TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 14/17 pp90-94
Volume 16, Number 1, February 2011

Small-Space Microphone Array Fractional Delay Algorithm Based on FIR Filter for Cochlear Implant*

CHEN Yousheng (陈又圣), GONG Qin (宫 琴)**

Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China

Abstract: Directional speech enhancement of signals from microphone arrays is an effective way to improve speech recognition for cochlear implant users. The strict implant size limitation results in a short distance between microphones. The fractional delay problem due to the short distance between microphones is solved by a maximal flat (Maxflat) finite impulse response (FIR) filter, using the Maxflat error criteria at a low frequency containing most of the speech information and energy. The fractional Maxflat FIR filter approximates the ideal digital fractional filter at the magnitude response, phase response, and phase delay characteristics, and is also very low order. The results demonstrate that the Maxflat FIR filter accurately and effectively solves the fractional digital delay and is very suitable for real-time speech processing in practical cochlear implant products.

Key words: microphone array; fractional delay; Maxflat FIR filter; directional speech enhancement system

Introduction

Currently, clinical cochlear implant has only one omnidirectional microphone or one simple directional microphone, which can not meet the sound quality requirement in complex environments. Cochlear implants have good speech recognition in quiet environments, but poor speech recognition in noisy environments[1-3]. A directional speech enhancement system based on the dual TP-microphones for a cochlear implant was developed to improve speech recognition. This directional system helps improve the speech SNR when the locations of the speech signal and noise signal differ. Then the cochlear implant turns to work in a "quiet" environment to improve the speech recognition. The two TP-microphones in the system were 1 cm apart to fit in the cochlear implant product design. Several beamforming methods for microphone array speech enhancement are presented in the literature[4-8]. The delay sum beamforming method was used in this research because of its low computational complexity for real-time processing in the cochlear implant. The three main parameters in this beamforming method are the delay, weight, and angle.

The digital delay is easily resolved when the microphones are far apart. However, in a cochlear implant, the microphones are only 1 cm apart. With such a short distance, a major problem is the accurate fractional delay for beamforming. The fractional digital finite impulse response (FIR) filter is used to obtain an effective fractional digital delay for the beamforming using the maximal flat (Maxflat) error criteria at a low frequency with most of the speech information and energy.

Received: 2009-12-22; revised: 2010-10-10

* Supported by the National Natural Science Foundation of China (Nos. 60871083 and 30800234), the Beijing Natural Science Foundation (No. 3082012), and the Key Technologies R&D Program of Ministry of Science and Technology of the People's Republic of China (No. 2008BAI50B08)

** To whom correspondence should be addressed.
E-mail: gongqin@mail.tsinghua.edu.cn; Tel: 86-10-62773014

1 Method

1.1 Beamforming and the fractional delay problem

The delay and beamforming method used for the microphone array speech enhancement system is based on the dual TP-microphone for cochlear implants as shown in Fig. 1.

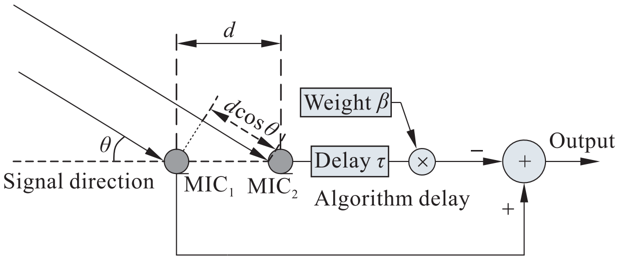

Fig. 1 Delay and beamforming method

In Fig.1, θ is the angle between the signal and the forward direction, d is the distance between microphones. Microphones MIC_1 and MIC_2 record the signals. τ and β are the delay and weight parameters for microphone MIC_2. The system magnitude response can then be described by

$$|H(\mathrm{j}\Omega)| = |1-\beta \mathrm{e}^{-\mathrm{j}\Omega(d\cos\theta/c+\tau)}| \tag{1}$$

where c is the speed of sound. The system magnitude response differs for different signal directions, θ, to generate the directional enhancement effect. Cochlear implant users focus more attention in front of the ears, with least interest to the rear side, and the lateral signal is possible between these two directions as they communicate with others. The beamforming method was then designed so that the cochlear implant provides different signal strengths from different directions. Signals coming from the front were at least 5 dB stronger than those coming from the rear and at least 3 dB stronger than those coming from the lateral side. The signal strength then uniformly decreased from the front to the lateral and then to the rear. The weight factor, β, was set to 1.1 and the delay, τ, was set to 1.1 d/c to provide this uniform decrease. The beam pattern of the beamforming delay algorithm is shown in Fig. 2.

The beam pattern in Fig. 2 shows the magnitude response for all orientation (0-360°), based on the optimized parameters of β=1.1 and τ=1.1 d/c. The signal modules from the front, lateral, and rear sides are 0.42, 0.24, and 0.1, so that the signal from the front side is stronger than from the lateral and rear by 5 dB and 12 dB, with a monotone decreasing from 0 to 180°.

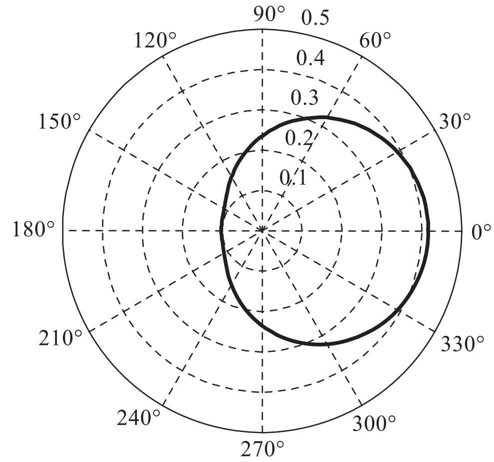

Fig. 2 Beam pattern to yield the directional effect

The digital delay of 1.1 d/c for the beam pattern of Fig. 2 was hard to achieve in the cochlear implant with the very small distance d. Here, the microphone distance d was 0.01 m and c is the velocity of sound in air, so the delay 1.1 d/c was only 32.35 μs. The speech processor sampling rate, f_s, was 44.1 kHz, so the delay for 1.1 d/c corresponds to only 1.427 samples. Since the delay of 1.427 sampling points can not be directly described by a digital delay of $x(n-1.427)$, this results in a fractional delay problem. For most applications with larger distances between the microphones, the digital fractional delay is easily handled by rounding the fractional delay to minimize the errors with little changes in the beam pattern. However, this method can not be used here due to the short distance and the small values of sampling in the delay, sometimes even less than one sampling point. Therefore, a low order FIR filter was designed to accurately realize the fractional delay.

1.2 Maxflat FIR filter incorporating speech characteristics

After the signal passed through the system a delay of D sampling points, the signal, $x(n)$, is then transferred to $y(n)=x(n-D)$. $X(\mathrm{e}^{\mathrm{j}\omega})$ and $Y(\mathrm{e}^{\mathrm{j}\omega})$ are the frequency responses for input and output signals, $H_{\mathrm{ideal}}(\mathrm{e}^{\mathrm{j}\omega})$ represents the ideal system frequency response. Then,

$$H_{\mathrm{ideal}}(\mathrm{e}^{\mathrm{j}\omega}) = \frac{Y(\mathrm{e}^{\mathrm{j}\omega})}{X(\mathrm{e}^{\mathrm{j}\omega})} = \frac{X(\mathrm{e}^{\mathrm{j}\omega})\cdot \mathrm{e}^{-\mathrm{j}\omega\times 1.265}}{X(\mathrm{e}^{\mathrm{j}\omega})} = \mathrm{e}^{-\mathrm{j}\omega D} \tag{2}$$

This represents an all-pass system with a linear phase, where D is equal to 1.427 in an actual system. Let $h(n)$ represent the actual impulse response of the