Real-time microphone selection in noisy reverberant environments for teleconferencing systems

Israel Cohen[1], Baruch Berdugo[2] and Joseph Marash[3]
[1]*Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, Israel*
[2]*MRD Technologies Ltd., Kibbutz Usha 30031, Israel*
[3]*Phoenix Audio Technologies, 2934 N. Naomi Street, Burbank, CA 91504, USA*
icohen@ee.technion.ac.il, bberdugo@phnxaudio.com, jmarash@phnxaudio.com

Abstract

In a teleconferencing application, it is sometimes desired to use more than one microphone for audio pickup in order to cover larger room setting. A major challenge is to monitor the perceived quality of each microphone signal and select, at any given point in time, the microphone with the best reception. We demonstrate a real-time system that comprises a few microphone clusters and a main audio unit to identify comparative features of output signals for each of the microphone clusters. The microphone selection contains two stages. The first stage is local: for each microphone cluster we compute some features of the local signals. The second stage is global: we select the least reverberant signal based on the features of the local signals. We show that local power and local power-ratio are reliable attributes which are sufficient for selecting the least reverberant signal out of the given set of measured signals.

1. Introduction and motivations

In a teleconferencing application, it is sometimes desired to use more than one microphone (or microphone cluster) for audio pickup in order to cover larger room setting. The challenge is to make the right selection by determining, at any given point in time, which microphone (or microphone cluster) has the best reception. In our solution, we assume that the level of reverberation received by the microphone is critical to the intelligibility and hence the quality of reception. Therefore our goal is to select the microphone (or microphone cluster) with the least amount of reverberation.

Direct approaches to the selection of the least reverberant signal are based on the signal power, or the signal-to-noise ratio [1,2,3]. The problem with using the microphone power as a measure of level of reverberation is that in some cases a microphone is positioned in a highly reverberant area in the room (close to a reflecting surface or next to a corridor) and its output power may be higher than another microphone positioned in a non-reverberant spot in the room. In addition the sensitivities of the microphones may vary from one microphone to the other; therefore using the measure of power level requires prior calibration of the microphones. The signal-to-noise ratio can be a reliable measure to quantify the level of reverberation only if the noise is uniform, and the power of late reverberation is uniform [1,2]. Unfortunately, in real applications, the noise cannot be assumed uniform, nor the late reverberation is uniform.

Naylor et al. [2] have addressed the measurement of reverberation in terms of the Direct-to-Reverberation Ratio (DRR) in the context of assessment of dereverberation algorithms for quantifying the level of reverberation before and after processing. They showed that a good estimate of DRR can be obtained from the measured signals using the Signal-to-Reverberant Ratio (SRR). However, the source signal has to be spectrally white and correctly normalized. Therefore, their method is not suitable for teleconferencing systems.

Goetze et al. [3] compared several objective measures using subjective listening tests for the assessment of dereverberation algorithms. They classified the objective measures into two different classes: measures that are based on the (i) impulse response or the transfer function of a system (system-based measures) and (ii) measures that are based on signals. In many applications (e.g., teleconferencing), the impulse responses of the room may not be accessible or not appropriate to apply those measures. Such situations restrict the number of applicable measures to those based on signals [4,5]. Unfortunately, most signal-based objective measures fail to judge the specific distortions that may be introduced by dereverberation algorithms like late reverberation since these artifacts are small in amplitude but perceptually relevant due to the loss of masking of the room impulse response. Only one signal-based measure, the so-called Perceptual Similarity Measure (PSM) [6], showed high correlation with subjective rating for the given test setup. Even this measure is not suitable for real-time quality monitoring of speech and audio signals, since it requires long segments of the signals, whereas in real-time quality monitoring we need to make the evaluation based on short time intervals.

In this demonstration, we present an innovative patent-pending research prototype of a real-time system for selecting, at any given point in time, a microphone signal, which has the least amount of reverberation amongst several signals acquired by microphones placed randomly in a conference room. We demonstrate state-of-the-art results by an economical, practical and innovative algorithm.

The system comprises a few microphone clusters and a main audio unit to identify comparative features of output signals for each of the microphone clusters. The microphone selection contains two stages. The first stage is local: for each microphone cluster we compute some features of the local signals. The second stage is global: we select the least reverberant signal based on the features of the local signals. We show that local power and local power-ratio are reliable attributes which are sufficient for selecting the least reverberant signal out of the given set of measured signals.

2. Problem formulation

Let $s(t)$ denote a source signal, and let $p_i = (x_i, y_i, z_i)$ denote spatial coordinates in a noisy and reverberant room ($i = 1,2,\dots,N$). Then the source signal measured at point p_i is given by

$$r_i(t) = s(t) * h_i(t) + n_i(t)$$

where $h_i(t)$ is the combined impulse response of the room and the microphone from the source to point p_i, and $n_i(t)$ is the noise measure at point p_i.

Our objective is to determine which signal out of a given set of measured signals $\{r_i(t) \mid i = 1,2,\dots,N\}$ is perceptually the least reverberant. Perception of the amount of reverberation in a given signal is closely related to the direct-to-reverberation ratio. For evaluating the direct-to-reverberation ratio, the impulse response $h_i(t)$ is split into two parts [7]:

$$h_i(t) = h_i^e(t) + h_i^l(t)$$

The early part $h_i^e(t)$ represents the direct path from the source to the microphone at point p_i, plus possibly some early reflections of the acoustic waves that propagate to point p_i. Early reflections include reflections whose path (from source to microphone) is not significantly longer than the direct path (these reflections lag behind the direct-path signal by no more than approximately 50 ms), and therefore are not perceived as reverberations. On the other hand, the late part $h_i^l(t)$ represents later reflections, which are perceived as reverberations. The direct-to-reverberation ratio is defined as the ratio between the power of the "direct signal" $s(t) * h_i^e(t)$ (including the early reflections) and the power of the "reverberations" $s(t) * h_i^l(t)$ (containing only the late reflections) [2]. Accordingly, our objective is to determine which signal out of the given set of measured signals $\{r_i(t) \mid i = 1,2,\dots,N\}$ has the greatest direct-to-reverberation ratio.

3. Implementation

Our system is based on clusters of uni-directional microphones, each looking at a different direction (for demonstration, we use four uni-directional microphones looking at direction 90 degrees apart). Alternatively one can use omni-directional microphones and create several listening audio beams by ways of beamforming techniques (see e.g. [8,9,10]). We then compare the signal received by each of the microphones (or beams) in a cluster (referred to as local) and compare it with the other local microphones (or beams). We derive a reverberation quality figure based on the assumption that direct signals are received with different levels by the local microphones, while indirect signals (reverberations) are received with a much closer level on all the local microphones (beams). We then compare the reverberation quality figure between all the clusters and select the audio source with the least amount of reverberation.

The proposed procedure contains two stages. The first stage is local: for each point we compute some features of the local signals. The second stage is global: we select the least reverberant signal based on the features of the local signals. The features include local power and local power-ratio. The local power is associated with the directional microphone (or directional beamformer output) that measures the strongest signal at a given point, compared to the signals that are measured by the other microphones at that point. The local power-ratio is defined as the ratio between the local maximum power and the local minimum power.

We demonstrate a real-time system that comprises a few microphone clusters of four microphone units, and a main audio unit to identify comparative features of output signals for each of the microphone clusters.

Based on the identified features, we monitor the perceived quality of each microphone signal and select, at any given point in time, the microphone with the best reception.

4. Conclusions and future developments

Microphones that are used in industrial applications are generally not calibrated. The sensitivities of different may be quite different. Therefore, the power is not reliable for a comparison between signals measured with different microphones. The signal-to-noise ratio is also not a reliable measure to quantify the level of reverberation, since in real applications, the noise cannot be assumed uniform, nor the late reverberation is uniform.

Our solution presents an innovative, economical and practical way of estimating the level of reverberation for selecting the microphone with the best reception amongst randomly placed microphones in a conference room. The system consists of clusters of uni-directional microphones, each looking at a different direction. Alternatively one can use omni-directional microphones and create several listening audio beams by ways of beamforming techniques. The reverberation quality figure relies on the local power, associated with the directional microphones, and the local power-ratio, associated with the ratio between the local maximum power and the local minimum power. Real-time monitoring of the reverberation quality figure enables a real-time comparison of all the microphone clusters, and finding the audio channel with the least amount of reverberation.

Future developments include integration of the proposed system with source direction indicators, to demonstrate a combination of real-time source localization and channel selection.

5. References

[1]. M. Wolf and C. Nadeu, “Channel selection measures for multi-microphone speech recognition”, *Speech Communication*, Vol. 57, Feb. 2014, pp. 170-180.

[2]. P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets, "Signal-based performance evaluation of dereverberation algorithms," *Journal of Electrical and Computer Engineering*, Vol. 2010, pp. 1-5.

[3]. S. Goetze, E. Albertin, M. Kallinger, a. Mertins, and K.-D. Kammeyer, "Quality assessment for listening-room compensation algorithms," in Proc. IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP-2010, 14-19 March 2010 pp. 2450-2453.

[4]. E. Albertin, J. Rennies and S. Goetze, “Objective quality measures for dereverberation methods based on room impulse response equalization,” in Proc. German Annual Conference on Acoustics (DAGA), Berlin, Germany, Mar. 2010.

[5]. J. Wen, N. Gaubitch, E. Habets, T. Myatt, and P. Naylor, “Evaluation of speech dereverberation algorithms using the MARDY database,” in Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, Sept. 2006.

[6]. R. Huber and B. Kollmeier, “PEMO-Q - A new method for objective audio quality assessment using a model of auditory perception,” *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, no. 6, Special Issue on Objective Quality Assessment of Speech and Audio, 2006.

[7]. E. A. P. Habets, S. Gannot and I. Cohen, “Late reverberant spectral variance estimation based on a statistical model,” *IEEE Signal Processing Letters*, Vol. 16, Issue 9, Sep. 2009, pp. 770-773.

[8]. M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin, Germany, 2001.

[9]. S. Gannot and I. Cohen, “Adaptive beamforming and postfiltering,” in Jacob Benesty, M. Mohan Sondhi and Yiteng (Arden) Huang (Eds.), *Springer Handbook of Speech Processing*, Springer, 2008, Part H, Ch. 47, pp. 945-978.

[10]. B. N. Gover , J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," *J Acoust. Soc. Am.*, Nov. 2002, pp. 1980-91.