

Yu Shiu

ISE 575 (Feb 17, 2005)

Review of “Polyphonic Music Transcription by Non-negative Sparse Coding of Power Spectra”

In this paper, the authors proposed a very mathematical system to tackle the problem of transcribing polyphonic music. The basic idea is to decompose the training data (one or more music pieces) into a collection of atomic features (the used feature is spectral profile), which are put in a matrix called as “dictionary matrix”. During the process of music transcription, the features of a testing music piece are represented as a linear combination of the atomic features with appropriate weight. In mathematic way, the description above could be represented as $x = As$ where x is n by 1 column vector, which represents features of music signals, A is the n by m matrix, each column of which represents one atomic vector, and s is m by 1 column vector, which presents the weight of the atomic features. In training phase, music signals x is provided to learn the atomic features in A . In transcription phase, the testing music signal x' is encoded optimally with s .

An assumption of statistical independence among the elements of s implies that, in order to make one note decision, few atomic features need to be examined. That is, “most observations can be encoded with only a few non-zero elements of s .” The assumption above leads to the use of Independent Component Analysis (ICA) or sparse coding. In addition, the authors articulated on the estimation of subspace variance and proposed generative model to explain the observation signal x given multiplicative noise v . For this generative model, the techniques for learning the dictionary matrix are also presented.

In the experiment, first of all, dictionary matrix A is initialized. It consists of roughly pitched spectra on a quarter-tone spacing. Second, training is performed by alternately optimizing the weight vector s and dictionary matrix A . Then, pitched atomic spectra are picked by visual inspection and assigned the pitch values. Atomic spectra with the same pitch are grouped together. The “activity” of certain pitch is calculated as the sum of the elements in s that belongs to the same pitch group. The result in figure 4 for a piano piece by Bach seems good but not finished. Figure 4 only presents the activity of each pitch group to the testing music piece.

I have some opinions about this paper. First, it seems very mathematical to me to understand. It has much advanced linear algebra material in merely 8-page paper. I could realize fully the content of this paper only if I am familiar with the many of important papers in the reference. When the authors mention something mathematical, they always advise the reader to see some papers in reference. Sadly, it seems inevitable because the piano transcription problem is already changed into a very linear-algebra-like basis decomposition problem. Second, as I know, there are some other papers in recent years dealing with the same problem, piano transcription. Is there any evaluation method to compare among different papers? For this paper, it might be an advantage to deal with polyphonic music instead of monophonic. How does the method in this paper work when dealing with other musical instruments? Piano music is always the first target for music transcription because it has a very clear onset when certain note is played. It is hard to predict how the result will be when dealing with other musical instruments. Third, this paper seems unfinished. It proposes a good method to detect the “activity” of certain pitch group and leave at that point. The problem of “music transcription” is supposed to include the detection of the starting and ending time for all notes, grouping of notes at the same time and some other characteristics like slur or measure. Current results show the result of intermediate process, which are like MIDI data if the authors further apply appropriate peak-picking algorithm.