Ideal singing posture: Evidence from behavioral studies and computational motion analysis

Geoff Luck
Department of Music, University of Jyväskylä, Finland
luck@campus.jyu.fi - http://www.cc.jyu.fi/~luck

Petri Toiviainen
Department of Music, University of Jyväskylä, Finland
ptoiviai@campus.jyu.fi - http://www.cc.jyu.fi/~ptoiviai

In: K. Maimets-Volt, R. Parncutt, M. Marin & J. Ross (Eds.)
Proceedings of the third Conference on Interdisciplinary Musicology (CIM07)

Background in music psychology. It is widely accepted among professional singers and vocal teachers that there is a relationship between a singer’s posture and the quality of their voice. Behavioral studies have shown that listeners are able to detect differences in the quality of the singing voice by listening to simple recordings (e.g., Mitchell & Kenny, 2007), while observers are very sensitive to postural information contained in point light displays (PLD’s) of human activity (e.g., Davis and Gao, 2004). Thus, we would expect that independently obtained ratings of the auditory and visual elements of a singer’s performance would reveal relationships between posture and voice quality.

Background in computing, mathematics, and statistics. High-resolution motion capture systems have been used in several studies to investigate the connection between music and movement (e.g., Palmer & Dalla Bella, 2004; Wanderley, Vines, Middleton, McKay, & Hatch, 2005; Luck & Toiviainen, 2006). There are a number of methodological challenges associated with this approach. In particular, these systems produce a vast amount of high-dimensional time-series data, the conversion of which into knowledge is an arduous task. The main problem is to extract descriptors of posture and motion that would capture features relevant to the question at hand. In the present study, we apply methods of inverse kinematics to quantify features related to the overall orientation of different body parts and the amount of their movement. Furthermore, we attempt to model these features computationally.

Aims. To explore relationships between bodily posture and singing quality, and identify key characteristics of best-practice postures.

Method. 15 singers individually performed a short song (Tuulantei, by Oskar Merikanto) without accompaniment. No instructions were given as regards to how participants should stand or move during the performance. Both the auditory and visual aspects of each performance were recorded, the latter with an eight-camera optical motion capture system. Each singer’s posture was analyzed by extracting a number of kinematic features from the motion capture data, along with a range of audio features from the vocal performance. Relationships between the spatial arrangement of the limbs and the audio features during the performance were then examined statistically. The motion-capture data was used to generate PLD’s of each singer’s movements and posture during their performance. In a pilot study, the auditory and visual performances were rated independently of each other by experts in the field (professional singing teachers), using a continuous response interface, and the relationship between the two ratings examined.

Results. The data indicates that there are relationships between a singer’s posture and certain qualities of their vocal performance. For example, features relating to timbre seem to be particularly associated with the lateral angle of the head. The frontal and lateral angles of the upper body, and the frontal angle and rotation of the head, were also important.

Conclusions. Relationships between a singer’s posture and their vocal performance have been identified experimentally. The data gathered so far suggests that, although there are some general themes among these relationships, they are also subject to large individual differences.

Implications. 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.

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. Types of music-related
movement studied include expressive (e.g., Davidson, 1993; Camurri, De Poli, Friberg, Leman, & Volpe, 2005) and ancillary (e.g., Wanderley et al., 2005) movements of instrumentalists, biological motion-related bases of musical timing (e.g., Friberg & Sundberg, 1999), and conductors’ gestures (e.g., Luck, 2000; Luck & Nte, in press; Luck & Toivainen, 2006), to name but a few, and a number of large-scale research initiatives have emerged, such as the Multi-sensory Expressive Gesture Applications (MEGA) project, and the Gesture CONtrolled Audio Systems (ConGAS) initiative. Despite this increasing amount of work, however, the body’s role in vocal production has received rather little attention in the literature, despite the widely held 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 perception of the singing voice, human movement, and quantification of audio and movement data, into an exploratory study of relationships between singers’ posture and the quality of their voice.

The singing voice

The quality of a singer’s voice depends upon a number of parameters, including the physical tools a vocalist has at their disposal, i.e., their physiology, and the way in which these tools are implemented, i.e., how a singer uses their body to produce sound. Research indicates that listeners can reliably perceive differences in voice quality, but that perceptual judgments do not always correlate well with acoustical properties of the sound. Wapnick and Ekholm (1997), for example, identified 12 features that listeners used when evaluating voice quality in classical singing. Ekholm, Papagiannis and Chagnon (1998), meanwhile, matched several of the 12 perceptual features identified by Wapnick and Ekholm to acoustical features of vocal performances. In both studies, analyses indicated that most of the perceptual features tapped into a single concept of ‘overall voice quality’, suggesting that, while overall voice quality is related to many different features, it may hard to perceptually distinguish these features’ individual contribution to quality. This redundancy suggests that a simpler rating of overall voice quality may be more appropriate in perceptual experiments.

Mitchell & Kenny (2007), on the other hand, found no consistent relationships between perceptual ratings and acoustic features of singers using the ‘open throat technique’. Mitchell & Kenny examined the long-term average spectra (LTAS) of singers’ voices, and compared two measurements based on the LTAS, singing power ratio (SPR) and energy ratio (ER), with perceptual ratings of voice quality. When the singers were ranked according to their rated voice quality, no relationships were found between ratings and acoustical features.

Thus, we might suppose that, while there is some merit in collecting both perceptual judgments and performing acoustic analyses, relationships between the two types of data may not necessarily relate to each other in a systematic fashion.

Posture and movement

Since Johansson (1973), the primary way of presenting human movement to observers in perceptual studies is to use a point-light display (PLD). PLD’s reduce movement to a series of small bright dots, which usually represent the joints of a person, moving against a dark background. The dots can be joined together to make a stick-figure, if desired, and this has the effect of highlighting the spatial arrangement of the limbs, especially if movement is minimal. A large number of studies have shown that a wealth of information can be perceived from such displays.

For example, people can perceive the gender of a person (e.g., Kozlowski & Cutting, 1977), emotional characteristics (e.g., Dittrich, Trosclair, Lea, & Morgan, 1996), one’s own walking pattern (e.g., Beardsworth & Buckner, 1981), and the dynamics which underly the kinematics of the movements (e.g., Runeson & Fryckholm, 1981, 1983). PLD’s are even being considered for use as a biometric alongside fingerprints and face-recognition techniques (e.g., Nixon & Carter, 2006).
Troje (2002) examined posture in an analysis and synthesis of human gait patterns, while Davis and Gao (2004) used posture as a component in a model for gender recognition. This suggests that differences in posture may contribute to observers’ perception of types of information described above. It is unclear, however, if differences in static or near-static posture, such as that of a performing singer, can be accurately perceived in PLD’s.

Quantification

The quantification of audio data is an emerging field of research, and there are a number of different approaches, typically based on principles of signal processing, machine learning, cognitive modeling, and visualization (Downie, 2003). A large number of studies have used such techniques in areas such as computational music analysis (e.g., Lartillot, 2004, 2005), automatic classification (e.g., Toivainen & 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 et al., 2006; Luck et al., in press). Quantification of the singing voice, meanwhile, has been undertaken extensively by Sundberg (see, for example, Sundberg, 1987).

The quantification of movement data parallels the development of audio feature-extraction techniques, and a number of studies have applied such methods to the analysis of performing musicians’ movements (e.g., Wanderley, Vines, Middleton, McKay, & Hatch, 2005) and conductors’ gestures (e.g., Luck, 2000; Luck & Nte, in press; Luck & Sloboda, 2007).

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), children’s rhythmic movement to music (Eerola, Luck, & Toivainen, 2006), and conductor-musician synchronization (Luck & Toivainen, 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. Following primary data collection, a small perceptual experiment was carried out in which experts rated the quality of each singer’s performance in terms of pitch accuracy and tension, as well as the quality of their posture. This pilot study indicated that several alterations were necessary a) to the point-light stimuli, and b) to the instructions given to participants. Thus, further data collection will be necessary before we can report the results of this experiment.

In addition, the movement and audio data were subjected to a computational feature-extraction process, and relationships between indicators of voice quality and posture examined by statistical analysis. The remainder of the paper will focus on this computational aspect of the study.

Given that this was the first study of its kind, we made no specific hypotheses regarding the types of relationships we would observe, other than that we expected some systematic relationships to emerge.

Method I: Data collection

Participants

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

Apparatus and procedure

Data collection took place in a professional recording studio to ensure that good quality recordings were made. Each singer performed two verses of "Tuulantei" by Oskar Merikanto, a song they were all familiar with and had sung before. Each singer was 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 about one minute.

Both the auditory and visual aspects of each performance were recorded. Singers’ movements were recorded with a Qualisys optical motion capture system at 120 fps using eight cameras to track reflective markers attached to key locations on body. The audio was recorded with ProTools using a high-quality microphone positioned two meters from the singer.

**Method II: Feature extraction**

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

**Audio features.** Four timbre-related features were extracted from the audio data using a one-second sliding window. To be conformant with the kinematic features described below, the sliding window was moved at steps of 1/120 second.

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

  \[ c = \frac{\sum a_i f_i}{\sum a_i} \]

  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

  \[ h = -\frac{\sum a_i \ln a_i}{\ln M} \]

  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.

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

  \[ A = \frac{1}{N} \sqrt{\sum y_i^2} \]

  where \( y_i \) denotes the amplitude of the \( i \)'th sample and \( N \) the number of samples in the window.

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

  \[ r = \frac{\sum (a_i - a_{i-1})^2}{A} \]

  This feature measures the jaggedness of the spectrum and has been found to be a perceptually relevant feature (e.g., Barhet et al. 2006).

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

- **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 angles 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 angles 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 angles 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, for all lateral and rotation angles the absolute values were used in the statistical analyses. PLD’s were also produced from the motion-capture data in Matlab for use in the pilot behavioral experiment. These were identical to the stick-figures shown in Figure 1.

Results

Relationships between the kinematic features and the audio features were investigated using stepwise ordinary least squares regression. Specifically, four separate regression analysis were carried out, in each of which the 14 kinematic features were entered as predictors of one of the audio features. 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 variables emerged. Consequently, it was decided to analyze each participant separately.

A second series of stepwise linear regression analyses were thus carried out, four analyses for each participant. In each analysis, the 14 kinematic features were entered as predictors of one of the audio features. This series of analyses did reveal some significant patterns in the data, but only for two of the audio features: Spectral irregularity and rms amplitude.

Tables 1 shows the number of participants for whom there was a significant relationship, either positive or negative, between spectral irregularity and the kinematic features. Table 2 shows the same thing with respect to rms amplitude.

Figure 1. Positions of the markers that were used in the analysis, frontal view on the left, lateral view on the right.
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 the head and upper body 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 thought of as ‘noisiness’ of the signal, increased when the head was tilted downwards. Meanwhile, rms amplitude, which, in perceptual terms, might be thought of as ‘loudness’, increased when the head was tilted up. These findings seem somewhat intuitive.

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 vocal apparatus, and permitting a greater flow of air.

The pilot study described briefly earlier in the paper is currently being modified. Feedback from the experts who served as raters caused us to re-think both the make-up of the PLD’s and the actual task required of them, resulting in a delay in data collection. It is hoped that more perceptual data will be collected in the near future.

As regards development of the computational analysis, future directions might include the extraction of a greater number 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.

The identification of relationships between a singer’s posture and quality of their vocal performance implies that singing teachers should stress the importance of maintaining correct posture in order to produce the best possible vocal performance. However, since the relationships between postural features and voice quality differed 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


synchronization task. Manuscript submitted for publication.


---

1 http://www.megaproject.org

2 http://www.cost287.org/index.html