Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

The interaction of long-term voice quality with the realisation of focus

Irena Yanushevskaya, Ailbhe Ní Chasaide, Christer Gobl

Phonetics and Speech Laboratory, Trinity College Dublin, Ireland

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

Abstract

Voice quality shifts have been shown to be associated with the realisation of accent, focus and deaccentuation. Mostly, accented and focally accented syllables are reported to exhibit a tenser mode of phonation than the unaccented, but accentuation with a laxer/breathier quality is also reported. Possibly, the long-term voice quality of the speaker (or of the utterance) influences the voice quality of accented/unaccented syllables. This paper examines the hypothesis that speakers who have a baseline tense phonatory setting, or who are using a tenser mode of phonation in a particular utterance, will signal accentuation more through breathy phonation than through increased tenseness. To test this hypothesis we analysed utterances produced by a single informant with three voice qualities; modal, breathy and tense. These utterances were produced with variations in the location of a focal accent. The utterances were manually inverse filtered and the voice source parameters F0, EE, UP, RG, CQ and RD were obtained by fitting the LF (Liljencrants-Fant) model to the glottal flow signal. The change in parameter settings in focally accented syllables was examined relative to the adjacent unaccented syllables. The hypothesis was not supported. Overall, the focally accented syllables were associated with tenser phonation, except in the utterance-initial position where they were associated with breathier voice irrespective of the sentence's intended phonation type. In these data, the voice source modulations associated with focal accentuation appear to depend on the position of the focal syllable in the phrase.

Index Terms: voice source, voice prosody, focus, accentuation, prominence, voice quality, phonation types, phrase-initial.

1. Introduction

This paper looks at how the baseline voice quality of an utterance affects the realisation of focal accentuation. This is part of a broader investigation of the 'prosody of the voice', i.e. the role of voice quality in shaping both linguistic and paralinguistic/expressive prosody [1], [2], [3]. Prior work has looked at voice source dynamics in signalling accentuation [4] [5], focus [6], and declination [7], and has explored how F0 and voice quality combine in the expression of prosodic 'events'.

There is increasing evidence that accentuation, and particularly focal accentuation is associated with shifts in phonatory quality. Our earlier studies [4], [6], [8], [5] suggest that a tenser mode of phonation is associated with both normally and focally accented syllables (the data included Irish-English, Irish and Swedish speech samples). This is also suggested by a number of other studies. The indirect source measures in the analyses in [9], [10], [11] have suggested that flatter spectral tilt (boosting of the higher frequency regions) and therefore, one assumes, tenser voice quality is associated with focal stress and prominence in Dutch and American English. There are similar findings in [12] which point to the overall intensity and spectral emphasis as reliable acoustic cues of focal accents in Swedish (spectral emphasis is a measure of a relative contribution of high frequency components to overall intensity). A study of the voice source correlates of stress using direct measures of source parameters [13] also points towards a tenser mode of phonation in stressed syllables: they are characterised by smaller OQ, lower RK (more skewed pulse), smaller DC offset (less glottal flow leakage), with higher EE and AC flow due to the more abrupt closure of the vocal folds. In a similar vein, [14], [15], [16] report voice source measures that also suggest a tenser phonation in accented syllables. All in all, the most frequent finding appears to be that stressed syllables involve changes in the glottal pulse shape that suggest greater tension in the mode of phonation.

However, there is some counter evidence to the generality of such findings. Although, broadly speaking, most studies would appear to point towards tenser voice quality being associated with accentuation, studies on Finnish [17], [18] found the opposite: they report higher NAQ values in focally accented syllables, as in stressed syllables. The NAQ measure [19] has been proposed as a global parameter, which correlates with the tense/lax dimension of vocal quality: a high NAQ value is indicative of lax or breathy voice, and a low NAQ value indicative of pressed or tense voice.

There are different kinds of explanations one could proffer for the seemingly contradictory findings. Languages could 'do accentuation' in different ways. Could it be that Finnish is just different from the other languages? Or might the apparent differences be an artefact of the differing methodologies that have been employed in different studies? One possible explanation is that the realisation of accentuation (and focal accentuation) may depend on the baseline voice quality that the speaker is using. The voice source is modulated to signal prominence, but what parameters are varied and to what extent could depend on what is available to the individual speaker – the speaker's baseline, neutral voice quality. If a speaker has (or temporarily uses) a tense voice quality, it might be difficult to effect a contrastively even tenser quality, and so, a breathier quality might be used to signal focal accentuation. Note that this would indirectly imply that Finnish speakers (or at least the speakers used in the particular studies mentioned above) use a long-term relatively tense phonatory setting.

 doi: 10.21437/SpeechProsody.2016-191

In this paper we test this hypothesis by analysing the realisation of focally accented syllables in utterances produced by the same speaker, with three voice quality settings: breathy, modal and tense voice. The sentence produced with the three different voice qualities was "WE were aWAY a YEAR ago" and it was produced over different repetitions so that the focal accent fell on one of the three syllables shown in capital letters. The prediction was that in the repetitions produced with tense voice, the focally accented syllables would exhibit a less tense (laxer, breathier) voice quality, while in the modal and breathy voiced repetitions, these same syllables would show rather a trend towards tenser phonation.

2. Material and method

2.1. Speech data

As mentioned, the data consisted of the all-voiced utterance 'We were away a year ago' produced by a male speaker of Irish-English with three phonation types, modal, breathy and tense voice. The speaker was producing these voice qualities in accordance with the Laver system of voice quality classification [20]. The utterance was repeated (as declaratives, with a falling F0) in each phonation type in such a way as to represent focal accentuation on all potentially accentable syllables, as shown below:

WE were away a year ago.

We WERE away a year ago*. (Not used here.)

We were aWAY a year ago.

We were away a YEAR ago.

The recording was made in a semi-anechoic room, using a Brüel & Kjær microphone and amplifier (B&K 4191 and B&K Nexus 2690) and a Roland A/D converter (Edirol UA-1000). This recording setup ensures a linear phase response as well as negligible amplitude distortion and noise. The microphone was held at 30 cm from the speaker. The utterances were recorded using a sampling frequency of 44.1 kHz. Prior to the inverse filtering the utterances were high-pass filtered (with a cut-off frequency of 40 Hz) and downsampled to 10 kHz.

Auditory analysis of the recorded voice qualities conducted by the authors confirmed that the utterances were consistently produced with breathy, modal or tense voice. It is worth mentioning that the speaker was an experienced speech researcher and an expert in voice source analysis.

2.2. Inverse filtering and source parameterisation

The 12 utterances were analysed using the software system described in [21]. 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, in order to get the best possible source approximation. Voice source parameterisation involved fitting the LF (Liljencrants-Fant) model of differentiated glottal flow [22] to the source signal derived from the inverse filtering. The voice source measures obtained and used in this study included F0, EE, UP, RG, OQ and RD. A brief description of these parameters is given below, see also [23] [24] [25]. These parameters were used in our previous studies and allow for comparability of results.

F0 (fundamental frequency), is defined as 1/T0 where T0 is the duration of one glottal cycle.

EE (excitation strength), is defined as the negative amplitude of the differentiated glottal flow at the time point of maximum waveform discontinuity.

UP (peak glottal flow), which is a measure of the maximum glottal airflow rate of the glottal pulse.

RG (normalised glottal frequency), which is a measure of the characteristic frequency of the glottal pulse (FG), normalised to F0. RG mainly affects the relative amplitudes of the first two harmonics of the source spectrum.

OQ (open quotient), is the duration of the open phase (excluding the return phase) of the glottal pulse normalised to the glottal period. Open quotient tends to be negatively correlated with RG. Changes in OQ affect mainly the amplitudes of the lower components of the source spectrum. CQ, the closed quotient, is derived from OQ as 1 – OQ, and it is this measure that is used in presenting results below.

RD (a global waveshape parameter) is derived from F0, EE and UP as follows: $(1/0.11)\times(F0\cdot UP/EE)$, where UP/EE is equivalent to the glottal pulse declination time during the closing phase of the glottal cycle. The scale factor (0.11^{-1}) makes the numerical value of RD equal to the declination time in milliseconds when f_0 is 110 Hz.

The accuracy of the analysis was tested by resynthesising the utterances (using the in house GlóRí voice analysis system [26]) and by conducting auditory comparison of the original and resynthesised utterances.

In the discussion below we show the results for WE, WAY and YEAR focally accented syllables. The calculations are based on parameter values smoothed by a moving average filter spanning three pulses.

The research question is: how do the voice source adjustments in signalling focal accentuation vary with the 'baseline' phonation type? What strategies are employed by the same speaker when he signals focus in three different types of phonation?

2.3. Visual representation of the data

As a measure of the phonatory shift associated with the focally accented syllable, we use the measure of 'parameter protrusion'. The contribution of source parameters to the salience of the focally accented syllable is shown here as its local 'protrusion' relative to the adjacent unaccented syllables, following [5]. A representative data sample was taken at vowel midpoint in each syllable, and the 'parameter protrusion' was measured as the difference between the parameter value in the focally accented syllable and the average of the parameter values in the adjacent unaccented syllables. This is a way to represent a local scaling of the excursion of each source parameter associated with the focally accented syllable. These values were then expressed as a percentage of the parameter range of each utterance, and these are shown as bars in Figure 1.

3. Results

Figure 1 gives the summary of 'parameter protrusions' associated with the focally accented syllables in the three focally accented syllables (WE, WAY, YEAR) in breathy (left panel, in blue), modal (mid panel, in black) and tense (right panel, in red) phonation types. Within each panel the leftmost bars are for the syllable WE, the mid bars are for the syllable WAY and the rightmost bars are for the syllable YEAR. Where protrusion bars for F0, EE, RG and CQ are positive, and where RD is negative, we infer a shift towards a tenser voice quality. The converse would suggest a laxer or breathier voice quality [27], [28].

3.1. Focal accentuation in tense phonation

Our hypothesis predicts that in a sentence produced with tense phonation, focal accentuation would be achieved by using a breathier voice quality. Looking at the results for our tense repetitions (rightmost panel with red bars in Figure 1), it is clear that the realisation of the focal accent is not identical across the three focus-carrying syllables. In the phrase-initial focused syllable, WE, there are clear indicators of a more breathy voice: the very low CQ, lowered RG, high UP and somewhat raised RD. When the focally accented syllable is non-initial, the measures point rather to a slightly tenser mode of phonation (drop in RD, rise in EE and slight rise in CQ) particularly for the syllable YEAR.

3.2. Focal accentuation in modal phonation

According to our hypothesis, focal accentuation would be realised with shifts towards a more tense voice in an utterance produced with modal phonation. In the present data for the modal utterances (mid panel with black bars) we see again that the realisation is not uniform across the phrase position. For the focally accented syllable YEAR, the raised CQ and RG and slightly lowered RD values point to a slightly more tense phonatory setting. This is also what one would infer from the lowered RD and slightly raised CQ in the focally accented syllable WAY. But, in the phrase initial WE, the focal accentuation seems rather to be marked by a more lax voice, suggested by the strikingly lowered RG and slightly lowered CQ values.

3.3. Focal accentuation in breathy phonation

Our hypothesis predicts that in a sentence produced with breathy phonation, focal accentuation would be achieved by using a tenser phonation. Looking at the results for the breathy repetitions (leftmost panel with blue bars in Figure 1), we find again that the phrase-initial case is different from the non-phrase-initial cases. In the latter there is a clearly tenser mode of phonation in the focally accented syllables (WAY and YEAR): this is shown by the substantially raised CQ and lowered RD values and by the raised RG. The values for the phrase-initial focally accented WE are different: the large positive protrusion of RD suggests again a more lax phonatory mode.

Although a tenser phonatory mode is not uniformly found for the focally accented syllable in the breathy voiced utterances, it is nonetheless striking that where it does occur (syllables WAY and YEAR) the degree of the shift is much greater than what is observed for these same syllables in the modal and tense utterances.

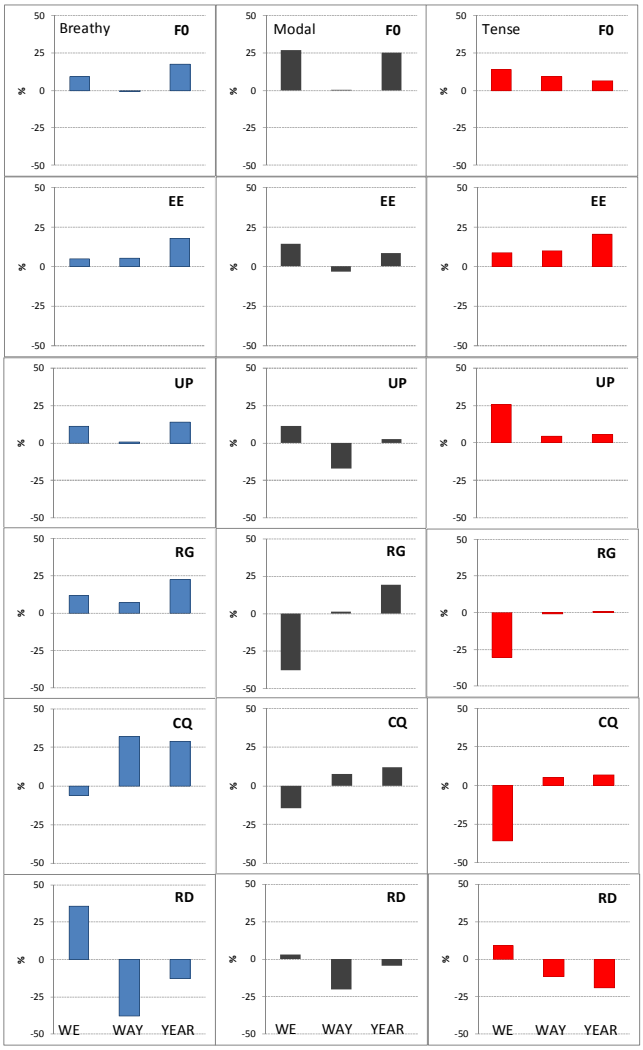

Figure 1. *Parameter levels associated with the focally accented syllables (relative to the adjacent unaccented syllables, expressed as % relative to the parameter ranges of each utterance).*

4. Discussion

The initial hypothesis that focal accentuation would be signalled by a more lax phonation in an utterance produced with a baseline tense voice is not supported by these results. With the exception of the phrase-initial case, most focally accented syllables tend to show a tenser phonatory mode.

In the tense and modal utterances, although the focally accented WAY and YEAR do show a tenser phonation mode, the size of the excursions is relatively small, and less than what was observed in different recordings analysed for this same speaker in experiments where the baseline voice quality was not controlled.

The present recordings do of course constitute a rather 'unnatural' task (for the non-phonetician at least): speakers do not consciously control their voice quality in this way in the course of typical discourse. We would speculate that the effort to maintain tense or modal voice might explain why the phonatory shifts towards tenser voice are here so slight in

these utterances. The fact that the breathy voiced renderings yield greater excursions towards tense voice in the syllables WAY and YEAR may simply reflect the fact that this overall lax voice setting allows the speaker more latitude to vary the vocal tension settings.

Although our initial hypothesis is not supported here, it might nonetheless be worth exploring it further in a somewhat modified version: that the speaker's intrinsic (long-term) voice quality will affect the realisation of focal accentuation, so that speakers with an intrinsically tense voice may choose breathiness as a marker of focal accentuation, while speakers with an intrinsically breathy voice will mark accentuation by using a tenser mode of phonation. To explore this further we propose to use speakers with different intrinsic voice qualities, but matched as far as possible in other respects. This would remove the potentially confounding factor of this study, where the need to 'control' the voice quality of the utterance as a whole may cut across the speaker's scope for modulating the voice for prosodic expression.

The broader question of how voice source modulation associated with prosody interacts with the speaker's long-term voice quality is an area that we are currently exploring. Clearly, the prosody related modulations of the voice, whether they express linguistic prosody (e.g., as here, focal accentuation, but also overall phrasing, boundary marking, etc.) or paralinguistic, affective prosody (e.g., boredom, anger, etc.) occur against the backdrop of the speaker's intrinsic, long-term voice quality. To date, little is known about how these dimensions of voice interact.

The most striking finding to emerge in this study is one that was not expected: the finding that the realisation of the phrase-initial focal accent differed from the others and was more lax, regardless of the speaker's intended overall voice quality. As this study involves a small sampling of data from a single speaker, one would not wish to make any claims that these results can be generalised. Nonetheless, they are striking and provide pointers for future work.

While the differences in the phrase-initial and non-phrase-initial realisations were unexpected, it is worth noting that this is not the first time that positional variation has been observed in the realisation of accentuation and focus. Other studies such as [29], [12], [30] have also found that prominence-lending voice source adjustments associated with focus appear to depend on the position of the focal syllable in the prosodic phrase.

We do not at this point have a likely explanation as to why the phrase-initial accent should behave differently from the non-phrase-initial case, and this is something we hope to investigate further. The fact that the accented syllable WE is absolutely phrase-initial and lacks an anacrusis may be relevant here. In further investigations we plan to examine whether varying the size of the anacrusis might have an effect.

The role of duration should also be included in a future study: the salience of an element mediated not only by its acoustic characteristics, but also by its temporal features.

Finally, although this study looked uniquely at the local effect of focal accentuation (the realisation of the focally accented syllable relative to the adjacent unaccented syllables), future studies will also address the global level of the utterance. As our earlier investigations have shown, an utterance-wide vocal highlighting/lowlighting effect contrasts the focally accented element with the post-focal (and sometimes the pre-focal) elements. This broader perspective will be important when trying to understand how the speaker's intrinsic voice quality interacts with the prosody of the utterance.

5. Conclusions

This paper tested the hypothesis that when using a baseline tense voice, speakers might signal focal accentuation by a shift towards breathiness, as opposed to the shift towards tenseness, which we and others have generally found. The hypothesis was not supported here. Given that directly controlling the voice quality of an utterance, as was done here, is inherently a rather untypical speech task, it is suggested that the hypothesis might still be worth exploring with different speakers, whose long-term voice quality setting is inherently tense or lax.

The most striking and unexpected finding to emerge in this study was the relatively lax phonation which characterised the focal accent when phrase-initial. This was quite unlike the non-phrase-initial cases, where a tenser phonation was observed. Although this study concerns a tiny speech sample from a single speaker, this is something that we feel warrants further exploration. In future studies we would hope to extend the size of the speech sample, and also consider whether the anacrusis has an effect. Other aspects of focal accent realisation will also be looked at, particularly the global adjustments in phonation mode that pertain to the utterance as a whole. Some of our past studies [4], [5], [6], have indicated that the postfocal (and sometimes prefocal) portions of the utterance can vary greatly, in a way that contributes to the salience of the focally accented element. The role of durational differences also need to be considered.

6. Acknowledgements

This research is supported by ABAIR (An Roinn Ealaíon, Oidhreachta agus Gaeltachta) and CabairE (An Chomhairle um Oideachas Gaeltachta & Gaelscolaíochta, COGG).

7. References

[1] 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.

[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] I. Yanushevskaya, C. Gobl, and A. Ní Chasaide, "Voice quality and f0 cues for affect expression: implications for synthesis," presented at the Interspeech 2005 - Eurospeech, Lisbon, Portugal, 2005.

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

[5] A. Ní Chasaide, I. Yanushevskaya, J. Kane, and C. Gobl, "The Voice Prominence Hypothesis: the interplay of F0

and voice source features in accentuation," presented at the Interspeech 2013, Lyon, France, 2013.
[6] 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.
[7] A. Ní Chasaide, I. Yanushevskaya, and C. Gobl, "Prosody of voice: declination, sentence mode and interaction with prominence," presented at the XVIIIth International Congress of Phonetic Sciences, Glasgow, UK, 2015.
[8] A. Ní Chasaide, I. Yanushevskaya, and C. Gobl, "Voice source dynamics in intonation," presented at the XVIIth International Congress of Phonetic Sciences, Hong Kong, China, 2011.
[9] A. M. C. Sluijter, V. J. van Heuven, and J. J. Pacilly, "Spectral balance as a cue in the perception of linguistic stress," *Journal of the Acoustical Society of America,* vol. 101, pp. 503-513, 1997.
[10] 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.
[11] 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.
[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] C. Mooshammer, "Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German," *Journal of the Acoustical Society of America,* vol. 127, pp. 1047-1058, 2010.
[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, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parameterization of the glottal flow," *Journal of the Acoustical Society of America,* vol. 112, pp. 701-710, 2002.
[20] J. Laver, *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press, 1980.
[21] C. Gobl and A. Ní Chasaide, "Techniques for investigating laryngeal articulation (Section B: Techniques for analysing the voice source)," in *Coarticulation: Theory, Data and Techniques*, W. J. Hardcastle and N. Hewlett, Eds., ed Cambridge: Cambridge University Press, 1999, pp. 300-321.
[22] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR,* vol. 4, pp. 1-13, 1985.
[23] 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.
[24] G. Fant, "The voice source in connected speech," *Speech Communication,* vol. 22, pp. 125-139, 1997.
[25] G. Fant, "The LF-model revisited: transformations and frequency domain analysis," *STL-QPSR,* vol. 2-3, pp. 119-156, 1995.
[26] J. Dalton, J. Kane, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "GlóRí - the glottal research instrument," presented at the Speech Prosody 2014, Dublin, Ireland, 2014.
[27] C. Gobl, "A preliminary study of acoustic voice quality correlates," *STL-QPSR,* vol. 30, pp. 9-22, 1989.
[28] C. Gobl and A. Ní Chasaide, "Acoustic characteristics of voice quality," *Speech Communication,* vol. 11, pp. 481-490, 1992.
[29] I. Yanushevskaya, C. Gobl, and A. Ní Chasaide, "Voice quality in affect cueing: does loudness matter?," *Frontiers in Psychology,* vol. 4:335, pp. 1-14, 2013.
[30] G. Fant, A. Kruckenberg, J. Liljencrants, and S. Hertegård, "Acoustic-phonetic studies of prominence in Swedish," *HTM-QPSR,* vol. 41, pp. 1-52, 2000.