Computer Engineering Group

Assistant Professor ChingShun Lin

Ph.D., Electrical Engineering, University of Southern California, U.S.A.

Fields of study: Immersive System, Multimedia Application, Spatial Audio Rendering, Multichannel Audio Signal Processing, Virtual Technology, Transformation between Image and Graphics, and Soft Computing.

Keywords: Virtual Microphone/ Loudspeaker, Audio-Visual Fusion, Soft Computing, Biomedical Pattern Recognition, and Machine Learning.

URL: http://homepage.ntust.edu.tw/chingshl/

Email: CHINGSHL@mail.ntust.edu.tw

Phone: 886-2-27376372(Voice), 886-2-27376424(Fax)

1. The Subject and Aims of Research

The goal of immersion system is to create the illusion of proximity for people physically in different areas. To achieve this, it is essential to pick-up and regenerate all the crucially visual and aural information that is perceptible by human senses. The major research directions would be the development of embedded multimedia system with emphases on immersive audio techniques and their applications:

- embedded multimedia system development,
- audio-visual fusion by machine learning,
- application of virtual technology to the Internet,
- image retrieval using graph modeling,
- modeling of augmented reality and immersive system,
- multichannel audio rendering and visualization,
- virtual acoustics for high-definition audio delivery,
- room acoustics/ equalization/ calibration,
- hand-free noise-control earphones and microphones,
- biomedical pattern recognition,
- soft computing,
- and their consumer electronics applications.

Applications of immersive audio include teleconferencing, home entertainment, air traffic control, pilot warning and guidance, distance learning, professional sound editing, and assisting people who are visually or aurally impaired.

2. Related Recent Research Topics

(1) EMD based near-field equivalence sound source imaging system:

The conventional microphone array near-field Fourier acoustic holography using Fast Fourier Transform (FFT) is able to efficiently reconstruct sound field and acquire an image of noise distribution. However, Fourier transform causes measuring error in practical applications, and people have to select primary frequency for observing sound field holography based on the spectrum of source signal. In this project, we use the ensemble empirical mode decomposition (EEMD) owing to its adaptive basis and low mode-mixing, which are able to decompose multiple sound sources in the time domain and acquire instantaneous frequencies by intrinsic mode functions (IMFs). Prior information about the primary frequency is not necessary by this approach that makes the simultaneous observation of each source possible. In addition, EEMD sound source imaging approach may be integrated into near-field equivalent source imaging (NESI) system, which includes a virtual microphone technology generally used for sound field image enhancement. Figure below shows the identification results of mixed signal of frequencies 1kHz and 450Hz.

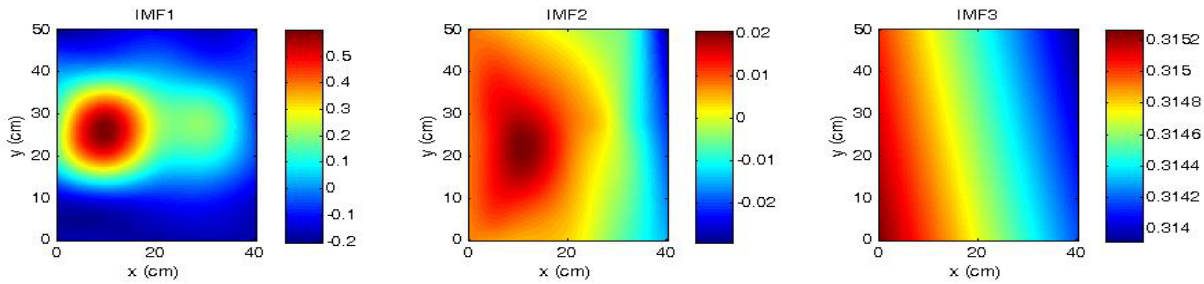

Fig. 1. Sound pressure images on the reconstruction surface for 1.2kHz and 0.5kHz signals extracted from the first three IMFs using the 2D EMD.

(2) Fast music information retrieval using PAT tree based dynamic time warping:

Recently, music retrieval is getting more and more attentions, especially for the query by humming (QBH), which is straightforward and convenient. However, the difference between humming phrase matching and the precision of singer humming increase the difficulty of retrieval. Apart from retrieval precision, retrieval time is another issue should be taken into consideration. In this research, we use two-stage searching approach for music retrieval. Since PAT tree (variant of Patricia tree, practical algorithm to retrieve information coded in alphanumeric) shows an excellent performance in matching partial sequence, in the first stage, the simply reduced note interval is adopted as a key phrase of PAT tree for fast indexing the candidate regions from the feature sequences in the database. In addition, the searching of partial sequences suffering from the insertion, deletion and transposition errors works well in the indexing of reduced note interval. The resulting diversity of candidates may be further reduced via a more complicated dynamic time warping (DTW) comparison in the second step. Our approach not only avoids the exhausted computation in the simple DTW approach, but preservers the fault-tolerance capability in the matching process.

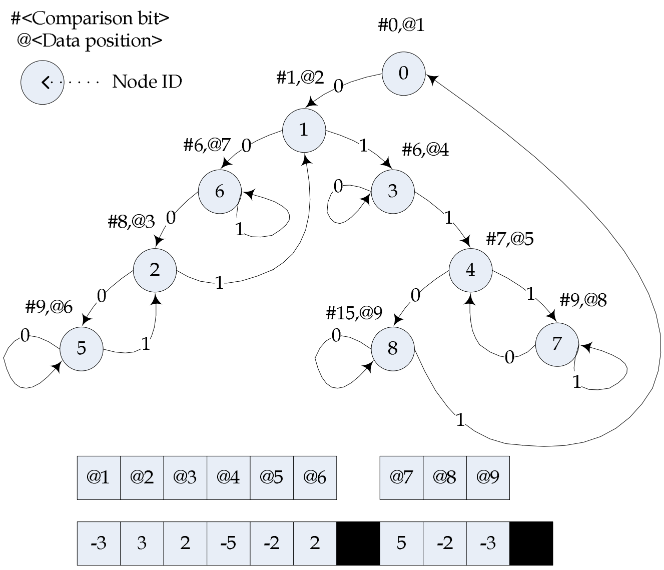

Fig. 2. Example of the structure of PAT tree and its searching instance.

(3) Instantaneously warping interval computation:

The development of fuzzy sets may be carried out on a stepwise and offline basis. However, this mechanism doesn't instantly reflect the real situation the fuzzy-based system is processing. Moreover, the intrinsic discontinuity of the stepwise functions usually results in the poor modeling of the processing data. This shortcoming may be removed through several more specific fuzzy sets that are properly associated with the fuzzy sets formed at the higher level. To fit the complexity of the granular descriptors, we propose a nonuniformly spaced interval calculation algorithm functioning in an instantaneous manner for more sophisticated fuzzy sets without losing reasonable interpretation. In this research, several examples are also performed to verify the usability of this EMD-based scale warping approach.

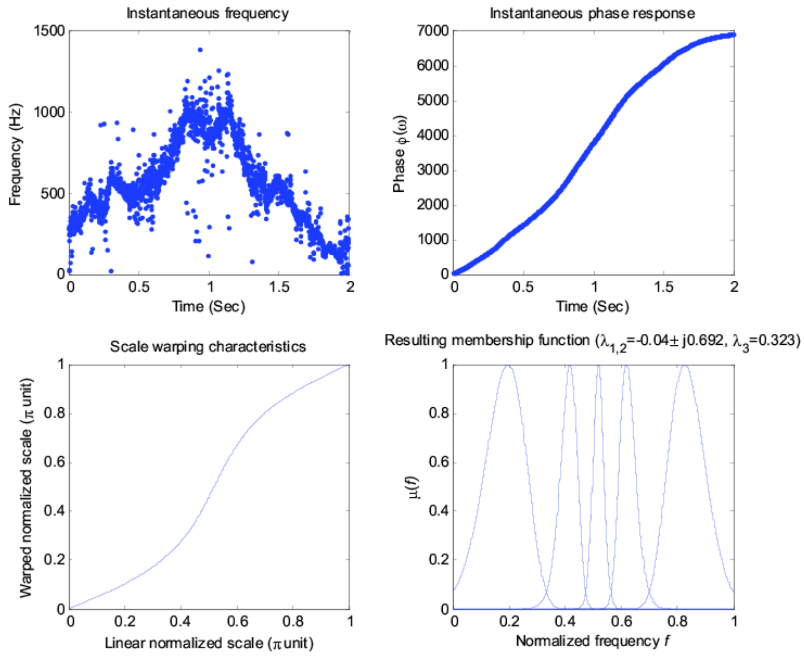

Fig. 3. Membership function for unsymmetrical chirp data with various warping coefficients after the third order all-pass transformation.

(4) Fast ensemble empirical mode decomposition for speech-like signal analysis:

Empirical mode decomposition (EMD) is one of the useful approaches for processing nonlinear and nonstationary signals. However, its shortcomings include mode mixing and end effects that usually appear in the decomposed bands. Although a noise-assisted data analysis (NADA) called ensemble empirical mode decomposition (EEMD) has been proposed to circumvent this problem, doing so also results in an inevitably long computation for alleviating the mode mixing. In this research, we use shaped noise instead of white noise as a disturbance for a fast convergence of EEMD. The signal-spectrum dependent noise (SSDN) is able to effectively randomize the targeted signal in time domain, and then significantly save the superfluous calculation around the corresponding energy-free frequencies. The experimental results also show that both pink noise and brown noise outperform the white noise in terms of computation for the EEMD of speech-like signal.

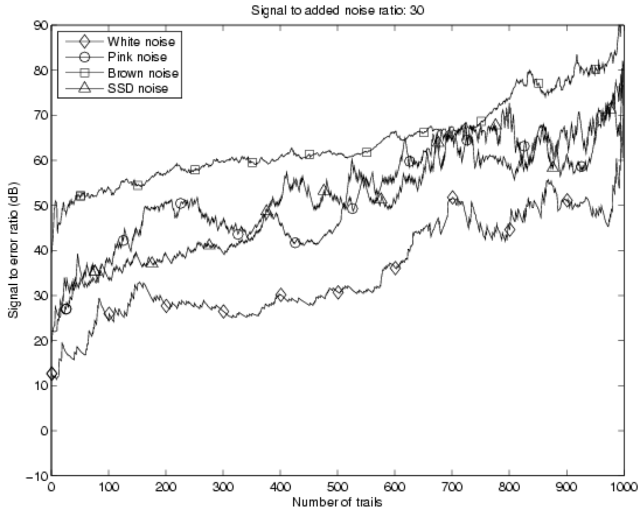

Fig. 4. Convergence rates for speech-like signal decomposition using different shaped noises.

(5) Spectrogram image encoding based on dynamic Hilbert curve routing:

In this research we propose an image-based biological classification system that can identify different creatures via their sounds. The overall system involves the relative spectral transform-perceptual linear prediction for spectrogram image extraction, cosine similarity measure for feature matching, dynamic Hilbert curve for spectrogram routing, and Gaussian mixture model for 1-D spectrogram classification. As an example of our approach, results for honk, dolphin, and whale classification are presented. This method works well on a wide variety of biosounds, especially for the highly self-repeated ones. Applications of this approach include biological signal analysis and spectrogram library establishment.

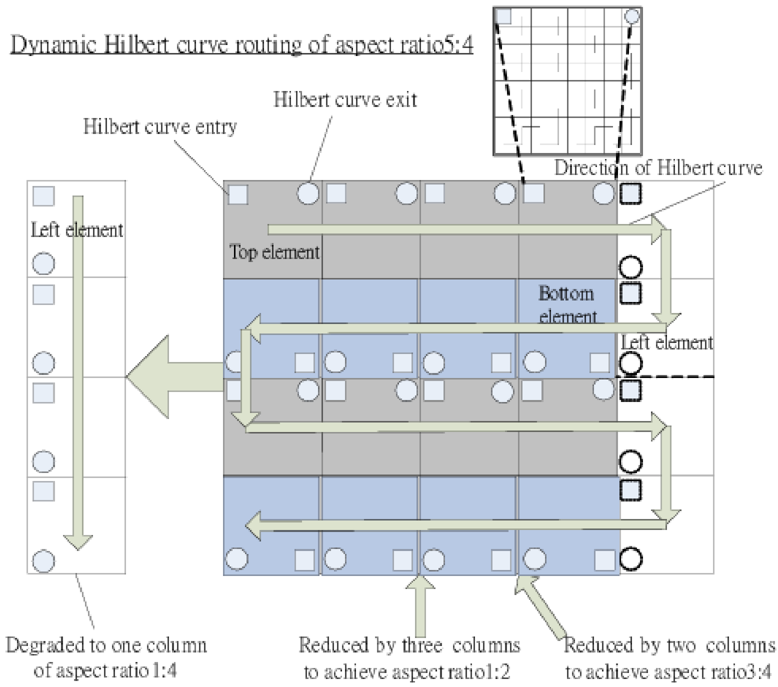

Fig. 5. Different routing strategies for dynamic Hilbert curve spectrogram image encoding.

(6) Multichannel audio rendering:

Multichannel audio is rapidly becoming the next step in the evolution of musical reproduction. Increasing the number of channels produces a more pleasing and immersive experience for listeners. However, only a relatively small number of multichannel musical recordings have been made. Furthermore, many older or historical recordings are seemingly destined to remain as one or two-channel renditions. One approach to up-converting these recordings to multichannel versions is to synthesize the necessary microphone signals that would normally be used to make a multichannel recording. We have proposed a morphing algorithm to synthesize signals based on the pair-wise reference signals. This approach is similar to the graphic morphing that allows signal not only being transformed from one to another, but incorporating with the appropriately acoustic characteristics.

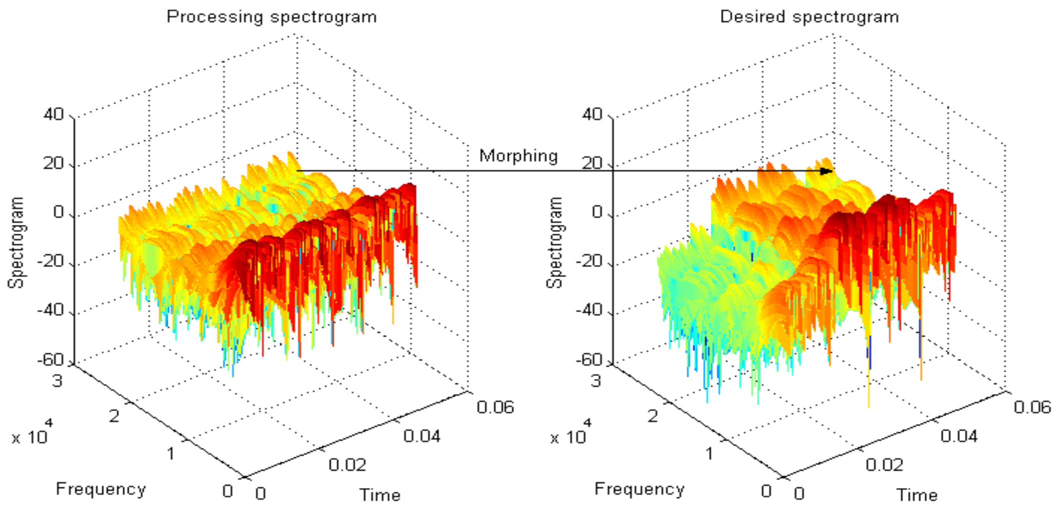

Fig. 6. Multichannel recording synthesis by spectrogram morphing.

(7) Removing additive noise via neuro-fuzzy based reinforcement learning:

In this research, a systematic treatment for developing a noise removal system based on the fundamental principle of reinforcement learning and fuzzy cerebellar model articulation controller (FCMAC) is presented. The proposed system improves its performance over time through two mechanisms. First, the modified stochastic real-valued (SRV) algorithm, learning from its own mistakes via the reinforcement signal and reinforcing its action to improve future performance, is used for searching the optimal noise spectrum for the overall training system. Second, system states associated with the positive reinforcement are memorized by FCMAC-based neurons where in the future, similar states will share the experiences already stored there and then lead the action to a more positive situation. In this work, FCMAC's intrinsically poor approximation of rapidly varying functions is solved by taking the complex semi-cepstrum. In addition, the FCMAC provides an improvement in accuracy of function approximation without losing the property of generalization, which makes the high fidelity DSP possible.

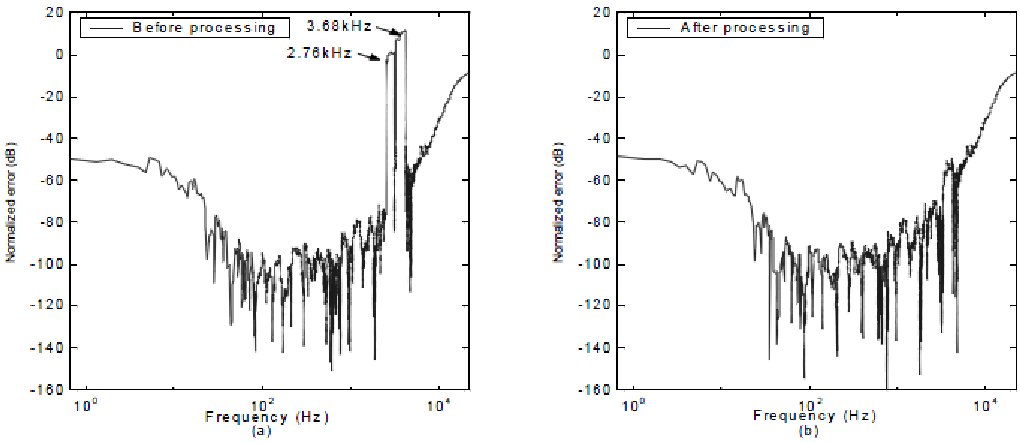

Fig. 7. Comparison based on 1/3 octave smoothing. (a) Normalized error between colored signal and original signal. (b) Normalized error between output signal and original signal.

3. Selected Publications and Projects

[1] **C. S. Lin** and Y. C. Chao, "Empirical mode decomposition based near-field equivalence source imaging system for sound identification," *8th International Conference on Information, Communications and Signal Processing*, Singapore, Dec. 2011. (EI)

[2] **C. S. Lin**, S. J. Hong, D. L. Shih, and Z. C. Cheng, "Instantaneously warping interval computation using low level intrinsic mode functions," *IEEE International Conference on Granular Computing*, Kaohsiung, Taiwan, Nov. 2011.

[3] **C. S. Lin**, J. S. Wang, and Z. C. Cheng, "Fast ensemble empirical mode decomposition for speech-like signal analysis using shaped noise addition," *4th International Conference on Interaction Sciences: IT, Human and Digital Content*, pp. 112-117, Busan, Korea, Aug. 2011. (EI)

[4] **C. S. Lin** and D. R. Wang, "Spectrogram image encoding based on dynamic Hilbert curve routing," *International Conference on Image Processing Theory, Tools, and Applications*, Paris, France, Jul. 2010.

[5] **C. S. Lin** and C. Kyriakakis, "Synthesizing multichannel recordings based on adaptive spectrogram morphing," *4th International Conference on Innovative Computing, Information and Control*, pp. 156-159, Kaohsiung, Taiwan, Dec. 2009.

[6] **C. S. Lin** and C. Kyriakakis, "Removing additive noise via neuro-fuzzy-based reinforcement learning," *Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 1026-1037, Aug. 2008. (SCI, EI)

[1] National Science Council — Audio Enhancement and Restoration Using Gaussian Based Granular Synthesis.

[2] National Science Council — Emotion Detection System Based on Audiovisual Signal Fusion.

[3] National Science Council — Development of Visual-Audio Aid Systems Based on Virtual Acoustics.

[4] Tri-Service General Hospital and National Taiwan University of Science and Technology — Computerized Auscultation of Bowel Sounds: Development of Electronic Stethoscope for Clinical Application.