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 forty 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 CSMfac (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 CSMmin (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 CSMmin 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 CSMfac 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 CSMfac 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
\[ X(z) = z^{-N} \sum_{n=0}^{N-1} x_n z^{-n} = \frac{x_0 \prod_{m=1}^{M_i} (z - z_{i,m}) \prod_{m=1}^{M_m} (z - z_{m,0})}{z^{(N-1)}} \], \quad x_0 \neq 0 \]
by \( z_1, z_2, \ldots, z_M \in \mathbb{C} \setminus \{0\} \); then \( X(z) \) is

- min phase if \( |z_m| < 1 \),
- max phase if \( |z_m| > 1 \), and
- mix phase if \( \exists i, j : |z_i| < 1 \land |z_j| > 1 \)

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 \( z_1, z_2, \ldots, z_M \in \mathbb{C} \setminus \{0\} \) such that
\[ X(z) = \sum_{n=0}^{N-1} x_n z^{-n} = 0 \] for \( 1 \leq i \leq M \).

Factorization of the z-transform yields (cf. def. 1)
\[ X(z) = x_0 \prod_{m=1}^{M_i} (z - z_{i,m}) \prod_{m=1}^{M_m} (z - z_{m,0}) \]
\[ z^{(N-1)} \] (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, \( |z_i| < 1 \), from those outside, \( |z_{o,m}| > 1 \) (cf. def. 1).

2.3. Complex cepstrum based separation (CSM) 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
\[ \hat{x}_n = F^{-1} \left\{ \log_{e} \left[ F \{ (x_n) \} \right] \right\} \]
\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log_{e} |X(e^{i\omega})| e^{i\omega n} d\omega \]

Where \( F \{ (x_n) \} \) is \( X(e^{i\omega}) = \sum_{n=0}^{\infty} x_n e^{-i\omega n} \) 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 \( \log_{e}(z) = \log_{10}(z) + i \arg(z) \) 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].

\[ \hat{x}_n = \left\{ \begin{array}{ll} -\sum_{m=1}^{M_i} x_{i,m}/n, & n > 0, \\ \sum_{m=1}^{M_m} x_{o,m}/n, & n < 0. \end{array} \right. \] (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 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. 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]
\[ e_o(t) = E_o e^{c (t - \epsilon(t - t_c))}, \quad 0 \leq t \leq t_c \]
\[ e_c(t) = E_c \epsilon(t - \epsilon(t - t_c)) e^{-c(t - t_c)}, \quad t_c < t \leq t_e \]
\[ e_s(t) = 0, \quad t_e < t \leq T \]

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 \( e_s(t) \) is min phase [8]. Possible sequences with zeros on the UC are removed. The following LF-GFM parameter variations are allowed.

\[ t_0 = 0, \quad t_e = 0.01 \]
\[ t_p \in [0.23; 0.72] t_c, \quad t_e \in [0.30; 0.80] t_c \]
\[ t_a \in [0.0004; 0.2523] t_c, \quad E_e = 1 \]

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 experiment \( t_e = 0.01 \), 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. 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]**

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

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. \( a_0 \geq a_1 \geq \ldots \geq a_N \geq 0 \) all in \( \mathbb{R} \), let \( a_N = 1, a_{N-1-i} = a_{N-i} + r \) for \( i \in [0; N-1] \) and \( r \sim U[0, 1] \). The continuous uniform distribution is denoted by \( U \). 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 \( \pm 1 \) - are convolved to obtain mix phase sequences of lengths \( N \in [40; 539] \); 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\textsuperscript{®} function `roots()` estimates eigenvalues of a polynomial’s companion matrix in time \( \mathcal{O}(n^3) \) [11]. In the remainder, Matlab\textsuperscript{®} function names are set with typewriter typeface. CSM\textsubscript{fft} is based on FFT, IFFT, phase-unwrapping, and the real logarithm; \( \text{fft}, \text{ifft} \in \mathcal{O}(n \log n) \) [12] and \( \text{unwrap}, \text{log} \in \mathcal{O}(n) \). Combined, this yields an asymptotic time complexity for CSM\textsubscript{fft} of \( \mathcal{O}(n \log n) \) as FFT and IFFT dominates. Evidently, the time complexity of CSM\textsubscript{fft} is lower than ZSM.

Fig. 1 illustrates running times (measured with `tic/toc`) of ZSM and CSM\textsubscript{fft} as functions of input sequence length during two consecutive executions of ZSM and CSM\textsubscript{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\textsubscript{fft} core functions `fft(), ifft(), unwrap()?, 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\textsubscript{fft} which is lower than the complexity of ZSM.

Measuring memory usage is deceptive when memory management is handled automatically; in this experiment Matlab\textsuperscript{®} 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\textsubscript{fft} cannot be isolated. However, it is still interesting, from a practical point of view, to illustrate and exemplify the memory used by Matlab\textsuperscript{®} 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.
than CSMfft which to some degree support theory; especially
for one of the runs, the usage is constant for CSMfft . 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, CSMfac (cf. (2.2)), based on com-
plex 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 CSMfft have been compared the-
oretically 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
CSMfft outperforms ZSM in both time and space complex-
ity; 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 co-
efficient 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 CSMfft .

<table>
<thead>
<tr>
<th>Alg.</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZSM</td>
<td>No phase unwrapping</td>
<td>In time $O(n^3)$</td>
</tr>
<tr>
<td></td>
<td>No aliasing</td>
<td>In space $O(n^2)$</td>
</tr>
<tr>
<td></td>
<td>No zeros must be on UC</td>
<td></td>
</tr>
<tr>
<td>CSMfft</td>
<td>In time $O(n \log n)$</td>
<td>Phase unwrapping</td>
</tr>
<tr>
<td></td>
<td>In space $O(1)$</td>
<td>Aliasing</td>
</tr>
</tbody>
</table>

Table 1. Pros and cons of ZSM and CSMfft.

6. REFERENCES

[1] Gunnar Fant, Acoustic Theory of Speech Production,

[2] Bruce P. Bogert, M.J.R Healy, and John W. Tukey,
“The quefrency analysis of time series for echoes:
Cepstrum, pseudo-autocovariance, cross-cepstrum and
sapphire cracking,” Proc. of the Symposium on Time Series

orphic analysis of speech,” IEEE Transactions on Audio

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

Kemmerait, “The cepstrum: A guide to processing,” Proceed-

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

of the complex cepstrum by factorization of the z-
transform,” in Acoustics, Speech, and Signal Process-
ing, 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.

model of glottal flow,” STL-QPSR 26-4, Dept. of
Speech, Music, and Hearing, Royal Institute of Tech-


and Sven Treitel, “Factoring very-high-degree-
20, no. 6, pp. 27–42, November 2003.


place radix-2 FFT,” in International Conf. on Acoustics,
Speech, and Signal Processing, San Diego, California,