Aalborg Universitet

Using Zeros of the z-transform in the Analysis of Speech Signals

Dalsgaard, Paul; Pedersen, Christian Fischer; Andersen, Ove Kjeld; Bayya, Yegnanarayana

Published in:
Proceedings of ISCA ITRW - Speech Analysis and Processing for Knowledge Discovery

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Using Zeros of the z-transform in the Analysis of Speech Signals

Paul Dalsgaard, Christian F Pedersen, Ove Andersen [1], *Yegnanarayana Bayya* [2]

[1] Department of Electronic Systems, Aalborg University, Denmark
[2] International Institute of Information Technology, Hyderabad, India
{pd,cfp,oa}@es.aau.dk, yegna@iiit.ac.in

Abstract

The objective of this study is to analyze speech signals using the zeros of the z-transform of the signal. Trajectories of the zeros are used to study the characteristics of speech production. The trajectories are obtained by varying the parameters of the window function used on the signal segment. A skew Poisson function is defined with three parameters to control the window function. The proposed method does not assume any model for the analysis. Both synthetic and natural speech signals are analyzed. The goal of this study is to demonstrate, and eventually separate the information of the source part from the vocal tract system part of the speech production process from the signal. This may provide new and additional insights into the speech production process over and above the existing methods such as the spectral and group delay methods.
The results from experiments with varying Poisson window parameters and speech signals are presented and discussed.
Index Terms: Zeros of the z-transform, Skew Poisson window, Zero trajectories, speech analysis.

1. Introduction

The objective in speech processing is to derive the information about speech production from the acoustic signal. Some popular speech processing methods include model-based spectral techniques and pitch synchronous analysis [1,2]. Some new methods exploit the zeros of the z-transform of a windowed signal samples [3-7]. These new methods are mostly non-model-based, as the zeros are calculated using only the windowed segment of the speech signal. Most of the results presented in [6] are based on observations of the zeros of the windowed signal samples. In the present study we try to provide theoretical justification for the behavior of the zeros of the z-transform of a windowed speech segment. The effect of the parameters of the window function on the zeros of the z-transform is explained in relation to the characteristics of speech production mechanism.

Section 2 gives the theoretical background for the interpretation of the zeros of the z-transform of a windowed signal. Section 3 gives the results of experiments on the trajectories of the zeros for a segment of synthetic speech signal. Section 4 presents the results of analysis for consecutive segment of real speech signal. Section 5 presents conclusions of this study, and possible extensions for further study.

2. Trajectories of zeros of the z-transform

The z-transform of a discrete time sequence

$\boldsymbol{x} = \{x[0], x[1], \cdots, x[N-1]\}$ is defined as $X(z) = \sum_{n=0}^{N-1} x[n] z^{-n}$

or by its rational form $X(z) = x[0] z^{-(N-1)} \prod_{m=1}^{N-1} (z - z_m)$

that for $x[0] \neq 0$ has (N-1) zeros z_m in the z-plane.

Since the objective of this study is to analyze the production characteristics of voiced speech, we consider the analysis with reference to one pitch period of a voiced segment. Each pitch period consists of two main sub-segments, separated by the glottal closure instant. One sub-segment corresponds to the open glottis region, and the other to the closed glottis region. To enable interpretation of the results of analysis in terms of these production characteristics, we propose a skew Poisson window function for the speech segment.

The skew Poisson function is defined as consisting of two parts as follows:

$$p_1[n] = \exp(-\alpha_1 |n - m_t + 1| / m_t), \quad n \in \{0, \cdots, m_t - 1\}$$
$$p_2[n] = \exp(-\alpha_2 |n - m_t + 1| / m_t), \quad n \in \{m_t, \cdots, N - 1\}$$

The parameters α_1 and α_2 enable control of the zero trajectories (see below) whereas m_t makes it possible to choose a setting of the function maximum value (relative to n=0) at the glottal closure instant within the segment for enabling separate analysis of the open and closed segments.

Multiplying the sequence **x** with the skew Poisson function leads to the following two-part discrete sequence

$$\begin{aligned} y &= \{x[0]p_1[0], \cdots, x[m_t - 1]p_1[m_t - 1], \\ &\quad x[m_t]p_2[m_t], \cdots, x[N-1]p_2[N-1]\} \\ &= \{y_1[0], \cdots, y_1[m_t - 1], y_2[m_t], \cdots, y_2[N-1]\} \\ &= \{y_1[n], y_2[n]\} \end{aligned}$$

that has the z-transform $Y_p(z) = Y_1(z) + Y_2(z)$ where

$$Y_1(z) = y_1[0]\, z^{-(m_t - 1)} \prod_{m=1}^{m_t - 1} (z - r_1 z_{1m}) \qquad (1)$$

$$Y_2(z) = y_2[m_t] r_2^{-1} z^{-(N-1)} \prod_{m=1}^{N - m_t - 1} \left(z - \frac{z_{2m}}{r_2}\right) \qquad (2)$$

with $r_1 = \exp(\frac{\alpha_1}{m_t})$ and $r_2 = \exp(\frac{\alpha_2}{m_t})$. Eq.1 and Eq.2 show the dependence of the trajectories of zeros of the z-transform to variable parameter values α_1 and α_2 and that the zeros z_{1m} and z_{2m} of the z-transform follow radial trajectories when parameter m_t is fixed.

The effects of choosing different settings of the parameters α_1, α_2 and m_t are analysed in details in the experiments conducted in Sections 3 and 4 encompassing synthetic and real speech respectively.

3. Analysis of synthetic speech

The synthetic speech signal is derived from a model of the vocal tract filter process defined by four formants. The vocal tract filter is excited by a sequence of pulses each equal to the Liljencrants-Fant glottal flow pulse model [7] with fixed pitch period. The synthetic signal is calculated using a state-space model based on polynomials corresponding to the zeros of the formants – defined by their radii relative to the unit circle and angles relative to the sampling frequency $f_s = 16000$ Hz. Each formant is modelled by a slowly varying sinusoid, and the four formant values are 600, 1400, 2400 and 3200 Hz respectively. The values used are taken from [6] and the pitch frequency is kept fixed at 100 Hz. A single pitch synchronous segment of the synthetic speech signal (N=160) samples covering the open and glottis phases is used as the basis for calculating the zeros of the two polynomials (1) and (2) representing this signal segment. The analysis results in a total number of $(m_t - 1)$ zeros z_{1m} (drawn as blue circles in the figures following) and (N- m_t -1) zeros z_{2m} (red circles). Space only allows showing a part of the results plotted in the z plane.

3.1 Trajectories of zeros of the z-transform

The experiments in section 3.1.1 show the trajectories of the zeros caused by varying the parameters α_1 and α_2 when the location parameter m_t is kept fixed at GCI**.** The experiment in section 3.1.2 is conducted with fixed α_1 and α_2 parameters, and the location parameter m_t varied in the range of ±5 samples around the GCI. The Glottal Closure Instants (GCI) and the Glottal Opening Instants (GOI) are known for the synthetic signal.

3.1.1 α_1 and α_2 varied and m_t fixed at GCI

The zeros in Figure 1 corresponding to $\alpha_1 = \alpha_2 = 0$ (i.e., window is rectangular) are (z_{1m} blue and z_{2m} red) close to the unit circle. It is shown in [6] that the zeros of the z-transform of the Liljencrantz-Fant glottal flow signal consist of two set of zeros. The zeros (blue circles) corresponding to the open phase glottal flow signal $y_1[n]$ are located outside the unit circle, whereas the zeros (red circles) corresponding to the closed phase glottal flow signal $y_2[n]$ are located inside the unit circle. When the α_1 and α_2 parameters are varied, the trajectories (in blue and red) have radial orientations. The experiments have been conducted with the following values: 0<α₁=α₂<5 in steps of 0.2 and m_t is kept fixed at the GCI.

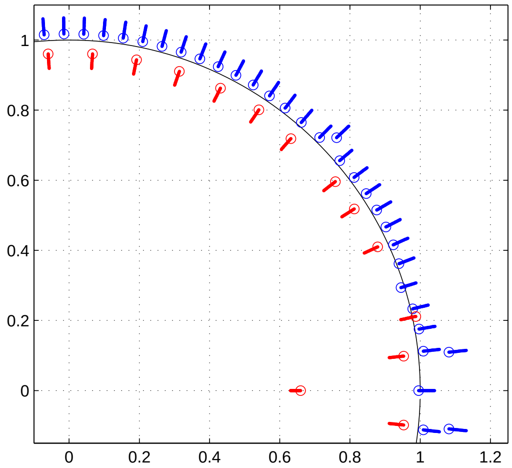

Figure 1: *Zero trajectories for variable α_1 and α_2 parameters and m_t at GCI*

3.1.2 α_1 and α_2 fixed and m_t is varied around GCI

The parameters $\alpha_1 = \alpha_2 = 5$ are fixed whereas m_t is varied in the range of ±5 samples (in steps of one sample) around GCI.

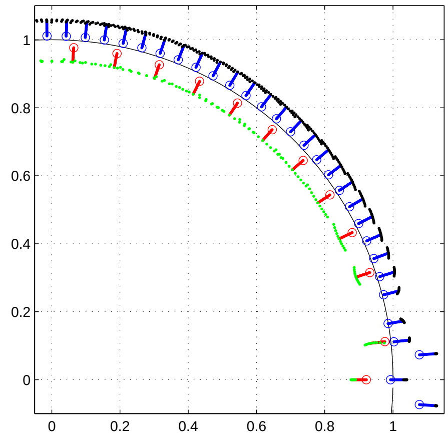

Figure 2: *Zero trajectories for fixed α_1 and α_2 parameters and m_t relocated around GCI*

The results are shown in Figure 2. The radial oriented trajectories are the trajectories of the zeros for z_{1m} (blue) and z_{2m} (red) repeated from Figure 1, whereas the black (contributed from $y_1[n]$) and green (contributed from $y_2[n]$) traces are caused by the relocations of m_t around GCI.

It is noted that the direction of the traces are very different from the trajectories caused by a variation in the α_1 and α_2 parameters. The most noticeable observation is that they no longer have radial orientation.

One explanation to the clear difference of these traces as compared to the trajectories in Figure 1 is that a relocation in m_t results in changes of both of the characteristic equations Eq.1 and Eq.2 used for calculating the zeros z_{1m} and z_{2m}. A relocation of the m_t at one sample away from GCI results in one less zero in z_{1m} and one additional zero in z_{2m} which are different from the number of zeros in z_{1m} and in z_{2m} when m_t is located at GCI. However, the total number of zeros for the entire segment remains equal to (N-1). According to the definition of a trajectory as given in Section 3 the traces caused by m_t - not located at GCI - are termed *pseudo trajectories*.

The experiments further show that *pseudo trajectories* reflect changes in angular value of some of the zeros which is not the case for radial trajectories.

It is noted that relocations of m_t around GCI cause changes in a number of the last or first coefficients (equal to signal samples) in the polynomials $Y_1(z)$ and $Y_2(z)$ and that the sample values at the GCI is relatively large. This may cause relatively large systematic changes in the position of some zeros in the z-plane. It is also observed that the relocation of m_t around the GCI seems to have larger effects on some of the zeros – both in terms of absolute value and angular value.

4. Analysis of real speech

In this section voiced segments of pitch synchronous real speech are analysed. Section 4.1 gives analysis on a segment of real speech corresponding to the analysis in section 3.2. In section 4.2 a real speech segment is shown together with the skew Poisson function for non-equal values of the α_1 and α_2 parameters and in section 4.3 zeros are shown for sequence of consecutive $y_2[n]$ closed phase signal segments.

The signal used is from a male speaker database containing Danish di- and triphone embedded in nonsense carrier words. The data are recorded in an anechoic room using a sampling rate of 44 kHz and later down sampled to 16 kHz.

Each segment consists of the speech samples between two consecutive Glottal Opening Instants. The m_t location within each segment is calculated as the distance (in samples) between the GOI and the GCI of the segment. The Poisson parameter m_t is located at the Glottal Closure Instant whereas the α_1 and α_2 parameters are varied as in section 3.1.1. The GOIs and GCIs of the signal are given by the DYPSA[1] algorithm [8].

4.1 Trajectories of zeros of the z-transform

Figure 3 shows the results from a single segment experiment using real speech. The trajectories are given by varying $0<\alpha_1=\alpha_2<5$ in steps of 0.2. The corresponding pseudo trajectories are with $\alpha_1 = \alpha_2 = 5$ whereas m_t is varied in the range of ±5 samples around the GCI. The results are very similar to those of the synthetic signal.

4.2 Details from open and closed phase signals

To demonstrate the effect of choosing different values of the α parameters, the following experiment was conducted with $\alpha_1 = 5$, $\alpha_2 = \alpha_1/10$ and m_t at GCI. Figure 4 show three curves and a (black) vertical line positioned at the GCI location within the signal. The original speech signal segment (in black), the resulting $y_1[n]$ and $y_2[n]$ signals (in blue) and the actual skew Poisson function (red).

For the chosen value of the α_2 parameter, the windowed closed phase signal segment is almost equal to the original signal $x_2[n]$. The windowed open phase signal can be extracted by choosing the α_2 parameter larger than the α_1 parameter.

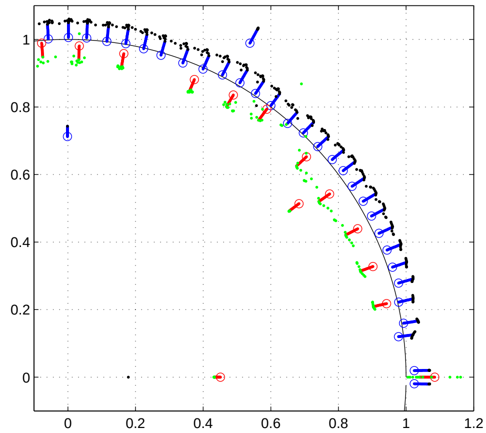

Figure 3: *Zero and pseudo trajectories for real speech segment*

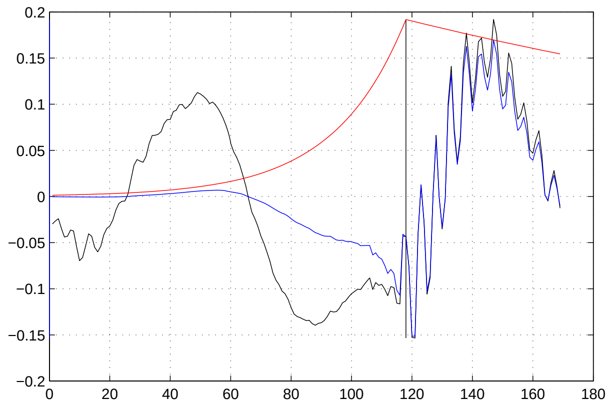

Figure 4: *Speech segment (black) and resulting (black) windowed signal parts from open and closed glottis phases. The red curve shows the actual Skew Poisson window.*

4.3 Consecutive speech segments

Zeros of the z-transform is calculated from a sequence of consecutive speech segments and in the following experiment. A set of non-overlapping voiced speech segments is established by choosing data from thirty-five consecutive voiced intervals. In each segment m_t is positioned at GCI. The α parameters are fixed with values $\alpha_1=5$ and $\alpha_2=\alpha_1/10$. With reference to the footnote in the previous column, the accuracy of the DYPSA placement of GCI and GOI instants is within ±4 samples (at 16 kHz sampling frequency), and it is noted that a possible inaccurate m_t location may result in zero relocations as has been seen in the two previous sections. This is however not taken into account here.

[1] The DYPSA algorithm is an automatic and reliable estimation of glottal closure instants (GCIs) in voiced speech. Reliable GCI estimation is essential for closed-phase speech analysis, from which can be derived features of the vocal tract and, separately, the voice source. DYPSA is automatic and operates using the speech signal alone without the need for an EGG or Laryngograph signal. DYPSA incorporates a new technique for estimating GCI candidates and employs dynamic programming to select the most likely candidates according to a defined cost function. Results for DYPSA show GCI detection accuracy to within ±0.25ms on 87% of the test database and fewer than 1% false alarms and misses.

The results plotted Figure 5 show a selected set of the z_{2m} zero locations in the z-plane from the thirty-five segments of voiced speech.
The results show characteristic clustering of zeros along the contour of the unit circle. It is to be noted that a number of zeros from the consecutive segments cluster at specific locations around the unit circle, at least for the zero clusters corresponding to the lower frequencies that are shown here.

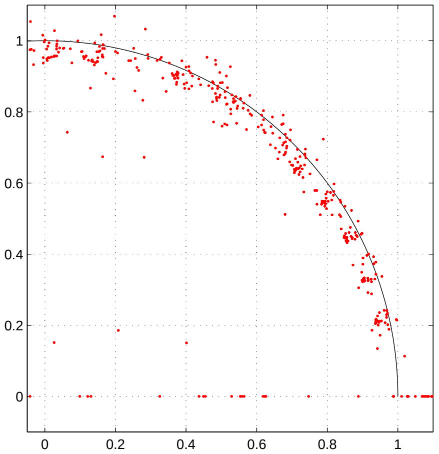

Figure 5: *Real speech analysis – m_t at GCI and α_1 and α_2 values fixed*

In general there seem to be some harmonic structure in the cluster locations. The spreading in zero locations (absolute value and angle) within a cluster represents changes in the $Y_2(z)$ polynomials over time in correspondence with the changes in speech samples (and pitch value) from segment to segment.

4.4 Spectrum of consecutive speech segments

The spectrum for the closed glottis phase zeros resulting from section 4.3 is shown in Figure 6.

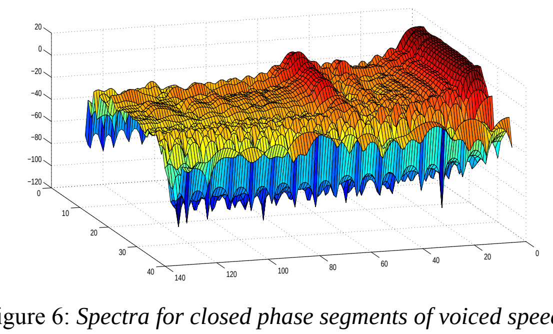

Figure 6: *Spectra for closed phase segments of voiced speech*

The magnitude spectra are calculated along the unit circle and using a resolution of M=256 bins on the basis of the following equation that is derived from the results in section 2.

$$S_2(k) = x[m_t] r_2^{-1} \prod_{m=1}^{N-m_t-1} \left| \exp\left(j2\pi \frac{k}{M}\right) - y_{2m} \right|$$

$$k \in \{0, \cdots, M-1\} \; and \;\; y_2[n] = r_2^{-n} x_2[n]$$

5. Conclusions

The objective of this study was to analyse whether the zeros of the z-transform can be used in establishing new ways of analysing and extracting inherent information from speech signals. One of the objectives was to demonstrate the dependence of the zero locations in the z-plane on the choice of the Poisson window parameters α_1 and α_2 that have the effect of radial contracting or extracting the zero locations. With the goal of enabling analysis of voiced speech segments separately in their open and closed phase parts, a skew Poisson function is defined with one extra parameter. This parameter allows a specific positioning of the window maximum value at a specific event. Here this is the sample of the speech signal corresponding to the glottal closure instant.
A number of experiments have been conducted to study the effect of different parameter settings, and mainly to demonstrate the effects on the zeros of the z-transform caused by relocations of the maximum value parameter away from the glottal closure instant. It is clearly demonstrated that this has a significant effect on the zero locations as relocations results in shifts in the angular position of zeros equivalent to a change in frequency. This has effects on the interpretation of results from further speech analysis when zeros are used for calculating the signal spectrum or the group delay.
It is emphasized that the experiments so far are only based on single segments of speech with limited speech material. More data need to be analysed to examine the usefulness of zeros of the z-transform for providing new ways of processing speech signals. It is also to be noted that the signals used in the experiments are noise-free.

6. References

[1] Bäckström T., Alku P., “All-pole modeling technique based on weighted sum of LSP polynomials”, IEEE Signal Processing Letters, Vol. 10, No. 6, pp. 180-183, June 2003.
[2] Alku P., Story B., Airas M., “Estimation of the voice source from speech pressure signals: Evaluation of an inverse filtering technique using physical modelling of voice production”. Folia Phoniatrica et Logopaedica, Vol. 58, No. 2, pp. 102-113, 2006.
[3] Doval B., D'Alessandro C., Henrich N., “The voice source as a causal/anticausal linear filter”, VOQUAL'03. Voice quality: functions, analysis and synthesis. Geneva, Switzerland, 2003
[4] Bozkurt B., Doval B., D'Alessandro C., Dutoit T.. “Appropriate windowing for group delay analysis and roots of z-transform of speech signals.”, 12th European Signal Processing Conference. EURASIP, 2004.
[5] Bozkurt B., Doval B., D'Alessandro C., Dutoit T., “Zeros of Z-transform (ZZT) decomposition of speech for source-tract separation”, Interspeech 2004-ICSLP, Korea, 2004.
[6] Bozkurt B., “Zeros of the z-transform (ZZT) representation and chirp group delay processing for the analysis of source and filter characteristics of speech signals”, PhD Dissertation, Faculté Polytechnique de Mons, 2005.
[7] Fant, G., ”The voice source in connected speech”, Speech Communication, Vol. 22, pp. 125–139, 1997.
[8] Kounoudes A., Naylor P. A., Brookes M., “The DYPSA Algorithm for estimation of glottal closure instants in voiced speech”, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. Volume 1, I-349 - I-352.