APPROPRIATE WINDOWING FOR GROUP DELAY ANALYSIS AND ROOTS OF Z-TRANSFORM OF SPEECH SIGNALS

Baris Bozkurt, Boris Doval, Christophe D'Alessandro, Thierry Dutoit

Faculté Polytechnique De Mons, TCTS Lab,Initialis Scientific Park, B-7000 Mons, Belgium, {bozkurt,dutoit}@tcts.fpms.ac.be
LIMSI,CNRS, Po Box 133 – F91403 Orsay, France, {doval,cda}@limsi.fr

ABSTRACT

This study discusses the difficulties of phase spectrum analysis of speech signals and shows that appropriate windowing is very crucial for obtaining reliable phase spectra. The main difficulties of phase based analysis stem from the domination of spiky effects of roots (zeros) of the signal z-transform close to the unit circle. We show how this problem is linked to windowing by discussing zero-patterns for speech signals. Once windowing is performed properly, group delay functions are much less noisy and reveal clearly formant information.

1. INTRODUCTION

Spectral analysis techniques are being used in speech processing for long years in many applications. However, it is often the amplitude component of the Fourier transform spectra that is used for analysis but not the phase component since it is rather difficult to analyse (although phase spectra contain as much information as the amplitude spectra). But recently, more and more studies discuss the importance of phase spectra [1] especially in the domain of speech perception and it is important to understand the sources of difficulties for analysis of phase spectra and find methods to reduce difficulties for further study of the phenomenon.

Very few studies issue analysis of phase spectra of speech signals. Often, the group delay function, which is the negative of differential phase spectrum, is studied since the spectral resonances observed as peaks on amplitude spectrum can also be observed in group delay functions and even with higher resolution. In [2,3], cepstral smoothing based methods are proposed for analysis of group delay functions 'obtained from amplitude spectra only'. However, there is strong evidence that speech signals are not minimum phase due to the anti-causal like character of glottal flow excitation signals [4,5] and group delay obtained from the phase component of the Fourier transform spectra are preferable.

As also discussed in [2,3], the main difficulty in group delay analysis is the domination of spikes due to roots (zeros) of z-transform (as presented in equation(1) where *X(z)* is the z-transform of a discrete time sequence *x(n)*, *Zm* are the roots of the z-transform and *G* is the gain factor). The contribution of zeros close to the unit circle in group delay function are spiky peaks since phase change of z-tranform values in a small range of frequency bins are high.

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n} = G z^{-N+1} \prod_{m=1}^{N-1} (z - Z_m) \quad (1)$$

The roots of the high degree z-transform polynomial can easily be obtained with enough precision for viewing purposes using the *roots* function of Matlab, which finds the eigen values of the associated companion matrix as roots.

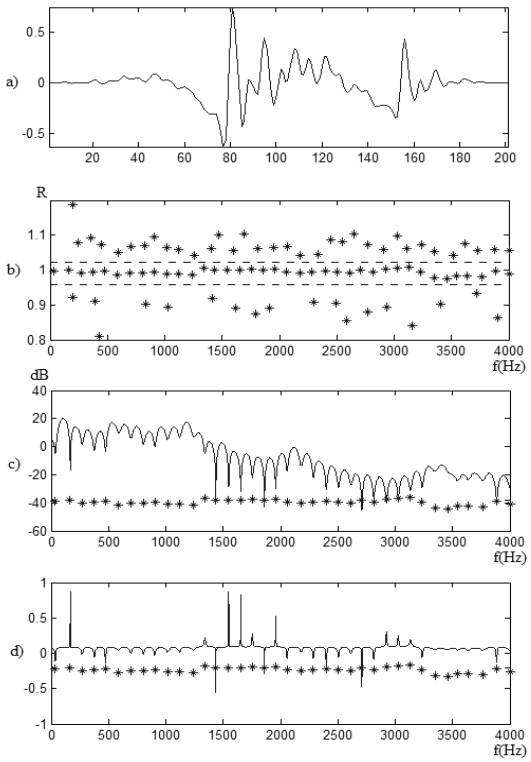

Figure 1: Speech frame taken from a natural utterance (phoneme "*a*" in the word '*party'*), a)Hanning windowed speech frame, b)roots of z-transform polynomial plotted on z-plane in polar coordinates (unit circle corresponds to a line at R=1 in polar coordinates), c)amplitude spectrum, d)group delay function computed from phase spectrum. The zeros close to the unit circle are superimposed on c) and d) to show their effect.

Discrete Fourier transform is the z-transform calculated on unit circle in z-plane. The effect of zeros close to the unit

circle can easily be observed both on amplitude spectra and group delay functions (Fig. 1). The general shape of the group delay function is mainly dictated by the zeros, which makes visual inspection for observing formants useless. However, the effect on amplitude spectrum is much lower and we can still observe formants. This is probably the main reason why amplitude spectrum is much more frequently used than phase spectrum or group delay function in speech processing. Additionally, the values of sharp peaks are reported to be unreliable [2] and for this reason unwrapping of phase spectrum is very problematic since for these frequency points, the derivative of the phase spectrum cannot be reliably estimated.

Due to the strong link between location of zeros and group delay functions, our study discusses the zero-patterns of the speech z-transform (in section 2), the effect of windowing on zero-patterns and shows that windowing is very crucial for obtaining reliable group delay functions for speech analysis (in section 3). With a proper choice of window function, size and position, group delay functions reveal clearly formant information, thus leading to new ways of speech analysis by phase spectra analysis.

2. ZERO PATTERNS FOR SOURCE-FILTER MODEL OF SPEECH

According to the well-known source-filter model for speech, speech signals are produced by exciting the vocal tract system by periodic glottal flow signals. The most widely accepted model for the glottal flow signals is the LF model [6] where the signal is supposed to be composed of two parts: an increasing exponential multiplied by a sinusoid and a decreasing exponential function (both functions are truncated to obtain a one pitch period size data). The periodic version of this function convolved with a time-varying vocal tract system gives the synthetic speech signal. In Fig. 2, zero plots for an LF model signal is presented.

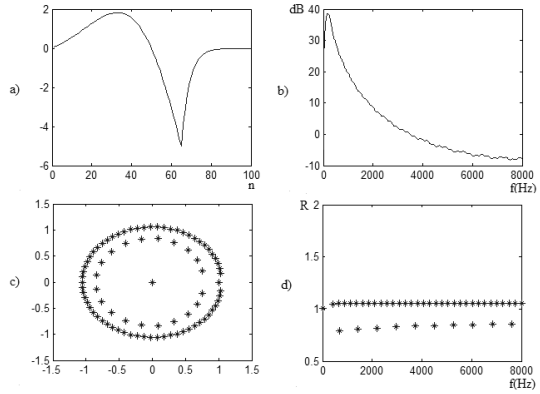

Figure 2: Typical differential LF signal, a)waveform, b)amplitude spectrum, c)zeros on z-plane in cartesian coordinates, d) zeros in polar coordinates

Each exponential component of the differential LF function contributes to the zero plot by a group of zeros lined in parallel to unit circle and the distance of these lined zeros to the unit circle is proportional to the exponential decay coefficient. Analytically, for a simple truncated exponential function, all the roots, *Zm*, of the z-transform polynomial *X(z)* (equation (3)) calculated for the signal *x(n)* (equation (2)) are equally spaced on a single circle with radius *R=a* (equation (4)) (and the zero on the real axis is cancelled by the pole at the same location).

$$x(n) = a^n, n = 0,1...N-1 \quad (2)$$

$$X(z) = \sum_{n=0}^{N-1} a^n z^{-n} = \frac{1-(\frac{a}{z})^N}{1-(\frac{a}{z})} \quad (3)$$

$$Z_m = ae^{j2\pi m/N}, m = 1,2...N-1 \quad (4)$$

Once the glottal flow signal is passed through the vocal tract filter, synthetic speech signal is obtained. In Fig. 3, amplitude spectrum, group delay function and zero plot for a synthetic signal obtained by convolving it with an all-pole filter response (with resonances at 600Hz, 1200Hz, 2200Hz and 3200Hz) is presented. As can be seen, the main contribution of the all-pole filter is observed on the zero pattern (circle) inside the unit circle, and zero-absent regions are due to the resonances of the filter.

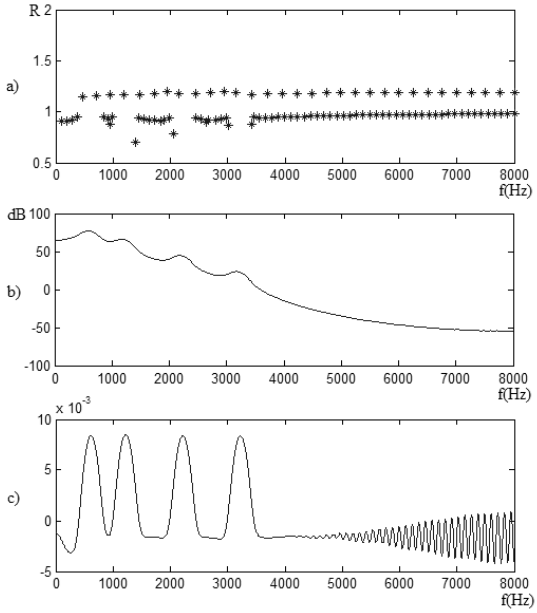

Figure 3: Synthetic speech signal, a)zeros on z-plane in polar coordinates, b)amplitude spectrum, c)group delay function

Both the amplitude spectrum and the group delay function are smooth and provide obvious resonance peaks since the zeros are at a distance from the unit circle. The spiky effects are not observed except at high frequency region of the group delay function (since zeros are closer to unit circle for that region). Each resonance due to a pole-pair inside the unit circle contributes with a positive peak in the group delay function. The advantage of the group delay function over the amplitude spectrum is obvious for this example: resonances are separable with a quite higher resolution than in the amplitude spectrum and the spectral tilt effect is removed thus peak picking based methods can be used effectively for tracking resonances.

Our target is to be able to obtain similar group delay plots from real speech signals. We will show in the next section that proper windowing is very important since windowing changes the location of zeros and therefore the group delay functions to a great extent.

3. WINDOWING EFFECT TO ZERO-PATTERNS

The effect of windowing in zero-patterns is drastic. A term-wise multiplication in time-domain corresponds to term wise multiplication of z-transform polynomial coefficients and how roots of the polynomial are displaced after this operation is an issue very hard to predict analytically. For this reason, we studied the zero-patterns of windowed data by observations rather than mathematical analysis of various examples and we will present some representative examples to explain our conclusions here. We will again use synthetic signals to present the windowing effects on zero-patterns and the group delay functions and then present real speech examples just after.

Both the window size-location and the window function are important. In Fig. 4, zero-plots and group delay functions for windowed synthetic speech (created by exciting a vocal tract all-pole filter by a periodic LF signal) data are presented. Comparing the zero plots with group delay functions: it is obvious that once zeros close to unit circle exist, the group delay functions are noisy, dominated by zero spikes. The group delay function that provides best representation of the vocal tract filter resonances is the one obtained by Blackman window of two period size centered at glottal closure instant (GCI) in Fig. 4f. GCI synchronized windowing has been shown to be effective in speech analysis recently [7]. Here we claim that the main reason is existence of a "zero gap" around unit circle (in other words, grouping of zeros in and out of unit circle but not on it). The source of lining is the glottal excitation zero-pattern presented in Fig. 2. Once the window is placed such that the increasing exponential part is multiplied with the first half of the window, which is also increasing, and the decreasing exponential part is multiplied with the second half of the window, which is also decreasing, the zero-pattern still includes the zero gap around unit circle. When the window is not centered at the increasing-decreasing function change point, the zero-pattern is destroyed (Fig. 4e) resulting in a noisy group delay function.

As size of the window gets larger than two pitch periods (T0), periodicity results in additional zeros close to unit circle which is observed as regularly spaced dips in both amplitude spectra (resulting in harmonic peaks observed on amplitude spectra) and group delay functions. This effect can be observed by comparing Fig 4d and 4f; it is as if additional zeros are placed around unit circle and this is mainly due to the window size (window size being larger than two pitch periods in Fig 4d).

Windowing function is also important but comparatively less important than window size and location once we limit ourselves with commonly used window functions. Due to space limitations, we presented effects of two different windowing functions on the zero-patterns and group delay functions: Blackman and hamming in Fig. 4f and 4g. It is clear that Blackman window is a better choice. In our study with various windowing functions, we observed that three types of windowing functions provide best group delay functions: Blackman, Gaussian and Hanning-Poisson. Hanning-Poisson windows provide the smoothest group delay functions since the Poisson contribution of the window is composed of exponential functions. The windowing with a Hanning-Poisson results in multiplication of exponentials of the Poisson function and the speech signal, thus addition of decay coefficients (in equation (2) in section 2), which shifts zeros further away from the unit circle. For this reason, Hanning-Poisson window is preferable in group delay based analysis methods.

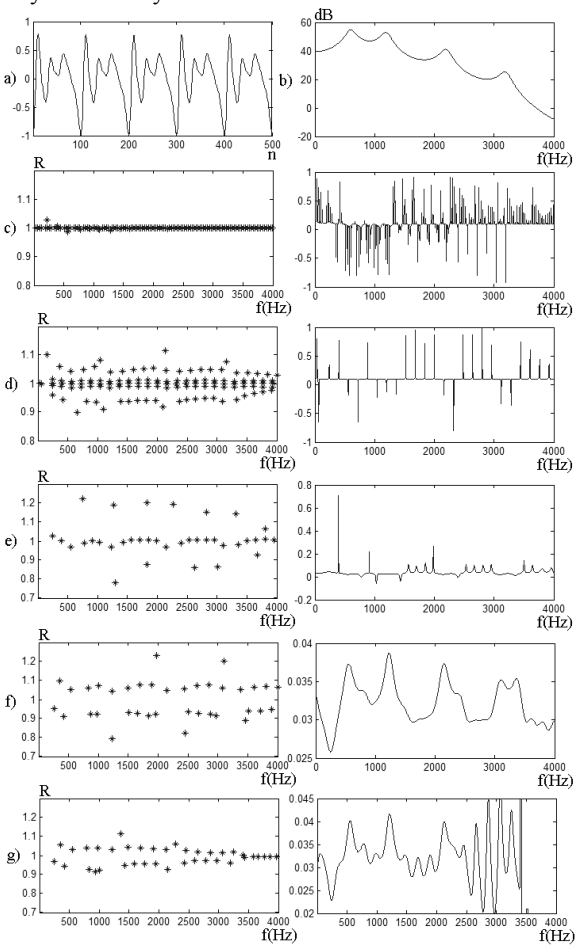

Figure 4: Windowing effects to group delay functions, a)synthetic speech waveform, b)frequency (amplitude) response of the vocal tract filter used, zero plots (first column) and group delay functions (second column) of windowed data; c)five T0 length rectangular window, d)five T0 length Blackman window, e)two T0 length Blackman window not centered at GCI, f)two T0 length Blackman window centered at GCI, g)two T0 length hamming window centered at GCI.

Finally, we present how group delay functions, computed on data with proper windowing provide formant structure on a real speech example (BrianNormal3.wav from Voqual 03 database, for which the uttered sentence is "*she has left for a great party today*" with modal phonation).

In Fig. 5, zero plot and group delay function for a single voiced real speech frame (of two pitch period size, centered at GCI) are presented. The zero-pattern is very similar to the one in Fig. 4f: zeros are lined in and out of the unit circle and a zero gap exists on unit circle. Therefore the group delay function is free of spikes. The formant peaks are easily observed at expected frequency locations for this vowel '*a*' taken from the word '*party*'.

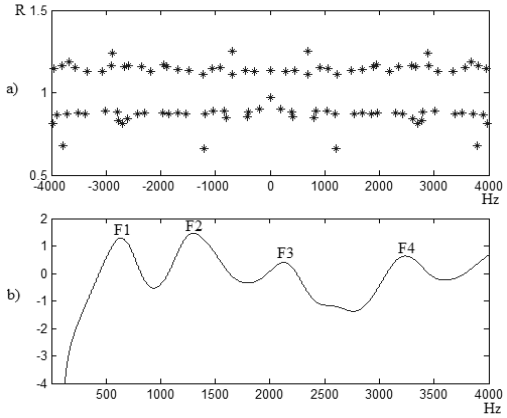

Figure 5: Windowing with Hanning-Poisson (alfa=6) for group delay analysis, a) zero plot of windowed data, b)group delay function of windowed data

In addition, we obtained spectrogram like plot from the positive part of the group delay functions computed for voiced frames of the complete speech data. Fig. 6 shows the spectrogram obtained by group delay functions and amplitude spectra and their correlation is obvious for the formant tracks. This figure shows that the group delay functions indeed carry resonance information of the signal once windowing is properly performed.

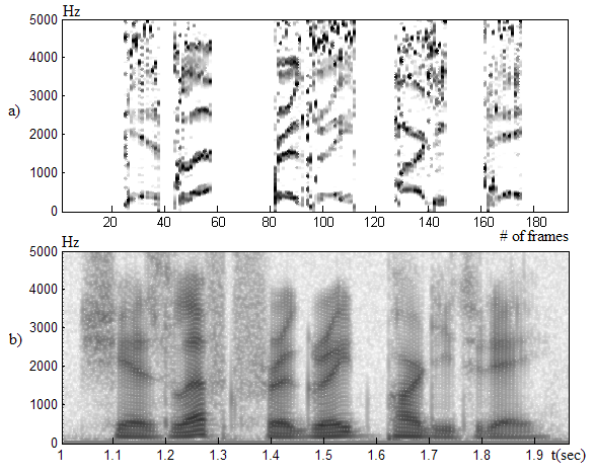

Figure 6: (a) Group delay and (b) amplitude spectrogram for the sentence "*she has left for a great party today*"

4. CONCLUSION

In this study, we have shown that reliable and clean group delay functions revealing resonance information can be obtained from speech signals if windowing is properly performed. This is an important step in speech analysis since phase characteristics of speech signals are known to be important for perception but analysis has been always reported to be very difficult. In some of the studies concerning speech coding and synthesis, the obscurity of phase information is reported and for improving naturalness of re-constructed speech, phase randomisation techniques are tried (often by trial and error methodologies). The issue is still obscure due to lack of reliable tools for phase spectrum analysis.

Windowing is a secondary issue (in importance) even in most of the speech analysis studies. Here we showed that for phase analysis, it is indeed one of the key issues. We expect that the outcomes of this study will be improvements in effectiveness of speech analysis in cases where phase information is valuable, therefore leading to improvement in those technologies (like speech coding and synthesis).

Our further research has shown that glottal flow parameter estimation and formant tracking can be satisfactorily performed on group delay functions obtained in the way presented here. The results of these studies will soon be published.

5. ACKNOWLEDGEMENT

This study is funded by Region Wallonne, Belgium, grant FIRST EUROPE #215095. We also would like to thank Hideki Kawahara, who stimulated this study by his presentation and discussions in ISCA ITRW VOQUAL 2003,Geneva.

REFERENCES

[1] K. K. Paliwal, and L. Alsteris, "Usefulness of phase spectrum in human speech perception," in *Proc. Eurospeech 2003*, Geneva, Switzerland, Sept. 2003, pp. 2117–2120.

[2] B. Yegnanarayana and H. A. Murthy, "Significance of group delay functions in spectrum estimation" *IEEE Trans. on Signal Processing*, vol.40, no.9, pp. 2281-2289, September 1992.

[3] H. A. Murthy and B. Yegnanarayana, "Speech processing using group delay functions" *Signal Processing*, vol.22, no.3, pp. 259-267, March 1991.

[4] B. Doval, C. d'Alessandro, and N. Henrich, "The voice source as a causal/anticausal linear filter," in *Proc. ISCA ITRW VOQUAL 2003*, Geneva, Switzerland, Aug. 2003, pp. 15–19.

[5] B.Bozkurt and T. Dutoit, "Mixed-Phase Speech Modeling and Formant Estimation, Using Differential Phase Spectrums," in *Proc. ISCA ITRW VOQUAL 2003*, Geneva, Switzerland, Aug. 2003, pp. 21–24.

[6] G. Fant, "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.

[7] P. Zolfaghari, T. Nakatani, T. Irino, H. Kawahara, and F. Itakura,, "Glottal closure instant synchronous sinusoidal model for high quality speech analysis/synthesis," in *Proc. Eurospeech 2003*, Geneva, Switzerland, Sept. 2003, pp. 2441–2444.