2. M. Dietz, L. Liljeryd, K. Kjorling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," 112th AES convention, Munich, Germany, May 2002.
3. Chi-Min Liu, Wen-Chieh Lee, and Han-Wen Hsu "High Frequency Reconstruction for Band-Limited Audio Signals", DAFX'03, London, UK, September 2003.
4. Manish Arora, Joonhyun Lee, and Sangil Park "High Quality Blind Bandwidth Extension of Audio for Portable Player Applications", 120th AES convention, Paris, France, May 2006.
5. Chatree Budsabathon, Akinori Nishihara "Bandwidth Extension with Hybrid Signal Extrapolation for Audio Coding", IEICE Trans. Fundamentals, Vol. E90-A, No. 8, August 2007.
6. N. Enbom and W.B.Klein "Bandwidth Expansion of Speech Based on Vector Quantization of the Mel Frequency Cepstral Coefficients", in IEEE Workshop on Speech Coding, Porvoo, Finland, 1999.
7. M. Nilsson, H. Gustafsson, S.V. Andersen, and W.B. Kleijn. "Gaussian mixture model based mutual information estimation between frequency bands in speech." ICASSP '02, IEEE International Conference on, 2002.
8. Neil Gershenfeld, "Cluster-Weighted Modeling: Probabilistic Time Series Prediction, Characterization and Synthesis", Nature of Mathematical Modeling, MIT Press, 1998.

BANDWIDTH EXTENSION FOR AUDIO SIGNALS USING CLUSER-WEIGHTED MODELING

Lyubimov N., Lukin A.

Moscow Lomonosov State University, Dept. of Computational Mathematics and Cybernetics
Laboratory of Mathematical Methods of Image Processing

Introduction

Bandwidth extension is the process of re-synthesizing missing frequency components in order to improve the subjective quality of the audio signal. Bandwidth extension methods can be found in modern perceptual audio coding standards, such as mp3PRO and AAC+. Such methods can be blind, when no information about missing signal components is available, and non-blind, when certain information about missing components is available during the synthesis stage.

A typical algorithm flowchart for both blind and non-blind methods looks as follows:

1. Time-frequency decomposition,
2. "Rough" generation of high-frequency spectral content,
3. Shaping of the energy spectrum envelope of high-frequency content,
4. Synthesis of the resulting signal from a time-frequency representation.

Algorithm description

In this work, a new algorithm for blind bandwidth extension is proposed. It is capable of accurate prediction of high-frequency energy envelopes using a Cluster-Weighted Model for MFCC coefficients of the audio signal.

A bandwidth-reduced audio signal x(t) is input to the algorithm. It is transformed using STFT with 30-ms Hann windows that overlap by 15 ms. A nonlinear distortion (waveshaping) is used to generate "rough" high-frequency components in time domain: $z(t) = |x(t)|^s$. Aliasing is reduced by the use of oversampling. These high-frequency components are transformed using a similar STFT filter bank.

To finally shape the resulting high-frequency signal, its energy is computed in 24 critical bands, and a regression model is developed to predict the shape of high-frequency energy envelope from a low-frequency energy envelope. In this work, a Cluster-Weighted Modeling is proposed for such prediction. The model is trained on the original full-bandwidth audio signals to determine a set of clusters of low-frequency envelopes and corresponding high-frequency envelopes. During evaluation stage, each input low-frequency envelope is represented as a weighted sum of several "cluster" envelopes, and the corresponding high-frequency envelope is predicted as a weighted sum of modeled high-frequency envelopes.

Once the desired high-frequency envelope is calculated, the shape of a "rough" high-frequency signal is transformed to match the desired envelope using STFT filter bank.

Results and conclusion

The algorithm has been evaluated on speech and music using a subjective evaluation protocol with several listeners. The proposed method has been compared with other methods, such as linear extrapolation of high-frequency envelope (see full references within the paper). The proposed algorithm has shown the highest subjective quality results.

Evaluations have shown that the algorithm is more effective on speech and solo music than on polyphonic music. A possible cause of this effect is introduction of intermodulation components by the process of nonlinear distortion. In our future work, we are planning to apply source separation techniques for individual processing of signal components.