An Algorithm to Estimate Anticausal Glottal Flow Component from Speech Signals

Baris Bozkurt, François Severin, Thierry Dutoit

TCTS Lab. Faculté Polytechnique de Mons, Initialis Sci. Park, B-7000 Mons, Belgium
{baris.bozkurt, francois.severin, thierry.dutoit}@tcts.fpms.ac.be

Abstract. In this paper, we define an algorithm with low complexity which performs a new use of the linear prediction analysis (covariance method) to retrieve the maximum-phase component of speech signals. First, we study the mixed-phase model of speech through a new representation named the Zeros of Z-Transform (ZZT) in the z-plane, which is an all-zero representation of the z-transform of a discrete time signal. Then, based on the properties of the mixed-phase model, we introduce an algorithm to estimate the anticausal glottal flow component from speech signals. LP-covariance analysis is used to estimate a pole pair outside the unit circle corresponding to the anticausal poles of the source signal component in the mixed-phase speech model. Given the pair of anticausal poles, a procedure to resynthesize the anticausal part of the glottal flow, and then an open quotient estimation method, are proposed. Evaluations show that the method is high quality for analyzing synthetic speech but lacks robustness in analysis of natural speech.

1   Introduction

This study targets estimation of the anticausal component in speech due to the first phase of the glottal flow (i.e. the glottal flow signal without the return phase). Our study is based on the mixed-phase speech model, which assumes that the speech signal is produced by convolution of a maximum phase glottal excitation signal with a minimum phase vocal tract filter impulse response [1].

When an all-pole model is studied for such a mixed phase signal, some of the poles fall outside the unit circle. In speech processing, poles outside the unit circle are most of the time (if not all) avoided/reflected due to the minimum-phase assumption. In this study, we follow the inverse path : we try to find outside poles for estimation of glottal flow characteristics.

The minimum-phase assumption relies on two properties : stability and causality. All the poles of a signal that is causal and stable must lie inside the unit circle on the z-plane. However, in the mixed-phase speech model, our assumption is : the speech signal is obtained by convolving an anticausal and stable glottal flow signal with a causal and stable vocal tract filter. The resonances due to the glottal flow signal correspond to poles outside the unit-circle on the z-plane but these poles are anticausal, and therefore still stable. The mathematical background for the glottal flow poles that are outside the unit circle can be found in [2].
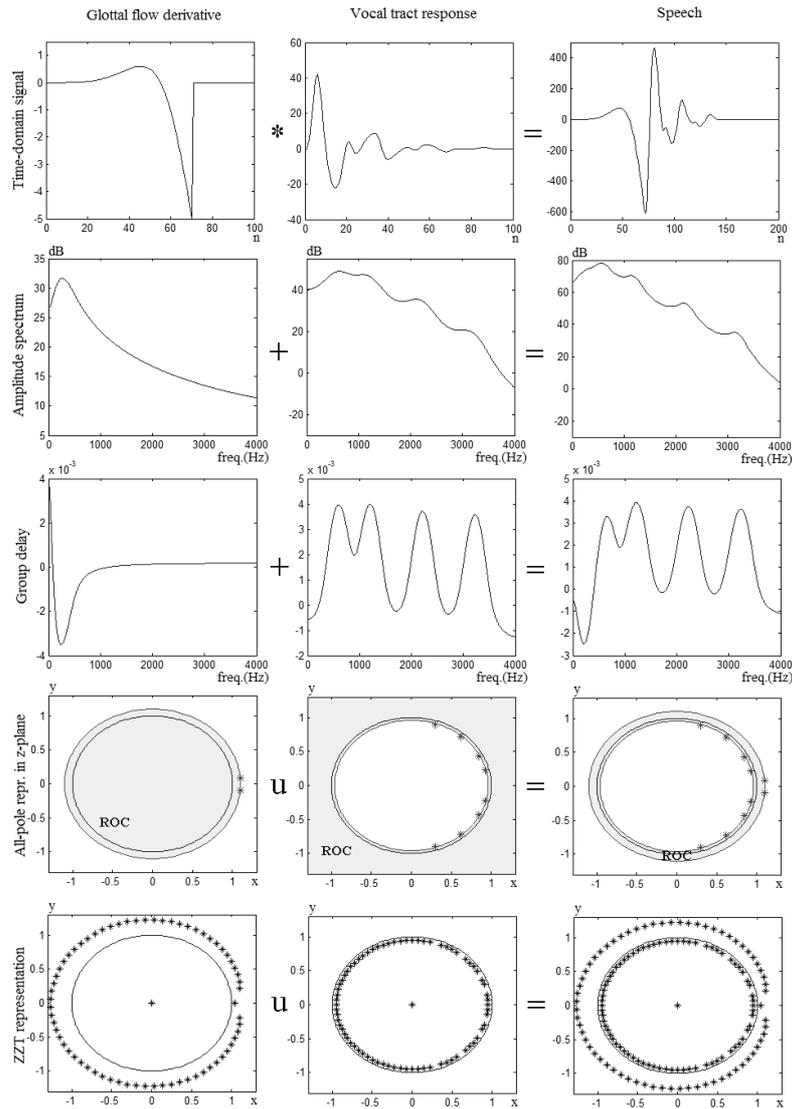
Fig. 1. The mixed-phase speech model

In Fig. 1, we present the mixed-phase speech model in the time domain, in the frequency domain through amplitude and group delay spectra, in the z-plane through all-pole representation and zeros of z-transform (ZZT) representation. ZZT representation is a new representation that serves as a domain to study z-transform characteristics of an actual discrete time signal. The ZZT representation, is defined as the set of roots/zeros (which can be found by some numerical method) of the Z-transform polynomial for a discrete time signal [3]. The last row of Fig. 1 includes the ZZT representations (roots plotted on the z-plane) of the glottal flow derivative, of the truncated vocal tract impulse response and of the speech signal obtained by convolution of these two signals (a convolution operation in the time domain corresponds to the union of ZZT sets). It is necessary to use such a representation in a practical framework where actual speech signals are to be analyzed, since the existence of the mixed-phase characteristics on speech data (therefore the existence of poles outside the unit circle) depends on the windowing applied. Systematically studying the ZZT of windowed speech signals, we showed that the windowing needs to be properly performed (a Blackman, Gaussian or Hanning-Poisson window of less than two pitch periods size, centered at glottal closure instants) to be able to extract speech data which have the same mixed-phase ZZT structure as that of the theoretical signal presented in Fig. 1 [3].

The ZZT representation is composed of N-1 zeros plotted on the z-plane. It is important to note that the ZZT of glottal flow are located outside the unit circle (with an exception at the origin) surrounding zero-gaps at the location of the poles presented in the all-pole representation. The ZZT of speech signal is organized such that ZZT of glottal flow and ZZT of vocal tract fall on opposite sides of the unit circle in z-plane.

Mixed-phase characteristics are best observed on group delay spectra since causality/anticausality of a resonance cannot be observed on the amplitude spectra. As seen on the third row of Fig. 1, the group delay spectrum of the glottal flow includes a negative peak that also contributes to the speech signal group delay as a negative peak at low frequency part of the spectrum since convolution in time domain corresponds to addition in group delay domain. The anticausal (outside of the unit circle) poles of glottal flow signal which causes negative group delay peak is presented in the all-pole representation on the fourth row of Fig. 1. The region of convergences (ROC) are also indicated on the all-pole representations, which is also linked to causality-stability of the signals. All of the three signals are stable since unit circle is included in the ROC.

Based on the processing of the ZZT of speech signals, we have developed a glottal flow parameter estimation method [4]. However methods including computation of ZZT are computationally heavy since roots of high order polynomials need to be computed, therefore computationally more efficient methods are needed. This study investigates utilization of linear prediction methods to track outside poles due to the glottal flow contribution in the speech signals.

2   The MixLP algorithm

The proposed Mixed-Phase Linear Prediction (MixLP) algorithm, for detecting a pole pair outside the unit circle corresponding to the contribution of the maximum phase glottal flow signal, is presented in Fig. 2a. First, a glottal closure instant (GCI) synchronous windowing is applied to the speech signal and a single pitch period length signal in-between two consecutive GCI marks is extracted with the method explained in [5]. Obtained speech frame is integrated, to remove the lip radiation contribution. LP-covariance analysis [6] is applied to this signal, which is expected to result in a pole pair outside the unit circle and several other pole pairs inside the unit circle. This is a particular property of the LP covariance analysis, usually considered as a sign of unstability of the estimation algorithm [7].
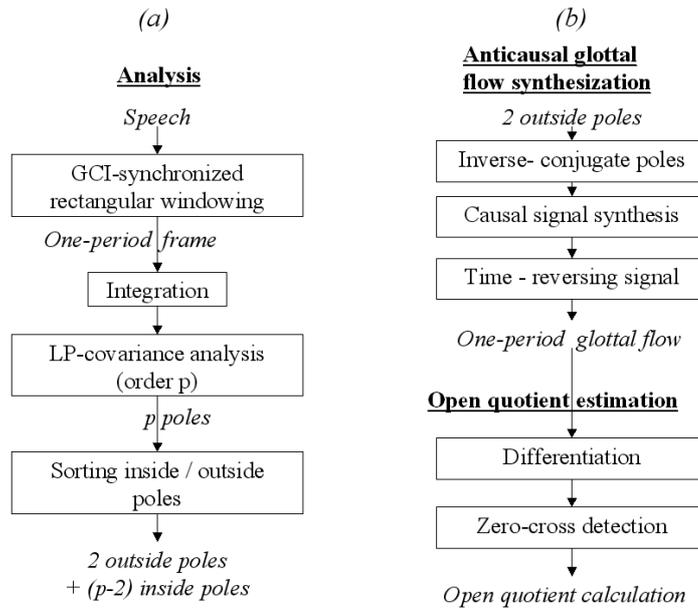

Fig. 2. MixLP algorithm flow diagram

It is interesting to mention some of the investigations performed during the design of the algorithm. Our first investigation was to find optimum windowing since the existence of poles outside the unit circle heavily depends on the applied windowing. Although we have shown through ZZT representation that mixed-phase characteristics can be observed on the Fourier transform of the windowed speech signal when the window is centred at GCI, such windowing is not appropriate for estimation of poles outside the unit circle with LP-covariance. By applying sliding window analysis and checking correctness of estimates on synthetic speech signals, we have observed that the end of the window must be synchronized with the GCI (a few samples before the GCI is a good choice for safety) and including even a few data samples after the GCI results in no poles outside the unit circle most of the time. A second investigation was on the order of the LP analysis, tested in the range [2-32] for 16000Hz synthetic speech signals (for which the LF model was used to synthesize glottal flow excitation and filtered by a four pole-pair all-pole vocal tract filter). The LP degree which provided best estimates is 14 or higher.

3   Analysing synthetic and natural speech

In order to test the MixLP algorithm we have designed a method to estimate the open quotient (Fig. 2b) from poles outside the unit circle. This includes the resynthesis of the glottal flow from the poles, which is achieved by : synthesis of a causal signal by computing the impulse response of a two-pole filter with the inverse-conjugate poles, and time reversion of this signal. A differentiation provides the differentiated glottal flow. In Fig. 3 we present an example of a glottal flow estimate using the MixLP method, together with a glottal flow estimate using a well-known inverse filtering algorithm (PSIAIF [8]). The open quotient is estimated on the differentiated glottal flow with a zero-cross detection method.
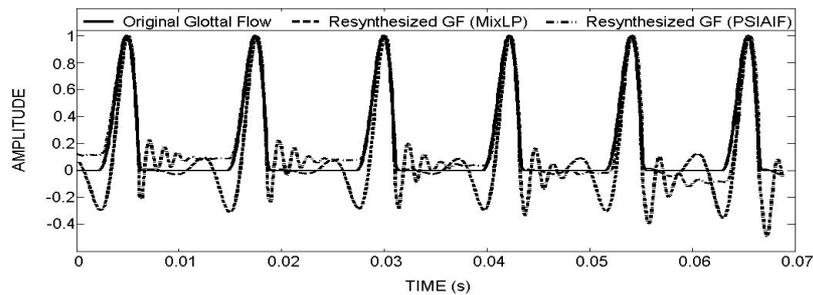

Fig. 3. Resynthesized glottal flow signals obtained with the MixLP and PSIAIF algorithms
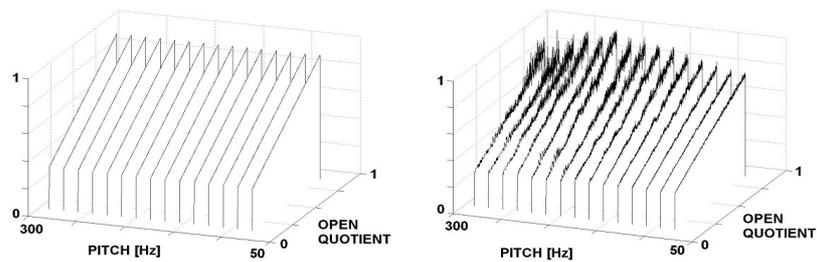

Fig. 4. Real and estimated open quotient , based on synthetic speech (a sustained vowel with constant first formant and return phase) for several values of the pitch

For evaluation of the open quotient estimation method, tests were conducted on synthetic speech signals, in which several parameters (pitch, spectral tilt, first formant frequency and open quotient) were varied systematically and higher formant frequencies are kept constant. Due to space limitations, we only present the output of our test for checking the robustness of estimation to pitch variations in Fig. 4. Some conclusions are : the error is small when the open quotient is higher than 0.7, and otherwise it is negligeable.

Moreover, the open quotient is better estimated if the return phase is short, and especially if the pitch is high. This open quotient estimation method was also compared to a well-known algorithm ([9]). Both methods provide similar results but the MixLP estimation method is more effective when the first formant frequency is small.

The open quotient estimation on natural speech was also tested. As a reference for the open quotient estimation tests, we used open quotient estimates obtained from differential electro-glotto-graph signals by using a thresholding method. Observations on a few natural utterances showed that the MixLP estimation method is not robust as the estimation error depends on the phonetic context.

4   Conclusions

In this paper, we have discussed the mixed-phase characteristics of windowed speech signals through zeros of z-transform (ZZT) representations which corresponds to the set of roots of the z-transform polynomial for a discrete time signal. ZZT representation appears to be an effective representation for studying mixed-phase characteristics of signals in the Z-domain. For speech signals, the observation of mixed-phase characteristics depends on the applied windowing and GCI synchronous windowing is necessary. A linear method was presented here for estimating the maximum phase glottal flow signal and the open quotient. Tests showed that open quotient estimation can be successfully performed on synthetic signals with LP-covariance analysis but the method lacks robustness when real speech signals are analyzed.

References

1. Bozkurt, B., Dutoit, T.: Mixed-Phase Speech Modeling and Formant Estimation, Using Differential Phase Spectrums. Proc. ISCA ITRW VOQUAL, Geneva, Switzerland (2003) 21–24.
2. Doval, B., d'Alessandro, C., Henrich, N.: The Voice Source As A Causal/Anticausal Linear Filter. Proc. ISCA ITRW VOQUAL, Geneva, Switzerland (2003) 15–19.
3. Bozkurt, B., Doval, B., d'Alessandro, C., Dutoit, T.: Zeros of Z-Transform (ZZT) Decomposition Of Speech For Source-Tract Separation. Proc. ICSLP, Jeju Island, Korea (2004).
4. Bozkurt, B., Doval, B., d'Alessandro, C., Dutoit, T.: A Method For Glottal Formant Frequency Estimation. Proc. ICSLP, Jeju Island, Korea (2004).
5. Kawahara, H., Atake, Y., and Zolfaghari, P.: Accurate vocal event detection method based on a fixed-point to weighted average group delay. Proc. ICSLP, Beijing, (2000) 664–667.
6. Makhoul, J.: Linear Prediction : A Tutorial Review. Proc. IEEE (1975) 561-580.
7. Makhoul, J.: Lattice Methods For Linear Prediction, IEEE Trans. On Acoustics, Speech, And Signal Processing, Vol. ASSP-25 (1977) 423-428.
8. Alku, P.: Glottal Wave Analysis With Pitch Synchronous Iterative Adaptive Inverse Filtering. Speech Communication 11 (1992) 109-118.
9. Hanson, H.M.: Glottal Characteristics Of Female Speakers. Ph.D. Thesis, Harvard University (1995).