The Importance of Phase on Voice Quality Assessment

Maria Koutsogiannaki[1,2], *Olympia Simantiraki*[1,2], *Gilles Degottex*[1,2], *Yannis Stylianou*[1]

[1]Multimedia Informatics Lab, CSD, University of Crete, Greece
[2]Institute of Computer Science, FORTH, Crete, Greece

mkoutsog@csd.uoc.gr, simantir@csd.uoc.gr, degottex@csd.uoc.gr, yannis@csd.uoc.gr

10.21437/Interspeech.2014-391

Abstract

State of the art objective measures for quantifying voice quality mostly consider estimation of features extracted from the magnitude spectrum. Assuming that speech is obtained by exciting a minimum-phase (vocal tract filter) and a maximum-phase component (glottal source), the amplitude spectrum cannot capture the maximum phase characteristics. Since voice quality is connected to the glottal source, the extracted features should be linked with the maximum-phase component of speech. This work proposes a new metric based on the phase spectrum for characterizing the maximum-phase component of the glottal source. The proposed feature, the Phase Distortion Deviation, reveals the irregularities of the glottal pulses and therefore, can be used for detecting voice disorders. This is evaluated in a ranking problem of speakers with spasmodic dysphonia. Results show that the obtained ranking is highly correlated with the subjective ranking provided by doctors in terms of overall severity, tremor and jitter. The high correlation of the suggested feature with different metrics reveals its ability to capture voice irregularities and highlights the importance of the phase spectrum in voice quality assessment.

Index Terms: Voice quality, Phase Distortion, Glottal shape, Dysphonia

1. Introduction

Reliable voice quality evaluation involves the conduction of subjective listening tests, a task which is considered time consuming and costly. Thus, the necessity of developing objective measures highly correlated with the subjective evaluations is necessary.

Depending on the application, various objective measures have been proposed for speech quality evaluation. For example, in speech modeling the distortion metrics used are based on time-domain features (Signal-to-Error Reconstruction Ratio, Signal-to-Noise Ratio) or frequency domain features extracted from the amplitude spectrum (Spectral Distance, Cepstrum distance measures). For the quality assessment of natural speech, besides amplitude-spectrum-based features computed on the speech signal (i.e. Harmonic-to-Noise Ratio), many techniques focus on glottal features [1] and amplitude-spectrum based features computed on the glottal signal (HRF [2], H1-H2 [3]) as they are linked to speech quality [4, 5, 6]. In voice pathology the estimation of these features becomes more complex. The feature extraction in time or in frequency domain from a non-harmonic amplitude spectrum is a difficult problem for disordered voices [7], while the glottal source estimation is a rather complex and delicate problem [8].

This work proposes a different objective measure for voice quality assessment based on the Phase spectrum. Phase is systematically neglected by the minimum phase assumption in speech processing, even though previous studies have shown the link of the maximum-phase component of speech with the glottal source [9, 10, 11, 12] and its importance on maintaining the perceived quality of speech [13, 14, 15]. Reasons that contribute to the slight of the phase, is the phase wrapping due to the linear phase shift [13], which prevents the disclosure of the phase structure. However, in [13] the notion of the center of gravity has been introduced as an attempt to remove the linear phase mismatch which is attributed to the excitation phase and reveal the phase structure in speech, leading to its successful incorporation in various applications [16, 17]. Moreover, in [18] the phase difference between two frequency components has been shown to have perceived characteristics. Furthermore, in [19] the Phase Distortion (PD) is used to characterize the shape of periodic pulses of the glottal source independently of other source-filter characteristics, like the duration of glottal pulse, the position of analysis window and the influence of minimum phase component of speech. As the glottal shape is connected with voice quality, the phase distortion could then be used as quality assessment metric.

The main goal of this paper is to reveal the importance of the phase in voice quality assessment with application in voice pathology. Specifically, this work suggests a new phase representation which can automatically detect voice irregularities. The suggested methodology is based on Phase Distortion proposed in [18] but the estimation of the phase features is done by a harmonic model, thus giving to PD similar to the group-delay characteristics [20, 21, 22, 17]. In our work, the estimation of the instantaneous phases is performed by the adaptive Harmonic Model (aHM) [23]. Then, for revealing the phase structure of speech, the linear phase shift [13, 14] and the Phase Distortion (PD) [19, 24, 25, 18] are estimated on the signal after removing its minimum phase component. The PD alleviates the phase wrapping effect and, after the removal of the minimum-phase component, is also highly correlated to the maximum-phase component [19, 24]. Our proposed feature which describes voice disorders better than PD, is its standard deviation computed over time for each harmonic. PDD describes the phase variability of the voice source [26] which is more evident in pathological voices. The advantage of the proposed technique over other phase-based techniques [17] is that eliminates the necessity of reliable estimation of the glottal closure instants (GCI). PDD is evaluated in two databases consisting normophonic and dysphonic speakers with spasmodic dysphonia [27]. The database of the normophonic speakers is used as a learning database to derive an one-dimensional description of PDD, namely the Regularity Ratio (RR), which can quantify normophonicity. Then, the Regularity Ratio is used to objectively rank the severity of dysphonic speakers.

This paper is organized as follows. Section 2 presents the algorithm description of PDD. Section 3 evaluates PDD on nor-

mophonic and dysphonic speakers and compares the efficiency of PDD with another amplitude-spectrum-based metric [28]. Section 4 discusses the results and finally Section 5 concludes the paper.

2. Estimation of the Phase Distortion from a harmonic model

Analysis of the signal is performed, using the adaptive Harmonic Model [23], to extract the instantaneous amplitudes a_k and the instantaneous phases ϕ_k of each harmonic k from the speech waveform $s(t)$:

$$s^i(t) = \sum_{k=1}^{K^i} a_k^i \cdot e^{j(k\phi_0(t)+\phi_k^i)} \quad (1)$$

where i is the frame index, $K^i = \lfloor \frac{f_s}{2f_0(t^i)} \rfloor$, f_s the sampling frequency and $\phi_0(t)$ is the integral of $f_0(t)$:

$$\phi_0(t) = \int_{t^i}^{t} \omega_0(\tau)d\tau \qquad \omega_0(t) = f_0(t) \cdot 2\pi/f_s \quad (2)$$

In this work the fundamental frequency curve $f_0(t)$ is computed by STRAIGHT. From the estimated features $\{a_k, \phi_k, f_0\}$ we need to extract features that can describe the maximum-phase component of speech. To that purpose, we use a phase model similar to [29]:

$$\phi_k^i = \theta_k^i + k\int_{t_c^i}^{t^i} \omega_0(\tau)d\tau + \angle V^i(k\omega_0(t_i)) \quad (3)$$

where the extracted phases ϕ_k from the speech waveform in (1), are modeled as a summation of i) the phase of the glottal pulses θ_k, ii) the linear phase imposed by the delay of the center of the analysis window t_c and the position of the glottal pulse and iii) the minimum phase component which models the vocal tract influence, $\angle V(k\omega_0(t))$. Aiming on describing the source shape, the minimum phase component should be eliminated. To remove the influence of the vocal tract, the amplitude spectral envelope is first estimated through linear interpolation across frequency of the amplitude parameters a_k of the harmonic model in (1). Then, the minimum phase response corresponding to this amplitude spectral envelope is estimated through cepstrum liftering [30] and subtracted from the measured phase ϕ_k using subtraction in log-spectral domain (i.e. deconvolution in time domain). The phases $\widetilde{\phi_k}$ that derive from the subtraction of the minimum phase component from the estimated instantaneous phases ϕ_k are given by:

$$\widetilde{\phi_k^i} = \theta_k^i + k\int_{t_c^i}^{t^i} \omega_0(\tau)d\tau \quad (4)$$

The linear phase $k\int_{t_c^i}^{t^i} \omega_0(\tau)d\tau$ is still present, preventing the phase structure of the glottal pulse to appear. In [13, 14] the relative phase shift (RPS) has been suggested as the appropriate metric which discards this linear phase component. Using the definition of RPS and applying it on $\widetilde{\phi_k}$ the linear phase is discarded:

$$\begin{aligned}\widetilde{RPS}_k^i &= \widetilde{\phi}_k^i - k\widetilde{\phi}_0^i \quad (5)\\ &= \theta_k^i + k\int_{t_c^i}^{t_i} \omega_0(\tau)d\tau + -k\cdot(\theta_1^i + \int_{t_c^i}^{t_i} \omega_0(\tau)d\tau)\\ &= \theta_k^i - k\theta_1^i\end{aligned}$$

The obtained phase still depends on the harmonic index k. This dependency should be removed otherwise the variance of RPS will increase for the higher frequencies. To that purpose, the finite difference of RPS is used, namely the Phase Distortion (PD, [19]):

$$\begin{aligned}PD_k^i =& \underset{k}{\Delta}\widetilde{RPS}_k^i = (\widetilde{\phi}_{k+1}^i - (k+1)\widetilde{\phi}_1^i) - (\widetilde{\phi}_k^i - k\widetilde{\phi}_1^i) \quad (6)\\ =& \theta_{k+1}^i - \theta_k^i - \theta_1^i\end{aligned}$$

The advantage of using PD is that it is related to the shape of the glottal source without the need of accurate glottal source estimation.

2.1. Introducing the Deviation of the Phase Distortion

The relation of PD with the shape of the pulses of the glottal source suggest that PD can be a good quality indicator for voice signals. A first analysis on a normophonic and a dysphonic speaker enforces this argument. Figure 1 shows the estimation of PD on a sustained vowel /a/ uttered by a normophonic (Fig.1a) and a dysphonic (Fig.1b) speaker who suffers from spasmodic dysphonia [27]. The PD spectrum of the normophonic speaker appears stable, with stable values across harmonics and across time. On the contrary, in the case of the dysphonic speaker (Fig.1b), PD varies across time and harmonics. This suggests that the variance of the PD may be a better descriptive characteristic than PD for the voice regularity. In the context of sustained vowels, the variance of phase distortion is more interesting than the phase distortion itself: in sustained phonation the variance of the PD should be close to zero as the shape of the glottal pulse is preserved in time. Therefore, we propose a new metric, namely the Phase Distortion Deviation which is defined by the following equation taking into account the circular domain of the phase [31]:

$$\sigma_{PD_k}^i = \underset{i}{std}(PD_k^i) = \sqrt{-2\ln\left|\frac{1}{M}\sum_{m\in B_i} e^{jPD_k^m}\right|} \quad (7)$$

where B_i is a window centered at sample i and M is the size of the window B_i. The above equation computes the short-term standard deviation in time of the PD values estimated on the sliding window M for each harmonic separately. Specifically, the term $\frac{1}{M}\sum_{m\in B_i} e^{jPD_k^m}$ is the center of gravity [13] of the phase distortion values on the z-plane. If the PD values inside a window of size M are low, the modulus of the center of gravity approaches to 1, resulting on a low PDD. In the opposite case, if the PD values are significant and variable inside the window M, the modulus of the center of gravity will approach to zero, resulting in a high PDD. Therefore, PDD as define above can capture the variance of the PD.

Figure 2 shows the deviation of the PD depicted in Fig.1, for a fixed window of $100ms$ duration. Compared to Fig.1, Fig.2 is more informative as it shows the time characteristics of the PD for each harmonic. Unlike the dysphonic speaker, the normophonic speaker has almost zero PDD, revealing the correlation of PDD with the deviation of the glottal shape in time.

3. Evaluations

PDD seems to be an appropriate metric of the voice quality, as Fig. 2 suggests. To enforce this argument, the evaluation of the PDD is performed in a ranking task of disordered speech

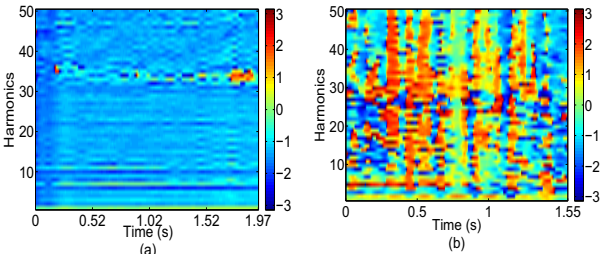

Figure 1: *The Phase Distortion (PD) on sustained phonation /a/: (a) normophonic male speaker ($\overline{f_0} = 125Hz$) and (b) dysphonic male speaker ($\overline{f_0} = 142Hz$)*

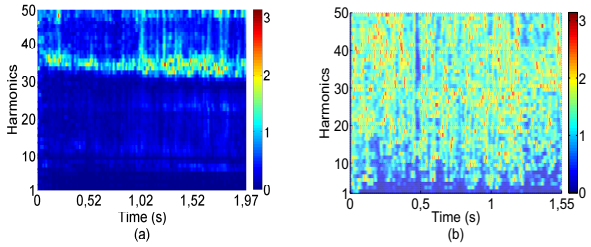

Figure 2: *The Phase Distortion Deviation (PDD) on sustained phonation /a/ using a fixed window of $100ms$: (a) normophonic male speaker ($\overline{f_0} = 125Hz$) and (b) dysphonic male speaker ($\overline{f_0} = 142Hz$)*

samples. Unlike binary classification, ranking is a more rigorous task. The advantage of ranking vs. classification is that it does not require a big dataset. The corpora used for our evaluation consist of two databases, one database of 16 normophonic subjects and one database of 20 speakers with spasmodic dysphonia. It should become explicit that no classification is made between normophonic and dysphonic speakers. However, the database of normophonic speakers is used to learn which are the PDD values that characterize normophonicity. Then, the objective ranking performed by PDD on the database of dysphonic speakers is compared with the subjective ranking performed by medical doctors[1] [27].

3.1. Objective PDD ranking method

Evaluation of PDD is performed on a database of sustained vowels uttered by speakers who suffer from spasmodic dysphonia [27]. Speakers with voice disorders are evaluated by doctors on sustained phonation, since the deviation of features in an entire phoneme can characterize voice pathologies. In order to capture such changes in the entire phoneme, the window length M in the calculation of PDD (Eq.(7)) is chosen to be the maximum possible. Small window would lead to an underestimation of the standard deviation of the PD. As the speech signals have variable length, the window length is proportional to the duration of the speech signal and specifically its length is chosen as the $1/3$ of the signal duration. This window length is the maximum possible, since there is an order restriction imposed by our zero-phase moving average filter that implements the summation in Eq.(7). An example of the estimated PDD can be seen in Fig.3 for normophonic and dysphonic speakers. In this representation the deviation of the PD for each harmonic is more prominent. The two dimensions of PDD create an informative visualization tool. However, it is difficult to manipulate a 2-D representation for ranking. Therefore, we propose an one-dimensional metric that describes the PDD feature. This feature is called Regularity Ratio and is based on the PDD distribution. Specifically, instead of using the PDD estimations of the whole spectrum, only the first 4 harmonics are considered to provide reliable estimations. Then, the distribution of PDD values in time for the 4 harmonics is estimated. Figure 4 shows the time-harmonic distribution of PDD (4 harmonics) computed in sustained vowels (/a/) of sixteen healthy subjects. The majority of the PDD values gather from [0, 0.4]. This area is defined as the normophonic area where the PD has low variance, while $(0.4, \infty)$ is characterized as the noisy area with high variance of PD. The Regularity Ratio (RR) is defined as:

$$RR = \log_{10}\left(\frac{P_1}{\sum_{i=2}^{\infty} P_i}\right), 1 \leq i < \infty \qquad (8)$$

where i is the bin index with width 0.4. P_1 is the probability of the first bin in the histogram which corresponds to interval $[0, 0.4]$ and the rest of bins cover the area $(0.4, \infty)$ of PDD values. The 0.4 value of the bin width is selected as the biggest possible value that gathers in the first bin the majority of PDD values for the normophonic speakers with a restriction imposed by the denominator of RR which should not be zero, giving infinite score of RR. For all the normophonic speakers of our database, RR is positive. In case of high variance of PD, we expect that RR will take negative values and will be able to rank patients according to their severity of spasmodic dysphonia. Indeed, Fig.4b depicts a typical distribution of PDD for a dysphonic speaker. PDD is much higher for the dysphonic speaker than in the normophonic case and the RR value is negative ($RR = -4.723$).

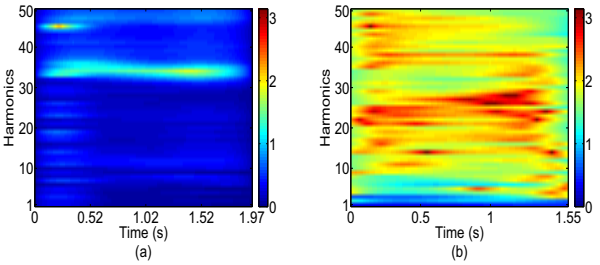

Figure 3: *The Phase Distortion Deviation (PDD) on sustained phonation /a/ estimated with the maximum possible window length: (a) normophonic male speaker ($\overline{f_0} = 125Hz$) and (b) dysphonic male speaker ($\overline{f_0} = 142Hz$)*

3.2. Performance evaluation

Evaluation of PDD is performed on a database of sustained vowels uttered by speakers who suffer from spasmodic dysphonia [27]. Subjective evaluations were performed by medical doctors who ranked the patients according to three features:

- Jitter: the cycle-to-cycle period perturbation. It is evaluated objectively by [27] using Ampex software [32] and the speech samples are ranked from high to low jitter.
- Tremor: rhythmic change in pitch and loudness. Tremor is subjectively estimated by medical doctors and the

[1] In the paper, medical doctors refer to doctors specialized on quantifying the quality of voice i.e., otorhinolaryngologists trained for judging voice quality.

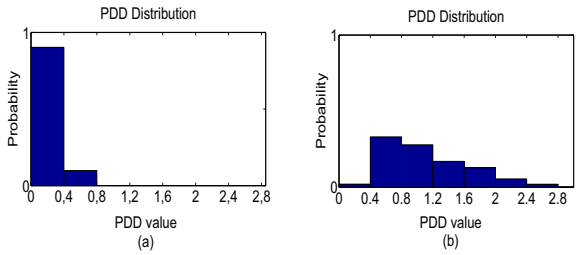

Figure 4: *The distribution of Phase Distortion Deviation (PDD) on sustained phonation /a/: (a) 16 normophonic speakers and (b) a typical distribution in case of spasmodic dysphonia.*

speech samples are ranked from high to low tremor value.

- Overall severity of spasmodic dysphonia of the speech samples in a descending order as evaluated by the doctors

For each patient, RR is estimated on sustained vowels /a/. The goal is to propose an objective ranking of the speech samples of the patients using RR and examine if our proposed objective measure is correlated with the three features described above. Moreover, RR is compared with another objective metric, called WMTV [28], which quantifies the severity of spasmodic dysphonia. The metric is based on tremor estimation on the speech signals and is extracted from minimum phase component of speech. Table 1 shows the correlation between

	Jitter (Ampex)		Tremor		Overall Severity	
	S	P	S	P	S	P
RR	-0.65 (0.0082)	-0.78 (0.0006)	-0.68 (0.005)	-0.70 (0.0037)	-0.81 (0.003)	-0.82 (0.0002)
WMTV	0.50 (0.0585)	0.44 (0.0991)	0.75 (0.0012)	0.72 (0.0024)	0.67 (0.0066)	0.68 (0.0053)

Table 1: *Pearson's (P) and Spearman's (S) correlation coefficient of the ranking of RR and WMTV with the ranking of subjective evaluations (Tremor, Overall severity) and objective evaluations (Jitter) provided by [27]. 15 speakers are ranked.*

RR and WMTV with the subjective evaluations provided by the medical doctors, except the Jitter estimation which is computed objectively by Ampex software, as reported above. For comparison reasons with WMTV, 15 out of the 20 dysphonic speakers are ranked as some speech samples had less than $1s$ duration and WMTV imposes the restriction of at least $1s$ duration of the speech signal in order to extract the tremor characteristics. Two correlation coefficients are used to evaluate our proposed metric, Pearson's correlation coefficient (P acronym on the Table 1) and Spearman's rank correlation coefficient (S acronym on the Table 1). The advantage of the latter is that it captures the monotonic relation between variables, even if their relationship is not linear. As we can see from Table 1, the ranking based on our proposed objective measure RR is significantly correlated with the subjective ranking of Tremor and Overall severity of spasmodic dysphonia (p values are provide in parenthesis). The objective measure based on PDD outperforms WMTV by 14% ($r_{S,RR} = -0.82$ vs $r_{S,WMTV} = 0.68$) on the estimation of the overall severity of spasmodic dysphonia. Moreover, RR is significantly correlated with jitter whereas WMTV is not, revealing that PDD can detect various features of voice irregularity. Another merit of RR vs. WMTV is that it can be estimated in speech signals independent of duration. Therefore, Table 2 shows the correlation between RR with the subjective evaluations provided by medical doctors for all 20 patients. The correlation between our ranking and subjective ranking on the overall severity of spasmodic dysphonia is significant; the correlation coefficient is -0.82 and the p-value is less than 0.0001.

	Jitter (Ampex)		Tremor		Overall Severity	
	S	P	S	P	S	P
RR	-0.62 (0.0033)	-0.71 (0.0004)	-0.59 (0.0067)	-0.64 (0.0023)	-0.82 (<0.0001)	-0.82 (<0.0001)

Table 2: *Pearson's (P) and Spearman's (S) correlation coefficient of the ranking of RR with the ranking of subjective evaluations (Tremor, Overall severity) and objective evaluations (Jitter) provided by [27]. All speakers are ranked.*

4. Discussion

This work emphasizes the importance of the phase spectrum as an objective measure for voice quality assessment. The phase information has not received much attention as a quality indicator in the literature. The linear phase influence due to misalignments of the glottal closure instants with the position of the window analysis, contributed to the noisinesses and incoherence of the phase spectrum. Removing these linear phase terms [13] and the phase contribution of the vocal tract, the remaining phase spectrum can provide useful information about the regularity of the glottal signal. The idea of the Phase Distortion proposed in [19] is combined with an adaptive Harmonic Model [23] to derive our propose scheme. Analysis on normophonic and dysphonic speakers revealed that the time deviation of PD is an efficient descriptor for voice pathologies. Even though no classification is performed between normophonic and dysphonic speakers, the evaluation of our proposed metric is performed on a more difficult and by more useful task, that of the objective ranking among speakers of the same category, that is dysphonic speakers who suffer from spasmodic dysphonia. The high correlations with the subjective evaluations indicate that PDD not only captures but also quantifies the noisy and harmonic part of speech, suggesting that PDD may be extended on other applications like speech synthesis.

5. Conclusions

In this paper the information of the phase spectrum is used to describe the level of voice pathology in speakers with spasmodic dysphonia. The proposed phase representation is the time deviation of the Phase Distortion (PDD). PDD is free from the linear phase influence and the phase contributions of the vocal tract and therefore, can characterize the regularity of the glottal source. The advantage of the proposed technique is that eliminates the necessity of detection of the GCI or reliable estimation of the glottal source through inverse filtering and can be used for voice detection irregularities. PDD is evaluated in a database of dysphonic speakers with spasmodic dysphonia. The objective ranking performed by PDD is highly correlated with the subjective ranking from medical doctors.

6. Acknowledgments

The authors would like to thank Pr. Dejonckere from Utrecht University for providing us the database and the subjective evaluations of the dysphonic speakers.

7. References

[1] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Amer.*, vol. 49, pp. 583–590, 1971.

[2] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis synthesis, and perception," *J. Acoust. Soc. Amer.*, vol. 90, no. 5, pp. 2394–2410, 1991.

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

[4] D. Klatt and L. Klatt, "Analysis, synthesis and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Amer.*, vol. 87, pp. 820–857, 1990.

[5] J. Kane and C.Gobl, "Evaluation of glottal source closure instant detection in a range of voice qualities," *Speech Commun*, vol. 55, no. 2, pp. 295–314, 2013.

[6] J. Kreiman, Y.-L. Shue, G. Chen, M. Iseli, B. Gerratt, J. Neubauer, and A. Alwan, "Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation," *J. Acoust. Soc. Amer.*, vol. 132, pp. 2625–2632, 2012.

[7] K. Shama, A. Krishna, and N. N. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology," *EURASIP Journal on Advances in Signal Processing*, 2007.

[8] 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.

[9] A. Oppenheim, G. Kopec, and J. Tribolet, "Signal analysis by homomorphic prediction," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 4, pp. 327–332, 1976.

[10] B. Doval, C. Alessandro, and N. Henrich, "The voice source as a causal/anticausal linear filter," *VOQUAL, Geneva*, 2003.

[11] B. Bozkurt, B. Doval, C. Alessandro, and T. Dutoit, "Zeros of z-transform representation with application to source-filter separation on speech," *IEEE signal processing letters*, vol. 12, no. 4, 2005.

[12] T. Drugman, B. Bozkurt, and T. Dutoit, "Complex cepstrum based decomposition of speech for glottal source estimation." *Interspeech, Brighton*, 2009.

[13] Y. Stylianou, "Removing linear phase mismatches in concatenative speech synthesis," *IEEE Trans. Speech Audio Process*, vol. 9, no. 3, pp. 232–239, 2001.

[14] I. Saratxaga, I. Hernaez, M. Pucher, and I. Sainz, "Perceptual importance of the phase related information in speech," *Proc. Interspeech. ISCA*, 2012.

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

[16] P. D. Leon, M. Pucher, J. Yamagishi, I. Hernaez, and I. Saratxaga, "Evaluation of speaker verification security and detection of HMM-based synthetic speech," *Audio, Speech and Language Pocessing, IEEE transactions*, vol. 20, no. 8, pp. 2280–2290, 2012.

[17] T. Drugman, T. Dubuisson, and T. Dutoit, "Phase-based information for voice pathology detection," *ICASSP*, pp. 4612–4615, 2011.

[18] S. P. Lipshitz, M. Pocock, and J. Vanderkooy, "On the Audibility of Midrange Phase Distortion in Audio Systems," *J. Audio Eng. Soc*, vol. 30, no. 9, pp. 580–595, 1982.

[19] G. Degottex, A. Roebel, and X. Rodet, "Function of phase-distortion for glottal model estimation," *in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4608–4611, 2011.

[20] H. Banno, K. Takeda, and F. Itakura, "The effect of group delay spectrum on timbre," *Acoustical Science and Technology*, vol. 23, no. 1, pp. 1–9, 2002.

[21] H. A. Murthy and B. Yegnanarayana, "Speech processing using group delay functions," *Elsevier Signal Processing*, vol. 22, pp. 259–267, 1991.

[22] R. Smits and B. Yegnanarayana, "Determination of instants of significant excitation in speech using group delay function," *IEEE Trans. SpeechAudio Processing*, vol. 3, pp. 325–333, Sep 1995.

[23] G. Degottex and Y. Stylianou, "Analysis and synthesis of speech using an adaptive full-band harmonic model," *IEEE Trans. on AudioSpeech and Lang. Proc.*, vol. 21, no. 10, pp. 2085–2095, 2013.

[24] G. Degottex, A. Roebel, and X. Rodet, "Phase minimization for glottal model estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1080–1090, 2011.

[25] M. Tahon, G. Degottex, and L. Devillers, "Usual voice quality features and glottal features for emotional valence detection," *Proc. International Conference on Speech Prosody*, pp. 693–696, 2012.

[26] A. Fourcin and M. Ptok, "Closing and opening phase variability in dysphonia," 2003.

[27] P. Dejonckere and C. Manfredi, "Long-term follow-up of patients with spasmodic dysphonia repeatedly treated with botulinum toxin injections," *International Journal of Phonosurgery and Laryngology*, vol. 1, no. 2, pp. 57–60, July-December 2011.

[28] M. Koutsogiannaki, Y. Pantazis, Y. Stylianou, and P. Dejonckere, "Tremor in speakers with spasmodic dysphonia," *MAVEBA*, 2011.

[29] Y. Agiomyrgiannakis and Y. Stylianou, "Wrapped gaussian mixture models for modeling and high-rate quantization of the phase data of speech," *IEEE Trans. on Audio, Speech and Lang. Proc.*, vol. 17, no. 4, pp. 775–786, 2009.

[30] A. V. Oppenheim and R. Schafer, *Digital Signal Processing*, 2nd ed. Prentice-Hall, 1978.

[31] N. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, Oct. 1995.

[32] L. V. Immerseel and J. Martens, "Pitch and voiced/unvoiced determination with an auditory model," *J. Acoust. Soc. Amer.*, vol. 90, no. 5, pp. 2394–2410, 1991.