



# Modelling the timing and scaling of nuclear pitch accents of Connaught and Ulster Irish with the Fujisaki model of intonation

*Maria O'Reilly, Ailbhe Ní Chasaide*

Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences  
Trinity College Dublin, Ireland

moreil12@tcd.ie, anichsid@tcd.ie

## Abstract

Connaught and Ulster Irish (Gaelic) use two diverse intonation patterns in declaratives: Connaught typically uses a sequence of falling (H\*+L) accents, while Ulster typically employs a sequence of rising (L\*+H) accents. In this paper the Fujisaki model is used to simultaneously capture the timing and scaling of these nuclear (i.e. IP-final) accents with the timing (T) and amplitude (Aa) accent command parameters. The speech materials include a set of simple declaratives with a nuclear syllable followed by 0, 1 and 2 tail syllables (henceforth N0, N1 and N2). The results demonstrate that the nuclear accents in both Connaught and Ulster Irish are sensitive to tail length with respect to timing but not scaling. Thus, the f0 inflection is timed earlier, i.e. pushed leftwards, in the absence of a tail. The scaling is rather impervious to tail length. The timing and scaling parameters of the accent command were compared with hand-measured contour-derived measurements. Overall, high correlations between the two sets of measurements indicate that the Fujisaki model adequately captures the fine-grained aspects of the nuclear accent realisation in varying tail length conditions.

**Index Terms:** Connaught Irish, Ulster Irish, nuclear accent, timing and scaling, Fujisaki model of intonation, accent command

## 1. Introduction

This paper describes the timing and scaling of nuclear accents in two rather diverse dialects of Irish, using the timing (T) and amplitude (Aa) parameters of the Fujisaki model [1-3]. This is part of a wide-ranging investigation of the prosody of Irish.

The analysis deals with the following two dialects of Irish: the Connaught dialect of Cois Fharráige (henceforth referred to as CF), and the Ulster dialect of Gaoth Dobhair (henceforth referred to as GD). The locations of the dialects are shown by the red arrows in Figure 1.

The intonation of declaratives and questions, and tonal alignment in the Irish dialects are described in recent work [4-9]. Although discussion in this paper focuses uniquely on the Fujisaki modelling of these accents, we have been using the Fujisaki model in parallel with a more traditional linguistic AM analysis, using the ear and IViE labelling. Throughout, the Fujisaki modelling is constrained to capture only linguistically relevant events (in terms of the number and location of accents and IPs).

The Fujisaki modelling is deemed important for this work on Irish for a number of reasons. It is helpful to provide

quantification of the intonational characteristics of these dialects. As in parallel research we are developing multi-dialect text-to-speech for Irish dialects [10, 11], this explicit quantitative modelling will, we hope, facilitate the generation of nuanced, dialect-specific prosody in synthesis.

The exercise also allows us to explore areas where there may be a possible mismatch in the outputs of our two descriptive approaches, e.g., cases where the quantitative model might fail to capture an aspect of the prosodic contour that could turn out to be linguistically important. We are also interested in using the model to frame hypotheses concerning the underlying control of prosody in these dialects.

This paper asks the following questions:

- (1) How does tail length affect the timing (T) and scaling (Aa) of the nuclear accents in the two dialects of Irish, CF (Connaught) and GD (Ulster)?
- (2) Are there differences in the effects of varying tail length between the two diverse dialects?
- (3) How well does the Fujisaki model capture the timing and scaling dimensions of the nuclear accents?

We address this last question by comparing the T and Aa model parameters to parallel contour-based measurements.

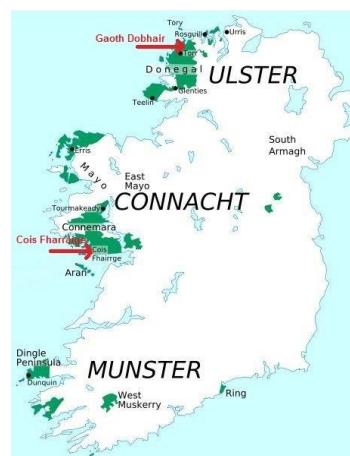


Figure 1: Map showing dialect locations: Cois Fharráige in Connaught and Gaoth Dobhair in Ulster.

## 2. Methods and materials

The nuclear accent data were obtained from nine informants representing the two dialects. For Connaught, the four informants are from Cois Fharráige on the coast of County

Galway (Figure 1). For Ulster, the five speakers are from Gaoth Dobhair in County Donegal (Figure 1). All informants are native speakers of the local variety with Irish as their first language. All speakers were between 20 and 60 years of age at the time of recording. None of them are professional speakers (e.g. newsreaders, actors), and use Irish on a daily basis.

The speech materials include three simple phrases in which the nuclear (N) syllable is followed by a systematically varied number of unstressed syllables (from 0 to 2, zero denoting the absence of tail). The target phrases are presented in Table 1. These two-accent phrases were designed to elicit the production of a single intonation phrase (IP) and were embedded in mini-dialogues. Four repetitions (five for two Cois Fharrage speakers) of each target phrase were obtained, thus giving 114 tokens in total.

Table 1: Target phrases for CF and GD in the nuclear accent dataset with their translations into English. The accented syllables are marked in bold font.

| Tail length | N0       | (CF)  | Bhí an carr seo <b>lán</b> .<br><i>Was the car this full</i><br>‘This car was <b>full</b> .’         |
|-------------|----------|---|--|
|             |          | (GD)  | Tá mo <b>mhála lán</b> .<br><i>Is my bag full</i><br>‘My bag is <b>full</b> .’                       |
|             | N1       | (CF)  | Tá'n citeal <b>lán a'm</b> .<br><i>Is the kettle full at me</i><br>‘I've the kettle <b>filled</b> .’ |
|             | (GD)     | Tá'n citeal <b>lán de</b> .<br><i>Is the kettle full of</i><br>‘The kettle's <b>full</b> of it.’        |  |
| N2          | (CF, GD) | Tá'n citeal <b>lán aici</b> .<br><i>Is the kettle full at her</i><br>‘She's the kettle <b>filled</b> .’ |  |

The Cois Fharrage and Gaoth Dobhair recordings were carried out in a semi-anechoic room at the Phonetics and Speech Laboratory in Trinity College Dublin. The randomised prompts were presented on a computer screen. The data were recorded at the 44 kHz sampling frequency onto a PC using the Audacity software [12].

The f0 contours were processed in Praat [13]. F0 was extracted using the autocorrelation method; any octave jump and segmental f0 errors were accordingly hand-corrected.

The prosodic annotation of the data was carried out with the IViE labelling system [14, 15]. The IViE analysis enabled grouping of the data according to tune type for the subsequent Fujisaki model analysis. The IViE annotation of the nuclear data is illustrated in Figure 2.

The IViE analysis revealed that 98% of the Cois Fharrage data contained a nuclear fall ( $H^*+L\%$ ). In the Gaoth Dobhair data 85% of the nuclear contours featured a rise ( $L^*+H\%$ ); the remaining 15% contained a nuclear rise-fall ( $L^*+HL\%$ ) – the latter were produced exclusively by one speaker and were excluded from further analysis. The selected nuclear data include the following tunes:  $H^*+L$   $H^*+L\%$  and  $H^* H^*+L\%$  in

CF;  $L^*+H$   $L^*+H\%$  and  $H^* L^*+H\%$  in GD. The timing and scaling measurements for the nuclear accents in these tunes are covered in the Results section.

Next, the data was analysed with the Fujisaki model using the FujiParaEditor tool [16]. The modelling is illustrated in Figure 2. To summarise it briefly, the model decomposes an f0 contour into a set of phrase commands and accent commands which can be related to prosodic phrases and accents, respectively. The phrase and accent commands are superimposed onto the base frequency asymptote, Fb.

The Fujisaki modelling approach adopted in our work is explained in more detail in [17, 18]. In short, Fb is treated as utterance-dependent, similarly to [19] and set at -0.5 semitones below the global f0 minimum (usually found IP-finally). The phrase and accent component constants are set to 3.0/s and 20.0/s, respectively [2, 3]. Only one phrase command is permitted per IP. Each accent command corresponds to a pitch accent or a high IP boundary. Negative accent commands are only permitted in nuclear  $L^*+H$  accents (as in GD, Figure 2), for which a negative-positive command pair is used in order to accurately capture the depth of the scoop.

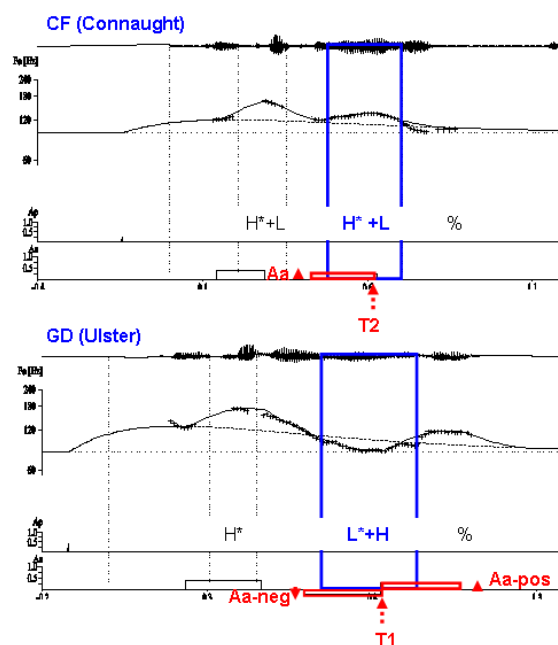


Figure 2: IViE annotation and Fujisaki modelling of the nuclear data in CF (upper panel) and GD (lower panel). Each panel shows (from top to bottom): waveform, extracted (crosses) and model-matched (solid line) f0 contours, IViE annotation, Fujisaki model input phrase command and accent commands (red boxes).

In this paper only the accent command parameters are of interest. These include the timing (T) and amplitude (Aa) of the accent command (Figure 2). In our approach, T models the timing of the f0 inflection pertaining to an accent. Thus, the onset of the fall is captured with T2 ( $H^*+L$  in Figure 2); the onset of the rise is captured with T1 of the positive accent command ( $L^*+H$  in Figure 2). Aa models the f0 excursion size of an accent: Aa of the single positive command for the nuclear fall ( $H^*+L$  in Figure 2), or the sum of absolute Aa

values of the negative-positive pair for the nuclear rise ( $L^*+H$  in Figure 2).

In order to objectively evaluate the performance of the Fujisaki model accent component in the nuclear accent data, parallel contour-derived measurements were taken for comparison. The two sets of measurements are presented in Figure 3.

The timing of the  $f_0$  inflection of the nuclear fall (or rise, depending on dialect) is estimated with the relevant T of the accent command: T2 (CF, Figure 2), or T1 of the positive accent command (GD, Figure 2). The T parameter is expressed in relative terms, i.e. as a percentage of the accented syllable duration (T-rel in Figure 3). The parallel contour-derived  $t^*$  measurement (Figure 3) similarly quantifies the timing of the  $f_0$  inflection which is also expressed as a percentage of the accented syllable duration.

The size of the nuclear fall (or rise) is estimated with Aa of the single positive accent command (CF), or the sum of absolute Aa values of the negative and positive command pair (GD) (Figure 2). The  $f_0$  excursion size of the nuclear accent (Aa-total in Figure 3) is compared to its contour-measured counterpart ( $f_0$  size in Figure 3).

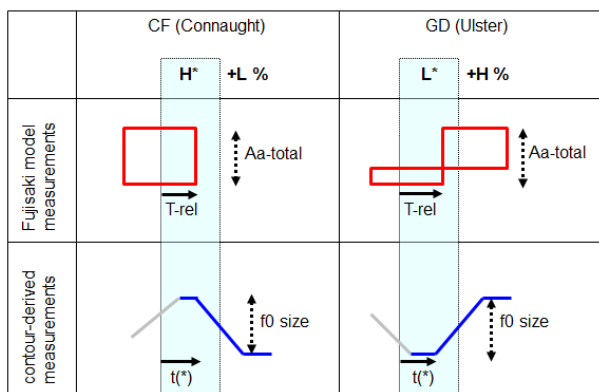


Figure 3: Schematic representation of the Fujisaki model measurements and the corresponding contour-derived measurements in the nuclear accent data. The light blue panels represent the accented syllable.

### 3. Results

This section presents the timing and scaling results for the nuclear accents in CF (Connaught) and GD (Ulster). The ANOVA tests of T-rel and Aa-total are done with factors Dialect and Tail Length. For the purpose of the ANOVA tests, the repetitions are averaged for each Tail Length condition for each speaker (four CF speakers and four GD speakers), thus disregarding the type of the preceding accent. Exact  $p$  values are reported for the ANOVA results. Only the significant ( $p < .05$ ) differences of the post-hoc Tukey HSD and Student's  $t$ -tests are presented.

#### 3.1. Timing of $f_0$ inflection of fall and rise

Figure 4 presents the results for the timing of the onset of the nuclear fall (CF) and rise (GD) estimated with T-rel.

The results in Figure 4 demonstrate that the timing of the  $f_0$  inflection in the nuclear accent is affected by tail length. Notably, T-rel is earlier in N0 than in N1 and N2 in both

dialects. If one disregards the effect of the preceding accent type, and simply looks at the effect of tail length, one finds that the leftward shift of T-rel in N0 amounts to approximately 20% of the accented syllable duration in all cases.

The timing of the  $f_0$  inflection in the nuclear accent is also influenced by the immediately preceding accent type, particularly in CF. As can be seen in Figure 4, the onset of the nuclear fall occurs earlier when the preceding accent is  $H^*$  than when it is  $H^*+L$ . The difference is quite striking (median T-rel values in N0-N2 range from 35% to 63% for  $H^*+L$   $H^*+L$  vs. from 16% to 32% for  $(H^*) H^*+L$ ). This observed difference seems likely to be related to the presence vs. absence of the intervening L valley. In GD, the preceding accent type has a comparatively minor influence on the timing of the  $f_0$  inflection of the nuclear rise.

The ANOVA test of T-rel confirms the significant effect of Tail Length ( $F(2, 18) = 9.05, p = .002$ ); T-rel in N0 differs significantly from N1 and N2 by 22% and 20%, respectively. The non-significant interaction ( $F(2, 18) = 0.26, n.s.$ ) confirms that CF and GD respond uniformly to tail length.

Looking at the two dialects, the onset of the nuclear fall in CF occurs relatively early in the syllable (frequently in the first half) compared to GD, where the onset of the nuclear rise occurs relatively late in the syllable (typically in the second half). This difference is confirmed by the significant ANOVA result for Dialect ( $F(1, 18) = 17.65, p < .001$ ). The mean T-rel difference between CF and GD amounts to 20% of the accented syllable duration.

Comparing the Fujisaki model T-rel and contour-based  $t^*$ , strong correlations between the two measurements were found ( $r = 0.89$  for CF and  $r = 0.87$  for GD).

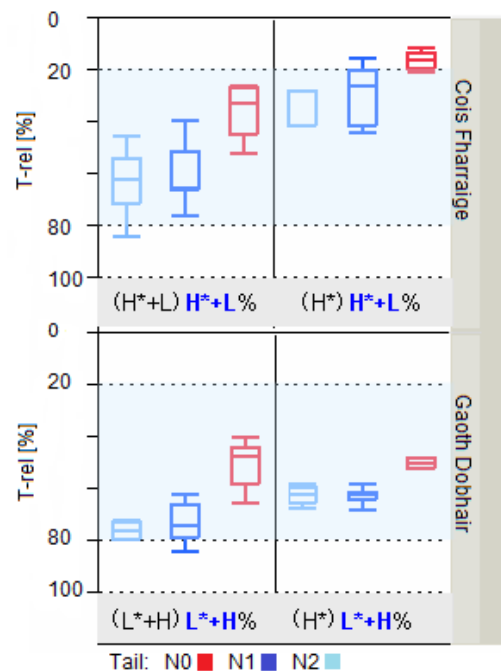


Figure 4: T-rel in the nuclear accent data expressed as a percentage of the accented syllable duration (0-100%). The blue panels represent the accented vowel (20-80%).

### 3.2. Scaling – f0 excursion size of fall and rise

Figure 5 presents the results for the scaling of the nuclear fall (CF) and rise (GD) captured with Aa. Note that this parameter is expressed on a logarithmic scale in the Fujisaki model. In the Irish nuclear data, the Aa-total values frequently fall somewhere between 0.2 and 0.4.

It is clear from Figure 5 that tail length has no appreciable effect on the f0 excursion size of the nuclear accent in either CF or GD. In the ANOVA test neither Tail Length ( $F(2, 12) = 2.50$ , n.s.) nor the interaction ( $F(2, 12) = 0.75$ , n.s.) are significant.

A considerable difference in Aa-total is observed between the two dialects. The f0 excursion size is notably greater for GD rises than for CF falls. This difference is significant in the ANOVA result for Dialect ( $F(1, 6) = 12.96$ ,  $p = .011$ ). The mean difference in Aa-total between the two dialects amounts to a rather substantial value of 0.2.

The f0 excursion size of the nuclear accent estimated with Aa-total was compared to the contour-derived counterpart. A very high positive correlation was found for GD ( $r = 0.92$ ). For CF the correlation is somewhat lower ( $r = 0.79$ ), implying some degree of underestimation of the f0 excursion size in the nuclear falls.

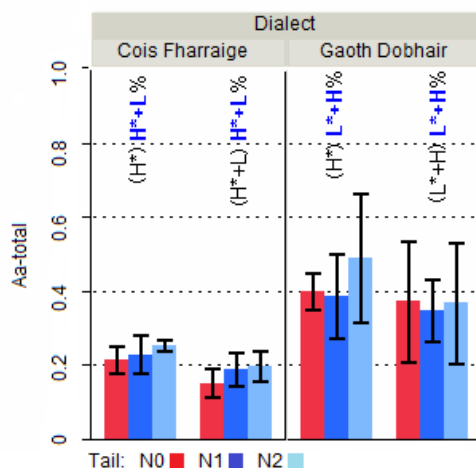


Figure 5: Aa-total in the nuclear accent data.

## 4. Conclusions

This paper described the timing and scaling of the nuclear accents of Connaught and Ulster Irish in varying tail length conditions. The Fujisaki model T and Aa parameters were used to flesh out the realisation of the IP-final H\*+L accents of Cois Fharráige Irish and L\*+H accents of Gaoth Dobhair Irish.

The results show that in both dialects the timing of f0 inflection (T-rel) of the nuclear accent is affected by tail length, while the f0 excursion size (Aa-total) is not. Also, some cross-dialect differences in the realisation of the IP-final accent were found (overall earlier timing of T-rel in CF, and higher Aa magnitudes in GD).

The performance of the two accent command parameters in the nuclear accent data was evaluated by comparing T-rel and Aa-total to parallel contour-derived measurements. The considerably high correlations between the two sets of measurements demonstrate that the accent component captures

the timing and scaling of the nuclear accents rather accurately. A nearly 90% agreement is found between T-rel and t(\*) in both dialects. The correspondence between Aa-total and f0 size is higher in GD, where the use of the negative-positive command pair allows greater precision of modelling the low portion of the nuclear accent than a single positive command does (as in the case of CF).

A good agreement between the accent command parameters and the contour-derived counterparts indicates that the model captures the essential characteristics of the nuclear accents in varying tail length conditions. If the current parameter settings were used in synthesis, they would produce very close approximations.

It is promising that the Fujisaki model is suitable for Irish. It generates appropriate contours that capture the fine details regarding the timing and scaling aspects in the two diverse dialects. With a view to these data being usable within TTS, perception experiments would be recommended to establish to what extent the possible mismatches of the Fujisaki model are (i) audible, (ii) likely to undermine the dialect-specific prosodic characteristics.

## 5. Acknowledgements

We thank our Cois Fharráige and Gaoth Dobhair informants, who gave so generously of their time for these investigations. We would also like to acknowledge partial funding for this work from Govt. of Ireland, Dept. of Arts, Heritage and the Gaeltacht (ABAIR project), and An Chomhairle um Oideachas Gaeltachta & Gaelscolaíochta (CabairE project).

## 6. References

- [1] H. Fujisaki and S. Nagashima, "A model for the synthesis of pitch contours of connected speech," *Annual Report of the Engineering Research Institute, University of Tokyo*, vol. 28, pp. 53-60, 1969.
- [2] H. Fujisaki and K. Hirose, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *Journal of the Acoustical Society of Japan (E)*, vol. 5, pp. 233-242, 1984.
- [3] H. Fujisaki, "A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour," in *Vocal Fold Physiology: Voice Production, Mechanisms and Functions*, O. Fujimura, Ed. New York: Raven Press, 1988, pp. 347-355.
- [4] M. Dalton and A. Ní Chasaide, "Modelling intonation in three Irish dialects," in *15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2003, pp. 1073-1076.
- [5] M. Dalton and A. Ní Chasaide, "Tonal alignment in Irish dialects," *Language and Speech: Special Issue on Intonation in Language Varieties*, vol. 48, pp. 441-464, 2006.
- [6] M. Dalton and A. Ní Chasaide, "Melodic alignment and micro-dialect variation in Connaught Irish," in *Tones and tunes: Studies in word and sentence prosody*. vol. 2, C. Gussenhoven and T. Riad, Eds. Berlin: Mouton de Gruyter, 2007, pp. 293-315.
- [7] M. Dalton, "The Phonetics and Phonology of the Intonation of Irish Dialects," Ph. D. dissertation, Trinity College Dublin, 2008.

- [8] M. O'Reilly and A. Ní Chasaide, "Realisation of sentence mode in the intonation of Connemara Irish," in *6th Celtic Linguistics Conference, Dublin, Ireland*, 2010.
- [9] A. Dorn, M. O'Reilly, and A. Ní Chasaide, "Prosodic signalling of sentence mode in two varieties of Irish (Gaelic)," in *17th International Congress of Phonetic Sciences*, Hong Kong, China, 2011, pp. 611-614.
- [10] "abair.ie – An Sintéiseoir Gaeilge," Available at [www.abair.tcd.ie](http://www.abair.tcd.ie).
- [11] A. Ní Chasaide, N. Ní Chiaráin, H. Berthelsen, C. Wendler, and A. Murphy, "Speech technology as documentation for endangered language preservation: The case of Irish," in *18th International Congress of Phonetic Sciences*, Glasgow, UK, 2015.
- [12] "Audacity Version 2.0.2 [Open source software for recording and editing sounds]," Available at <http://audacity.sourceforge.net/>.
- [13] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (Versions 4.0 and 5.2.11) [Computer program]," Retrieved from <http://www.praat.org/>, 2011.
- [14] E. Grabe, F. Nolan, and K. Farrar, "IViE - A comparative transcription system for intonational variation in English," in *5th International Conference on Spoken Language Processing*, Sydney, Australia, 1998, pp. 1259-1262.
- [15] E. Grabe, "The IViE labelling Guide," Available at <http://www.phon.ox.ac.uk/IViE/guide.html>, 2001.
- [16] H. Mixdorff, "FujiParaEditor: a Windows environment for Fujisaki model parameter extraction [Computer program]," Retrieved from <http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html>, 2006.
- [17] M. O'Reilly, A. Ní Chasaide, and C. Gobl, "Cross-dialect Irish prosody: linguistic constraints on Fujisaki modelling," in *Interspeech 2008*, Brisbane, Australia, 2008, pp. 836-839.
- [18] M. O'Reilly, "Sentence mode, alignment and focus in the intonation of Cois Fharraige, Inis Mór and Gaoth Dobhair Irish - a dual approach," Ph. D. dissertation, Trinity College Dublin, 2014.
- [19] B. Möbius, *Ein quantitatives Modell der deutschen Intonation: Analyse und Synthese von Grundfrequenzverläufen*. Tübingen: Niemeyer, 1993.