ON THE APPLICATION OF THE AM-FM MODEL FOR THE RECOVERY OF MISSING FREQUENCY BANDS OF TELEPHONE SPEECH

Hesham Tolba *Douglas O'Shaughnessy*

INRS-Télécommunications, Université du Québec
16 Place du Commerce, Verdun (Île-des-Soeurs),
Québec, H3E 1H6, Canada
{tolba, dougo}@inrs-telecom.uquebec.ca

ABSTRACT

This study presents a novel technique to reconstruct the missing frequency bands of band-limited telephone speech signals. This technique is based on the Amplitude and Frequency Modulation (AM-FM) model, which models the speech signal as the sum of N successive AM-FM signals. Based on a least-mean-square error criterion, each AM-FM signal is modified using an iterative algorithm in order to regenerate the high-frequency AM-FM signals. These modified signals are then combined in order to reconstruct the broad-band speech signal. Experiments were conducted using speech signals extracted from the NTIMIT database. Such experiments demonstrate the ability of the algorithm for speech recovery, in terms of a comparison between the original and synthesized speech and informal listening tests.

1. INTRODUCTION

The problem of processing band-limited (telephone) speech signals has received considerable attention in order to improve the subjective speech quality, especially with the significant increase in speech applications over telephone lines. Telephone speech is usually limited to frequency components in the range 300-3400 Hz, to maintain a minimum intelligibility of the signal while permitting economy of the frequency spectrum in the public switched telephone network. Several techniques have been proposed in the literature to solve this problem; however such techniques suffer from several drawbacks [1, 3, 4]. In this study, we present a novel speech recovery technique, which is based on a multiband analysis of the speech signal which is modeled using the AM-FM model. We show through experiments that, using such a model, we are capable of restoring the missing frequency bands of telephone speech using a Mean Square Error (MSE) criterion as shown in section 3. The obtained weights from the MSE algorithm are then used to modify the N AM-FM narrow-band signals in order to reconstruct the broad-band speech signal.

This novel technique reconstructs the missing frequency bands (i.e., recovers broad-bandwidth speech) of the band-limited telephone speech signals via the reconstruction of high-frequency bands of its spectrum using an iterative approach. In this proposed approach, the amplitude and frequency (AM-FM) modulation model [6] and a multi-band analysis scheme [7] are applied to the speech signal to extract the spectrum of each AM-FM signal at the successive resonances of the speech signal. This process is applied to training broad-band speech data in order to extract the isolated AM-FM signals, which are used to compute a weighting function ω_i that is used to modify the corresponding narrow-band AM-FM signals based on a MSE criterion. The obtained weighting functions serve in the prediction of the high-frequency missing frequency bands. The estimated signals are then used to regenerate the broad-band speech. To reconstruct the broad-bandwidth speech signal, the modified AM-FM signals are accumulated together in the time domain to reconstruct the speech on a frame-by-frame basis. Then these speech frames are concatenated in order to regenerate the broad-band speech signal.

Although broad-band speech recovery using multiband analysis was investigated in [4], in this paper we attack the problem using the AM-FM framework, which facilitates the analysis by limiting the number of the filters required to perform the analysis. Moreover, the synthesis of the broad-band speech is performed by a procedure that uses the inverse fast Fourier transform to compute each AM-FM signal's contribution, easily in the time-domain, rather than sets of oscillator functions, as in other sinusoidal models.

The outline of this paper is as follows. In section 2, an overview of the AM-FM Modulation Model and the energy operator is given. Then in section 3, we describe the proposed approach for speech recovery and the system designed for such an approach. Experimental results that demonstrate the effectiveness of our algorithm are presented in section 4. Finally, in section 5 we conclude and discuss our present and future work.

2. THE AM-FM MODULATION MODEL FRAMEWORK

2.1. Analysis-Synthesis

Motivated by several nonlinear and time-varying phenomena during speech production, Maragos, Quatieri and Kaiser [6] proposed an AM-FM modulation model that represents each single speech resonance (formant) as an AM-FM signal. This model represents each resonance of a speech signal as a signal with a combined amplitude modulation (AM) and frequency modulation (FM) structure. Then, the speech signal $x(t)$ is modeled as the sum of N such AM-FM signals, one for each formant, as follows:

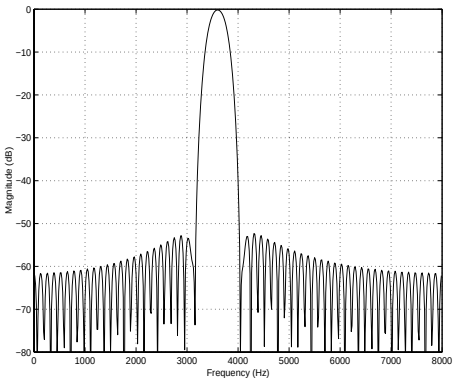

Figure 1: Spectral Magnitude of the BPF (bandwidth = 400 Hz) used to isolate each AM-FM signal component.

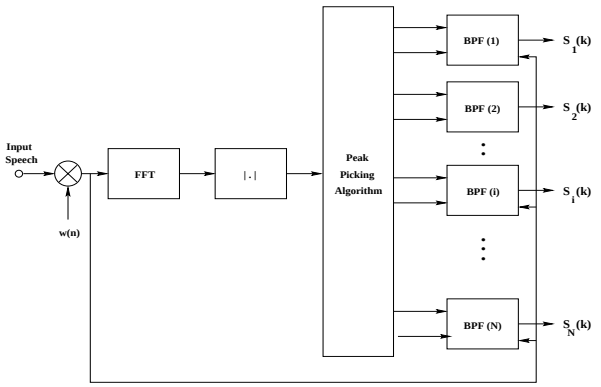

Figure 2: Block Diagram of the multiband AM-FM-based Analysis broad-bandwidth speech recovery system.

$$x(t) = \sum_{i=1}^{N} a_i(t)\, cos\Big(2\pi[f_{c,i}t + \int_0^t q_i(\tau)\, d\tau] + \theta_i\Big), \quad (1)$$

where $f_{c,i}$ is the center value of the i^{th} formant frequency, $q_i(t)$ is the frequency modulating signal, and $a_i(t)$ is the time-varying amplitude. The i^{th} instantaneous formant frequency signal is $f_{inst,i}(t) = f_{c,i} + q_i(t)$. In the discrete-time domain the i^{th} discrete-time AM-FM signal is defined as

$$x_i(n) = a_i(n)\, cos\Big(\Omega_{c,i}\, n + \Omega_m \int_0^n q_i(k)\, dk\Big), \quad (2)$$

where $a_i(n)$ is the discrete-time amplitude envelope, $\Omega_{c,i}$ is the carrier frequency, and Ω_m is the modulation frequency. The digital instantaneous frequency of the discrete-time AM-FM signal, $\Omega_{inst,i}$, is defined as

$$\Omega_{inst,i}(n) = \Omega_{c,i} + \Omega_m q(n). \quad (3)$$

Modeled as a sum of AM-FM signals, the speech signal can be processed easily to estimate several parameters such as the amplitude of the envelope, the instantaneous frequency of each resonance (peak) at each time instant t, the discrete-time energy operator DEO, the tracking of the formants and pitch extraction.

2.2. Multi-Band AM-FM Demodulation

By the selection of an appropriate bandpass filter, isolation of an AM-FM signal component could be possible [7]. This is due to the fact that the wideband FM signal could be restricted to a limited-band signal without losing the original signal if the limited-band signal contains more than 98% of the total power of the original FM signal. Moreover, noise components not falling within the vicinity of the desired local AM-FM component could be rejected.

To isolate the local modulation energy of an AM-FM signal component, it is necessary to utilize a bank of bandpass filters centered at each peak of the speech signal with an appropriate bandwidth. A neighboring spectral peak that has not been eliminated through bandpass filtering can seriously affect the exact reconstruction of the synthesized signal. 400 Hz was found in [6] to be a reasonable value for the bandwidth of such a filter in order to avoid the effects of neighboring formants.

Resonance isolation is performed using a bank of bandpass filters. These filters are implemented using a truncated, discretized BPF FIR filter with impulse response

$$g_\omega(n) = g(n)\, \omega_H(n), \quad (4)$$

where $\omega_H(n)$ is the Hamming window function

$$\omega_H(n) = \begin{cases} 0.54 - 0.46\cos\Big(2\pi\, \frac{n}{L-1}\Big), & 0 \le n \le L-1, \\ 0, & \text{otherwise}, \end{cases} \quad (5)$$

which is used to modify and truncate the ideal impulse response $g(n)$ given by:

$$g(n) = \frac{\sin[\omega_H(n - L/2)]}{\pi(n - L/2)} - \frac{\sin[\omega_L(n - L/2)}{\pi(n - L/2)]}, \quad (6)$$

where ω_H and ω_L define the low and high edges of the passband, respectively, and $L/2$ is the delay required for causality.

3. WIDE-BAND SPEECH RECOVERY SYSTEM

The wide-band speech recovery system proposed in this paper is based upon a multi-band analysis described in section 2.2 applied to the speech signal modeled using the AM-FM speech model described in section 2. The idea of this proposed approach is to modify the spectrum of the high-frequency N successive bands of the bandpass filtered narrow-band speech signal in order to restore the missing frequency bands. Such a modification is obtained through an iterative adaptation technique based on a least square estimate during a training phase. The coefficients obtained in such a training are then used in order to modify the narrow-band speech spectrum. These spectrums, in turn, are then added together in order to synthesize the wide-band speech signal. Such a reconstruction is performed in the time domain on a frame-by-frame basis.

3.1. High-Frequency Spectral Estimation

The AM-FM speech analysis window used in this study consists of three parts: the first and the third parts are half Hamming

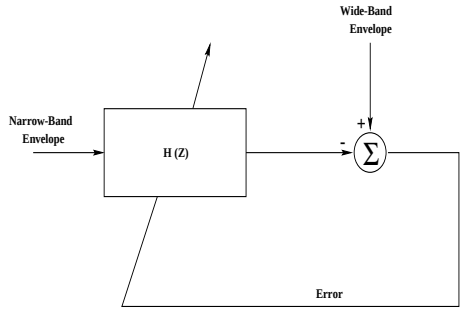

Figure 3: Adaptive Filter Elements.

window, whereas the second part is a rectangular window. This window function is given by:

$$\omega_T(n) = \begin{cases} 1, & 1 \leq n \leq L_\omega - 1, \\ \omega_H, & L_\omega \leq n \leq 2L_\omega - 1, \\ 1, & 2L_\omega \leq n \leq N_\omega - 1, \end{cases} \quad (7)$$

where ω_H is the Hamming window function (Eq. 5), N_ω is the window length and $L_\omega = N_\omega/3$. The frame size, N_ω, is chosen equal to 30 ms, with 10 ms displacement. The AM-FM analysis window is applied to an equal number of samples (160 samples) from the past speech frame, the present speech frame and the future speech frames, respectively. Then, each windowed speech frame is transformed into the frequency domain using a 1024-point FFT. Then the spectrum of the N AM-FM signals is computed for each peak in the Fourier spectrum by isolating each peak using a filter bank of the bandpass filters (described in section 2.2) centered around each peak. The bandwidth for such bandpass filters was chosen to be approximately 400 Hz as mentioned in section 2.2.

Once we isolate each AM-FM signal, the spectra for these successive resonances are then modified using an adaptive filter as shown in Fig. 3. The coefficients of such a filter are obtained during a training phase, using a least-mean-square (LMS) adaptive algorithm given the IFFT of the spectral obtained from the narrow-band speech signal, $s_{k,i}^{NB}(n)$, as the input signal to such a filter and the IFFT of the spectral obtained from the broad-band speech signal, $s_{k,i}^{WB}(n)$, as the desired signal. The output of such a filter is given by:

$$\hat{s}_{k,i}^{WB}(n) = \sum_{j=1}^{M} \omega_i(j) s_{k,i}^{NB}(n-j), \quad (8)$$

where $\hat{s}_{k,i}^{WB}(n)$ is the estimate of the recovered AM-FM signal around the i^{th} peak for the k^{th} frame, M represents the filter's length and ω_i represents the filter coefficients for such a peak. In our experiments, we used filters consisting of 41 taps.

3.2. Broad-Band Speech Synthesis

Once we have obtained the modified AM-FM signals based on the above mentioned LMS algorithm, the synthesis of the broad-band speech is performed by the summation of the modified AM-FM signals, $\hat{s}_{k,i}^{WB}(n)$, in order to obtain the speech signal for each frame k, $\hat{s}_k^{WB}(n)$. Then, for each of these signals, only the second 160 samples, which represent the current frame, are selected, then concatenated together in order to generate the broad-band speech signal.

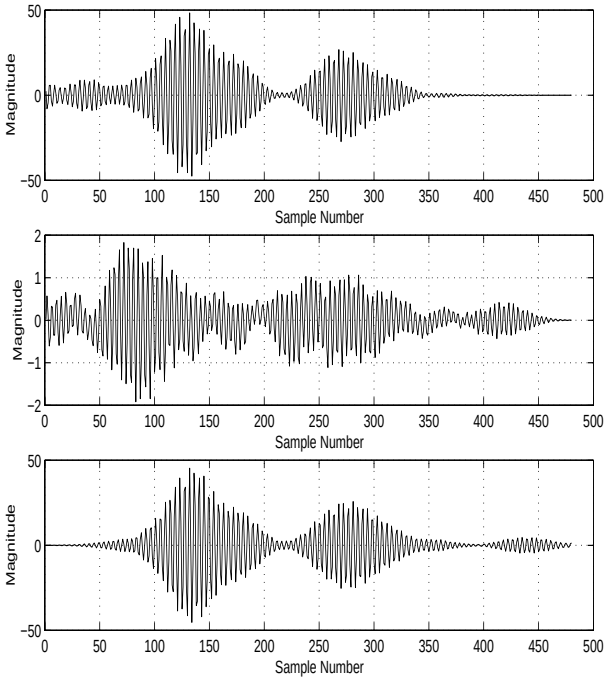

Figure 4: Comparison of a single AM-FM time waveform of (f_c=3969 Hz; BW ≈ 400 Hz) (a) the original broad-band speech, (b) the narrow-band speech and the (c) the estimated speech.

4. EXPERIMENTAL RESULTS

The speech corpus for this experiment is a subset of the NTIMIT database [8]. The NTIMIT database is the telephone-bandwidth version of the widely-used TIMIT database, which was sampled at 16 kHz. Besides band limitation, several kinds of noise can be found in the NTIMIT database, such as broadband noise, band-limiting, low frequency hum, crosstalk, dial pulses, shot noise and sharp pulses. Utterances from both the TIMIT and NTIMIT databases were selected in order to train our algorithm to obtain the weighting coefficients ω_i. However, sentences outside the training set were used for the evaluation of the algorithm.

Figs. 4 and 5 show for different resonance (peak) values that using the adaptive filters per peak restores the broadband AM-FM signal from the narrow-band one around that peak. The more training data that we use, the more the filter is accurate and our estimation is better.

Fig. 6 illustrates a typical example of the spectrograms of a speech utterance uttered by a male, selected from the subset dr1 from the TIMIT database, the same utterance uttered by the same speaker selected from the NTIMIT database and the reconstructed broad-band speech utterance using the above-mentioned algorithm. This figure indicates that most of the missing frequency bands have been reconstructed. The amount of accuracy of such a reconstruction depends on several factors such as: the amount of training data used to design the adaptive filters (in other words, how accurate is the adaptive filter), the number of

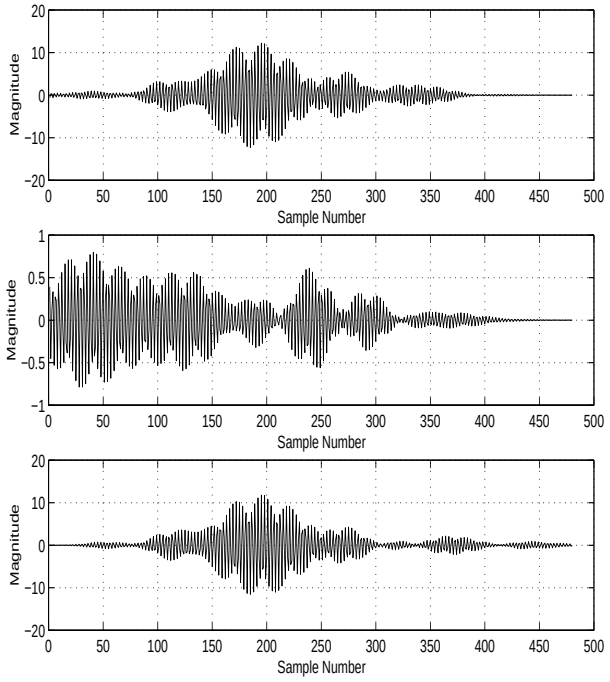

Figure 5: Comparison of a single AM-FM time waveform (f_c=5547 Hz; BW $\approx$ 400 Hz) of (a) the original broad-band speech, (b) the narrow-band speech and the (c) the estimated speech.

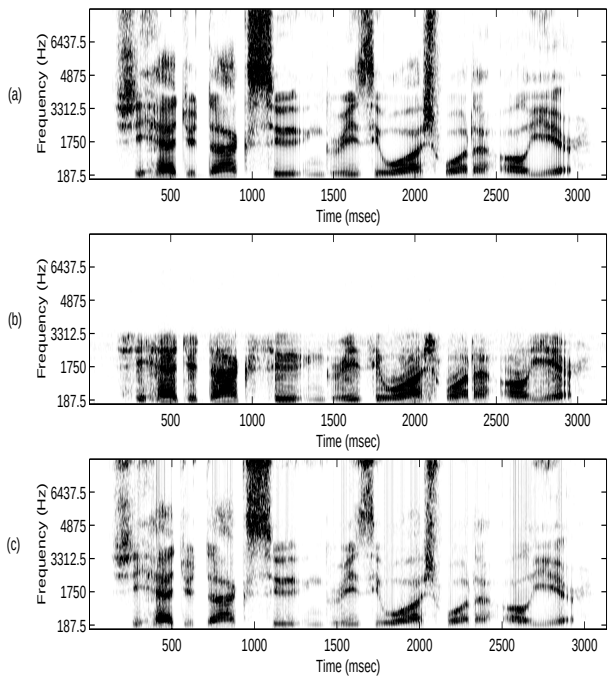

Figure 6: Comparison of the spectrograms of (a) the original broad-band speech, (b) the narrow-band speech and the (c) the recovered speech.

the AM-FM signals, N, consequently the number of filters that have been used in the analysis and synthesis of the speech signal and the associated bandwidths (i.e., the design of the filter bank that is used for analysis and synthesis of the speech signal).

5. CONCLUSION

In this paper, we have presented a novel speech recovery approach. The recovery algorithm was presented in the framework of the AM-FM speech modulation model. We have examined several possibilities for the recovery of the missing frequency bands of a bandlimited signal in terms of the parameters obtained from the AM-FM speech model. We found that the characteristics of the telephone channel for each AM-FM signal extracted based on a LMS criteria using both the broad-band and the narrow-band speech signals can play a role in the recovery of the missing bands. In these respects, an algorithm has been built and tested on telephone speech. Comparisons between the original and synthesized speech indicated that such an algorithm is able to restore and recover missing frequency bands in the telephone speech for different speakers. Moreover, informal listening tests indicate that the subjective quality of bandlimited speech is enhanced using our proposed algorithm.

One possibility for the future research is to improve the recovery algorithm described above by using a set of contiguous bandpass filters, whose passbands increase in width with increasing frequency. That is, applying analysis which is similar to the frequency analysis made by the ear that is characterized by the so-called *critical bands*.

6. REFERENCES

[1] Y. M. Cheng, D. O'Shaughnessy and P. Mermelstein, "Statistical Recovery of Wideband Speech from Narrow-band Speech", Proc. ICSLP–92, pp. 1577–1580, 1992.

[2] J. Makhoul and M. Berouti, "High-Frequency Regeneration in Speech Coding Systems", Proc. ICASSP-79, 428-431, 1979.

[3] Carlos Avendano, H. Hermansky and Eric A. Wan, "Beyond Nyquist: Towards the Recovery of Broad-Bandwidth Speech from Narrow-Bandwidth Speech", Proc. EUROSPEECH–95, pp. 165-168, 1995.

[4] H. Hermansky, Eric A. Wan and Carlos Avendano, "Speech Enhancement Based on Temporal Processing", Proc. ICASSP–95, pp. 405-408, 1995.

[5] James F. Kaiser, "On a Simple Algorithm to Calculate the Energy of a Signal", Proc. ICASSP-90, pp. 381-384, 1990.

[6] P. Maragos, T. F. Quatieri and J. F. Kaiser, "On Amplitude and Frequency Demodulation Using Energy Operators", IEEE Trans. on Signal Processing, Vol. 41, No. 4, pp. 1532-1550, April 1993.

[7] A. Bovik, P. Maragos and T. Quatieri, "AM-FM Energy Detection and Separation in Noise Using Multi-band Energy Operators", IEEE Trans. on Signal Processing, SP–41(12), pp. 3245-3265, December, 1993.

[8] Charles Jankowski, Ashok Kalyanswamy, Sara Basson and Judith Spitz, "NTIMIT: A Phonetically Balanced, Continuous Speech, Telephone Bandwidth Speech Database", Proc. ICASSP-90, pp. 109-112, 1990.