REVIEW OF WIDEBAND SPEECH NOISE REDUCTION TECHNIQUES

Malay Gupta, Chris Forrester, and Sean Simmons
Research in Motion, 295 Phillip St., Waterloo Ontario, Canada, N2L 3W8 mgupta@rim.com

1. INTRODUCTION

Wideband speech has been touted as the next great leap in telephone voice quality – analogous some would say to the difference between AM and FM radio. Wideband telephony extends the transmitted speech bandwidth from 3.4 kHz to 7.0 kHz. Wideband speech potentially offers improved speech quality and intelligibility and a more life-like voice experience. Realizing the potential of high quality speech communications, VoIP-based enterprise systems have been first to deploy wideband speech in a significant way in desktop phones within the office environments. Of course the public telephone networks are still narrowband, but with the advent of high-powered digital signal processors in mobile phones, increased network bandwidth and the standardization of wideband speech codecs, the deployment of wideband voice into the wireless public network appears to be on the horizon.

One challenge facing wireless phone designers is to realize the potential of wideband speech in noisy environments. With the possibility of enterprise mobile phones replacing the desktop phones in the near future [1], the removal of background acoustic noise has become very important in order to fully realize the benefits offered by wideband speech codecs. Narrowband systems constrain the low frequency bandwidth to 200-300 Hz and hence inherently filter out significant low frequency noise. Wideband systems have the potential to transmit low frequency noise down to 50-100Hz. This additional one to two octaves of low frequency noise can significantly degrade the speech quality unless great care is taken in suppressing it. This paper outlines the challenges involved and presents possible wideband noise suppression approaches.

1.1 Bandwidth of transmission

A bandwidth of 200 Hz to 3400 Hz has traditionally been considered sufficient to transmit the information contained in human speech over the telephone networks. However, this constrained bandwidth actually transmits only 20 percent of the frequencies that are present in the human speech [2]. This constrained bandwidth therefore results in a low quality/intelligibility speech, causing confusion and fatigue between the near-end and far-end users.

Increased bandwidth has been shown to be a critical factor determining the speech intelligibility and naturalness [2]. According to a Siemens 2003 press release, *wideband transmission can reduce speech ambiguities by as much as 90 percent, increasing conversational intelligibility and reducing listener fatigue*. Polycom reported in its 2006 technology white paper that *for single syllables, 3.3 KHz yields an accuracy of only 75 percent, as opposed to over 95 percent with 7 KHz bandwidth* [2]. However, this increased bandwidth also opens up a new challenge for noise suppression algorithms as now more noise will also be subsequently transmitted along with the near-end speech.

1.2 Wideband speech structure

The transmitted narrowband speech is mainly voiced speech mostly consisting of vowel sounds. On the other hand, a wideband signal in addition to vowel sounds has a significant amount of consonant energy. Consonants are unvoiced high frequency speech components contributing mainly towards the speech intelligibility. Consonants are often low in intensity when compared to the voiced low frequency content of the signal and have a noise-like structure in them. Therefore, consonant sounds are actually more prone to noise distortions. The increased susceptibility of wideband systems for noise distortions call for careful development of noise reduction algorithms for wideband.

2. NOISE REDUCTION IN WIDEBAND

For many decades, noise reduction algorithm developers have focused on the development and optimization of algorithms suitable only for the narrowband systems. Because of this huge inertia, very little information is available in the literature about noise reduction techniques suitable for the wideband frequency range. Extending a narrowband noise reduction algorithm to wideband is a non-trivial task, and requires a deep understanding of differences between the narrowband and wideband signals and systems. As such, algorithms that seem to work for narrowband systems are not guaranteed to work well in wideband systems even when accompanied with an increased computational complexity.

A majority of narrowband speech enhancement algorithms process the noisy speech in frequency domain and are based on short-time analysis-synthesis technique and follow the principles of spectral subtraction and short-time spectral amplitude (STSA) estimation [3]. In these techniques, the incoming noisy signal is processed in successive frames by computing the FFT of the windowed frames, and then estimating the clean speech spectrum from the noisy speech spectrum by using various techniques [3]. Potentially, the underlying principles behind short-time analysis-synthesis are applicable to wideband systems with certain modifications. In order to develop noise reduction algorithms for wideband systems, algorithm developers have two choices, first is to modify the existing narrowband noise reduction algorithms for wideband applications, and the second choice is to account for structural differences between narrowband and wideband signals and device different algorithmic strategy for wideband signals.

2.1 Extending narrowband algorithms to wideband

The transition from narrowband to wideband involves the doubling of sampling frequency from 8 KHz to 16 KHz. For a fixed number of frequency bins, this results in a loss of spectral resolution in wideband as compared to narrowband systems. Hence in order to keep the spectral resolution same as narrowband, the length of FFTs in wideband processing must be doubled. One obvious impact of this change is increased computational complexity; however, that is not the only impact. In order to keep the algorithm behavior same as narrowband, various other parameters must also be carefully examined. These parameters may include attack and decay constants commonly used for speech activity based gain computations, smoothing parameters for periodograms etc., just to name a few [4]. The frequency-domain changes might also impact the time-domain evolution of the algorithm. For example, frequency-domain parameter changes might adversely impact the convergence time of the algorithm. Hence a proper compromise needs to be made between time-domain and frequency-domain performances when modifying narrowband algorithms for use in wideband systems.

2.2 New algorithms for wideband

New algorithms specifically tailored for wideband applications generally demand different processing strategies for different parts of the wideband signals. For example, a typical solution would consist of a *split-band* approach by dividing the signal spectrum into a low-frequency part and a high-frequency part as has been proposed in [5] for automatic speech recognition (ASR) applications. Different processing strategies could be devised for the low-frequency part and high-frequency part in terms of FFT size, spectral gain computations, or in terms of parameters/thresholds that are needed in the algorithm.

Structural/spectral richness of wideband signals could also be exploited to develop multi-state speech models, in which states represent the presence or absence of fricative, stop, vowel, glide, and nasal speech sounds [6]. An elaborate speech enhancement procedure would then consist of first identifying the state and then applying a state specific filter to the noisy speech frame. Because of the presence of more speech states, perceptual models may also have a greater impact in noise reduction process. Recent algorithms incorporating perceptual models can be found in [7]-[10].

2.3 Processed speech expectations

User expectations of a much better speech quality compared to narrowband raise the bar for algorithm developers in terms of preserving the speech quality in noisy conditions. It is more important than ever not to introduce any spectral distortion in the processed speech. It has also been noted that the quality of background noise transmission in a wideband system impacts overall speech quality differently than a narrowband system [11]. Hence preserving background noise characteristics also appears to be an important task in a wideband system. Fully characterizing the parameters that affect the overall speech quality in a wideband system is an ongoing process, with a specialist task force 294 (STF-294) of ETSI specifically working on wideband speech applications.

3. CONCLUSIONS

The standardization of wideband speech codec has made the deployment of wideband speech a reality in public telephone domain. In this paper, we have reviewed the current state-of-the-art of noise reduction algorithms suitable for wideband applications. We presented possible strategies for the development of noise reduction algorithms for wideband applications. We discussed few modifications that are needed for the application of narrowband noise reduction algorithms to wideband. Also due to the structural richness of wideband signals some new algorithms tailored for wideband applications have been mentioned. Lastly, we have outlined speech quality expectations resulting from narrowband to wideband transition.

REFERENCES

[1] Gartner Press Release (IT section), Feb. 2009.

[2] Rodman J., The effect of bandwidth on speech intelligibility. Polycom white paper, Sept. 2006.

[3] Loizou, P., Speech enhancement: Theory and Practice. CRC Press, Boca Raton, FL, 2007.

[4] Beaugeant C., et al., Challenges of 16 KHz in acoustic pre- and post-processing for terminals. IEEE Communications Magazine, pp. 98-104, May 2006.

[5] Macho D., et al., On the use of wideband signal for noise robust ASR. Proc. IEEE ICASSP, vol. 2, pp. 109-112, April 2003.

[6] Drucker H., Speech processing in a high ambient noise environment. IEEE Trans. Audio Electroacoust., vol. AU-16m, pp. 165-168, June 1968.

[7] Thiemann J., et al., Noise suppression using a perceptual model for wideband speech signals. Proc. Biennial Symp. Commun., Kingston, ON, pp. 516-519, June 2002.

[8] Jelinek M., et al., Noise reduction method for wideband speech coding. Proc. EUSIPCO., pp. 1959-1962, 2004.

[9] Kato M., et al., A wideband noise suppressor for the AMR wideband speech codec. Proc. IEEE ICASSP, pp. 156-158. 2002.

[10] C. Beaugeant et al., Noise reduction and preprocessing for the AMR-WB. ETSI STQ Wksp. Wideband Speech Signal Processing, June 2005.

[11] Gierlich H.W., Wideband speech communication–The quality parameters as perceived by the user. ForumAcusticum, 2005.