Separating Voices in Polyphonic Music:
A Contig Mapping Approach (Chew, Wu) *Naru Sundar, ISE 575*

This paper describes a voice separation algorithm based on the notion of contig mapping. Contig mapping is a DNA sequencing technique where overlapping fragments of a particular DNA sequence are reassembled to form the original sequence. The thrust of this paper involves constructing contigs of voice fragments and growing them to form the overall voices. The growing phase of the algorithm was implemented and shown to run in $O(n^2)$ which was better than the more complicated dynamic programming or heuristically driven stochastic methods of Temperley and Kilian & Hoos.

The algorithm was based on two main perceptual rules. The first was the pitch proximity rule which included aspects like avoiding wide pitch leaps and maintaining chordal relationships. The second perceptual rule was that of avoiding stream crossing encapsulated by avoiding shared pitches between voices and avoiding crossings with respect to pitch. These two rules represented assumptions about the tendencies of the input material. Two more assumptions were made. First, that each voice can sound at most one note at a time, and second that all voices will sound synchronously at some point in time.

The paper then defined a contig as a collection of overlapping fragments such that the overlap depth is constant. A fragment was defined as a sequence of notes in the same voice. The input piece was segmented into fragments based on either a change in voice count or a change in voice status. The fragments are then reassembled into contigs. Noting that all reassemblies are equally valid, a cost function was used to pick which fragments to merge. Given notes q_X and p_Y from fragments X and Y such that q and p are at the end and the beginning of X and Y respectively, the cost function was designed as follows. If the two notes are segments of the same longer note then the cost function is -2^{31}. If one of the two notes is empty (that is, no such note existed at that slot), then the cost is extremely high, 2^{31}. The remaining case is computed as the absolute difference between the two pitches. Intuitively the cost function merges fragments that are likely continuations of each other, and stays away from segments that have no adjacent notes in common. In the absence of either case, the pitch proximity is used as the cost selection.

As stated before, this is an $O(n^2)$ algorithm, mainly because we have to compute the cost function for each possible fragment pair. As the algorithm grows each contig, we add more and more fragments and finally we are left with the contigs which should represent the individual voices in the piece. The algorithm was evaluated using three measures. The first was average fragment consistency, which measured how well the fragments were identified. The second was average connection consistence which measured how well the fragments were correctly paired. The last was average voice consistency which measured how well the notes matched the actual voicing. It seemed that the latter was the most direct measure of success, whereas the former two were useful for identifying problem cases in more detail. The paper wasn't necessarily clear on whether the correct voicings were computed by hand or from some other source. I surmised that they were in a meta-data form that specifically identified voices since hand assigning of the voices seemed tedious.

The experimental results seemed to indicate that the algorithm performed quite well on the Bach contrapuntal pieces in the data set. One question that came to mind was whether the lack of stream crossing was a feature prominent to Bach's work or to contrapuntal work in general. The two main weaknesses noted was the rare case where shortest pitch distance combined two wrong fragments, and stream crossing due to quantization error. It might be possible to fix the quantization error by selectively not quantizing notes that quantize into aligned positions. Since inter-fragment notes don't matter, we could use the heuristic that a quantized note that overlaps an existing note would replace the next note in the fragment. We can store the original note information and map the note to voicing number for all notes post priori. The other issue with cases that break the pitch distance rule is more of an underlying problem since it deals with a break from the assumptions.