VoSA - Voice Separation Analyzer

1. Research Team

Project Leader: Prof. Elaine Chew, *Industrial Systems Engineering*

Graduate Students: Xiaodan Phillip Wu

2. Statement of Project Goals

The objective of the VoSA project is to develop a computer program to separate voices in polyphonic music using basic principles of music perception. The main applications of this research are:

- Music information retrieval: Most of the music we hear is polyphonic in nature, consisting of multiple voices sounding synchronously. Voice separation is necessary for matching melodic (single-voice) queries to polyphonic (multi-voice) databases, for example, in query-by-humming applications.

- Pitch spelling: accurate pitch spelling is dependent on contextual rules as well as melodic rules. A good voice separation algorithm will result in better implementation of voice-leading rules.

3. Project Role in Support of IMSC Strategic Plan

This research supports IMSC's research in music information management through systematic use of music perception principles to design robust algorithms for voice separation. This interdisciplinary project combines computer science and music perception and expands IMSC's use of creative technologies to computer-assisted research in the humanities. This project augment's the center's research agenda in information management, sensory interfaces and user centered sciences.

4. Discussion of Methodology Used

The Voice Separation Analyzer, VoSA, is developed under Java JDK1.4.2. It is platform-independent software, which could be run under Windows, Mac OS and UNIX. The current version of VoSA handles only MIDI file input.

Our approach takes advantage of the general music perception principles (see [1]) detailed below which we use as axioms for the algorithm:

1. All the voices will sound synchronously at some time.
2. Each voice can only sound one note at any time.
3. Voices tend to avoid big leaps in pitch.
4. Voices don't cross.

These axioms translate to the following algorithmic rules:

1., 2. & 4. → All the voices will sound synchronously in a well-behaved manner at some time. In these maximal voice contigs, we know the voice assignments for each note.

3. & 4. → We can use distance-minimizing rules to connect voices between contigs. The connections grow outwards from the maximal voice contigs where the voice assignments are most certain (see Figure 1).

The inspiration for this approach comes from the field of computational biology (in DNA sequencing) where fragments are assembled into overlapping contigs that need to be connected into a whole.

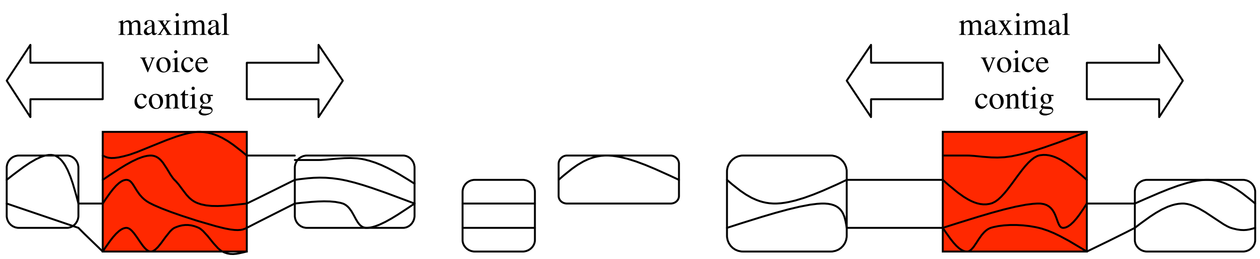

Figure 1: Minimum distance voice connections grow out from the maximal voice contigs.

5. Short Description of Achievements in Previous Years

N/A

5a. Detail of Accomplishments During the Past Year

1. The Voice Separation Analyzer, VoSA, a JAVA program, has been developed.
2. The voice separation technique, as an important part of the music retrieval system and of the human-machine interaction, was presented in the Ph.D. seminar in Daniel J. Epstein Department of Industrial Systems Engineering in November 2003.
3. VoSA was demonstrated at the Integrated Media Systems Center's Scientific Advisory Board/ Board of Councilors meeting on November 20, 2003.

<u>Graphical User Interface</u>

VoSA provides a graphical user interface for the user to analyze the performance of the voice separation algorithm. The GUI shows the voice count (lower part of Figure 2) at any point in time, the segmentation and voice fragments (upper part of Figure 2). Erroneous voice assignments are highlighted, thus displaying the result of the voice separation. Also included is a MIDI playback function.

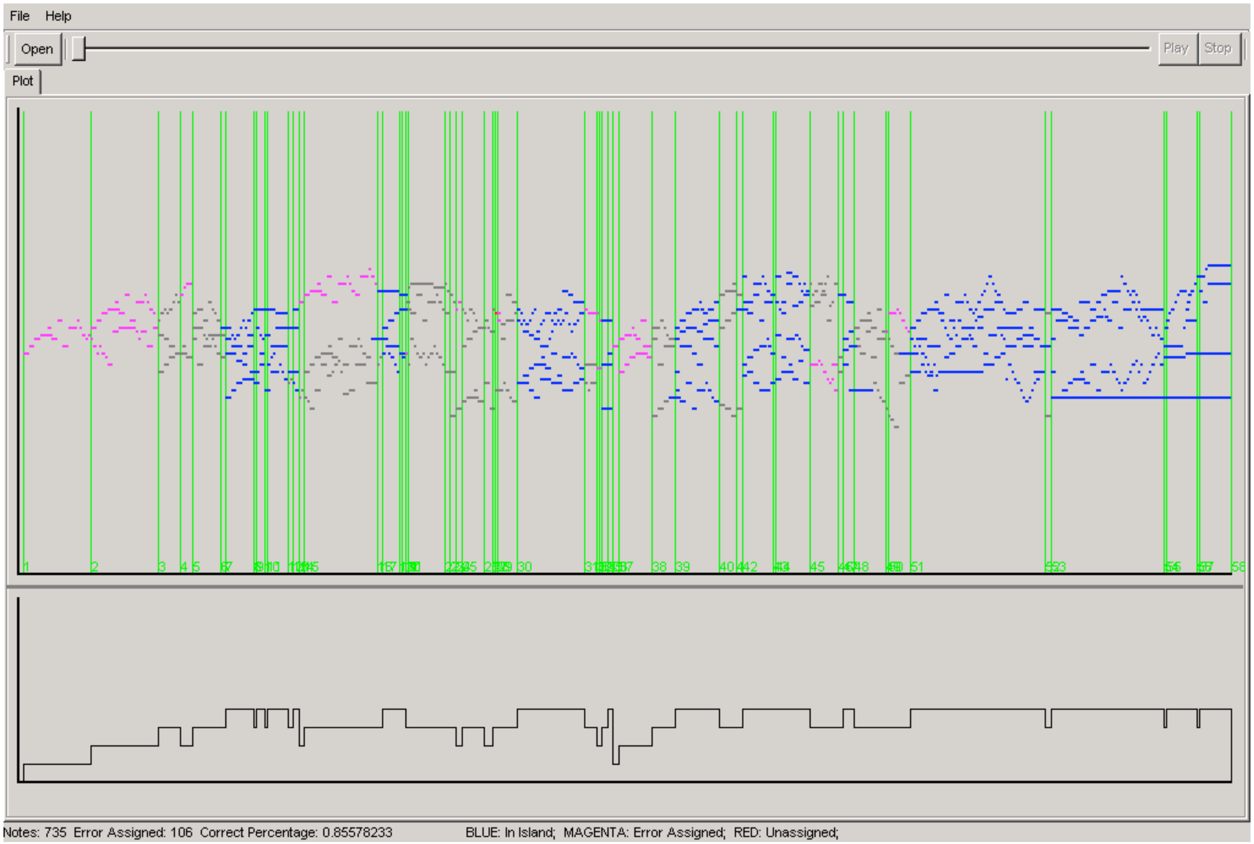

Figure 2: Screenshot of VoSA

Flowchart

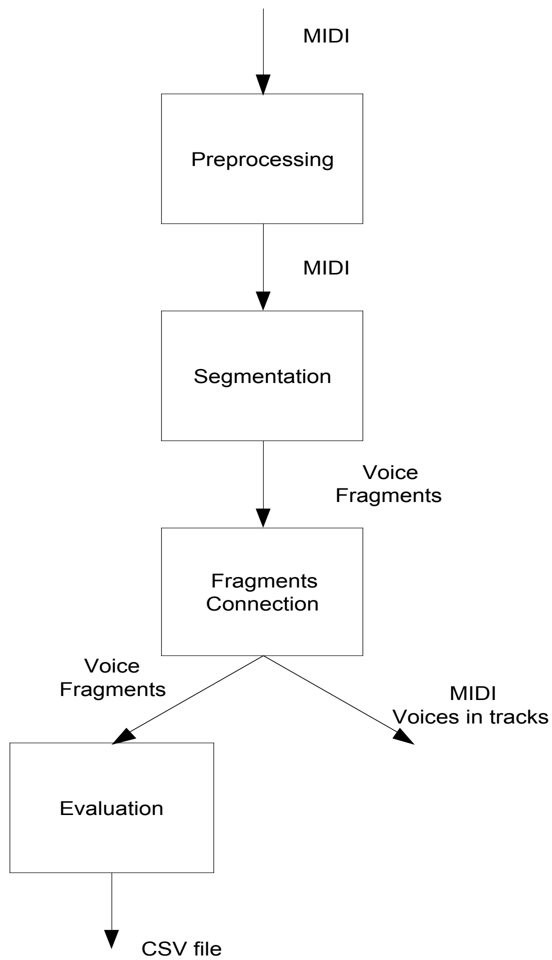

Figure 3: Flowchart for VoSA

Quantizing Technique

We pre-process the MIDI data to get clean input for the algorithm. Performance artifacts and rounding errors produce overlapping notes from the same voice or gaps in the data. A selective snap-to-grid procedure was developed to handle such situations so as to improve the algorithm's performance. A snap-to-grid was only invoked when the time difference between any two boundaries was less than a given threshold (we used 40ms).

Data Export

The voices separated by VoSA can be saved as MIDI files, which can be easily edited and analyzed by the third-party audio editing software. And the results of the evaluation process, which measure the performance of the algorithm, can be exported as worksheet files.

Performance of VoSA and Evaluation

VoSA has been tested on J. S. Bach's Inventions, Sinfoniettas and Fugues, and Shostakovich's Fugues. The average voice consistency in each fragment was well above 99% (that is to say, all notes in the same fragment are most likely from the same voice), and the correct connection rate was approximately 90% (the rate of connecting each fragment correctly to its contiguous strand). The high fragment voice consistency rate and the 10% connection error resulted in an average of 87% notes connected to the correct voices. The edit distance would be lower than the 87% correct assignment rate might indicate.

A description of our voice separation algorithm and the VoSA system has been submitted to the 2004 Computer Music Modeling and Retrieval Conference: Chew, Elaine and Wu, Xiaodan. "Separating Voices in Polyphonic Music: A Contig Mapping Approach." Submitted to the International Computer Music Modeling and Retrieval Conference, Esbjerg, Denmark, May 2004.

6. Other Relevant Work Being Conducted and How this Project is Different

Other researchers have studied the problem of voice separation: Huron [1] from a musicological point of view, and Kilian & Hoos [2] and Temperley [3] from a computational point of view. We merge the two views to produce VoSA, which incorporates musicological concepts in a computational framework.

Previous work has also been lacking in evaluation procedures. We map voice fragments to contigs and use maximal voice contigs to ensure the highest performance possible in voice assignments. This approach produces highly quantifiable results, thus presenting a rigorous way to measure the performance analysis of voice separation algorithms.

7. Plan for the Next Year

Possible avenues for further research include:

1. Determining when voice crossings should be allowed.
2. Determining if pattern matching can improve the performance of VoSA.
3. Applying the results of VoSA to music retrieval systems.

8. Expected Milestones and Deliverables

1. VoSA, the software and manual,
2. Algorithms for voice separation research.
3. More musicological tests to validate approach.

9. Member Company Benefits

See Section 8.

10. References

[1] David Huron, Tone and Voice: A Derivation of the Rules of Voice-leading from Perceptual Principles, *Music Perception*, Vol. 19, No. 1 (2001) pp. 1-64.

[2] Kilian, J. & Hoos, H. (2002) "*Voice Separation – A Local Optimization Approach*." In Proceedings of the 3rd International Conference on Music Information Retrieval, p.39-46.

[3] David Temperley, *The Cognition of Basic Musical Structures*, 2001, pp. 85-114