Dalla Bella (2004) studied the effect of tempo on the amplitude of pianists' finger movements. Eerola, Luck and Toiviainen (2006) investigated toddlers' corporeal synchronization with music, and Luck and Toiviainen (2006) studied conductor-musician interaction.

It is clear, then, that the human body's role in the perception and production of music has attracted a steadily increasing amount of attention by researchers in recent years. Despite this, however, the body's role in vocal production has received rather little attention in the literature, despite the generally accepted view that a singer's voice quality is at least in part affected by their bodily movements and general posture. The aim of this paper is to synthesize basic research on the singing voice, human movement, and quantification of audio and movement data, into an exploratory study of relationships between singer's bodily movements and the quality of their voice.

Quantification of audio and movement

Audio data quantification techniques have undergone considerable development in recent years, and a number of different approaches have emerged. These different approaches are typically based on principles such as signal processing, machine learning, cognitive modeling, and visualization (Downie, 2003). A large number of studies have used such techniques in areas including computational music analysis (e.g., Lartillot, 2004, 2005), automatic classification (e.g., Toiviainen & Eerola, 2006), organization (e.g., Rauber, Pampalk, & Merkl, 2003), and content-based retrieval (Lesaffre et al., 2003), and the present authors have also applied such techniques to the analysis of music therapy improvisations (Luck & Riikkilä et al., 2006; Luck & Toiviainen et al., 2008). Quantification of the singing voice, meanwhile, has been undertaken extensively by Sundberg (see, for example, Sundberg, 1987), and, more recently, in a series of studies by Mitchell, Kenny, and colleagues, focusing on the practice known as "open throat" technique (see, for example, Mitchell & Kenny, 2007).

Movement data quantification techniques have developed in parallel with the audio techniques mentioned above, and frequently utilise high-quality motion-capture data. Motion-capture systems record movement with high temporal and spatial resolution, and provide a three-dimensional (3D) picture of the activity in question. These features, combined with the nature of the output data – precise spatial coordinates of specific bodily locations – make motion-capture recordings particularly amenable to computational analysis. A number of studies have applied such methods to the analysis of performing musicians' movements (e.g., Wanderley et al., 2005) and conductors' gestures (e.g., Luck, 2000; Luck & Nte, 2008; Luck & Sloboda, 2007, 2008).

The movement- and audio-based approaches have been combined in several studies on topics such as expressiveness in audio and movement (Camurri, De Poli, Friberg, Leman, & Volpe, 2005; Camurri, Lagerlöf & Volpe, 2003), children's rhythmic movement to music (Eerola, Luck, & Toiviainen, 2006), and conductor-musician synchronization (Luck & Toiviainen, 2006). There appear, however, to be no studies which have combined the audio and movement approaches in an investigation of singers' vocal production.

The present study

We recorded the movements and vocal performance of singers in an exploratory study of relationships between singers' posture and the quality of their voice. The movement and audio data were subjected to a computational feature-extraction process, and relationships between indicators of voice quality and spatial arrangement of the limbs examined statistically. As regards the types of relationships we expected to find, given that this was the first study of its kind, we made no specific hypotheses other than that we expected some systematic relationships to emerge.

Method I: data collection

Participants

Fifteen singers participated in this study, all of whom were in receipt of singing tuition at the time of data collection. All participants were current music degree students at the University of Jyväskylä or Jyväskylä University of Applied Sciences.

Apparatus and procedure

In order to obtain high-quality audio recordings, data collection took place in a professional recording studio. The audio was recorded with ProTools using a high-quality microphone positioned two meters from the singer. Each singer performed two verses of *Tuulantei* (op. 13, 1899) by Finnish composer Oskar Merikanto, a song they were all familiar with and had sung before. Participants were recorded separately and unaccompanied, and no instructions were given as to how they should stand or move during the session. The total length of each performance was approximately one minute.

Singers' posture and movements were simultaneously recorded with a Qualisys optical motion capture system at 120 fps using eight cameras to track reflective markers attached to key locations on the body. It should be noted that, while no instructions were given as to how participants should stand or move while singing, the use of a motion-capture system in any study cannot help but draw a participant's attention to the movements they make. Thus, it must be acknowledged that the use of a motion-capture system may have potentially impacted upon the data collected.

Method II: feature extraction

Using Matlab, a series of audio and kinematic features were extracted from the data. These were as follows:

Audio features. Four timbre-related features were extracted from the audio data using a one-second sliding window. In order to be consistent with the frame rate of the motion-capture recordings and subsequent kinematic feature-extraction, the sliding window was moved at steps of 1/120$^{\text{th}}$ second.

- Spectral centroid. This feature was calculated according to the formula

  $$c = \frac{\sum_i a_i f_i}{\sum_i a_i}$$

  where a_{i} and f_{i} denote the amplitude and the frequency corresponding to the i'th bin of the amplitude spectrum. Perceptually, spectral centroid corresponds to the degree of brightness of sound.

- Spectral entropy. This feature was calculated according to the formula

  $$h = -\frac{\sum_i a_i \ln a_i}{\ln M}$$

  where M stands for the total number of bins in the amplitude spectrum. Spectral entropy is a measure of degree of noisiness of sound. In particular, high spectral entropy indicates a high degree of noisiness.

- Spectral irregularity. This feature was calculated according to the formula

  $$r = \frac{\sqrt{\sum_i (a_i - a_{i-1})^2}}{A}$$

  This feature measures the jaggedness of the spectrum and has been found to be a perceptually relevant feature (e.g., Barthet, Kronland-Martinet, Ystad, 2006).

- RMS amplitude. This feature was calculated according to the formula

  $$A = \frac{1}{N}\sqrt{\sum_i y_i^2}$$

where y_i denotes the amplitude of the i'th sample and N the number of samples in the window. In perceptual terms, RMS amplitude might be considered as the loudness of the signal.

Kinematic features. Fourteen kinematic features were extracted from the motion-capture data based on the marker positions shown in Figure 1. These were as follows:

- Leg angle (frontal and lateral). To calculate the leg angles, the leg vector was first defined as the vector pointing from the midpoint of the ankle markers to the midpoint of the knee markers. Subsequently, the frontal and lateral leg angles were calculated as the angles between the vertical direction and the projections of the leg vector on the frontal and lateral planes, respectively.

- Knee angle (frontal and lateral). To calculate the knee angles, the thigh vector was defined as the vector pointing from the midpoint of the knee markers to the midpoint of the hip markers. Subsequently, the frontal and lateral knee angles were calculated as the angle between the thigh and leg vectors projected on the frontal and lateral planes, respectively.

- Hip angle (frontal and lateral). To calculate the hip angles, the torso vector was defined as the vector pointing from the midpoint of the hip markers to the midpoint of the shoulder markers. Subsequently, the frontal and lateral hip angles were calculated as the angle between the torso and thigh vectors projected on the frontal and lateral planes, respectively.

- Shoulder angle (frontal and lateral). To calculate the shoulder angles, the neck vector was defined as the vector pointing from the midpoint of the shoulder markers to the midpoint of the four head markers. Subsequently, the frontal and lateral shoulder angles were calculated as the angle between the neck and torso vectors projected on the frontal and lateral planes, respectively.

- Head angle (frontal and lateral). The frontal head angle was defined as the angle between the transverse plane and the projection onto the frontal plane of the vector pointing from the midpoint of the right-side head markers to the midpoint of the left-side head markers. Similarly, the lateral head angle was defined as the angle between the transverse plane and the projection onto the lateral plane of the vector pointing from the midpoint of the back head markers to the midpoint of the front head markers.

- Knee rotation. This feature was defined as the angle between the projections onto the transverse plane of the vector pointing from the right knee marker to the left knee marker, and the vector pointing from the right ankle marker to the left ankle marker.

- Hip rotation. This feature was defined as the angle between the projections onto the transverse plane of the vector pointing from the right hip marker to the left hip marker, and the vector pointing from the right knee marker to the left knee marker.

- Shoulder rotation. This feature was defined as the angle between the projections onto the transverse plane of the vector pointing from the right shoulder marker to the left shoulder marker, and the vector pointing from the right hip marker to the left hip marker.

- Head rotation. This feature was defined as the angle between the projections onto the transverse plane of the vector pointing from the midpoint of the right head markers to the midpoint of the left head markers, and the vector pointing from the right shoulder marker to the left shoulder marker.

For reasons of body symmetry, absolute values for all lateral and rotation angles were used in subsequent statistical analyses.

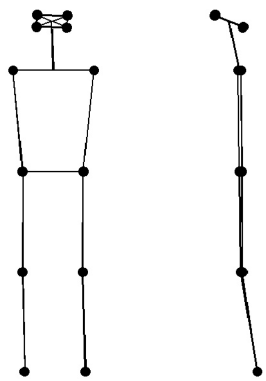

Figure 1. Positions of the markers that were used in the analysis, frontal view on the left, lateral view on the right.

Results

Relationships between the kinematic features and the audio features were investigated using ordinary least squares regression. Initially, all participants were analysed together. Four separate regression analyses were carried out, in each of which the 14 kinematic features were entered simultaneously as predictors of one of the audio features. However, this series of analyses yielded no significant results. Thus, when all participants were analyzed together, no consistent pattern of relationships between the kinematic and audio features emerged. Consequently, each participant was analyzed separately.

A second series of linear regression analyses were thus carried out, four analyses for each participant. In each analysis, the 14 kinematic features were entered simultaneously as predictors of one of the audio features. This series of analyses revealed some clearer patterns in the data for two of the audio features: spectral irregularity and RMS amplitude. All models for these two features were statistically significant, and the amount of variance they explained ranged from 13% to 38% for spectral irregularity, and from 14% to 36% for RMS amplitude. The results of all 30 analyses for spectral irregularity and RMS amplitude are summarised in Tables 1 and 2, respectively.

It can be seen that, for most participants, features related to the shoulders and head were most strongly related to these two audio features (as shown by the highest beta coefficients). For spectral irregularity, there was a generally positive relationship with lateral shoulder angle, and a negative relationship with lateral head angle. For RMS amplitude, however, this pattern was reversed. In practical terms, this means that tilting the neck backwards from the shoulders was more associated with an increase in spectral irregularity, while tilting it forward was more related to an increase in RMS amplitude. Meanwhile, angling the head downwards was more associated with an increase in spectral irregularity, while angling the head upwards was more related to an increase in RMS amplitude.

However, it is clear that these are trends in the data, and that the relationships between the audio and kinematic features were complex. For some participants, for example, features related to the lower limbs were most strongly related to the audio features. For other participants, there were several kinematic features from different parts of the body that were strongly related to the audio features, with no clear ‘winner’. Finally, it can be seen that, of the kinematic features extracted, it was primarily angular as opposed to rotational features that were important; head rotation was the only rotational feature with the highest beta value, and this occurred for only one participant in relation to spectral irregularity.

Table 1. Regression models for spectral irregularity for each of the 15 participants (*P1-P15*). The table shows beta values, which indicate the strength and direction of the relationship between each postural feature and spectral irregularity, as well as F ratios, *df*, and R squared values for each model. Each participant's highest Beta coefficient is in **bold** text.

	P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*	*P11*	*P12*	*P13*	*P14*	*P15*
Leg angle front.	-.145**	-.077**	-.099**	.104**	.110**	.408**	-.028*	-.094**	-.068**	-.071**	-.198**	-.298**	-.254**	-.128**	.012†
Leg angle lat.	-.165**	.218**	.107**	**.508****	.585**	.377**	.007†	-.377**	.348**	.077**	.145**	.150**	.462**	.335**	.149**
Knee angle front.	.203**	.211**	.096**	.143**	-.077**	.057**	-.046**	.004†	-.084**	.274**	.011†	.251**	.296**	.118**	-.040**
Knee angle lat.	.030†	.350**	.124**	.255**	.565**	.293**	-.254**	-.113**	**.692****	.206**	.269**	.303**	**.599****	.322**	.072**
Hip angle front.	-.081**	-.140**	.071**	.102**	-.017†	.251**	-.081**	-.020†	-.020†	.002†	**.276****	.103**	.166**	.032*	-.149**
Hip angle lat.	-.116**	.276**	.251**	.292**	**.625****	**-.517****	-.060†	-.007†	.176**	.008†	-.113**	.068**	.390**	.088**	-.139**
Shoulder angle front.	.121**	.070**	-.117**	.067**	.162**	.055*	-.329**	.008†	-.010†	.116**	.118**	.130**	-.014†	.157**	.196**
Shoulder angle lat.	.014†	.343**	.564**	.148**	.256**	.023†	.421**	**.443****	.063*	**.588****	.242**	**.405****	.240**	-.144**	**-.223****
Head angle front.	-.155**	.032†	-.036**	.018†	.038*	-.214**	.385**	.162**	-.102**	-.063**	.002†	.086**	.132**	**-.405****	-.078**
Head angle lat.	-.304**	**-.702****	**-.886****	-.206**	-.147**	.129**	**-.645****	-.324**	-.160**	-.395**	.226**	-.207**	-.482**	-.156**	.015†
Knee rotation	.060**	-.094**	.170**	.136**	.170**	-.120**	-.077**	.085**	.085**	-.054**	-.099**	.028†	-.114**	.104**	.143**
Hip rotation	-.078**	-.148**	-.093**	-.248**	-.008†	-.132**	-.012†	-.087**	.241**	-.244*	.042**	-.049**	-.005†	.219**	.137**
Shoulder rotation	-.051**	.089**	-.017†	-.057**	.097**	-.104**	.095**	.061**	-.050**	-.067**	-.036**	-.117**	.023†	.072**	-.027*
Head rotation	**.354****	-.087**	-.119**	-.045**	-.073**	.035*	-.035**	.043**	.132**	-.082**	.164**	.184**	.015†	.105**	.070**
F ratio	145.23**	163.55**	125.06**	79.35**	155.51**	272.91**	223.01**	70.05**	141.95**	103.37**	292.85**	81.35**	128.12**	179.64**	98.44**
df	14, 5801	14, 6400	14, 6986	14, 5068	14, 5737	14, 6373	14, 5807	14, 6691	14, 6829	14, 8161	14, 6885	14, 6442	14, 7782	14, 5503	14, 7967
R squared	.26	.26	.20	.18	.28	.38	.35	.13	.23	.15	.37	.15	.19	.31	.15

†not significant. $^{*}p < .05$. $^{**}p < .01$.

Table 2. Regression models for RMS amplitude for each of the 15 participants (*P1-P15*). The table shows beta values, which indicate the strength and direction of the relationship between each postural feature and RMS Amplitude, as well as *F* ratios, *df*, and *R* squared values for each model. Each participant's highest Beta coefficient is in **bold** text.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
Leg angle front.	.150**	.022†	-.022†	.104**	-.061**	-.155**	-.001†	.078**	.047**	-.010†	-.003†	.413**	.010†	-.041*	-.511**
Leg angle lat.	.243**	-.172**	.255**	-.193**	-.528**	-.475**	-.402	.213**	-.463	-.563**	-.645**	.196**	-.372**	.331**	.114**
Knee angle front.	-.207**	-.180**	.037*	-.189**	.137**	.033*	-.001†	.132**	.094**	-.244**	.078**	-.381**	-.241**	.014†	.319**
Knee angle lat.	.235**	-.147**	.009†	.169**	-.287**	-.491**	-.248**	.192**	-.647**	-.565**	-.532**	.305**	-.351**	.387**	-.045**
Hip angle front.	.002†	.141**	.034*	-.118**	.033*	.022†	-.068**	.033†	-.007†	.047*	-.216**	-.160**	-.043**	-.059**	.312**
Hip angle lat.	.281**	-.283**	-.112**	-.297**	-.350**	.330**	-.105**	-.003†	-.364**	-.301**	-.371**	.033*	-.102**	.163**	.082**
Shoulder angle front.	.069*	-.154**	.088**	-.130**	-.004†	-.063*	.157**	.020†	.205**	-.042**	-.152**	-.058**	-.010†	-.176**	-.210**
Shoulder angle lat.	.032†	-.604**	.178**	-.535**	-.047†	-.175**	-.526**	-.369**	-.334**	-.583**	-.684**	-.301**	.326**	.042†	.476**
Head angle front.	-.228**	.097**	.046**	.072**	-.105**	-.022†	.029†	-.125**	-.009†	.143**	-.059**	-.013†	.065**	.405**	-.044**
Head angle lat.	.253**	.793**	.093*	.577**	-.001†	.131**	.600**	.155**	.365**	.353**	.067**	.560**	-.179**	.417**	-.115**
Knee rotation	-.195**	-.110**	-.158**	-.219**	-.227**	-.027†	-.142**	-.119**	-.050**	-.117**	.034**	-.223**	.082**	-.231**	.236**
Hip rotation	-.022†	.183**	-.090**	-.136**	.095**	-.022†	.120**	.190**	-.195**	.313**	.091**	.106**	.046**	-.193**	.053**
Shoulder rotation	.081**	.013†	.040**	-.086**	-.155**	-.045**	-.025†	-.052**	-.092**	.080**	.221**	.035**	.037**	-.021†	-.001†
Head rotation	-.056**	.259**	.221**	.218**	.043**	.006†	-.046**	-.074**	-.050**	.007†	-.184**	-.232**	.037**	.150**	.218**
F ratio	189.65**	149.91**	189.10**	193.37**	141.43**	221.63**	156.64**	79.27**	87.48**	177.31**	271.16**	219.42**	116.81**	213.32**	254.90**
df	14, 5801	14, 6400	14, 6986	14, 5068	14, 5737	14, 6373	14, 5807	14, 6691	14, 6829	14, 8161	14, 6885	14, 6442	14, 7782	14, 5503	14, 7967
R squared	.31	.25	.28	.35	.26	.33	.27	.14	.15	.23	.36	.32	.17	.35	.31

†not significant. *$p < .05$. **$p < .01$.

Discussion

This paper offers some preliminary data on relationships between singers' posture and the quality of their voice. A computational analysis of four timbre-related audio features and 14 kinematic features indicated that arrangement of the head and neck had the most profound effect on voice quality, but that there were large individual differences in relationships overall. Spectral irregularity, which, in perceptual terms, might be considered as the noisiness of the signal, tended to increase when singers angled their neck back and tilted their head downwards. Meanwhile, RMS amplitude, which, in perceptual terms, might be thought of as 'loudness', tended to increase when singers angled their neck forwards and tilted their head up.

These findings seem somewhat intuitive. For example, tilting the head downwards may obstruct the vocal apparatus, thus causing more noisiness in the signal. Tilting the head upwards, on the other hand, could have the opposite effect, freeing up the vocal apparatus, and permitting a greater flow of air.

In terms of the regression models, it can be seen that they were moderately successful in explaining relationships between the audio and movement features, but that much of the variance was still left unexplained. Clearly, there is room for improvement in our approach. One obvious development would be to extract a greater range of audio features, not just those related to timbre. The statistical technique employed, linear regression, combined with the large amount of data collected for each singer, could easily accommodate an increase in the number of features analyzed. Likewise, the extraction of alternative movement features might also be explored.

Moreover, future work could investigate temporal relationships between changes in movement and changes in sound quality during a performance. This might offer a more comprehensive picture of how different parts of the human body are employed during vocal production. It might also be interesting to examine performances of non-classical singers, such as rock, pop, folk, or gospel singers to see if relationships between movement and sound production generalise or are genre-specific.

Finally, an investigation of the relationships between movement of the body and structural and expressive elements of the music being performed would enhance our understanding of the body's role in expressive performance. Indeed, the present study has already started down this path since timbre is one feature which can be manipulated by a performer to enhance the expressivity of their performance.

In music educational terms, the identification of relationships between a singer's bodily movements and quality of their vocal performance implies that singing teachers should stress the importance of using the body in an optimal manner in order to produce the best possible vocal performance. However, since the relationships between movement features and voice quality seem to differ between singers, singers should be assessed and advised on an individual basis. More work is needed in this area to better understand the impact of kinematics of the body on vocal production.

Acknowledgments

This research was supported by the Academy of Finland (project number 110576).

References

Barthet, M., Kronland-Martinet, R., & Ystad, S. (2006). Consistency of timbre patterns in expressive music performance. In P. Depalle & V. Verfaille (Eds.), *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)* (pp. 19–25). Montreal: Canada. Available online at: http://www.dafx.ca/proceedings/allinone.pdf

Boone, R. T., & Cunningham, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of Nonverbal Behaviour*, 25(1), 21–41.

Camurri A., Lagerlöf I., Volpe G., (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1–2), 213–225.

Camurri, A., Mazzarino, B., & Volpe, G. (2004). Expressive gestural control of sound and visual output in multimodal interactive systems. In C. Agon & G. Assayag (Eds.), *Proceedings of Sound and Music Computing 2004 (SMC'04)* (pp. 237–244). Paris: France.

Camurri, A., De Poli, G., Friberg, A., Leman, M., & Volpe, G. (2005). The MEGA project: Analysis and synthesis of multisensory expressive gesture in performing art applications. *Journal of New Music Research*, 34(1), 5–21.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21, 103–113.

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic light-point displays represented in dance. *Perception*, 25, 727–738.

Downie, J. S. (2003). Music information retrieval. In B. Cronin (Ed.), *Annual Review of Information Science and Technology 37* (pp. 295–340). Medford, NJ: In-formation Today.

Eerola, T., Luck, G. & Toiviainen, P. (2006). An investigation of pre-schoolers' corporeal synchronization with music. In M. Baroni, A. R. Addessi, R. Caterina & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception & Cognition (ICMPC9)* (pp. 472–476). Bologna, Italy.

Friberg, A., & 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), 1469–1484.

Godøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo*, 36(4), 317–319.

Krumhansl, C. L., & Schenk, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchines's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae*, 1, 63–83.

Lartillot, O. (2004). A musical pattern discovery system founded on a modelling of listening strategies. *Computer Music Journal*, 28(3), 53–67.

_________. (2005). Multi-dimensional motivic pattern extraction founded on adaptive redundancy filtering. *Journal of New Music Research*, 34(4), 375–393.

Leman, M., & Camurri, A. (2006). Understanding musical expressiveness using interactive multimedia platforms. *Musicae Scientiae*, Special issue 2005/06, 209–234.

Lesaffre, M., Tanghe, K., Martens, G., Moelants, D., Leman, M., De Baets, B., De Meyer, H., & Martens, J.-P. (2003). The MAMI query-by-voice experiment: Collecting and annotating vocal queries for music information retrieval. In *Proceedings of the 4th International*

Conference on Music Information Retrieval (ISMIR'03). Baltimore: USA. Available online at http://ismir2003.ismir.net/papers/Lesaffre.PDF

Luck, G. (2000). Synchronizing a motor response with a visual event: the perception of temporal information in a conductor's gestures. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda (Eds.), *Proceedings of the 6th International Conference on Music Perception and Cognition (ICMPC6)*. Keele, UK: Keele University Department of Psychology.

Luck G., & Nte, S. (2008). A new approach to the investigation of conductors' gestures and conductor-musician synchronization, and a first experiment. *Psychology of Music*, 36(1), 81–100.

Luck, G., & Sloboda, J. (2007). Synchronizing with complex biological motion: An investigation of musicians' synchronization with traditional conducting beat patterns. *Music Performance Research*, 1(1), 26–46. Available online at http://www.mpr-online.net/pdfs/MPR0002.pdf

_________. (2008). Exploring the spatio-temporal properties of simple conducting gestures using a synchronization task. *Music Perception*, 25(3), 225–239.

Luck, G., & Toiviainen, P. (2006). Ensemble musicians' synchronization with conductors' gestures: An automated feature-extraction analysis. *Music Perception*, 24(2), 189–200.

Luck, G., Riikkilä, K., Lartillot, O., Erkkilä, J., Toiviainen, P., Mäkelä, A., Pyhäluoto, K., Raine, H., Varkila, L., & Värri, J. (2006). Exploring relationships between level of mental retardation and features of music therapy improvisations: a computational approach. *Nordic Journal of Music Therapy*, 15(1), 30–48.

Luck, G., Toiviainen, P., Erkkilä, J., Lartillot, O., Riikkilä, K., Mäkelä, A., Pyhäluoto, K., Raine, H., Varkila, L., & Värri, J. (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music*, 36(1), 25–46.

Mitchell, H. F. & Kenny, D. T. (2007). Open throat: Acoustic and perceptual support for pedagogical practice. In: K. Maimets-Volt, R. Parncutt, M. Marin & J. Ross (Eds.), *Proceedings of the the 3rd Conference on Interdisciplinary Musicology (CIM07)*. Tallinn: Estonia. Available online at http://www-gewi.uni-graz.at/cim07/CIM07%20Proceedings/CIM07_Mitchell-Kenny_Open%20throat%20technique.pdf

Palmer, C., & Dalla Bella, S. (2004). Movement amplitude and tempo change in piano performance. *Journal of the Acoustical Society of America*, 115, 2590.

Port, R. F., & van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: Bradford Books/MIT Press.

Rauber, A., Pampalk, E., & Merkl, D. (2003). The SOM-enhanced JukeBox: Organization and visualization of music collections based on perceptual models. *Journal of New Music Research*, 32(2), 193–210.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131–141.

Schmidt, E. (2002). Tension between invention and convention in jazz performance: some effects on the listener. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition (ICMPC7)*. Adelaide, Australia: Causal Productions.

Seddon, F. A. (2005). Modes of communication during jazz improvisation. *British Journal of Music Education*, 22(1), 47–61.

Sundberg, J. (1987). *The science of the singing voice*. Dekalb, IL: Northern Illinois University Press.

Todd, N. P. McAngus, O'Boyle, D. J., & Lee, C. S. (1999). A sensory-motor theory of rhythm, time-perception and beat induction. *Journal of New Music Research*, 28(1), 5–28.

Toiviainen, P., & Eerola, T. (2006). Autocorrelation in meter induction: The role of accent structure. *Journal of the Acoustical Society of America*, 119(2), 1164–1170.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Wanderley M. M., & Depalle, P. (2004). Gestural control of sound synthesis. In G. Johannsen (Ed.), *Proceedings of the IEEE*. Special Issue on Engineering and Music – Supervisory Control and Auditory Communication. 92(4), 632–644.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34(1), 97–113.

Williamon, A. (2000). Coordinating duo piano performance. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda (Eds.), *Proceedings of the 6th International Conference on Music Perception and Cognition (ICMPC6)*. Keele, UK: Keele University Department of Psychology.

Williamon, A., & Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, 6(1), 1–17.