Improved Inverse Filtering based Glottal Source Separation using Weighed Linear Prediction

Liang Chen

Abstract
The object of this research is to improve the estimation accuracy of glottal source, and weighed linear prediction (WLP) is applied into iterative adaptive inverse filtering (IAIF), whose output is the estimated vocal tract filter and voice source signal. In the WLP-modulated version of IAIF, the contribution of glottal closure instants is downgraded when outputting vocal tract parameters, while stressed when computing glottal excitation spectrum. Evaluation of this method is discribed in the latter part of this paper.

Keywords
inverse filtering, source-filter separation, weighed linear prediction

Contents

Introduction 1
1 Methods 1
1.1 Weighed Linear Prediction(WLP) 1
1.2 Iterative Adaptive Inverse Filtering(IAIF) 1
1.3 Combined Method 1
2 Evaluations 2
2.1 Rough Evaluation 2
2.2 Further Evaluation 2
3 Conclusion and Future Work 2

Introduction

Spectral envelope extracted from time - frequency domain mostly comes from the voice production and recognition system, using vocal chords as an excitation and the mouth and nose as a resonator. Voiced signals produce a harmonic spectrum and such feature influences our recognization of sounds remarkably. This model is called source-filter model. We can make use of a source-filter model for an audio signal and modify this model in order to achieve different digital audio effects.

However, this problem of extracting a spectral envelope from a spectrum is hard to achieve, and leaves a bad estimation of these harmonics, which are also called formants, and of glottal excitation.

One way to estimate the spectral envelope of a sound is based on a simple sound-production model, in which sound is produced by passing an excitation source (source signal) through a synthesis filter. The filter is an all - pole filter for it only models the resonances. This widely used approach is called linear predictive coding (LPC).

Besides, there are lots of models for glottal source estimation, like two - mass model, glottal area model and LF - model and I will use LF - model to construct my research.

1. Methods

1.1 Weighed Linear Prediction(WLP)

In the weighed linear prediction method, a window should be used when calculating residuals to change the effect of glottal flow at glottal closure instants (GCI).

$$E = \sum_{n=n_1}^{n_2} {e_n}^2 W_n = \sum_{n=n_1}^{n_2} \left(s_n - \sum_{i=1}^{p} a_k s_{n-k}\right)^2 W_n \tag{1}$$

$$\frac{\partial E}{\partial A} = 0 \tag{2}$$

The weight and the LF - model waveform is showed below.

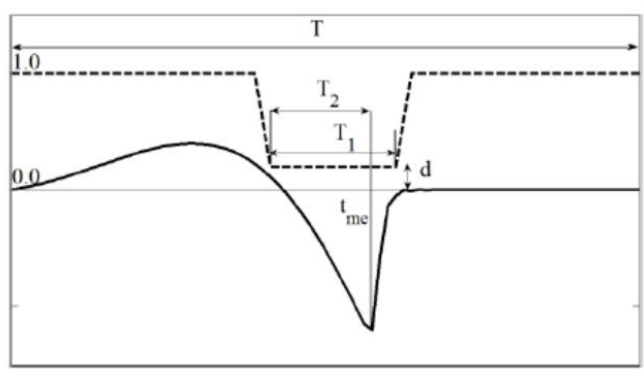

Figure 1. weight (dashed); LF - model waveform (solid)

1.2 Iterative Adaptive Inverse Filtering(IAIF)

Iterative Adaptive Inverse Filtering (IAIF) is a glottal wave analysis method. In this algorithm, the glottal contribution is first estimated with an iterative structure. The vocal tract transfer function is then modeled after eliminating the average glottal contribution. The glottal excitation is obtained by cancelling the effects of the vocal tract and lip radiation by inverse filtering.

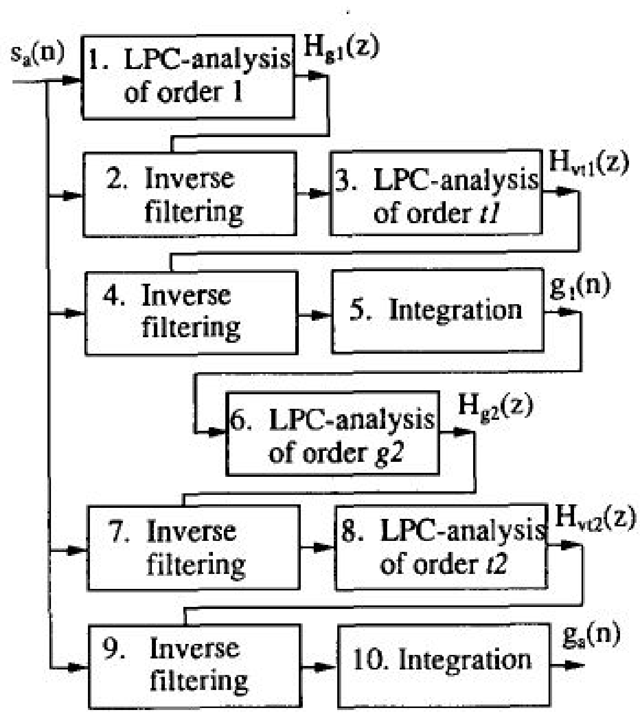

Figure 2. IAIF algorithm

1.3 Combined Method

WLP is applied into the approach iterative adaptive inverse filtering (IAIF), and the contribution of glottal closure instants is downgraded when outputting vocal tract parameters, in Block 3 and 8, Figure 2, while stressed when computing glottal excitation spectrum, in Block 1 and 6.

2. Evaluations

2.1 Rough Evaluation

In a preliminary as well as rough evaluation, I proceeded a piece in the song, Songbie, through conventional IAIF and WLP - modulated IAIF to obtain glottal flows. The spectrum of two glottal flows showed in Figure 3.

2.2 Further Evaluation

Using a LPC synthesizer, vowels with different fundamental frequencies and various formant parameters are synthesized. I proceeded these vowels and computed formant distortion between original formants and obtained formants to evaluate the efficiency of this approach. The result is showed in Figure 4, where the fundamental frequency is setted to 400Hz.

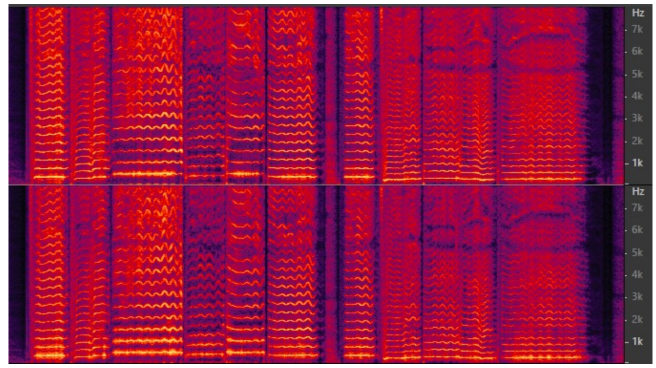

Figure 3. upper:conventional;lower:WLP - modulated

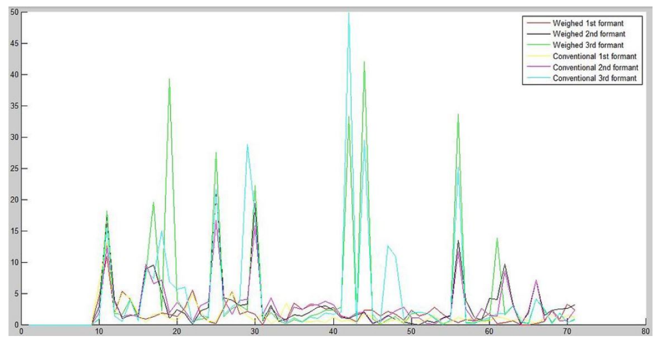

Figure 4. Formant Distortion

3. Conclusion and Future Work

The rough result(Figure 3) indicates that the lower frequency has been improved, while the higher part has been slightly mitigated. Although the results don't indicate a remarkable improvment of glottal flow, the method is effective to some extent.

The formant distortion figure shows no big difference between two methods, however, the numeric result indicates there exists slight improvement using weighed linear prediction. However, the synthesized sound used in this experiment is not perfectly suitable, which may result in inaccurate conclusion, so I am struggling with other methods to generate suitable sounds.

More audio examples can be found in
http://liangchen1ce.github.io/eng/projects/glottalflow/.

References

[1] DAFX: digital audio effects[M]. New York: Wiley, 2002.
[2] Alku P, Pohjalainen J, Vainio M, et al. Improved formant frequency estimation from high-pitched vowels by downgrading the contribution of the glottal source with weighted linear prediction[C]//INTERSPEECH. 2012.
[3] Alku P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering[J]. Speech communication, 1992, 11(2): 109-118.