# On the detection of the level of attention in an orchestra through head movements

Giorgio Gnecco*

IMT – Institute for Advanced Studies,
Piazza S. Ponziano 6,
55100 Lucca, Italy
and
DIBRIS Department,
University of Genoa,
Via Opera Pia, 13,
16145 Genova, Italy
E-mail: giorgio.gnecco@imtlucca.it
E-mail: giorgio.gnecco@unige.it
*Corresponding author

Donald Glowinski

NEAD – Swiss Centre for Affective Sciences,
Campus Biotech – Uni Dufour,
Rue Général Dufour 24,
1211 Genève 4, Switzerland
and
DIBRIS Department,
University of Genoa,
Via Opera Pia, 13,
16145 Genova, Italy
E-mail: donald.glowinski@unige.ch

Antonio Camurri and Marcello Sanguineti

DIBRIS Department,
University of Genova,
Via Opera Pia, 13,
16145 Genova, Italy
E-mail: antonio.camurri@unige.it
E-mail: marcello.sanguineti@unige.it

**Abstract:** Results from a study of non-verbal social signals in an orchestra are presented. Music is chosen as an example of interactive and social activity, where non-verbal communication plays a fundamental role. The orchestra is adopted as a social group with a clear leader (the conductor) of two groups of musicians (the first and second violin sections). It is shown how a reduced set of simple movement features – head movements – can be used to measure the levels of attention of the musicians with respect to the conductor and the music stand under various conditions (different conductors/pieces/sections of the same piece).

**Keywords:** automated analysis of non-verbal behaviour; head ancillary gestures; level of attention.

**Reference** to this paper should be made as follows: Gnecco, G., Glowinski, D., Camurri, A. and Sanguineti, M. (2014) 'On the detection of the level of attention in an orchestra through head movements', *Int. J. Arts and Technology*, Vol. 7, No. 4, pp.316–338.

**Biographical notes:** Giorgio Gnecco received his Laurea (MSc) degree Cum Laude in Telecommunications Engineering and PhD in Mathematics and Applications, both from the University of Genoa in 2004 and 2009, respectively. He received his Diploma in Violin at the Livorno Higher Music School and Diploma in Viola at the Piacenza Conservatory in 2000 and 2001, respectively. After having been a postdoctoral researcher at the DIBRIS Department at the University of Genoa, he is currently an Assistant Professor in Systems Control and Optimisation at IMT – Institute for Advanced Studies, Lucca, Italy. His current research topics include: network optimisation, optimal control, neural networks, statistical learning theory, game theory, and affective computing.

Donald Glowinski received his MSc in Cognitive Science from the École des Hautes Etudes en Sciences Sociales (EHESS), MSc in Music and Acoustics from the Conservatoire National Superieur de Musique et de Danse de Paris (CNSMDP), MSc in Philosophy from the Sorbonne-Paris IV, and PhD in Computing Engineering from the InfoMus International Research Centre – Casa Paganini, Genoa, Italy, under the direction of Professor Antonio Camurri, where he was a research fellow from 2009 to 2013. Currently, he is a scientific collaborator at University of Geneva with Professor Didier Grandjean. His research interests include user-centric, multimodal, and social aware computing. He works in particular on the modelling of automatic gesture-based recognition of emotions in real-word scenarios. He is a member of the IEEE.

Antonio Camurri holds a PhD in Computer Engineering, and he is an Associate Professor at the University of Genova, where he teaches human computer interaction and multimodal systems for human computer interaction. He is founder and Scientific Director of InfoMus Lab and of Casa Paganini – InfoMus International Research Centre, and a founding member of the Italian Association for Artificial Intelligence. He is an author of more than 150 international scientific publications. He is coordinator and local project manager of more than 20 EU projects, co-owner of patents on software systems, and responsible for University of Genoa of industry contracts. His research interests include: multimodal intelligent interfaces and interactive systems; sound and music computing; kansei information processing; computational models of non-verbal expressive gesture, emotion, and social signals; interactive multimodal systems for theater, music, dance, museums; interactive multimodal systems for therapy, rehabilitation, independent living.

Marcello Sanguineti received his Laurea (MSc) cum Laude in Electronic Engineering and PhD in Electronic Engineering and Computer Science from the University of Genova, Italy, where he is currently an Associate Professor of Operations Research. He covers also a research associate position at the National Research Council of Italy. He authored or coauthored more than 200 research papers in archival journals, book chapters, and international conference proceedings. His main research interests include: infinite-dimensional programming, non-linear programming in learning from data, network and team optimisation, neural networks for optimisation, and affective computing. He is an Associate Editor of various international journals and member of the programme committees of several conferences. From 2006 to 2012, he was an Associate Editor of the *IEEE Trans. on Neural Networks*.

## 1 Introduction

Music is a well-known example of interactive and social activity where non-verbal communication plays a fundamental role. Several works have shown how the movements of a player can carry information about a music performance (e.g., by conveying different expressive intentions).

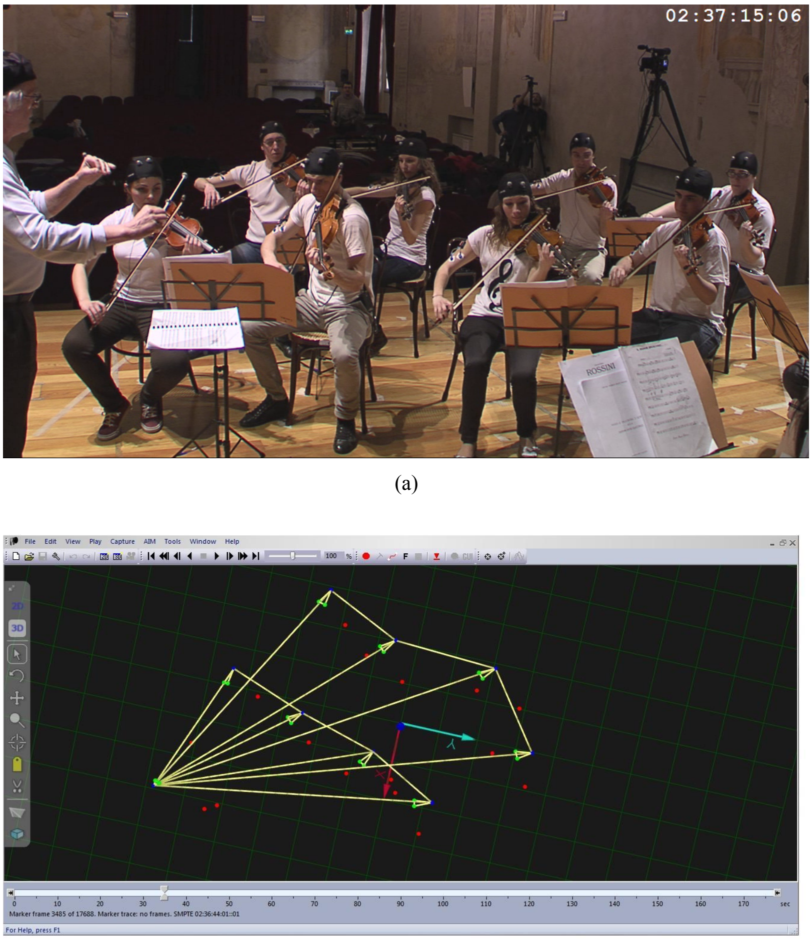
Among visual features, in this paper we focus on head movements, which are instances of the so-called *ancillary* or *accompanist gestures* (Wanderley, 2002), i.e., movements of a music instrument or of the body of a music player, not directly related to the production of the sound (vs. *instrumental* or *effective gestures*, which are directly involved in sound production). For instance, the movements of the bows of string players are (mainly) instrumental gestures, whereas the movements of their heads are ancillary gestures. Some movements of the hands of a harpist during and after string plucking are classified as ancillary gestures (Chadefaux et al., 2012). The movements of the bell of a clarinet are often classified as such, too (Wanderley, 2002), since they are performed spontaneously by the music player – although they play a direct role in the production of sound (being movements of a sound source, the clarinet). Obviously, instrumental gestures are informative: without them, musicians would not be able to express the different musical ideas they want to communicate. Ancillary gestures are informative, too, as they often allow one to recognise different expressive intentions, without looking at the instrumental gestures/listening to the performance. For instance, for the case of a piano player, Davidson (1993) claimed that visual information alone is sufficient to discriminate among performances of the same piece of music played with different expressive intentions (inexpressive, normal and exaggerated), and that the larger the amplitude of the movement, the deeper the expressive intention (Davidson, 1994). This was confirmed by other studies. Among others, Castellano et al. (2008) investigated the discriminatory power of several movement-related features for the case of a piano player and Palmer et al. (2009) showed how the movement made by the bell of a clarinet is larger when the player performs more expressive interpretations of the same piece. These works focus on a performance by merely one player. More recent studies address non-verbal communication in larger musical ensembles such as a string quartet (Varni et al., 2010) and a section of an orchestra (D'Ausilio et al., 2012). Among ancillary gestures, head movements are particularly significant. They are known to play a central role in non-verbal communication, in general (Glowinski et al., 2011), and in music, in particular (Dahl et al., 2009). They may express, e.g., the way how musicians understand the phrasing and breathing of the music and so provide information about the high-level emotional structures in terms of which the players are interpreting the music. Head movements have been investigated, e.g., in Glowinski et al. (2013) to estimate the position of a common point of interest of string players in a quartet – or more generally, a group of people (Stiefelhagen, 2002; Camurri et al., 2013) – and in Gnecco et al. (2013) to study how they depend on the presence/absence of a such a common point of interest. In principle, eye-gazes would be better suited than head directions for these applications. However, still nowadays eye-gaze tracking equipment is intrusive and costly. Moreover, previous studies have shown that often head direction and eye-gaze are correlated (Stiefelhagen, 2002; Stiefelhagen and Zhu, 2002; Stiefelhagen et al., 2002; Ba and Odobez, 2006). In Camurri et al. (2013), the role of each player in estimating his/her contribution in the determination of the position of the point of interest has been evaluated using head movements combined with cooperative game theory.

The present study, which is an improved and extended version of Gnecco et al. (2013), is aimed at investigating how the head movements of a group of players in an orchestra can be used to measure the levels of attention toward the conductor and the music stand under various conditions (different conductors/music pieces/sections of the same piece). In Section 2, the experimental methodology is described. In Section 3, the data analysis is presented. In Section 4 the obtained results are shown and discussed. Finally, Section 5 contains some conclusive remarks.

## 2 Experimental methodology

The experiments took place in a 250-seat auditorium, an environment similar to a concert hall. Figure 1 illustrates the setting. Two violin sections of an orchestra and three orchestra's professional conductors were involved in the study. Each section counted four players and was equipped with passive markers of a Qualisys motion capture system. More specifically, for each musician two markers were placed above the eyes and one on the nape (back of the neck). The violinists of each section were disposed in a single row. Additional markers, not considered in this analysis, were placed on the bows of the players and on the baton of the conductors. The musicians in the other sections of the orchestra played in all the recordings but their movements were not tracked. Various experimental conditions were tested, which differ by the presence of a different conductor and the music piece that was performed: about one minute of music excerpts from the Overture to the opera 'Il signor Bruschino' by G. Rossini, and about 90 seconds of music excerpts from the third movement of the 'Vivaldiana' for orchestra by G.F. Malipiero). Each experimental condition was repeated three times, for a total of 18 recordings. The recordings belonging to the same experimental condition were executed one after the other. The frames were recorded at a frame rate of 100 frames per second. The present study focuses on measuring the levels of attention of the musicians toward the conductor and the music stand, resp., through an analysis of the movements of their heads under the various experimental conditions.

**Figure 1** (a) Locations of the players and the conductor (b) A snapshot of the positions of the head markers of the players and of the conductor (see online version for colours)



Marker frame 3485 of 17688. Marker trace: no frames. SMPTE 02:36:44:01::01

For Help, press F1

(a)

(b)

Notes: Triangles correspond to positions and directions of heads. The other markers are represented by red dots in the online version.

## 3 Data analysis

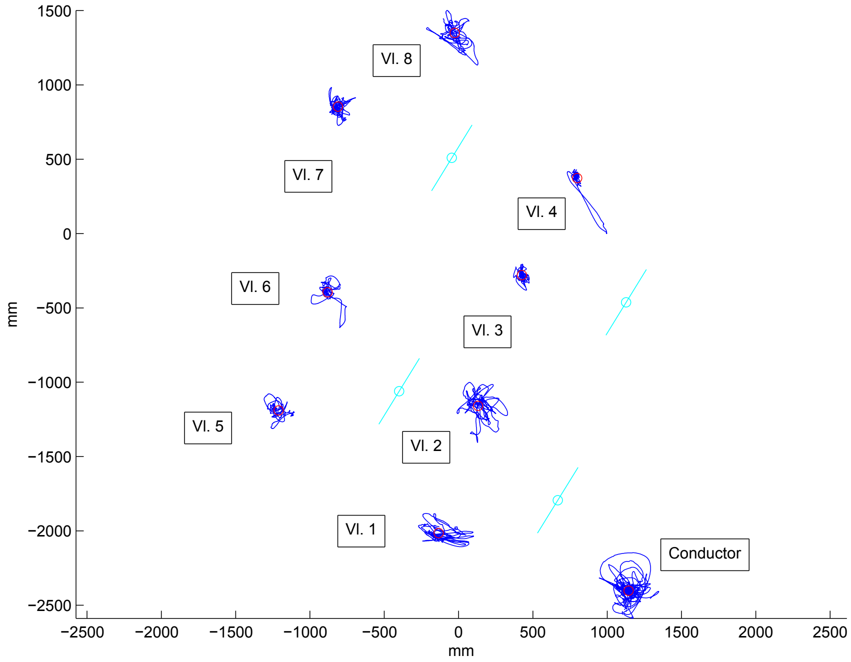
Movement data were collected by using a Qualisys motion capture system equipped with seven cameras, integrated with the EyesWeb extended multimodal interaction (XMI) software platform to obtain synchronised multimodal data. The data analysis was performed using MATLAB 7.7. A reduced dataset, describing the positions of three reflective markers associated with the heads of the musicians in the 18 recordings, was extracted from the collected data and head movement features were automatically computed. For each row, the violinists have been numbered from left to right, from

1 to 4 for the first section (first row) and from 5 to 8 for the second section (second row).

- *Choice of the data*. In the data analysis, we have considered only movement features associated with the heads of the musicians. One reason for taking into account only the movements of the heads is that they are purely ancillary gestures and they are not prescribed by the music score to the same extent as the movements of the bows.
- *Choice of the features*. The following features have been computed in the data analysis. Their computation has been made possible by the Qualisys tracking manager (QTM) representation of each marker, which provides its position in each frame, apart from the case of frames in which the marker was undetected or unlabelled, so its position was not determined. All the geometric features have been defined taking into account the projections of the motion-capture data on the horizontal plane, thus discarding the vertical component (then, 2-dimensional vectors have been considered). Indeed, for each musician the two frontal markers have been positioned much above his/her eyes, so the vertical component of the positions of such markers was misleading, e.g., in determining the direction of the head. It is important to observe that the horizontal component of the head direction can be recovered from the data associated with the horizontal movements of the head marker data, as long as the heads perform rotations only around the vertical axis (panning) and the sagittal one (tilting). Indeed, this was the case in our recordings, as no significant rotations of the heads around the frontal axes were observed through a visual inspection of the video recordings (likely because the violinists used the shoulder rest to hold the violin, thus reducing the amplitudes of such rotations). This allows the use of a simplified model, in which only 2-dimensional vectors are considered. Another possible approach – not followed in the paper – consists in estimating all the three components of the head directions. This could be achieved, e.g., by introducing into the data analysis individual corrections to the positions of the head markers, obtained by estimating their relative positions with respect to the eyes. We have adopted the first approach, which discards the vertical components of the positions of the markers, since it is simpler, well-motivated in the present application (as discussed above), and does not require the estimation of such relative positions. Finally, we remark that the strategy of looking at head movements in the horizontal plane has been used in several works (see, e.g., Stiefelhagen, 2002; Stiefelhagen and Zhu, 2002; Stiefelhagen et al., 2002).

First, for the frames in which the positions of all the markers associated with the heads were determined, the positions of the barycenters of the heads of the musicians have been computed. Each of them is defined as the barycenter of the three markers associated with the head of the musician. Figure 2 shows the trajectories of such barycenters for a particular recording, together with their average positions with respect to all the frames of the respective performance. Of course, the frames before the beginning of the performance were excluded from the computation of the average, as well as the ones after its end and the frames for which at least one of the three markers was undetected or unlabelled, so the position of the barycenter was not determined. In the figure, the asymmetry of the movement patterns between the left-side player and the right-side player associated with each music stand is due to the fact that the left-side player was responsible of turning pages during the performance.

**Figure 2** Estimated locations of the music stands (cyan), average locations of the heads (red) of the eight violinists and of the conductor, and their trajectories (blue) for one of the recordings (see online version for colours)



In Figure 2, the music stands have been represented by segments in the horizontal plane. According to the experimental setup, it was decided in advance to place the music stands in a parallel fashion. However, before the beginning of some recordings, some musicians moved accidentally the music stands. So, for each recording, the locations of the music stands have been estimated using the following method (which we have developed and discussed with some professional violinists with orchestral experience):

1 First, the horizontal position of the chair of each violinist has been estimated as the average horizontal position of the barycenter of his/her head.

2 Then, the mid-point of the segment between the estimated horizontal positions of the chairs of the two violinists associated with the same music stand has been determined.

3 Subsequently, a first estimate of the position of the mid-point of the music stand has been found by starting from the point determined in item 2) and moving forward – as the music stand was in front of the two musicians – along the direction orthogonal to the segment in item 2) by 70 cm, which is an estimate of the typical distance of a music stand from such a point (of course, the obtained

estimate was by construction equidistant from the violinists, as this was the original displacement of the music stand, in absence of its re-positioning by one of the two violinists).

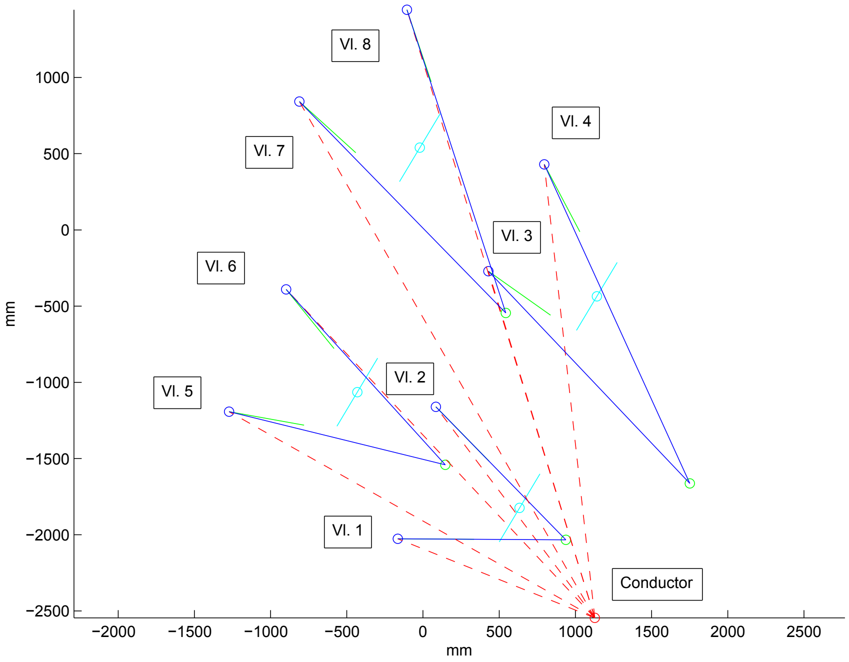
4 An additional correction to the position of the music stand has been inserted for the case in which – by looking at the videos – the music stand was found to be significantly closer to one violinist than to the other. For each music stand, such a correction was the same for all the video-recordings belonging to the same experimental condition (i.e., any fixed pair conductor/piece). Indeed – the few times this accidentally happened – the music stands were moved only in the downtimes successive to each group of the three recordings performed (consecutively) under the same performance conditions.

5 Finally, starting from the just-determined estimate of the position of the mid-point of the music stand, its extreme points have been estimated by moving by 26 cm (an estimate of the half of the width of the music score) in each of the two senses along the average direction of the four vectors joining the estimated chairs of the left-sided players to the ones of the right-sided players, for each music stand. This procedure was followed since, in such a way, the estimated music stands were automatically placed in a parallel fashion.

In spite of the various estimates used inside the procedure described above, the final displacements of the music stands were in good agreement with their actual displacements observed in the video recordings [compare, e.g., Figures 1(a) and 2].

Then, for the frames in which the position of the barycenter of the head is determined, the (2-dimensional) direction of the head of each violinist has been estimated as the unit-norm vector joining the barycenter of the head with the mid-point of the segment between the two frontal markers. A correction to such a direction [i.e., a rotation by a specific angle around the vertical axis) has been inserted for the case in which the three markers on the head of the violinist were misplaced (this happened, for instance, for the displacement of the markers on the head of the violinist nearest to the conductor in Figure 1(a)]. Again, such a correction has been obtained by looking at each video-recording (in particular, searching for a frame containing a frontal view of each violinist, in some cases even a few seconds before or after the actual performance), and was the same for all the video-recordings belonging to the same experimental condition. Indeed, we remark that the musicians did not re-position their head markers during each performance, but – the few times this happened – only in the downtimes after each group of three consecutive recordings, performed under the same conditions. Then, for the frames for which the positions of both barycenters are determined, also the segments joining the barycenters of the heads of the violinists to the barycenter of the head of the conductor have been determined (see Figure 3). Finally, for each violinist, the average oriented angle between the $x$-axis and the corrected direction of the head has been evaluated, where the average has been performed with respect to all the frames of the performance for which the corrected direction of the head has been determined. Then, by rotating counter-clockwise the $x$-axis by such an angle, the average corrected direction of the head of the violinist has been obtained (see again, Figure 3).

**Figure 3** Corrected directions of the heads of the violinists (blue) and segments (dashed; red) joining the barycenters of their heads to the barycenter of the head of the conductor, for a particular frame of one of the recordings (see online version for colours)



Notes: The average corrected directions of the heads are also shown (green), together with the estimated displacements of the music stands (cyan). To help the visualisation, for each pair of violinists associated with the same music stand, also the intersection of the two half-lines having the barycenters of their heads as origins and directed as the corrected head directions is shown.

Starting from the features above, the following four higher-level *individual* features have been evaluated for each violinist and each frame.

- *Level of attention of the violinist toward the conductor*:

  a Equal to 1 if the angle between the corrected direction of the head of the violinist and the vector joining the barycenter of the head of the violinist with the barycenter of the head of the conductor is equal to or smaller than a given threshold. This has been chosen to be equal to a musician-dependent value, which is the sum of two terms: the first one is a constant (here, chosen as 12°), which takes (directly) into account a possible misalignment between the corrected head direction and the direction of the eye gaze, when looking at the conductor. The second one is musician-dependent as it is inversely proportional to the distance between the musician and the conductor, and – as the violinist changes – varies between 3° and 9° (so, the maximum threshold is 21°). The reason for such a second term is the following: it takes into account the fact that, when such distance is smaller, the musician can see the conductor under a larger angle.

b Equal to 0 if the angle above is larger than the threshold.

c Not determined if the position of the barycenter of the head of the violinist or of the conductor is not determined in that frame.

- *Level of attention of the violinist toward the music stand*:

  a Equal to 1 if the half-line starting from the barycenter of the head of the violinist and having the corrected direction of the head of the same violinist intersects the segment that models the music stand in front of the violinist. Also for this feature, a possible misalignment between the corrected head direction and the direction of the eye gaze when reading the music part has been taken (indirectly) into account, since the feature is equal to 1 for a whole range of oriented angles between the $x$-axis and the corrected head direction.

  b Equal to 0 if they do not intersect.

  c Not determined if the position of the barycenter of the head of the violinist is not determined in that frame.

- *Distance of the barycenter of the head of the violinist from its average position*:

  a Not determined if the position of the barycenter of the head of the violinist is not determined in that frame.

- *Oriented angle between the average corrected direction of the head and the corrected direction of the head*:

  a belonging to the interval $[-\pi, \pi)$, for the frames in which the corrected direction of the head is determined (of course, such an oriented angle does not depend on the choice of the $x$-axis)

  b not determined, otherwise.

For illustrative purposes, the following Figures 4 to 7 refer to the same recording. Figures 4 and 5 show the two levels of attention defined above for the violinists of the first section and those of the second section, respectively, whereas Figures 6 and 7 show respectively, for each violinist, the distance of the barycenter of the head from its average position and the oriented angle between the average corrected direction of the head and the corrected direction of the head in each frame. Of course, only the frames of the actual performance have been considered in defining the quantities above. It is interesting to observe from Figures 4 and 5 that for some musicians, the frames for which the musician's head is directed toward the conductor approximately coincide with the frames for which it is not directed toward the music stand, and vice-versa. However, this remark does not extend to all the musicians, due to their different displacements (see Figure 2). We refer to Section 5 for more details on a possible way to overcome this issue.

**Figure 4** Level of attention toward the conductor (above; blue) and the music stand (below) for each violinist of the first section, for one fixed recording (see online version for colours)

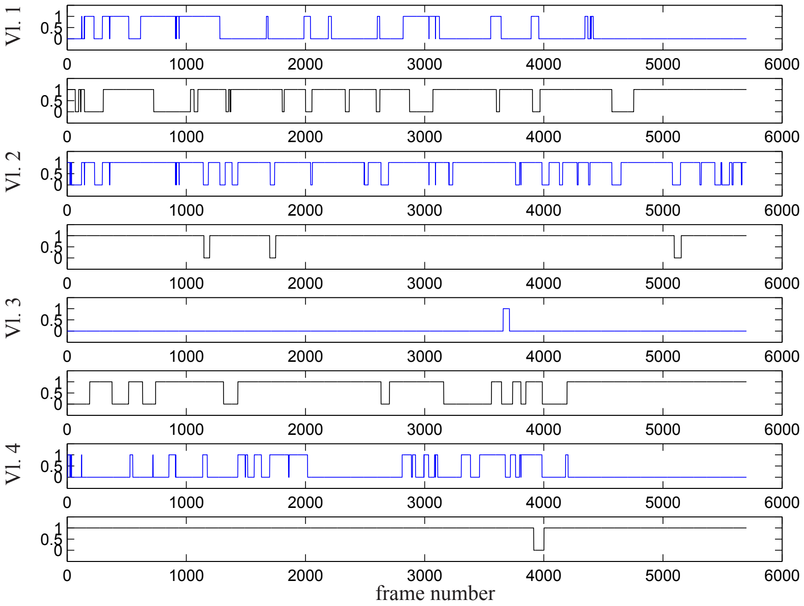


**Figure 5** Level of attention toward the conductor (above; blue) and the music stand (below) for each violinist of the second section, for one fixed recording (see online version for colours)

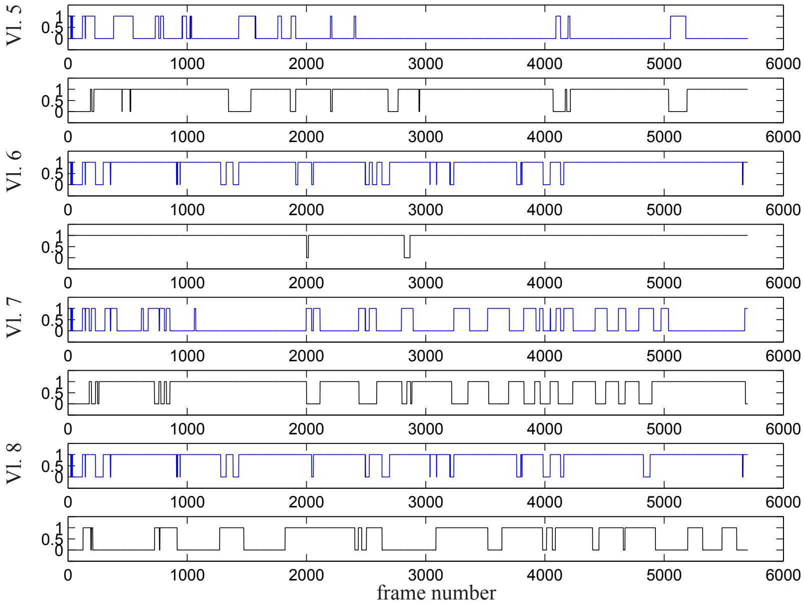


**Figure 6** Distance (in mm) of the barycenter of the head from its mean position for each violinist, for one fixed recording (see online version for colours)

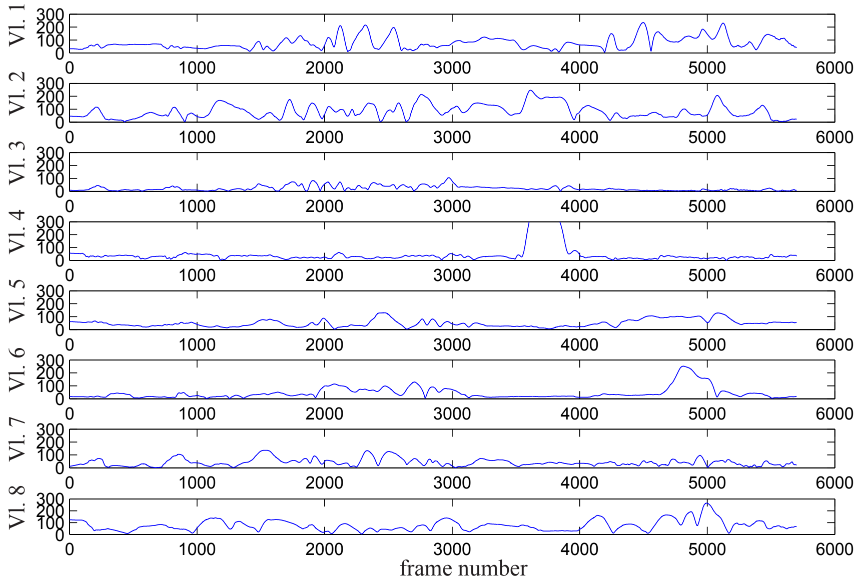
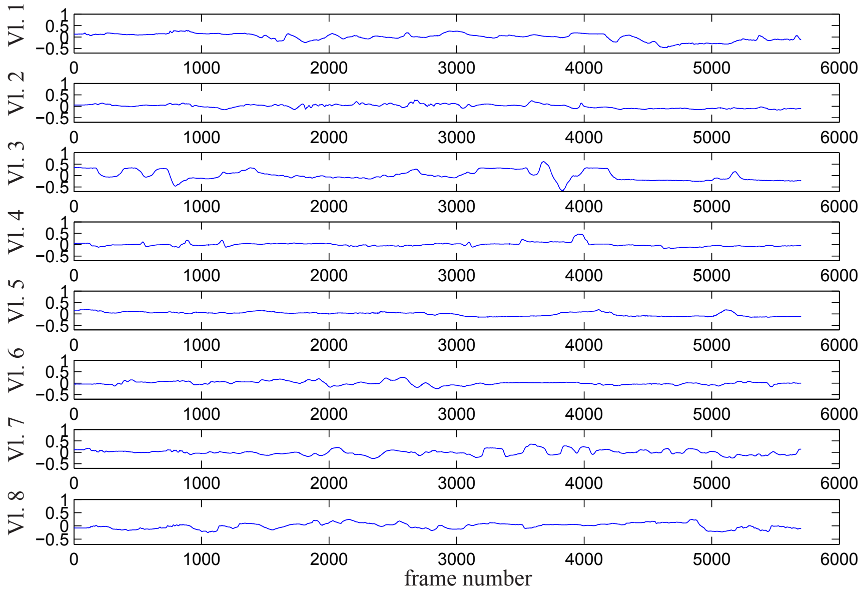


**Figure 7** Oriented angle (in rad) between the average corrected direction of the head and the corrected direction of the head for each violinist, for one fixed recording (see online version for colours)



Finally, for each section, each recording, and a given set of consecutive frames, the following *group* features have been evaluated.

- *Feature $A$: average level of attention of the musicians of the section toward the conductor*. The average is performed with respect to all the musicians of the section and the given set of frames.
- *Feature $B$: average level of attention of the musicians of the section toward the music stand*. The average is performed with respect to all the musicians of the section and the given set of frames.
- *Feature $C$: average of the distances of the barycenters of the heads of the violinists of the section from their average positions*. The first average is performed with respect to all the musicians of the section and the given set of frames; for each musician, the average position of the barycenter of the head is obtained considering the given set of frames.
- *Feature $D$: standard deviation of the average of the oriented angles between the corrected head directions of the musicians of the section and their average corrected directions*. For each violinist, the average corrected direction is evaluated averaging with respect to the given set of frames; the average of the oriented angles is computed frame-by-frame, by performing the average with respect to all the musicians of the section; the standard deviation is performed with respect to the given set of frames.

Of course, we have excluded from the computation the frames in which some of the quantities to be averaged are not determined, thus reducing the effective number of frames on which the averages are evaluated.

## 4 Results

We first report the values assumed in the available recordings by the features defined in Section 3. Then, we present the results of a statistical analysis for the feature $A$. Being conscious of the possible presence of residual errors after performing the calibration processes described in Section 3, in order to obtain meaningful insights from the available data when comparing various conditions we have decided to focus on comparisons in which all the factors possibly ‘difficult to be calibrated’ are fixed (if necessary, with estimated musician-dependent values) for each of the conditions (‘treatments’) to be compared, when they correspond to the same ‘block’ of observations. In particular, we have used non-parametric statistical tests such as the Friedman test (Bewick et al., 2004) and the Wilcoxon signed-rank test (Whitley and Ball, 2002).

Two choices of the set of consecutive frames have been made in the definitions of the features $A$, $B$, $C$ and $D$ provided in Section 3: all the frames of the performance (case 1), and the frames corresponding to the beginning of the performance (case 2), whose duration was assumed to be equal to eight seconds (starting from about one second before the attack of the piece by the conductor).

**Table 1** For each performance of the two pieces, average level of attention of the musicians of each section toward the conductor (feature $A$)

*Piece 1*

*(a) Whole performance*

| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
|---|---|---|---|
| *Section 1/Recording 1* | 0.386 | 0.403 | 0.688 |
| *Section 1/Recording 2* | 0.362 | 0.355 | 0.657 |
| *Section 1/Recording 3* | 0.330 | 0.383 | 0.574 |
| *Section 2/Recording 1* | 0.496 | 0.547 | 0.827 |
| *Section 2/Recording 2* | 0.466 | 0.499 | 0.657 |
| *Section 2/Recording 3* | 0.480 | 0.524 | 0.672 |

*(b) Beginning of the performance (first 8 seconds)*

| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
|---|---|---|---|
| *Section 1/Recording 1* | 0.468 | 0.577 | 0.871 |
| *Section 1/Recording 2* | 0.483 | 0.376 | 0.830 |
| *Section 1/Recording 3* | 0.395 | 0.410 | 0.882 |
| *Section 2/Recording 1* | 0.546 | 0.694 | 0.970 |
| *Section 2/Recording 2* | 0.532 | 0.537 | 0.793 |
| *Section 2/Recording 3* | 0.569 | 0.549 | 0.891 |

*Piece 2*

*(c) Whole performance*

| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
|---|---|---|---|
| *Section 1/Recording 1* | 0.439 | 0.293 | 0.240 |
| *Section 1/Recording 2* | 0.401 | 0.252 | 0.198 |
| *Section 1/Recording 3* | 0.373 | 0.274 | 0.238 |
| *Section 2/Recording 1* | 0.660 | 0.706 | 0.558 |
| *Section 2/Recording 2* | 0.708 | 0.633 | 0.596 |
| *Section 2/Recording 3* | 0.687 | 0.703 | 0.602 |

*(d) Beginning of the performance (first 8 seconds)*

| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
|---|---|---|---|
| *Section 1/Recording 1* | 0.471 | 0.282 | 0.393 |
| *Section 1/Recording 2* | 0.308 | 0.258 | 0.290 |
| *Section 1/Recording 3* | 0.358 | 0.275 | 0.349 |
| *Section 2/Recording 1* | 0.717 | 0.807 | 0.633 |
| *Section 2/Recording 2* | 0.7455 | 0.682 | 0.528 |
| *Section 2/Recording 3* | 0.730 | 0.788 | 0.734 |

**Table 2** For each performance of the two pieces, average level of attention of the musicians of each section toward the music stand (feature $B$)

| | | | |
|---|---|---|---|
| *Piece 1* | | | |
| *(a) Whole performance* | | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.910 | 0.904 | 0.817 |
| *Section 1/Recording 2* | 0.879 | 0.869 | 0.772 |
| *Section 1/Recording 3* | 0.902 | 0.878 | 0.734 |
| *Section 2/Recording 1* | 0.930 | 0.665 | 0.661 |
| *Section 2/Recording 2* | 0.927 | 0.740 | 0.866 |
| *Section 2/Recording 3* | 0.867 | 0.779 | 0.813 |
| *(b) Beginning of the performance (first 8 seconds)* | | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.856 | 0.695 | 0.752 |
| *Section 1/Recording 2* | 0.762 | 0.653 | 0.663 |
| *Section 1/Recording 3* | 0.936 | 0.822 | 0.626 |
| *Section 2/Recording 1* | 0.944 | 0.420 | 0.530 |
| *Section 2/Recording 2* | 0.967 | 0.672 | 0.761 |
| *Section 2/Recording 3* | 0.885 | 0.631 | 0.806 |
| *Piece 2* | | | |
| *(c) Whole performance* | | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.522 | 0.415 | 0.582 |
| *Section 1/Recording 2* | 0.409 | 0.412 | 0.603 |
| *Section 1/Recording 3* | 0.324 | 0.359 | 0.644 |
| *Section 2/Recording 1* | 0.750 | 0.637 | 0.816 |
| *Section 2/Recording 2* | 0.752 | 0.637 | 0.726 |
| *Section 2/Recording 3* | 0.715 | 0.657 | 0.746 |
| *(d) Beginning of the performance (first 8 seconds)* | | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.161 | 0.237 | 0.298 |
| *Section 1/Recording 2* | 0.189 | 0.269 | 0.546 |
| *Section 1/Recording 3* | 0.017 | 0.294 | 0.474 |
| *Section 2/Recording 1* | 0.724 | 0.456 | 0.808 |
| *Section 2/Recording 2* | 0.706 | 0.474 | 0.690 |
| *Section 2/Recording 3* | 0.764 | 0.486 | 0.674 |

**Table 3** For each performance of the two pieces, average of the distances of the barycenters of the heads of the violinists of each section from their average positions (feature $C$)

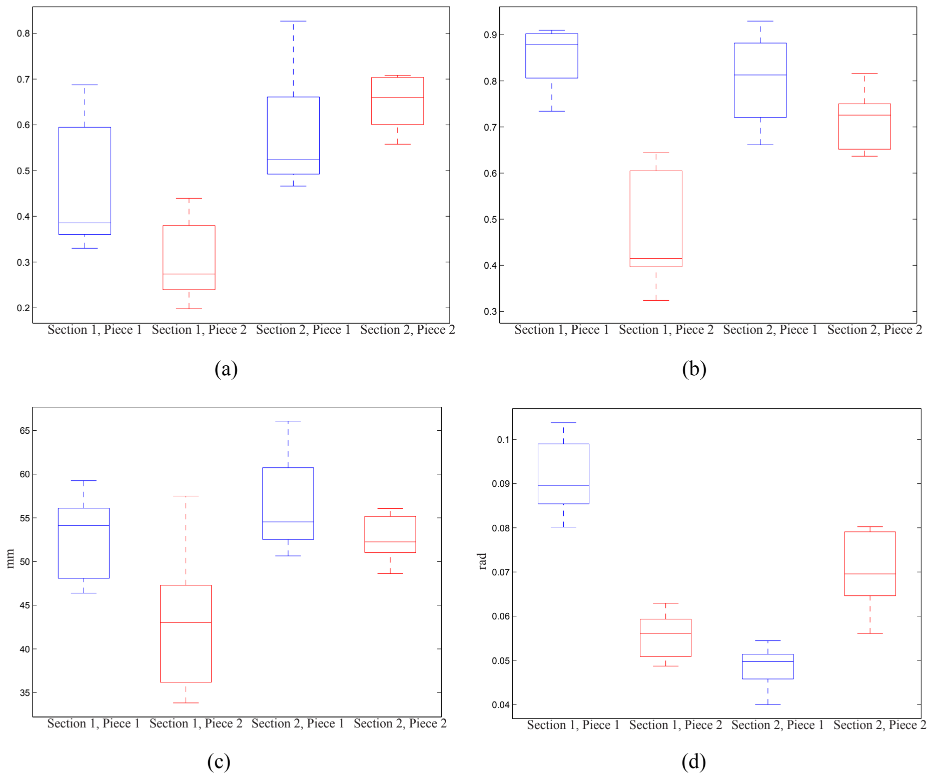
| | *Piece 1* | | |
|---|---|---|---|
| | *(a) Whole performance* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 51.1 | 46.4 | 55.6 |
| *Section 1/Recording 2* | 54.2 | 59.3 | 46.5 |
| *Section 1/Recording 3* | 48.6 | 57.7 | 54.1 |
| *Section 2/Recording 1* | 54.5 | 50.6 | 51.9 |
| *Section 2/Recording 2* | 54.5 | 58.4 | 61.4 |
| *Section 2/Recording 3* | 60.5 | 52.7 | 66.1 |
| | *(b) Beginning of the performance (first 8 seconds)* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 33.2 | 20.9 | 32.0 |
| *Section 1/Recording 2* | 40.5 | 34.6 | 30.4 |
| *Section 1/Recording 3* | 28.9 | 26.3 | 45.0 |
| *Section 2/Recording 1* | 22.3 | 19.7 | 31.7 |
| *Section 2/Recording 2* | 22.6 | 19.0 | 47.7 |
| *Section 2/Recording 3* | 30.6 | 25.6 | 28.5 |
| | *Piece 2* | | |
| | *(c) Whole performance* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 43.0 | 57.5 | 39.0 |
| *Section 1/Recording 2* | 50.0 | 35.5 | 33.8 |
| *Section 1/Recording 3* | 45.3 | 36.4 | 46.4 |
| *Section 2/Recording 1* | 48.6 | 52.3 | 49.7 |
| *Section 2/Recording 2* | 53.8 | 56.1 | 51.5 |
| *Section 2/Recording 3* | 55.9 | 51.5 | 54.9 |
| | *(d) Beginning of the performance (first 8 seconds)* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 20.0 | 39.1 | 28.9 |
| *Section 1/Recording 2* | 28.3 | 26.3 | 19.5 |
| *Section 1/Recording 3* | 27.8 | 17.3 | 14.8 |
| *Section 2/Recording 1* | 40.3 | 42.5 | 26.6 |
| *Section 2/Recording 2* | 34.9 | 40.4 | 49.7 |
| *Section 2/Recording 3* | 21.1 | 49.7 | 39.5 |

**Table 4** For each performance of the two pieces, standard deviation of the average of the oriented angles between the corrected head directions of the musicians of each section and their average corrected directions (feature $D$)

| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
|---|---|---|---|
| | *Piece 1* | | |
| | *(a) Whole performance* | | |
| *Section 1/Recording 1* | 0.104 | 0.090 | 0.080 |
| *Section 1/Recording 2* | 0.098 | 0.097 | 0.087 |
| *Section 1/Recording 3* | 0.081 | 0.102 | 0.089 |
| *Section 2/Recording 1* | 0.040 | 0.046 | 0.051 |
| *Section 2/Recording 2* | 0.049 | 0.051 | 0.054 |
| *Section 2/Recording 3* | 0.050 | 0.052 | 0.045 |
| | *(b) Beginning of the performance (first 8 seconds)* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.041 | 0.033 | 0.037 |
| *Section 1/Recording 2* | 0.054 | 0.033 | 0.049 |
| *Section 1/Recording 3* | 0.072 | 0.050 | 0.033 |
| *Section 2/Recording 1* | 0.027 | 0.022 | 0.034 |
| *Section 2/Recording 2* | 0.015 | 0.034 | 0.053 |
| *Section 2/Recording 3* | 0.037 | 0.027 | 0.031 |
| | *Piece 2* | | |
| | *(c) Whole performance* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.053 | 0.063 | 0.051 |
| *Section 1/Recording 2* | 0.056 | 0.058 | 0.049 |
| *Section 1/Recording 3* | 0.058 | 0.062 | 0.050 |
| *Section 2/Recording 1* | 0.080 | 0.061 | 0.070 |
| *Section 2/Recording 2* | 0.073 | 0.066 | 0.069 |
| *Section 2/Recording 3* | 0.079 | 0.056 | 0.079 |
| | *(d) Beginning of the performance (first 8 seconds)* | | |
| | *Conductor 1* | *Conductor 2* | *Conductor 3* |
| *Section 1/Recording 1* | 0.063 | 0.058 | 0.058 |
| *Section 1/Recording 2* | 0.082 | 0.049 | 0.045 |
| *Section 1/Recording 3* | 0.070 | 0.063 | 0.035 |
| *Section 2/Recording 1* | 0.046 | 0.017 | 0.061 |
| *Section 2/Recording 2* | 0.030 | 0.048 | 0.050 |
| *Section 2/Recording 3* | 0.018 | 0.026 | 0.038 |

For each conductor/piece/section, Tables 1 to 4 show the values of the features $A$, $B$, $C$ and $D$ obtained in each performance, under both cases 1 and 2. Then, Figures 8(a) to 8(d) illustrates respectively, for each piece, the boxplots of the features $A$, $B$, $C$ and $D$ for each of the two sections, under both cases 1 and 2. The boxplots above have been obtained using the data shown in Tables 1 to 4. The dependence on the conductor has not been considered to generate the boxplots, in order to increase the number of samples (the same) used to draw each boxplot. Inspection of the data in Tables 1 to 4 and of the boxplots in Figures 8(a) to 8(d) show that, for each fixed conductor, there is usually a dependence on the piece of the feature $A$, evaluated on each whole performance. Such a dependence appears to be more pronounced in the case of the first section.

**Figure 8** For each section and piece, boxplot of, (a) the average level of attention of the musicians of the section toward the conductor (feature $A$) (b) the average level of attention of the musicians of the section toward the music stand (feature $B$); the average of the distances of the barycenters of the heads of the violinists of the section from their average positions (feature $C$); the standard deviation of the average of the oriented angles between the corrected head directions of the musicians of the section and their average corrected directions (feature $D$) (see online version for colours)



(a)

(b)

(c)

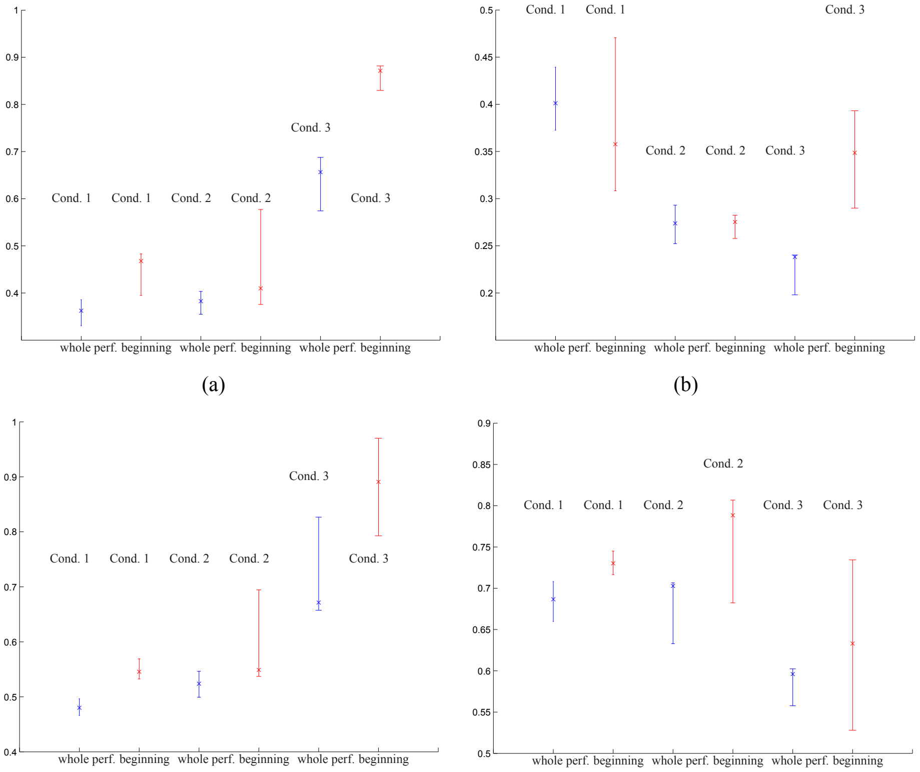
(d)

Let us now consider in more detail the case of the feature $A$, examining the entries in parts (a) and (c) of Table 1, which refers to the case 1 defined above. Interestingly, for a fixed conductor and a fixed piece, inspection of the corresponding entries in the table shows that, for the first section, the average level of attention toward the conductor, evaluated on each whole performance, has usually a larger value in the first recording than in the successive ones (although in general it is not a decreasing function of the recording number). In a sense, the first section 'memorises' the behaviour of the conductor from the first execution of a piece to the last one. This effect arises in the first section, likely because it is the nearest to the conductor. In order to validate this finding, we have performed – for each section – a comparison of repeated measures, implemented by a Friedman test, in which each block is constituted by the values assumed by the feature $A$ evaluated in the case 1 (whole performance) in the first, second and third recording under each pair conductor/piece (so, the recording numbers correspond to the 'treatments' of the test). We have chosen to use a Friedman test since, for each violinist, the corrections that have been introduced in the evaluation of the head directions – and their possible residual errors – are the same for all the three recordings belonging to the same pair conductor/piece (this motivates the dependence assumption of the test for the observations belonging to the same block). The Friedman test has provided test statistics (modelled by $\chi^2$ distributions with 2 degrees of freedom, no adjustments for ties were required) equal to 9 and 2.33 for the first and the second section, resp., and $p$-values equal to 0.011 and 0.311, resp., allowing to reject with a significance level $0.0125$ – for the case of the first section – the null hypothesis that the different samples have been drawn from three distributions with the same median. Moreover, to control the inflation of type I error probability due to multiple comparisons, $0.0125$ has been adopted as the significance level instead of $0.05$, using the Bonferroni correction (Shaffer, 1995) with parameter $n = 4$, which gives $0.05/n = 0.05/4 = 0.0125$. Indeed, here and in the following we have performed a total of four tests (two Friedman tests, and two Wilcoxon signed-rank tests).

A comparison with the entries in parts (b) and (d) of Table 1 shows that in general, the average level of attention of each section toward the conductor is larger when evaluated at the beginning of the performance than on the whole performance. This is also illustrated in Figures 9(a) to 9(d), in which the median values and the error bars of such average levels of attention are plotted and compared for each conductor. This finding can be interpreted taking into account that the role of the conductor is, of course, particularly important at the beginning of the performance (and of course, also in other parts of the performance, which may be identified by an analysis of the music score). Indeed, at the beginning of the performance, looking at the conductor is the only way for the musicians to synchronise themselves (no audio feedback from the other musicians of the orchestra is available in such a moment). With the aim of validating this finding, we have performed – for each section – a Wilcoxon signed-rank test, pairing the value assumed by the feature $A$ in case 1 (whole recording) with the one assumed in case 2 (first eight seconds of the same recording). Also for this case, the dependence assumption of the test inside each block is motivated by the fact that, for each recording, the frames corresponding to the beginning of the performance form a subset of the whole set of frames, and also the (possibly musician-dependent) calibrations are the same for the two cases. The Wilcoxon signed-rank test has provided a test statistics, $z$-value and $p$-value equal, resp., to $(18, -2.94, 0.003)$ for the first section and $(9, -3.33, 0.001)$ for the second section, allowing to reject with a significance level

0.0125 – for both sections – the null hypothesis that the median difference between the pairs is 0.

**Figure 9** Medians and error bars of the average level of attention toward the conductor (feature $A$), evaluated at the beginning of the performance and on the whole performance, (a) for the case of: the first section and the first piece (b) the first section and the second piece (c) the second section and the first piece (d) the second section and the second piece (see online version for colours)



(a) (b)

(c) (d)

For each conductor, in general the features $B$, $C$ and $D$ resulted smaller at the beginning of the performance as compared to the whole performance (in this case, the error bars are not shown, but this can be still obtained in a similar way as before). This means that, as compared to the whole performance, at the beginning of the performance the musicians of each section tend to reduce, respectively, their average level of attention toward the music stand, the amplitude of the movements of their heads, and the amplitude of the angular movements of the directions of their heads. Finally, we note that – in the same recordings – there are some differences in the tables between features evaluated on one section and the same features evaluated on the other section. However, in this case one cannot infer that such differences do really depend on the sections (for instance, due to the different parts performed by the two sections) since the features may be also dependent on the locations and the calibrations, which, in general, are different for different musicians (see Section 5 for a discussion about these topics).

## 5 Discussion

Behavioural features have been investigated for the movements of the heads of the violinists in an orchestra, in order to study their dependence on the conductor/piece/segment of a piece/number of times the same experimental condition is repeated. In particular, the average level of attention toward the conductor of the first violin section has shown to depend on the number of times each piece has been already performed, and also on the particular segment of the piece that is taken into consideration (e.g., the average level of attention toward the conductor at the beginning of the performance is in general larger than in the whole performance). Although the results reported in the paper refer to specific choices of some parameters (e.g., the thresholds in the definitions of the two levels of attention of each violinist toward the conductor and the music stand, respectively), similar results in terms of rankings of the values of the features under different conditions were obtained for a few other choices of such parameters (not reported here due to space limitations). It has also to be remarked that the 'one second anticipation' in the definition of the 'beginning of the performance' has been introduced merely with the aim of simplifying the manual procedure used to identify the initial frames of each performance. It is likely that such an anticipation introduces a bias in the definitions of the various features considered in the paper. However, this has no significant consequences in our analysis, as we mainly focus on the differences in the values of the features in different situations.

The aim of this study is to provide ways to measure the two levels of attention and to obtain some insights on their dependencies (particularly, for the case of the level of attention toward the conductor) on various factors whose influence can be determined in spite of the presence of residual errors in the performed calibrations. For instance, interesting results of the data analysis – investigated also from a statistical significance point of view – are the emergence of the 'memorisation effect' described in Section 4, and the comparison between the average levels of attention toward the conductor at the beginning of the performance and on the whole performance. Of course, various improvements are possible. In order to make more unlikely the occurrence – for the same musician – of simultaneous 1's in the features 'individual level of attention toward the conductor' and 'individual level of attention toward the music stand', the conductor may be positioned in a different way, e.g., standing on a higher floor than the violinists. Slightly different setups may be considered: e.g., one may place the musicians in a 'more symmetric way' (e.g., equally angular-spaced on two concentric arcs), in order to reduce the dependence of some features from the location. The procedure followed in the paper may be improved by making it more automatic; this would reduce residual errors. This may be achieved, e.g., by applying more sophisticated computer-vision techniques, thus reducing the need for the visual inspections used in this work to estimate some quantities. At the same, time, a fully-automatic procedure would also improve the precision and accuracy of such estimates and would be able to process a larger amount of data in less time. It would also allow to reduce or eliminate the above-mentioned 'one second anticipation' in the definition of the 'beginning of the performance'.

The features considered in this work are mainly 'attentional' features, since they aim at revealing how much the attention of each section is focused toward particular points of interest (e.g., the conductor and the music stand). Among directions of research we mention: the investigation of 'expressive' visual features, able to discriminate, e.g., between the levels of expressivity of different pieces, or between different interpretations of the same piece [see, again, Glowinski et al. (2013) for such a kind of study, in the case of a string quartet] and the investigation of possible correlations among the selected attentional features. Other possible extensions in the analysis include: the investigation of relations among the proposed features and the music score; the analysis of speed, acceleration and coordination of head movements [see, e.g., Glowinski et al. (2013) for such a kind of study on coordination, in the case of a string quartet]; the use of tools commercially available in the future, such as Google glasses, to obtain even better estimates of both levels of attention (and also estimates of all the three components of the head directions, possibly using suitable image processing techniques); the study of relations among the movements of the baton of the conductor and the level of attention toward the conductor.

## Acknowledgements

The EU ICT SIEMPRE project acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 250026-2. We thank the conductors P. Borgonovo and S. Tokay, the Orchestra of the Music Conservatory of Genoa directed by A. Tappero Merlo, and our colleagues at Casa Paganini – InfoMus Research Centre. The authors were partially supported by the 'Progetto di Ricerca di Ateneo 2012' 'Models and Computational Methods for High-Dimensional Data', granted by the University of Genoa.

## References

Ba, S. and Odobez, J-M. (2006) 'A study on visual focus of attention recognition from head pose in a meeting room', in St. Renals, S. Bengio and J.G. Fiscus (Eds.): *Machine Learning for Multimodal Interaction*, Vol. 4299 of *Lecture Notes in Computer Science*, pp.75–87, Springer, Berlin Heidelberg.

Bewick, V., Cheek, L. and Ball, J. (2004) 'Statistics review 10: further nonparametric methods', *Critical Care*, Vol. 8, No. 3, pp.196–199.

Camurri, A., Dardard, F., Ghisio, S., Glowinski, D., Gnecco, G. and Sanguineti, M. (2013) 'Exploiting the Shapley value in the estimation of the position of a point of interest for a group of individuals', *Procedia – Social and Behavioral Sciences*, to appear.

Castellano, G., Mortillaro, M., Camurri, A., Volpe, G. and Scherer, K. (2008) 'Automated analysis of body movement in emotionally expressive piano performances', *Music Perception*, Vol. 26, No. 2, pp.103–120.

Chadefaux, D., Wanderley, M., Le Carrou, J-L., Fabre, B. and Daudet, L. (2012) 'Experimental study of the musician/instrument interaction in the case of the concert harp', in *Proc. of the 11th Congrès Français d'Acoustique and the 2012 Annual IOA Meeting*, pp.1645–1650.

D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., Aloimonos, Y. and Fadiga, L. (2012) 'Leadership in orchestra emerges from the causal relationships of movement kinematics', *PLoS one*, Vol. 7, No. e35757, pp.1–6.

Dahl, S., Bevilacqua, F., Bresin, R., Clayton, M., Leante, L., Poggi, I. and Rasamimanana, N. (2009), 'Gestures in performance', in Godøy, R. I. & Leman, M. (Eds.): *Musical Gestures: Sound, Movement, and Meaning*, pp.36–68, Routledge, New York.

Davidson, J.W. (1993) 'Visual perception of performance manner in the movements of solo musicians', *Psychology of Music*, Vol. 21, No. 2, pp.103–113.

Davidson, J.W. (1994) 'What type of information is conveyed in the body movements of solo musician performers?', *J. of Human Movement Studies*, Vol. 6, pp.279–301.

Gnecco, G., Badino, L., Camurri, A., D'Ausilio, A., Fadiga, L., Glowinski, D., Sanguineti, M., Varni, G. and Volpe, G. (2013) 'Towards automated analysis of joint music performance in the orchestra', in *Arts and Technology, 3rd Int. Conf., ArtsIT*, Milan, Italy, March 21–23, Revised Selected Papers, Vol. 116 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST) Series*, pp.120–127, Springer, Berlin Heidelberg.

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M. and Scherer, K. (2011) 'Toward a minimal representation of affective gestures', *IEEE Trans. on Affective Computing*, Vol. 2, No. 2, pp.106–118.

Glowinski, D., Gnecco, G., Camurri, A. and Piana, S. (2013) 'Expressive non-verbal interaction in string quartet', in *Proc. 5th IEEE Int. Conf. on Affective Computing and Intelligent Interaction (IEEE ACII)*, to appear.

Palmer, C., Koopmans, E., Carter, C., Loehr, J.D. and Wanderley, M. (2009) 'Synchronization of motion and timing in clarinet performance', in *Proc. 2nd Int. Symp. on Performance Science*, pp.159–164.

Shaffer, J.P. (1995) 'Multiple hypothesis testing', *Annual Review of Psychology*, Vol. 46, pp.561–584.

Stiefelhagen, R. (2002) 'Tracking focus of attention in meetings', in *Proc. 4th IEEE Int. Conf. on Multimodal Interfaces*, pp.273–280, IEEE.

Stiefelhagen, R. and Zhu, J. (2002) 'Head orientation and gaze direction in meetings', in *CHI '02 Extended Abstracts on Human Factors in Computing Systems, CHI EA '02*, pp.858–859, ACM, New York, NY, USA.

Stiefelhagen, R., Yang, J. and Waibel, A. (2002) 'Modeling focus of attention for meeting indexing based on multiple cues', *IEEE Trans. on Neural Networks*, Vol. 13, No. 4, pp.928–938.

Varni, G., Volpe, G. and Camurri, A. (2010) 'A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media', *IEEE Trans. on Multimedia*, Vol. 12, No. 6, pp.576–590.

Wanderley, M.M. (2002) 'Quantitative analysis of non-obvious performer gestures', in *Proc. of the Gesture Workshop*, Vol. 2, pp.241–253.

Whitley, E. and Ball, J. (2002) 'Statistics review 6: nonparametric methods', *Critical Care*, Vol. 6, No. 6, pp.509–513.