Investigating voice function characteristics of Greek speakers with hearing loss using automatic glottal source feature extraction

Anna Sfakianaki, George P. Kafentzis

Speech Signal Processing Laboratory, Department of Computer Science, University of Crete, Greece
{asfakianaki, kafentz}@csd.uoc.gr

Abstract

The current study investigates voice quality characteristics of Greek adults with normal hearing and hearing loss, automatically obtained from glottal inverse filtering analysis using the Aalto Aparat toolkit. Aalto Aparat has been employed in glottal flow analysis of disordered speech, but to the best of the authors' knowledge, not as yet in hearing impaired voice analysis and assessment. Five speakers, three women and two men, with normal hearing (NH) and five speakers with prelingual profound hearing impairment (HI), matched for age and sex, produced symmetrical /'pVpV/ disyllables, where V=/i, a, u/. A state-of-the-art method named quasi-closed phase analysis (QCP) is offered in Aparat and it is used to estimate the glottal source signal. Glottal source features were obtained using time- and frequency-domain parametrization methods and analysed statistically. The interpretation of the results attempts to shed light on potential differences between HI and NH phonation strategies, while advantages and limitations of inverse filtering methods in HI voice assessment are discussed.

Index Terms: hearing loss, voice assessment, glottal inverse filtering, Greek

1. Introduction

Hearing loss, especially when occurring prelingually, can have detrimentral effects on various speech production parameters, such as articulation, respiration and phonation [1, 2]. Inaccurate interarticulatory coordination, resulting from glottal airflow mismanagement as well as problematic vocal fold movement and velopharyngeal valving, can lead to faulty segmental and suprasegmental production [3]. Inappropriate pausing at linguistic boundaries and decreased syllable production per breath unit as well as inefficient vocal fold vibration patterns, abduction/adduction gestures and larynx control [1] have been reported to cause pitch and loudness issues, excess breathiness, strain, roughness and vocal fatigue to speakers with hearing impairment (henceforth HI) [4, 5, 6, 7, 8].

Vocal function can be examined using various instrumental methods. Electrolaryngography (ELG) and electroglottography (EGG) are non-invasive techniques commonly used for vocal fold vibration monitoring and voice quality assessment [9, 10]. Two gold plated electrodes consisting of an inner disk surrounded by an outer guard-ring, are placed on either side of the thyroid cartilage and held in position by an elastic neckband. Electrical conductance between the electrodes is measured so as to examine vocal fold vibration. Besides ELG and EGG signal analysis, vocal fold movement can also be observed via laryngeal endoscopic imaging, such as videoendoscopy and videostroboscopy [11]. These methods involve the insertion of a long tube in the speaker's throat in order to visualise vocal fold activity. Although they are considered successful at providing objective documentation and assessment of vocal fold behaviour [12], they are both invasive and expensive. Alternatively, measurements directly from the glottal airflow velocity signal of recorded speech can be made using glottal inverse filtering (GIF). GIF is heavily based on the source-filter paradigm introduced by Fant [13], where speech can be considered as the outcome of a linear filtering operation, with the source signal being the glottal excitation signal and the filter being the vocal tract. GIF introduces the idea of inversion according to which, the effects of vocal tract and lip radiation are cancelled from the speech signal [14]. Thus, by analyzing the speech signal we can estimate the glottal excitation. The usefulness of GIF in pathological speech analysis has been demonstrated in the literature [15, 16, 17, 18, 19]. However, GIF analysis is not a trivial task to perform from scratch and it is not included in most commercial or freely available speech analysis software for immediate assessment of voice quality.

2. Related Work and Aims of the Study

The assessment of glottal aerodynamics of speakers with HI can provide useful information about vocal fold movement and glottal airflow during speech [20]. Such information should contain suitable measures for detection of HI voice deviations and measures for examination of differences in vocal adjustments of speakers with HI as compared with those of speakers with normal hearing (henceforth NH) [21, 22, 23, 24]). Several glottal characteristics have been associated with HI voice disorders. For example, variations in F0 and its amplitude indicate breathiness, roughness, and hoarseness [24, 25, 26] while close-to-open phase ratio and steepness of glottal closure has been associated with breathiness [7]. Furthermore, the extent of vocal fold abduction and glottal efficiency have been related to reduced HI vocal fold mobility and oscillatory efficiency [23]. However, these findings have been obtained using invasive methods and/or specialized, expensive equipment.

Although there are few studies on HI voice characteristics using such instrumental methods [23](Furcin, 2000), to the best of our knowledge, there are no studies on HI glottal source features extracted directly from the acoustic signal via inverse filtering using Aalto Aparat or any other related software. Hence, the present paper aims at examining automatically extracted voice features of Greek speakers with NH and with prelingual profound HI using the GIF program of Aalto Aparat [27] and discussing the results in relation to existing literature. The clinical value of inverse filtering in atypical voice research as well as the advantages and limitations of the application of freely available tools and algorithms in HI voice assessment will also be discussed.

The rest of the paper is structured as follows: Section 3 presents the dataset used and the analysis part, while Section 4 describes the statistical analyses and discusses the results. Finally, Section 5 concludes the work and suggests future research directions.

3. Methodology

3.1. Dataset

A small dataset was selected from a corpus recorded in order to examine the articulation of Greek speakers with NH and HI [28]. The dataset includes recordings of symmetrical /'pVpV/ disyllables with the corner vowels /i, a, u/ with stress on the first syllable in the carrier phrase "'Lejje ... 'pali" ("Say ... again"). Each disylalble was produced 10 times by five speakers with NH, three women and two men, and five speakers with HI, matched for age and gender. All participants were $18 - 35$ years old and native speakers of Greek. Speakers with HI had prelingual, profound (average > 90 dB HL at $500, 1000$ and 2000 Hz) hearing loss which was diagnosed before the age of 2. They had all been fitted with hearing aids by the age of 3 and had received more than 4 years of speech therapy at the time of the recording. Speech was recorded at 22050 Hz and downsampled to 8000 Hz for the analysis.

3.2. Analysis

The analysis was conducted on the stressed first-syllable vowels /i, a, u/ of every repetition. Thus a total number of 300 items was analysed. The first 6 glottal cycles of the first syllable vowel in the sentence were used for feature extraction. In cases where the first 6 cycles were problematic (e.g. uneven, incomplete, or missing altogether), 6 cycles from a more steady-state portion of the vowel were manually chosen for analysis. Such issues sometimes arose mainly for speakers with HI. Hence, 1800 measurements were conducted in total (6 cycles x 3 disyllables x 10 repetitions x 10 speakers) Instead of using standard GIF methods, we decided to employ Quasi-Closed Phase (QCP) analysis [29] for voice source estimation, as provided by Aalto Aparat. Aalto Aparat is a voice source analysis toolkit developed at Aalto University. QCP is a method inspired from closed phase (CP) analysis, that is the estimation of the vocal tract during the closed phase of the glottis. This is an important task since vocal tract estimation during closed phase is free from nonlinear source-filter interactions. However, direct estimation of the glottal closed phase is problematic [14]. Compared to CP-based methods, the proposed technique does not utilize the covariance method of linear prediction to estimate the vocal tract filter but takes advantage of weighted linear prediction (WLP) in order to exploit all the samples of an analysis frame of successive pitch periods, emphasizing on the samples which are located in the closed phase. The default QCP parameter values as provided by Aparat have been selected for our purpose. An example of GIF application on vowel /a/ of a speaker with NH and a speaker with HI is illustrated in Figure 1.

After voice source estimation using QCP, we extracted four time domain parameters and three frequency domain parameters using Aparat. Time domain parameters include the normalized amplitude quotient (NAQ), the closing quotient (CQ - ClQ in Aparat), and the speed quotient (SQ - SQ1 in Aparat), whereas the frequency domain ones are the harmonic richness factor (HRF), the difference between spectral amplitudes of the fundamental and the second harmonic (H1-H2), and the parabolic spectral parameter (PSP). A review of all these parameters used in glottal inverse filtering can be found in [14].

Specifically, NAQ is a parameter that describes the glottal closing phase. Amplitude quotient (AQ) is defined by the ratio of the maximum of the glottal flow over the minimum of its derivative, thus NAQ is AQ normalized with respect to the period of the waveform. In [30], the authors demonstrate its robustness and efficiency to discriminate between different phonation types. Another very widely used parameter, the CQ, measures the ratio of the duration of the closing phase to the glottal cycle and reflects the abruptness of vocal fold closure. Finally, SQ measures the ratio of the opening phase duration over the closing phase duration and has been associated with glottal efficiency and vocal tract mobility.

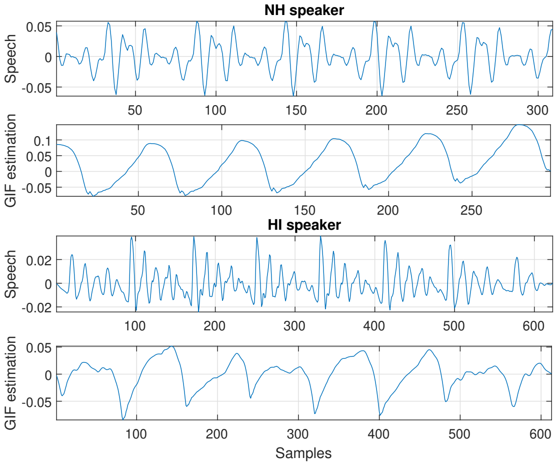

Figure 1: *Glottal source estimation based on the analysis of the stressed vowel /a/ in the disyllable /'papa/ produced by a speaker with NH and a speaker with HI. First panel: speech signal (NH), second panel: glottal source signal (NH), third panel: speech signal (HI), fourth panel: glottal source signal (HI).*

Regarding frequency domain parameters, the H1-H2 difference (H1-H2) is defined as the difference between the fundamental frequency log-amplitude and the log-amplitude of the second harmonic of the glottal source. H1-H2 is a very well known and widely used measure of voice quality characterization and is considered to be a rough measure of the spectral decay. The Harmonic Richness Factor (HRF) quantifies the amount of harmonics in the spectrum of the glottal source signal. HRF is defined as the ratio of the sum of the amplitudes of harmonics over the amplitude of the fundamental frequency. Usefulness in voice quality characterization has been demonstrated in [31, 32]. The PSP fits a second-order polynomial to the glottal source log-spectrum over a single glottal cycle. PSP has been associated with phonation type.

Moreover, driven by the fact that it has extensively been used in the literature [33, 34], F0 information has been extracted using SWIPE pitch estimator [35].

4. Results and Discussion

Statistical analyses of variance (ANOVA) were run for the three time-based (CQ, NAQ, SQ), the three frequency-based parameters (H1H2, HRF, PSP) obtained via Aparat and the mean F0 values obtained via SWIPE, of the vowels /a, i, u/, vs the factors gender and hearing status. Although the time-based parameter OQ was calculated as well, it was excluded from the statistical analyses, as the automatically obtained values displayed a lot of variability. A great deal of fine tuning GIF parameters is required in order to receive interpretable results regarding

V	G	HS	NAQ	CQ	SQ	H1H2	HRF	PSP
/a/	M	NH	0.13 (0.004)	0.34 (0.01)	1.77 (0.1)	12.99 (1.33)	-1.81 (0.81)	0.27 (0.012)
		HI	0.12 (0.003)	0.25 (0.01)	*2.86 (0.09)	*7.49 (1.33)	-1.00 (0.81)	0.24 (0.012)
	F	NH	0.12 (0.002)	0.28 (0.009)	2.38 (0.07)	12.16 (1.09)	-1.03 (0.66)	0.19 (0.008)
		HI	*0.14 (0.002)	0.34 (0.009)	*1.96 (0.07)	8.75 (1.09)	-0.16 (0.66)	*0.23 (0.008)
/i/	M	NH	0.15 (0.003)	0.40 (0.008)	1.42 (0.05)	12.85 (1.17)	-5.10 (1.02)	0.41 (0.01)
		HI	0.15 (0.003)	*0.29 (0.008)	*2.21 (0.05)	9.05 (1.17)	-2.04 (1.02)	*0.33 (0.01)
	F	NH	0.14 (0.003)	0.31 (0.006)	2.23 (0.04)	9.39 (0.96)	-1.19 (0.83)	0.28 (0.009)
		HI	*0.19 (0.003)	*0.41 (0.006)	*1.40 (0.04)	11.25 (0.96)	-2.52 (0.83)	0.25 (0.009)
/u/	M	NH	0.17 (0.004)	0.38 (0.009)	1.62 (0.06)	15.03 (1.49)	-5.98 (1.07)	0.31 (0.012)
		HI	*0.15 (0.005)	*0.32 (0.01)	*2.07 (0.07)	*6.46 (1.49)	*1.09 (1.07)	*0.43 (0.018)
	F	NH	0.17 (0.004)	0.36 (0.008)	1.79 (0.05)	10.38 (1.22)	-3.40 (0.88)	0.29 (0.014)
		HI	*0.15 (0.003)	*0.30 (0.007)	*2.28 (0.05)	9.79 (1.22)	-0.73 (0.88)	0.30 (0.012)

Table 1: *Mean and SE values of all parameters in vowels (V) /i,a,u/ with statistical comparisons within gender (G: M for Male and F for female) and between hearing status (HS). Statistically significant differences between speakers with NH and HI of the same gender are denoted with an asterisk (*$p < .05$*) before the HI mean value.*

OQ. The statistical analyses of the seven parameters showed that gender was significant for CQ, PSP and F0 in all vowels, and additionally for SQ in /u/ and NAQ in /i/, thus Tukey post-hoc tests between the two genders within hearing group were conducted. Besides the expected lower F0 values for male vs female speakers with NH, additional gender differences were found, such as higher CQ and PSP values and lower SQ values in /a/ and /i/, and lower HRF values in /i/. Gender differences within speakers with HI were also observed in all aforementioned parameters. Hence, comparisons between speakers with NH vs HI were subsequently conducted separately for each gender. Hearing status was not found statistically significant for the F0 parameter. F0 in HI speech is reported deviant in some studies and within normal range in others (**??** for a review). Instead of F0 mean values, F0 variance or other measures might be more promising as features (Bolfan, 2007). The hearing status factor was found significant for all parameters in vowel /u/, while in vowels /a/ and /i/ hearing status was significant for SQ, NAQ, H1H2, and for NAQ, PSP correspondingly. Table 1 summarises the mean values and standard error (SE) of all parameters according to gender group and hearing status in each vowel. Statistical comparisons were conducted between speakers with NH and HI of the same gender.

Regarding the time domain, significant differences were located between speakers with NH and HI in all three parameters. CQ reflects the abruptness of vocal fold closure [36]. This parameter was found lower in male speakers with HI denoting more abrupt vocal fold closure than normal. In addition, CQ reflects changes in glottal source due to intensity and phonation type. Thus, lower CQ in HI male speakers may either indicate higher intensity or more pressed phonation than normal. Female speakers with HI seem to assume a more gradual vocal fold closure than normal at least for vowels /i/ and /a/. NAQ as a measure is highly correlated with CQ, although shown to be more robust [14]. In our data, the two parameters indeed follow similar trends. In reference to SQ, the picture is variable regarding gender and vowel. Male speakers with HI show significantly higher SQ in all vowels, as do female speakers in vowel /u/, while in vowels /i/ and /a/ female speakers with HI show significantly lower SQ than normal. SQ has been associated with glottal efficiency and vocal fold mobility [23]. A lower value may indicate a less gradual opening phase or a more abrupt closing phase than normal. On the other hand, higher values than normal might suggest a more abrupt opening or less precipitous closing of the vocal folds. Although male and female HI trends are opposite, and vowel type seems to also play a role, as also reported in the literature [37], both male and female HI patterns are significantly different than normal, indicating differential skewness patterns for speakers with HI. Tenseness or stress have also been reported to influence skewness, with stressed or more intense vowels being more symmetrical and thus presenting lower SQ values [20] (or, inversely, higher SQ in EGG or photoglottographic studies).

The majority of frequency-based differences between NH and HI speech were located in male speakers, as shown in Table 1. Frequency domain measures reflect changes in phonation type and vocal quality of the speaker. HIH2 is a rough measure of spectral decay. High values denote steeper decay, towards breathy phonation, whereas low values indicate gradual decay towards pressed phonation [22]. H1H2 is also strongly correlated with OQ, which reflects the extent of vocal fold abduction. In our data, male speakers with HI display significantly lower H1H2 values for vowels /a/ and /u/ than normal. No significant differences were located for female speakers in H1H2 for any vowel. HRF depicts the amplitude relationships of higher harmonics to the F0 amplitude [31]. This parameter was found higher than normal for male speakers with HI only in vowel /u/, again suggesting steeper decay and consequently more pressed phonation for male speakers with HI, while no differences were located in other vowels or for female speakers. PSP has been associated with phonation type. Lower values indicate pressed phonation while higher values show breathy phonation [38]. PSP values were significantly higher for both male and female speakers with HI, although there was variability depending on the vowel.

Overall, although patterns display some variability depending on gender and vowel, significant differences in a number of glottal source characteristics as investigated via time- and frequency-based parameters provided by Aalto Aparat, have been located in HI vs NH speech, suggesting differential laryngeal adjustment than normal for HI speakers in agreement with previous literature [23]. Results in most parameters indicate more pressed phonation than normal, at least for male speakers with HI, which could also be associated with problematic placement of stress and intensity control. Gender differences in the HI group may also be related to differences in intelligibility

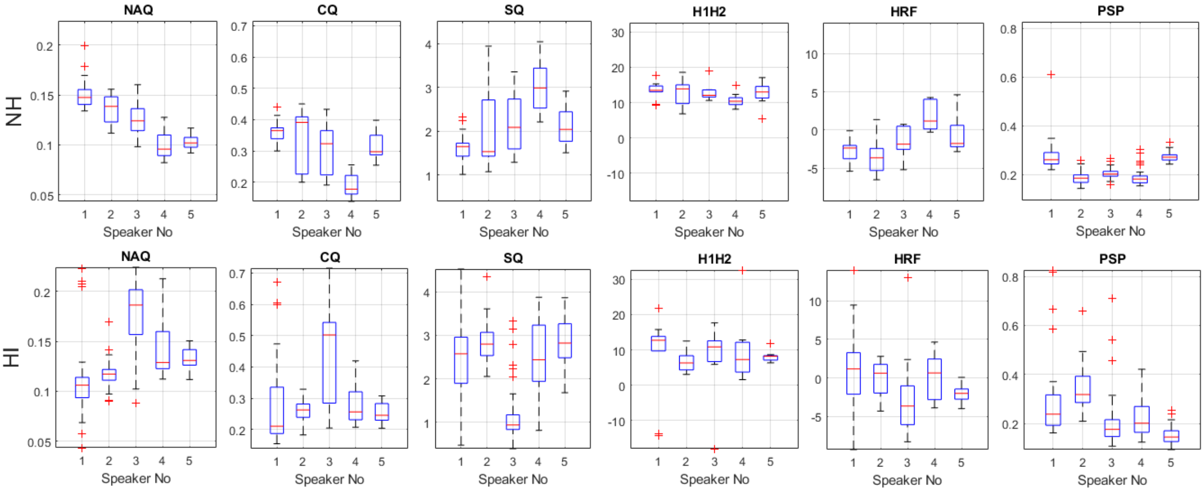

Figure 2: *Boxplots of all six parameters for NH (top) and HI (bottom) speakers obtained from analyzing vowel /a/.*

level and individual strategies. Looking at Figure 2, individual variability is evident in most parameters. Hence this factor could also be involved in our investigation. Among the three corner vowels /i, a, u/, most differences in glottal source parameters between speakers with HI and NH are located in vowel /u/. This is an interesting observation, as vowel /u/ is a high back vowel that has been found significantly fronted for Greek speakers with HI [39]. Thus it seems that articulation and phonation issues are both evident for this vowel in HI Greek speech.

5. Conclusions and Future Work

In this paper, we examined the use of automatic glottal source feature extraction in the context of voice function assessment of speakers with hearing loss. A user-friendly and fully non-invasive tool, Aalto Aparat, was used to automatically extract the glottal source signal from speech recordings and to obtain glottal parameters that characterise HI vs NH speech. Although variability was located in the results, the main trends observed include

Aalto Aparat is a very convenient, interactive, user-friendly tool for performing glottal source analysis. However it has certain limitations. Namely, the glottal analysis assumes speech is stationary. This means that the vocal tract configuration does not change inside the analysis time frame. This is not the case for speakers with HI but also for many speakers with NH as well. Hence, time-varying GIF can be utilized for increased robustness and accuracy of the results. In addition, critical time instants extraction can be problematic in some cases; therefore, more robust time domain parameters can be used like the so-called quasi-quotients (for example, the Quasi-Open Quotient (QOQ)). Finally, GIF methods are improving, including deep neural network (DNN)-based strategies [40, 41]. One can suggest the estimation of the glottal source using DNNs rather than plain source-filter based methods and perform feature extraction on that source waveform.

A next step in our work would be conducting speaker by speaker statistical comparisons in order to find out to what extent differences from speakers with NH can be observed in individual speakers with HI. Individual strategies in vocal adjustments and intelligibility level could influence measurements in the chosen parameters as also highlighted in [23]. Future work could also incorporate the association of time- and frequency-based parameters with HI voice quality ratings by speech pathologists, such as general voice quality, breathiness, hoarseness and laryngeal strain or the components of the widely-used GRBAS scale [42], so as to identify which parameters significantly correlate with specific perceptual attributes.

!!!According to literature, HI voice deviations have not been reliably based on XXXXXXXXXXXXXXX (ref 22), while deviant HI laryngeal features have not been significantly correlated to voice quality characteristics (Metz, Whitehead, Whitehead, 1984)!!!. In addition, sustained vowels vs VCV sequences with different consonants can be incorporated in the analysis. ELG signal analyses of connected speech (i.e. phonetically balanced sentences) has been reported not as suitable as that of sustained vowels for detecting deviating voice quality in deaf speech (ref 22). Hence sustained phonation data could be analysed and compared with analyses from /pVCV/ disyllables including also fricative consonants. Peak flow has been shown to differ in anticipation of the voiceless fricative /s/ than before /p, b, v/ (ref 40). Voiceless consonants are produced with a glottal abduction gesture that has to be coordinated in time with the making and breaking of the oral closure/constriction for stops and fricatives [43]. As interarticulator programming has been found deviant in HI speech, we expect that an investigation of voice features involving production of different voiceless consonants may present great interest.

The usefulness of GIF has been documented in atypical voice analysis [44, 45, 46, 47, 48]. However there is still paucity of research on disordered voices as compared to healthy voices [14] and more research is needed specifically on the voice characteristics of speakers with HI either using hearing aids or cochlear implants, as the influence of hearing loss and remediation on many aspects of voice and speech is yet to be defined [6, 49]. Computerised assessment of voice quality can aid and complement the diagnosis and treatment of voice disorders by speech therapists or ENT doctors (e.g. ARE THESE ARTICLES SUITABLE???? Jalalinajafabadi, 2016, Cesari et al, 2018). However more research is required in clinical voice assessment [50]. Limitations in GIF application to pathological voices as accuracy of GIF methods deteriorates when speech signal becomes irregular and weak.

6. References

[1] J. M. Pickett, “Producing speech: Contemporary issues, for katherine safford harris,” *The Journal of the Acoustical Society of America*, vol. 100, no. 4, pp. 1931–1934, 1996.

[2] I. Hochberg, H. Levitt, and M. J. Osberger, *Speech of the hearing impaired : research, training, and personnel preparation / edited by Irving Hochberg, Harry Levitt, and Mary Joe Osberger.*, ser. Perspectives in audiology series. University Park Press, 1983.

[3] M. J. Osberger and N. S. McGarr, “Speech production characteristics of the hearing impaired,” ser. Speech and Language, N. J. LASS, Ed. Elsevier, 1982, vol. 8, pp. 221–283.

[4] D. Calvert, “Deaf voice quality: a preliminary investigation,” *Volta Review*, vol. 64, pp. 402–403, 1962.

[5] C. Thomas-Kersting and R. L. Casteel, “Harsh voice: Vocal effort perceptual ratings and spectral noise levels of hearing-impaired children,” *Journal of Communication Disorders*, vol. 22, no. 2, pp. 125–135, 1989.

[6] A. C. Coelho, D. M. Medved, and A. G. Brasolotto, “Hearing loss and the voice,” in *Update On Hearing Loss*. Rijeka: IntechOpen, 2015.

[7] B. Maasen and D.-J. Povel, “Glottal determinants of deaf voice quality,” in *Proceedings of the XIth International Congress of Phonetic Sciences*, 1987, pp. 377–380.

[8] J. Subtelny, R. L. Whitehead, and N. Orlando, “Description and evaluation of an instructional program to improve speech and voice diagnosis of the hearing impaired.” *Volta Review*, vol. 82, no. 2, pp. 85–95, 1980.

[9] V. Ball, “Computer-based tools for assessment and remediation of speech,” *British Journal of Disorders of Communication*, vol. 26, no. 1, pp. 95–113, 1991.

[10] D. Howard, *Electroglottography (EGG) / Electrolaryngography (ELG)*. Sage Knowledge, 2019.

[11] D. D. Deliyski, R. E. Hillman, and D. D. Mehta, “Laryngeal high-speed videoendoscopy: Rationale and recommendation for accurate and consistent terminology,” *Journal of Speech, Language, and Hearing Research*, vol. 58, no. 5, pp. 1488–1492, 2015.

[12] J. G. Švec, F. Šram, and H. K. Schutte, “Videokymography in voice disorders: What to look for?” *Annals of Otology, Rhinology & Laryngology*, vol. 116, no. 3, pp. 172–180, 2007.

[13] G. Fant, *Acoustic theory of speech production*. Walter de Gruyter, 1970, no. 2.

[14] 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, 2011.

[15] I. V. McLoughlin, O. Perrotin, H. Sharifzadeh, J. Allen, and Y. Song, “Automated assessment of glottal dysfunction through unified acoustic voice analysis,” *Journal of Voice*, 2020.

[16] P. Corcoran, A. Hensman, and B. Kirkpatrick, “Glottal flow analysis in parkinsonian speech,” in *12th International Conference on Bio-inspired Systems and Signal Processing*, 2019, pp. 116–123.

[17] J. Hanratty, C. Deegan, M. Walsh, and B. Kirkpatrick, “Analysis of glottal source parameters in parkinsonian speech,” in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2016, pp. 3666–3669.

[18] P. Barche, K. Gurugubelli, and A. K. Vuppala, “Towards Automatic Assessment of Voice Disorders: A Clinical Approach,” in *Interspeech*, 2020, pp. 2537–2541.

[19] S. R. Kadiri and P. Alku, “Analysis and detection of pathological voice using glottal source features,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 367–379, 2020.

[20] R. J. Baken, *Clinical measurement of speech and voice*, 2nd ed. San Diego: Singular Thomson Learning, 2000.

[21] R. Mora, B. Crippa, E. Cervoni, V. Santomauro, and L. Guastini, “Acoustic features of voice in patients with hearing loss,” *Journal of otolaryngology - head and neck surgery*, vol. 41, pp. 8–13, 2012.

[22] N. Arends, D.-J. Povel, E. V. Os, and L. Speth, “Predicting voice quality of deaf speakers on the basis of glottal characteristics,” *Journal of Speech, Language, and Hearing Research*, vol. 33, no. 1, pp. 116–122, 1990.

[23] J. Mahshie and A. Oster, “Electroglottograph and glottal air flow measurements for deaf and normal-hearing speakers,” *Speech Transmission Laboratory Quarterly Progress and Status Report*, vol. 32, pp. 19–27, 1991.

[24] N. Jaganathan and B. Kanagaraj, “Analysis of deaf speakers’ speech signal for understanding the acoustic characteristics by territory specific utterances,” *Circuits and Systems*, vol. 7, pp. 1709–1721, 2016.

[25] R. B. Monsen, “Voice quality and speech intelligibility among deaf children,” *American Annals of the Deaf*, vol. 128, no. 1, pp. 12–19, 1983.

[26] V. I. Wolfe and T. M. Steinfatt, “Prediction of vocal severity within and across voice types,” *Journal of Speech, Language, and Hearing Research*, vol. 30, no. 2, pp. 230–240, 1987.

[27] P. Alku, H. Pohjalainen, and M. Airaksinen, “Aalto aparat—a freely available tool for glottal inverse filtering and voice source parameterization,” *Subsidia: Tools and Resources for Speech Sciences*, pp. 21–23, 2017.

[28] A. Sfakianaki, “An acoustic study of coarticulation in the speech of greek adults with normal hearing and hearing impairment,” Ph.D. dissertation, Aristotle University of Thessaloniki, Greece, 2012.

[29] 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.

[30] M. Airas and P. Alku, “Comparison of multiple voice source parameters in different phonation types,” in *Interspeech*, 2007, pp. 1410–1413.

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

[32] P. Alku, C. Magi, S. Yrttiaho, T. Bäckström, and B. Story, “Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering,” *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3289–3305, 2009.

[33] C. Manfredi, M. D’Aniello, P. Bruscaglioni, and A. Ismaelli, “A comparative analysis of fundamental frequency estimation methods with application to pathological voices,” *Medical engineering and physics*, vol. 22, pp. 135–47, 2000.

[34] H. Strik and L. Boves, “Control of fundamental frequency, intensity and voice quality in speech,” *Journal of Phonetics*, vol. 20, no. 1, pp. 15–25, 1992.

[35] A. Camacho and J. G. Harris, “A sawtooth waveform inspired pitch estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1638–1652, 2008.

[36] M. Airas, “Tkk aparat: An environment for voice inverse filtering and parameterization,” *Logopedics, phoniatrics, vocology*, vol. 33, pp. 49–64, 2008.

[37] C. Bickley and K. Stevens, “Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies,” *Journal of Phonetics*, vol. 14, no. 3, pp. 373–382, 1986, voice Acoustics and Dysphonia Gotland, Sweden, August 1985.

[38] P. Alku, “Parabolic spectral parameter — a new method for quantification of the glottal flow,” *Speech Communication*, vol. 22, no. 1, pp. 67–79, 1997.

[39] A. Sfakianaki, K. Nicolaidis, and A. Okalidou, “Vowel production and intelligibility in hearing-impaired speech: Evidence from greek,” *Glossologia*, vol. 24, pp. 75–92, 01 2016.

[40] M. Airaksinen, T. Raitio, and P. Alku, “Noise robust estimation of the voice source using a deep neural network,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 5137–5141.

[41] N. Narendra, M. Airaksinen, B. Story, and P. Alku, “Estimation of the glottal source from coded telephone speech using deep neural networks,” *Speech Communication*, vol. 106, pp. 95–104, 2019.

[42] N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, M. Blanco-Velasco, and F. Cruz-Roldan, “Automatic assessment of voice quality according to the grbas scale,” in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 2478–2481.

[43] A. Löfqvist and H. Yoshioka, “Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production,” *Speech Communication*, vol. 3, no. 4, pp. 279–289, 1984.

[44] J. Deller, “Evaluation of laryngeal dysfunction based on features of an accurate estimate of the glottal waveform,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 7, 1982, pp. 759–762.

[45] R. H. Colton, D. W. Brewer, and M. Rothenberg, “Evaluating vocal fold function.” *The Journal of otolaryngology*, vol. 12, no. 5, pp. 291–294, 1983.

[46] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, “Phonatory function associated with hyperfunctionally related vocal fold lesions,” *Journal of Voice*, vol. 4, no. 1, pp. 52–63, 1990.

[47] P. Howell and M. Williams, “Acoustic analysis and perception of vowels in children’s and teenagers’ stuttered speech,” *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1697–1706, 1992.

[48] B. Hammarberg, “Voice research and clinical needs,” *Folia phoniatrica et logopaedica*, vol. 52, pp. 93–102, 2000.

[49] A. Gautam, J. Naples, and S. Eliades, “Control of speech and voice in cochlear implant patients,” *The Laryngoscope*, vol. 129, 01 2019.

[50] N. Roy, J. Barkmeier-Kraemer, T. Eadie, M. Sivasankar, D. Mehta, D. Paul, and R. Hillman, “Evidence-based clinical voice assessment: A systematic review,” *American journal of speech-language pathology / American Speech-Language-Hearing Association*, vol. 22, 11 2012.