

A COHERENCE-BASED ALGORITHM FOR NOISE REDUCTION IN DUAL-MICROPHONE APPLICATIONS

Nima Yousefian, Kostas Kokkinakis and Philipos C. Loizou

Center for Robust Speech Systems,
Department of Electrical Engineering, University of Texas at Dallas,
800 West Campbell Road, Richardson, TX 75080, USA
nimayou@student.utdallas.edu, kokkinak@utdallas.edu, loizou@utdallas.edu

ABSTRACT

In this paper, we present a novel coherence-based dual-microphone noise reduction approach and show how the proposed technique can capitalize on the small microphone spacing in order to suppress coherent noise present inside a realistic reverberant environment. Listening tests with normal-hearing subjects conducted in a two-microphone array configuration, reveal that the proposed method outperforms the generalized sidelobe canceller (GSC), which is commonly used in suppressing coherent noise.

1. INTRODUCTION

Noise is detrimental to speech recognition. In real-life signal processing, speech is often disturbed by additive noise components. Single microphone speech enhancement algorithms are favored in many applications because they are relatively easy to apply. Their performance, however, is limited, especially when the noise is non-stationary. In recent years, with the significant progresses seen in digital signal processors, two-microphone configurations are receiving a lot attention for tasks such as directional audio capturing, noise reduction and even ‘blind’ speech dereverberation (e.g., see [2, 5, 6, 12]).

There are three types of noise fields: (1) incoherent noise caused by the microphone circuitry, (2) coherent noise generated by a single well-defined directional noise source and (3) diffuse noise which is characterized by uncorrelated noise signals of equal power propagating in all directions simultaneously. In coherent noise fields, the noise signals captured by the microphone array are highly correlated. In this scenario, the performance of methods that may work well in diffuse fields starts to degrade. This has prompted many to suggest techniques for noise reduction in coherent noise fields.

One of the most popular techniques, known to be extremely powerful in suppressing coherent noise is the generalized sidelobe canceller (GSC) [8] which is an adaptive noise cancellation technique that can null out the interfering noise source. The authors in [3] have shown that the noise reduction performance of GSC theoretically reaches infinity for coherent noise.

Another technique widely used for reduction of uncorrelated noise, which was first proposed in [1], is to use the coherence function of noisy signals. The premise behind coherence-based methods is that speech signals in the two channels are correlated, while the noise signals are uncorrelated. Indeed, if the amplitude of the coherence function between the noisy signals at the two channels is one or close to one the speech signal is predominant and it must be passed without distortion. Although coherence-based methods work well when the noise components are uncorrelated, they are deficient when dealing with coherent noise [11]. In recent years, many authors have proposed approaches that can suppress coherent noise by simply relying on the cross-power spectral density of the noise components at the two microphone channels (e.g., see [4, 9, 11, 14]).

In this paper, we propose a new coherence-based dual-microphone noise reduction method, which is capable of reducing coherent noise substantially. Listening tests conducted with normal-hearing listeners reveal that the proposed method outperforms the conventional generalized sidelobe canceller (GSC).

2. OVERVIEW OF COHERENCE-BASED METHODS

Let us consider the scenario in which the noise and target speech signals are spatially separated. The listener is wearing a behind-the-ear (BTE) hearing aid (or cochlear implant) equipped with two microphones, with small spacing between them. In this case, the noisy speech signals, after delay compensation, can be defined as:

$$y_i(m) = x_i(m) + n_i(m) \quad (i = 1, 2) \quad (1)$$

where i indicates the microphone index, m is the sample index and $x_i(m)$ and $n_i(m)$ represent the (clean) speech components and noise components at each sensor, respectively. After applying a short-time discrete Fourier transform (DFT) on both sides of Eq. (1), the signals captured by the two microphones are expressed in the frequency-domain as follows:

$$Y_i(f, k) = X_i(f, k) + N_i(f, k) \quad (i = 1, 2) \quad (2)$$

where f is the frequency bin and k is the frame index, respectively. Assuming that the noise and speech components are

This work was supported by Grants R03 DC008882 (K. Kokkinakis) and R01 DC007527 (P. C. Loizou) awarded from the National Institute on Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health (NIH).

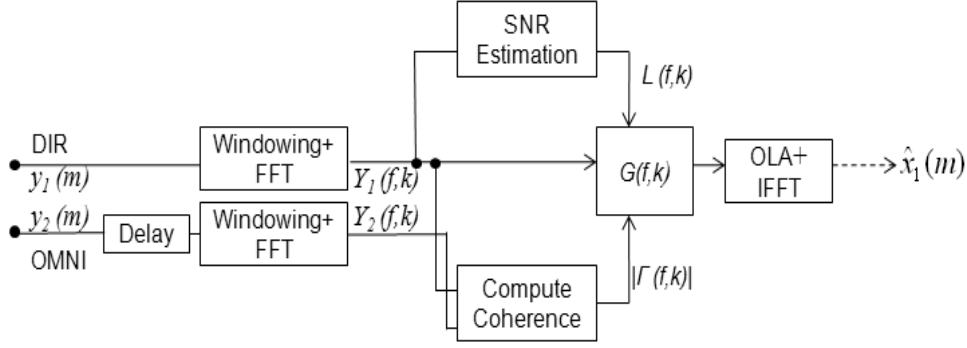


Figure 1: Block diagram of the proposed two-microphone speech enhancement technique.

uncorrelated, the cross-power spectral density of the noisy signals, can be written as:

$$P_{Y_1 Y_2}(f, k) = P_{X_1 X_2}(f, k) + P_{N_1 N_2}(f, k) \quad (3)$$

where $P_{uv}(f, k)$ denotes the cross-spectral density defined as $P_{uv}(f, k) = E[U(f, k)V^*(f, k)]$. In situations where the speech signals are correlated (e.g., when reverberation is present), and the noise sources are uncorrelated, one can use the coherence function as an objective criterion to determine if the target speech signal is present or absent at a specific frequency bin. The magnitude coherence function between the signals $y_1(t)$ and $y_2(t)$ is defined as:

$$|\Gamma_{Y_1 Y_2}(f, k)| = \frac{|P_{Y_1 Y_2}(f, k)|}{\sqrt{P_{Y_1}(f, k)}\sqrt{P_{Y_2}(f, k)}} \quad (4)$$

The coherence function has been used in several recent studies (e.g., see [4, 9, 14]) to suppress uncorrelated frequency components, while allowing correlated components (presumably containing target speech information) to pass.

The above technique leads to effective noise reduction in diffuse noise fields and in scenarios wherein the distance between the microphones is large. Theoretically for ideal diffuse noise fields, the coherence function assumes the shape of a sinc function with the first zero crossing at $f_c = c/(2d)$ Hz, where c is the speed of sound and d is the microphone spacing [13]. Clearly, the smaller the spacing the larger the range of frequencies for which the coherence is high (near one). For our hearing aid application at hand, where the distance between the two microphones is fairly small (~ 2 cm), the above approach might not be always effective in reducing noise. A different approach is discussed next.

3. PROPOSED DUAL-MICROPHONE NOISE REDUCTION METHOD

Before describing the proposed suppression function, we first derive the relationship between the coherence of the noisy and noise-source signals. After dividing both sides of Eq. (3) by $\sqrt{P_{Y_1} P_{Y_2}}$ and after omitting the f and k indices for better clarity, we obtain:

$$\Gamma_{Y_1 Y_2} = \frac{P_{X_1 X_2}}{\sqrt{P_{Y_1} P_{Y_2}}} + \frac{P_{N_1 N_2}}{\sqrt{P_{Y_1} P_{Y_2}}} \quad (5)$$

which can be re-written as:

$$\Gamma_{Y_1 Y_2} = \Gamma_{X_1 X_2} \left(\frac{\sqrt{P_{X_1} P_{X_2}}}{\sqrt{P_{Y_1} P_{Y_2}}} \right) + \Gamma_{N_1 N_2} \left(\frac{\sqrt{P_{N_1} P_{N_2}}}{\sqrt{P_{Y_1} P_{Y_2}}} \right) \quad (6)$$

After using Eq. (3), Eq. (6) becomes:

$$\begin{aligned} \Gamma_{Y_1 Y_2} &= \Gamma_{X_1 X_2} \sqrt{\frac{P_{X_1}}{P_{X_1} + P_{N_1}}} \sqrt{\frac{P_{X_2}}{P_{X_2} + P_{N_2}}} \\ &+ \Gamma_{N_1 N_2} \sqrt{\frac{P_{N_1}}{P_{X_1} + P_{N_1}}} \sqrt{\frac{P_{N_2}}{P_{X_2} + P_{N_2}}} \end{aligned} \quad (7)$$

Now let SNR_i be the true speech-to-noise ratio at the i -th channel, which is defined by:

$$\text{SNR}_i = \frac{P_{X_i}}{P_{N_i}} \quad (8)$$

Substituting this expression in Eq. (7), we obtain:

$$\begin{aligned} \Gamma_{Y_1 Y_2} &= \Gamma_{X_1 X_2} \left(\sqrt{\frac{\text{SNR}_1}{1 + \text{SNR}_1}} \sqrt{\frac{\text{SNR}_2}{1 + \text{SNR}_2}} \right) \\ &+ \Gamma_{N_1 N_2} \left(\sqrt{\frac{1}{1 + \text{SNR}_1}} \sqrt{\frac{1}{1 + \text{SNR}_2}} \right) \end{aligned} \quad (9)$$

This last equation reveals that the coherence function between the noisy signals is, in fact, dependent on both the coherence of the target speech and noise signals. Given the small microphone spacing in our application, we can further make the assumption that the SNR values at the two channels are nearly identical, such that $\text{SNR}_1 \simeq \text{SNR}_2$. Based on this assumption, we can conclude that at higher SNRs the coherence of the noisy signals is affected primarily by the coherence of the speech signals, while at lower SNRs it is affected by the coherence of the noise signals. Put differently, we can deduce the following:

$$|\Gamma_{Y_1 Y_2}| \approx \begin{cases} |\Gamma_{X_1 X_2}|, & \text{if } \text{SNR} \rightarrow +\infty \\ |\Gamma_{N_1 N_2}|, & \text{if } \text{SNR} \rightarrow 0 \end{cases} \quad (10)$$

The above equation suggests that the desired suppression function needs to account for the dependence of SNR and coherence of speech and noise signals. In our hearing aid application, we assume that the spacing between the two microphones is small. We further assume that the target speech signal originates from the front (0°), typically at a distance of 1 m from the listener, while the noise source(s) originate

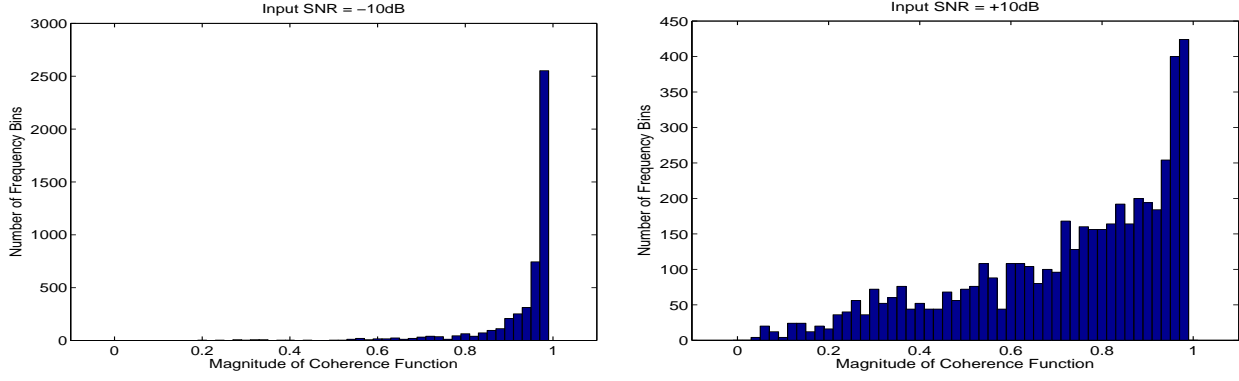


Figure 2: Distribution of the amplitude of the coherence function $|\Gamma_{Y_1 Y_2}|$ estimated for 10 successive frames of a noisy signal at SNR = -10 dB (left) and SNR = +10 dB (right). The noise source is speech-shaped noise located at 90° azimuth.

from either of the two hemifields (e.g., at 90°). Under these assumptions, we noted that at low SNR levels, the noise sources are correlated, and thus they have a coherence close to one. To demonstrate this, the histogram of the coherence function (accumulated for all frequencies) for 10 successive frames of noisy signals at SNR = -10 dB and SNR = +10 dB is compared in Figure 2. By observing Figure 2, it quickly becomes apparent that at low SNR levels the coherence function assumes values near 1, while at higher SNR levels the coherence values span across the whole range of [0,1].

The aforementioned observations suggest the use of a suppression function, which at low SNR levels attenuates the frequency components (dominated presumably by noise) having a coherence close to 1, while allowing the remaining frequency components (dominated by the target speech) to pass. We thus consider the following suppression function:

$$G(f, k) = 1 - |\Gamma_{Y_1 Y_2}(f, k)|^{L(f, k)} \quad (11)$$

where $\Gamma_{Y_1 Y_2}(f, k)$ is the coherence of the noisy signals at the two sensors and $L(f, k) \geq 1$ is a parameter that depends on the estimated SNR at frequency bin f . Figure 3 shows a plot of function $g(x) = 1 - x^L$ for different values of L and for all $0 \leq x \leq 1$. As it can be seen, for small values of L , and subsequently small values of SNR, the function $g(x)$ attenuates all frequency components with coherence near one. On the other hand, for large values of L , and subsequently large values of SNR, the function $g(x)$ allows the frequency components to pass.

In the present study, the parameter $L(f, k)$ in Eq. (11) is set to be proportional to the estimated SNR and is computed as follows:

$$L(f, k) = \begin{cases} 1 & \text{if } \xi(f, k) < -20 \text{ dB} \\ 2 \left\lfloor \frac{\xi(f, k)}{5} \right\rfloor + 5 & \text{otherwise} \\ 512 & \text{if } \xi(f, k) > +20 \text{ dB} \end{cases} \quad (12)$$

where $\xi(f, k)$ is the estimated *a priori* SNR in frame k and bin f computed using the ‘decision-directed’ approach [7]:

$$\begin{aligned} \xi(f, k) &= a \left[\frac{|G(f, k-1) Y_1(f, k-1)|^2}{\hat{N}_1^2(f, k-1)} \right] \\ &+ (1-a) \max[\gamma(f, k) - 1, 0] \end{aligned} \quad (13)$$

where parameter $a = 0.98$, $G(f, k-1)$ represents the suppression function at frame $k-1$ and frequency bin f , $\hat{N}_1^2(f, k-1)$ is the estimate of the noise power spectrum and $\gamma(f, k) = Y_1^2(f, k) / \hat{N}_1^2(f, k)$.

Note that in this work, we resort to the noise estimation algorithm proposed in [15] for estimating $\hat{N}_1^2(f, k)$. To further reduce the variance (across frequency) of $\xi(f, k)$ in Eq. (13), we divided the spectrum into four bands (0-1 kHz, 1-2 kHz, 2-4 kHz and 4-8 kHz) and averaged the corresponding $\xi(f, k)$ values in each band. The averaged $\xi(b, k)$ values in band b , where $b = 1, 2, 3, 4$, were subsequently used in Eq. (12) to compute the band $L(b, k)$ values. The $L(b, k)$ values were subsequently smoothed over time with a forgetting factor of 0.995. This was done to reduce musical-noise type of distortion typically associated with sudden changes in the suppression function $G(f, k)$.

The block diagram of the proposed two-microphone speech enhancement algorithm is depicted in Figure 1. The signals collected at the two microphones are first processed in 30 ms frames with a Hanning window and with a 50% overlap between successive frames. After computing the short-time Fourier transform of the two signals, the cross-power spectral density $P_{Y_1 Y_2}$ is computed based on the following recursive averaging:

$$P_{Y_1 Y_2}(f, k) = \lambda P_{Y_1 Y_2}(f, k-1) + (1-\lambda) Y_1(f, k) Y_2^*(f, k) \quad (14)$$

where $\lambda = 0.6$. A more thorough discussion on optimal settings of the parameter λ can be found in [9]. The cross-spectral density $P_{Y_1 Y_2}$ is used in Eq. (4) to compute the magnitude of the coherence function, which is in turn used in Eq. (11). Next, Eq. (13) is used to estimate the SNR at time-frequency cell (f, k) , from which the power exponent $L(f, k)$ is derived according to Eq. (12). The resulting suppression function $G(f, k)$ described in Eq. (11) is then applied to $Y_1(f, k)$, corresponding to the Fourier transform of the noisy input signal captured by the directional microphone. To reconstruct the enhanced signal in the time-domain, we apply an inverse FFT and we synthesize it using the overlap-add (OLA) method.

4. EXPERIMENTAL RESULTS

Modern hearing aid devices come furnished with more than one microphone and thus offer the capacity to integrate intelligent dual-microphone noise reduction strategies in order to

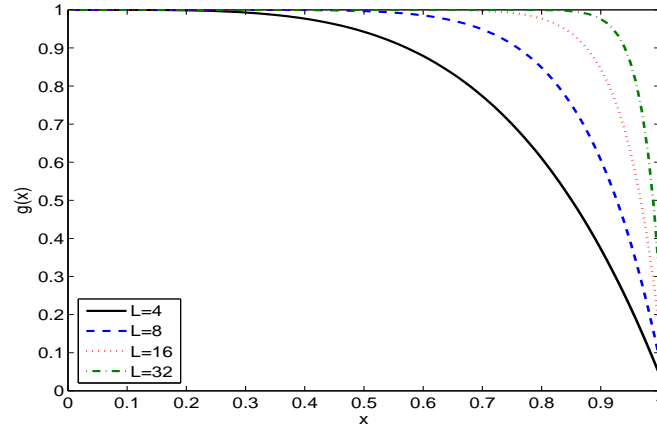


Figure 3: The proposed suppression function $g(x) = 1 - x^L$.

enhance noisy incoming signals. A number of studies have shown that the overall improvement that can be achieved in terms of SNR with the use of an additional directional microphone alone can be 3–5 dB when compared to processing with just an omni-directional microphone [2, 16]. Beamformers can be considered an extension of differential microphone arrays, where the suppression of noise is carried out by adaptive filtering of the noisy signals. An attractive realization of adaptive beamformers is the generalized side-lobe canceller (GSC) structure [8]. In this paper, we compare our coherence-based method with the adaptive beamforming technique proposed in [16]. Throughout the remainder of this paper, GSC refers to the implementation of the technique implemented in [16].

The speech stimuli used in our experiment were sentences from the IEEE database [10]. The IEEE speech corpus contains phonetically-balanced sentences (approximately 7–12 words each) and was designed specifically for assessment of speech intelligibility. Two types of noise were used in the present study: (1) speech-shaped noise and (2) multi-talker babble noise. The noisy stimuli at the pair of microphones were generated by convolving the target and noise sources with a set of HRTFs measured inside a mildly reverberant room ($T_{60} \simeq 300$ ms) with dimensions 5.50 m \times 4.50 m \times 3.10 m (length \times width \times height). The HRTFs were measured using identical microphones to those used in modern hearing aids. In our simulation, the target speech sentences originated from the front of the listener (0° azimuth) while the noise source originated from the right of the listener (90° azimuth). Although, in this work, we only report simulation results obtained for 90° azimuth, similar outcomes were observed for other angles as well.

The noisy sentence stimuli at SNR = -10, -5 and 0 dB were processed using the following conditions: (1) the input to the directional microphone, (2) the GSC algorithm and (3) the proposed coherence-based algorithm. The performance obtained with the use of the directional microphone alone will be used for baseline purposes to assess relative improvements in performance when no processing is taking place. The GSC algorithm is an adaptive beamforming algorithm, which has been used widely in both hearing aid and cochlear implant devices [2, 16]. In our implementation, we used a 128-tap adaptive filter and also a fixed FIR filter as a spatial pre-processor as proposed in [16]. The array configuration used in [16] is the same as in the present study, i.e., it consists of a front directional microphone and a rear omni-directional microphone.

A total of seven normal-hearing listeners, all native speakers of American English, were recruited for the listening tests. In total, there were 18 different listening conditions (3 algorithms \times 3 SNR levels \times 2 types of noise). Two IEEE lists (20 sentences) were used for each condition. The processed sentences were presented to the listeners via headphones at a comfortable level. The mean intelligibility scores, obtained by computing the total number of words identified correctly are shown in Figure 4.

As shown in Figure 4, the proposed coherence-based algorithm outperforms the GSC algorithm, particularly at low SNR levels (-10 dB and -5 dB) and for both types of noise. A substantial improvement in intelligibility was obtained with the proposed coherence-based algorithm relative to the baseline condition (directional microphone input) in all conditions. The intelligibility scores at -10 dB SNR (multi-talker babble) improved from near 0% with the directional microphone and from 28% with the GSC algorithm to near 60% with the proposed coherence-based algorithm. The overall improvement in intelligibility with the coherence-based algorithm was maintained in multi-talker babble (non-stationary) conditions. Such conditions are particularly challenging for the GSC algorithm, since the adaptive filter needs to track sudden changes to the noise signals in the background.

5. CONCLUSIONS

In this work, we have developed a novel coherence-based technique for dual-microphone noise reduction. Although, coherence-based techniques are more often used for suppressing uncorrelated noises, we have shown that such methods can be also used for coping with coherent noises. Suppressing coherent noises is a challenging problem, which has been thoroughly addressed in this paper. The simplicity of our implementation and the positive outcomes in terms of intelligibility make this method a potential candidate for future use in commercial hearing aid and cochlear implant devices.

REFERENCES

- [1] J. B. Allen, D. A. Berkley and J. Blauert, “Multi-microphone signal processing technique to remove room reverberation from speech signals,” *J. Acoust. Soc. Amer.*, vol. 62, pp. 912–915, October 1977.
- [2] J. V. Berghe and J. Wouters, “An adaptive noise canceller for hearing aids using two nearby microphones,” *J. Acoust. Soc. Amer.*, vol. 103, pp. 3621–3626, June 1998.

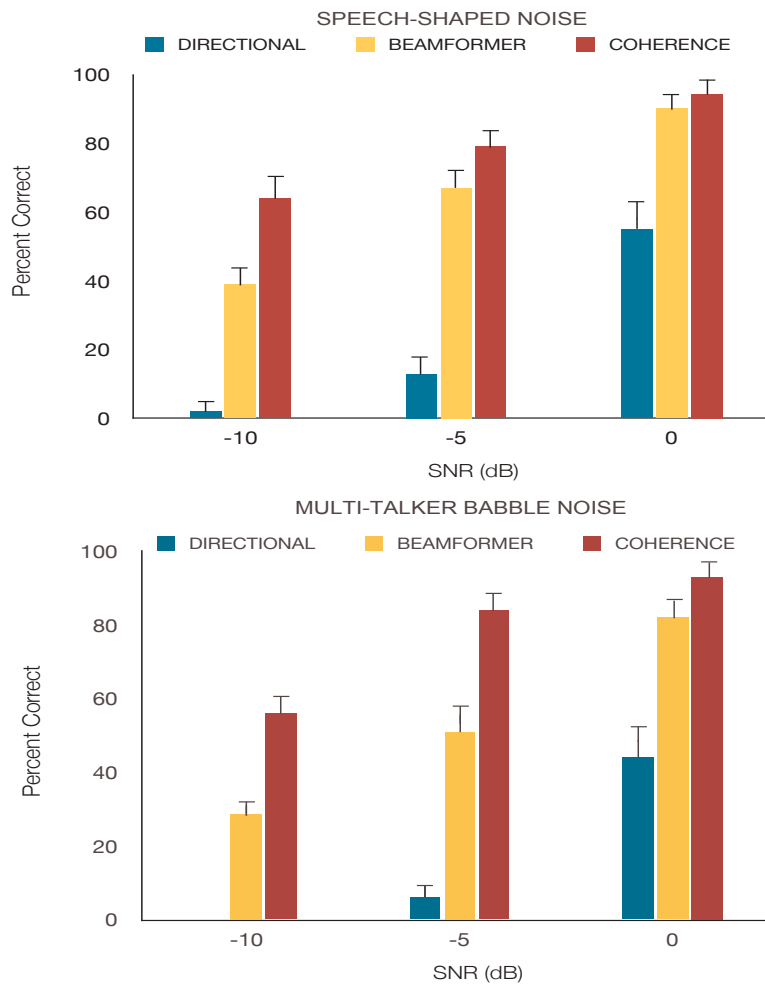


Figure 4: Mean percent word recognition scores for seven normal-hearing listeners tested on IEEE sentences embedded in speech-shaped noise (top) and multi-talker babble noise (bottom) at SNR = 0 dB, -5 dB and -10 dB. Scores for sentences processed through a directional microphone only are shown in blue. Scores for sentences processed through the GSC beamformer are plotted in yellow. Scores for sentences processed through the proposed coherence-based algorithm are shown in red. Error bars indicate standard deviations.

- [3] J. Bitzer, K. U. Simmer and K.-D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *Proc. ICASSP 1999*, Phoenix, AZ, March 15–19, 1999, pp. 2956–2968.
- [4] R. L. Bouquin-Jeannès, A. A. Azirani and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 484–487, September 1997.
- [5] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer Verlag, 2001.
- [6] J. Chen, K. Phua, L. Shue and H. Sun, "Performance evaluation of adaptive dual microphone systems," *Speech Communication*, vol. 51, pp. 1180–1193, December 2009.
- [7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 32, pp. 1109–1121, December 1984.
- [8] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagation*, vol. 30, pp. 27–34, January 1982.
- [9] A. Guérin, R. L. Bouquin-Jeannès and G. Faucon, "A two-sensor noise reduction system: Applications for hands-free car kit," *EURASIP J. Applied Signal Process.*, vol. 11, pp. 1125–1134, March 2003.
- [10] IEEE Subcommittee, "IEEE recommended practice speech quality measurements," *IEEE Trans. Audio Electroacoust.*, vol. 17, pp. 227–246, September 1969.
- [11] J. M. Kates, "On using coherence to measure distortion in hearing aids," *J. Acoust. Soc. Amer.*, vol. 91, pp. 2236–2244, April 1992.
- [12] K. Kokkinakis and P. C. Loizou, "Selective-tap blind de-reverberation for two-microphone enhancement of reverberant speech," *IEEE Signal Process. Lett.*, vol. 16, pp. 961–964, November 2009.
- [13] H. Kuttruf, *Room Acoustics*, Elsevier Science Publishers Ltd, 1991.
- [14] M. Rahmani, A. Akbari and B. Ayad, "An iterative method for cross-PSD noise estimation for dual microphone speech enhancement," *Applied Acoustics*, vol. 70, pp. 514–521, March 2009.
- [15] S. Rangachari and P. C. Loizou, "A noise estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 48, pp. 220–231, February 2006.
- [16] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. Van Dijk, A. Van Wieringen and J. Wouters, "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom cochlear implant system," *Ear Hearing*, vol. 28, pp. 62–72, February 2007.