Proc. Biennial Symp. Commun. (Kingston, ON), pp. 524-527, June 2002

PSEUDO-WIDEBAND SPEECH RECONSTRUCTION FROM TELEPHONE SPEECH

Yasheng Qian and Peter Kabal
Dept. of Electrical and Computer Engineering, McGill University,
3480 University Street, Montreal, Quebec H3A 2A7

ABSTRACT

The bandwidth of telephone speech is limited to a 300 – 3400 Hz bandwidth. The sound quality is much lower than for broadcast radio and audio compact discs. We present an algorithm to regenerate the missing highband components (3.4–7 kHz). The highband spectrum recovery is based on a Line Spectrum Frequency (LSF) VQ codebook mapping from the narrowband speech to the high frequency components. The highband excitation employs a substitute of a bandpass (2–3 kHz) envelope modulated Gaussian White Noise. The modulation gain of the excitation exploits a lowband LSF VQ mapping codebook. Spectrograms demonstrate that the reproduced speech has obtained most of missing components. Subjective tests show that the reconstructed speech is significantly more pleasant and natural than the conventional telephone speech.

Summary

The present public telephone networks limit human voice transmission to a narrow bandwidth (300–3400 Hz) for compatibility with the existing infrastructure. Although speech with this bandwidth is highly intelligible, it has lost most of the high frequency components over 3.4 kHz which give the speech a “presence”. The sound quality is not comparable to common broadcast radio, TV and compact discs. The wider bandwidth better preserves speaker identity.

3G wireless systems can deliver wideband speech (up to 7 kHz). However, the extra quality will be lost in the connection to the existing narrowband PSTN. Pseudo-wideband expansion techniques can recreate the missing highband components of narrowband speech from the PSTN to the wideband terminal. This provides a pragmatic approach to enhance the quality presented to the owner of a wideband phone.

The challenging task is to recover the signal in the 3.4–7 kHz range from the narrowband signal. We employ a speech synthesis model which uses an excitation signal and a spectral envelope. The excitation signal can be modelled as white noise for unvoiced phonemes, a pulse train for voiced phonemes and their mix for transitions. The missing high frequency components of the excitation can be generated in a number of ways from the narrowband residual of an LP (Linear Prediction) filter. The highband excitation can be formed in a manner similar to LPC coding by modelling the excitation as a pulse train

or a noise sequence [1]. Another approach employs spectral folding by downsampling and upsampling the available narrow band residual [2]. The highband spectrum envelope can be reconstructed from a codebook mapping or a statistical modelling approach. The spectrum envelope can be represented by linear prediction coefficients (LPC), linear frequency cepstrum coefficients (LFCC) or mel frequency ceptstrum coefficients (MFCC), etc.

Our new approach for highband excitation recovery is to generate an envelope from the bandpass (2–3 kHz) components of the original residual signal. This envelope modulates high frequency Gaussian white noise. The use of a relatively narrow portion of the original residual is motivated by the observation that this frequency region contains good noise-like components for unvoiced speech and a strong envelope modulation for voiced speech. A modulation gain parameter is introduced to give the resynthesized highband components the appropriate energy. The gain factor for the excitation exploits a lowband spectrum VQ mapping codebook. The modulating gain is defined as the square root of the energy ratio of the original highband to the resynthesized one. The codebook consists of the centroids of the corresponding narrowband quantized LSF vectors of the training data. The gains corresponding to the LSF vectors supply the gain for the highband excitation. The codebook is trained from a large set of gains and LSF data.

The missing highband spectrum is reproduced by a narrowband to highband band LSF VQ mapping codebook. Because of the well-known properties (ordering and quantization error constraints) of LSF for representing the speech spectrum, 14 LSFs are utilized to represent the narrow band spectrum and 10 LSFs are used for the highband. From the narrowband to highband LSF VQ mapping codebook, the highband LSF vector is generated from the narrowband LSFs. The training of the VQ codebook is done by clustering the highband LSFs to the corresponding lowband LSFs counterparts. The centroids of the highband LSFs cells form the mapping VQ codebook.

Spectrograms of the pseudo-wideband speech, narrow band speech and original speech are shown in Fig. 1. The test sentences are drawn from a phonetically balanced set of utterances (Harvard database). The specific sentence shown is “Hats are worn to tea and not to dinner”. The top trace shows the original wideband speech. The bottom trace shows the narrowband speech that is input to the pseudo-wideband algorithm. The middle trace shows the wideband speech reconstructed entirely from the narrowband input. The spectrograms demonstrate that most of missing highband components have been reconstructed. For voiced segments, the periodicity has been successfully extended to the high band. For unvoiced speech, the high frequency reconstructed components are also noise-like. The level of the reconstructed components is generally less than the original high frequency components. It is better to underestimate these components than to supply unnaturally strong components. Informal listening shows that the proposed pseudo-wideband reconstruction algorithm is much more natural than the narrowband telephone speech.

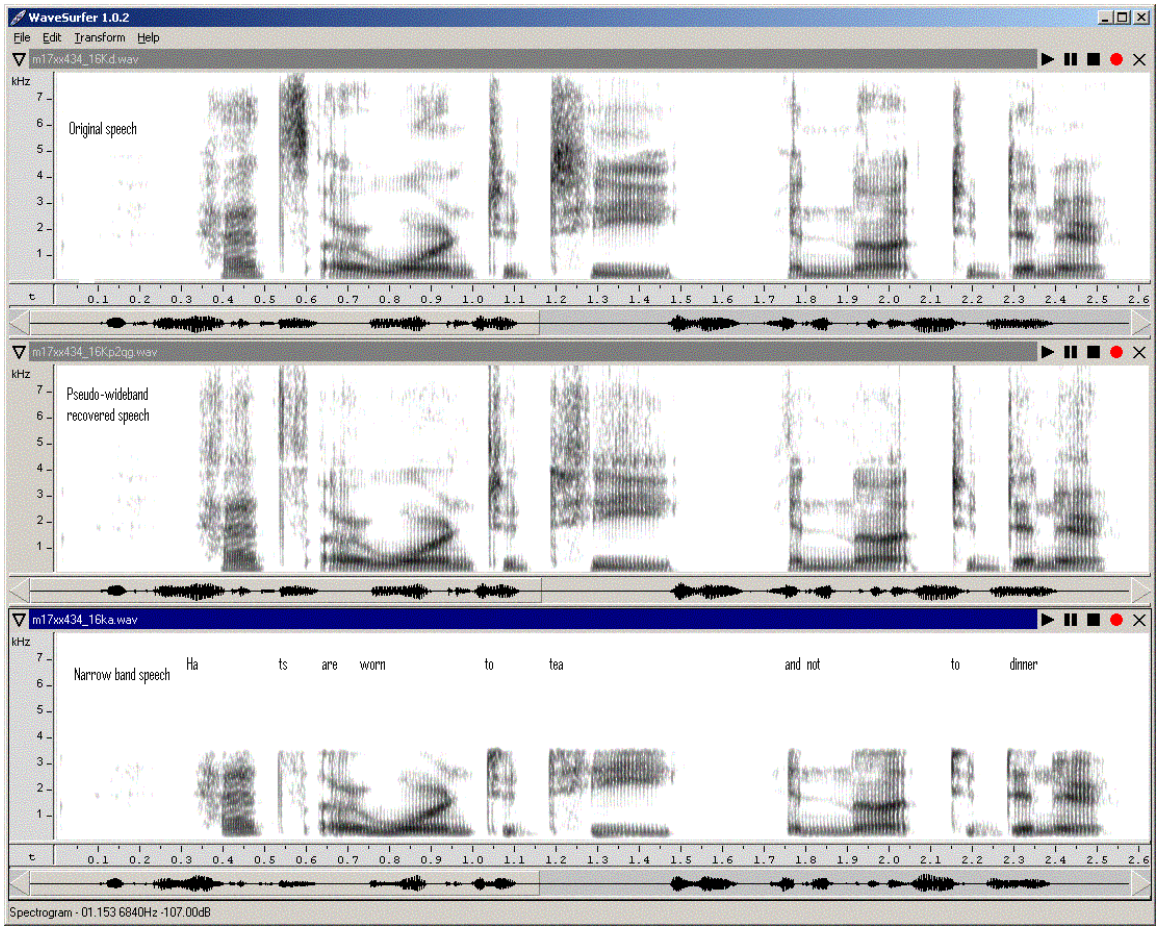

Fig.1 Spectrograms for (a) an original wideband speech utterance: “Hats are worn to tea and not to dinner.”, (b) the pseudo-wideband reconstructed speech, and (c) the telephone speech.

References

[1] Y. Yoshida and M. Abe, “An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping”, *Proc. Int. Conf. Spoken Language Processing*, pp.1591–1594, 1994.

[2] M. Nilsson and W. B. Kleijn, “Avoiding over-estimation in bandwidth extension of telephony speech”, *Proc. IEEE Int. Conf. Acoust. Speech Sig. Process.*, pp. 869–872, 2001.