Applied Acoustics 129 (2018) 126–134

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

A robust interaural time differences estimation and dereverberation algorithm based on the coherence function

Yi Fang [a,b,*], Haihong Feng [b], Youyuan Chen [b]

[a] University of Chinese Academy of Sciences, PR China
[b] Shanghai Acoustics Laboratory, Chinese Academy of Sciences, PR China

ARTICLE INFO

Article history:
Received 23 March 2017
Received in revised form 28 June 2017
Accepted 18 July 2017
Available online 28 July 2017

Keywords:
Precedence effect
Sound localization
Dereverberation
Coherence

ABSTRACT

A novel scheme of binaural sound localization and dereverberation in reverberation environment is present in this paper. The performance of cross-correlation based traditional time-delay estimation method is degraded sharply in a reverberation environment. Some precedence effect models have been proposed to apply in cross-correlation functions, but these models are parameter-sensitive and the front-end processes are very complex. This paper firstly proposes a simple and effective time-delay estimation method based on a coherence function in which the absolute values of coherence function is used to judge the reliability of the frequency-domain signal. And then the estimated time-delay values were applied to the coherent-to-diffuse power ratio (CDR) estimator, which can be used for reverberation suppression. Experimental results showed that the proposed scheme has higher localization accuracy than traditional methods and achieve a higher PESQ scores than other CDR estimators.

1. Introduction

Room reverberation reduces the ability of sound source localization and also degrades the quality and intelligibility of speech. Several contributions have been made in the past to reduce the effects in reverberant rooms and hence to increase the localization ability and the quality of speech.

In the human system, the interaural time difference (ITD) and interaural level difference (ILD) are considered to be the main localization cues [1]. Most of the binaural sound localization methods [2–4] are based on ITD estimation since ITD yields position estimations with smaller standard deviation compared to ILD [5]. So we concentrate on the reliable estimation of ITDs in reverberation environment in this paper. The most commonly used methods for computing ITDs are inspired by Jeffress's model [6], which represents a physiologically based theory that describes the ITD estimation based on binaural coincidences in the auditory system [7]. Many methods have been developed based on this mechanism [8–11], most of which are designed to identify the peak value of the cross-correlation function. These methods have performed satisfactorily in free-field conditions. However, audio devices are usually performed in rooms containing large reverberation. Some precedence effect based models have been suggested to increase localization accuracy in reverberation environments. The "precedence effect" refers to phenomena that reveal the human localization mechanism in reverberation environments [12]. Psychoacoustic researchers have determined that although a sound reaches the ear through several paths in such an environment, the first-arriving signal plays a dominant role in human localization [13–15]. The echo-avoidance model proposed by Jie Huang is one of the most well-known algorithms [16]. In this model, a generalized pattern of impulse response is used to estimate the sound-to-echo ratio in each frequency band. High ratios regions only are then utilized for ITDs. The performance of this model is highly dependent on parameter selection in the generalized pattern of impulse response. Unfortunately, these parameters vary from room to room. Martin proposed another model [17] as an implementation of Zurek's [18] precedence account. The algorithm first detects a sharp onset and an inhibition signal is generated after approximately 1 ms, with a gradual time scale recovery of approximately 10 ms. The inhibitory signal is then applied in the cross-correlation function [17]. In Martin's model, the origin signals are decomposed into different frequency bands and then passed through the Meddis-Hewitt inner haircell synapse (IHC) model [19]. These front-end processes require significant computational power, and onset detection is a challenge in noisy environments. Faller and Merimaa proposed a novel interaural-coherence (IC) based binaural cues selection model [20] in which high coherence cues in sub-bands indicate true localization cues, including ITD and

* Corresponding author at: University of Chinese Academy of Sciences, PR China.
E-mail address: m15249967745@163.com (Y. Fang).

http://dx.doi.org/10.1016/j.apacoust.2017.07.019
0003-682X/

ILD. The normalization of the cross-correlation function is performed to obtain an IC estimate, and only maximum IC values above a certain threshold are used for ITD estimation. Otherwise the signals are thought to be unreliable and abandoned. Recent psychoacoustic experiments have also shown a strong relationship between the IC and the listener's ability to discern the direction of sounds [21–22]. However, this model requires the calculation of cross-correlation functions in each frequency band and is sensitive to IC thresholds. More details regarding precedence effect models can be found in our Ref. [23]. Although these models provide a variety of approaches to extracting reliable signals among noisy sounds, they are difficult to apply in practical situations due to the complexity of the computation and the difficulty in selecting parameters.

Coherence-based dual-channel dereverberation approaches have been widely researched during the past decades [24–26]. The main idea of these methods is that direct sounds can be seen as coherent speech and the reverberation can be considered as diffuse noise (non-coherent). Recently, the concept named coherent-to-diffuse power ratio (CDR) have been proposed. Marco Jeub proposed a CDR estimator under the assumption of no time-delay between two-channel signals [27]. In Ref. [28], the direction of arrival (DOA) of the target speech was taken into account in their CDR estimator. In Ref. [29], Schwarz and Kellermann proposed the improved estimators, including the case of known DOA and noise coherence, unknown DOA, and unknown noise coherence. However, the above CDR estimators are presented based on a free-field model. Chengshi Zhen proposed a binaural dereverberation algorithm that the head shadowing effect is taken into account for CDR estimation [30]. But this estimator requires prior knowledge of ITDs. But the ITDs estimation in reverberation environment is also a difficult task.

In this work, a coherence-based binaural cues selection model was proposed for ITD estimation and dereverberation. Firstly, we introduce a high coherence cues based ITD estimation method that is robust in reverberation environments and has extremely low computational complexity. And then we applied the estimated ITDs to the CDR estimator. At last, the CDR values are used for reverberation suppression.

2. System overview

The proposed coherence based signal processing model is show in Fig. 1, where the high-coherence cues based ITD estimation was introduced in Section 3.1 and the CDR estimators were present in Section 3.2.

2.1. Definition of coherence function

In the time-domain, the signals received by the two microphones can be represented as follows:

$$x_i = s_i(t) + n_i(t) \quad i = 1,2 \tag{1}$$

where t denotes the sample-index, and $s_i(t)$ and $n_i(t)$ represent the target sound source and interference noise components. The time-domain signal is converted to the frequency domain using a short-time discrete Fourier transform (STFT):

$$X_i(\lambda,\mu) = S_i(\lambda,\mu) + N_i(\lambda,\mu) \quad i = 1,2 \tag{2}$$

where λ and μ are the frame index and the discrete frequency bin, respectively. The coherence function between the two channels is defined as:

$$\Gamma_{X1X2}(\lambda,\mu) = \frac{P_{X1X2}(\lambda,\mu)}{\sqrt{P_{X1X1}(\lambda,\mu)\cdot P_{X2X2}(\lambda,\mu)}} \tag{3}$$

where $P_{X1X1}(\lambda,\mu)$ and $P_{X2X2}(\lambda,\mu)$ are the auto-power spectral density (APSD) of $X_1(\lambda,\mu)$ and $X_2(\lambda,\mu)$, respectively. And $P_{X1X2}(\lambda,\mu)$ is the cross-power spectral density (CPSD) of $X_1(\lambda,\mu)$ and $X_2(\lambda,\mu)$. The CPSDs and APSDs are computed based on the following formula:

$$P_{X1X1}(\lambda,\mu) = \alpha\cdot P_{X1X1}(\lambda-1,\mu) + (1-\alpha)\cdot |X_1(\lambda,\mu)|^2 \tag{4}$$

$$P_{X2X2}(\lambda,\mu) = \alpha\cdot P_{X2X2}(\lambda-1,\mu) + (1-\alpha)\cdot |X_2(\lambda,\mu)|^2 \tag{5}$$

$$P_{X1X2}(\lambda,\mu) = \alpha\cdot P_{X1X2}(\lambda-1,\mu) + (1-\alpha)\cdot X_1(\lambda,\mu)\cdot X_2^*(\lambda,\mu) \tag{6}$$

3. Implementation of model

3.1. High coherence cues for ITD estimation

3.1.1. Ideal coherence function

We can obtain the ideal coherence function of the two input signals using the case described in Fig. 2 [31–32]. The ideal coherence function means that there is only one sound source and no other interference:

$$\Gamma_{ideal}(\lambda,\mu) = e^{j\cdot\omega\cdot fs\cdot\tau} \tag{7}$$

Thus, the Eq. (7) can be rewritten as:

$$\Gamma_{ideal}(\lambda,\mu) = \cos(\omega\cdot fs\cdot\tau) + \mathrm{j}\cdot\sin(\omega\cdot fs\cdot\tau) \tag{8}$$

where τ is the time-delay between two microphones. The real (Real) and imaginary (Imag) parts of the coherence function are expressed as follows:

$$\mathrm{Real} = \cos(\omega\cdot fs\cdot\tau) \tag{9}$$

$$\mathrm{Imag} = \sin(\omega\cdot fs\cdot\tau) \tag{10}$$

In Eqs. (9) and (10), we can find that the absolute value of coherence function is equal to one.

$$|\Gamma_{ideal}(\lambda,\mu)| = \sqrt{\cos(\omega\cdot fs\cdot\tau)^2 + \sin(\omega\cdot fs\cdot\tau)^2} = 1 \tag{11}$$

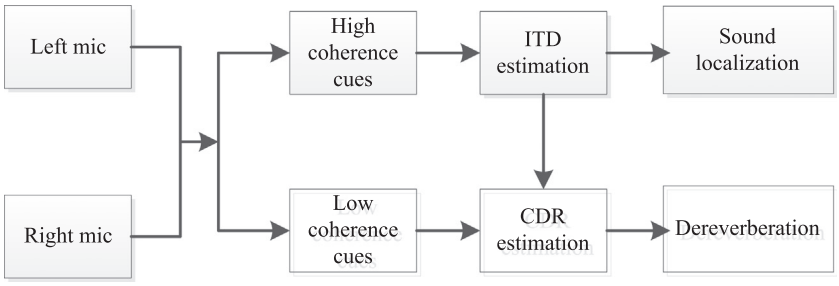

Fig. 1. The proposed coherence based binaural cues selection system.