Aalborg Universitet

Separation of Mixed Phase Signals by Zeros of the Z-transform

A Reformulation of Complex Cepstrum Based Separation by Causality

Pedersen, Christian Fischer; Andersen, Ove; Dalsgaard, Paul

Published in:
I E E E International Conference on Acoustics, Speech and Signal Processing. Proceedings

DOI (link to publication from Publisher):
10.1109/ICASSP.2010.5495060

Publication date:
2010

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

SEPARATION OF MIXED PHASE SIGNALS BY ZEROS OF THE Z-TRANSFORM - A REFORMULATION OF COMPLEX CEPSTRUM BASED SEPARATION BY CAUSALITY

C.F. Pedersen, O. Andersen, P. Dalsgaard

Department of Electronic Systems, Aalborg University, Aalborg, Denmark

{cfp,oa,pd}@es.aau.dk

ABSTRACT

In recent studies, a non-parametric speech waveform representation (rep.) based on zeros of the z-transform (ZZT) has been proposed. The ZZT rep. has successfully been applied in separating mixed phase signals, e.g. pitch-synchronously windowed speech, into min/max phase by using the unit circle as discriminant. As the ZZT rep. is obtained by factorization of the z-transform, relations to the complex cepstrum (CC) exist. The present paper interrelates the ZZT rep. with the CC via factorization of the z-transform, and demonstrates that unit circle discrimination of a ZZT rep. can be formulated as a CC based separation by causality. A numerical experiment supplements theory by separating a range of LF glottal flow waveforms into their opening and closing phase constituents. Further, randomized mixed phase sequences are separated. As the CC based separation also can be obtained via FFT it has a lower time and space complexity than the ZZT based counterpart.

Index Terms— Zeros of the z-transform, complex cepstrum, mixed phase separation

1. INTRODUCTION

Production of human speech is commonly considered as either quasi-periodic or randomized sources of energy, i.e. airflow through the glottis in the larynx, modulated by a time-varying filter function determined by the shape of the supralaryngeal vocal tract; this model, referred to as the *source-filter* model, is often attributed to [1]. Separation of the source and filter components in voiced speech has been a subject of study for several years in the speech science community. A prevalent separation method, that dates back about fourty years, is by homomorphic deconvolution via complex cepstrum (CC) [2, 3, 4]; cf. [5] for a survey of CC analysis techniques and application domains.

Recently, a non-parametric speech waveform representation (rep.) based on zeros of the z-transform (ZZT) has been proposed [6]. The ZZT rep. has successfully been applied in separating mixed (mix) phase signals into minimum (min) and maximum (max) phase, e.g. separation of source and filter components in pitch-synchronously windowed speech, by using the unit circle (UC) in the complex plane as discriminant [6]. A tight relation between the ZZT rep. and the CC exist as they both can be obtained by factorization of the z-transform [6, 7].

The present paper demonstrates the relationship between the ZZT rep. and the CC via factorization of the z-transform, and demonstrates that UC discrimination of a ZZT rep. can be formulated as CC based separation by causality. The two separation methods are denoted ZSM and CSM_{fac} (ZZT/CC based mix phase signal separation method); subscript *fac* indicates the CC is obtained by factorization. Thereby, the recently proposed ZSM is interrelated with a well-known and developed body of theory.

To supplement the analysis, a numerical experiment is conducted. Employing the ZSM and CSM_{fft} (subscript *fft* denotes that the CC is obtained by FFT), a range of glottal flow waveforms, generated by the Liljencrants-Fant (LF) glottal flow model (GFM), are separated; a LF GFM sequence is mix phase, the opening part is max and the closing part is min phase [8]. Also, a range of randomized mixed phase sequences are separated. The CSM_{fft} has a lower time and space complexity than the ZSM; this is supplemented by measuring running times and memory usage during the experiments.

The remainder of this paper is organized as follows. The relationship between the ZSM and CSM_{fac} is demonstrated in section 2, and in section 3 mix phase separation is exemplified by a numerical experiment. The results are presented in section 4, and in section 5 the results are discussed along with future perspectives.

2. SEPARATION OF MIXED PHASE SIGNALS

In this section, the ZSM and CSM_{fac} are established and interrelated. First, the concepts of min, max and mix phase sequences are defined.

2.1. Minimum, maximum and mixed phase signals

A signal is min, max or mix phase if its z-transform is min, max or mix phase respectively. This leads, by factorization of the z-transform polynomial, to the following definition.

Definition 1 *Min, max, and mix phase polynomials*
Denote the zeros of the complex polynomial

$$X(z) = \sum_{n=0}^{N-1} x_n z^{-n} = \frac{x_0 \prod_{m=1}^{N-1}(z - z_m)}{z^{(N-1)}}, \quad x_0 \neq 0$$

by $z_1, z_2, ..., z_M \in \mathbb{C} \setminus \{0\}$*; then* $X(z)$ *is*

$$\begin{array}{ll} \textit{min phase if} & |z_m| < 1, \\ \textit{max phase if} & |z_m| > 1, \ \textit{and} \\ \textit{mix phase if} & \exists i, j : |z_i| < 1 \wedge |z_j| > 1 \end{array}$$

where $m, i, j \in [1; M]$

2.2. Zeros of the z-transform based separation (ZSM)

The ZZT rep. is defined as an all-zero rep. of the z-transform of a signal sequence, i.e.

Definition 2 *Zeros of the z-transform*
The zeros of the z-transform of a sequence $(x_n)_{n=0}^{N-1} \subset \mathbb{R}$ *are defined as* $z_1,\ z_2, ..., z_M \in \mathbb{C} \setminus \{0\}$ *such that* $X(z_i) = \sum_{n=0}^{N-1} x_n z_i^{-n} = 0$ *for* $1 \leq i \leq M$.

Factorization of the z-transform yields (cf. def. 1)

$$X(z) = \frac{x_0 \prod_{m=1}^{M_o}(z - z_{o,m}) \prod_{m=1}^{M_i}(z - z_{i,m})}{z^{(N-1)}} \tag{2.1}$$

Provided $x_0 \neq 0$.

Hence, the ZZT is an unordered sequence of the zeros of the assumed polynomial function in the numerator deducted by any poles, i.e. zeros at zero in this case, as these lead to an undefined z-transform. The min/max phase separation is done by separating the zeros inside the UC, $|z_{i,m}| < 1$, from those outside, $|z_{o,m}| > 1$ (cf. def. 1).

2.3. Complex cepstrum based separation (CSM$_{\text{fac}}$)

The complex cepstrum is defined as

Definition 3 *Complex cepstrum [4, chap. 12]*
The complex cepstrum, $(\hat{x}_n) \subset \mathbb{R}$*, of* $(x_n) \subset \mathbb{R}$ *is defined as*

$$\begin{aligned} (\hat{x}_n) &= F^{-1}\Big\{log_e\Big[\ F\{(x_n)\}\Big]\Big\} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} log_e[X(e^{i\omega})]e^{i\omega n}\, d\omega \end{aligned}$$

Where $F\{(x_n)\} = X(e^{i\omega}) = \sum_n x_n e^{-i\omega n}$ is the discrete-time Fourier transform, $F^{-1}\{X(e^{i\omega})\} = (x_n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{i\omega})e^{i\omega n}\, d\omega$ is the inverse ditto, and $log_e(z) = log_e|z| + i\ arg(z)$ is the complex logarithm where $arg(z)$ is the continuous phase function.

By expressing the z-transform as in (2.1), the CC can be computed by factorization of the z-transform polynomial; this method circumvents phase unwrapping [4, chap. 12] [7].

$$(\hat{x}_n) = \begin{cases} -\sum_{m=1}^{M_i} z_{i,m}^n/n, & n > 0, \\ \sum_{m=1}^{M_o} z_{o,m}^n/n, & n < 0. \end{cases} \tag{2.2}$$

Where $z_{i,m}$ are zeros inside the UC and $z_{o,m}$ are zeros outside. For completeness, $(\hat{x}_n) = ln|A|$ for $n = 0$.

Hence, the min/max phase separation is achieved by causality of the CC. The CC is causal, i.e. $(\hat{x}_n) = 0$ for $n < 0$, iff (x_n) is min phase. Equivalently, the CC is anti-causal, i.e. $(\hat{x}_n) = 0$ for $n > 0$, iff (x_n) is max phase.

Finally, by comparing (2.1) and (2.2) the relationship between the ZSM and the CSM$_{\text{fac}}$ can be established. In both methods, zeros within the UC explain min phase signals and zeros outside the UC explain max phase signals.

3. NUMERICAL EXPERIMENT

The experiment is to compare time and space complexity, running time, and memory usage of the ZSM and CSM$_{\text{fft}}$. The dataset of mix phase sequences, on which to execute the algorithms, is divided in two halves; one half is LF GFM sequences as the origin of the study is source-filter separation and one half is, to generalize the experiment, randomized sequences constructed by convolving min and max phase sequences. Further, the latter allow easy verification of the separation by reconstructing the min/max phase constituents.

3.1. Data material

3.1.1. LF Glottal flow sequences

The LF GFM is defined by the derivative of the glottal flow;

Definition 4 *Liljencrants-Fant glottal flow model [9]*

$$\begin{array}{lll} e_o(t) & = E_0 e^{\alpha t} sin(\omega_g t), & t_0 \leq t \leq t_e \\ e_c(t) & = -\frac{E_e}{\epsilon t_a}\left(e^{-\epsilon(t-t_e)} - e^{-\epsilon(t_c-t_e)}\right), & t_e < t \leq t_c \\ e_s(t) & = 0, & t_c < t \leq T \end{array}$$

Where $e_o(t)$, $e_c(t)$ and $e_s(t)$ are the opening, closing and shut parts respectively. The LF GFM is used to generate 1000 mix phase sequences; $e_o(t)$ is max and $e_c(t)$ is min phase [8]. Possible sequences with zeros on the UC are removed. The following LF-GFM parameter variations are allowed.

$$\begin{array}{ll} t_0 = 0, & t_c = 0.01 \\ t_p \in [0.23; 0.72]t_c, & t_e \in [0.30; 0.80]t_c \\ t_a \in [0.0004; 0.2523]t_c, & E_e = 1 \end{array}$$

The parameter ranges - each sampled ten times equidistantly - span the predominant varieties of normal speech quality [8]. Typical sampling frequencies in speech rep.s are in $[4; 24]kHz$; in this experimement $t_c = 0.01$, thus the range of coefficient sequence lengths are $N \in [40; 240]$; however, this is expanded to $N \in [40; 539]$ to illustrate the asymptotic behaviour of ZSM and CSM$_{\text{fft}}$. With 500 elements in $[40; 539]$ two different sequences per length exist.

3.1.2. Randomized sequences

Th. 1 is employed to supplement the data material.

Theorem 1 *Eneström-Kakeya [10]*

$$\text{If} \quad p(a,z) = \sum_{n=0}^{N} a_n z^n \quad \text{with} \quad a_0 \geq a_1 \geq \ldots \geq a_N > 0,$$

then all the zeros of $p(a,z)$ lie outside the open unit disk. Conversely, if $a_N \geq a_{N-1} \geq \ldots \geq a_0 > 0$, then all the zeros of $p(a,z)$ lie in the closed unit disc.

For max phase sequence generation, i.e. $a_0 \geq a_1 \geq \ldots \geq a_N > 0$ all in $\mathbb{R}$, let $a_N = 1$, $a_{N-1-i} = a_{N-i} + r$ for $i \in [0; N-1]$ and $r \sim \mathcal{U}[0,1]$. The continuous uniform distribution is denoted by $\mathcal{U}$. Equivalently, min phase sequences are generated by reversing the coefficient ordering. 1000 min and 1000 max phase sequences are generated; possible sequences with zeros on the UC are removed. The min/max phase sequences - same lengths ± 1 - are convolved to obtain mix phase sequences of lenghts $N \in [40; 539]$; again, two different sequences per length exist.

3.2. Comparison of time complexity and running time

The drawback of the ZSM is time complexity of factorizing high-degree polynomials; the Matlab© function `roots()` estimates eigenvalues of a polynomial's companion matrix in time $\mathcal{O}(n^3)$ [11]. In the remainder, Matlab© function names are set with typewriter typeface. CSM_{fft} is based on FFT, IFFT, phase-unwrapping, and the real logarithm; $\texttt{fft}(), \texttt{ifft}() \in \mathcal{O}(n \log n)$ [12] and $\texttt{unwrap}(), \texttt{log}() \in \mathcal{O}(n)$. Combined, this yields an asymptotic time complexity for CSM_{fft} of $\mathcal{O}(n \log n)$ as FFT and IFFT dominates. Evidently, the time complexity of CSM_{fft} is lower than ZSM.

Fig. 1 illustrates running times (measured with `tic/toc`) of ZSM and CSM_{fft} as functions of input sequence length during two consecutive executions of ZSM and CSM_{fft} on the dataset. As four different sequences, two from each dataset, with the same length exist, the average running time per sequence length is reported. By visual inspection, practice illustrate theory. The sudden jump and offset (persistent through ten additional tests) in the ZSM curve at 2^9 pertain presumably to the `eig()` implementation.

3.3. Comparison of space complexity and memory usage

The function `roots()` has a space complexity of $\mathcal{O}(n^2)$ [11]. The CSM_{fft} core functions `fft()`, `ifft()`, `unwrap()` and `log()` are all in $\mathcal{O}(1)$, i.e. the space needed per function is constant wrt. input sequence length. For `fft()` and `ifft()` in-place algorithms keep space complexity constant, e.g. [12, 13]. Combined, this yields a constant space complexity for CSM_{fft} which is lower than the complexity of ZSM.

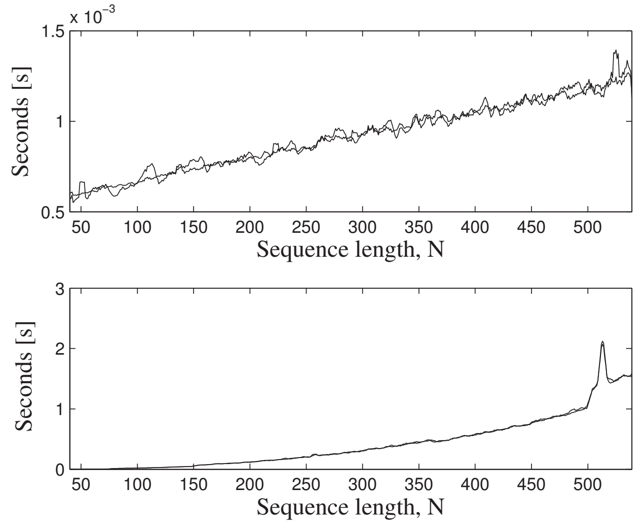

Fig. 1. Running times in two consecutive executions of CSM_{fft} (upper) and ZSM (lower) on the dataset.

Measuring memory usage is deceptive when memory management is handled automatically; in this experiment Matlab© determines allocation timing/quantity and potential time-space-tradeoffs. When executing a function, the function and its context is pushed into memory, so bare consumption of ZSM and CSM_{fft} cannot be isolated. However, it is still interesting, from a practical point of view, to illustrate and exemplify the memory used by Matlab© during execution of the algorithms; this is done in Fig. 2 where memory is measured with `memstats`. As in section 3.2 the average per sequence length is found.

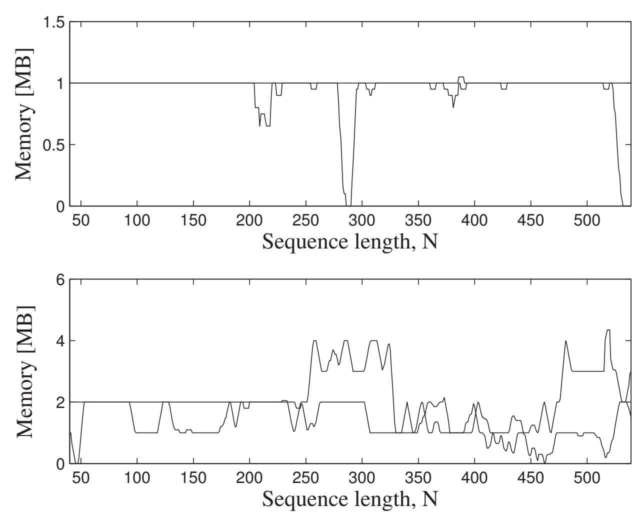

Fig. 2. Memory usage in two consecutive executions of CSM_{fft} (upper) and ZSM (lower) on the dataset.

By visual inspection, ZSM uses on average more memory than CSM_{fft} which to some degree support theory; especially for one of the runs, the usage is constant for CSM_{fft} . The asymptotic behaviour is not expressed.

4. RESULTS

A recent mixed phase separation method, ZSM (cf. (2.1)), based on zeros of the z-transform has been interrelated to a well established method, CSM_{fac} (cf. (2.2)), based on complex cepstrum causality. It is demonstrated that both methods rely on z-transform factorization and the distribution of zeros on either side of the UC in the complex plane; zeros within the UC explain the min phase signal component and zeros outside the UC explain the max phase component.

Further, the ZSM and CSM_{fft} have been compared theoretically wrt. time and space complexity (cf. section 3.2 and 3.3 respectively) and practically wrt. running time (cf. fig. 1) and memory usage (cf. fig. 2). It is shown that the CSM_{fft} outperforms ZSM in both time and space complexity; this is underpinned by the running time experiment and to some degree by the memory usage experiment.

5. DISCUSSION

As the ZZT rep. rely on factorization of the z-transform, time complexity impede real time operation in continuous source-filter separation. To alleviate the burden, it would be relevant to investigate methods for re-estimation of zeros based on coefficient pertubations, i.e. utilize a-priori knowledge about coefficients.

Table 1 summarizes pros/cons - based on the results from this paper and related literature - for the ZSM and CSM_{fft} .

Alg.	Pros	Cons
ZSM	No phase unwrapping	In time $\mathcal{O}(n^3)$
	No aliasing	In space $\mathcal{O}(n^2)$
		No zeros must be on UC
CSM_{fft}	In time $\mathcal{O}(n \log n)$	Phase unwrapping
	In space $\mathcal{O}(1)$	Aliasing

Table 1. Pros and cons of ZSM and CSM_{fft} .

6. REFERENCES

[1] Gunnar Fant, *Acoustic Theory of Speech Production*, Mouton and Co., Hague, Netherlands, 1960.

[2] Bruce P. Bogert, M.J.R Healy, and John W. Tukey, “The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking,” *Proc. of the Symposium on Time Series Analysis (M. Rosenblatt, ed.), Chap. 15*, vol. John Wiley and Sons, Inc., New York, pp. 209–243, 1963.

[3] Allan V. Oppenheim and Ronald W. Schafer, “Homomorphic analysis of speech,” *IEEE Transactions on Audio and Electroacoustics*, vol. AU-16, no. 2, pp. 221–226, June 1968.

[4] Alan V. Oppenheim and Ronald W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall Inc., international edition, 1989.

[5] Donald G. Childers, David P. Skinner, and Robert C. Kemerait, “The cepstrum: A guide to processing,” *Proceedings of the IEEE*, vol. 65, no. 10, pp. 1428–1443, October 1977.

[6] Baris Bozkurt, *Zeros of the z-transform (ZZT) representation and chirp group delay processing for the analysis of source and filter characteristics of speech signals*, Ph.D. thesis, Faculté Polytechnique de Mons, Belgium, October 2005.

[7] Kenneth Steiglitz and Bradley Dickinson, “Computation of the complex cepstrum by factorization of the z-transform,” in *Acoustics, Speech, and Signal Processing*, Hartford, Connecticut, USA, May 1977, IEEE, vol. 2, pp. 723– 726.

[8] Christian Fischer Pedersen, Paul Dalsgaard, and Ove Andersen, “On separability of the opening/closing phases of the LF glottal flow model by zeros of the z-transform,” *IEEE Transactions on Audio, Speech, and Language Processing*, submitted for publication 2009.

[9] G. Fant, J. Liljencrants, and Q. Lin, “A four-parameter model of glottal flow,” STL-QPSR 26-4, Dept. of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, Sweden, 1985, pages 1-13.

[10] P. Borwein and T. Erdélyi, *Polynomials and Polynomial Inequalities*, Springer, 1995.

[11] Gary A. Sitton, C. Sidney Burrus, James W. Fox, and Sven Treitel, “Factoring very-high-degree-polynomials,” *IEEE Signal Processing Magazine*, vol. 20, no. 6, pp. 27–42, November 2003.

[12] Herbert S. Wilf, *Algorithms and Complexity*, A.K. Peters, Ltd., 2nd edition, 2002, ISBN: 1-56881-178-0.

[13] H.W. Johnson and C.S. Burrus, “An in-order, in-place radix-2 FFT,” in *International Conf. on Acoustics, Speech, and Signal Processing*, San Diego, California, USA, March 1984, IEEE, vol. 9, pp. 473–476.