INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Automatic glottal inverse filtering with non-negative matrix factorization

Manu Airaksinen[1], *Lauri Juvela*[1], *Tom Bäckström*[2], *Paavo Alku*[1]

[1]Aalto University, Finland
[2]International Audio Laboratories Erlangen, Friedrich-Alexander University (FAU), Germany
manu.airaksinen@aalto.fi, paavo.alku@aalto.fi

Abstract

This study presents an automatic glottal inverse filtering (GIF) technique based on separating the effect of the glottal main excitation from the impulse response of the vocal tract. The proposed method is based on a non-negative matrix factorization (NMF) based decomposition of an ultra short-term spectrogram of the analyzed signal. Unlike other state-of-the-art GIF techniques, the proposed method does not require estimation of glottal closure instants.

The proposed method was objectively evaluated with two test sets of continuous synthetic speech created with a glottal vocoding analysis/synthesis procedure. When compared to a set of reference GIF methods, the proposed NMF technique shows improved estimation accuracy especially for male voices.
Index Terms: speech analysis, glottal inverse filtering, non-negative matrix factorization

1. Introduction

The glottal volume velocity waveform, or the *glottal flow*, is the main acoustical excitation in production of voiced speech. The study of glottal excitations is an important tool in many areas of speech research, such as in fundamental research of speech, medicine (e.g. occupational voice or speech pathology), phonetics (e.g. prosody), and neuroscience (e.g. brain responses evoked by speech). In addition, application of glottal excitation estimation has recently gained momentum in speech technology, especially in speech synthesis [1].

Glottal inverse filtering (GIF) is a computational method for estimating the glottal flow from a recorded microphone signal. This approach assumes the so-called *source-filter model* [2] of speech production, which is most commonly presented as a linear cascade of three processes: (1) a time-domain input that represents the glottal flow, (2) a digital filter representing the vocal tract transfer function, and (3) a differentiator that models the lip radiation effect (i.e. transform of flow at lips into pressure in free field). GIF is performed by blindly applying antiresonances to the recorded acoustic pressure signal so that the effects of the vocal tract and lip radiation are cancelled, ideally leaving the glottal flow intact. The practice is effective and non-invasive which is key for automated solutions.

Several digital GIF methods have been developed since the 1970's. (For further details, see reviews in [3, 4]). Some of the most well-known previous methods were recently compared with two novel GIF techniques proposed by the current authors, Quasi-closed phase analysis (QCP) [5] and Quadratic programming GIF (QPR) [6]. Our experiments in [5, 6] indicate that both QCP and QPR show very good accuracy in glottal flow estimation. Both of these techniques are based on the principle of the Closed Phase Covariance (CP) analysis, that is, computing the vocal tract auto-regressive (AR) model from excitation-free speech samples that are located in the closed phase of the glottal cycle. In QCP, this principle was developed further by using temporally weighted linear prediction (WLP) [7] as a vocal tract modelling technique. QCP uses a special type of a temporal weighting function, the attenuated main excitation (AME) waveform [8], which enables attenuating the contribution of samples located in the vicinity of the glottal main excitation in computation of the vocal tract model. QCP takes advantage of all the samples of the analysis frame in the computation of the vocal tract AR model instead of just those few that are located in a single closed phase as in conventional CP. The QPR method expanded this idea by proposing an approach in which the conventional CP based optimization is computed jointly with the AME-based optimization by using quadratic programming. Both QCP and QPR have shown improved estimation accuracy of the glottal flow particularly for high-pitched voices. Both of these new GIF-techniques, however, require extraction of the glottal closure instants (GCIs) which may be a source of error particularly when processing noisy or spontaneous speech.

In the present study, we propose a novel GIF technique based on non-negative matrix factorization (NMF). The technique, NMF-GIF, uses the principles of AME modeling in the computation of the vocal tract. Differently from conventional CP analysis, QCP, and QPR, however, no GCI extraction is needed in NMF-GIF thereby overcoming performance degradation caused by erroneous GCI estimates. This is achieved by performing a rank-2 NMF decomposition for an ultra short-term spectrogram consisting of ≈ 5 ms frames with 1 sample shifts. This decomposition is capable of separating those areas of the spectrogram that are greatly influenced by the glottal excitation from the areas less affected, thereby justifying the use of NMF in glottal inverse filtering.

The proposed method is based on the source-filter model and NMF processing, whose basics are reviewed in Section 2. The proposed approach, where NMF is applied on a short-term convolution matrix, is presented in Section 3. Our experimental evaluation in Section 4 shows that the proposed method improves GIF estimation accuracy especially for male voices.

2. Background

2.1. Source-filter model

The source-filter model of speech production is defined in the z-domain as

$$S(z) = G(z)V(z)L(z), \tag{1}$$

where $S(z)$ is the speech signal, $G(z)$ is the glottal excitation, $V(z)$ is the vocal tract transfer function, and $L(z)$ is the transfer function of the lip radiation effect. As the transfer function of the lip radiation effect is usually assumed to be known and of

http://dx.doi.org/10.21437/Interspeech.2016-338

the form

$$L(z) = 1 - \alpha z^{-1}, \tag{2}$$

where α is a constant within the range $[0.96, 1[$, GIF methods are left with the task of accurately estimating the vocal tract transfer function $V(z)$ to obtain an estimate of $G(z)$. This is also the principle approach taken in the proposed method.

2.2. Non-negative Matrix Factorization

Non-negative matrix factorization (NMF) [9] is a popular method for multivariate analysis of non-negative data, such as spectrograms [10], images [11], and text [12]. The task of NMF is to find, given a non-negative matrix $\mathbf{X} \in \mathbb{R}^{m \times N}$, two non-negative matrix factors $\mathbf{W} \in \mathbb{R}^{m \times k}$ and $\mathbf{H} \in \mathbb{R}^{k \times N}$ so that:

$$\mathbf{X} \approx \mathbf{WH}. \tag{3}$$

By denoting the length of an observation vector $\mathbf{x_i}$ by m, the number of observations by N, and the rank of the decomposition by k, $\mathbf{W}$ contains basis vectors as its columns. Moreover, each column of $\mathbf{X}$ can be represented as $\mathbf{x_i} \approx \mathbf{Wh_i}$, meaning that each column of $\mathbf{X}$ can be approximated as a linear combination of the columns of $\mathbf{W}$ weighted by the non-negative components of $\mathbf{H}$.

The most common way to optimize NMF is to minimize the Euclidean distance between $\mathbf{X}$ and $\mathbf{WH}$:

$$\min_{\mathbf{W},\mathbf{H}} ||\mathbf{X} - \mathbf{WH}||^2 \qquad \text{s.t. } \mathbf{W}, \mathbf{H} \geq 0 \tag{4}$$

Efficient algorithms based on multiplicative iterative updates on $\mathbf{W}$ and $\mathbf{H}$ were introduced in [13]. Further advancements in NMF optimization include, for example, conditions to encourage sparsity in $\mathbf{H}$ [14], convolutive NMF [15], convex NMF [16], and orthogonal NMF [17]. These recent advancements were considered during the development of the proposed GIF method, but they failed to yield any significant performance gains. Therefore, the conventional least squares NMF was selected as the framework for the remainder of the study. This choice is justified by the simplicity and efficiency of the least squares algorithm, and also by the method's close ties to K-means clustering [18].

3. Glottal Inverse Filtering with Non-negative Matrix Factorization

3.1. NMF Application to GIF

As discussed in Section 2.1, the source-filter model assumes that voiced speech is produced by convolving the glottal flow with the impulse responses of the vocal tract and lip radiation. The glottal flow and the lip radiation effect can be combined into the effective driving excitation (glottal flow derivative). Therefore, the model simplifies into an excitation (glottal flow derivative) and filter (vocal tract). During the glottal closed phase, the excitation is (close to) zero, and the resulting speech waveform corresponds mainly to the decaying response of the vocal tract. This phenomenon is the main idea behind the GIF techniques (e.g., [19], [5]) based on the conventional CP analysis. However, at instants of the main excitation of the vocal tract, which happen during glottal closing phases, the glottal excitation has a strong effect on the produced speech signal. This effect can be also seen in a spectrogram, if the spectral analysis is computed over a frame whose duration is less than one glottal cycle. Figure 1 (a) presents a spectrogram computed from a 30 ms frame of pre-emphasized speech. The length of the

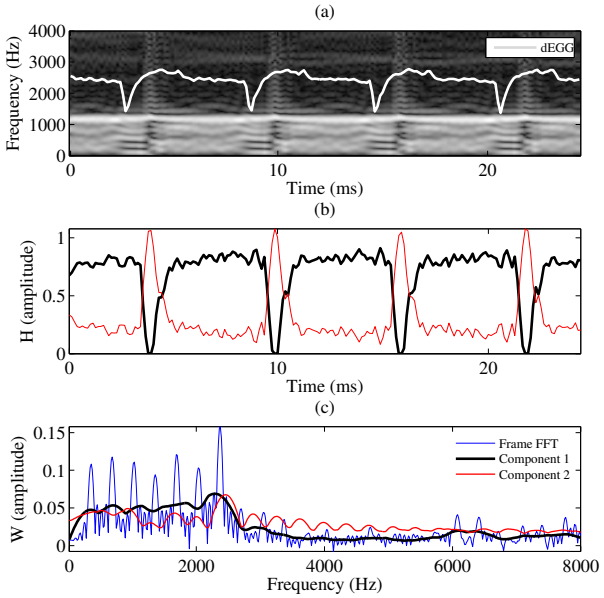

Figure 1: *(a) Spectrogram of a frame of pre-emphasized speech superposed with the corresponding differentiated electroglottography (dEGG) signal. DFT length is 45 samples (zero-padded to 1024), and hop size is 1 sample. (b) NMF activation functions ($\mathbf{H}$) computed from the given spectrogram. (c) Obtained NMF basis vectors ($\mathbf{W}$).*

DFT window (computed with 8-kHz sampling) was 45 samples (5.5 ms), and the hop size was 1 sample (0.125 ms). Superposed with the spectrogram is the corresponding differentiated electroglottography (EGG) signal. It can be seen from the spectrogram that in vicinity of GCIs, which correspond to the negative peaks of the differentiated EGG, the spectral properties of the speech signal are different from the spectra computed away from these peaks. The most distinct spectral features at GCIs are the more prominent high-frequency contents (corresponding to a smaller spectral tilt), and the appearance of the harmonic comb structure for the multiples of the fundamental frequency (f_0). Both of these properties are caused solely by the glottal excitation.

The strategy taken by the conventional CP analysis is to identify the GCIs and glottal opening instants (GOIs), and to compute a vocal tract model using a covariance criterion based linear prediction (LP) analysis from samples between one GCI and the next GOI. Though effective, this method defines an AR model from a small number of samples which makes the analysis sensitive to the accurate estimation of the GCIs and GOIs [19]. In the quasi-closed phase (QCP) method [5], robustness is improved as only the GCIs are needed and the AR model is defined from the data samples of the entire analysis frame using the AME weighting. Poor GCI estimates caused by non-ideal conditions, however, still cause problems for accuracy of QCP [20, 5].

To the best of our knowledge there are only few techniques that estimate the vocal tract transfer function over the closed glottal phase without explicitly detemining GCIs. One example of this kind of a technique is weighted linear prediction (WLP) [7], or the more recent stabilized weighted linear prediction (SWLP) [21], with the short-term energy (STE) weighting function [7].

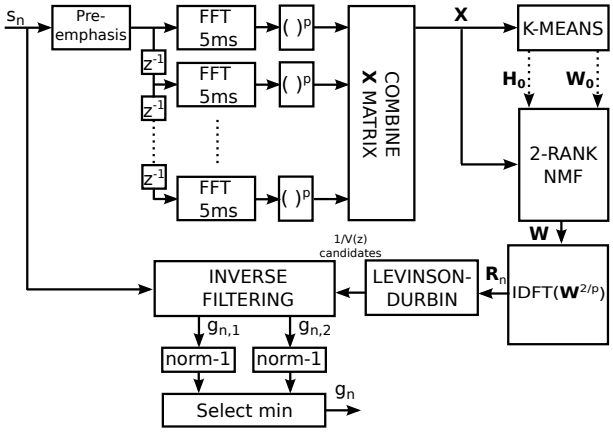

Figure 2: *Block diagram of the proposed NMF-GIF method.*

In the present study, we aim to factorize, using NMF, a spectrogram similar to that of Figure 1 (a) into two components: (1) the one that corresponds to the glottal main excitation and (2) the one outside the main excitation. In other words, we perform NMF with rank 2 on a magnitude spectogram with a short (≈5 ms) DFT length and 1 sample hop size, which yields two basis vectors on the matrix $\mathbf{W}$, and their activation functions on the matrix $\mathbf{H}$. The activation functions and basis vectors are depicted in Figures 1 (b) and (c), respectively. The basis vectors correspond to the average spectral envelopes during the glottal main excitation and outside the main excitation. As our goal is the estimation of the spectral envelope of the signal that is decoupled from the glottal main excitation, one of these basis vectors is exactly what we are looking for. It can be observed from Fig. 1 (b), that the activation function of "Component 2" corresponds remarkably well with the spectrum at GCIs (i.e. at instants of main glottal excitation) seen in the spectrogram, and the activation function of "Component 1" has orthogonal behavior to "Component 2".

3.2. Proposed method

As the main principles of applying NMF to GIF were presented in Section 3.1, the in-depth description of the implementation is presented next. The block diagram of the proposed method is shown in Figure 2.

First, a speech frame is pre-emphasized with a first-order differentiator ($H(z) = 1 - z^{-1}$) to roughly cancel out the spectral tilt of the glottal excitation [2]. The signal is then Fourier-transformed in ≈ 5 ms sub-frames which are shifted by 1 sample to obtain a series of magnitude spectra that are raised to the power of p to compress the spectrum similarly to a log function, while still maintaining non-negativity. This procedure has been observed to be useful e.g. in robust AR model estimation for speaker recognition [22]. In the present study, $p = 0.55$ is used based on informal test.

NMF is known to be a non-convex technique [13], which means that the optimization algorithms are only guaranteed to arrive at local minima. As a result of this, the correct initialization has a strong effect on the overall performance of the method. As our problem is essentially to cluster the given spectra into "with main excitation" and "without main excitation" components, we used K-means clustering [23] to initialize the vectors for $\mathbf{W}$ by selecting the obtained centroid vectors as the initial values. For $\mathbf{H}$, the activation functions were selected as the normalized Euclidean distances from the corresponding centroid vectors.

After the initialization, the NMF decomposition can be computed yielding the basis vectors and their activation functions. The power spectra of the basis vectors are raised next to $\frac{1}{p}$ to cancel the effect of compression and the resulting spectra are inverse Fourier transformed to obtain the corresponding autocorrelation sequences. The Levinson-Durbin algorithm is then applied to the m first autocorrelation coefficients to obtain the vocal tract inverse filter candidates $\mathbf{a}_1$ and $\mathbf{a}_2$.

The final task is to detect the inverse filter corresponding to the best glottal flow estimate. This is done by perfoming glottal inverse filtering according to Eq. 1, and normalizing the obtained *glottal flow* estimates between [0, 1]. The normalized estimate that yields the smallest norm-1 value is selected as the final estimate.

4. Experiments

4.1. Test setup

Evaluation of GIF methods is known to be problematic, because the reference glottal excitation cannot be acquired from real speech [3]. The most common way to circumvent this problem is to utilize synthetic speech where the excitation signal is known. Most commonly, the synthesized speech signals are produced as sustained vowels using a parametric (e.g. the Liljencrants-Fant (LF) model [24] with an all-pole vocal tract model) or a physical modelling based approach [5]. Test vowels synthesized with these techniques, however, are unrealistically stationary, which might add unknown bias to the evaluation experiments.

In the present study, we propose to use continuous synthetic speech based on the source-filter model of speech production (see Section 2.1). The synthetic continuous speech with a known glottal excitation is produced with a tweaked version of the GlottHMM vocoder [1], in which real speech samples are transformed with an analysis/synthesis process. The analysis/synthesis process is performed as follows. First, glottal vocoding analysis is performed on the given speech signal. This includes the estimation of feature vectors for each analysis frame that consist of the f_0, harmonic-to-noise ratio (HNR), and vocal tract spectral envelope. The spectral tilt parameters are omitted. Also, GlottHMM's original GIF method, Iterative Adaptive Inverse Filtering (IAIF), is replaced with a more straightforward pre-emphasized LP analysis to ensure that the features of the re-synthesized signals do not correspond to any of the compared methods' typical results. Second, the speech signal is resynthesized according to the feature vectors by first constructing the glottal excitation based on the anti-aliased LF model by Kawahara [25]. For simplicity, fixed LF parameters were used to produce a constant phonation type, but the f_0 and HNR were modified to match the vocoder parameters. The excitation waveform for the whole utterance was saved, and then filtered according to the vocal tract filter trajectory to produce the final speech waveform.

Two sets of test speech was used in the evaluation: Sustained real speech vowels of varying phonation types, and continuous normal speech. Both test sets were also divided according to the gender of the talker. The data used for sustained vowels included recordings of all eight Finnish vowels ([a], [e], [i], [o], [u], [y], [ae], and [oe]) from three male and four female speakers using a breathy, modal, or pressed phonation repeated three times. In total, the data contained 84053 test frames.

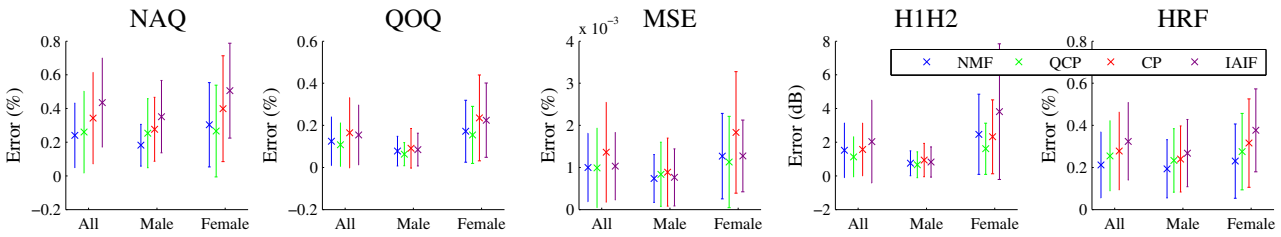

Figure 3: *Estimation error in five objective measures (see Section 4) for sustained vowels. Mean values denoted with 'x's, and the lines denote the standard deviation.*

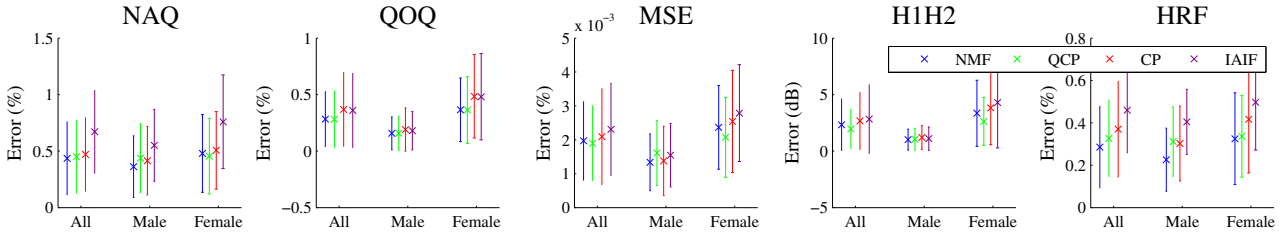

Figure 4: *Estimation error in five objective measures (see Section 4) for continuous speech. Mean values denoted with 'x's, and the lines denote the standard deviation.*

On the vocoded samples, the LF parameters were crudely adjusted to the target phonation type by selecting the typical LF parameters for each phonation type from [26]. The continuous speech dataset utilized 100 sentences from high-quality female ("Nancy" [27]) and male ("Nick" [28]) voices, resulting in a total of 103640 test frames. The LF parameters were set for modal phonation for all samples.

For each frame of the test sets, error between the estimated and reference glottal flows was computed using the following objective measures: Normalized amplitude quotient (NAQ) [29], quasi-open quotient (QOQ) [30], mean-squared error (MSE), H1-H2 [31], and harmonic richness factor (HRF) [32]. NAQ measures the relative length of the glottal closing phase, QOQ measures the approximate length of the glottal open quotient, and HRF and H1H2 are measures that are used to depict the spectral tilt of the glottal source waveform. The proposed method (denoted as "NMF") was compared to the following GIF methods: Quasi-closed phase analysis (QCP) [5], closed phase covariance analysis (CP) [19], and iterative adaptive inverse filtering (IAIF) [33]. Out of the compared methods, QCP and CP require GCI (and GOI for CP) estimation, whereas NMF and IAIF do not. The GCI and GOI estimation was performed with the SEDREAMS algorithm [34]. For the analysis, we used a sampling rate of 8 kHz with a 25 ms analysis frame length (with 5ms pre-frame buffer), and a vocal tract filter order $m = 10$. The CP method was implemented with the covariance criterion in LP analysis, by using two pitch-period analysis for frames with F0 $\geq$ 200Hz. For the QCP method we used the fixed AME parameters of PQ = 0.01, DQ = 0.7, and Nramp = 7 [5]. For IAIF we used the spectral tilt prediction order of $g = 4$ [33].

4.2. Results

The results for the tests described in Section 4.1 are presented in Figures 3 and 4 for the sustained and continous data sets, respectively. In most cases, the overall score of the NMF method is tied with the QCP method for the best score. The distinction between the methods is that the NMF method can be seen to perform better with male voices, whereas QCP is the best method for female voices that are known to have a higher f_0 than male voices. This can be explained by the use of the fixed-length ultra-short analysis window (5.5 ms), that for high f_0s contains over one glottal cycle of data. Lower frame durations were experimented with, but it was concluded that considerably smaller short-frame sizes had too little data for good results. Also, an important thing to note is that the GCI estimation-free NMF method clearly outperforms IAIF, the other comparable method in this regard.

5. Discussion

This study presented a novel approach to glottal inverse filtering with non-negative matrix factorization (NMF). The method is based on computing ultra short-term spectrograms of the speech signal, and then using the rank-2 NMF decomposition to acquire spectral envelope estimates corresponding to the glottal main excitation-free areas of the speech signal. The method can be applied robustly in a completely automatic manner that does not require external parameter estimations e.g. for the glottal closure instants.

The proposed method was evaluated on a vocoded real speech based dataset of sustained vowels and continuous speech where the LF-model based glottal excitation signal is known. The results indicate that the proposed method is on par with the state-of-the art methods, and even outperforms them for male speech. These properties contrast the proposed method from our previous work (Quasi-closed phase analysis (QCP) [5] and Quadratic programming GIF (QPR) [6]) in the sense that (1) the NMF-based method is GCI estimation free, and (2) QCP and QPR were shown to perform best for very high valued f_0s.

6. Acknowledgements

The research leading to these results has received funding from the Academy of Finland (project no. 256961, 284671).

7. References

[1] T. Raitio, A. Suni, J. Yamagishi, H. Pulakka, J. Nurminen, M. Vainio, and P. Alku, "HMM-based speech synthesis utilizing glottal inverse filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 153–165, 2011.

[2] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*, ser. Prentice-Hall signal processing series. Prentice-Hall, 1978.

[3] P. Alku, "Glottal inverse filtering analysis of human voice production – A review of estimation and parameterization methods of the glottal excitation and their applications," *Sadhana*, vol. 36, no. 5, pp. 623–650, 2011.

[4] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," *Computer Speech & Language*, vol. 28, no. 5, pp. 1117–1138, 2014.

[5] M. Airaksinen, T. Raitio, B. Story, and P. Alku, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 596–607, 2014.

[6] M. Airaksinen, T. Bäckström, and P. Alku, "Glottal inverse filtering based on quadratic programming," in *Proc. Interspeech*, 2015.

[7] C. Ma, Y. Kamp, and L. Willems, "Robust signal selection for linear prediction analysis of voiced speech," *Speech Communication*, vol. 12, no. 1, pp. 69 – 81, 1993.

[8] P. Alku, J. Pohjalainen, M. Vainio, A.-M. Laukkanen, and B. H. Story, "Formant frequency estimation of high-pitched vowels using weighted linear prediction," *Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1295–1313, 2013.

[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788 – 791, 1993.

[10] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

[11] I. Buciu, "Non-negative matrix factorization, a new tool for feature extraction: Theory and applications," *nt. J. of Computers, Communications & Control*, vol. 3, pp. 67 – 74, 2008.

[12] Y. Liu, R. Jin, and L. Yang, "Semi-supervised multi-label learning by constrained non-negative matrix factorization," in *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*, ser. AAAI'06. AAAI Press, 2006, pp. 421–426.

[13] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *In NIPS*. MIT Press, 2000, pp. 556–562.

[14] J. Eggert and E. Korner, "Sparse coding and nmf," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 4, 2004, pp. 2529–2533 vol.4.

[15] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *International Symposium on ICA and BSS*, 2004.

[16] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[17] S. Choi, "Algorithms for orthogonal nonnegative matrix factorization," in *IEEE International Joint Conference on Neural Networks*, 2008.

[18] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *SIAM International Conference on Data Mining*, 2005.

[19] D. Wong, J. Markel, and A. Gray Jr., "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 4, pp. 350 – 355, 1979.

[20] T. Raitio, A. Suni, J. Pohjalainen, M. Airaksinen, M. Vainio, and P. Alku, "Analysis and synthesis of shouted speech," in *Proc. Interspeech*, 2013.

[21] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilised weighted linear prediction," *Speech Communication*, vol. 51, no. 5, pp. 401 – 411, 2009.

[22] R. Saeidi, P. Alku, and T. Backstrom, "Feature extraction using power-law adjusted linear prediction with application to speaker recognition under severe vocal effort mismatch," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 1, pp. 42–53, 2016.

[23] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif., 1967, pp. 281–297.

[24] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, vol. 26, no. 4, pp. 1 – 13, 1985.

[25] H. Kawahara, K.-I. Sakakibara, H. Banno, M. Morise, T. Toda, and T. Irono, "Aliasing reduction in l-f model implementation for an interactive tool applicable to speech science education," *IEICE technical report*, vol. 115, no. 169, pp. 1–6, 2015.

[26] C. Gobl, "The voice source in speech communication - production and perception experiments involving inverse filtering and synthesis," Ph.D. dissertation, KTH Royal Institute of Technology, Speech Transmission and Music Acoustics, 2003.

[27] S. King and V. Karaiskos, "The blizzard challenge 2011," in *Proc. of Blizzard Challenge 2011*, 2011.

[28] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Hurricane natural speech corpus," 2013, LISTA Consortium. [Online]. Available: http://dx.doi.org/10.7488/ds/140

[29] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," *Journal of the Acoustical Society of America*, vol. 112, no. 2, pp. 701–710, 2002.

[30] T. Hacki, "Klassizierung von glottisdysfunktionen mit hilfe der elektroglottographie," *Folia phoniatrica*, vol. 41, no. 1, pp. 43 – 48, 1989.

[31] G. Fant, "The LF-model revisited. Transformations and frequency domain analysis," *STL-QPSR*, vol. 36, no. 2-3, pp. 119 – 156, 1995.

[32] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394 – 2410, 1991.

[33] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Communication*, vol. 11, no. 23, pp. 109 – 118, 1992.

[34] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit, "Detection of glottal closure instants from speech signals: A quantitative review," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 994 – 1006, 2012.