SPEECH DEREVERBERATION BASED ON A STATISTICAL MODEL OF LATE REVERBERATION USING A LINEAR MICROPHONE ARRAY

Emanuël A.P. Habets

Technische Universiteit Eindhoven
Department of Electrical Engineering
Signal Processing Systems group, bldg. EH 3.27
PO box 513, 5600 MB Eindhoven, The Netherlands
Email: e.a.p.habets@tue.nl

Introduction

In general, acoustic signals radiated within a room are linearly distorted by reflections from walls and other objects. These distortions degrade the fidelity and intelligibility of speech, and the recognition performance of automatic speech recognition systems. We have investigated the application of signal processing techniques to improve the quality of speech distorted in an acoustic environment.

Early room echoes mainly contribute to coloration or spectral distortion. The most important effect of reverberation on speech results from the non-stationarity of speech. An effect of reverberation is a lengthening of the speech units, so that the reverberation 'tails' of previous sounds overlap the subsequent sounds. This phenomenon is referred to as *overlap-masking*. Reverberation reduction processes may generally be divided into single or multiple microphone methods and into those primarily affecting coloration or those affecting reverberant tails.

Lebart et.al. presented a single-channel speech dereverberation method based on Spectral Subtraction to reduce *overlap-masking* [1]. The instantaneous power spectrum of the late reverberation can be estimated based on Polack's statistical model of late reverberation. Polack's model describes the Room Impulse Response (RIR) as one realization of a non-stationary stochastic process. A simplified version of this model can be expressed as

$$h(t) = \begin{cases} 0 & t < 0 \\ b(t)e^{-\alpha t} & t \geq 0 \end{cases},$$

where $b(t)$ is a white zero–mean Gaussian stationary noise and α is linked to the reverberation time T_r through $\alpha \triangleq \frac{3\ln(10)}{T_r}$. The energy envelope of the RIR can be expressed as

$$E_h\{h^2(t)\} = \sigma^2 e^{-2\alpha t} \tag{1}$$

where $E_h\{\cdot\}$ denotes ensemble averaging over h, and σ^2 denotes the variance of $b(t)$.

The received microphone signal can be divided into a direct signal and a late reverberant signal. Using Polack's stochastic model it can be shown that:

1. The instantaneous short-term Power Spectral Density (PSD) of the late reverberant signal consists of a delayed and attenuated instantaneous short-term PSD of the received microphone signal.
2. The short-term cross-correlation between the direct signal and the late reverberant part is zero.

Unfortunately these relations do not hold in case of time-averaged short-term correlations and related short-term power spectral densities which will be used in practice.

In [2] we have presented an extension of Lebart's single-channel algorithm to the multi-channel case. We have shown how the estimate of the instantaneous power spectrum of late reverberation can be improved using multiple microphone signals. Additionally, the fine-structure of the speech signal is partially restored due to spatial averaging of the received amplitude spectra. It has been shown that different realizations of the same stochastic process are obtained by varying the position of the receiver with a fixed source position or by varying the position of the source with a fixed receiver position (or by varying both positions). This implies that we can assume ergodicity and evaluate the ensemble average in (1) by spatial averaging.

In [2] we assumed that all source-receiver distances were equal. This constraint on the microphone array geometry limits the applicability of the proposed method due to the restrictions on the source position. In this workshop we will explain the proposed algorithm in detail and show under which assumptions a linear array can be used. We will also show comparison results of the reverberation reduction performances of the delay & sum beamformer and the proposed multi-channel dereverberation algorithm using a linear array.

Experiment

Some results are shortly described in this section. We have used a uniform linear array of 7 omni-directional microphones, the inner-microphone spacing was 10 cm. The source-receiver distance was 3 m. All RIRs have been generated using a modified Allen and Berkley's image method, to ensure inner-microphone phase relations. The reverberation time was 437 ms. The source signal consists of a clean female speech fragment of 8.5 seconds, sampled at 8 kHz. In this experiment all reflections arriving later than 40 ms after the arrival time of the direct wave are assumed to be late reverberant and will be reduced by the proposed algorithm.

The spectrogram of the first 3 seconds of the reverberant signal, delay & sum beamformer output and the output of the proposed algorithm are depicted in Figure 1. The Local Reverberation Reduction (LRR) in each frame of 32 ms (75% overlap) was calculated using the following expression:

$$LRR(i) = 10\log_{10}\left(\sum_{n=0}^{L-1} r_{in}^2(iRL+n)\right) - 10\log_{10}\left(\sum_{n=0}^{L-1} r_{out}^2(iRL+n)\right)$$

where $r_{in}(n)$ denotes the late reverberant part of the received signal (center microphone), $r_{out}(n)$ the processed late reverberant part and R denotes the frame rate (0.25). The LRR for the proposed method (solid line) and delay & sum beamformer (dotted line) are shown in Figure 2 [1]. The dashed line in Figure 2 represents speech activity. The global reverberation reduction (GRR), defined as the average LRR during periods of silence, of the proposed solution was 11 dB while the GRR for the delay & sum beamformer was approximately 1 dB.

The distortion is measured using the Log Spectral Distance (LSD) over the periods of speech. The LSD is calculated with respect to the log spectrum of the anechoic signal. The LSD of the proposed solution is comparable to that of the delay & sum beamformer. Both signals contain less distortion, i.e. smaller LSD, than the unprocessed reverberant speech signal.

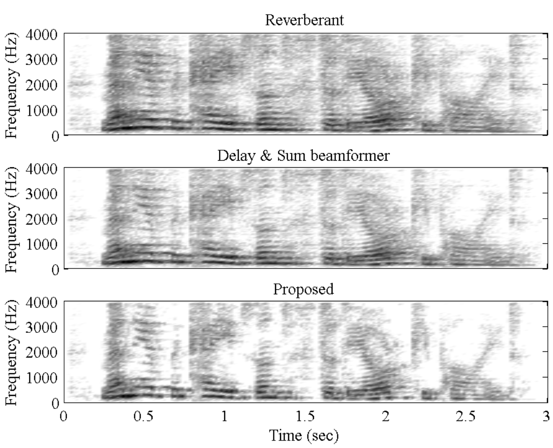

Fig.1 Spectrograms

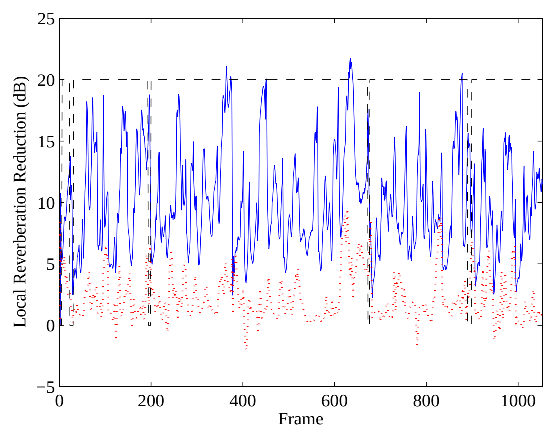

Fig.2 Local Reverberation Reduction

References

[1] K. Lebart and J.M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, pp. 359–366, 2001.

[2] E.A.P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," *ICASSP 2005, accepted for publication*, 2005.

[1] The results are available for listening on the following web page: *http://www.sps.ele.tue.nl/members/e.a.p.habets/hscma05/hscma05.html*