

Ananda Jacobs
ISE 599
April 8, 2004
Paper review

An Abstract Control Space for Communication of Sensory Expressive Intentions in Music Performance by Sergio Canazza, Giovanni De Poli, Antonio Roda, and Alvis Vidolin

After touching upon the importance of emotional and expressive quality in musical performances, Canazza et al. describe two approaches to modeling expressive intentions in music. One is categorical, the other is dimensional. Categorical models are used to describe which emotion is being conveyed, using descriptive words like “happy” or “afraid.” It is assumed that all emotions can be derived from a basic set of emotions. The other model, dimensional, uses a more intuitive method for describing emotions. The authors describe this as having two axes: valence, which refers to positivity or negativity of an emotion and activation level, or the tendency of the emotion to incite a person to action. An example of this model is the PerformanceWorm.

The authors justify using the activation-evaluation space in their mention of Plutchik’s work. Plutchik, said to hold a relatively conservative view of what emotion is, defines hundreds of working emotional adjectives. These can be difficult to implement into computerized models, thus, a system with only two axes would be much more suited to emotional modeling. I also tend to agree with this view. When asking listeners to rate a piece of music in terms of its emotional content, it can be quite cumbersome and debilitating for the listener to have to choose a particular word. Words can have very different connotations for each individual. Also, perhaps most importantly, music is temporal. Therefore, if listeners are asked to rate the emotional content of music in real-time, that is, as each emotion occurs, there is scarcely any time to stop and think about adjectives. Rather, an easier way would be to “draw” what happens on continuous scales of valence/activation, for example.

The present study defines four layers of performance understanding. The first is physical signals (audio), the second is low-level features, the third is mid-level features and patterns, and the fourth is concepts and structures, built from the lower levels. The authors’ description of these four levels was not nearly as clear as it should have been. It was unclear exactly how these distinctions of levels are made. The diagram in Figure 2 was not very helpful in mapping how the levels interact.

In the current experiments, the authors looked at various ways of rating emotions. Subjects (both musicians and non-musicians) were asked to rate pieces based on brightness, heaviness, softness, etc. The musicians tended to cluster together in agreement. A three-dimensional space was also used, where listeners imagined a space including Kinematics (tempo), and Energy. In general, subjects did in fact recognize the performers intentions when the performer altered the supposed emotional content of the piece.