Title	連続発話音声中に含まれる男声・女声知覚に寄与する音響特徴量に関する研究
Author(s)	柴田，武志
Citation	
Issue Date	2008-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/4356
Rights	
Description	Supervisor:赤木正人，情報科学研究科，修士

A study on acoustic features contributing to gender perception in continuously uttered speech signals

Takeshi Shibata (610043)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 7, 2008

Keywords: gender perception, STRAIGHT, Temporal Decomposition, multidimentional scaling, dynamic and static features.

1 Introduction

In the study on gender perception, fundamental frequency has been regarded as one of dominant cues for gender perception. However, recent studies on speech production mechanism have revealed more subtle differences between male and female speech utterances. It is well-known that speech signals are characterized by both static and dynamic features. Although the effect of the static features on gender perception has been deeply discussed, only few work have done in examining the dynamic features from the gender perception viewpoint. Therefore, it is also important to investigate the effect of the dynamic feature on gender perception.

The aim of this study is to clarify relationships between acoustic features in contributing to gender perception of continuously uttered speech signals.

2 voice conversion model

First of all, we propose a new voice conversion method to synthesize speech stimuli, in which both static and dynamic features are exploited. In our system, both static and dynamic features that show differences between male and female speech are modified.

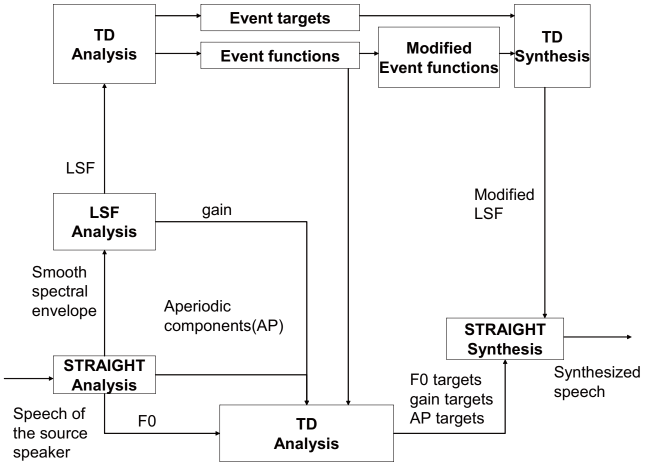

Figure 1: Fremework of voice conversion using TD.

The block diagram of our proposed voice conversion system is shown in Figure 1. First, STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum) decomposes input original speech signals into spectral envelopes, fundamental frequency (F0) information, and aperiodic components (AP). The spectral envelopes can be further analyzed into LSF parameters. Modified Restricted Temporal Decomposition (MRTD) is employed in the next step to decompose LSF parameters into event targets (static deatures) and event functions (dynamic features). In the next step, the modified event function are synthesized as dynamic spectral envelopes by TD synthesis. On the other side, the modified event target are synthesized as static spectral envelopes by TD synthesis. The fundamental frequency contour is modified by F0 target. Finally, STRAIGHT synthesis is employed to synthesize modified speech. Thus, static and dynamic features are modified in our proposed voice conversion system. In the following sections, we explain each component in detail.

3 Proposed method

These components obtained from STRAIGHT are further processed by MRTD which is able to decompose speech components into so-called event targets (i.e. static features) and event functions (i.e. dynamic features). Fitting a polynomial function to each event function is done in the non-linear least square sense. Finally, a new event function is generated by using the values approximated using the fitting. The synthesized stimuli can be further used to investigate the effect of static and dynamic features of speech on the gender perception.

4 analysis

Parameters (Averaged F0, averaged spectral envelopes, gain and F0 contours, spectral movement, and , duration) are analyzed by MDS so that each parameter is confirmed whether it relates to male and female.

5 Psychoacoustic experiment

Experiments are carried out using synthesized stimuli. The stimuli are averaged voice in which each parameter value is averaged in male and female voices and replaced voices in which each parameter value in the averaged voice is replaced with the corresponding parameter value of male or female voice.

6 Result

As the results of both the psychoacoustic experiment and MDS, it was clarified that the order of influence to gender perception is averaged fundamental frequency, averaged spectral envelope, F0 counter and duration, spectral movement, and gain in descending order.