Ching-Hua Chuan Feb 22 2006

Synopsis and Response Report of Visualizing Expressive Performance in Tempo-Loudness Space by Jörg Langer and Werner Goebl

This paper presents an integrated display of tempo and loudness variations in expressive music performances. The visualization technique analyzes music performances in MIDI files and audio recordings, representing the smoothed information on a two-dimensional space of tempo and loudness. A trajectory generated by the visualization can be interpreted as an intrinsic performance path, and snapshots of these trajectories can be used for detailed performance analyses. Two sets of visualizing results are demonstrated in the paper: two expert performances of Chopin's E major Etude (op. 10, No. 3) compared with the one performed by Maurizio Pollini, and an algorithmic performance of Schubert's G flat major Impromptu (D. 889, No. 3) compared with the one performed by Alfred Brendel.

In the introduction, Langer and Goebl explain their focus of performance research and the motivation of their visualization technique. Performance research in the last decade mainly concentrate on expressive piano performance since its expressive parameters are relatively few (timing, dynamics, and articulation), and are relatively easy to obtain. Most of these studies focus on one of these parameters, timing specifically. However, both performers and listeners experience these parameters in an integral way. In order to clarify the dependence between these expressive parameters, Langer and Goebl design a visualization program that can display tempo and loudness simultaneously, synchronized with the sound of the performance.

Their program can take MIDI files and audio recordings as inputs. For MIDI files, each performed onset is mapped to a symbolic score of the given piece so that the *track level* (an unit of score time, e.g., quarter-note, eighth-note) can be automatically determined. For audio recordings, an interactive software, automatic beat detection developed by Dixon in 2001, is used to generate basic information of beats. Then the timing data of audio recordings is adjusted manually to remove errors. The loudness information of MIDI and audio is obtained from a recording of the played back MIDI file or the audio file, using Zwicher's loudness model in Matlab which generates loudness envelope in *sone*. A smoothing process using overlapped Gaussian windows is then applied to tempo and loudness obtained in previous steps to produce smooth motion in the two-dimensional visualization space.

To exemplify the properties of the visualization technique, Langer and Goebl examine several performances of the first 21 bars of Chopin's Etude op. 10, No. 3. The choice of this piece is based on its homogeneous texture – sixteenth notes go through the whole excerpt. Using the visualization technique, the authors are able to make many interesting observations regarding the similarities and differences between the performances. The most noteworthy observation is that at the beginning of a phrase, the performers are likely to increase the tempo first and then the loudness. Langer and Goebl further apply the technique to elucidate relations between tempo and dynamics. They visualize two performances (a hybrid performance by Windsor and Clarke and a performance by Alfred Brendel) of Schubert's G flat major Impromptu D.889 No.3. The hybrid performance shows a diagonal line in the tempo-loudness space, which demonstrates 'the faster, the louder' rule proposed by Todd, although the result is not a perfect diagonal line. However, Brendel's interpretation indicates a asymmetric phrase concept, which disobey Todd's model of symmetrical grouping structure.

This visualization technique definitely provides a handy tool for studying expressive performances. Tempo and loudness are the most apparent parameters of performances. Although these two are not descriptive enough, they are representative for the first visualization tool. In my opinion, this version of the technique (they propose an updated one later) has four drawbacks. First, the timing data of audio recordings needs to be adjusted by hand. Secondly, the size of the processing window depends on onset-score match (need to know the score). Thirdly, as the authors also indicated, the smooth function shifts turning points or peaks, resulted in a less sensitive response or mismatch to listeners. At last, the representation makes it harder to trace the dynamic changes among the whole piece.