INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Time to Frequency Domain Mapping of the Voice Source: the Influence of Open Quotient and Glottal Skew on the Low End of the Source Spectrum

Christer Gobl, Ailbhe Ní Chasaide

Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences
Trinity College Dublin, Ireland
`cegobl@tcd.ie, anichsid@tcd.ie`

Abstract

This paper explores the mapping of time and frequency domain aspects of the voice source, focussing on the low end of the source spectrum. It refines and extends an earlier study, where the LF model was used to explore the correspondences between the open quotient (O_q), glottal skew (R_k) and harmonic levels of the source spectrum, including the $H1$-$H2$ measure, widely assumed to reflect differences in O_q. Here we use a different model (the F-model) as it better reflects the effective open quotient and glottal skew in certain conditions. As in the earlier study, a series of glottal pulses were generated, keeping peak glottal flow constant, while systematically varying O_q and R_k. Results suggest that the effects of R_k on the low harmonics is considerably less than estimated in the earlier study, and its main impact is on the level of $H2$ (and consequently $H1$-$H2$) when O_q is relatively high. The conclusion remains that the $H1$-$H2$ is not simply a direct reflection of O_q. However, for O_q values of up to about 0.6, it maps closely to $H1$-$H2$: beyond this point, $H1$-$H2$ reflects a more complex interaction of open quotient and glottal skew.

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

1. Introduction

The dynamic modulation of the voice source is a fundamental aspect of speech communication which shapes two essential dimensions of prosody. Firstly, it is an integral part of the linguistic prosody, such as the variations in accentuation, prominence, declination and phrasing that cue information structure. This allows the listener to segment the stream of speech for words and phrases, to identify the important items to attend to in an utterance, etc. The role of voice source modulation in these aspects of linguistic prosody is explored in [1-4]. Secondly, voice modulations carry the paralinguistic prosody, which signals interpersonal information concerning the speaker's state (mood and emotion), attitude to the interlocutor and to the discourse context. These aspects are explored in [5-8].

Despite its communicative importance, this aspect of speech is poorly understood – largely due to the difficulties in obtaining reliable measures of the voice source. The gaps in our understanding have many implications for speech technology. For example, having developed synthetic voices for dialects of Irish [9, 10], the plan to deploy them in interactive educational games and dialogue systems [11, 12] will require being able to approximate the essential linguistic and paralinguistic aspects of voice prosody.

Parameters used in voice analysis are defined either in the time domain or in the frequency domain. Each yields particular insights: time domain measures relate closely to production aspects of the voice, while frequency domain measures relates more closely to perception [13]. Time domain parameters are difficult to estimate reliably, and being phase sensitive, require stringent recording conditions. (For a discussion of analysis difficulties, see [14]). Factors such as these mean that more easily obtained spectral measures are often used as proxy measures of the voice source.

The $H1$-$H2$ spectral measure is widely used, often assumed to be an indicator of the pulse open quotient (O_q) [15], and taken as a measure of breathiness [16-21]. An increase in O_q generally leads to an increase in the amplitude of $H1$, thus increasing $H1$-$H2$ [22]. (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 [23, 15].)

Ideally, one would wish to be able to map reliably between the time and frequency domains to characterise the source (see, for example [21, 13, 24, 25, 26]). In this paper, we build on an earlier study [26], which focussed on the low end of the source spectrum, and which used LF model [27] simulations to explore the correlation of time domain parameters with the levels of the first three harmonics. The time domain parameters considered included the peak flow (U_p) – described by Fant and Lin [24] as the principle determinant of $H1$ – and the pulse shape parameters open quotient (O_q) and glottal pulse skew (R_k). O_q is the duration of the open phase as a proportion of the glottal period ($O_q = (t_p - t_o)/T_0$) and R_k is a measure of glottal pulse symmetry, given by the duration of the closing branch of the pulse relative to the duration of the opening branch. Thus, a lower R_k value corresponds to a more skewed pulse ($R_k = (t_e - t_p)/(t_p - t_o)$, see Figure 1).

By maintaining a constant U_p and by varying O_q and R_k systematically, the aim was to illuminate how these parameters affect $H1$ and $H2$, as well as the $H1$-$H2$ measure. Although O_q emerged as the main determinant of the amplitudes of $H1$ and $H2$, there was a strong influence of glottal skew, particularly on $H2$. With regard to $H1$-$H2$, although a broad correspondence with O_q, emerged, glottal skew was found to have a strong effect, particularly at high O_q values. It is clear that these effects can potentially invalidate direct inferences from $H1$-$H2$ on O_q and on voice quality.

However, a re-examination of the LF model simulations in [26] lead us to reconsider these findings, and motivate the present study. As illustrated in Figure 1, an LF pulse with very high O_q (0.85) and high skew (low R_k of 0.15) yields an effective O_q that is considerably lower than the numerical value of the model (0.5 rather than 0.85). This is due to the fact that during the initial part of the opening phase of the LF pulse, the flow is very close to zero – a consequence of the exponentially

http://dx.doi.org/10.21437/Interspeech.2019-2888

growing sinusoidal segment used to model the open phase in the LF model (for details, see further [26-29]).

In Figure 1, an estimate of the *effective* open quotient (OQ_e) and skew (RK_e) for these LF settings is obtained by fitting a triangular pulse to the LF glottal pulse. This triangular pulse is made up of two line segments. The line for the opening branch intersects with the model at the point of the maximum slope and at the point of peak glottal flow. Similarly, the line for the closing branch intersects with the point of maximum slope and the peak glottal flow. There are of course many possible stylisations, but this simple stylisation is deemed adequate to capture the effective open quotient.

As already pointed out, Figure 1 (left panel) highlights how much smaller the effective open quotient ($OQ_e = 0.50$) of the LF pulse is than its numeric value ($O_q = 0.85$) would imply. Differences also emerged in the glottal skew, where the effective R_k values are higher than the specified values in the model.

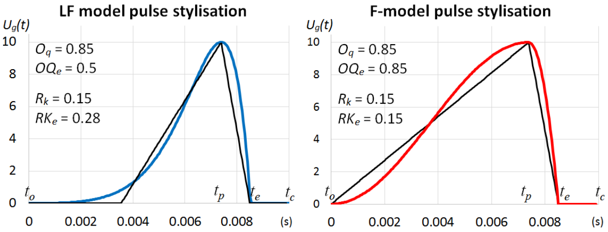

Figure 1: *LF model pulse (left panel) and F-model pulse (right panel) with triangular stylisations.*

For the reanalysis in this paper, we used the three-parameter model described in [30, 31] (see also the specification of the model in Section 2 below). In [27] this model is referred to as the *F-model*, a term which is adopted henceforth.

A similar stylisation of the glottal pulse for the F-model is also shown in Figure 1 (right panel) using the same settings for O_q and R_k as were used for the LF pulse (left panel). As can be observed there, the F-model generates pulses (red) where O_q and R_k values are always the same as the effective open quotient and glottal skew values of the stylised pulse (black). It is striking, for the same O_q and R_k settings, how divergent the effective open quotients are in the two models in this case.

2. Methodology

In [26] we used the LF voice source model to analyse the relationships between the glottal pulse shape and the amplitudes of source spectral components. However, for reasons outlined above, we here carry out essentially the same analysis, but instead using a different source model.

Thus, we used the F-model to generate a range of glottal pulses for which the open quotient and glottal pulse skew were varied in controlled steps by changing the O_q and R_k parameters, while the peak amplitude ($U_p = 10$) and pulse duration were kept constant.

We also reanalysed the data from our earlier study [26] using OQ_e and RK_e for the LF model (instead of O_q and R_k) for the correlation with $H1$ and $H2$ amplitudes.

The F-model (see Figure 1, right panel) is determined by the two expressions in (1), which generate the glottal pulse of the opening phase and closing phase respectively. The model does not include a return phase after the main excitation.

$$U_g(t) = \begin{cases} \frac{U_p}{2}\left(1 - \cos\omega_g t\right) & t_o \le t \le t_p \\ U_p\left(1 - K\left(1 + \cos\omega_g t\right)\right) & t_p < t \le t_e \\ 0 & t_e < t < t_c \end{cases} \tag{1}$$

In addition to f_0, the F-model has three parameters: U_p, ω_g and K. $\omega_g = 2\pi F_g$ where F_g is the characteristic frequency of the pulse. K determines the rate by which the flow drops during the closing branch of the pulse. Neither O_q nor R_k are parameters of the F-model, but ω_g and K can be derived from O_q and R_k using the following formulas: $\omega_g = \pi f_0 O_q^{-1}(1 + R_k)$ and $K = (1 - \cos \pi R_k)^{-1}$.

As in [26], glottal pulses were generated with nine different O_q settings, ranging from 0.15 to 0.95 in steps of 0.1. For each O_q setting, nine pulses with different R_k values were generated. The R_k values also ranged from 0.15 (high skew) to 0.95 (low skew) in steps of 0.1, thus covering most of the possible R_k range.

The sampling frequency used was 20 kHz, which ensured that the effect of aliasing on the lower end of the source spectrum would be negligible. Each pulse was repeated five times in order to produce a harmonic spectrum. For each of the 81 glottal waveforms, a 1000-point (50 ms, rectangular window) DFT spectrum was calculated and the amplitudes of the first two harmonics were extracted. The window length was chosen so that the output frequency samples would coincide with the harmonic frequencies, thus avoiding potential rounding errors.

3. Results

The left panels of Figure 2 show the $H1$ and $H2$ amplitudes as a function of O_q and R_k, when derived from the F-model. The previously reported estimates [26] using the LF model are shown in the mid panels. Shown in the right panels are the values obtained for the LF model when OQ_e and RK_e are used instead of O_q and R_k for the mapping to $H1$ and $H2$. OQ_e and RK_e are the effective values of the open quotient and glottal skew respectively, according to the stylisation in Figure 1.

It is clear that the effects of R_k on the $H1$ and $H2$ amplitudes are reduced when the source spectrum is derived from the F-model rather than from the LF model. Note for $H1$ that the R_k-lines are much closer together, and while there is still some R_k influence, the differences are very small compared to the differences found for the LF model spectra, particularly when R_k is between 0.15 and 0.35. Note also that in the F-model spectra the peak in $H1$ occurs close to $O_q = 0.79$ regardless of R_k, quite unlike in the LF model spectra where the peak occurs at increasing O_q values as R_k decreases.

Similarly for $H2$, the effects of R_k are much smaller in the F-model spectra compared to the LF model spectra. The peak $H2$ values for the F-model spectra occur consistently around $O_q = 0.37$, which is strikingly different to what was previously found in the LF model spectra (lower mid panel) where the peak in $H2$ occurs at increasing O_q values when R_k is reduced.

All in all, what emerges is a simpler and more predictable correlation of the time and frequency dimensions. Nonetheless, there are still some clear differences in $H2$ levels, depending on the degree of glottal skew, particularly for higher O_q values.

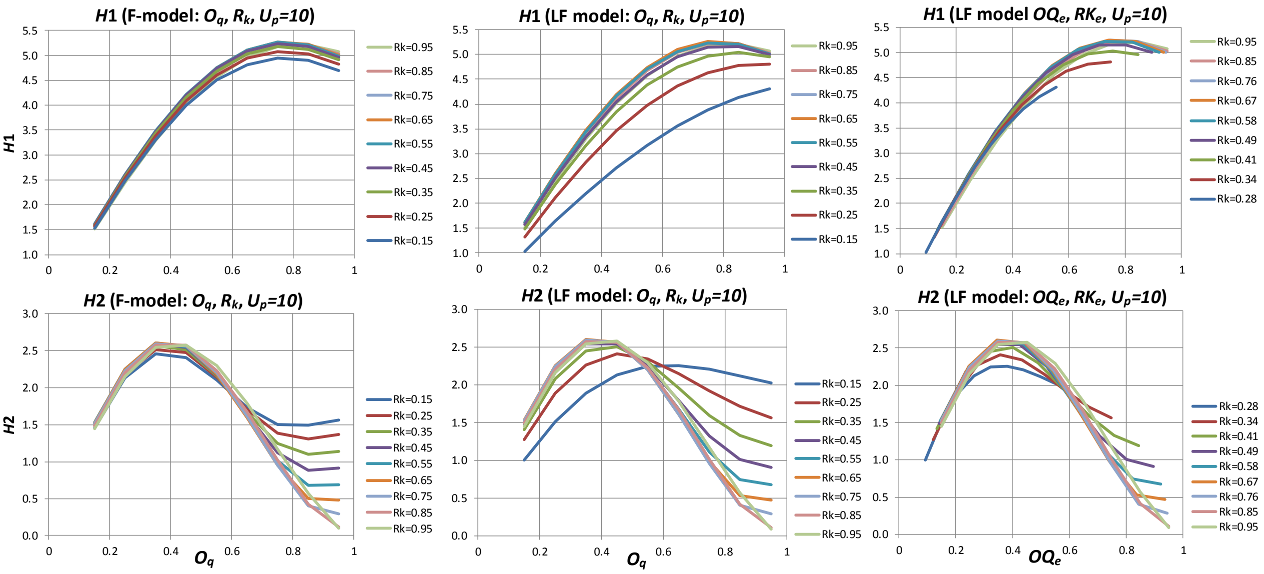

Figure 2: *Variation in H1 and H2 as a function of O_q and R_k for the F-model (left panels) and for the LF model (mid panels). The right panels show H1 and H2 variation as a function of the effective open quotient and glottal skew measures OQ_e and RK_e (see Figure 1) for the LF model pulses. U_p = 10 in all cases.*

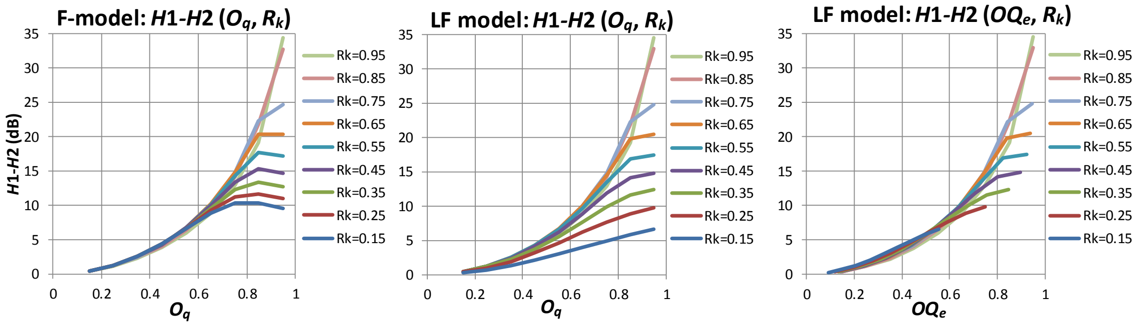

Figure 3: *Variation in H1-H2 as a function of O_q and R_k for the F-model (left panel) and for the LF model (mid panel). The right panel shows H1-H2 variation as a function of the effective open quotient and glottal skew measures OQ_e and RK_e for the LF model pulses.*

In the right panels of Figure 2, the replotted LF model data for *H*1 and *H*2 are shown. Using the effective values of open quotient and glottal skew (OQ_e and RK_e) the *H*1 and *H*2 variation that emerges is much more like that found for the F-model.

3.1. *H*1-*H*2 as a measure of the open quotient

Figure 3 illustrates the *H*1-*H*2 variation as a function of O_q and R_k derived from the F-model spectra (left panel) and from the LF model spectra (mid panel). The right panel of Figure 3 shows the *H*1-*H*2 variation as a function of OQ_e and RK_e derived from the LF model. In the case of the spectra derived from the LF model, the *H*1-*H*2 measure is highly correlated with O_q, but there is also a strong interaction with R_k, as shown previously also in [26, 32, 33]. In comparison, for the F-model spectra, the *H*1-*H*2 value is essentially determined by O_q up to about $O_q = 0.6$, with little influence of R_k. For higher O_q values one again notes the strong interaction with R_k. What this effectively means is that if the *H*1-*H*2 value is below 8 dB, it can be reasonably assumed to reflect O_q differences. For level differences above 8 dB these simulations imply a combined influence of O_q and R_k: unless one is given, the other cannot be inferred. As in Figure 2, we note that when replacing original O_q and R_k values with the effective values OQ_e and RK_e, the LF model data mirror the F-model data more closely.

4. Estimating the correction factor, *k*

As discussed in Fant and Lin [24], the amplitude of *H*1 is proportional to U_p and can be derived according to (2), where $|R(f)|$ represents the radiation function and *k* is a correction factor which depends on the glottal pulse shape.

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

In [26] we used spectral measurements from the LF model simulations to determine the *k* factor. It was found that the variation in *H*1 due to changes in O_q closely follows part of a parabolic curve. Thus, a quadratic function was presented for the prediction of the *k* factor according to:

$$k = a_2 O_q^2 + a_1 O_q + a_0 \tag{3}$$

The coefficients of the polynomial in (3) depend on R_k, and the variation in the three *a*-coefficients were shown to also closely match quadratic functions. Thus, the *a*-coefficients can be determined by the expressions in (4).

$$\begin{aligned} a_2 &= b_2 R_k^2 + b_1 R_k + b_0 \\ a_1 &= c_2 R_k^2 + c_1 R_k + c_0 \\ a_0 &= d_2 R_k^2 + d_1 R_k + d_0 \end{aligned} \tag{4}$$

If we now use instead the spectral measurements from the F-model, we find that the expressions in (3) and (4) are still valid ($R^2 = 0.97$ or higher). The values of the nine coefficients in (4), which are used to calculate the *a*-coefficients, are shown in Table 1.

Table 1: *Values of the coefficients for the equations in (4) derived from the F-model voice source data.*

b_2	b_1	b_0
1.04	–1.03	–1.58
c_2	c_1	c_0
–1.17	1.27	2.52
d_2	d_1	d_0
0.0505	–0.0619	–0.0567

Note, however, that the R_k-dependent variation in the *a*-coefficients is quite different when the data are based on the F-model: the variation is much smaller, particularly for a_1 and a_2 (see Figure 4). It is therefore feasible to use constant *a*-coefficients, independent of R_k. By calculating the mean values over the full R_k range we get the constants shown in (5), which can be used without losing too much accuracy in the calculation of the *k* factor.

$$\begin{aligned} a_2 &= -1.8 \\ a_1 &= 2.8 \\ a_0 &= -0.07 \end{aligned} \tag{5}$$

To test the accuracy of the of *k* correction factor as determined by (3) and (4), the *H*1 prediction errors were calculated on a different set of 64 F-model pulses, where both O_q and R_k varied from 0.2 to 0.9 in steps of 0.1. The correlation between the estimated and actual *H*1 amplitudes is very high ($R^2 = 0.9995$) with mean and maximum errors in the *H*1 estimate being 0.07 dB and 0.28 dB respectively. Using the constant coefficients in (5), the corresponding error values are 0.19 dB (mean error) and 0.47 dB (max error) and $R^2 = 0.994$.

5. Conclusions

This paper refines and extends an earlier study [26] concerning the mapping of time to frequency domain measures of the voice source. It is found that the influence of the glottal skew on the first two harmonics is not as great as was thought in [26], but was overestimated due to the way the open phase of the LF model is generated, which results in lower *effective* open quotient values in certain conditions. In the present revised estimates, it emerges that *H*1 can be determined reasonably accurately from U_p and O_q, independently of glottal skew. The appropriate *k* factor for estimating *H*1, based on these revised calculations is presented.

*H*2 is also less dependent on the glottal skew than was thought in [26], but there is still a clear influence, particularly when the open quotient is high.

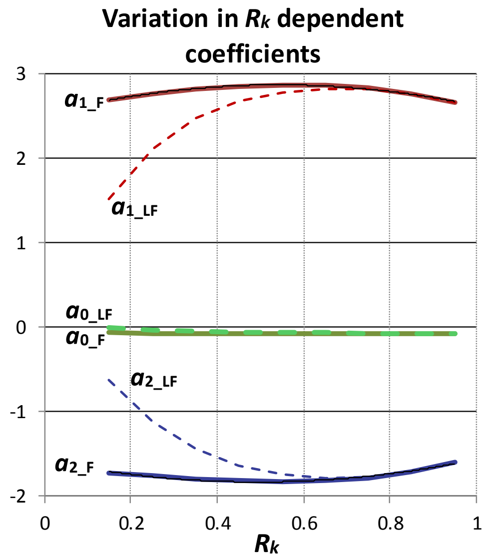

Figure 4: *Variation in the R_k dependent a_2, a_1 and a_0 coefficients of (4) for the F model (solid lines) and the LF model (dashed lines).*

With regard to *H*1-*H*2, when the difference is below about 8 dB, the correlation with O_q is straightforward, and there is little influence of glottal skew. When the difference exceeds this value, there is a clear interaction with the pulse skew, where increasing skew counteracts the effects of increasing open quotient.

Thus, in broad terms our findings agree with previous studies in that *H*1-*H*2 is not uniquely attributable to O_q, but reflects a more complex interaction of open quotient and glottal skew. However, the present results show a more nuanced picture of this interaction, showing that the influence of glottal skew may in fact be negligible when the open quotient is relatively low.

This study draws attention to differences between voice source models in the way that parameters such as O_q and R_k are realised, which can yield differences in their effective values. It may therefore be useful to work with effective measures of open quotient and glottal skew in controlling the source model. For a comparison of different measures of open quotient, see [34].

The focus here is limited to the parameters that shape the low end of the source spectrum. This is thus only a small step towards a broader enterprise, which is the overall mapping of the time-to-frequency correspondences in the glottal source. Future work will address parameters that govern the mid and upper parts of the source spectrum.

To conclude, an eventual mapping between time and frequency dimensions of the source would confer many benefits. It would extend our understanding of the production and perception of voice quality and bring us closer to developing more robust voice source analysis techniques, crucial to the goal of modelling voice prosody in human speech and replicating it in technology.

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] 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.

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

[3] 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, 2013, pp. 3527-3531.

[4] I. Yanushevskaya, A. Murphy, C. Gobl and A. Ní Chasaide, "Perceptual salience of voice source parameters in signaling focal prominence," *INTERSPEECH 2016*, San Francisco, California, 2016, pp. 3161-3165.

[5] 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.

[6] I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "Cross-language differences in how voice quality and f0 contours map to affect," *Journal of the Acoustical Society of America*, vol. 144, pp. 2730-2750, 2018.

[7] I. Yanushevskaya, C. Gobl and A. Ní Chasaide, "Voice parameter dynamics in portrayed emotions," *Proc. 6th Int. Workshop: Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA*, Florence, Italy, 21-24, 2009.

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

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

[10] 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, 2017, pp. 2113-2117.

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

[12] 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, 2016, pp. 3429-3435.

[13] J. Kreiman, B. R. Gerratt, M. Garellek, R. Samlan, Z. Zhang, "Toward a unified theory of voice production and perception," Loquens 1, e009, 2014

[14] 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.

[15] M. Iseli, Y-L Shue, and A. Alwan, "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *Journal of the Acoustical Society of America*, vol. 121, pp. 2283-2295, 2007.

[16] 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.

[17] M. K. Huffman, "Measures of phonation type in Hmong," *Journal of the Acoustical Society of America*, vol. 81, pp. 495-504, 1987.

[18] 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.

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

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

[21] 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," *Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2625-2632, 2012.

[22] D. H. Klatt and L.C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers," *Journal of the Acoustical Society of America*, vol. 87, pp. 820-856, 1990.

[23] H. M. Hanson, "Glottal characteristics of female speakers: acoustic correlates," *Journal of the Acoustical Society of America*, vol. 101, pp. 466–481, 1997.

[24] 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, vol. 4, pp. 1-21, 1988.

[25] 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, 1999, pp. 2437-2440.

[26] C. Gobl, A. Murphy, I. Yanushevskaya, and A. Ní Chasaide, "On the relationship between glottal pulse shape and its spectrum: correlations of open quotient, pulse skew and peak flow with source harmonic amplitudes," *INTERSPEECH 2018*, Hydrabad, India, 2018, pp. 222-226.

[27] 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, vol. 4, pp. 1-13, 1985.

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

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

[30] G. Fant, "Glottal source and excitation analysis," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 20 (1), pp. 85-107, 1979.

[31] G. Fant, "Vocal source analysis – a progress report," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 20 (3-4), pp. 31-53, 1979.

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

[33] B. Doval, d'Alessandro, C. and N. Henrich, "The spectrum of glottal flow models," Acta Acustica united with Acustica, vol. 92, pp. 1026-1046, 2006.

[34] J. Kane, S. Scherer, L.-P. Morency, and C. Gobl, "A comparative study of glottal open quotient estimation techniques. *INTERSPEECH 2013*, Lyon, France, 2013, pp. 1658-1662.