

MIXED-PHASE SPEECH MODELING AND FORMANT ESTIMATION, USING DIFFERENTIAL PHASE SPECTRUMS

Baris Bozkurt, Thierry Dutoit

TCTS Lab. Faculté Polytechnique De Mons,
Initialis Sci. Park, B-7000 Mons, Belgium
{bozkurt, dutoit}@tcts.fpms.ac.be

Abstract

This paper introduces a new speech model, termed as the mixed-phase model, based on the assumption that the speech signal is produced by convolution of a maximum phase glottal excitation signal with a minimum phase vocal tract filter impulse response. The glottal excitation signal is assumed to be an anti-causal stable signal and the vocal tract filter is assumed to be causal and stable. For estimating resonances of the maximum phase signal (source) and the minimum phase filter (vocal tract filter), use of differential phase spectrums of z-transforms is proposed.

1. Introduction

Source-tract separation has been an interesting problem in all areas of speech processing for long years. This study addresses the issue of estimating the resonance frequencies of two systems: vocal source and vocal tract.

The resonances of the source signal and the tract system are convolved and are hardly separable. Linear predictive (LP) analysis[1] is very widely used for estimation of resonances for a speech signal with the assumption that speech signals can be modeled by an all-pole model. Once the resonances are estimated, the problem reduces to relating source and tract resonances respectively, a difficult and important problem in speech processing technology. Additionally, there are many difficulties and inefficiencies of LP estimation due to various problems, like non-linear source-tract interaction, or dependency on the degree of linear prediction, or on the position of the analysis frame relatively to the glottis closure instant.

Various methods have been proposed for source-tract separation using linear prediction. One of the well-known algorithms is the PSIAIF [2], which tries to perform the separation by an iterative linear prediction analysis. There also exist methods based on the linear prediction analysis together with glottal flow models [3,4]. All of these techniques, however, suffer from the deficiencies of the LP approach because LP estimation is hard-coded in these techniques.

In this document, we introduce the use of differential phase plots of the z-transform of a speech signal for detecting resonances of source and tract separately, without any inverse filtering. The analysis is based on an assumption that the speech signal is mixed phase, therefore the second section of this paper is dedicated to the mixed phase model. In the third section, the analysis based on differential phase plots will be presented with examples on synthetic mixed phase signals.

2. Mixed-phase model of speech

In all areas of speech processing, the basic source-filter speech model is very frequently used. It mainly assumes that the speech signal is produced by exciting a filter (corresponding to vocal tract) by an excitation produced by the lung pressure and larynx (source signal).

Being output of a physical system, the speech signal is assumed to be stable. Together with the causality feature, this assumption draws important guidelines for speech analysis. Once it is also assumed that the speech signal is causal, we end up with the minimum-phase speech model: all the poles

of a signal that is causal and stable must lie inside the unit circle on the z-plane. LP estimation automatically finds such poles.

Here, we propose a mixed-phase model of speech, where we assume that speech is obtained by convolving an anti-causal and stable source signal with a causal and stable vocal tract filter. In this model, some resonances of the signal correspond to poles outside the unit-circle on the z-plane but these poles are anti-causal, and therefore still stable. These anti-causal poles correspond to resonances of the glottal source signal, while the causal-stable poles (inside the unit circle on z-plane) correspond to the vocal tract resonances. The speech signal is a mixed-phase signal obtained by exciting a minimum-phase system (vocal tract system) by a maximum-phase signal (glottal source signal).

This assumption is based on the characteristics of glottal flow models (LF[5], KLGLOTT88[6]). In Fig.1, an example glottal flow signal is presented. As seen in the figure, the glottal flow signal looks like a time-reversed causal filter response.

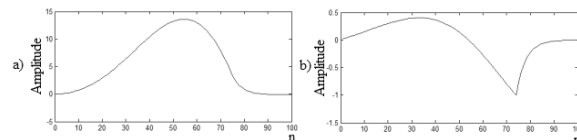


Figure 1: LF Glottal Pulse Signal in discrete time axis, a) the glottal flow signal, b) differential glottal flow signal

Anti-causality of the glottal flow signal has been discussed within the context of spectrum of glottal waveform models in [7] and the authors point the similarity of the phase spectrum of KLGLOTT88 signal to an anti-causal filter phase spectrum. The authors compare the impulse response of an anti-causal all-pole system with KLGLOTT88 synthesized glottal flow signal and the main difference is reported to be the oscillations in due to truncation. In [8], a glottal open quotient estimation method is proposed, which uses an all-pole system where all of the poles are anti-causal for the source signal.

For stability of an anti-causal all-pole system, all of the poles have to be out of the unit circle and therefore the system has to be maximum phase. In contrast, the mixed-phase model proposed here assumes that speech signals have two types of resonances; anti-causal resonances of the glottal source signal and causal resonances of the vocal tract filter.

2.1. Detection of causal and anti-causal resonances of a mixed phase signal with group delay spectrums

Let $X(\omega)$ be Fourier Transform of a signal $x(t)$, and $D(\omega)$ the associated group delay function :

$$D(\omega) = -\frac{d(\arg(X(\omega)))}{d\omega} \quad (1)$$

The causality feature of a resonance is best observed on group delay spectrums since a reversal of a signal in time domain corresponds to no change in power spectrum of the signal but the group delay spectrum is inverted horizontally (see Fig. 2).

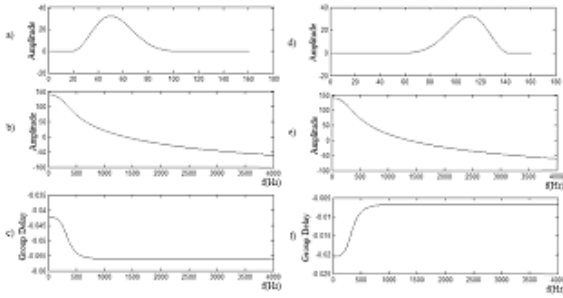


Figure 2: Causal and anti-causal single pole filter response plots; a)causal impulse response, b)log-amplitude spectrum of a, c)group delay spectrum of a, d)anti-causal impulse response, e)log-amplitude spectrum of d, group delay spectrum of d

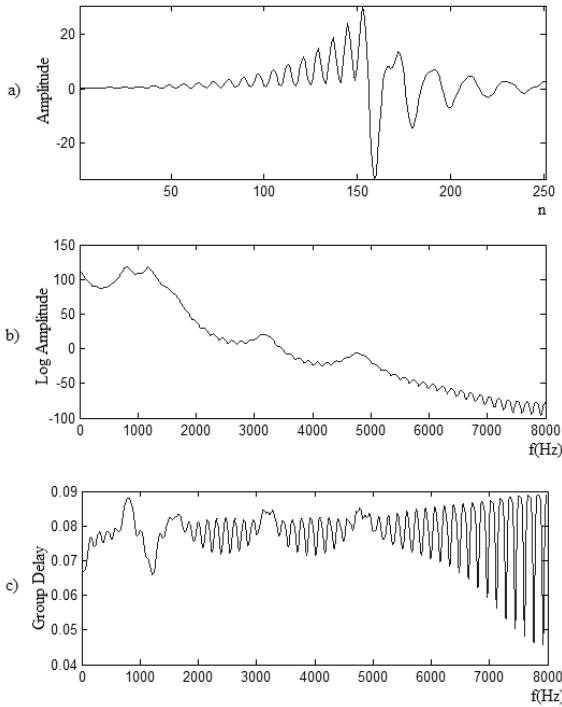


Figure 3: A mixed phase signal with causal resonances at 800, 1600, 3200 and 4800 Hz and anti-causal resonances at 0 and 1200 Hz. a)time domain signal, b)log-amplitude spectrum, c)group delay function

In Fig.3, we present a mixed phase signal (synthesized by filtering a maximum phase signal through an all-pole model) and its group delay spectrum. The causal and anti-causal resonances appear as peaks with opposite direction on the group delay spectrum, whereas causality or anti-causality cannot be observed on the amplitude spectrum.

3. Differential Phase Plots For Formant Tracking

In Fig. 3, we have presented how causality can be observed on group delay spectrums. Observation of these opposite direction peaks on group delay spectrums for real speech signals are not easy due to existence of zeros very closely located on the z-plane. In Figure 4, we present all-pole and all-zero representations of a single period speech data on the z-

plane, and its group delay function. The all-zero representation is obtained by finding all the roots of the polynomial $X(z)$ corresponding to the z-transform of the signal(2) and the all-pole representation is obtained by LP analysis:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (2)$$

Clearly, a simple observation on group delay do not provide the information we aim to get, the plots being usually too noisy due to the zeros close to unit circle.

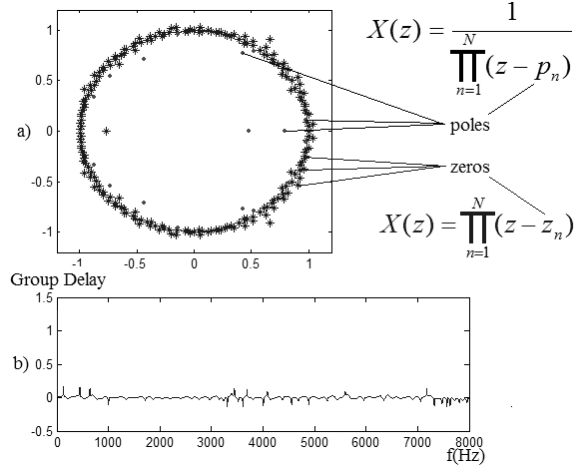


Figure 4: Effect of zeros on Group Delay. a)all-pole and all-zero plots for a single period speech data, b)group delay function

To get rid of the noisy structure, we propose here to use differential phase spectrums calculated on circles with radius different then 1 (the unit circle).

It is well known that the Discrete Fourier Transform corresponds to computing the Z-transform of a signal on the unit circle, i.e., for $z = \exp(i*\omega)$, where i is the complex value ($i = \sqrt{-1}$) and ω is the angular frequency variable. Therefore, the group delay function is differential phase function of z-transform calculated on the unit circle on z-plane. Here we propose the use of differential phase plots calculated at circles centered at the origin but with radius different than 1 (i.e., $z = R*\exp(i*\omega)$). As we shall see, with such a computation on z values not too close to the zeros around the unit circle, smoother plots can be obtained. In Fig. 5 an example is presented.

In Fig. 5, we see that some peaks are easily observed on the differential phase spectrum. In further experiences with various speech data (which we do not present here due to space limitations), high correlation was observed with LP estimated formant locations and peaks observed on differential phase plots. This suggests that frequencies and causality of resonances can be detected by tracking peaks on differential phase plots calculated at various circles on the z-plane. Frequencies of resonances can be tracked by processing positive and negative peaks on the differential phase spectrums. For detecting causality, sign change of peaks shall be tracked; if a pole is inside the circle for which the differential phase is calculated, the observed peak has positive direction and if the pole is outside this circle, the peak is at negative direction. For a single pole pair (single resonance), Fig. 6 and Fig. 7 exemplify this property.

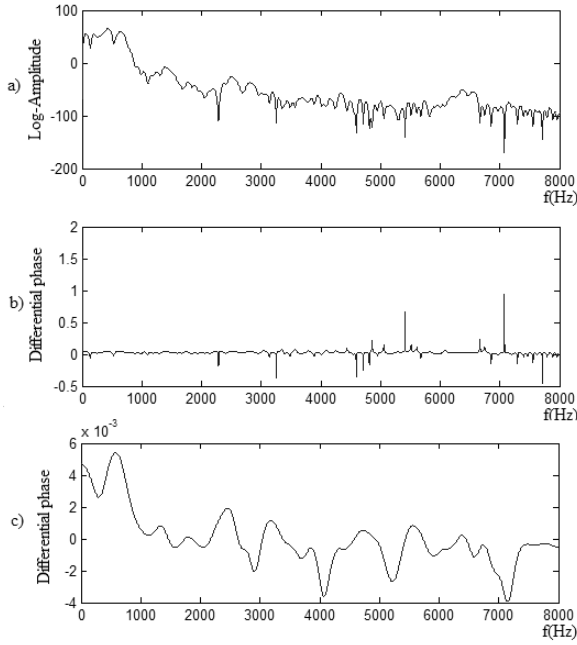


Figure 5: Comparison of Group Delay with differential phase spectrum. a) log-amplitude spectrum at $R=1$, b) group delay spectrum (differential phase spectrum at $R=1$), c) differential phase spectrum calculated at $R=0.9$ (reverse is plotted for simplicity of comparison)

For estimation of source and tract formants of a speech signal, we thus propose the following approach:

- Detection of frequencies of formants by processing several differential phase spectrums calculated on various circles on z-plane;
- Grouping these into two sets according to 'being inside or outside the unit circle'. The poles grouped to be outside the unit circle (anti-causal and stable poles) shall correspond to source signal and the poles grouped to be inside (causal and stable poles) shall correspond to vocal tract system.

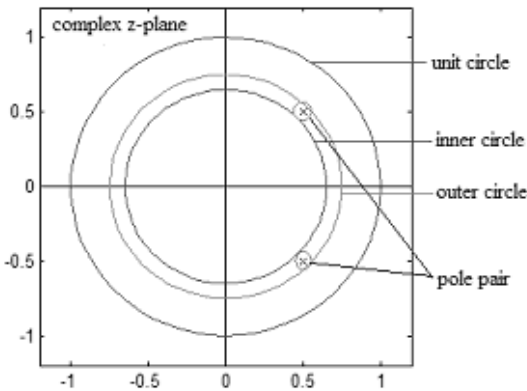


Figure 6: Circles on the z-plane for pole location estimation using differential phase spectrums

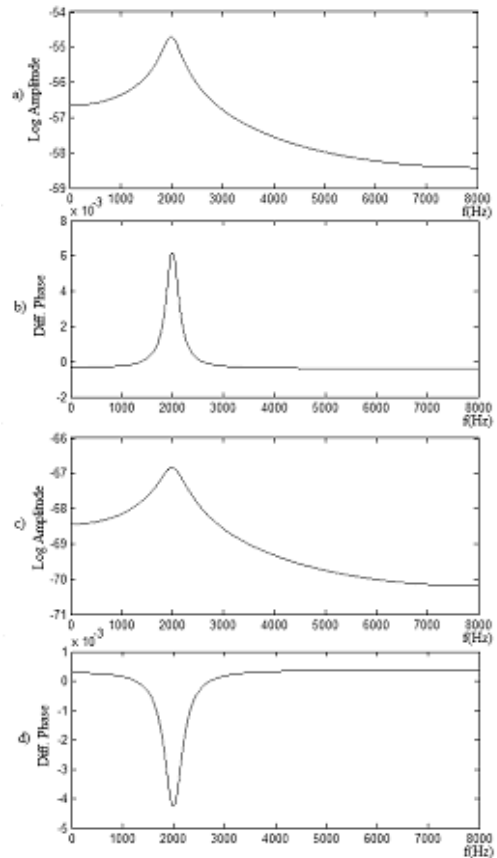


Figure 7: Z-transform plots for a single resonance (all-pole system with a single pole-pair at $f=2000\text{Hz}$, $R=0.7$) calculated on circles different than unit circle, a) log-amplitude spectrum at $R=0.65$, b) differential phase plot at $R=0.65$, c) log-amplitude spectrum at $R=0.75$, d) differential phase plot at $R=0.75$

In Fig. 8, we present the output of the method for a mixed-phase synthetic signal. The speech signal is synthesized using all-pole models for both the source signal and the vocal tract filter. The source signal poles were located at 0 and 4000 Hz and the vocal tract poles are placed at 1600, 3200, 4800 and 6400 Hz. The source signal was first synthesized as a causal all-pole filter response and then time reversed to make it anti-causal.

The resonances detected are marked by circles on the plots. The observed peaks appear at 0, 1600, 3200, 4000, 4800 and 6400 Hz. From plot b), we conclude that all poles are inside the circle $R=1.2$. From plot c), we conclude that two resonances are outside the circle $R=1.05$ (at 0 and 4000 Hz) which means they shall correspond to the maximum phase part of the signal. Plot d) shows that, there are 2 resonances outside and 4 resonances inside the circle $R=0.99$. In plot e), there is only one peak marked as positive, but a very close negative peak also exists. For these cases, no decision is taken about sign of the peak detected but only the frequency of resonance is marked. We conclude that all resonances are outside of the circle $R=0.85$ from the differential phase plot c). Final results are: two resonances (at 0 and 4000 Hz) are found outside the unit circle and classified as source signal resonances, four resonances (at 1600, 3200, 4800 and 6400 Hz) are found inside the unit circle and classified as tract resonances.

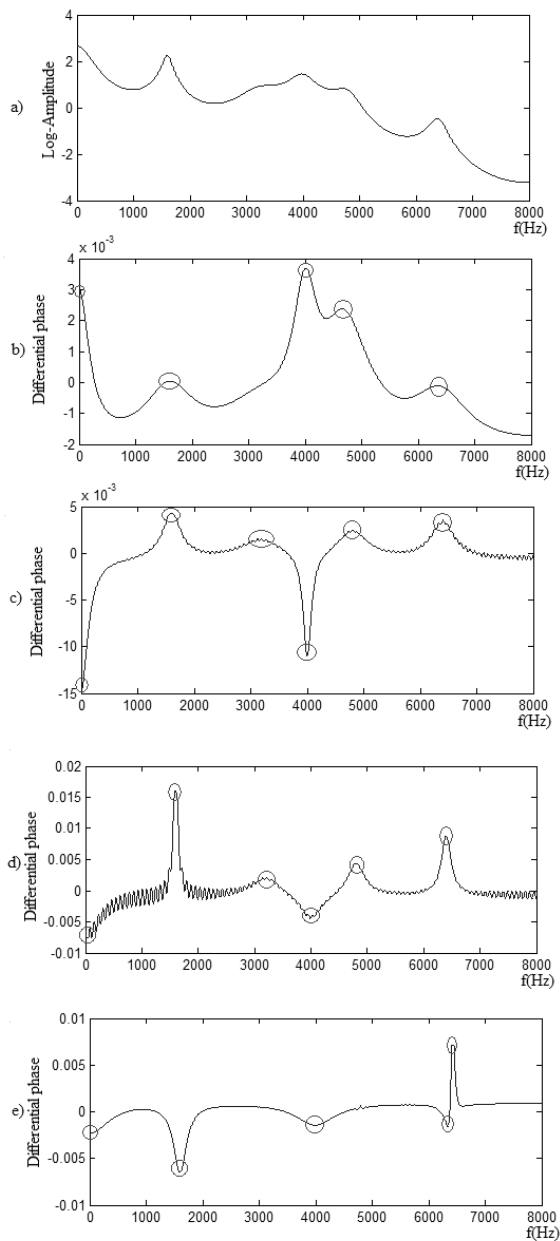


Figure 8: Detecting resonances by processing differential phase spectrums calculated at various R on z -plane, a) amplitude spectrum of the mixed-phase signal, b) differential phase plot at $R=1.2$, c) differential phase plot at $R=1.05$, d) differential phase plot at $R=0.99$, e) differential phase plot at $R=0.85$

4. Conclusions

Due to the noisy character of phase spectrums and group delay functions, only a few speech applications/analysis tools utilize the phase component of the frequency domain representation of a signal. In this paper, we have presented the potential of differential phase spectrums for formant estimation and source-tract separation applications. We have used one controlled test (analysis of a synthetic signal) to demonstrate this potential. Further effort will be dedicated to designing

automatic analysis methods for processing recorded speech signals.

We have also presented the mixed-phase model for speech, emphasizing the fact that the source-tract separation problem needs to be handled with tools, which can detect causality of resonances. In the frequency domain, this information is available in the phase component, therefore analysis tools for processing phase spectrums are of great importance.

5. Acknowledgements

This study is funded by Region Wallonne, Belgium, grant FIRST EUROPE #215095.

6. References

- [1] Makhoul, J., "Linear prediction: A tutorial review", *Proc. IEEE* 63, pp 561-580, 1975.
- [2] Alku, P., "Glottal Wave Analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering," *Speech Communication*, vol. 11, no. 2-3, pp. 109-117, 1992.
- [3] Milenkovic, P., "Glottal inverse filtering by joint estimation of an AR system with a linear input model", *IEEE Trans. on Acoustics, Speech and Signal Proc.*, 1986.
- [4] Riegelsberger, E.L. and Krishnamurthy, A.K. "Glottal source estimation: Methods of applying the LF model to inverse filtering", *Proc. ICASSP 93*, pp 542-545, Minneapolis, 1993.
- [5] Fant, G., "The LF-model revisited. Transformation and frequency domain analysis", *Speech Trans. Lab. Q. Rep.*, *Royal Inst. of Tech. Stockholm*, vol.2-3, pp 121-156, 1995.
- [6] Klatt, D. and Klatt, L., "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *JASA*, Vol.87, pp 820-857, 1990.
- [7] Doval, B., and d'Alessandro, C., "Spectral correlates of glottal waveform models: an analytic study", *Proc. ICASSP 97*, Munich, 446-452.
- [8] Henrich, N., Doval, B. and d'Alessandro, C., "Glottal open quotient estimation using linear prediction", *Proc. International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Sept. 1999.