STUDY ON PHASE PERCEPTION IN SPEECH

Tong Ming  Bian Zhengzhong  Li Xiaohui*  Dai Qijun  Chen Yanpu**
(Xi'an Jiaotong University, Xi'an 710049)
* (Xidian University, Xi'an 710049)
**(Xi'an Communication College, Xi'an 710106)

Abstract  The perceptual effect of the phase information in speech has been studied by auditory subjective tests. On the condition that the phase spectrum in speech is changed while amplitude spectrum is unchanged, the tests show that: (1) If the envelop of the reconstructed speech signal is unchanged, there is indistinctive auditory perception between the original speech and the reconstructed speech; (2) The auditory perception effect of the reconstructed speech mainly lies on the amplitude of the derivative of the additive phase; (3) \( t_d \) is the maximum relative time shift between different frequency components of the reconstructed speech signal. The speech quality is excellent while \( t_d < 10 \text{ms} \); good while \( 10 \text{ms} < t_d < 20 \text{ms} \); common while \( 20 \text{ms} < t_d < 35 \text{ms} \), and poor while \( t_d > 35 \text{ms} \).

Key words  Speech signal; Auditory perception; Phase spectrum; Additive phase

I. Introduction

The auditory perception of the phase in speech signal has been chronically neglected\(^1\)\(^-\)\(^5\), but with the increasing requirement to speech quality, it is of most importance to quantificationally research the perceptual influence of the phase information in speech on auditory perception. So, the perceptual effect of the phase information in speech has been studied through auditory subjective tests by directly changing the phase spectrum in continuous speech of whole sentence.

The original speech signals in time domain are decomposed into a series of frames with Hanning window, and half of sample points are overlapped between the adjacent frames. The STFT (Short Time Fourier Transform) spectrum is deduced by carrying FFT (Fast Fourier Transform) on each frame with windows:

\[
A_k = |A_k| \exp(j\phi_k), \quad k = 0, 1, \cdots, N - 1
\]

So, the STFT amplitude spectrum of speech signals are \( \{|A_k|, k = 0, 1, \cdots, N - 1\} \) and phase spectrum are \( \{\phi_k, k = 0, 1, \cdots, N - 1\} \).

In the experiment, the sample frequency of speech signal \( f_s \) equals 11.025kHz \((f_s = 11.025\text{kHz})\) and frame length \( N \) equals 256 \((N=256)\). The reconstructed speech signal varied in phase can be obtained by keeping the speech single STFT amplitude spectrum from varying, transforming the STFT phase spectrum by some way, combining those two spectrums, then carrying on IFFT and adding the half overlapped frames. The phase in-

\(^1\)Manuscript received date: April 4, 2003; revised date: April 24, 2003.
Supported by the National Natural Science Foundation of China (No.60071029)
formation affection on subjective auditory can be discovered by comparing the subjective auditory perception test of the original speech with that of the reconstructed speech.

II. The Unchanged Conditions of the Amplitude Spectrum of Full Time Domain Signal

When varying the phase of each frame signal, the amplitude spectrums of the reconstructed signal in full time domain are the same as that of the original signals only if the phase difference between each frame is unchanged. That is

\[ \phi_m(\omega) - \phi_{m-1}(\omega) = \phi_m(\omega) - \phi_{m-1}(\omega) \]  

where, \( \phi_m(\omega) \) represents the phase of the m-th frame signal at the frequency \( \omega \); \( \phi_{m-1}(\omega) \) represents the phase of the \( m-1 \)-th frame signal when changed in phase at the frequency \( \omega \). Because the values of \( m \) and \( \omega \) are random, the condition making the equation valid must be

\[ \phi_m(\omega) = \phi_m(\omega) + \phi_d(\omega) \]  

where, \( \phi_d(\omega) \) is only related to \( \omega \), called additive phase.

III. Additive Phase Effects on Waveform in Time Domain

First, in the case of a simple additive phase

\[ \phi_d(\omega) = \alpha + \beta \omega \]  

where, \( \alpha \) and \( \beta \) are any constants. The second term will not cause the waveform distortion of a speech signal and only cause the time drift \( t_d = b \) in time domain. But the first term usually causes the waveform distortion of the reconstructed signal. As \( \phi_d(\omega) = \alpha \), the Fourier transformation of the analytical signal of the reconstructed signal is

\[ Z'(j\omega) = \begin{cases} 2S(j\omega)\exp(j\alpha), & \omega > 0 \\ 0, & \omega < 0 \end{cases} \]  

So, the time domain form of the analytical signal is

\[ z'(t) = e(t)\exp[j\gamma(t) + j\alpha] \]  

Fig.1 is the effect of the additive phase on the waveform and the envelope of reconstructed speech signal in time domain. It shows that the envelope of the reconstructed signal in time domain is still \( |e(t)| \), and is not related to \( \alpha \).

To any phase spectrum, for example, the additive phase (Continuous Wave, CW) changing with the frequency \( \omega \),

\[ \phi_d(\omega) = \pi \sin(0.008\omega) \]  

The envelope and waveform of the reconstructed speech are different from the original speech.

IV. Perception Experiment of Additive Phase by Ears

1. Constant additive phase

In the experiment, we use constant additive phase \( \phi_d(\omega) = \alpha \) to correct the phase, then restructure the speech signal. \( \alpha \) gets 12 test points at the range of \(-\pi \sim \pi\), i.e., \( \pi, -5\pi/6, \ldots, 2\pi/3, 5\pi/6 \). The experiment indicates that no matter what the speech is, e.g. male,