NOISE SUPPRESSION IN CELLULAR TELEPHONY

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

1. INTRODUCTION

Noise suppression in cellular telephony is one of the several measures (ITU-T P.330) taken to improve the speech quality and intelligibility at the far end. With cellular/mobile phones increasingly becoming outdoor devices, noise suppression has become an integral part of the audio signal processing chain used in cellular telephony today. The term speech enhancement, which in principle has a broader scope, has been used synonymously with noise suppression and has been an active area of research since the late 1960's. Suppressing various types of background acoustic noise without adversely impacting the speech quality makes this problem particularly challenging.

Examples of noise environments that a typical cell phone has to encounter include: home, office, car, street, café etc. A noise suppression algorithm is desired to enhance speech quality/intelligibility in these environments and is expected to operate reliably in a broad range of stationary/non-stationary signal-to-noise-ratio (SNR) environments ranging from -3 dB to 30 dB (ITU-T G.330). With worldwide mobile sales beating almost any optimistic forecast, a good noise suppression algorithm is one of the key factors to keep voice quality high and meet customer requirements.

2. NOISE SUPPRESSION METHODS

The goal of noise suppression system is to aid speech encoder by presenting it a signal with reduced background noise (BGN) level in comparison to the original noisy signal. According to the number of microphones used in the noise suppression system, we can broadly group noise suppression techniques into two groups. Group 1 utilizes a single-microphone and Group 2 utilizes an array of microphones. In the next two subsections we briefly discuss these techniques and refer to some of the important references in the present context

2.1 Single Microphone Noise Suppression

Most of the techniques proposed in this category function in the frequency-domain and use the short-time analysis-synthesis technique. These techniques apply a frequency-dependent gain function to the spectral components of the noisy speech to attenuate the components with greater noise content. This gain function could be a simple thresholding function or could be a much more sophisticated perceptually motivated spectral weighting function. These methods collectively belong to the class of spectral subtraction or short-time spectral amplitude (STSA) estimation methods [1]-[2].

With the changing BGN environment in cell phone applications, the most critical component of these algorithms is the estimation of BGN power spectrum. There are two distinct set of approaches towards BGN spectrum estimation. The first approach makes use of an explicit voice activity detector (VAD) to update noise spectrum. However, VAD based techniques are difficult to tune and become unreliable in low-SNR and non-stationary noise conditions. The other approach is based on minimum statistics and probabilistic update [3] of noise spectrum has been proposed for improved speech quality.

Further improvements in single microphone noise suppression were reported in 1995 [4] with the introduction of subspace processing in noise suppression applications. These methods decompose the noisy speech subspace in to two subspaces, one corresponding to the noise-plus-speech subspace and one corresponding to the noise subspace. Noise suppression is performed by removing the noise subspace and estimating the signal from the remaining subspace. Extensions of subspace algorithms to non-stationary and non-Gaussian noise cases can be found in [5].

2.2 Multi-Microphone Noise Suppression

Significant progress in digital signal processor (DSP) technology has enabled multi-microphones based noise suppressors in cell phones. The main advantage of having multiple microphones is their ability to utilize spatial dimension in addition to temporal dimension to effectively suppress the BGN. The popularity of multi-microphones in cellular telephony appears to be growing with several dual-microphone noise suppression solutions already available in the marketplace.

A common goal of these techniques is to optimally combine the multi-sensor data to achieve improved noise suppression. The procedure is more commonly known as the beamforming. Beamforming has been a well studied area of signal processing for over three decades [6]; in particular, the delay-and-sum (DS) beamformer and the adaptive beamformer (ABF) are the most commonly used techniques suitable for BGN suppression purposes. ABF generally performs better than the DS, but requires *a priori* information about the *look direction*, and *speech absence* interval for ABF coefficient update.

Recently there has been a lot of interest in microphone array processing algorithms that do not require such *a priori* information. These approaches, under the assumption of linear speech/noise mixing are also known as blind source separation (BSS) and have been used for speech separation applications [7]. These approaches are based on one of the following criteria, 1.) Statistical independence of speech and noise, 2.) Signal diversities present in speech, e.g., time-frequency diversity, 3.) Temporal characteristics of speech, and 4.) Non-stationarity of speech signal. The techniques employing independence criteria are particularly popular and are collectively known as the Independent Component Analysis (ICA) techniques.

3. CHALLENGES

Single microphone solutions are easy to understand and low in computational complexity. However, most single-microphone solutions proposed report musical noise phenomenon associated with the processed speech. Musical noise manifests itself as randomly occurring pure tones in the processed speech and is very unpleasant to listen to. None of the solutions proposed so far claim complete elimination of musical noise in low SNR and non-stationary noise scenarios, despite the huge amount of research already been devoted to the problem. Hence effective noise suppression with a single microphone is still a challenge.

Multi-microphone systems have the potential of better noise suppression. In theory, a large array with many microphones is required for effective noise suppression either by beamforming or BSS methods. In beamforming for a fixed number of microphones, a larger array results in narrower mainlobe beamwidth, and for a fixed array length, more microphones result in lower sidelobe levels in the beampattern. Similarly, with BSS the system must have at least as many microphones as the number of sound sources for effect separation of the desired speech source from noise sources. For a fixed array geometry and fixed number of microphones, the dependency of beampattern on the signal frequency further complicates the design of such systems. Another limitation of these techniques is the heavy dependency on the application environment and the nature of the noise field. There have been recent performance studies of dual microphone systems under different microphone configurations and varying noise scenarios including directional and diffuse noise conditions in [8]-[9]. It remains, however, to study the amount of gain that can be obtained by (on-line) beamforming or BSS methods with small number of microphones under test conditions/signals specified by ETSI in ETSI EG 202 396-1 v1.2.3 (2009-03).

3.1 Performance evaluation

Assessing speech quality and intelligibility of the speech processed by the noise suppression system in the presence of different types of noise environments is an important but difficult task. A standardized set of test cases is one of the most critical aspects in noise suppression algorithm evaluation. Telecommunication standardization bodies like ITU/ETSI/TIA all have recommended test cases for evaluating the noise suppressor performance and its impact on speech quality. The noise suppression performance is generally measured in terms of ambient noise rejection (ANR) value as described in ETSI TS 126 132 v4.1.0 (2001-09). Tests for algorithm impact on loudness ratings can be found in ITU-T P.79. Processed voice quality subjective and objective guidelines can be found in ITU-T P.800/P.835 and P.862 respectively. Recommendations for acceptable algorithmic delays, convergence time, active speech level changes etc, are described in 3GPP TS 26.077 V4.0.0 (2001-03) and ITU-T G.114. Artificial test signals are also described in ITU-T P.50 in order to reduce the variability associated with real speech. Note that most of the above mentioned test cases involving noise sources assume noise source stationarity. The development of test cases suitable for non-stationary noise sources is still underway and new cases are constantly being added (ITU-T P.835 Amendment 1: New Appendix III) to existing test cases.

4. CONCLUSIONS

This paper describes the current state-of-the-art of noise suppression algorithms suitable for cellular telephony. An overview of operating principles behind both single-microphone and multi-microphone systems was provided. Associated challenges along with performance guidelines were discussed. This paper aims to give algorithm developers an overview of challenges involved in the design of noise suppression algorithms for cellular telephony.

REFERENCES

[1] Boll S. F., Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoustics, Speech, Signal Processing, vol. 2, 113-120, 1979.

[2] Ephraim Y. et al., Recent advancements in speech enhancement. The Electrical Engg. Handbook, CRC Press, 2004.

[3] Cohen I., Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. IEEE Trans. Speech, Audio Processing, Vol. 11, 466-475, 2003.

[4] Ephraim Y., et al., A signal subspace approach for speech enhancement. IEEE Trans. Speech, Audio Processing, Vol. 4, 251-266, 1995.

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

[6] Krim H., et al., Two decades of array signal processing research. IEEE signal processing magazine, vol. 13, no. 4, 1996.

[7] Makino S., et al., Blind Speech Separation, Springer, 2007.

[8] Kocinski J., Speech Intelligibility improvement using convolutive blind source separation assisted by denoising algorithms., Speech Communication, vol. 50, no. 1, Jan. 2008.

[9] Chen J., et al., Performance evaluation of adaptive dual microphone systems. Speech Communication, *In Press*.