INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Cross-speaker Variation in Voice Source Correlates of Focus and Deaccentuation

Irena Yanushevskaya, Ailbhe Ní Chasaide, Christer Gobl

Trinity College Dublin, Ireland

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

Abstract

This paper describes cross-speaker variation in the voice source correlates of focal accentuation and deaccentuation. A set of utterances with varied narrow focus placement as well as broad focus and deaccented renditions were produced by six speakers of English. These were manually inverse filtered and parameterized on a pulse-by-pulse basis using the LF source model. Z-normalized F0, EE, OQ and RD parameters (selected through correlation and factor analysis) were used to generate speaker specific baseline voice profiles and to explore cross-speaker variation in focal and non-focal (post- and prefocal) syllables. As expected, source parameter values were found to differ in the focal and postfocal portions of the utterance. For four of the six speakers the measures revealed a trend of tenser phonation on the focal syllable (an increase in EE and F0 and typically, a decrease in OQ and RD) as well as increased laxness in the postfocal part of the utterance. For two of the speakers, however, the measurements showed a different trend. These speakers had very high F0 and often high EE on the focal accent. In these cases, RD and OQ values tended to be raised rather than lowered. The possible reasons for these differences are discussed.

Index Terms: voice source, focus, accentuation, deaccentuation, cross-speaker variation

1. Introduction

As part of a broader examination of the voice source correlates of linguistic prosody, previous studies have examined the detailed voice source correlates of accentuation [1, 2], and focal accentuation [3-5]. Accented syllables tend towards tenser phonation setting than unaccented syllables [2]. In the case of focal accentuation rather similar switches in phonation mode were found, but affecting the entire realization of the utterance. The parameter values for the focally accented syllable exhibited shifts suggesting a rather tenser mode of phonation, while the postfocal tail exhibited shifts indicative of increasingly lax phonation [3, 4].

Broadly speaking, these voice source measurements accord with findings in a number of other studies where spectral measurements of the speech waveform appeared to indicate increased phonatory tension associated with accentuation [6-8] and with focal accentuation. However, not all studies are unanimous on this matter. An analysis of focal accentuation of Finnish [9] suggests focal accentuation to be associated with relatively more lax phonatory settings – a finding that runs directly counter to our previous findings and those of other researchers. It is difficult to comment on the observed differences: as the studies are of different languages and employ rather different methodologies, one must be cautious in drawing conclusions. On the one hand, one could be dealing with cross-language differences. On the other hand, it is possible that differences reported could be due to differences in the methodologies used. Or, perhaps as suggested by [10] the differences might arise out of interactions between F0 (especially in the high F0 range) and other source parameters. Effectively, outside certain ranges of F0, the relationship of certain source measures and auditory phonatory tenseness/laxness may not hold, and are ripe for misinterpretation.

Our earlier studies were primarily based on careful manual (pulse by pulse) inverse filtering of the speech data, with similar manual source parametrization using the LF voice source model [11] (see below). Due to the considerable time it takes to generate such data, the analyses focus on limited materials, often based on a single speaker. Such studies yield rich insights into the underlying control mechanisms that may be involved, but leave open the question of whether the precise mechanisms observed are generalized to the population.

Variations in the data led to the Voice Prominence Hypothesis (also referred to as the ‘six of one, half a dozen of the other’ hypothesis) [2]. It proposes that prosodic prominence (whether to differentiate accented from unaccented syllables generally or to effect focus on a particular item in an utterance) can be achieved by different adjustments to the voice source parameters, i.e. parameters like F0, EE, OQ, RA, and RD (i.e. affecting the pitch, strength and phonatory quality of the voice) working synergistically. It further proposes that the extent to which one or other parameter is exploited is likely to vary, and that greater exploitation of one parameter will tend to entail the lesser use of another – essentially a trading relationship. Finally, it proposes that this variation may depend on speaker-specific strategies, and prosodic factors (e.g., nuclear vs. prenuclear accentuation) which may constrain their operation.

In the earlier studies of focal accentuation [3, 5], given the very labor intensive nature of manual pulse by pulse inverse filtering and model matching, a very limited dataset based on a single speaker was used. The present paper is also based, like many of the earlier studies, on manually analyzed (inverse filtered and source parameterized) data, but uses a dataset of six speakers to focus particularly on cross-speaker variation. Two aspects are of importance. Firstly, we are interested in the intrinsic differences in the individual speakers’ baseline voices as this is likely to have an impact on how they use their voice in prosodic signaling. Secondly, and of primary interest here, we look at how speakers differ in terms of their use of F0 and other source parameters to mark differential salience of focal and postfocal portions of the utterance.

2. Materials and method

This section describes the data collection, the analysis methods (inverse filtering and model matching procedure) and the approaches adopted for data normalization and representation.

http://dx.doi.org/10.21437/Interspeech.2017-1535

2.1. Speech material

The all-voiced utterance 'We were away a year ago' was elicited with different focal placement from six male speakers of English using short dialogues. The realizations included broad focus (BR), narrow focus on the potentially accentable syllables WE, WERE, WAY and YEAR, as well as a deaccented rendition (DEAC). The narrow focus utterances were realized with falling and rising pitch on the focally accented syllables. The recording was done in a semi-anechoic room directly to a PC, using a Brüel & Kjær microphone and amplifier (B&K 4191 and B&K Nexus 2690), with the sampling frequency of 44.1 kHz. The distance to the microphone and the amplification were kept the same for all speakers. This recording setup ensures a linear phase response as well as negligible amplitude distortion and noise. Prior to the inverse filtering the utterances were highpass filtered at 50 Hz and downsampled to 10 kHz. In total, six speakers recorded 10 utterances each: BR, DEAC, (WE, WERE, WAY, YEAR) x 2 (Fall, Rise). One speaker was not able to produce rising pitch naturally (rising pitch is in fact uncommon in Southern Irish English varieties [12]), so the utterances with rising pitch were excluded for this speaker. Overall, the dataset comprised 56 utterances (6 speakers × 6 utterances × 1 falling pitch + 5 speakers × 4 utterances × 1 rising pitch).

2.2. Source analysis and parameterization

The utterances were analyzed using the software system described in [13]. First, semi-automatic inverse filtering was carried out, based on closed-phase covariance LPC. Subsequently, the inverse filtering was fine-tuned manually, pulse by pulse, to achieve the best possible source approximation. Voice source parameterization involved fitting the LF (Liljencrants-Fant) model of differentiated glottal flow [11] to the source signal derived from the inverse filtering. From this source modelling, the following voice source measures were derived: F0, EE, UP, RK, RG, OQ, RA, FA and RD. These parameters were used in earlier studies; they are briefly described in Table 1. For further details, see [14-16].

Parameter values were smoothed using a moving average filter spanning three pulses. Since one of the speakers did not produce utterances with rising pitch, only the utterances with falling pitch were selected for further analysis here (6 speakers x 6 utterances = 36 utterances; 4327 glottal pulses). To compare the voice source analysis data across speakers, z-score normalization was used.

2.3. Selection of source parameters: factor analysis

Prior to cross-speaker comparison, a principal component analysis (PCA) with varimax rotation was conducted to establish the underlying parameter grouping in order to reduce the number of parameters and to identify the parameters most implicated in focal signaling. The correlation matrix including the eight source parameters F0, EE, RA, FA, RG, RK, OQ, UP and RD (see Table 2) was inspected. Parameters RK, FA, UP and RG showed multiple correlations below 0.3 or above 0.9 and so were excluded from the subsequent factor analysis, as recommended by [17]. F0 showed low correlation with other source parameters, but was included in the PCA as it is a commonly studied measure in the prosodic analysis of focus.

Factor loadings after rotation are shown in Figure 1. Results suggest that source variation in the analyzed utterances can be described in terms of two underlying components that account for 83% of the variance. RA, OQ, RD have high loadings on the same component (component 1) and reflect source variation along the tense/lax continuum; F0 is heavily loaded on factor 2 (frequency). RA was highly correlated with RD and OQ, and was excluded. The final set of parameters used in this study comprises EE, F0, OQ and RD.

Table 1*: Source parameters.*

	Description
F0	The voice fundamental frequency, 1/T0 where T0 is the fundamental period, i.e. the duration of one glottal cycle.
EE	The strength of the main excitation during the glottal cycle, defined as the negative amplitude of the differentiated glottal flow at the maximum waveform discontinuity.
RA	The normalized effective duration of the return phase of the glottal pulse after the main excitation. RA relates to the source spectral slope (increased RA = greater spectral slope).
FA	The frequency characteristics of the exponential function of the return phase are approximately those of a first order low pass filter. The cutoff frequency, FA, is inversely correlated with the amount of dynamic leakage: FA = F0/(2πRA).
RG	The glottal frequency, FG, normalized to F0, where FG is the characteristic frequency of the glottal pulse during the open phase of the glottal cycle. Mainly affects the relative amplitudes of the first two harmonics of the source spectrum.
RK	A measure of the glottal pulse skew: the smaller the RK value, the more asymmetrical the glottal pulse.
OQ	The duration of the glottal open phase relative to the duration of the whole glottal cycle. Mainly affects the lower components of the source spectrum.
UP	The peak amplitude of the glottal flow pulse.
RD	A global waveshape parameter derived from F0, EE and UP as follows: (1/0.11) · (F0·UP/EE).

Table 2: *Parameter correlations (initial analysis).*

	EE	F0	RA	FA	RG	RK	OQ	UP
F0	.34	1.00						
RA	-.35	.29	1.00					
FA	.40	.09	-.81	1.00				
RG	.47	-.31	-.38	.19	1.00			
RK	-.18	-.14	.08	-.12	.25	1.00		
OQ	-.58	.31	.46	-.26	-.95	.02	1.00	
UP	.95	.23	-.19	.19	.46	.001	-.50	1.00
RD	-.51	.28	.86	-.68	-.54	.41	.71	-.32

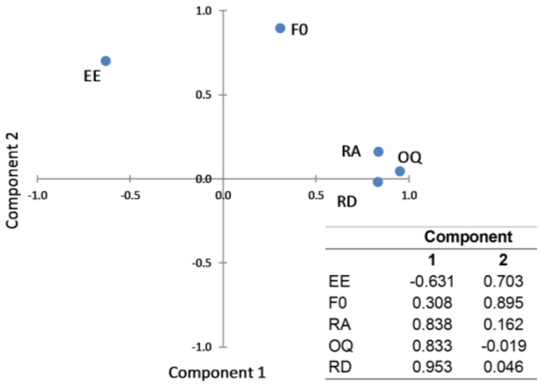
3. Results and discussion

3.1. Speaker-specific baseline voice profiles

To examine global trends in the voice source correlates of focus and deaccentuation across the six speakers, mean values were calculated for z-normalized parameter levels in three sections of each utterance: the prefocal material, the focally accented syllable and the postfocal material.

First, speaker specific baseline voice profiles were plotted using the average z-normalized parameter values relative to the group mean (Figure 2, left panels). The group mean is the zero line – shown in Figure 2 as a black dotted line. The mean parameter levels for each speaker relative to the group mean are shown as a blue solid line. Table 3 shows non-normalized values (means and standard deviations) for each speaker.

In the left panels in Figure 2, where each speaker's baseline voice profile is shown, we note some striking differences among them. Speaker TO has a very low F0 and rather tense phonatory settings (low OQ and RD). Speaker BO has similarly low F0 but otherwise phonatory settings that are very much



	Component	
	1	2
EE	-0.631	0.703
F0	0.308	0.895
RA	0.838	0.162
OQ	0.833	-0.019
RD	0.953	0.046

Figure 1. *Factor loadings of parameters after rotation.*

Table 3: *Mean and standard deviation (in brackets) parameter values for individual speakers and group.*

Speaker	EE (dB)	F0 (Hz)	OQ (%)	RD
TO	63.4 (3.8)	93 (9)	46.5 (4.9)	0.85 (0.2)
BO	61.4 (3.4)	94 (9)	54.6 (6.3)	1.03 (0.22)
SL	59.1 (5.3)	118 (13)	60.7 (4.5)	1.03 (0.19)
JK	68.4 (4.9)	116 (12)	50.3 (4.6)	0.86 (0.15)
LP	65.6 (3.8)	121 (16)	55.1 (7.1)	1.05 (0.23)
JD	61.1 (3.6)	115 (19)	57.7 (5.9)	1.07 (0.23)
Group	63.1 (5.3)	110 (17)	54.4 (7.3)	0.98 (0.22)

at the average for this group. Speaker SL has a higher OQ and lower EE suggesting a more lax baseline phonatory setting, and a rather weak glottal excitation. Speaker JK has a relatively tense phonatory setting with high EE, low OQ and RD. Speakers LP and JD have setting that are fairly close to the average for this group (the dotted zero-line), but differ in that LP has a slightly raised EE and F0, whereas speaker JD veers slightly towards breathier phonation (raised RD and OQ).

3.2. Parameter levels in focal and postfocal syllables

In the rightmost panel of Figure 2 are shown for each speaker the parameter levels in the focally accented syllable (red) relative to the postfocal portion of the utterance (grey) for the utterance 'We were aWAY a year ago' (focal accent on WAY). As in the left panel, the z-normalized mean values are shown.

Cross-speaker differences clearly emerge in the realization of focus, as can be seen in the relationship of the red and grey outlines. For speaker TO, who has an intrinsically tense phonatory setting, focalization appears to involve relative shifts in the phonatory tension setting in the utterance, and particularly a more lax phonatory setting in the postfocal material where EE is lowered, OQ and RD are raised and F0 slightly lowered. For speaker BO, the effective salience contrast entails similar parameter differences, but relative to this speaker's baseline setting, entail changes both to the focal and postfocal portions: increased tension settings in the focally accented syllable (raised EE, lowered OQ and RD and raised F0), and shifts in the opposite direction in the postfocal material. The overall effect on the utterance in terms of the balance of salience is very much as for TO – with phonation strength, tension settings and F0 diverging in focal and postfocal material.

For speaker SL, there are particularly dramatic shifts in the EE parameter, which is both relatively raised in the focally accented syllable and relatively very much weaker in the postfocal part of the utterance. Though less dramatic, there is also salience-lending differentiation in the OQ, RD and F0 values. For speaker JK, who has an intrinsically rather tense baseline, the contrast between the focal and the deaccentuated postfocal part of the utterance appears to rely to a large extend on EE differentiation, while there is also some difference in F0. For speaker LP, the differentiation seems most noticeable in the EE and F0 parameters, but with some slighter differences also in RD and OQ.

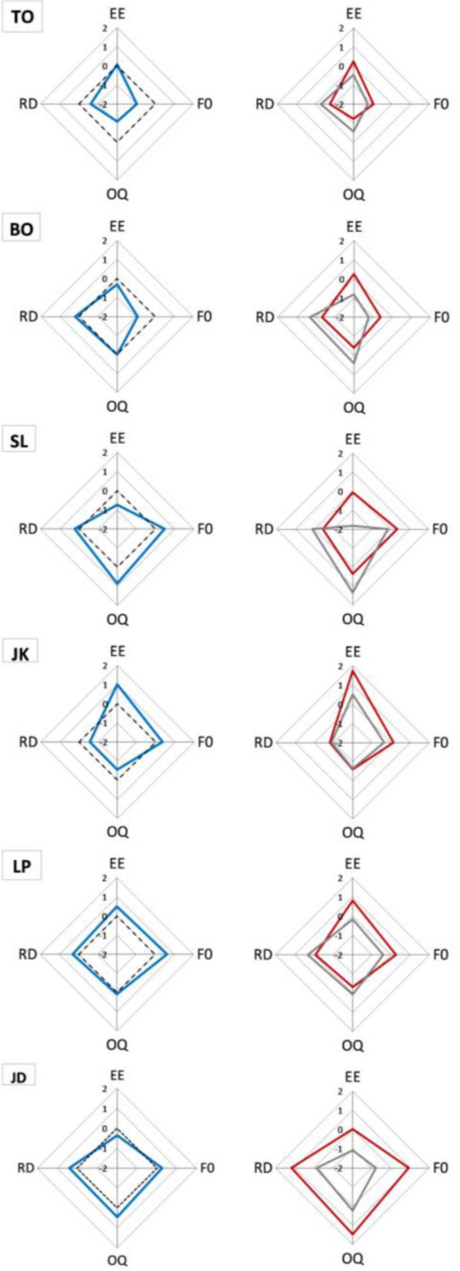


Figure 2: *Left panels: baseline voice profiles for individual speakers (blue) relative to the group mean (black dashed line); right panels: parameter levels in the focally accented WAY (red) and postfocal material (grey). Plotted are z-normalized mean values.*

The final speaker, JD shows the most striking divergence from the rest of the group. For JD, the contrast between the focally accented syllable and the following postfocal material entails a more extreme difference in F0. There is, as with the other speakers, a considerable differentiation in EE. Thus, the focus and the deaccented portions of the utterance are differentiated by the pitch and the strength of the source pulse. However, in contrast to the other speakers, RD and OQ values are markedly higher in the focal than in the postfocal part of the utterance. This appears to indicate a laxer mode of phonation in the focally accented syllable, something that runs counter to general trend for the other speakers.

There are different possible interpretations that might be offered here. It may simply be that some speakers use a relatively laxer phonation for focal accentuation whereas others employ a tenser phonation (see the Finnish study [9] mentioned above). The fact that this speaker relies more heavily on F0 differentiation in signaling focal accentuation/postfocal deaccentuation may mean that tension settings, otherwise important, become irrelevant.

It was pointed out in [10] that the parameters which are usually taken as indicative of phonatory tenseness/laxness may need to be interpreted with caution at very high F0 values, as the changes to the glottal pulse shape with very high F0 may entail that it no longer correlates in a straightforward way with phonatory (and auditory) tenseness. With the F0 values here such a factor may not hold, but this is an area where clearly more work is needed to establish the interactions between F0 and the glottal pulse shape parameters.

Figure 3 shows essentially the same information as Figure 2 (right panels): mean parameter levels for the six speakers in the focally accented syllable and postfocal deaccented syllables in a way that may make it easier to see the general trends and cross-speaker differences.

Overall, the speakers show similar changes in phonatory settings to signal focus and deaccentuation. The most striking parameter consistently associated with focalization/deaccentuation is EE, showing, as already observed in the individual cases, a strengthening of the glottal excitation in the focal syllable, and a relative attenuation in the postfocal material. Thus, the realization of focal accentuation in an utterance appears to consistently (for this group of speakers at least) entail modulation in the strength of the voice. Both laryngeal tension and respiratory effort are likely to contribute here in effecting this EE difference, to manipulate the relative salience of focal and postfocal material in the sentence. The differentiation of F0 is also consistent across the speakers. This is hardly surprising, as this is long established in the intonation literature on focus. However, substantial cross-speaker differences emerge in the extent to which F0 marks focus. Speaker TO appears to differentiate rather little on the basis of F0, whereas speaker JD differentiates a great deal – more than any of the other speakers. As for OQ and RD, generally interpreted as indicators of tenseness/laxness in the voice, the general trend for speakers TO, BO, SL and JK is for increased tension in the focally accented syllable, and a more lax setting postfocally. This pattern does not hold for two of the speakers. LP exhibits similar OQ and RD values both focally and postfocally.

As discussed above, the trend for JD is the converse of the trend observed for the first four speakers: the focally accented syllable has higher OQ and RD values. This would seem to point to a more lax phonation in the focally accented syllable (as suggested in [9] for Finnish). However, the fact that this counter-trend occurs with the speaker who has the highest F0 and the most extensive use of F0 in differentiating focally accented and postfocally deaccented parts of the utterance suggests that different factors may be at work. Given the extensive use of the F0 parameter in effecting the salience difference, and given the fact that large EE differences are also contributing, it may be the case that the tension setting differences associated with OQ and RD become relatively unimportant. Or it might be the case that F0 being high comes with a change to the shape of the source pulse, so that OQ and RD are less reliable indicators of phonatory tenseness/laxness [10].

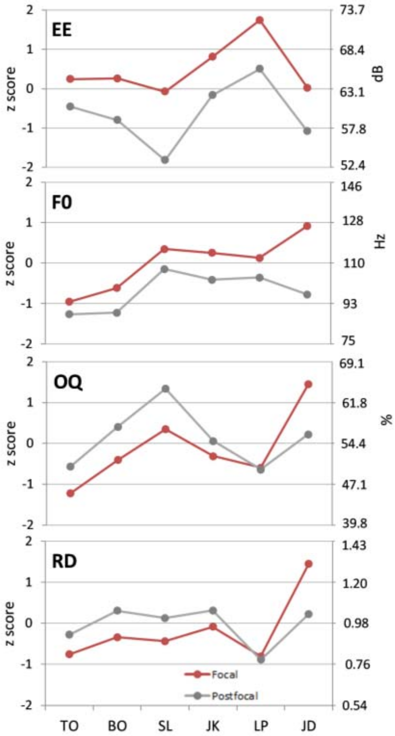


Figure 3: *Mean parameter levels in focally accented syllable WAY (red) and postfocal material (grey). Values are shown as z scores (left y axis) and as corresponding group mean +/- 2 SD values (right y axis).*

4. Conclusions

Cross-speaker analysis of voice source correlates of focal accentuation and deaccentuation for the six speakers analyzed reveals considerable differences in the individual baseline settings and in the realization of focal accentuation/postfocal deaccentuation. While the single most consistently large differentiation appears to be achieved in terms of the strength of the glottal excitation, EE, greater or lesser shifts in F0 clearly are also key. Differences in the tenseness/laxness setting also appear to be a trend, but not evidenced in every case. The most striking counterexample is for the speaker with largest F0 excursion. More work will be required to assess the interaction of F0 with glottal shape parameters.

The present study supports the Voice Prominence Hypothesis [2], as it pertains to cross-speaker variation. The extent to which different source parameters and F0 contribute to focal salience does appear to differ from speaker to speaker. The speaker most exploiting F0 to realize a contrast in salience relies less on other parameters such as OQ and RD.

5. Acknowledgements

This research is supported by the Dept. of Arts, Heritage, Regional, Rural and Gaeltacht Affairs (*ABAIR* project).

6. References

[1] C. Gobl, "Voice source dynamics in connected speech," *STL-QPSR,* vol. 1, pp. 123-159, 1988.

[2] A. Ní Chasaide, I. Yanushevskaya, J. Kane, and C. Gobl, "The Voice Prominence Hypothesis: the interplay of F0 and voice source features in accentuation," in *Interspeech 2013*, Lyon, France, 2013, pp. 3527-3531.

[3] I. Yanushevskaya, C. Gobl, J. Kane, and A. Ní Chasaide, "An exploration of voice source correlates of focus," in *Interspeech 2010*, Makuhari, Japan, 2010, pp. 462-465.

[4] A. Ní Chasaide, I. Yanushevskaya, and C. Gobl, "Voice source dynamics in intonation," in *XVIIth International Congress of Phonetic Sciences*, Hong Kong, China, 2011, pp. 1470-1473.

[5] I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "The interaction of long-term voice quality with the realisation of focus," in *Speech Prosody 2016*, Boston, MA, 2016, pp. 1-5.

[6] M. Heldner, "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," *Journal of Phonetics,* vol. 31, pp. 39-62, 2003.

[7] J. Koreman, "The effects of stress and f0 on the voice source," in *PHONUS 1*, Saarbrücken: Institute of Phonetics, University of Saarland, 1995, pp. 105-120.

[8] A. M. C. Sluijter and V. J. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," *Journal of the Acoustical Society of America,* vol. 100, pp. 2471-2485, 1996.

[9] M. Vainio, M. Airas, J. Järvikivi, and P. Alku, "Laryngeal voice quality in the expression of focus," in *Interspeech 2010*, Chiba, Japan, 2010, pp. 921-924.

[10] C. Gobl and A. Ní Chasaide, "Amplitude-based source parameters for measuring voice quality," in *VOQUAL'03*, Geneva, Switzerland, 2003, pp. 151-156.

[11] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR,* vol. 4, pp. 1-13, 1985.

[12] J. C. Wells, *Accents of English*. Cambridge: Cambridge University Press, 1982.

[13] C. Gobl and A. Ní Chasaide, "Techniques for analysing the voice source," in *Coarticulation: Theory, Data and Techniques*, W. J. Hardcastle and N. Hewlett, Eds., Cambridge: Cambridge University Press, 1999, pp. 300-321.

[14] G. Fant, "The LF-model revisited: transformations and frequency domain analysis," *STL-QPSR,* vol. 2-3, pp. 119-156, 1995.

[15] G. Fant, "The voice source in connected speech," *Speech Communication,* vol. 22, pp. 125-139, 1997.

[16] C. Gobl and A. Ní Chasaide, "Voice source variation and its communicative functions," in *The Handbook of Phonetic Sciences*, W. J. Hardcastle, J. Laver, and F. E. Gibbon, Eds., 2 ed Oxford: Blackwell Publishing Ltd, 2010, pp. 378-423.

[17] A. Field, *Discovering statistics using SPSS*, 3 ed. London: Sage, 2009.