PROSODY OF VOICE: DECLINATION, SENTENCE MODE AND INTERACTION WITH PROMINENCE

Ailbhe Ní Chasaide, Irena Yanushevskaya, Christer Gobl

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

ABSTRACT

This paper looks at voice source correlates of declination. This follows a line of research that proposes that prosody is primarily signalled by the temporal modulation of the voice source, and not just F0. Inverse filtering and source parameterisation was carried out on twelve 3-accent sentences of Connemara Irish (Declaratives, WH questions and Yes/No questions). All have a falling intonation and a H* H* H*+L pattern. The measured voice source parameters indicate a declination of other aspects of the voice: a decline in the strength of the source excitation (EE), an increasingly lax phonation over the course of the sentence, as indicated by the decline in the closed quotient (CQ) and the normalised glottal frequency (RG). These indicators of 'source declination' were found across sentence modes. However, there appear to be source differences between the declaratives and the question sentences in terms of the overall levels of EE, CQ, RG and F0. The interaction with the accentuation pattern of the sentence is discussed.

Keywords: voice source, prosody, declination.

1. INTRODUCTION

We are investigating the 'prosody of the voice', pursuing the proposition that prosody is primarily carried by the dynamic, temporal variation of the voice source, encompassing not only melodic characteristics (F0 dynamics) but also other dimensions, that would be regarded as relating to the quality and loudness of the source signal. At one end prosody carries the paralinguistic signalling, and shifts in voice quality are associated with the signalling of one's mood, attitude and interpersonal stance [6], [9], [21]. Paralinguistic prosody can involve rather major obvious shifts in voice quality – in layman terms, tone of voice. Rather less obvious is the fact that even for the most affectively neutral utterance, there is extensive modulation of the voice, modulation that is in our view an expression of the essentially linguistic prosody. Some earlier studies [4], [8], [13-16], [19], [20] examined the voice source correlates of linguistic prosody, looking at source parameters that may be associated with phenomena such as accentuation, focus, deaccentuation, etc., typically thought of and defined in terms of their F0 correlates, but, we suggest, depend on the properties of the entire voice source. Including voice source parameters additional to F0 in the prosodic description should provide new insights into the nature of linguistic prosody, and should furthermore establish the underlying prosodic scaffolding on which paralinguistic expression sits. Ultimately, the aim is a more holistic account of prosody, not only in terms of its production and acoustic specification, but also in treating linguistic and paralinguistic aspects within a single framework.

In this paper, we explore the voice source correlates of declination. Declination in intonation analysis is defined as the downward trend of F0 over the course of an utterance [12], [17], [18]. It is observed in many languages mainly in declaratives (where the pitch is most typically falling) and may be suspended in utterances realised with a rising pitch, e.g., questions in many languages. The present study looks specifically at voice source correlates of declination, building on an earlier analysis of the correlates of accentuation [16]. We return to this earlier study in our discussion section below.

2. MATERIALS AND METHOD

2.1. Utterances analysed

The utterances analysed were elicited from a female speaker of Connemara Irish. The sentences were of three sentence modes: Declaratives, WH questions and Yes/No questions. There were four utterances for each sentence mode, i.e. 12 sentences in total, and these were elicited within scripted mini-dialogues. The recordings were conducted in a quiet room using a high quality head-mounted microphone. The sentences and their word-for-word translation are shown below, with the accented syllables in caps.

DEC Bhí CIAN ag CAINT leo sa MArgadh.
[vʲiː ciənˠ ə kaɲtʲ lʲɔː sə marˠəgə]
'was Cian at speaking with them in the market'

WH-Q (Is) CÉ bheas á MBAiliú ón MArgadh?
[(ɪs) ceː vʲes ɑː maʎu ɔn marˠəgə]

'and who will be at their collecting from the market'

YN-Q Raibh CIAN ag CAINT léi sa MArgadh.
[ɾˠo ciən ə kaɲt̪ʲ lʲɔ: sə marˠəgə]
'was Cian at speaking with them in the market'

All these utterances were realised with three accents, they have an overall falling pitch, and would in virtually all cases have been transcribed as H* H* H*+ L.

2.2. Inverse filtering and source parameterisation

Central to research on the voice source is the development of analysis tools and algorithms that will yield robust measures of voice source dynamics [10], [11]. A flexible analysis system GlóRí [2] is currently being developed and an initial version of this system is used in the present analysis. The voice source analysis was done using the automatic component of the GlóRí system and the procedure can be described as follows. Initially, glottal closure instants (GCIs) are detected using the SE-VQ algorithm [10]. Inverse filtering is then carried out using the iterative and adaptive inverse filtering method (IAIF) developed by P. Alku [1], 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 [5]. For the model fitting, we use the recently developed DyProg-LF method, which has been shown to be more robust than previous methods [11]. 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, which is a measure of the degree of symmetry of the glottal pulse; RG, the normalised glottal frequency, which is the characteristic frequency of the glottal pulse normalised to F0; OQ, the open quotient, which is the proportion of the glottal cycle for which the glottis is open (note that OQ is determined by RK and RG); CQ, the closed quotient is derived from OQ as 1 - OQ. Together, RG, RK and OQ are indicators of the levels of the low frequencies of the source spectrum. For further discussion of these parameters, see [7].

2.3. Data representation

In each syllable of each of the 12 utterances, a representative data sample was taken approximately in the middle of the vowel, away from the consonantal perturbations. These sample points were averaged across the utterances per sentence mode. The declination trends for F0 and each measured source parameters were modelled by fitting regression lines to these data points for each sentence mode. Figure 1 (right panel) illustrates the time-normalised declination for the different sentence modes for the following individual parameters, F0, UP, EE, RG, and CQ. Note that RG and CQ are expressed here as percentages. Declaratives are shown with a dotted red line; WH questions are indicated with a solid black line and YN questions are shown with a dashed blue line. The slope and the R^2 (goodness of fit) values are shown in Table 1. In the left panel of Figure 1 is shown the extent of the declination in each sentence mode, as Hz per second for F0, as dB per second for UP and EE, percentage points per second for RG and CQ.

3. RESULTS AND DISCUSSION

3.1. Declination

As is clear from Figure 1, declination is realised not only in F0 but also in parameters of the source. Mirroring the declination of F0 in all three sentence modes, there is a declination of EE, indicating a reducing level in the voice source excitation strength. Along with this, we see a declination in RG and CQ. Thus, in the course of these utterances, there is a gradual decrease in the level of the source excitation and an increase in the relative dominance of the lower end of the source spectrum, as would be indicative of an increasingly lax mode of phonation. UP is largely unchanging, and only in the case of the Yes/No questions is there any declination trend.

3.2. Declination and sentence mode

Voice source declination (including clear F0 declination) is observed across the different sentence modes. There are however some striking differences in the realisation for declaratives and questions.

Figure 1: Declination trends for selected voice source parameters.

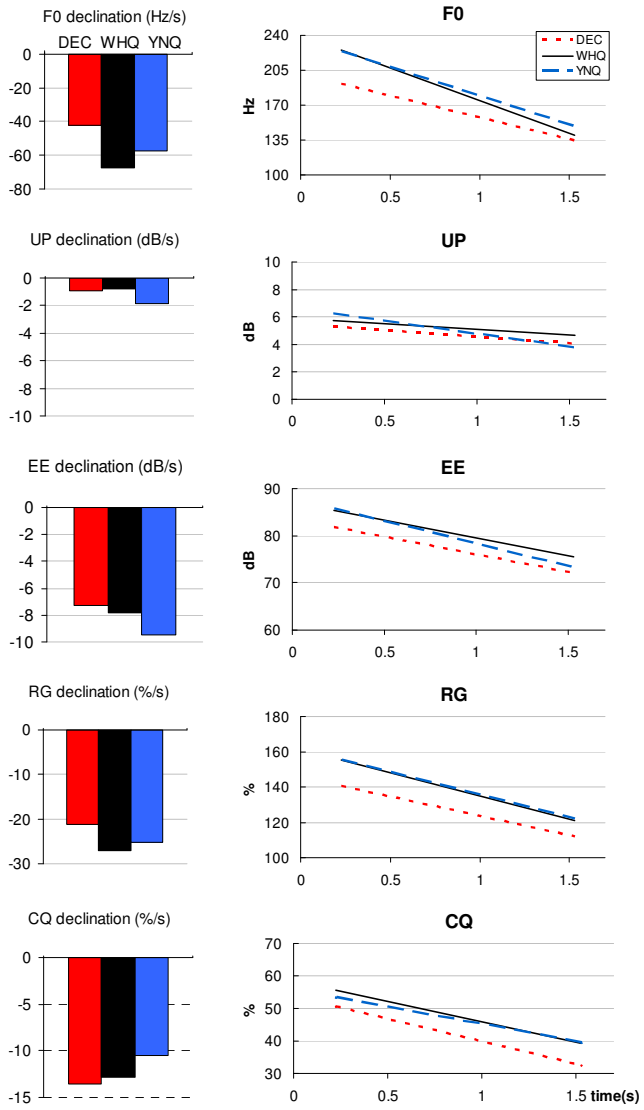


Table 1: Declination slopes and R^2 goodness of fit values for selected voice source parameters.

	DEC	WHQ	YNQ
F0 (Hz/s)	-42.4 $R^2 = 0.80$	-67.5 $R^2 = 0.90$	-57.4 $R^2 = 0.76$
UP (dB/s)	-0.92 $R^2 = 0.07$	-0.83 $R^2 = 0.06$	-1.9 $R^2 = 0.26$
EE (dB/s)	-7.3 $R^2 = 0.49$	-7.8 $R^2 = 0.62$	-9.5 $R^2 = 0.48$
RG (%/s)	-21.2 $R^2 = 0.48$	-27.0 $R^2 = 0.41$	-25.2 $R^2 = 0.35$
CQ (%/s)	-13.5 R^2=0.53	-12.9 R^2=0.45	-10.5 R^2=0.27

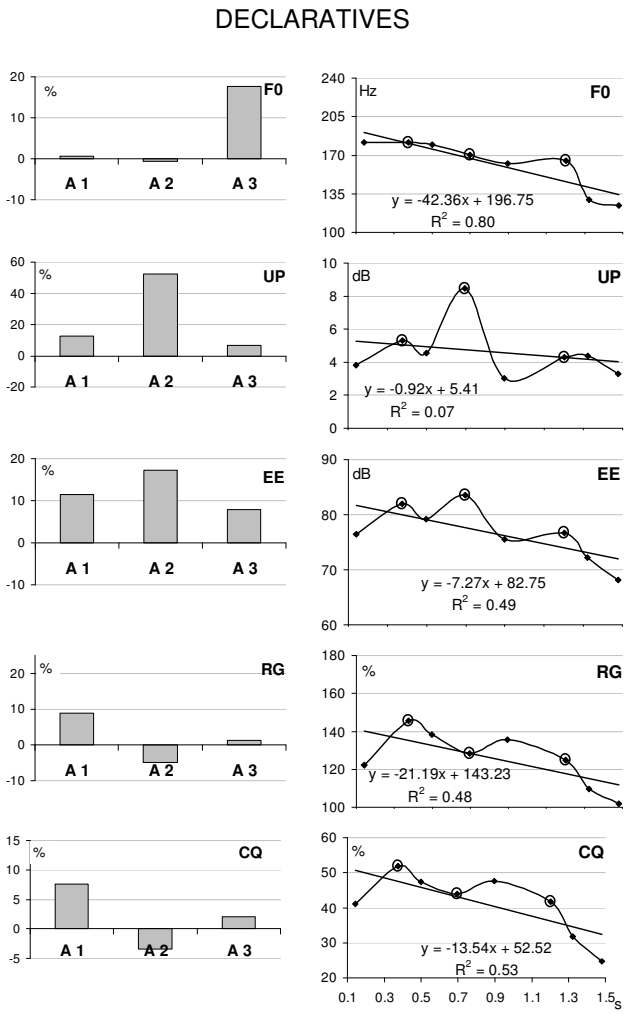
Although the amount of data for each sentence mode is limited, it seems worth noting these differences here. Sentence mode appears to occasion differences in the overall level of the source parameters – including the level of F0. It is striking that in both types of questions, WH and Yes/No questions, the level of EE, RG and CQ are raised, suggesting a subtle differentiation in terms of phonation mode. This co-occurs with a raised level of F0. It should be mentioned that in Irish, questions are not differentiated from the declarative forms in terms of the melodic contour: in all cases the basic contour is a falling one [3]. The present data suggest that the overall levels of the source parameters may be important in the differentiation of declaratives and questions.

3.3. Declination and syllable prominence

An earlier study [16] looked at the source correlates of accentuation in these same data. In this study the authors sought support for the Prominence Hypothesis. This hypothesis proposes that prominence arises from the contribution of different source parameters, and that the extent to which a particular parameter contributes can vary. Thus, for example, accentuation of syllables which have no pitch prominence is signalled by other parameters of the source. The present data set was of particular relevance to testing the hypothesis, as in all of these sentences – heard as having three accents (and transcribed as sequences of H* H* H*+L, a flat-hat pattern) – it was striking that the middle H* accent was not associated with any particular pitch excursion that might lend it prominence. According to the authors' hypothesis, the middle accent should achieve salience through excursions of voice source parameters other than F0. The results of that study supported the hypothesis, and the middle accent in all cases had prominence-lending excursions to other source parameters.

Figure 2 (left panel) illustrates a portion of the results of the earlier study. For each of the three accented syllables (A1, A2 and A3) is shown the size of the excursion of a particular parameter (F0, UP, EE, RG and CQ), measured as the distance between the value of the accented syllable relative to that of the adjacent unstressed syllables. In each case the value shown is relative to the overall range of values in a particular parameter for the whole utterance. Note that in the dataset shown here neither A1 nor A2 have pitch prominence as such, and the prominence derives from the other source adjustments. For further elaboration, see [16].

Figure 2: Selected voice source parameters in signalling prominence and declination.



The right-hand panel in Figure 2 shows the declination trend in these same parameters for the same data. The thing to note is that the source is simultaneously contributing to the multiple aspects of the sentence prosody. The local perturbations that are part and parcel of the accentuation signalling sit on top of declination baseline. This declining baseline is common to the different sentence modes. However, in the case of these Irish data, sentence mode differentiation appears to implicate the overall levels of the parameters, and perhaps sometimes also the declination slope.

4. CONCLUSIONS

These data support the initial suggestion that declination is a phenomenon which is manifested, not just in the pitch contour of the utterance, but in the different dimensions of the voice source. Although the amount of data for each sentence mode is rather small, the results would prompt us to suggest that sentence mode differentiation in this variety of Irish also exploits the many dimensions of the voice source, manifested by differing levels of F0 and other source parameters for questions compared to declaratives.

The present data add to the picture that emerged in an earlier analysis of voice source variation in accentuation, where it was also clear that accentuation equally draws on many aspects of the voice source signal. Prominence-lending perturbations can be observed in F0, EE, UP, RG and CQ, even though a particular parameter need not always be implicated. Importantly, the prominence-lending modulations of these individual parameters occur against a declining baseline, and an overall level setting, which we tentatively suggest reflects sentence-mode.

Taken together, these data support the underlying proposition of the present work: that prosody is signalled by the entire voice source, and that an account of intonation featuring only F0 presents but a partial account that can potentially mislead.

As mentioned, these studies are based on a limited data set, and a priority for future work will be to extend the work to more and more varied data. Perceptual testing of our hypothesis and of the relevance of these findings is another avenue we hope to explore.

5. ACKNOWLEDGMENTS

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, intonation transcription and for the discussion of the data.

6. REFERENCES

[1] Alku, P. 1992. Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering. *Speech Communication,* 11, 109–118.

[2] Dalton, J., Kane, J., Yanushevskaya, I., Ní Chasaide, A., Gobl, C. 2014. GlóRí - the glottal research instrument. *Proc. Speech Prosody 2014*, Dublin, 944–948.

[3] Dorn, A., O'Reilly, M., Ní Chasaide, A. 2011. Prosodic signalling of sentence mode in two varieties of Irish (Gaelic). *Proc. XVIIth ICPhS*, Hong Kong, 611–614.

[4] Epstein, M. 2003. Voice quality and prosody in English. *Proc. XVth ICPhS* Barcelona, 2405–2408.

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

[6] Gobl, C., Ní Chasaide, A. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication,* 40, 189–212.

[7] Gobl, C., Ní Chasaide, A. 2010. Voice source variation and its communicative functions. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds.), *The Handbook of Phonetic Sciences*, 2 ed. Oxford: Blackwell Publishing, 378–423.

[8] 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. Interspeech 2006 - ICSLP*, Pittsburgh, Pennsylvania.

[9] Juslin, P. N., Scherer, K. R. 2005.Vocal expression of affect. In: Harrigan, J., Rosenthal, R., Scherer, K. R. (eds.), *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford: Oxford University Press, 65–135.

[10] Kane, J., Gobl, C. 2013. Evaluation of glottal closure instant detection in a range of voice qualities. *Speech Communication,* 55, pp. 295–314.

[11] Kane, J., Gobl, C. 2013. Automating manual user strategies for precise voice source analysis. *Speech Communication,* 55, 397–414.

[12] Ladd, D. R. 2008. *Intonational Phonology*, 2 ed. Cambridge: Cambridge University Press.

[13] Ní Chasaide, A., Gobl, C. 2004. Decomposing linguistic and affective components of phonatory quality. *Proc. Interspeech 2004*, Jeju Island, Korea, 901–904.

[14] Ní Chasaide, A., Gobl, C. 2004. Voice quality and f0 in prosody: towards a holistic account. *Proc. Speech Prosody 2004*, Nara, Japan, 189–196.

[15] Ní Chasaide, A., Yanushevskaya, I., Gobl, C. 2011. Voice source dynamics in intonation. *Proc. XVIIth ICPhS*, Hong Kong, 1470–1473.

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

[17] Nolan, F. 1995. The effect of emphasis on declination in English intonation. In: Windsor Lewis, J. (ed.), *Studies in English and General Phonetics*. London: Routledge, 241–254.

[18] 't Hart, J., Collier, R., Cohen, A. 1990. *A Perceptual Study of Intonation: an Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

[19] Vainio, M., Airas, M., Järvikivi, J., Alku, P. 2010. Laryngeal voice quality in the expression of focus. *Proc. Interspeech 2010*, Makuhari, Japan, 921–924.

[20] Yanushevskaya, I., Gobl, C., Kane, J., Ní Chasaide, A. 2010. An exploration of voice source correlates of focus. *Proc. Interspeech 2010*, Makuhari, Japan, 462–465.

[21] Yanushevskaya, I., Ní Chasaide, A., Gobl, C. 2011. Universal and language-specific perception of affect from voice. *Proc XVIIth ICPhS*, Hong Kong, 2208–2211.