Development and Evaluation of a Method for the Separation of Musical Duet Signals

by
Robert C. Maher
Computer Music Project
University of Illinois
1114 W. Nevada
Urbana, IL 61801
(217) 244-1207

ABSTRACT

This paper describes a signal processing approach to the problem of analysis, identification, tracking, and resynthesis of a specified timbre from a digital recording of a musical duet. Analysis is performed using a quasi-harmonic sinusoidal representation of the constituent signals, based on short-time Fourier transform (STFT) methods. The procedure is tested using both real and artificial test signals. Applications include signal restoration, musicology, digital editing and splicing, musique concrete, noise reduction, and time-scale compression/expansion.

INTRODUCTION

This paper describes an approach for extracting a single musical voice from a recorded duet. Unlike many previous investigations which emphasize signal segmentation for transcription into musical notation [cf., for example, Moorer, 1975; Schloss, 1985; Chowning and Mont-Reynaud, 1986; etc.], the goal of the research reported here has been to resynthesize the separate signals while retaining as much of the original material as possible. For example, we might like to extract the 'violin' part from a monaural recording of a violin and cello duet. Some approaches related to this task have been reported for simultaneous speech [cf., for example, Everton, 1975; Parsons, 1976; Naylor and Boll, 1987; Lee and Childers, 1988; Danisewicz and Quatieri, 1988; etc.] and for musical signals [Stockham, 1971; Wold and Despain, 1986].

DESCRIPTION

In order to limit the complexity of the separation procedure, the following rules govern the input signal: First, only two separate voices (musical duets) are allowed. Second, we wish to determine the approximate frequencies of each partial of the two musical voices, so the duet voices are assumed to be quasi-harmonic. Third, the fundamental frequencies of the two voices are not allowed to cross during the signal segment under consideration. Finally, the signal is assumed to be free from reverberation and other correlated noise sources, at least to the extent that these features may not represent additional background signals in the recording, violating the duet assumption.

Two fundamental research questions guide this project:

1) How may we automatically obtain accurate estimates of the time-variant fundamental frequency of each voice from a digital recording of a duet?

2) Given time-varying fundamental frequency estimates of each voice in a duet, how may we identify and separate the interfering partials (overtones) of each voice?

Question (1) treats the problem of estimating the time-variant frequencies of the spectral components contributed by each voice, while question (2) involves the fundamental constraints on simultaneous time and frequency resolution in the analysis process. Note that it is possible to confront question (2) without solving question (1) if the fundamental frequencies of the duet can be identified manually.

The complete separation procedure may be summarized as follows:

1) The duet signal is provided, including specification of the non-overlapping fundamental frequency ranges for each voice of the duet.

2) The short-time Fourier transform (STFT) is calculated for the input signal, and estimates of the two fundamental frequencies are determined using a "two-way mismatch" error procedure.

3) The STFT is reduced to a sinusoidal representation [McAulay and Quatieri, 1986; Smith and Serra, 1987]. Any partials with frequency spacing smaller than the resolution bandwidth of the analysis are identified using the frequency estimates from step 2. These colliding partials may be corrupted by crosstalk, so they are marked for further processing. The remaining uncollided partials are segregated into two lists according to the harmonic series of the two fundamental frequencies.

4) For the marked (colliding) partials of step 3, several different strategies are employed to estimate the individual components from the composite measurement.

5) The repaired partials from step 4 are inserted into the appropriate segregated lists from step 3, and the separated signals are regenerated using an additive synthesis procedure.

The procedure has been successful for the separation of several duet examples. However, the separated signals contain occasional audible artifacts, indicating the need for further development. Audio examples of the separation process will be presented.

REFERENCES

J. Chowning and B. Mont-Reynaud, "Intelligent analysis of composite acoustic signals," (NSF proposal) Stanford University, Dept. of Music Rep. STAN-M-36, 1986

R.G. Danisewicz and T.F. Quatieri, "An approach to co-channel talker interference suppression using a sinusoidal model for speech," MIT Lincoln Laboratory Technical Report 794, February 1988

J. Everton, "The Separation of the Voice Signals of Simultaneous Speakers," PhD thesis, Dept. of Computer Science, Univ. of Utah, 1975

C.K. Lee and D.G. Childers, "Cochannel speech separation," J. Acoust. Soc. Am., vol. 83, no. 1, pp. 274-280, January 1988

R.J. McAulay and T.F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, no. 4, pp. 744-754, Aug. 1986

J.A. Moorer, "On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer", PhD dissertation, Stanford University Dept. of Music Rep. STAN-M-3, 1975

J.A. Naylor and S.F. Boll, "Techniques for suppression of an interfering talker in co-channel speech," Proc. 1987 ICASSP, vol. 1, pp. 205-208, 1987

T.W. Parsons, "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am., vol. 60, no. 4, pp. 911-918, Oct. 1976

W. A. Schloss, "On the Automatic Transcription of Percussive Music--From Acoustic Signal to High-level Analysis," PhD dissertation, Stanford University, Dept. of Music Rep. STAN-M-27, 1985

J.O. Smith and X. Serra, "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," ICMC-1987: Proc. Int. Computer Music Conf., pp. 290-297, Computer Music Assn., August 1987

T.G. Stockham, Jr., "Restoration of old acoustic recordings by means of digital signal processing," Preprint, 41st Convention, Audio Engineering Society, New York, October 1971

E. H. Wold and A. M. Despain, "Parameter Estimation of Acoustic Models: Audio Signal Separation," Proc. 1986 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE, Sept. 1986