

## Review of Separating Voices in Polyphonic Music by Chew and Wu

The paper presents a contig-based mapping approach to voice separation. The approach only uses pitch height and event boundaries, with no more requirements on user-defined parameters. The author also proposes three metrics for quantifying the quality of voice separation result.

The incentive of this research is that many music applications require the matching of monophonic queries to polyphonic databases. The one to do that is first separate each piece of voice into its component voices prior to matching the melody. Also there are requirements in automatic detection and categorization of music by meter, voice separation is valuable tool in these musical retrieval solutions.

Main principles used in this contig-based mapping approach are: The pitch proximity principle and the stream crossing principle. The pitch proximity principle claims “the coherence of an auditory stream is maintained by close pitch proximity in successive tones within the stream”. That is the key idea the author adopted by using shortest distance to connect voices between segments. The stream crossing principle claims that concurrent ascending and descending streams of the same timbre will avoid crossing each other. There are two main assumptions also adopted in this model. First, each voice can only sound at most one note at a time. Second, there is some time that all voices will sound synchronously (maximal voice contig). The second one is more important because the model will catch these maximal voice contig as the start point of voice separation process. The total number of voices are distinguished at these points, and then using the minimum distance between each voice in maximal voice contig with the neighbor segment voice to concatenate each voice.

The process of voice separation starts in segmentation procedure, which is based on voice count. The output of this procedure is contigs of original music that voice count remains constant in each contig. After the segmentation procedure, maximal voice contigs seed the connection process. All fragments in each maximal voice contig are ordered by pitch height and assigned voice numbers according to their ordering. The author argue that because the number of voices is relatively small, thus there is no intelligent algorithm used in computing the smallest distance between segments, enumeration is instead used.

The approach is implemented on VoSA, and three data sets are tested on the system. The performance of this model is evaluated by the average fragment consistency, the correct fragment connection rate and the average voice consistency. The result shows a good performance of this model.

Several things need to be considered. First, the performance segmentation procedure is very important because it decides the maximal voice contigs. Second, seems it is not sufficient to adopt only the smallest distance in connecting the segments. Also different type of music will have different connection pattern. To improve this, one possible way is to adopt some intelligent search like A\* search, which is a cost-based search method to

Haojun Wang

[haojunwa@usc.edu](mailto:haojunwa@usc.edu)

Week 7 Review

predict the minimum cost. Another possibility is to use some heuristic capturing the connection type on different genre of music. Also current method assumes that the number of voices is relatively small, but that's not always the case. To handle a large number of voices at the same time, more efficient algorithms will be need.