Improved Methoed in Obtaining Glottal Flow from High-Pitched Voices

Liang Chen

Beihang University

liangchen@ee.buaa.edu.cn

September 11, 2013

Achknowledgement

- My supervisors
- Supporting students
- Help from Internet

Overview

Introduction to Glottal Source Models

In speech processing, the speech waveform can be used as a speech model. However, a more suitable and convenient speech model than the recorded speech waveform is often employed in speech applications, such as the extraction of linguistic information from the speech signal.

A speech analysis method is used to convert the speech sig- nal into a different representation, which decomposes the speech signal into the source and lter components, which are considered to be independent.

For example, the acoustic model of speech production typically represents the source as the derivative of the signal produced at the glottis and the lter as the vocal tract system.

Glottal Source Models

Physical Models

- Two-mass Model(Ishizaka and Flanagan, 1972)
- Three-mass Model(Story, 2003)
- Adapted Two-mass Model (Pelorson, 1994)

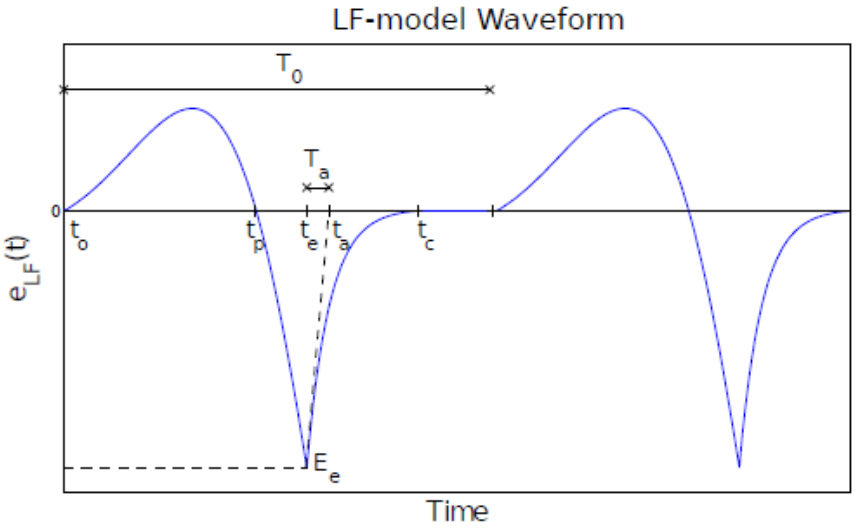
Glottal Area Models

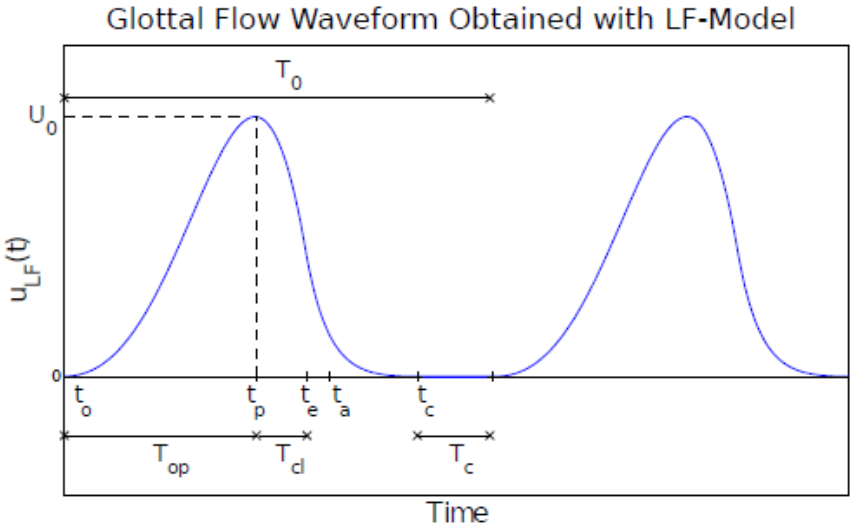
- The glottal area model of Titze(Titze, 1984)

Acoustic Models

- Liljencrants model(Fant et al., 1985)
- LF-model(Fant et al., 1985)

LF-model


LF-model Waveform
T0
Ta
0
to
tp
te
ta
tc
eLF(t)
Ee
Time


Glottal Flow Waveform Obtained with LF-Model
T0
U0
0
to
tp
te
ta
tc
uLF(t)
Top
Tcl
Tc
Time

Extraction of Glottal Flow

- Electroglottography
- Inverse Filtering

Inverse Filtering

Most glottal source estimation techniques are based on an inverse ltering process. These methods rst estimate a parametric model of the vocal tract, and then obtain the glottal ow by removing the vocal tract contribution via inverse ltering. The methods in this category differ by the way the vocal tract is estimated. This estimation could be computed in

- glottal closed phase, or by
- an iterative and/or adaptive procedure(IAIF)

Limitations of IAIF

IAIF is a widely used technique in estimating glottal flow from the original voices, and linear prediction (LP) is a foundation of this method to estimate the vocal tract resonances, the formants. However, the performance of conventional LP is known to deteriorate for high-pitched speech . In particular, the estimates of the lowest formants are biased, due to the error criterion used in conventional LP.

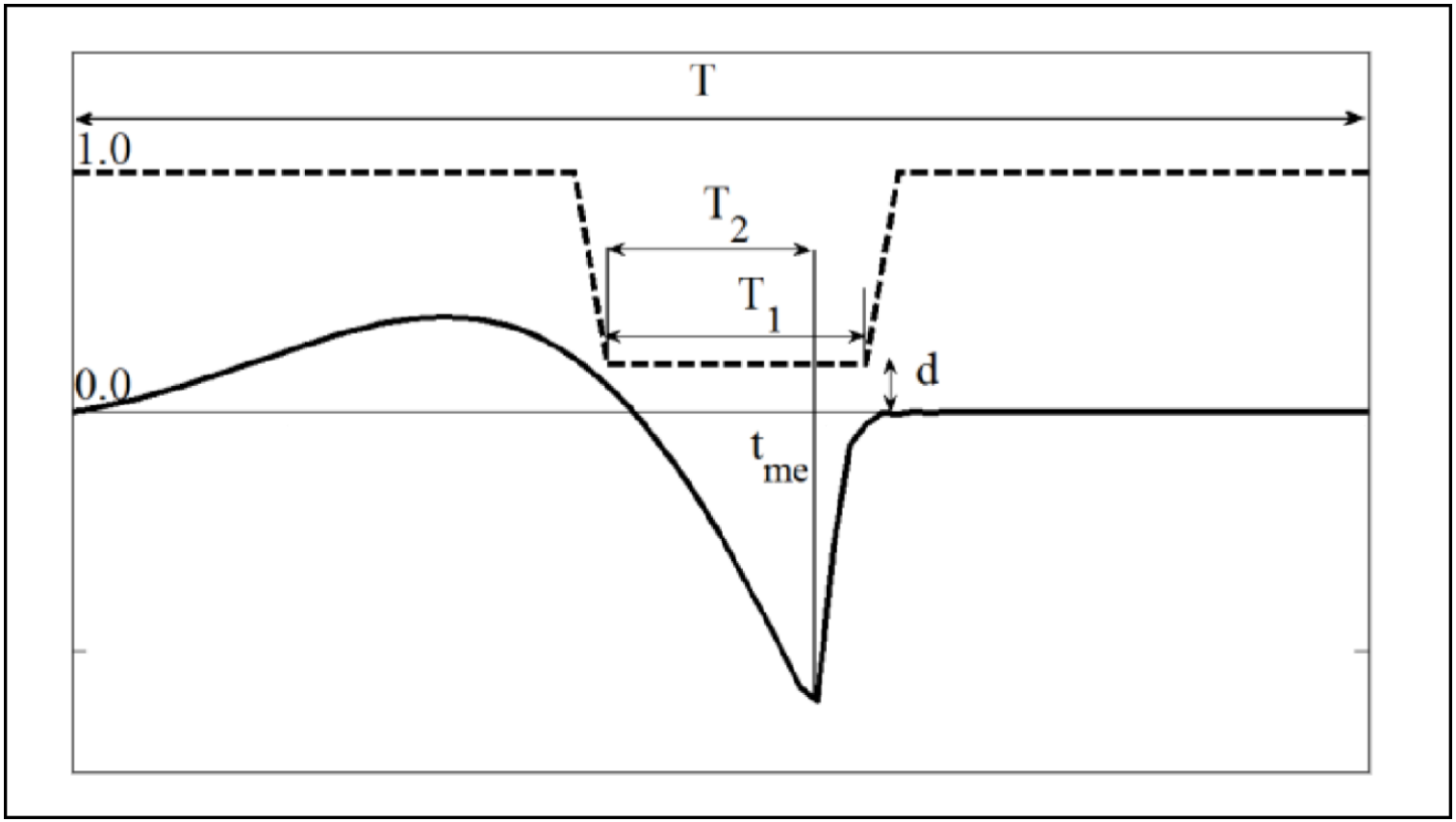
Improved Method

Theorem (error criterion)

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

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

Improved Method



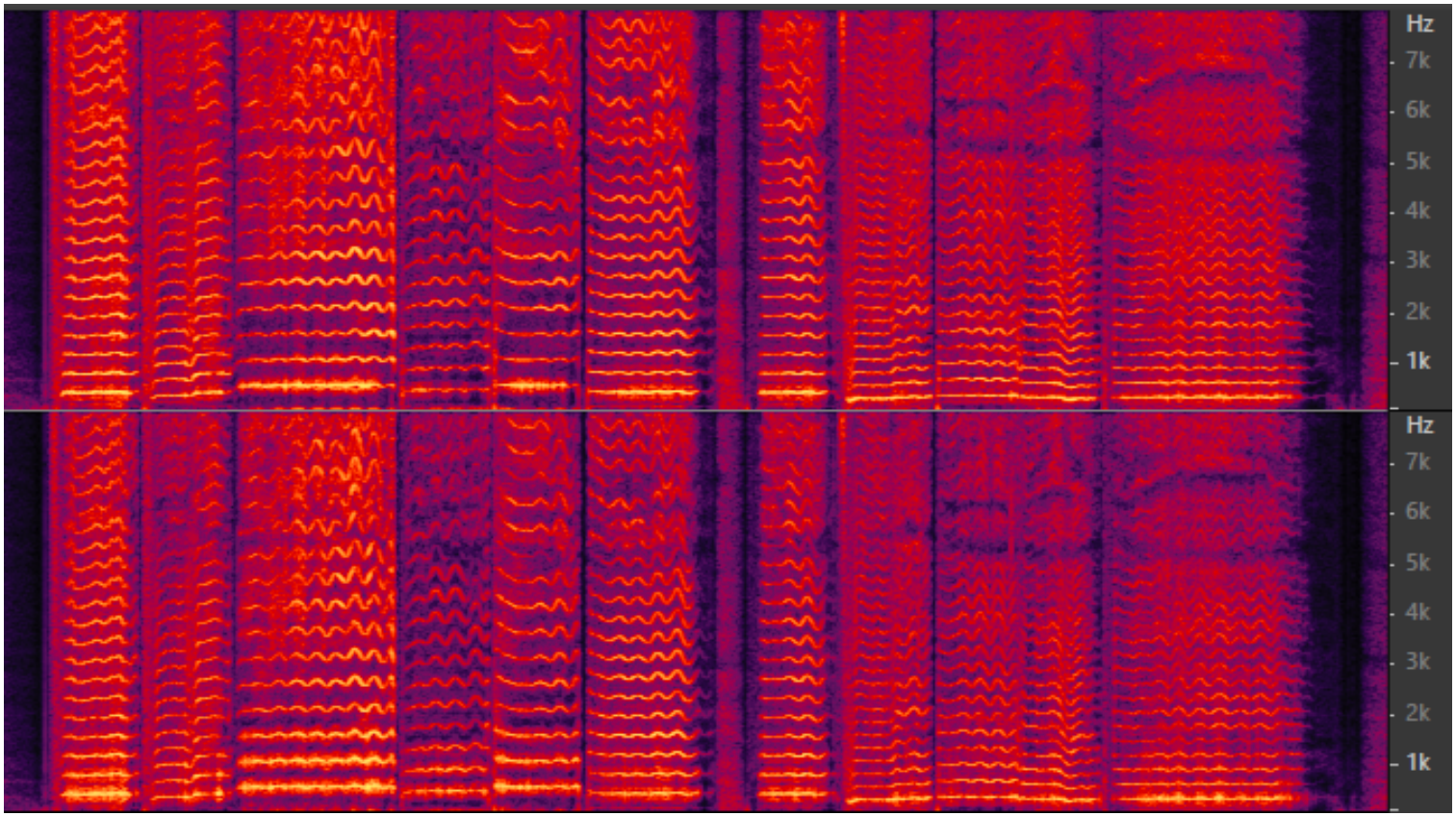
Example

Example (Code)

```
for i0 = 1:m+p,
    for i1 = 1:p,
        if i0-i1>0,
            s(i1,i0) = xe(i0-i1);
        else
            s(i1,i0) = 0;
        end
    end
end

for i3 = 1:m+p,
    Equ1 = Equ1 + w(i3) * s(:,i3) * s(:,i3)';
    Equ2 = Equ2 + w(i3) * xe(i3) * s(:,i3);
end
```

Experiment



Conclusion

The results of the experiment show 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 truly effective.

References

Makhoul, J. (1975)

Linear prediction: A tutorial review

Proc. IEEE 12(3), 63:561-580.

El-Jaroudi, A. and Makhoul, J. (1991)

Discrete all-pole modeling

IEEE Trans. Signal Process. 12(3), 39:411423, 1991.

Fant, G. (1982)

Linear The voice source - acoustic modeling

Thank You