VOICE SOURCE DYNAMICS IN INTONATION

Ailbhe N íChasaide, Irena Yanushevskaya & Christer Gobl

Phonetics and Speech Laboratory, CLCS, Trinity College Dublin, Ireland
anichsid@tcd.ie; yanushei@tcd.ie; cegobl@tcd.ie

ABSTRACT

A qualitative analysis of voice source dynamics is presented for utterances varying in focal placement, with falling (*F*) and rising (*R*) pitch. Source parameters, F0, EE, UP, RG, OQ and RD were obtained by manual inverse filtering and parameterization. Results for *F* sentences suggest that focalization depends on the balancing of focal and non-focal portions of the entire IP. Local prominence-enhancing shifts (towards tenser phonation) are complemented by deaccentuation of post-focal material. While the broad trends are similar in *F* and *R* contours, the rising F0 of the *R* sentences appears to counteract the deaccentuation pattern found with the *F* contours. Thus, focal accentuation yields less source differentiation in the later portions of the IP.

Keywords: voice source, prosody, intonation, F0, focus

1. INTRODUCTION

The paper is concerned with voice source dynamics that are correlated with sentence intonation. Our starting assumption is that voice source dynamics are an inherent dimension of intonation, and that intonational entities such as tonal accent, boundary tone, intonational phrase, etc. involve the modulation of the entire source signal [5], not just F0. We are further interested in how tone-of-voice signals affect (paralinguistic prosody), and feel that to fully understanding it, we need to understand the more narrowly linguistic aspect of source prosody.

This qualitative study looks at voice source dynamics within utterances where the focal accentuation is varied and where the pitch contour is falling (*F*) or rising (*R*). Of particular interest are the following questions. (i) Is focal prominence simply a local perturbation, or does it involve the dynamic relationships of the source parameters over the entire IP? (ii) Are there consistent source correlates of focalization (whether as a local or global IP phenomenon)? (iii) Are these correlates consistent across falling and rising pitch contours?

Although there is limited data on the topic, past studies on the source correlates of focus involving both direct [2, 6-8, 10] and indirect measures [9] have tended on the whole to suggest tenser phonation with focally accented vowel. However, diametrically different results (laxer phonation) have also been reported [10] or indeed no correlates [8]. The direct measures studies have been based on a rather limited sampling, e.g., a single glottal pulse (concatenated for further analyses) or short vowel segments. This is a compromise prompted by the difficulties and effort required to obtain reliable source data (see discussion in [5]). Similarly here, the compromise between quality and quantity of data involved choices. As the major interest here is on source variation over the entire IP, analysis was carried our on all of 8 sentences, pulse-by-pulse.

2. MATERIALS AND METHOD

8 repetitions of 'WE WERE aWAY a YEAR ago' were produced by a male speaker of Irish-English with the narrow focus on each of the four potentially accented syllables (in caps), with both falling (*F*) and rising (*R*) pitch. Using the system described in [3], the signal was inverse filtered and fine-tuned manually, pulse by pulse (1156 pulses). Voice source parameterisation involved fitting the LF model of differentiated glottal flow [1] to the inverse filtered data, yielding measures for: **F0, EE, RG, OQ, UP** and **RD**, described in [5]. **EE** is the strength of the main excitation, defined as the negative amplitude of the differentiated glottal flow. **RG** is the glottal frequency, FG, normalized to F0. **OQ** is the duration of the glottal open phase relative to the whole cycle. **UP** is the peak amplitude of the glottal flow pulse. **RD** is a global waveshape parameter derived from F0, EE and UP. Resynthesis enabled auditory verification of analysis accuracy.

3. RESULTS AND DISCUSSION

Fig. 1 shows trajectories for EE/F0, OQ, RG and RD taken from the sentences with focally accented WERE (black = F, red = R contours). Regression lines fitted to the raw data from the time point of

the F0 peak/dip allows comparison of the slope of parameter values.

Figure 1: Parameter tracks for EE/F0, OQ, RG and RD for the WERE sentence, with regression lines fitted from the time point of F0 peak/dip. Black = *F*; red = *R* sentences. F0 shown in upper panel with dotted lines.

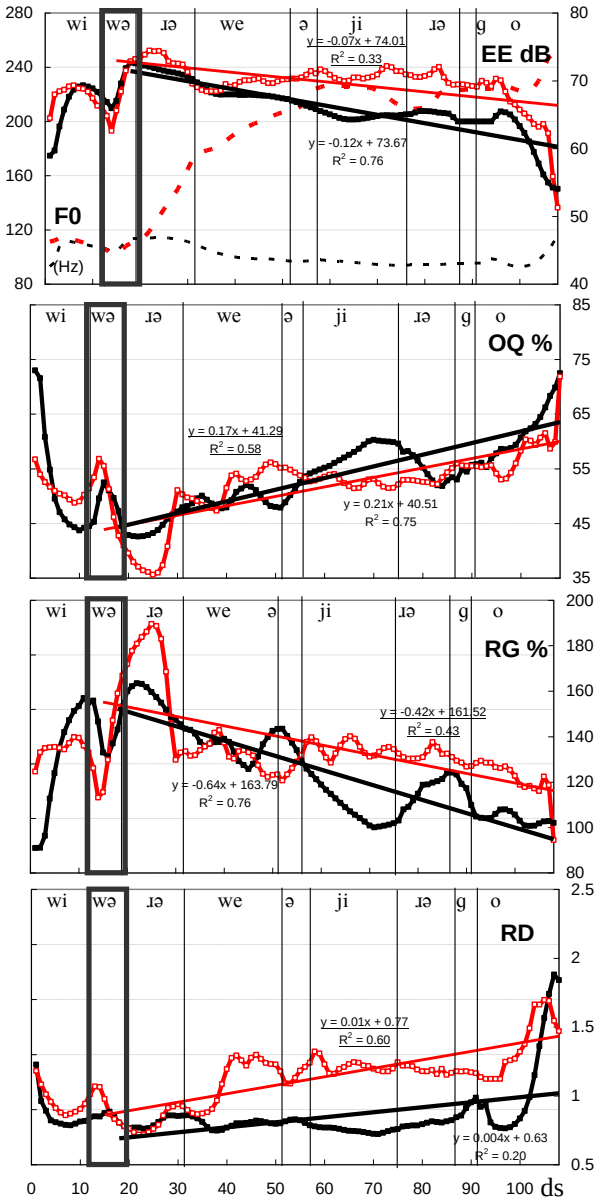


Fig. 2 provides a stylized overview of the data by showing source values (peaks or dips) associated with each syllable, for the sentences with *F* (left panel) and *R* (right panel) contours. Circles/ellipses indicate the values associated with the focal accent in each utterance. The ellipse indicates where a peak or dip, associated with the focal accent is located in the post-focal syllable. Peak delays are well attested in the intonation literature for F0, and these data show similar delays with other source parameters. Solid circles/ellipses are used when the values are the most extreme (highest peak/lowest dip) for a given utterance; broken circles/ellipses show that this was not the case.

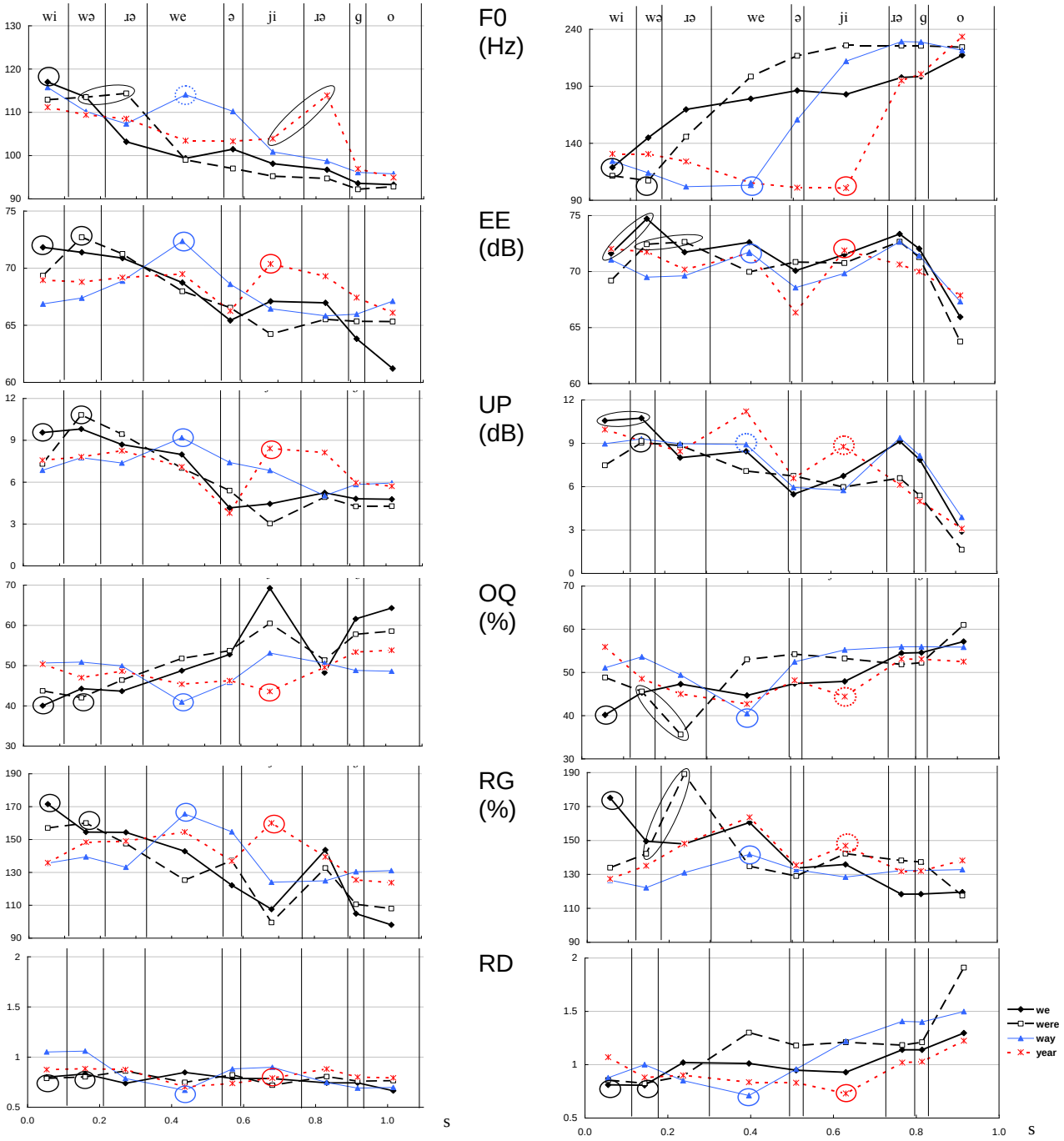
Falling contours (left panel) Focal accent is associated with a high falling tone (H*L), with peak delay in two cases (WE, YEAR). It is also associated with higher EE and UP (a stronger source excitation, and overall spectral boosting). A concomitant RG peak along with a trough in OQ (delayed in the WERE sentence, see also Fig. 1) suggest a boosting of the lower end of the source spectrum. Taken together, these source features suggest a rather tenser mode of phonation as a marker of focal accent, with peaks/troughs sometimes achieved in the post-focal syllable. The raised UP with the lower OQ suggests increased respiratory effort, along with increased laryngeal tension. The RD parameter, as would be expected, shows little differentiation.

Although focal accent yields the most extreme source values within each utterance, it is striking that the relative dominance of the focal element is greater in the earlier (WE and WERE) than in the later parts of the utterance. This suggests an underlying declination effect, where, as with F0, the accent-related peaks are relative to an overall downdrift.

Deaccentuation of the post-focal material is also evident, not only in F0 but also in the other source parameters. Along with the F0 lowering, one sees post-focal lowering of EE, UP, RG and a rising of OQ (a laxer mode of phonation). Focalization appears to be achieved by a dual process: local enhancement of the focally accented element and dis-enhancement in (particularly) the post-focal material. Consequently, in comparing across utterances, although the focus-related source values are in all cases the most extreme, the cross-utterance differences between focalized/non-focalized items is much more marked in the later portions of the utterance, where focal accentuation is being contrasted with deaccentuation. Thus, differences across utterances are greater for aWAY and YEAR than for WE and WERE.

Rising contours (right panel) The *R* sentences exhibit broadly similar trends, but to a lesser degree. The focally accented syllable is associated with a low rising tone (L*H), whose elbow is located within the accented syllable, with the rise continuing through the post-focal material. The

Figure 2: Stylized voice source parameter tracks in the *F* (left panel) and *R* (right panel) sentences. Circles show parameter values in the focally accented syllables.



post-focal ***deaccentuation*** trends observed in the *F* sentences are present but greatly attenuated in the *R* set. The regression lines (Fig. 1) show the trend in the source parameters, estimated from the F0 dip in *R*, or the F0 peak in *F* to the end of the IP. The regression lines are shallower in the *R* set. For EE, UP, RG the slopes remain negative, but much less dramatically so. However, the updrift of OQ associated with deaccentuation in *R* contours is about the same for the *F* sentences.

Thus, despite some similar trends (particularly for OQ), the sharply rising F0 in the post-focal material appears to counter many of the effects of deaccentuation. From a physiological point of view, this makes sense: considerable laryngeal tension must be required to effectuate the very high rise observed here in *R* sentences, which have

a range of 90-240 Hz, compared to the range of 90-120 Hz in the *F* sentences. It seems logically predictable that the necessary increase in laryngeal tension would counter deaccentuation-related tendencies during the rise.

One striking way the *R* sentences differ from the *F*, is in the rise of the RD parameter which appears to be related to the rise in F0. The rising RD is unlikely to be an indicator of deaccentuation (note, in contrast the stability of RD in *F* sentences). Rather, it seems to corroborate the findings and suggestions in [4], that, at high pitches, the shape of the glottal pulse alters in a way that is not necessarily indicative of a laxer phonation. One cannot thus presume that RD and NAQ are reliable indicators of phonation quality over large variations in F0, as often assumed.

Broadly speaking, the ***local*** effects of focal accentuation tend to involve similar source changes to what was found for the *F* sentences, but to a lesser degree. It is striking (as with the *F* set) that focal accentuation is more clearly marked when it occurs earlier than in the later portions of the utterance. In focalized WE and WERE, the trends are rather like those observed for *F*. Within any utterance, EE, UP, RG values all appear to peak, while OQ values are at their lowest. For the syllable WAY these observations mostly still hold, but the peaks/troughs are less extreme and the focal values are not always the most extreme across the four focal conditions. For the final focalized YEAR source parameters are neither consistently more salient within a given utterance, nor across utterances/focal conditions. This, we surmise, largely reflects the suspension of deaccentuation of post-focal material in the *R* set.

4. CONCLUSIONS

Results suggest that voice source dynamics are part and parcel of focalization and that a dual process is involved which potentially encompasses the entire IP: (1) ***local*** source prominence enhancement (peaks in EE, UP and RG along with a trough in OQ) which entail a somewhat more tense mode of phonation, and (2) source prominence dis-enhancement, or more lax mode of phonation, involving falling EE, UP and RG along with a rising OQ over the post-focal material in the utterance. As regards the local enhancement, the peaks/troughs may sometimes be fully achieved in the post-focal syllable. A difference also emerged between realizations of focalization in falling and rising contours. In *R*, the sharply rising F0 appears to largely counteract the second of the above processes (deaccentuation), and in a related way, source correlates of focal accentuation become increasingly less in evidence the nearer one gets to the end of the IP.

To conclude, we suggest that source dynamics are inherently part and parcel of intonation. This particular study, being of limited utterances by a single speaker, will need a fuller elaboration with more data. Finally, we would emphasize that the dynamic shifts in mode of phonation do not entail auditory shifts in voice quality, but are simply part of the prosody in an utterance that would be described as modal.

5. ACKNOWLEDGEMENTS

The research was 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] Fant, G., Liljencrants, J., Lin, Q. 1985. A four-parameter model of glottal flow. *STL-QPSR* 4, 1-13.

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

[3] Gobl, C., N íChasaide, A. 1999. Techniques for analysing the voice source. In Hardcastle, W.J., Hewlett, N. (eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge: CUP, 300-321.

[4] Gobl, C., N íChasaide, A. 2003. Amplitude-based source parameters for measuring voice quality. *Proc. of the VOQUAL'03*, Geneva, Switzerland.

[5] Gobl, C., N íChasaide, A. 2010. Voice source variation and its communicative functions. In Hardcastle, W.J., Laver, J., Gibbon, F.E. (eds.), *The Handbook of Phonetic Sciences* (2nd ed.). Oxford: Blackwell, 378-423.

[6] Iseli, M., Shue, Y.L., Epstein, M. A., Keating, P., Kreiman, J., Alwan, A. 2006. Voice source correlates of prosodic features in American English. *Proc. of the Interspeech 2006.* Pittsburgh, Pennsylvania, USA.

[7] Koreman, J. 1996. *Decoding \Llinguistic Information in the Glottal Airflow.* Ph.D. thesis. University of Nijmegen.

[8] Murphy, P. 2007. Voice source change during fundamental frequency variation. *Verbal and Non-Verbal Communication Behaviours.* Berlin: Springer, 165-173.

[9] Sluijter, A.M.C., van Heuven, V.J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustic Soc. of America* 100, 2471-2485.

[10] Vainio, M., Airas, M., Järvikivi, J., Alku, P. 2010. Laryngeal voice quality in the expression of focus. *Proc. of the Interspeech 2010*.