Interspeech 2018
2-6 September 2018, Hyderabad

On the relationship between glottal pulse shape and its spectrum: correlations of open quotient, pulse skew and peak flow with source harmonic amplitudes

Christer Gobl, Andy Murphy, Irena Yanushevskaya, Ailbhe Ní Chasaide

Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences
Trinity College Dublin, Ireland

cegobl@tcd.ie, murpha61@tcd.ie, yanushei@tcd.ie, anichsid@tcd.ie

Abstract

This paper explores the relationship between the glottal pulse amplitude (U_p) and the amplitude of the first harmonic ($H1$), as well as the combined effects of U_p, the open quotient (O_q) and degree of pulse asymmetry/skew (R_k) on the low end of the source spectrum. This serves to elucidate their relationship to the $H1$-$H2$ estimate, widely used to make inferences on changes in O_q and voice quality. It has been suggested that $H1$ is mainly determined by U_p and that the pulse shape has a relatively small impact. To investigate this, a series of glottal pulses were generated using the LF model, where U_p was kept constant, while O_q and R_k were systematically varied. The resulting harmonic amplitudes of these pulses show that U_p is not the sole determinant of $H1$. Rather, $H1$ is highly dependent on O_q and to a certain degree also on R_k. Although the effects of these parameters on the lowest harmonics is rather complex, we find that the $H1$-$H2$ measure is broadly correlated with O_q. However, there is also a strong effect of differences in glottal skew, particularly at high O_q values, which could invalidate inferences on O_q and voice quality from estimates of $H1$-$H2$.

Index Terms: voice source, glottal flow, LF model, open quotient, skew, $H1$-$H2$, frequency domain

1. Introduction

The present study is part of an ongoing investigation of the detailed mapping of time and frequency domain dimensions of the voice source, important to our understanding of voice source variation in speech. Each domain yields different insights: whereas the time domain illuminates production, the frequency dimension relates more directly to perception.

Our past research has entailed detailed analyses of the voice source, using inverse filtering techniques and the modelling of the glottal waveform [1] using the Liljencrants-Fant (LF) model [2]. These have examined the source correlates of voice qualities [3, 4], the variation of the voice in prosody, to signal linguistic [5-7] and affective information [8-11].

Such time domain voice source analyses are necessarily limited in scope due in part to the fact that they require stringent recording conditions, e.g., to ensure phase-linearity (for a fuller discussion of the limiting factors, see [1]). Consequently, there is a relatively limited body of analytic data available, and researchers have tended to use spectral information as proxy measures of the source. Such a measure is $H1$-$H2$, which is widely taken to be an indicator of the open quotient (O_q) of the glottal pulse [12], and long used as a measure of the breathy-tense dimension of voice quality [13-18]. An increase in O_q is assumed to lead to an increase in the amplitude of $H1$, thus increasing $H1$-$H2$ [19]. (Note that when the measure is based on the speech waveform, the H1*-H2* measure is often used, to correct for the vocal tract resonances [20, 12].)

Although a full time-frequency mapping is our objective, the focus of this paper is on the lower end of the source spectrum. We explore the relationship of the amplitude of the glottal flow pulse (i.e. the peak flow, U_p) and the amplitude of the first harmonic of the source spectrum ($H1$). Furthermore, we examine the combined effects of U_p, the open quotient (O_q) and pulse asymmetry/skew (R_k) on the low end of the source spectrum, looking in detail at how $H1$, $H2$, $H3$ and the $H1$-$H2$ estimate relate to O_q.

Fant and Lin [21] provide a theoretical basis for the relationship between time and frequency domain representations of the voice source and they show that the amplitude of the first harmonic of the source spectrum can be derived according to (1). Here the effect of the vocal tract filter is disregarded (or assumed to have been properly eliminated by inverse filtering) and $|R(f)|$ represents the effect of the radiation transfer on the amplitude spectrum of the speech output at a specific distance from the lip opening of the speaker.

$$H1 = k \cdot \frac{U_p}{2} \left| R(f) \right| \tag{1}$$

If we ignore the radiation factor, we end up with $H1$ being determined by $k \cdot U_p/2$, i.e. half the peak flow scaled by the value of k, which is a correction factor. In this case $H1$ would correspond to the amplitude of the first harmonic of the waveform produced at the glottis, and that is how the term $H1$ is used in the following text.

As pointed out by Fant and Lin [21], for approximate calculations, it is convenient to set the correction factor $k = 1$. Although it is explained that k does vary with O_q, no detailed information is presented on precisely how the glottal pulse shape affects k (and $H1$).

It is hoped that a growing understanding of the time-frequency mapping of source dimensions will facilitate the development of more robust analysis techniques, not constrained by factors such as the recording conditions. We further see it as key to establishing how source dimensions map to the auditory perception of voice quality (see for example [22]), as these are not well understood. Ultimately the goal is to understand and model how we use modulation of the voice for the rich nuancing of prosody in speech communication.

In a parallel research strand, we are developing text-to-speech synthesis for the dialects of Irish (see [23, 24]), and plan to exploit them in educational games [25] and dialogue systems [26]. Our basic research on voice modelling thus goes

10.21437/Interspeech.2018-2532

hand in hand with our goal to provide more adequate descriptions of voice modulation in prosody, which is essential for this kind of application where speaker affect is important.

2. Methods

To analyse the relationships between the glottal pulse shape and the amplitude of source spectral components, the LF model [2] was used to generate glottal pulses with constant peak amplitude (U_p) and pulse duration, while the open quotient and the degree of pulse asymmetry were varied in controlled steps by changing the O_q and R_k parameters, as defined in Figure 1.

The LF model (see Figure 1) is determined by the two expressions in (2), which generate the differentiated glottal waveform of the open phase and return phase respectively. Note that as our focus here is on O_q and R_k, the pulses generated did not have a return phase (i.e. $T_a = 0$).

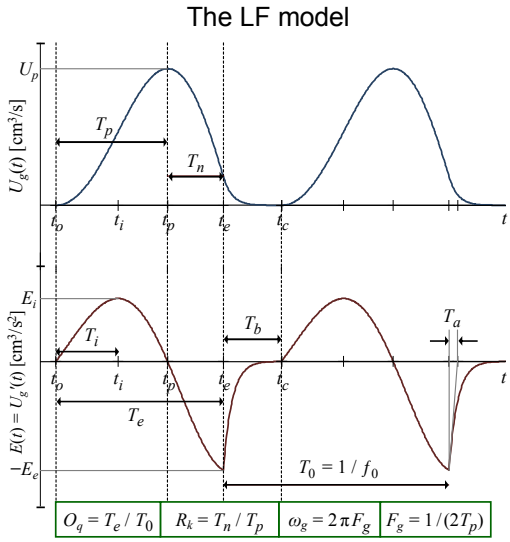

Figure 1. *Two LF model pulses and parameter definitions (for details, see [1]). Glottal flow (top), flow derivative (bottom).*

$$U_g'(t) = \begin{cases} E_0\ e^{\alpha t} \sin \omega_g t & t_o \le t \le t_e \\ \dfrac{-E_e}{\varepsilon T_a}\left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon T_b} \right) & t_e < t < t_c \end{cases} \quad (2)$$

In this work we are particularly interested in the flow pulse, rather than the flow derivative, and therefore the alternative expressions shown in (3) are used, which are obtained by the integration of (2). They produce the open phase and return phase of the LF glottal flow pulse (see also [27, 28]).

As pointed out in [29], the actual LF model parameters are E_e, T_e, ω_g, α, ε and T_b. E_0 and T_a in (2) and (3) are only auxiliary parameters: E_0 is determined by E_e, α, T_e, and ω_g, while T_a is determined by ε and T_b.

E_e is the excitation strength, T_e is the duration of the open phase which determines O_q, ω_g is the angular frequency ($2\pi F_g$) of the glottal pulse in the open phase and α relates to its bandwidth. Apart from these parameters defining the LF pulse, the model requires that there is area balance [2], i.e. that the area of the positive part of the differentiated glottal pulse equals the area of the negative part (see the lower panel of Figure 1), so as to avoid any drift in the zero-flow line. To achieve this, α is allowed to 'float', i.e. it is implicitly determined by the other parameters in order to obtain area balance.

It is clear from the above that U_p is not a parameter of the LF model. Therefore, as for α, U_p is typically allowed to float, and will attain a value entirely determined by the settings of the other parameters.

$$U_g(t) = \begin{cases} \dfrac{E_0\ e^{\alpha t} \sin\left(\omega_g t - \arctan \dfrac{\omega_g}{\alpha} \right)}{\sqrt{\alpha^2 + \omega_g^{\,2}}} + \dfrac{E_0\ \omega_g}{\alpha^2 + \omega_g^{\,2}} & t_o \le t \le t_e \\ \dfrac{E_e}{\varepsilon^2 T_a}\left[e^{-\varepsilon(t-t_e)} + \varepsilon\, e^{-\varepsilon T_b}\left(t - \left(t_c + \dfrac{1}{\varepsilon} \right) \right) \right] & t_e < t < t_c \end{cases} \quad (3)$$

This does not serve the present purpose where we wish to keep U_p constant while varying O_q and R_k. To do this, we use a version of the iterative procedure described in [29], where the direct control of ω_g is sacrificed, so that U_p can be directly controlled.

However, we also wish to control the glottal skew by varying R_k in a systematic way, and since ω_g is allowed to float, this also means that R_k cannot be controlled (see Figure 1). Therefore, the algorithm in [29] was modified so that E_e would float instead of ω_g. This was achieved by replacing the two partial derivatives with respect to ω_g in [29] with the following partial derivatives with respect to E_e, shown in (4) and (5).

$$\frac{\partial \mathrm{f}_1}{\partial E_e} = \pi \omega_g \left(e^{-\alpha T_e} + e^{-\alpha\left(T_e - \pi \omega_g^{-1} \right)} \right) \quad (4)$$

$$\frac{\partial \mathrm{f}_2}{\partial E_e} = 0 \quad (5)$$

Using this modified parameter control of the LF model, 81 different flow pulses were generated. U_p remained fixed and was arbitrarily set to 10. The glottal period (i.e. $T_e + T_b$, see Figure 1) was kept constant at 10 ms.

Nine different O_q settings were used, ranging from 0.15 to 0.95 in steps of 0.1, covering a very wide range of O_q values. For each of the O_q settings, nine pulses with different R_k values were produced, also ranging from 0.15 to 0.95 in steps of 0.1 – again, covering most of the possible range of R_k values.

The sampling frequency was 20 kHz, which was deemed sufficiently high to avoid any impact of aliasing on the lower end of the source spectrum. Each pulse was repeated five times in order to produce a harmonic spectrum. A 1000-point (50 ms, rectangular window) DFT spectrum was calculated for each of the 81 glottal waveforms, and the amplitudes of the first three harmonics were extracted. The window size was chosen so as to ensure that the output frequency samples would coincide with the harmonic frequencies, thus avoiding potential rounding errors.

3. Results of Spectral Analysis

3.1. Influence of O_q and R_k on $H1$, $H2$ and $H3$

Figure 2 shows how the amplitudes (in dB) of the first three harmonics vary as a function of O_q. Each panels shows this variation for different R_k settings, 0.15, 0.35 and 0.65 in the top, mid and lower panels respectively. Figure 3 shows in separate panels the variation in $H1$ and $H2$ (on a linear scale) with O_q, with the nine different R_k settings superimposed. 0 dB corresponds to $k = 1$, which here is the (linear) value of 5.

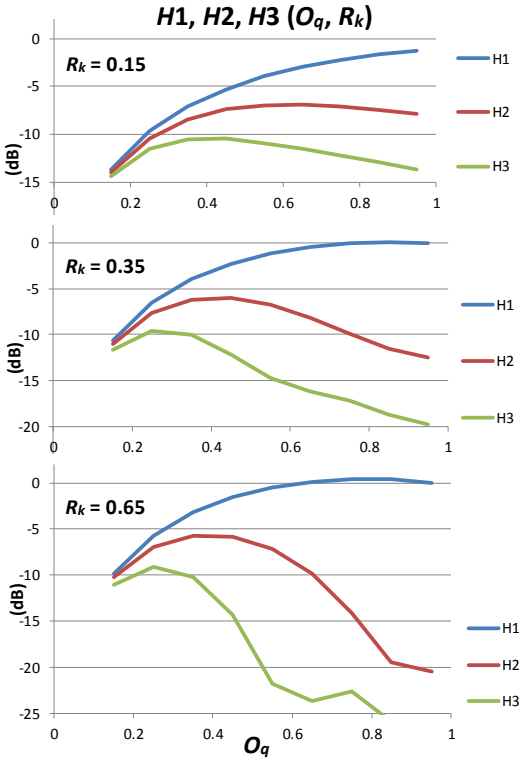

Figure 2: *H1, H2 and H3 amplitude levels as a function of O_q for three different R_k settings.*

Figures 2 and 3 show that $H1$ varies considerably as a function of O_q and that there is also a substantial influence of R_k. At relatively low O_q values $H1$ increases with increasing O_q. However, this increase plateaus and $H1$ even drops (Figure 3) at very high O_q (except where R_k is very low). Once the plateau is reached, the further changes in $H1$ are relatively minor.

It is also clear in the upper panel of Figure 3 that R_k has a strong influence on $H1$, particularly when the pulse is skewed, in the range of 0.15 to 0.35. The influence of R_k extends up to about 0.45.

The lower panel of Figure 3 shows how $H2$ is influenced by the pulse shape parameters. At very low O_q values $H2$ rises with O_q, but drops thereafter, as O_q continues to rise. The O_q value where this drop occurs and its steepness depends on the R_k setting, being more extreme with high R_k (low skew). $H3$ also drops sharply with rising O_q values, and also shows a strong influence of R_k.

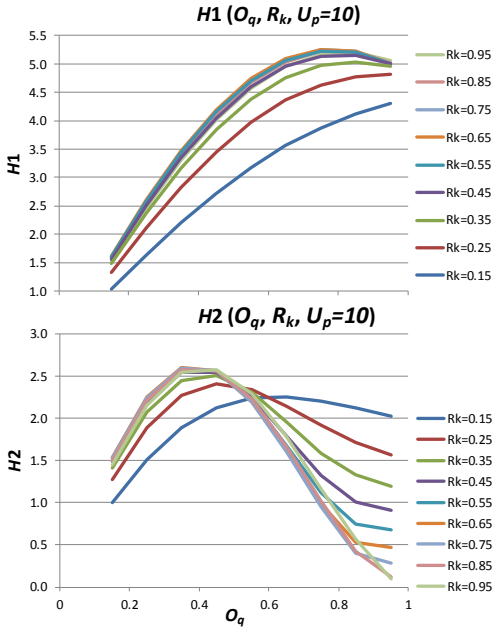

Figure 3: *H1 and H2 as a function of O_q and R_k. $U_p = 10$ in all cases.*

3.2. $H1$-$H2$ as a measure of O_q

As mentioned in the Introduction, the $H1$-$H2$ measure has been widely interpreted in the literature as an indirect measure of O_q. However, it is clear from Figures 2 and 3 that neither $H1$ nor $H2$ have a straightforward correlation with O_q, even if R_k is held constant. Nonetheless, the difference between them is more closely correlated to O_q. And although the $H1$-$H2$ difference is generally assumed to be the consequence of a rise in $H1$ with O_q, Figure 2 suggests that the drop in $H2$ is likely to be an important factor.

A comparison of the three panels of Figure 2, shows that, despite a broad, positive correlation of O_q with $H1$-$H2$, this correlation is greatly affected by the pulse skew R_k. This last is elaborated in detail in Figure 4, which charts how the $H1$-$H2$ correlation with O_q is impacted by variation in pulse skew. As typical values of R_k tend to range from about 0.2 to 0.6, for much of running speech, the $H1$-$H2$ measure is neither a direct nor unique indicator of the open quotient. This supports suggestions by [30] that R_k is likely to have an important impact on the lower part of the source spectrum and on the $H1$-$H2$ measure.

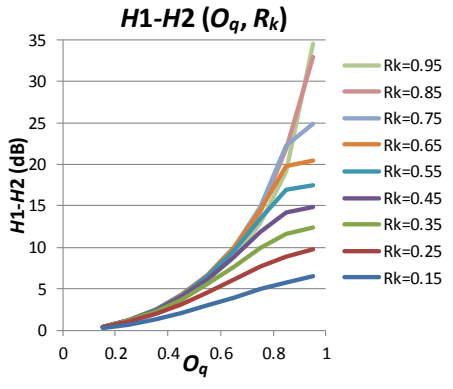

Figure 4: *The $H1$-$H2$ as a function of O_q and R_k.*

To sum up, while there are clear correspondences between the spectral measure $H1$-$H2$ and the open quotient, given the clear impact of other factors (particularly here, the pulse skew), one should be cautious in making inferences on O_q on the basis of $H1$-$H2$. Though it may capture broad trends much of the time, its potential limitations need to be borne in mind.

The present analysis has been limited to the joint influences of U_p, O_q and R_k, as these are considered to be the primary determinants of the lower end of the source spectrum. The analysis does not extend to consideration of other possible influences, such as variation of the return phase (T_a, see Figure 1), which primarily determines the shaping of the upper end of the source spectrum. Nonetheless, interactions are likely, and these will need to be considered at a later point.

4. Predicting *H*1 and deriving the *k* correction factor

We use these spectral measurements to predict more precisely the k factor. The variation in $H1$ due to changes in O_q, closely follows part of a parabolic curve (see Figure 3). Second order polynomial fitting yields R^2 values close to 1 for the nine R_k values (R^2 = 0.9992 or higher). Thus, we can use a quadratic function to derive an estimate of the correction factor k according to equation (6), and, given the amplitude of the peak flow (U_p), $H1$ can be approximated according to equation (1) in the Introduction.

$$k = a_2 O_q^2 + a_1 O_q + a_0 \tag{6}$$

However, the coefficients of the function in (6) depend on R_k. The variation in the three a-coefficients of the polynomial in (6) also closely match quadratic functions (see Figure 5, left panel). Although the fit is very good, the R^2 values were somewhat lower in this case (but not lower than 0.975). The three equation for the a-coefficients as a function of R_k are shown in (7), and the numerical values of the nine coefficients are shown in Table 1. The values have been rounded to three-digit precision, with negligible effect on the accuracy of predictions.

$$\begin{aligned} a_2 &= b_2 R_k^2 + b_1 R_k + b_0 \\ a_1 &= c_2 R_k^2 + c_1 R_k + c_0 \\ a_0 &= d_2 R_k^2 + d_1 R_k + d_0 \end{aligned} \tag{7}$$

Table 1: *The values of the coefficients in equation (7) used to derive the coefficients for the quadratic equation (6) estimating the correction factor, k.*

b_2	b_1	b_0
3.95	–5.42	0.0307
c_2	c_1	c_0
–4.35	6.01	0.811
d_2	d_1	d_0
0.161	–0.243	0.0203

To test the accuracy of the of k correction factor as determined by (6) and (7), the $H1$ prediction errors were calculated on a different set of 64 LF pulses, where both O_q and R_k varied from 0.2 to 0.9 in steps of 0.1. As can be ascertained in the right panel of Figure 5, there is a high correlation between the estimated and actual $H1$ amplitudes. The average error in the $H1$ estimate was 0.14 dB with a maximum error of 0.35 dB.

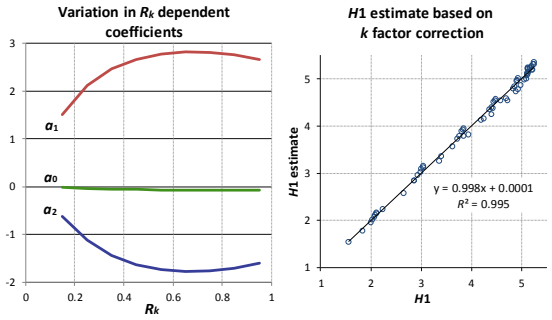

Figure 4: *Variation in the R_k dependent a_2, a_1 and a_0 coefficients of (6) (left panel). H1 estimates vs. actual H1 amplitudes using the k correction factor as defined by equation (6) (right panel).*

5. Conclusions

The approach taken in this paper – using adjustments to the LF model, to allow control of different source parameters – has permitted a detailed exploration of the influence of U_p, O_q and R_k on the shaping of the low end of the source spectrum, and has illustrated the complex interaction of these parameters. A formula is presented that allows a precise estimation of the k factor and prediction of $H1$.

The spectral analyses also illuminate how these parameters affect $H2$, and the impact on the $H1$-$H2$ measure. Despite the broad correspondence with O_q, glottal skew was found to have a strong effect, particularly at high O_q values. These effects could potentially invalidate direct inferences from $H1$-$H2$ on O_q and on voice quality.

This work is seen as part of a larger enterprise of mapping time and frequency dimensions of the source. An extension of the present study will explore how the return phase of the glottal pulse, usually seen as mainly influencing the higher end of the source spectrum, impacts on the low frequency end.

The quadratic correspondence found between U_p and $H1$ as the glottal pulse shape varies, means that there is not a one-to-one mapping: the same harmonic amplitudes can be produced by different glottal pulse configurations. Although the basic framework of Fant and Lin [21] is very useful, the more detailed elaboration of the precise correlations here indicate that the inverse prediction of time domain measures from the frequency domain will be complex and challenging. Future work will explore the potential of deriving the peak glottal flow from the low harmonics, by exploiting some of the correlations elaborated here to impose constraints on possible solutions.

Progress towards comprehensive time-frequency mapping would greatly further our ability to explore the perception of voice source variation. It would facilitate more flexible and robust voice source analysis [31], allowing us greater access to aspects of speech communication that are little understood. This we believe to be vital, if speech synthesis [32, 33] and related technologies are to be adequate for many applications.

6. Acknowledgements

This research was supported by funding from the Department of Culture, Heritage and the Gaeltacht, Government of Ireland (*ABAIR* project).

7. References

[1] C. Gobl and A. Ní Chasaide, "Voice source variation and its communicative functions," in *The Handbook of Phonetic Sciences (Second Edition)*, eds. William J. Hardcastle, John Laver and Fiona E. Gibbon, Oxford, Blackwell, pp. 378-423, 2010.

[2] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 4, pp. 1-13, 1985.

[3] C. Gobl, "A preliminary study of acoustic voice quality correlates," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 4, pp. 9-21, 1989.

[4] C. Gobl and A. Ní Chasaide, "A. Acoustic characteristics of voice quality," *Speech Communication*, vol. 11, pp. 481-490, 1992.

[5] A. Ní Chasaide, I. Yanushevskaya, and C. Gobl, "Prosody of voice: declination, sentence mode and interaction with prominence," *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[6] I. Yanushevskaya, C. Gobl and A. Ní Chasaide, "Cross-speaker variation in voice source correlates of focus and deaccentuation," *INTERSPEECH 2017*, Stockholm, pp. 1034-1038, 2017.

[7] A. Ní Chasaide, I. Yanushevskaya, J. Kane and C. Gobl, "The voice prominence hypothesis: the interplay of F0 and voice source features in accentuation," *INTERSPEECH 2013*, Lyons, pp. 3527-3531, 2013.

[8] C. Gobl, and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication*, vol. 40, pp. 189-212, 2003.

[9] I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "Universal and language-specific perception of affect from voice," *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, pp. 2208-2211, 2011.

[10] A. Ní Chasaide, I. Yanushevskaya, and C. Gobl, "Voice-to-affect mapping: inferences on language voice baseline settings," *INTERSPEECH 2017*, Stockholm, pp. 1258-1262, 2017.

[11] C. Ryan, A. Ní Chasaide, and C. Gobl, " Voice quality variation and the perception of affect: continuous or categorical?," *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, pp. 2409-2412, 2003.

[12] M. Iseli, Y-L Shue, and A. Alwan, "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *J. Acoust. Soc. Am.*, vol. 121, pp. 2283–2295, 2007.

[13] M. Garellek and P. A. Keating, "The acoustic consequences of tone and phonation interactions in Jalapa Mazatec," *Journal of the International Phonetic Association*, vol. 41, no. 4, pp. 185-205, 2011.

[14] M. K. Huffman, "Measures of phonation type in Hmong," *J. Acoust. Soc. Am.*, vol. 81, pp. 495-504, 1987.

[15] P. L. Kirk, P. Ladefoged, and J. Ladefoged. "Using a spectrograph for measures of phonation types in a natural language," *UCLA Working Papers in Phonetics*, vol. 59, pp. 102-113, 1984.

[16] P. Ladefoged and N. Antonanzas-Barriso, "Computer measures of breathy phonation," *UCLA Working Papers in Phonetics*, vol. 61, pp. 79-86, 1985.

[17] J. Kreiman, B. R. Gerratt, and S. D. Khan, "Effects of native language on perception of voice quality," *J. Phonetics*, vol. 38, pp. 588-593, 2010.

[18] J. Kreiman, Y-L. Shue, G. Chen, B. R. 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. Am.*, vol. 132, no. 4, pp. 2625-2632, 2012.

[19] D. H. Klatt and L.C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers," *J. Acoust. Soc. Am.*, vol. 87, pp. 820–856, 1990.

[20] H. M. Hanson, "Glottal characteristics of female speakers: acoustic correlates," *J. Acoust. Soc. Am.*, vol. 101, pp. 466–481, 1997.

[21] G. Fant and Q. Lin, "Frequency domain interpretation and derivation of glottal flow parameters," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 4, pp. 1-21, 1988.

[22] C. Gobl and A. Ní Chasaide, "Perceptual correlates of source parameters in breathy voice," *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, pp. 2437-2440, 1999.

[23] www.abair.ie. TTS systems for three dialects of Irish.

[24] A. Ní Chasaide, N. Ní Chiaráin, C. Wendler, H. Berthelsen, A. Murphy, and C. Gobl, "The ABAIR initiative: bringing spoken Irish into the digital space," *INTERSPEECH 2017*, Stockholm, Sweden, pp. 2113-2117, 2017.

[25] N. Ní Chiaráin and A. Ní Chasaide, "The Digichaint interactive game as a virtual learning environment for Irish," *EUROCALL 2016*, Limassol, Cyprus, pp. 330-336, 2016.

[26] N. Ní Chiaráin and A. Ní Chasaide, "Chatbot technology with synthetic voices in the acquisition of an endangered language: motivation, development and evaluation of a platform for Irish," *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, Portorož, Slovenia, pp. 3429-3435, 2016.

[27] C. Gobl, "The Voice Source in Speech Communication: Production and Perception Experiments Involving Inverse Filtering and Synthesis," PhD thesis, KTH, Stockholm, Sweden, 2003.

[28] C. Gobl, "Modelling aspiration noise during phonation using the LF voice source model," *Proceedings of the 8th International Conference on Spoken Language Processing, INTERSPEECH 2006*, Pittsburgh, Pennsylvania, pp. 965-968, 2006.

[29] C. Gobl, "Reshaping the transformed LF model: generating the glottal source from the waveshape parameter R_d, *INTERSPEECH 2017*, Stockholm, Sweden, pp. 3008-3012, 2017.

[30] M. Swerts and R. Veldhuis, "The effect of speech melody on voice quality," *Speech Communication*, vol. 33, pp. 297-303, 2001.

[31] J. Dalton, J. Kane, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "GlóRí - the Glottal Research Instrument", *Proceedings of the 7th International Conference on Speech Prosody,* Dublin, Ireland, pp. 944-948, 2014.

[32] C. Gobl, E. Bennett, and A. Ní Chasaide, "Expressive synthesis: how crucial is voice quality," *Proceedings of the IEEE Workshop on Speech Synthesis*, Santa Monica, California, paper 52, pp. 1-4, 2002.

[33] A. Ní Chasaide and C. Gobl, "Voice quality and the synthesis of affect," in E. Keller, G. Bailly, A. Monaghan, J. Terken and M. Huckvale (Eds.) *Improvements in Speech Synthesis,* Wiley and Sons, pp. 252-263, 2002.