An Exploration of Voice Source Correlates of Focus

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

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

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

Abstract

This pilot study explores how the voice source parameters vary in focally accented syllables. It examines the dynamics of the voice source parameters in an all-voiced short declarative utterance in which the focus placement was varied. The voice source parameters F0, EE, UP, OQ, RG, RA, RK and RD were obtained through inverse filtering and subsequent parameterisation using the LF-model. The results suggest that the focally accented syllables are marked not only by increased F0 but also by boosted EE, RG and UP. The non-focal realisations show reduced values for the above parameters along with a tendency towards higher OQ values, suggesting a more lax mode of phonation.

Index Terms: voice quality, dynamics, pitch, voice source analysis, focus

1. Introduction

Although tone of voice and the voice source signal are known to be central in speech communication, rather little is known about them. Furthermore, current linguistic analysis of prosody focuses almost exclusively on intonation (variation in fundamental frequency, F0) and timing. Ní Chasaide and Gobl [1, 2] have suggested a holistic approach to prosodic analysis, arguing that (i) prosody is fundamentally the temporal patterning of the entire source, not just of F0, (ii) the dynamic patterning of the source is as important to the linguistic functions of prosody as it is to the paralinguistic function of expressing emotion and attitude, and (iii) the source dynamics involved in the paralinguistic may best be treated as perturbations to the linguistically relevant patterning of the source.

A preliminary study is presented here, which looks at how the voice source parameters change with linguistically relevant shifts in focus. Focus in spoken communication is generally described as a means to emphasise a particular piece of information relative to the information already shared by the conversation participants. Focus can be achieved by grammatical or prosodic means: in the case of prosodic focus it is widely assumed to be primarily signalled by F0 variation (by assigning pitch accents to the syllables that are lexically stressed or by extending the range of F0), e.g., Xu [3, 4] and references therein. For example, in English, a narrow focus is generally characterised by a high falling (H*L) nuclear pitch accent.

The studies of voice source variation related to focus are rather scarce. Gobl [5] reported changes in the strength of the glottal excitation (EE), where the dynamic range was increased in focal context, enhancing the vowel-consonant distinction, with EE being higher for the vowel and weaker for the surrounding voiced consonants. Murphy [6] measured the open quotient (OQ) extracted from the first derivative of the electroglottographic signal (EGG) in a short sentence 'The bat sat here' with focus on the words 'bat' or 'sat', but reported no consistent differences for the vowels in focal and non-focal syllables. Iseli et al. [7] investigated how a number of voice source parameters, e.g., F0, EE, glottal pulse symmetry (RK), and the amplitude difference of the first two harmonics of the source spectrum (H1*-H2*), vary depending on pitch accents, stress, and sentence type. They reported increased F0 and EE and decreased values of RK and H1*-H2* in high pitch accented syllables. Other studies exploring the relationship between F0 and other source parameters, although not related to focus, include [8-12].

2. Materials and methods

This study focuses on detailed source analysis of the short all-voiced utterance 'We were away a year ago' in which the focus placement was varied. The analysis involved inverse filtering to derive an estimate of the voice source signal and source model matching to derive measures for a range of source parameters.

The utterances were produced as declaratives by a male speaker of Irish English. In various renditions of the utterance, the focus was placed in turn in each of the capitalised syllables 'WE WERE aWAY a YEAR ago' with a falling F0 in the nuclear accents. The utterances for the voice source analysis were selected out of several recorded repetitions. According to the auditory analysis, the utterances selected all had one nuclear pitch accent, which was realised as H*L (with the exception of YEAR, where WAY was stressed but not accented), while all the post-nuclear material was deaccented.

The recording was conducted in a semi-anechoic chamber using a Brüel & Kjær microphone and amplifier (B&K 4191 and B&K Nexus 2690) and a Roland A/D converter (Edirol UA-1000). The microphone was held at 30 cm from the speaker and utterances were recorded using a sampling frequency of 44.1 kHz. This recording setup ensures a linear phase response as well as negligible amplitude distortion and noise. Prior to the inverse filtering the utterances were high-pass filtered and downsampled to 10 kHz. The high-pass filter used was a high order FIR filter (linear phase) with a cut-off frequency of 50 Hz.

The four utterances were analysed 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, in order to attain the best possible source approximation. Voice source parameterisation involved fitting the LF (Liljencrants-Fant) model of differentiated glottal flow [14] 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 and RD. A brief description of these parameters is given in Table 1. For further details, see [1, 12, 13].

The extracted parameters were used for copy synthesis of the utterances. In the resynthesis, source and filter parameter values were smoothed by a moving average spanning five pulses. The auditory quality of the resynthesis closely mimicked the original sound, suggesting a reasonably accurate representation of the source and the filter.

For the visual presentation of results and to allow cross-utterance comparison, the utterances were segmented into syllables and the smoothed parameter trajectories were plotted over time. The parameter trajectories were further stylised by taking the extreme parameter values within each syllable (the 'outlier' values at the utterance boundaries were excluded). The time axis was normalised by averaging the duration of each syllable across the four utterances. These stylised parameter trajectories are plotted in Fig. 1.

Table 1: *Source parameters selected for the analysis.*

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.
UP	The peak amplitude of the glottal flow pulse.
RK	A measure which captures the glottal pulse skew: the smaller the RK value, the more asymmetrical the glottal pulse.
RG	The glottal frequency, FG, normalised to F0, where FG is the characteristic frequency of the glottal pulse during the open phase of the glottal cycle.
OQ	The duration of the glottal open phase in relation to T0, i.e. the duration of the whole glottal cycle (OQ can be derived from RG and RK).
RA	RA is TA normalised to the fundamental period, i.e. TA/T0. TA is a measure of the effective duration of the return phase of the glottal pulse (after the main excitation, prior to full or maximum glottal closure) and relates to how instantaneous the glottal closure is. The duration of the return phase is a major determinant of the slope of the source spectrum.
RD	A global waveshape parameter derived from F0, EE and UP as follows: (1/0.11) · (F0·UP/EE).

3. Results and discussion

Fig. 1 illustrates source parameter values taken for the most part at points at the centre of the syllable. At the top of the figure is shown F0. Here, as expected, the focally accented syllable exhibited considerably higher F0 values compared to non-focal realisations. In the case of YEAR, note that the F0 peak is found not on the accented syllable as such, but rather on the following unaccented syllable. This peak delay has been frequently observed in other studies for English, e.g., [15]. The focal accent is always heard here as a high falling H*L tone. The F0 boosting is very much more dramatic in the case of the accented vowels of aWAY and YEAR.

Looking at the EE values, it is also striking that the focally accented syllable in every case shows the highest excitation value. This holds for within any individual repetition, but also for when the specific syllable with focal accent is compared to the realisations of the same syllable without focal accent. As with F0, there is a general tendency towards declination in EE, which tends to only set in after the focal accent. When the focal accent occurs early in the utterance, the downdrift sets in from the beginning of the utterance, and the point at which the downdrift begins gets later when the focal accent occurs later. In some general sense, EE is therefore positively correlated with F0. Note however that this correlation is not necessarily found throughout. Compare for example the F0 and EE traces in the accented syllable YEAR and the following unaccented schwa.

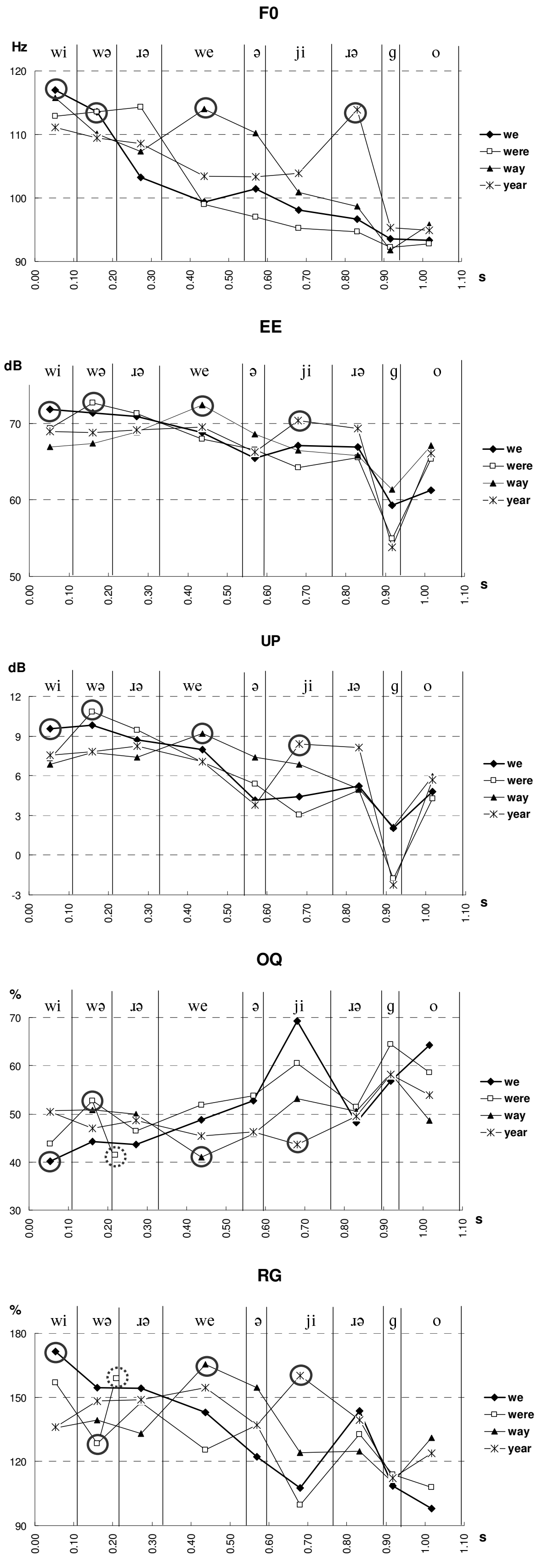

Figure 1: *Stylised voice source parameter trajectories over the course of the utterance. Circles show raised values (reduced for OQ) in focally accented syllables.*

UP values also show a striking correlation with EE values in these utterances with clear peaking in the focally accented syllable. As mentioned for EE, there is a broad tendency towards downdrift in the post-focal part of the utterance.

The RG parameter also shows up an effect of the focal accent, in that the focally accented syllable exhibits the highest RG values. Note that for the accented syllable WERE, the RG peak occurs towards the end of the syllable rather than in the middle, which in this instance exhibited a deep trough.

The OQ parameter showed a tendency towards very low values in the focally accented syllable. In a way that mirrors the RG dynamics, the OQ minimum associated with the accented WERE was delayed and coincided with the right edge of the syllable. Taken together, it is clear that the OQ and RG values have an inverse relationship. This suggests that the glottal pulse shape in the focally accented syllable shows a rather short open phase (low OQ), with a boosting of the second harmonic relative to the first harmonic of the source spectrum. If one remembers that the UP values in focally accented syllables are very high, while OQ values are at their lowest, one can deduce that greater respiratory effort must be major a contributory factor.

Note that for UP and RG there is considerable falloff after the focal accent while values either stay high up to the focal accent or rise to a peak on that syllable. The same can be said for OQ although the trend here is the opposite. The focally accented syllable is associated with the lowest point, while the post-focal material shows an upward trend. This post-focal declination in most source parameters is probably a major aspect of what is heard as deaccentuation which has been described in intonation studies as typical for post-focal material.

Rather surprisingly, the RA the RD parameters showed relatively little effects of focus (not shown in Fig. 1). We would tentatively suggest that, for these utterances at any rate, focal accentuation is being achieved more by the shaping of the lower end of the source spectrum than by any boosting in the higher end.

The spidergrams in Fig. 2 offer a visual way of pulling the various parameters together. In Fig. 2(A) we see values for the focally accented syllable in aWAY (solid black line) compared to the non-focally accented realisations in the other utterances. Similarly, in Fig. 2(B) parameter values for focally accented YEAR (solid black line) are compared to the non-focal realisations. It is evident that the focal realisation is marked by having a stronger excitation (EE), a higher RG and a higher F0. The focal realisations also have the highest UP values. The non-focal realisations, in addition to reduced values for the above parameters, show a tendency towards higher OQ values, suggesting a more lax mode of phonation.

Fig. 3 compares mean parameter values for the syllables WAY and YEAR in the focally accented and non-focally accented realisations. The rate of change of parameter values, calculated as the first order difference of the smoothed parameter values, is also shown in Fig. 3 as delta values. An independent-samples t-test was conducted separately for the WAY and YEAR syllables to compare the mean parameter values and mean delta values in the focally accented and non-focally accented realisations. In both syllables, mean parameter values were found significantly higher for F0, EE, UP and RG and significantly lower for OQ in the focally-accented realisations than in the non-focally accented realisations. The difference in the rate of change (delta values), however, was statistically significant only for EE and RG (higher deltas in the focally accented realisation) for the YEAR syllable, while in other cases the difference was not found to be statistically significant.

4. Conclusions

This study illustrates the suggestion made at the outset of the paper that source parameters are intimately involved as part of the prosody of an utterance. It is clear that in these instances for this speaker focally accented syllables involve increased respiratory effort, with stronger excitation, a smaller open quotient. Deaccentuation in the post-focal material of the utterance is contributed to likewise by a number of parameters: falling RG, UP, EE and rising OQ.

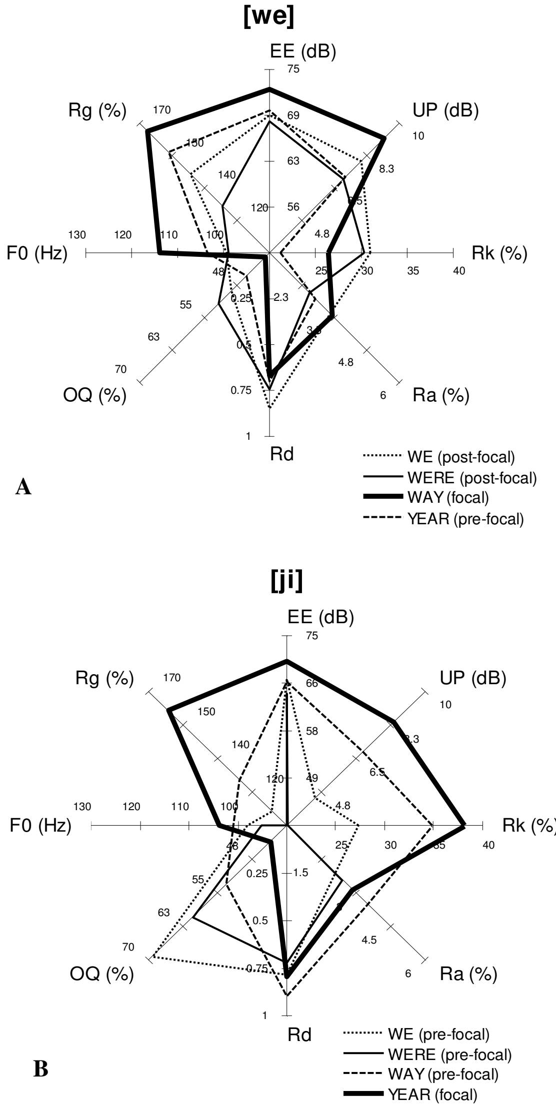

Figure 2: *Source parameter values in syllables WAY (panel A) and YEAR (panel B) in focal and non-focal realisations. Capitalised words indicate the focally accented syllable.*

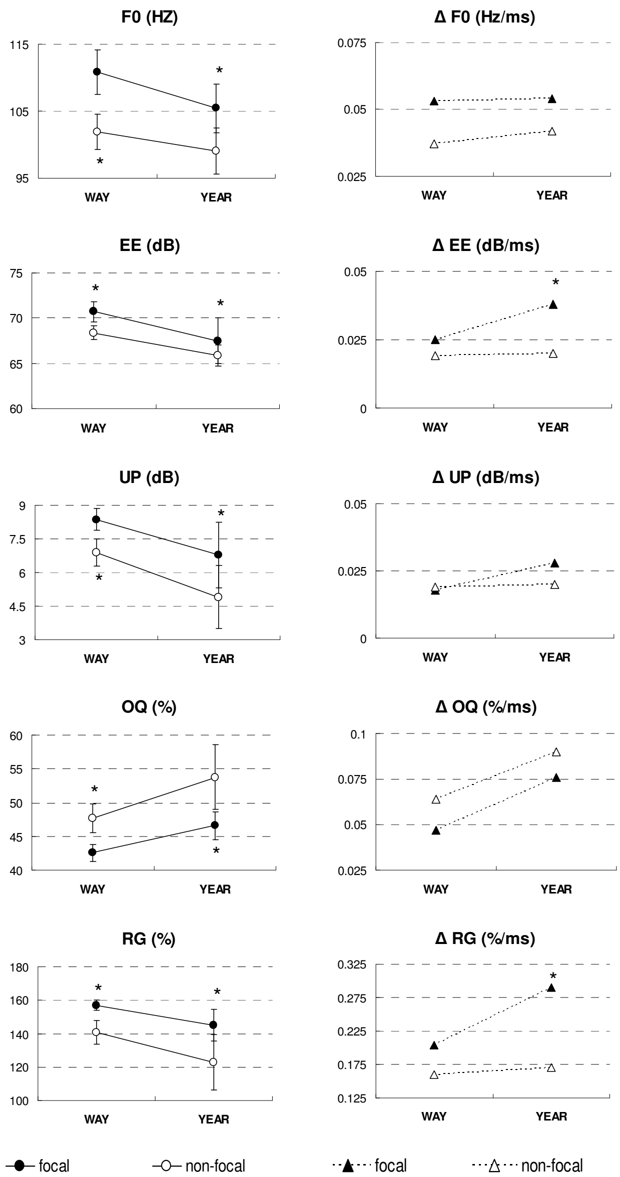

Figure 3: *Mean parameter values and mean rate of change (delta) values for syllables WAY and YEAR in focally and non-focally accented realisations. Statistically significant differences indicated by * (p <0.05).*

We would argue that it is also important to take account of source parameters other than F0 if we want to fully understand prosody. Traditionally, linguists have treated prosody as concerning the dynamic variation of F0, and the general consensus appears to be that intonation is just about the tunes or melodies, i.e. the contours. While pitch is undoubtedly of major perceptual importance, this narrow account is not only a very partial description of the phenomenon, but also often begs a number of questions. For example, if we have a contour where the nucleus has a high falling accent, it is often found that the F0 peak and the fall occurs not on the accented syllable (as heard by the listener), but on the post-accented syllable. Such a case is illustrated in the repetition 'We were away a YEAR ago' (see F0 trace in Fig. 1). One must wonder in a case such as this why it is that YEAR is heard as the focally accented syllable. The likely answer appears in the measures for the other source parameters: EE, RG, OQ, UP, which are all unambiguously enhanced in the syllable YEAR.

This study is a qualitative illustration of what we feel will be required for a fuller understanding of prosody and intonation. As mentioned at the outset, our goal is a holistic account which characterises the behaviour of the source in its entirety.

5. Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 07 / CE / I 1142) as part of the Centre for Next Generation Localisation (www.cngl.ie).

6. References

[1] A. Ní Chasaide and C. Gobl, "Voice quality and f_0 in prosody: towards a holistic account," in *Speech Prosody 2004*, Nara, Japan, pp. 189-196, 2004.

[2] A. Ní Chasaide and C. Gobl, "Decomposing linguistic and affective components of phonatory quality," in *Interspeech 2004*, Jeju Island, Korea, 2004, pp. 901-904.

[3] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics,* vol. 33, pp. 159-197, 2005.

[4] Y. Xu, "Effects of tone and focus on the formation and alignment of F0 contours," *Journal of Phonetics,* vol. 27, pp. 55-105, 1999.

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

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

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

[8] H. Strik, "Physiological control and behaviour of the voice source in the production of prosody. PhD thesis," University of Nijmegen, 1994.

[9] 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.

[10] J. B. Pierrehumbert, "A preliminary study of the consequences of intonation for the voice source," *STL-QPSR,* vol. 30, pp. 23-36, 1989.

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

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

[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. 300-320.

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

[15] D. R. Ladd, A. Schepman, L. White, L. M. Quarmby, and R. Stackhouse, "Structural and dialectal effects on pitch peak alignment in two varieties of British English," *Journal of Phonetics,* vol. 37, pp. 145-161, 2009.