Group Delay Function For Improved Gender Identification

Kye-Hwan Lee, Sang-Ick Kang, Ji-Hyun Song, Joon-Hyuk Chang

School of Electronic Engineering
Inha University
Incheon, Korea
{khlee,sikang,jhsong}@dsp.inha.ac.kr, changjh@inha.ac.kr

Abstract

One of the key issues in practical speech recognition is to achieve robust gender identification. Most conventional gender identification approaches use relevant features derived from the magnitude spectrum. In this paper, we propose a novel gender identification method using a group delay function (GDF). Based on the statistical analysis of the GDF, it is found that the GDF is an effective feature for gender identification. The experimental results demonstrate that the proposed method gives significant improvement compared to conventional methods.

Index Terms: Gender Identification, Group Delay, GMM

1. Introduction

Voice-based gender identification is an important task in human computer interaction and automatic speech recognition. Researches on gender identification have been advanced substantially due to recent progress in machine learning such as Gaussian mixture models (GMMs) and hidden Markov models (HMMs) [1], [2]. In current automatic gender identification systems [3], the cepstral features derived from the power spectrum are the most commonly used features [4], [5]. In most speech processing applications, the phase spectra have been totally ignored, which is perhaps due to the common understanding that the human ear is almost insensitive to phase [6]. In addition, Schroeder [7] and Oppenheim [8] concluded on the basis of experiments that the phase spectrum locks the information more than the magnitude spectrum in a short-time Fourier transform. On the other hand, Liu [9], Puliwal [10], and Alsteris [11] showed that the phase contour changes with different emotions in the speech signal. Also, Yeegnanaryana [13] informed that the group delay function (GDF) derived from the phase can be widely used in signal processing such as formant extraction, pitch extraction and spectrum estimation.

In this paper, we present an effective gender identification approach based on the GDF. Based on an investigation of the GDF characteristics, a novel approach that improves the performance of gender identification based on the GMM is proposed. Experimental results demonstrate that the GDF is an effective feature for determining gender difference in the speech signal. To evaluate the performance of the proposed method, experiments on gender identification are conducted under various conditions.

2. Review of Group Delay Function

Generally, the GDF is defined as a derivative of the phase function [14]. Given the discrete-time input signal \( x[n] \), the Fourier spectrum \( X(\omega) \) is given as

\[
X(\omega) = \sum_{n} x[n] e^{-j\omega n} \tag{1}
\]

\[
X_R(\omega) + jX_I(\omega)
\]

where \( X_R(\omega) \) and \( X_I(\omega) \) denote the real and imaginary parts of the Fourier transform. Also the Fourier spectrum \( X(\omega) \) can be represented by its magnitude \( |X(\omega)| \) and phase \( \theta(\omega) \) as follows:

\[
X(\omega) = |X(\omega)| e^{j\theta(\omega)}. \tag{2}
\]

If we use the logarithmic function to separate the magnitude and phase, then

\[
\log X(\omega) = \log |X(\omega)| + j\theta(\omega) \tag{3}
\]

Since the GDF is defined as \( \tau(\omega) = -\frac{d|\log X(\omega)|}{d\omega} \), the GDF can be computed as follows:

\[
\tau(\omega) = -\Im\left( \frac{d|\log X(\omega)|}{d\omega} \right) \tag{4}
\]

\[
= -\Im\left( \frac{-j \sum_{n} nx[n] e^{-j\omega n}}{X(\omega)} \right)
\]

\[
= \Re\left( \frac{Y(\omega)}{X(\omega)} \right)
\]

where \( Y(\omega) \) is given by

\[
Y(\omega) = \sum_{n} nx[n] e^{-j\omega n} \tag{5}
\]

\[
= Y_R(\omega) + jY_I(\omega).
\]

Finally, the GDF is derived as follows [15]:

\[
\tau(\omega) = \Re\left( \frac{[Y_R(\omega) + jY_I(\omega)][X_R(\omega) - jX_I(\omega)]}{X_R(\omega) + X_I(\omega)} \right) \tag{6}
\]

\[
= \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{X_R(\omega) + X_I(\omega)}.
\]

3. Proposed Gender Identification Algorithm

For a successful classification, a feature vector that clearly account for the discrimination between male and female
voice should be selected. For this, firstly, the input speech signal is sampled at 8 kHz and segmented into 15 ms frame. After the discrete fourier transform (DFT) analysis, the resulting spectrum was divided into 16 critical bands having the respective limits which is analogous to that of the IS-127 system [16]. Based on the experimental conditions, we conduct an extensive analysis such as scatter plot as depicted in Fig. 1 from which we can discover the superior capability of the GDF in discriminating gender. This is mainly due to the relative physical differences of the vocal tract where male or female produce speech. Different statistical characteristics of the features could be successfully described by the GMM. Based on this, we apply the GMM based on the expectation maximization (EM) algorithm to model arbitrary multi-modal densities [17]. If \( \mathbf{x} \) denotes the feature vector \( \mathbf{x} = \{x_1, x_2, \ldots, x_D\} \), the likelihood of the GMM with \( M \) mixture components is given by

\[
p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} p_i b_i(\mathbf{x}),
\]

where \( p_i \) is the weight for the \( i \)th mixture component, \( \mathbf{\mu}_i \) represents the mean vector and \( \Sigma_i \) is a covariance matrix. These parameters are collectively represented by the gender model \( \lambda_G \) such that

\[
\lambda_G = \{p_i, \mathbf{\mu}_i, \Sigma_i\}, \quad i = 1, \ldots, M.
\]

We then use the iterative EM algorithm to establish the gender model \( \lambda_G \), which is based on the maximum likelihood (ML) estimation. We then determine gender by comparing the male-likelihood and the female-likelihood according to the established gender model.

\[
\hat{G} = \arg \max_{1 \leq G \leq S} \sum_{i=1}^{N} \log p(\mathbf{x}|\lambda_G)
\]

\[
S = 2, \quad (1 : \text{male}, 2 : \text{female}).
\]

### 4. Experiments and Results

The performance of the GDF-based approach is evaluated and compared to that of the conventional features-based method [18]. Specifically, the GDF is combined with conventional features such as MFCC (13th), LPC (10th), RC (10th) and formant (10th) as a fusion scheme. The assessment is based on objective improvement in the equal error rate (EER) and the receiver operating characteristic (ROC). Gender models with 16 mixture component densities were trained using 200 data files from 10 male and 10 female speakers in the OGI database [19]. 2000 data files for the test were compiled from different 34 male and 34 female speakers.

<table>
<thead>
<tr>
<th>Feature Vector</th>
<th>EER</th>
<th>Feature Vector</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC</td>
<td>15.2 %</td>
<td>LPC</td>
<td>24.5 %</td>
</tr>
<tr>
<td>MFCC + GDF</td>
<td>10.1 %</td>
<td>LPC + GDF</td>
<td>22.5 %</td>
</tr>
<tr>
<td>RC</td>
<td>18.9 %</td>
<td>Formant</td>
<td>40.6 %</td>
</tr>
<tr>
<td>RC + GDF</td>
<td>17.6 %</td>
<td>Formant + GDF</td>
<td>24.5 %</td>
</tr>
</tbody>
</table>

Table 1: Equal error rate for feature vectors
Fig. 2 shows the ROC curves of the different features incorporating the GDF. From the results, we found that the GDF significantly improves the gender identification performance. Also, Table I shows the EER results, where a significant improvement of the error rate was achieved over all given features. Thus, it is evident that the GDF is an effective feature to characterize voice based human gender.

5. Conclusions

We have proposed a gender identification method using the GDF for automatic speech recognition. Based on a statistical analysis of the GDF, it has been demonstrated that the GDF provides an efficient way for gender identification from a human’s voice.

6. Acknowledgements

This work was supported in part by IITA through IT Leading R&D Support Project.

7. References


