INTERSPEECH 2013

The Voice Prominence Hypothesis: the Interplay of F0 and Voice Source Features in Accentuation

Ailbhe Ní Chasaide, Irena Yanushevskaya, John Kane, Christer Gobl

Phonetics and Speech Laboratory, Centre for Language and Communication Studies
Trinity College Dublin, Ireland

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

Abstract

This paper explores the interplay of source correlates of accentuation, examining a hypothesis (the Voice Prominence Hypothesis) that different source parameters are involved and may serve as equivalent. It predicts that where accentuation is not marked by pitch salience there will be more extensive changes in other source parameters. This follows our assumption that prosodic entities such as accentuation, focus, declination, etc. involve adjustments to the entire voice source and not simply to F0. Twelve 3-accent sentences of Connemara Irish (declaratives, WH questions and Yes/No questions) were analysed. These are typically produced and transcribed as H* H* H*L. Of particular interest were the second accents: although they are heard as accented, there are no particular pitch excursions that would account for their salience. Inverse filtering and subsequent source parameterisation was carried out to yield measures for a range of source parameters. Results support the voice prominence hypothesis: as predicted, the most striking source adjustments were found in the second accent. Even where there is substantial pitch movement (final accent), parameters other than F0 appear to be contributing to the salience of the accented syllable. The precise source changes associated with accentuation varied across sentence types and within the prosodic phrase.

Index Terms: voice source, accentuation, prominence.

10.21437/Interspeech.2013-759

1. Introduction

This paper explores voice source correlates of prosodic accentuation, as part of a wider study of the prosody of the voice. Our underlying contention is that prosody concerns the temporal, dynamic patterning of the voice, and not just of F0. It follows that we are claiming that prosodic entities (such as accentuation, focus, declination, etc.), typically thought of and defined in terms of F0 correlates, are more fundamentally properties of the entire voice source [1, 2].

In this paper we examine the Voice Prominence Hypothesis, suggested by our earlier work, e.g., [3, 4] that the prominence attributed to accented syllables may depend to a greater or lesser degree on 'excursions' in a number of source parameters, of which F0 is only one. It suggests a degree of perceptual equivalence among source parameters that can signal salience. Thus the speaker might exhibit variability in the use of parameters across utterances or contexts. Furthermore, speakers might vary in terms of how individual parameters would be exploited. So for example, one speaker might have a preference to achieve prominence by using large F0 excursions, with relatively modest change in the phonatory settings, while another might use relatively limited F0 excursions, and rely more heavily on changes to other source parameters.

Of particular interest in terms of our hypothesis are accented syllables where F0 does not appear to be contributing much. This is often found to be the case in sentences with three accents, with a flat hat pattern, which have typically been transcribed as H* H* H*+L, and where the middle accent shows no particular pitch perturbations that would render it prominent. According to our hypothesis, the middle accent should achieve salience through excursions of voice source parameters other than F0. For this study we have selected utterances which conform to this general 3-accent pattern, where the middle accent is clearly heard by the four trained phoneticians as accented, but where close inspection shows that prominence is unlikely to be conferred by pitch.

It should be pointed out that the contribution of the voice source to speech prosody has gone largely undescribed. One reason for this is that it is hard to quantify: the lack of robust and accurate tools for measuring voice source characteristics has been a major roadblock. Parallel research of our group is therefore focussing on the development of more robust analytic tools and algorithms [5, 6] and an initial version of our in-house analysis tool GlóRí is exploited here.

In the intonation literature, F0 cues have traditionally been regarded as the primary marker of accentuation, with duration and intensity also playing a role. Note that supralaryngeal articulation is also implicated [7]. There is not an extensive literature on voice source correlates of accentuation, and particularly few studies where direct measurements were made of voice source parameters. However, proxy measures of the voice source have been used, i.e. spectral measurements of the speech waveform which could reflect source behaviour. For example, in [8-11] on the basis of such measurements, it is suggested that a less steep spectral slope (boosting of the higher frequency regions) and consequently tenser voice is associated with focal stress and prominence in Dutch and American English. Similar findings are reported in [12].

Direct measurements of source parameters are presented in [13-18]. The results of these studies are not fully conclusive. Whereas [13-15] suggest a tenser mode of phonation in accented syllables, results in [17, 18] suggest the opposite, i.e., a laxer phonation mode. Some earlier explorations by the present authors have illustrated source changes on focally accented syllables that are consistent with a tenser mode of phonation [3, 4]. However, these studies did not reveal a boosting of the higher frequencies of the source spectrum, such as suggested in some of the studies referred to above [8-11].

2. Materials and method

2.1. Selected utterances

The utterances analysed were elicited from a female speaker of Connemara Irish. The sentences were of three sentence modes, declaratives (DEC), WH questions (WH-Q) and Yes/No questions (YN-Q), four utterance of each sentence mode, 12 sentences in total. They were elicited within scripted mini-

25 – 29 August 2013, Lyon, France

dialogues. The sentences and their word-for-word translation are shown below, with the accented syllables in caps.

DEC Bhí **CIAN** ag **CAINT** leo sa **MA**rgadh.
[vʲi ciənˠ ə kaɲt̠ʲ lʲɔ sə marˠəgə]
'was Cian at speaking with them in the market'

WH-Q Is **CÉ** bheas á **MBA**iliú ón **MA**rgadh?
[ɪs ce vʲes ɑː maʎu ɔn marˠəgə]
'and who will be at their collecting from the market'

YN-Q Raibh **CIAN** ag **CAINT** léi sa **MA**rgadh.
[ɾˠo ciən ə kaɲt̠ʲ lʲɔ sə marˠəgə]
'was Cian at speaking with her in the market'

All these utterances were realised with three accents, and would in virtually all cases have been transcribed as H* H* H*+L, often referred to as a flat hat pitch pattern.

2.2. Inverse filtering and parameterisation

The voice source data were obtained using the analysis facilities available in the GlóRí system. The procedure can be described as follows. Initially, glottal closure instants (GCIs) are detected using the SE-VQ algorithm [6]. Inverse filtering is then carried out using the iterative and adaptive inverse filtering method (IAIF) developed by P. Alku [19], analysing frames containing two glottal pulses, centred on the GCIs. Parameterisation of the glottal source signal obtained from the inverse filtering is done by fitting the Liljencrants-Fant (LF) glottal source model [20]. For the model fitting, we use the recently developed DyProg-LF method [5], which has been shown to be more robust than previous methods. The DyProg-LF method employs a dynamic programming algorithm to determine an optimal path of RD values through the whole utterance, where RD is a global shape parameter of the glottal pulse. Finally, an optimisation procedure to refine the model fit is carried out.

A full set of parameters were derived from this procedure including F0, the fundamental frequency; EE, the excitation strength, which is a measure of the amplitude of the main glottal excitation, and which is an indicator of the overall amplitude level of the source spectrum; UP, the peak flow of the glottal pulse; RA, which is a measure of the return phase of the glottal cycle following the main excitation, and which is a major indicator of source spectral tilt affecting the relative levels of the higher frequencies of the source spectrum; RK, glottal skew, is a measure of the degree of symmetry of the glottal pulse; RG, the normalised glottal frequency, is the characteristic frequency of the glottal pulse normalised to F0; OQ, the open quotient, is the proportion of the glottal cycle for which the glottis is open (note that OQ is determined by RK and RG). Together, RG, RK and OQ are indicators of the levels of the low frequencies of the source spectrum, and along with EE and RA, indicate overall vocal intensity. For further discussion of these parameters, see [21].

2.3. Data representation

In each of the syllables of the 12 utterances, a representative data sample was taken approximately in the middle of the vowel, away from the consonantal perturbations. The F0 and voice source values at these sample points were averaged across the utterances per sentence mode. Averaged prosodic contours were then generated for each source parameter and for each sentence and are illustrated in the contour panels of Figure 1. Note that the values for the accented syllables are encircled and designated as A1, A2, A3 to refer to the initial, the middle and the final accented syllables.

Underneath the averaged contour lines for each parameter in Figure 1 we show also bar charts which help us to easily visualise the contribution of a given parameter to the 'salience' of a particular accented syllable. This is intended as a way of illustrating the local scaling of excursions of each source parameter associated with the accented syllables. These were derived from the averaged contours in the following way: the height of any bar was calculated as the difference between the value of the accented syllable and the average of the values of the two adjacent unaccented syllables. Where the parameter value in an accented syllable departs from those of the adjacent unaccented syllables, in a direction that would, in principle, be prominence lending, we can regard it as contributing to salience. We are cognisant of the fact that perceptual testing will be needed to establish to what extent source changes like these contribute to the perceived prominence of the syllable. The data represented in the bar charts have been normalised to the average range of sample values across the twelve inverse filtered utterances.

3. Results and discussion

As expected, the mid-phrase accent A2 did not appear to have pitch salience. Although the pitch contour for the three sentence modes was rather similar, there were nonetheless a number of differences and they will be looked at individually. It should be noted that over all these utterances, RA did not appear to be implicated in the differentiation of accented and unaccented syllables. RA values are not shown in Figure 1.

3.1. Declaratives

In these sentences neither the middle accent A2 nor the first (A1) appear to owe their prominence to F0: the accented syllable does not protrude from the surrounding unaccented syllables, unlike the third accent A3, where F0 drops sharply.

In A2, UP and EE levels are very high relative to the values of the surrounding syllables. This indicates greater airflow as well as a raised level of the source spectrum. As OQ is also somewhat elevated, these parameters would appear to show what has been termed 'flow phonation' [22] characterised by efficient voicing with little increase in laryngeal tension.

In these sentences A1, rather like A2 did not appear to have pitch salience, as the pitch on the initial unstressed syllable is already rather high. Like A2, it exhibits some (though less) EE raising, but the striking rise in UP, attested for A2 is not found here. A further difference is that OQ is lowered and RG raised: together these suggest a relative weakening of the low end of the source spectrum, as would be achieved with an increase in laryngeal tension. We would infer that the higher EE is here achieved through greater laryngeal tension, unlike the case of A2. In the case of A3, there is considerable F0 movement (it is heard as a H*L), and other than some raising of EE, there is little else that would lend extra salience in the source parameters.

3.2. WH Questions

Overall the pitch contours of WH questions are very similar to those of the declaratives. However, there are some differences. While the F0 of A2 does not protrude in any way that would

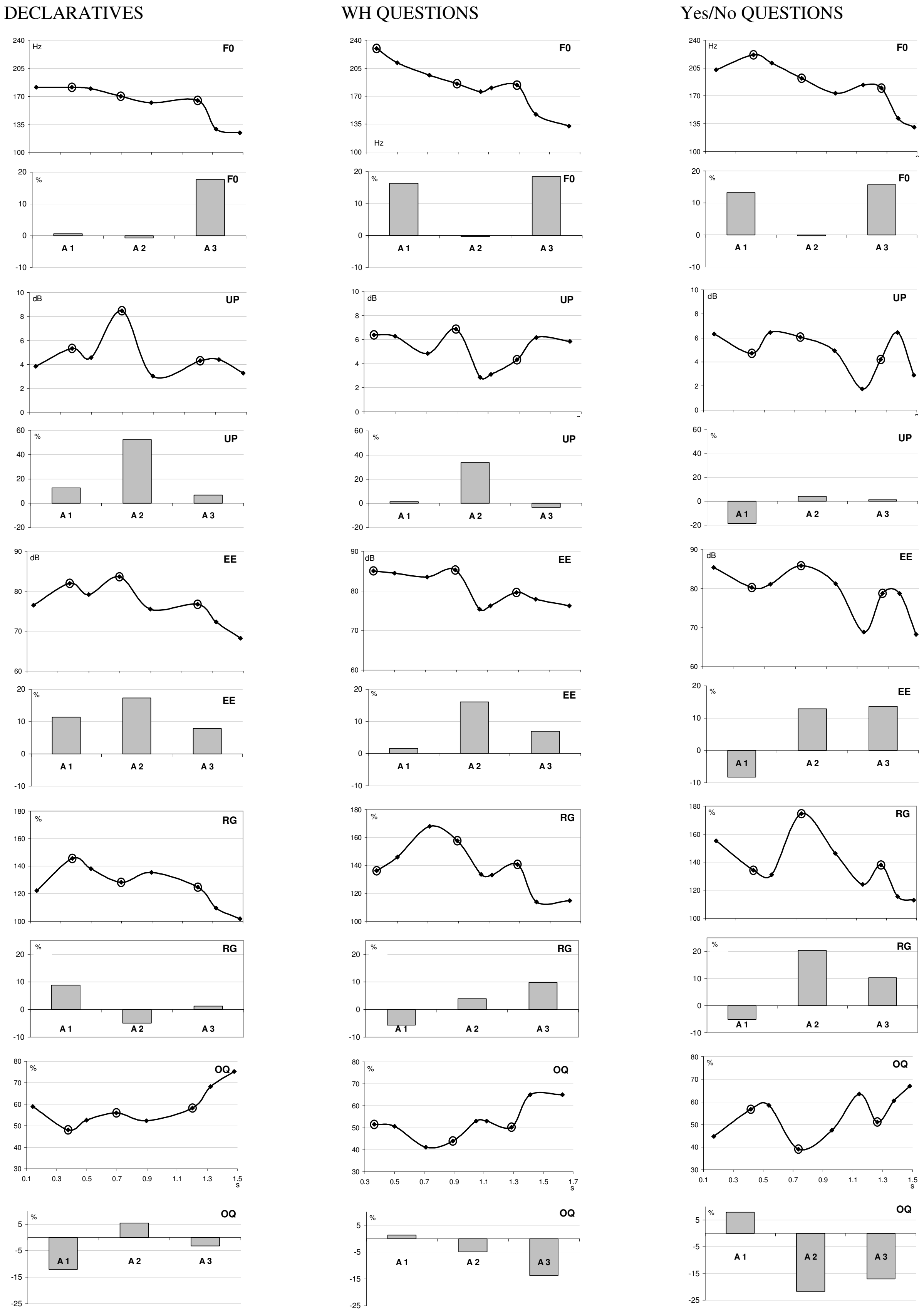

Figure 1. Voice source parameter dynamics in the three sentence modes. Bar charts show local scaling of excursions for each parameter associated with the three accented syllables: values are expressed relative to the overall ranges across the 12 utterances.

lend F0 salience, the initial H* accent, A1, has an elevated F0 level, in comparison with the initial H* of declarative sentences. Such raising of the first H* has been noted before for WH questions in both Irish and Irish English [23, 24]. The F0 salience of A1 (and of A3) shows up clearly in the bar chart in Figure 1.

Although A1 and A3 have pitch salience (see bar charts in Figure 1), we note, particularly for A3 that other source parameters are also implicated. EE is elevated, and the raised RGand lowered OQ point to an increase in laryngeal tension. Note that the raised RG and lowered OQ in A3 are momentarily overriding what is a trend towards an increasingly lax phonation in the utterance. This is essentially the 'source declination', evidenced here in the downdrift of RG and updrift of OQ, and is discussed again briefly below. All in all, the most striking involvement of source parameter salience in this sentence set is found with A2, which does not appear to have pitch salience. And while the accentuation of A3 (which has considerable pitch salience) does show involvement of other source parameters, the extent of their contribution appears to be less. This conforms to our expectations from the voice prominence hypothesis.

3.3. Yes/No Questions

As with the WH-questions, although the basic 'tune' remains the same, both A1 and A3 have considerable pitch salience (see bar charts in Figure 1). A2 does not have pitch salience, and is the main focus of our interest.

In striking contrast to the declarative and WH question sets, UP raising is not found here on A2. There is a degree of EE raising, but it is not as great as the raising found in the other sentence sets. The most striking salience-lending effect found for A2 in these sentences is in the high RG and concomitantly lowered OQ: these imply a tenser mode of phonation.

A3 has considerable pitch salience, but exhibits some similar source characteristics: a raised EE, along with raised RG and lowered OQ (the extent of the latter changes are less extensive than for A2). Curiously, for A1, a combination of a slightly lowered EE and UP, along with a very slight lowering of RG and raising of OQ point to a slightly laxer mode of phonation in this instance.

To sum up for the Yes/No questions: A2 achieves salience, not through raising of EE and UP (as in the other sentence sets) but rather through a more tense phonatory setting. In the case of A3 and A1, which have pitch salience, there are some differences. A3 has source changes rather similar to A2, but to a lesser extent. For A1 no such adjustments were found, but rather a slight trend in the opposite direction.

4. Discussion and Conclusions

These data conform to our basic perspective on prosody, that prosody is expressed through the temporal, dynamic variation of the entire voice source, of which F0 variation is a part. Here, source parameters other than F0 are implicated in accentuation, whether or not F0 salience is found.

These data further lend support to Voice Prominence Hypothesis, which suggests that the prominence of the accented syllables can be achieved through different, salience-lending adjustments of the source. Variability in the source parameters associated with accentuation in different instantiations is thus to be expected, and the hypothesis predicts that where pitch prominence is not used, there is likely to be a greater involvement of other source parameters. In the present data, for the accented syllables which do not have pitch salience (A2, and A1 of the declaratives) we find the most striking changes to other source parameters. Even for those accented syllables that do have clear pitch excursions (A3 of all sentences and A1 of the question sets) parameters other than F0 are clearly contributing, if to a lesser extent. Accentuation does appear to rely on potentially different source features, and there is a tendency to increase one, when another is not present.

It is also striking that the (non F0) source adjustments in accentuation may vary. Location in the prosodic phrase may be one determinant: note for example in the declaratives that the principal prominence-lending source adjustment differs from A1 to A2 to A3. And the source adjustments for a given accent (say A2) are not constant across the different sentence modes looked at here: in the declaratives and WH questions there is a major involvement of UP and EE; in the Yes/No questions this is not found but a considerable adjustments to RG and OQ are found. Even among the more similar declaratives and WH questions there were indicators that there are differences in the laryngeal tension settings.

It was mentioned above that there was evidence in these sentences of source-parameter involvement in declination. This can be seen in the downdrift tendency in parameters such as EE and RG, along with the concomitant updrift in OQ, all of which point to an increasingly lax mode of phonation over the course of the utterance. As with the F0 adjustments, source changes indicative of accentuation are superimposed on a baseline that shifts over the utterance. For some discussion of source declination, see [1, 2].

RA shifts were not found to be correlated with accentuation in any of the data sets examined here, and from this we infer that a boosting of the higher frequencies of the source spectrum is not involved in these cases. This is rather different from what has been suggested in studies such as [8-11] where source levels were inferred from measurements carried out on the speech output, but consistent with our own earlier findings [3, 25].

To sum up, the results of this study support our contention that all source parameters together provide the prosody of speech, and the hypothesis that specific prosodic targets such as accentuation or prominence can be achieved through different strategies so that different source parameters may be involved. To that extent, looking for a single invariant source correlate of prominence would be premature.

These data were intended to provide an initial perspective on how different source parameters can combine in the prosodic expression of accentuation. From here, we hope to proceed with more extensive measurements more speakers and different contexts. A priority in future work will be to examine through perception experiments how the differing source characteristics observed here contribute to the perceived prominence of accented syllables.

5. Acknowledgements

This research is supported by the Science Foundation Ireland Grant 09/IN.1/I2631 (FASTNET) and the Irish Department of Arts, Heritage and the Gaeltacht (ABAIR project). The authors would like to thank Maria O'Reilly for the data and intonation transcription as well as for the discussion of the data.

6. References

[1] A. Ní Chasaide and C. Gobl, "Voice quality and f0 in prosody: towards a holistic account," presented at the Speech Prosody 2004, Nara, Japan, 2004.

[2] A. Ní Chasaide and C. Gobl, "Decomposing linguistic and affective components of phonatory quality," presented at the Interspeech 2004, Jeju Island, Korea, 2004.

[3] A. Ní Chasaide, I. Yanushevskaya, and C. Gobl, "Voice source dynamics in intonation," presented at the XVII International Congress of Phonetic Sciences, Hong Kong, China, 2011.

[4] I. Yanushevskaya, C. Gobl, J. Kane, and A. Ní Chasaide, "An exploration of voice source correlates of focus," presented at the Interspeech 2010, Makuhari, Japan, 2010.

[5] J. Kane and C. Gobl, "Automating manual user strategies for precise voice source analysis," *Speech Communication,* vol. 55, pp. 397-414, 2013.

[6] J. Kane and C. Gobl, "Evaluation of glottal closure instant detection in a range of voice qualities," *Speech Communication,* vol. 55, pp. 295-314, 2013.

[7] T. Cho and P. Keating, "Effects of initial position versus prominence in English," *Journal of Phonetics,* vol. 37, pp. 466-485, 2009.

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

[9] A. M. C. Sluijter, S. Shattuck-Hufnagel, K. N. Stevens, and V. J. Van Heuven, "Supralaryngeal resonance and glottal pulse shape as correlate of stress and accent in English," presented at the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, 1995.

[10] A. M. C. Sluijter and V. J. van Heuven, "Acoustic correlates of linguistic stress and accent in Dutch and American English," presented at the ICSLP 96, Philadelphia, USA, 1996.

[11] A. M. C. Sluijter, V. J. van Heuven, and J. J. Pacilly, "Spectral balance as a cue in the perception of linguistic stress," *The Journal of the Acoustical Society of America,* vol. 101, pp. 503-513, 1997.

[12] 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.

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

[14] M. Epstein, "Voice Quality and Prosody in English. PhD thesis," PhD thesis, 2002.

[15] M. Iseli, Y.-L. Shue, M. A. Epstein, P. Keating, J. Kreiman, and A. Alwan, "Voice source correlates of prosodic features in American English," presented at the Interspeech 2006 - ICSLP, Pittsburgh, Pennsylvania, USA, 2006.

[16] P. Murphy, "Voice source change during fundamental frequency variation," in *Verbal and Non-Verbal Communication Behaviours*. vol. 4775, ed Berlin / Heidelberg: Springer, 2007, pp. 165-173.

[17] M. Vainio, M. Airas, J. Järvikivi, and P. Alku, "Laryngeal voice quality in the expression of focus," presented at the Interspeech 2010, Chiba, Japan, 2010.

[18] M. Airas, P. Alku, and M. Vainio, "Laryngeal voice quality changes in expression of prominence in continuous speech," presented at the 5th International Workshop on Models and Analysis of Vocal Emissions in Biomedical Applications (MAVEBA 2007), Florence, Italy, 2007.

[19] P. Alku, "Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering," *Speech Communication,* vol. 11, pp. 109-118, 1992.

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

[21] 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.

[22] J. Sundberg, *The Science of the Singing Voice*: Northern Illinois University Press, 1987.

[23] A. Dorn, M. O'Reilly, and A. Ní Chasaide, "Prosodic signalling of sentence mode in two varieties of Irish (Gaelic)," presented at the The XVII International Congress of Phonetic Sciences, Hong Kong, China, 2011.

[24] R. Kalaldeh, "Segmental and Intonational Analysis of Drogheda English. PhD thesis," Trinity College Dublin, 2011.

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