Multi-microphone Noise Reduction System Integrating Nonlinear Multi-band Spectral Subtraction

Radu Mihnea Udrea, Claudia Cristina Oprea(✉), and Cristian Stanciu

Telecommunication Department, University Politehnica of Bucharest,
Iuliu Maniu 1-3, 061071 Bucharest, Romania
{mihnea, cristina, cristian}@comm.pub.ro

Abstract. This paper presents a robust system to improve speech signals processed by communication systems. The system includes multi-microphone techniques, for which both spectral and spatial characteristics of the signal sources can be used. Also a spectral subtraction algorithm for noise reduction is integrated into the system. The modified spectral subtraction method takes into account the non-uniform effect of colored noise on the speech spectrum and improves the multi-microphone noise filtering.

Keywords: Speech enhancement · Multi-microphone noise reduction

1 Introduction

Speech signal is often accompanied by environmental noise. There are several negative effects during processing the degraded speech for applications such as: automobile speech communication systems, voice recognition systems, speech recognition, speaker authentication.

Improvement techniques can be classified as a single channel, dual-channel and multiple channel speech enhancement techniques. Techniques to improve single channel speech [1] apply to situations in which only one microphone is available. In dual channel enhancement techniques, a reference signal for the noise is available and therefore adaptive noise cancellation technique can be applied. Multiple channel techniques use microphone array [2] and take advantage of the availability of multiple signal inputs to our system to make possible to use the phase alignment to reject unwanted noise components [3]. Beamforming is a multi-microphone signal processing technique that achieves a more directional pattern than what could be obtained with only one microphone.

The spatial-filter-based beamformers have been developed for narrow-band signals [4], which can be characterized by a single frequency. For the speech signal, which has a broadband frequency domain, the beamformers will not yield the same model for different frequencies and the beamwidth decreases as frequency increases. If we use such a beamformer when the steering direction is different from the incident angle of the source, the source signal will be low-pass filtered. In addition, the noise coming

N. Oliver et al. (Eds.): MindCare 2016/Fabulous 2016/IIoT 2015, LNICST 207, pp. 133–138, 2018.
https://doi.org/10.1007/978-3-319-74935-8_19

from another direction, will not be attenuated evenly across its entire spectrum, resulting in some disturbing artifacts in the output array.

In this paper we propose integrating a nonlinear multi-band spectral subtraction noise reduction method into the beamforming system. The multi-band spectral subtraction applies different subtraction factors depending on the SNR in each frequency band. Because the beamformer will not attenuate uniformly the noise over entire spectrum, the proposed method will compensate this non-uniformity. The simulations are performed using a car environment with engine noise background.

2 Microphone Array Processing

Consider an array of M microphones in a reverberant and acoustical noisy environment. The i^{th} microphone output can be expressed as [2]:

$$y_i(n) = s(n) * h_i(n) \tag{1}$$

where $s(n)$ represents the clean speech signal, $h_i(n)$ denotes the impulse response between the speech source and the i^{th} microphone and * denotes convolution.

There are fixed and adaptive beamforming systems. Fixed beamformer focus on source direction and therefore captures less noise and reverberation arrive from a different direction than the source. Adaptive beamformer provides better noise reduction, but it doesn’t reduce the reverberation on other directions.

In a conventional delay-and-sum beamformer (DSB), $y_i(n)$ is first shifted by a time-delay n_i and then scaled by a corresponding weight w_i. The resulting delayed and scaled signals from all microphones are then summed to produce the beamformer output $z(n)$:

$$z(n) = s(n) * g(n) \tag{2}$$

where

$$g(n) = \sum_{i=1}^{M} w_i h_i(n - n_i) \tag{3}$$

The purpose of the delays n_i is to time-align the direct path components of the impulse responses $h_i(n)$ so as to steer the beamformer in the direction of the desired speech source. This way, the direct-path signals are phase-aligned and reinforced while echoes apart from the steering direction are attenuated.

The fixed spatial-filter beamformer directivity pattern will have a main lobe on the direction of the source speech signal and several secondary lobes. The characteristic changes depending of the frequency as shown in Fig. 1.

For the speech signal, which has a broadband frequency domain, such beamformers will not offer the same filtering model for different frequencies. As seen in Fig. 1 the beamwidth decreases as frequency increases. Therefore, the source signal will be low-pass filtered and the noise coming from another direction will not be attenuated evenly across its entire spectrum, resulting in some disturbing artifacts in the output array.

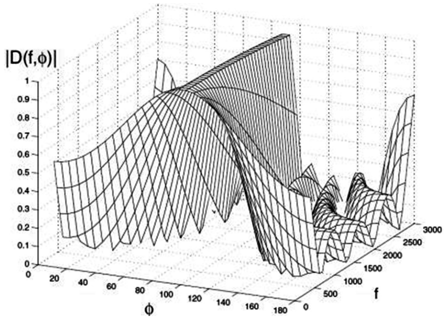


Fig. 1. The directivity characteristic of a fixed DS beamformer depending on the steering angle ϕ and frequency *f*.

3 The Spectral Subtraction Method

The basic assumption of the method is treating the noise as uncorrelated additive noise. Assume that a speech signal $s(n)$ has been degraded by the uncorrelated additive noise signal $d(n)$:

$$y(n) = s(n) + d(n) \tag{4}$$

Short time power spectrum of the noisy speech can be approximated by:

$$|Y(k)|^2 \approx |S(k)|^2 + |D(k)|^2 \tag{5}$$

The power spectral subtraction estimator results by replacing noise square-magnitude $|D(k)|^2$ with its average value taken during non-speech activity period.

$$\hat{\sigma}_d^2(k) \simeq E\left\{|D(k)|^2\right\} \tag{6}$$

Berouti [1] proposed an important variation of spectral subtraction for reduction of residual musical noise. An overestimate of the noise power spectrum is subtracted and the resulted spectrum is limited from going below a preset minimum level (spectral floor). The proposed algorithm could be expressed as:

$$\left|\hat{S}(k)\right|^2 = \begin{cases} |Y(k)|^2 - \alpha \cdot \hat{\sigma}_d^2(k), & \text{if } \left|\hat{S}(k)\right|^2 > \beta \cdot \hat{\sigma}_d^2(k) \\ \beta \cdot \hat{\sigma}_d^2(k), & \text{otherwise} \end{cases} \tag{7}$$

where α is the over-subtraction factor and β is the spectral floor parameter.

To reduce the speech distortion caused by large values of α, its value is adapted from frame to frame [5]. The basic idea is to take into account that the subtraction process must depend on the segmental noisy signal to noise ratio (NSNR) of the frame, in order to apply less subtraction with high NSNRs and vice versa.

In real environments, noise spectrum is not uniform for all the frequencies [6]. For example, in the case of engine noise the most of noise energy is concentrated in low frequency. To take into account the fact that colored noise affects the speech spectrum differently at various frequencies, a multi-band linear frequency spacing approach to spectral over-subtraction was proposed in [7]. The speech spectrum is divided into N non-overlapping bands, and spectral subtraction is performed independently in each band. The estimate of the clean speech spectrum in the *i*-th band is obtained by:

$$\left|\hat{S}_i(k)\right|^2 = |Y_i(k)|^2 - \alpha_i \cdot \hat{\sigma}_d^2(k), \quad v_i < k < v_{i+1} \tag{8}$$

where k is the frequency bin for the spectrum computed using the discrete Fourier transform, v_i and v_{i+1} are the beginning and ending frequency bins of the *i*-th frequency band and α_i is the over-subtraction factor of the *i*-th band.

The over-subtraction factor α_i can be calculated as:

$$\alpha_i = \begin{cases} 1 & \gamma_i \geq 20\,\mathrm{dB} \\ \alpha_0 - \frac{3}{20}\gamma_i & -5\,\mathrm{dB} \leq \gamma_i \leq 20\,\mathrm{dB} \\ 4.75 & \gamma_i \leq -5\,\mathrm{dB} \end{cases} \tag{9}$$

where $\alpha_0 = 4$ and the aposteriori NSNR γ_i of the *i*-th frequency band is:

$$\gamma_i(dB) = 10\log_{10}\frac{\sum_{k=w_i}^{w_{i+1}} |X_i(k)|^2}{\sum_{k=w_i}^{w_{i+1}} \hat{\sigma}_d^2(k)^2} \tag{10}$$

A nonlinear frequency spacing approach for multi-band over-subtraction factor estimation was proposed in [7] based on the fact that human ear sensibility varies nonlinear in frequency spectrum. A perceptual spectral estimation of critical bandwidth was involved, denoting the noise bandwidth limit at which the detection threshold of the signal (tone) ceased to increase. The noise power within the same critical band with the signal is then equal to the product of the measured power spectral density and the critical bandwidth of the band in question.

4 Implementation and Experimental Results

We simulate a microphone array configuration to enhance the speech signal inside an automobile. We considered a linear array with a variable number of 2 to 6 microphones equally spaced at a distance of 0.2. The speech signal source is placed in front of the array at a distance of 0.5 m. Two types of noise were used for experiments: Gaussian white noise and engine recorded noise. The white noise was uniformly added at different SNR over each microphones, while engine noise source was placed at a distance of 1 m behind the last microphone of the array.

The signals received through the microphones were applied to a fixed DS beamformer designed to enhance the direction of the desired speech source. For multi-band spectral over-subtraction we used nonlinear frequency spacing with a number of 4 bands that gives an optimal speech quality [7].

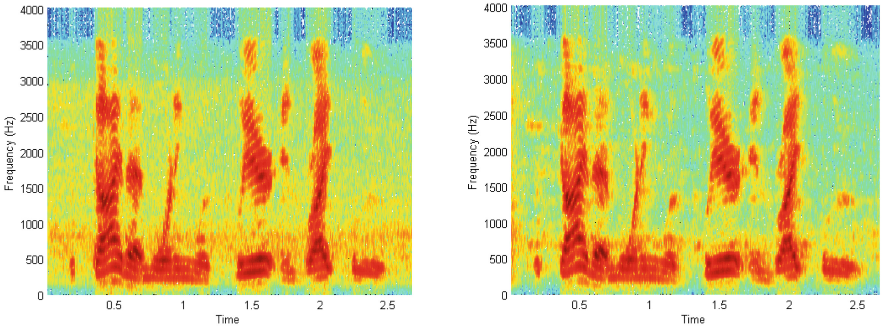
Objective and subjective quality evaluation methods were applied to establish the performance of the algorithms presented in this study. In Table 1 the simulations show that the Mean Opinion Score (MOS) computed from ITU-T Recommendation P.862 (PESQ) [8] is increasing when using more than two microphones. Increasing the number of microphones more than four does not give an increasing of quality.

Table 1. PESQ MOS evaluation for the enhanced speech.

Input SNR	Output of the DS beamformer	Bark spaced four multi-band spectral over-subtraction				
Number of microphones		2	3	4	5	6
0 dB	1.75	1.79	1.82	1.85	1.84	1.84
5 dB	1.85	1.90	1.99	1.97	1.98	1.97
10 dB	2.26	2.41	2.44	2.46	2.45	2.40
15 dB	2.82	2.86	2.88	2.90	2.89	2.84

Subjective listening tests indicate that, using the fixed DS beamforming followed by non-linear Bark spaced multi-band over-subtraction, a very good speech quality with less musical noise and with minimal speech distortion is obtained.

Figure 2 shows the spectrogram for speech signal "The sky this morning was clear and light blue" affected by car engine noise, at a SNR of 10 dB at the output of the DS beamformer and the spectrogram of the enhanced speech obtained using over-subtraction with four non-linear Bark spaced bands.



(a) (b)

Fig. 2. Spectrogram of speech signal "The sky this morning was clear and light blue" affected by car noise (a) at the output of the DS beamformer (b) after the multi-band spectral over-subtraction was applied (Color figure online)

5 Conclusions

This paper presents an improved noise reduction system using multi-microphone signal processing and a spectral subtraction method that takes into account the non-uniform effect of colored noise on the speech spectrum. The proposed method uses a nonlinear frequency spacing approach for multi-band over-subtraction factor estimation. This compensates the fact that the beamformers will not filter the noise for different frequencies since the beamwidth decreases as frequency increases.

The proposed method also reduces the residual musical tones that appear in the case of conventional power spectral subtraction. Simulations with different types of noise and different configurations for microphone arrays show a better quality for the enhanced speech when using the multi-band spectral subtraction method after the multi-microphone signal processing.

Acknowledgment. This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-1880.

References

1. Berouti, M., Schwartz, R., Makhoul, J.: Enhancement of speech corrupted by acoustic noise. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 208–211, April 1979
2. Benesty, J., Chen, J., Huang, Y.: Microphone Array Signal Processing. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-78612-2
3. Souden, M., Chen, J., Benesty, J., Affes, S.: An integrated solution for online multichannel noise tracking and reduction. IEEE Trans. Audio Speech Lang. Process. **19**(7), 2159–2169 (2011)
4. Cornelis, B., Doclo, S., van dan Bogaert, T., Moonen, M., Wouters, J.: Theoretical analysis of binaural multimicrophone noise reduction techniques. IEEE Trans. Audio Speech Lang. Process. **18**(2), 342–355 (2010)
5. Udrea, R.M., Ciochina, S.: Speech enhancement using spectral over-subtraction and residual noise reduction. In: International Symposium on Signals, Circuits and Systems, pp. 165–169. IEEE Press, Iasi, Romania (2003). https://doi.org/10.1109/SCS.2003.1226974
6. Kamath, S., Loizou, P.: A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE Press, Orlando (2002). https://doi.org/10.1109/ICASSP.2002.5745591
7. Udrea, R.M., Vizireanu, N., Ciochina, S., Halunga, S.: Nonlinear spectral subtraction method for colored noise reduction using multi-band Bark scale. Sig. Process. **88**(5), 1299–1303 (2008)
8. ITU-T, Perceptual evaluation of speech quality PESQ, an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, ITU-T Recommendation P.862 (2000)