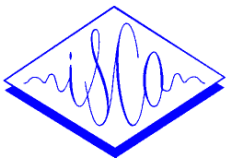


2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

LATERALIZATION OF SPEECH SOUNDS BY BINAURAL DISTRIBUTING PROCESSING

Qianjie FU, Deyu XIA, Ren Hua WANG

Department of Radio and Electronics
University of Science and Technology of China
230027,Hefei,Anhui,P.O.Box 4,P.R.China

ABSTRACT

This paper proposed a lateralization method of speech sounds that had no restrictions with regard to sound field conditions, including the stationary and moved multisignals under the real circumstance. It is specially designed to model the space perception function of the human auditory system , which is the basis for the further binaural processing method such as COCKTAIL PARTY PROCESSOR. This important aim was achieved by applying the knowledge of binaural space perception of the human auditory system. Combining the crosscorrelation processing between binaural channels with the distributing characteristics of the location information of the different sound sources, a integrating processing unit of the location information in the different auditory channels was introduced in this paper,then a composite judgement unit was introduced to simulate the location judgement function of the central nerve system. The individual steps of the processing scheme were deseribad and preliminary results were presented.

10.21437/ICSLP.1992-313

INTRODUCTION

With the development of binaural auditory nerve processing model, more accurate and effective models on the lateralization of sound source are required. Although there are a few model proposed on this aspect [1,2,3] , these models are essentially aimed at the lateralization of simple tones. As the real speech sounds are time varying and contain abundant frequency components , it is critical for the further study of binaural auditory nerve processing model to solve the lateralization of complex sounds.

Several parameters are proposed for the simulation such as binaural intensity, phase differences, and time differences, however, most physiological models of lateralization center on the concept of time. Von Hornbostel and Wertheimer discussed the importance of binaural time differences in [4], and suggested that binaural intensity differences were also converted into time differences because they believed that latency of neural firing was inversely related to stimulus intensity at least over the intensity range from threshold to 70 dB SL, and this hypothesis was supported by several experiments. So only binaural time differences are used in lateralization of speech sounds in this paper.

In 1908 Bowlker suggested a central mechanism to transform temporal differences into place differences. A more detailed model of this same mechanism was proposed by Jeffress in 1948 [5]. Jeffress envisaged a central neural complex , which tracts from both ears made overlapping synaptic connections, such that discrepancies in the arrival time of impulses from the ears focused on different loci within the neural complex and thereby triggered different postsynaptic fibers for each delay. The basic concept of Jeffress' hypothesis was following: Incoming neural signals from each ear, after filtering by critical bands, neural impules from each ear are sent to the brainstem. Each filter sends branches to cell bodies of neurons in an auditory nucleus. If the sound is delayed in one ear , then the neural impulses from the two ears will simultaneously converge on a cell body that is relatively close to the delayed ear. Thus, the location (place) of the neuron that is excited simultaneously provides information about the position of the sound source. Jeffress' hypothesis was supported by many studies of single cells in the medial and lateral superior olivary nuclei.

Based on the experiment analysis, we supposed that the convergent cell body in Jeffress' hypothesis was equivalent to the peak in the crosscorrelation function, that is to say, the crosscorrelation processing proposed in [1][2][3] was consistent with the processing proposed in Jeffress' hypothesis.So the crosscorrelation processing was introduced in this paper.

Moreover, single units have been found in the inferior colliculus of the cat that show regular changes in their rate of response as a function of binaural time differences, and other single units were found in which the maximum number of spikes always occured at a constant dichotic interval,regardless of frequency. These units seem to be capable of detecting an absolute binaural time differences[6]. So there is some information of binaural time differences in the different auditory channel more or less, if we could detect the information by simulating the function of the above single units and envisage a central nerve module to synthesize these information , enough information of binaural time differences could be obtained to locate the sound sources.

Based on these views, a novel method on lateralization of speech sound by binaural distributed processing was proposed in this paper. It mainly based on the following hypothesis: supposed that there are lots of units which were capable of detecting an absolute binaural time difference existing in the different auditory channels, and the CrossCorrelation Processing(CCP) function was used to simulate the function of these units. Some units of the individual auditory channel carry out the function proposed in Jeffress' hypothesis, so we can detect Interaural Delay Time(IDT) information of certain auditory channel by the equivalent CCP function, then a proposed central processor detect the absolute binaural time differences of all the auditory channels and synthesized the binaural time differences distributing in the different auditory channels to obtain the overall IDT information among the speech signals from the different directions, then we can locate the corresponding position of the different sound sources by certain transformation algorithm.

SYSTEMATIC STRUCTURE

The overall structure of the proposed system is shown in Fig.1, the individual processing steps are described in the following sections.

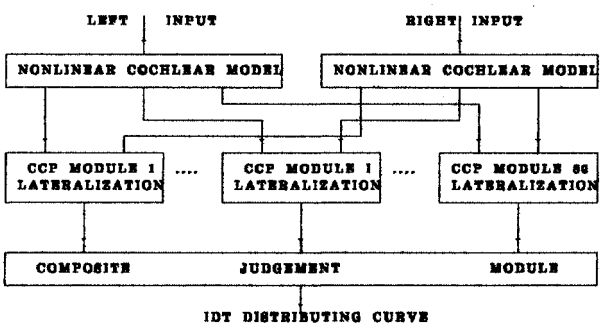


Fig.1: Struture of lateralization of Speech Sounds

① *The Nonlinear Cochlear Model*

The cochlear model in this paper simulates the cochlear function — the generation mechnism of the Instantaneous Firing Rate (IFR) of the nerve fibers in the human auditory system.The ouput of the cochlear model is the desired IFR curve.Since the performance of the cochlear model has a great influence on the performance of the proposed lateralization method, we apply the recent nonlinear cochlear model proposed by ourselves into the proposed model.

The nonlinear cochlear computational model introduced in this paper based on the double Q adaptive circuit .The start point of the model is to simulate the generation of the instantaneous firing rate of the similar nerve fiber.Using the double Q adaptive circuit made the model more effective to present the level dependent nonlinearites observed in the basilar membrane comparing with previous models ,the characteristic of the synchrony capture was improved significantly.

The overall nonlinear cochlear model is composed of 86 independent channels. Each channel includes a nonlinear bandpass filter following a nonlinear processing stage. The scope of frequency of the bandpass filter bank is from 200Hz to 4400Hz. The filter bank was designed on the basis of the physiological experimental data on basilar membrane and the tuning curve of the nerve fiber obtained in recent years. The nonlinear processing stage aims at simulating the transform of the vibration of the basilar membrane into the impulse discharge of the nerve fiber. The detailed design method of this nonlinear cochlear model was shown in [7][8].

② *Individual lateralization Module*

It was well known that almost all information of the input speech signal was included in the Instantaneous Firing Rate [9].So it is reasonable to suggest that the lateral information from two ears should be included in the IFR of the different auditory channel. As above, there are many perceptual units which is capable of detecting an absolute binaural time difference distributed in certain auditory nerve , such as in the inferior colliculus.Supposed that these units had the similar function of detecting IDT information proposed by Jeffress, meanwhile the function of CCP processing had a equivalent effect,therefore, we proposed an individual lateralization module to simulate the lateralization function of these units by CCP processing.

CrossCorrelation (CC) processing and similar CC function processing were proposed in previous paper [1][2][3], but these only aimed at solving the problem on lateralization of pure signals such as the simple frequency signal below 500Hz. As we know, since the content in the spectrum of speech signal is very abundant and variable,there are abundant harmonic components in certain auditory channel even for a single speech signal, furthermore, the spectrum of the overlapping speech signals would be more complicated, because the composite spectrum was the linear addition of the spectrum of the individual speech signal. For a single speech signal, as the strong inherent intercorrelation between the corresponding auditory channels of two ears, and the existance of the possible space perception units in all auditory channels, which was supported by physiological experiment, we assumed that it still retains the IDT in all auditory channels, and through lots of experimental processing analysis, we found it possible to detect IDT information using the CCP function .However for the overlapping speech signals, the situation is different, because of the influence of the linear addition in signal spectrum , all signal information is included in the same auditory channel. Is this CCP function still effective ? By analysizing the composite spectrum structure, we found that, a dominant signal in certain auditory channel is still present,that is to say, in this auditory channel, the spectrum of an individual signal plays a dominant function in the composite spectrum, according to this point, it is possible to detect IDT information of the auditory channel, moreover this IDT information mainly exhibited IDT information of the dominant signal. Because the dominant signal is different in the different auditory channel, IDT information obtained in the different auditory channel was also different.

The detailed methods on computational simulation is following:

$$IDT_i^j = MAX(AC_i^j(k)) \qquad k = 1,2,....N \tag{1}$$

$$AC_i^j(k) = \sum_{l=1}^{l=NUM} IFR_{Left}^j(l) \times IFR_{Right}^j(l+k) \tag{2}$$

Where IDT_i^j is the IDT information of ith auditory channel during jth frame, $AC_i^j(K)$is the corresponding autocorrelation function, andIFR_{Left}^j、IFR_{Right}^j is the instantaneous firing rate of left ear and right ear respectively.

③ *Composite Judgement Module*

In previous analysis, we know that the different auditory channel exhibits IDT information of the dominant signal. So as long as the individual signal plays a dominant function in some auditory channels more or less, IDT information of this individual signal can be detected.Each auditory channel has a different

or similar IDT information,so the central auditory system can perceive lots of IDT information.IDT information of the different signals may have a peak in IDT information distributing curve by integrating the spatial IDT information. So a composite judgement module is proposed to detect the peak of composite IDT distributing curve to obtain IDT information of the different signals, the detailed formula is following:

$$CIDT^j(IDT_i^j) = CIDT^j(IDT_i^j) + 1 \qquad i = 1,2,...,80 \qquad (3)$$

$$LI^j = PEAK(CIDT^j(k)) \qquad k = 0,1,...,N \qquad (4)$$

Where $CIDT^j$ is the composite IDT information of all auditory channels, LI^j is the location information of speech sounds during jth frame, and PEAK(F) stands for searching the peak value of F curve.

EXPERIMENTAL RESULT ANALYSIS

The further binaural nerve processing models need more accurate and effective lateralization of multi-signal to segregate the individual signal from the different direction, so lateralization of multi-signal is the most important case.Preliminary experiment on the lateralization of the crosstalker was carried out to prove the feasibilty of the proposed model.

The first experiment is that the overlapping speech signal was composed of two speech signals from the different directions. The position of the sound sources is stationary, that is to say, IDT information of these two signals is constant. It is difficult for us to obtain the composite multisignal with binaural information, so in this experiment, a synthetic multisignal with binaural information was introduced. The detailed processing is following:

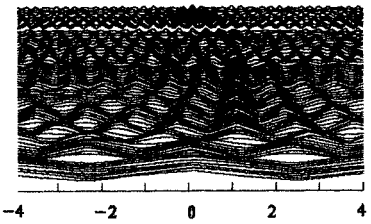
/ YI/ and / ER/ is two original pure digital signals spoken by two different male speakers with the 10kHz sampling rate, synthetic signal was composed as following:

Supposed that $S_{YI}(t)$and S_{ER} stand for the temporal sequences of signal / YI/ and / ER/ .

$$\begin{cases} S_{Right}(t) = S_{YI}(t) + S_{ER}(t) \\ S_{Left}(t) = S_{YI}(t+T) + S_{ER}(t) \end{cases} \qquad (5)$$

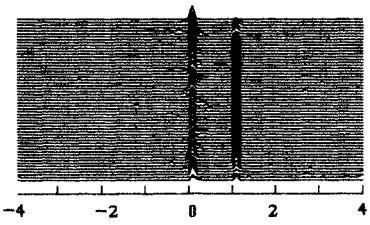
Where T is equal to 1ms in this experiment,that is to say, / YI/ has a 1.0ms Interaural Delay Time, while / ER/ has no interaural Different Delay time. Fig.2 shows the crosscorrelation function curve of auditory channels during 10th frame, where the frame length is 6.4ms. The horizon direction represents the length of cross correlation function, and the vertical direction stands for the CC function curve of every auditory channel. The IDT information distributing curve by this method is shown in Fig.3, the horizon direction represents the IDT infroamtion , while the vertical direction shows the trendency of location information along with the temporal sequency.

The second experiment is almost the same as the first one. But position of the sound sources is changing, that is to say, IDT information of these two signals is changed during the different temporal range. Like the previous experiment, the composite signal was composed of two original pure digital signal / YI/ and / ER/ ,and / YI/ has a 1.0ms+Ct (C is constant, t is the temporal information,that is T = Ct+1.0, C = 0.02) Interaural Delay Time, while / ER/ has no interaural Different Delay time. The CC function curve of auditory channels during 40th frame is shown in Fig.4.The IDT information distributing curve is shown in Fig.5.



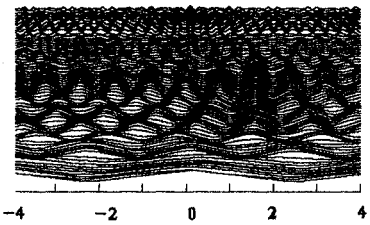
Cross Correlation Function (ms)

Fig.2 Cross Correlation function curve of all auditory channels during 10th frame of composite signal



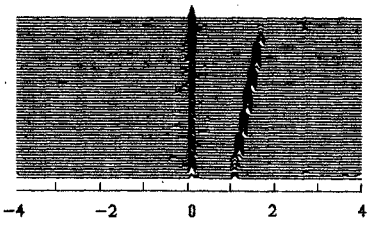
Interaural Delay Time (ms)

Fig.3 Position-stationary IDT distributing curve



Cross Correlation Function (ms)

Fig.4 Cross Correlation function curve of all auditory channels during 40th frame of composite signal



Interaural Delay Time (ms)

Fig.5 Position-moved IDT distributing curve

In Fig.2, we found that IDT information was constant during the different temporal range, the middle peak track represented IDT information of signal / ER/ , by this track, we could learn that there was no interaural delay time in this signal, so we

said that this signal was from the front or the back. In this experiment, the right ear information was regarded as the standard, the other IDT information indicated the left information is delayed, that is to say, this signal was from the direction adjacent to the right ear. In Fig.3, the case is similar, the different point was that the position of signal /YI/ was changed during the different temporal range, so we could see that the IDT information is changed, and this change exhibited that the left ear signal delay time became more and more , this signal was approaching to the right direction.

SUMMARY

In this paper, we explained the space perception exsiting in the human auditory system and proposed a lateralization method of speech sounds by binaural distributing processing. The proposed mehtod functionally detected the little discrepancy between the interaural instantaneous firing rate and the output of the model corresponded to the IDT information between binaural auditory channels. Preliminary experiments on stationary and moved multisignal testified the feasibility of the model.

The present work lays a foundation for further binaural nerve processing methods such as COCKTAIL PARTY PROCESSOR. However, the lateralization method in this paper is irrelevant to the other audiotry perception function such as PITCH perception and auditory attention processing, how to integrate effectively these function to form a composite auditory processing system and stronger lateralization of speech sounds under more complicated circumstance were the further research work.

REFERENCE

[1] W. Lindemann, " Extension of a biaural cross correlation model by contra lateral inhibition " , JASA , 1986 , pp1608–1622.

[2] Shihab A. Shamma ," Stereausis: Binaural processing without neural delays ", JASA, 86(3), 1989, pp989–1006.

[3] Raymond H. Dye, Jr, "The combination of interaural information across frequence: Lateralization on the basis of interaural delays", JASA, 88(5), 1990, pp2159–2169.

[4] Hornbostel, E. M. von. Das raumliche Horen, in Handbuck der normalen und pathologischen, Vol. II, A. Bethe (ed.), Berlin: Springer–Verlag, 1926.

[5] Jeffress, L. A. " A place theory of sound localization", J. Comp. Physiol. Psychol, 1948.

[6] W. Lawrence Gulick, George A. Gescheider, Robert D. Frisina, HEARING, Sound Localization, 1989, pp317–349.

[7] R.H. WANG, Q.J. FU, D.Y. XIA, " A nonlinear cochlear model based on double Q adaptive circuit", 14th ICA,1992,Beijing.

[8] R.H. WANG, Q. J. FU, D.Y. XIA, " Design of cochlear filter ", accepted by ACTA BIOPHYSICA, 1992.

[9] Shihab A.Shamma, "Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve", JASA, 78(5), pp1612–1621.