INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

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]. 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 [2]. Inappropriate pausing at linguistic boundaries and decreased syllable production per breath unit as well as inefficient vocal fold vibration patterns, abduction/adduction gestures and laryngeal adjustment [1] have been reported to cause pitch and loudness issues, excess breathiness, strain, roughness and vocal fatigue to speakers with hearing impairment (henceforth HI) [3, 4, 5].

Vocal function can be examined using various instrumental methods. Electrolaryngography (ELG)/electroglottography (EGG) is a non-invasive technique commonly used for vocal fold vibration monitoring and voice quality assessment [6]. 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/EGG signal analysis, vocal fold movement can also be observed via laryngeal endoscopic imaging, such as videoendoscopy and videostroboscopy [7]. These methods involve the insertion of a long tube in the speaker's throat in order to visualise vocal fold activity. Although they are recommended for the assessment of the vibratory function of the vocal folds during phonation [8], they are both invasive and expensive, and their use in clinical practice has been characterised as highly subjective [9]. 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 [10], 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 [11]. 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 [12, 13, 14, 15]. 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 [6]. 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) [16, 17, 18, 19]). Several glottal characteristics have been associated with HI voice disorders. For example, variations in $F0$ and its amplitude indicate breathiness, roughness, and hoarseness [19, 20, 21] while close-to-open phase ratio and steepness of glottal closure has been associated with breathiness [5]. Furthermore, the extent of vocal fold abduction and glottal efficiency have been related to reduced HI vocal fold mobility and oscillatory efficiency [18]. 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 [18], 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 [22] 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 describes the dataset used and the glottal flow estimation and feature extraction procedure, while Section 4 presents and dis-

http://dx.doi.org/10.21437/Interspeech.2021-870

cusses the statistical analysis 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 [23]. 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 disyllable 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. The recordings were conducted in a sound proof room used for audiological evaluations. Speech was produced at a comfortable rate and volume. It was recorded at 22050 Hz and downsampled to 8000 Hz for the analysis.

3.2. Analysis

A total number of 300 items was analysed. The analysis was conducted on 6 glottal cycles of the stressed first-syllable vowel of every repetition. The cycles were manually selected so as to avoid problematic cases (e.g. uneven, incomplete, or altogether missing cycles), which sometimes occurred 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 [24] 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 [11]. 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 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 and the default formant number, lip radiation coefficient, and low-pass cutoff frequency 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 three 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 quasi-open quotient (QOQ), whereas the frequency domain ones are the harmonic richness factor (HRF), the difference between log-spectral amplitudes of the fundamental and the second harmonic (H1-H2), and the parabolic spectral parameter (PSP). A review illustrating the capability of all these glottal source parameters in discriminating different voice qualities can be found in [11].

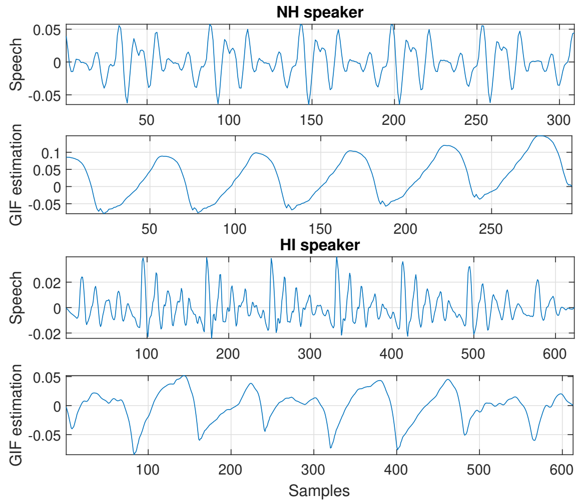

Figure 1: *Glottal source estimation based on the analysis of the stressed vowel /a/ in the disyllable /'papa/ produced by a speaker with NH and a speaker with HI. First panel: recorded speech signal (NH), second panel: estimated glottal source signal (NH), third panel: recorded speech signal (HI), fourth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

In brief, 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 [25], 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. Additionally, QOQ is defined as the ratio of the quasi-open time over the quasi-closed time of the glottis, normalized with respect to the period of the waveform. It quantifies the time interval when the glottal flow is above some threshold (usually 50%) of the difference between the maximum and minimum flow. QOQ correlates well to the so-called Open Quotient (OQ), but only amplitude measures are used for its computation, thus alleviating the infamous glottal opening instant (GOI) detection problem. It should be noted that although the time-based parameters Open Quotient (OQ) and Speed Quotient (SQ) were calculated as well, they were excluded from the statistical analyses, as the majority of the automatically obtained values were erroneous. In order to receive accurate OQ and SQ results, a great deal of fine tuning of the GIF settings is required.

Regarding frequency domain parameters, the H1-H2 difference (H1H2) 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 [26]. 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

Table 1: *Mean and SE values of all parameters in vowels /i,a,u/ with statistical comparisons within gender (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.*

Vowel	Gender	HS	CQ	NAQ	QOQ	H1H2	HRF	PSP
/a/	M	NH	0.34 (0.01)	0.13 (0.004)	0.51 (0.007)	12.99 (1.33)	-1.81 (0.81)	0.27 (0.012)
		HI	0.25 (0.01)	0.12 (0.003)	*0.45 (0.006)	*7.49 (1.33)	-1.00 (0.81)	0.24 (0.012)
	F	NH	0.28 (0.009)	0.12 (0.002)	0.38 (0.004)	12.16 (1.09)	-1.03 (0.66)	0.19 (0.008)
		HI	0.34 (0.009)	*0.14 (0.002)	*0.42 (0.005)	8.75 (1.09)	-0.16 (0.66)	*0.23 (0.008)
/i/	M	NH	0.40 (0.008)	0.15 (0.003)	0.61 (0.009)	12.85 (1.17)	-5.10 (1.02)	0.41 (0.01)
		HI	*0.29 (0.008)	0.15 (0.003)	*0.53 (0.009)	9.05 (1.17)	-2.04 (1.02)	*0.33 (0.01)
	F	NH	0.31 (0.006)	0.14 (0.003)	0.44 (0.007)	9.39 (0.96)	-1.19 (0.83)	0.28 (0.009)
		HI	*0.41 (0.006)	*0.19 (0.003)	*0.47 (0.007)	11.25 (0.96)	-2.52 (0.83)	0.25 (0.009)
/u/	M	NH	0.38 (0.009)	0.17 (0.004)	0.55 (0.009)	15.03 (1.49)	-5.98 (1.07)	0.31 (0.012)
		HI	*0.32 (0.01)	*0.15 (0.005)	0.56 (0.01)	*6.46 (1.49)	*1.09 (1.07)	*0.43 (0.018)
	F	NH	0.36 (0.008)	0.17 (0.004)	0.45 (0.008)	10.38 (1.22)	-3.40 (0.88)	0.29 (0.014)
		HI	*0.30 (0.007)	*0.15 (0.003)	0.45 (0.007)	9.79 (1.22)	-0.73 (0.88)	0.30 (0.012)

has extensively been used in the literature [27, 28], $F0$ information has also been extracted.

4. Results and Discussion

Statistical analyses of variance (ANOVA) were run for the three time-based (CQ, NAQ, QOQ) and the four frequency-based parameters (H1H2, HRF, PSP, $F0$) of the vowels /a, i, u/, vs the factors gender and hearing status. The results showed that gender was significant for CQ, QOQ, PSP and $F0$ in all vowels, and additionally for NAQ in /i/. Thus, Tukey post-hoc tests between the two genders within hearing group were conducted. Within the NH group, besides the expected lower $F0$ values for male vs female speakers, additional gender differences were found, such as higher QOQ values in all vowels, higher CQ and PSP values and lower SQ values in /a/ and /i/, and lower HRF values in /i/ for male vs female speakers. Gender differences within the HI group 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 ([4] for a review). Instead of $F0$ mean values, $F0$ variance or other measures might be more promising as features [29]. The hearing status factor was found significant for all parameters except QOQ in vowel /u/, while in vowels /a/ and /i/ hearing status was significant for NAQ, QOQ, H1H2, and for NAQ, QOQ, PSP correspondingly. Table 1 summarises the mean values and standard error (SE) of the time- and frequency domain 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 [30]. 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 [11]. In our data, the two parameters indeed follow similar trends. Since QOQ is a correlate of OQ, as related to voice quality it reflects the extent of vocal fold abduction; the higher its value, the more abducted the vocal folds. According to our data, the two genders present a different picture. Male speakers with HI have a lower QOQ in vowels /a/ and /i/ than their NH counterparts, while female speakers with HI have a higher QOQ than normal in the same vowels. This result again indicates more tense voice for male speakers with HI and more breathy voice for female speakers with HI, since vocal folds generally remain open for a greater portion of the cycle for female HI speakers in comparison with their NH counterparts.

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 [17]. 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 [26]. 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, while no differences were located in other vowels or for female speakers in any vowel. PSP has been associated with phonation type. Lower values indicate pressed phonation while higher values breathy phonation [31]. PSP values were significantly higher for female speakers with HI, while variable results were found for male speakers with HI depending on the vowel.

Overall, 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 [18]. Results in most parameters indicate more breathy voice for female speakers with HI and more pressed phonation than normal for male speakers with HI, which could

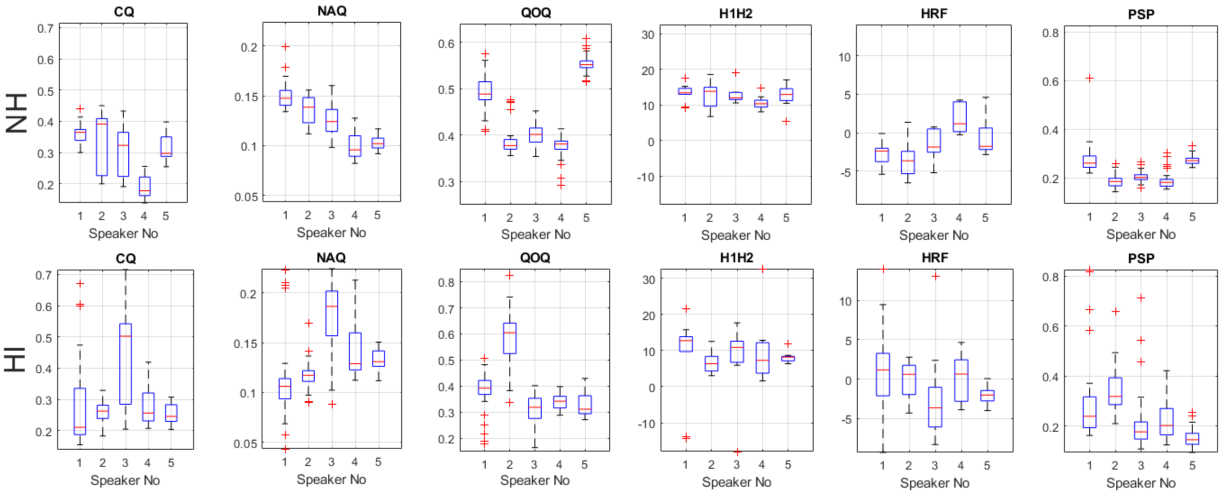

Figure 2: *Boxplots of all six parameters for NH (top) and HI (bottom) speakers obtained from analyzing vowel /a/.*

also be associated with problematic placement of stress and intensity control. Gender differences in the HI group may be related to differences in intelligibility level and individual strategies. Looking at Figure 2, individual variability is evident in most parameters. Hence this factor will be involved in the next step of 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 [32]. Therefore the existence of both deviant articulation and phonation in this vowel suggests a link between articulatory configuration and phonation in HI speech [2].

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 lower CQ, NAQ, QOQ and H1H2 values for male speakers with HI suggesting higher intesity or more pressed phonation, and higher CQ, NAQ, QOQ and PSP values for female speakers with HI indicating more breathiness than normal.

Aalto Aparat is a very convenient, interactive, user-friendly tool for performing glottal source analysis. However it has certain limitations. Namely, the analysis assumes that the vocal tract configuration does not change inside the analysis time frame. This is not necessarily true, not only 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. Moreover, GIF methods are improving, including deep neural network (DNN)-based strategies [33, 34]. One can suggest the estimation of the glottal source using DNNs rather than plain linear source-filter based methods and perform feature extraction on that source waveform. Additionally, more recordings of speakers would provide statistical robustness to the aforementioned results.

A next step in our work would also include 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 [18]. 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 [35], so as to identify which parameters significantly correlate with specific perceptual attributes. 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 [17]. 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/ [36]. 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 [36]. 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.

Computerised assessment of voice quality can aid and complement the diagnosis and treatment of voice disorders by speech therapists or ENT doctors [37]. GIF has been documented as a useful method in atypical voice analysis [38, 39]. However there is still paucity of research on disordered voices as compared to healthy voices [11, 9], 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 [4, 40].

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] M. J. Osberger and N. S. McGarr, "Speech production character-

istics of the hearing impaired," ser. Speech and Language, N. J. LASS, Ed. Elsevier, 1982, vol. 8, pp. 221–283.

[3] 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.

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

[5] 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.

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

[7] 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.

[8] R. Patel, S. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J. Svec, and R. Hillman, "Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function," *American Journal of Speech-Language Pathology*, vol. 27, pp. 1–19, 06 2018.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] N. Bolfan-Stosic and B. Simunjak, "Effects of hearing loss on the voice in children," *The Journal of Otolaryngology*, vol. 36, pp. 120–3, 04 2007.

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

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

[37] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "Voice disorder detection via an m-health system: Design and results of a clinical study to evaluate vox4health," *BioMed Research International*, vol. 2018, pp. 1–19, 08 2018.

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

[39] 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.

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