Cross-dialect Irish prosody: linguistic constraints on Fujisaki modelling

Maria O'Reilly, Ailbhe Ní Chasaide, Christer Gobl

Phonetics and Speech Laboratory,
School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland
moreil12@tcd.ie, anichsid@tcd.ie, cegobl@tcd.ie

Abstract

We describe here our approach to quantifying cross-dialect differences in Irish Gaelic, using the Fujisaki model. The basic principle is that the way in which the modelling is carried out respects a parallel linguistic (AM) analysis. The aims are: (1) to ensure that our modelling strategies permit a reliable cross-dialect comparison, (2) that the model-derived measurements can be related to meaningful linguistic dimensions and (3) that the analysis forms the basis for multi-dialect synthesis.
Index Terms: Irish Gaelic, Fujisaki model, speech synthesis.

1. Introduction

This paper discusses an approach currently being used to provide quantitative, cross-dialect comparisons of Irish intonation, using the Fujisaki model. The principal objective of this research is to tease out as many as possible of the intonational characteristics that define these dialects as being different from one another. This task involves primarily a linguistic-phonetic description, and it is therefore imperative from the outset to use the model in such a way as to ensure that the parameter values which are the output from the modelling can provide measures that we can correlate directly with linguistically important dimensions, such as timing and amplitude of pitch accents, declination rate, level and dynamic range of pitch variation.

The present work builds on the project *Prosody of Irish Dialects* [1]. From this project we already have an Autosegmental-Metrical analysis of the basic sentence types for the main dialects of Irish (Figure 1) [2–4] using the IViE labelling system [5]. The linguistic analyses have provided many insights, particularly concerning major structural similarities and differences across the dialects. However, the typical IViE (or ToBI) descriptions are in terms of the tone sequences that make up the intonational contour, and these are necessarily at a rather abstract level of description. As such, they do not capture the many finer-grained phonetic differences which may be important in differentiating the prosody of different dialects to the layperson's ear. Specific measurements were made of peak timing in nuclear and initial prenuclear accents for certain utterances, and these did reveal some striking cross-dialect differences [2, 3]. However, it was felt that there may be other kinds of cross-dialect differences which cannot easily be picked up without using a quantitative model for our comparisons. It is hoped that by employing the Fujisaki model we can come closer to capturing more of the finer-grained phonetic level of differentiation which is important, but often missed in traditional linguistic description, where the emphasis is on the phonological level.

In applying the Fujisaki model it is crucial for us that the modelling be carried out in such a way that the model parameters or parameter-derived measures can be correlated with linguistic dimensions. Much of the research with the Fujisaki model is geared towards speech technology, particularly towards speech synthesis, and the primary concern for many researchers is to be able to generate reasonable synthetic prosody for a particular language. With this in mind, the model has also been used to characterize the intonation of individual languages such as German [9, 10]. What is particular in the current approach is that our research goals require that we explicitly constrain how the model is used for cross-dialect comparisons. Effectively this means we are using the model *in parallel* with the AM-based linguistic description, and that the model parameters are intended to extend the phonetic coverage of the linguistic analysis.

Although our primary interest is that of a linguistic description, there is a further motivation for using the Fujisaki model for our analysis. A parallel strand of research in our research group involves the generation of multi-dialect synthesis of Irish Gaelic [6]. It is our intention that our description will provide the intonation models in such a way as to be directly useable for the synthesis of these dialects. We intend to generate a range of dialect-specific prosodic models and use the Fujisaki model to serve in their building. The dialects in question are indicated in Figure 1.

Figure 1: *Dialects of Irish Gaelic: Donegal (1), Mayo (2), Connemara (3), Aran (4) and Kerry (5).*

2. Fujisaki model of intonation

The well-known Fujisaki model [7, 8] has been applied to many languages over a number of years. It decomposes the fundamental frequency curve into a set of component curves, from which timing and frequency information on the contours can be extracted. The model components are related to the linguistic organisation of an utterance. The phrase component is the global component which models the overall f0 trend (declination of the baseline) in an intonational phrase (IP). The accent component is the local component which models the local f0 variations at the accentual level. Phrase and accent commands are superposed onto the base frequency (asymptote). Thus, an f0 contour is represented as the sum of the three terms on a natural logarithmic scale, and reproduces

phrase-level and accent-level f0 variations over time. The model parameters are shown in Figure 2.

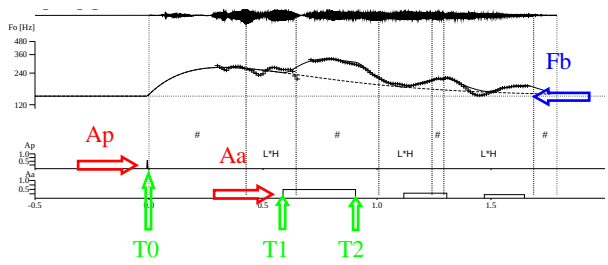

Figure 2: *Fujisaki model parameters: phrase command onset (T0) and amplitude (Ap), accent command onset and offset (T1 and T2) and amplitude (Aa) and base frequency (Fb).*

3. Setting linguistic constraints

As mentioned in the introduction, the Fujisaki modelling is carried out in parallel with the prior AM analysis, and is where appropriate guided and constrained by this linguistic description. There are some basic principles that were adopted from the outset. Firstly, we needed to incorporate a modelling strategy that could be used consistently across dialects so as to ensure comparability of measurements. Secondly, the results of the modelling were always hand-tailored to ensure that the linguistic analysis is not violated in any important way, particularly when mapping accent commands to the tonal events of the AM analysis.

Thirdly, as there are potentially too many degrees of freedom in how the Fujisaki phrase and accent commands can be used to match the contour of the original utterance, it was important to constrain the analyser's choices at the outset. In order to ensure comparability across the data, a single analytic strategy had to be maintained. For example, when hand-tuning the accent commands, there is considerable trade-off between their timing and amplitude. Furthermore, the settings used for one model component affect those of the other parameters (e.g., the base frequency versus phrase and accent commands). The risk is that the analyser would adopt slightly different strategies for different data, particularly if the immediate objective were simply to get the best exact match to the original contour. Any inconsistency on the part of the analyser in hand-tuning the model parameters would introduce a potential measuring error. Differences emerging in the data would thus be difficult to interpret with confidence. For that reason it was deemed necessary to fix on explicit matching strategies that could be maintained across the different dialects. Finally, we were from the outset concerned to ensure that our model-based measures could be correlated with intonational events.

The most important Fujisaki component for cross-dialect differentiation, certainly for broadly-focused short declaratives, is the accent command, as it captures important timing and amplitude characteristics of the accent and boundary tones. Setting the parameters for the accent commands requires that the base frequency and phrase command parameters be set first. Therefore, we will describe our modelling strategy in that order.

3.1. Base frequency and its interpretation

The base frequency of the Fujisaki model is the asymptote to which accent and phrase commands are added. It can be treated as a constant for a given dataset (speaker-dependent) [10]. Alternatively, it can be set to the minimum f0 value found for a particular contour (utterance-specific) [8, 9].

The approach chosen in our work is to allow for Fb to vary. In the context of Irish Gaelic, the base frequency coincides with the final lowering of f0 at the end of an IP where a phrase-final fall is observed. Where the contour has a phrase-final rise, Fb is placed slightly below the final accentual low. In this approach the base frequency in principle captures the bottom of the f0 range found in a given contour.

Our reasoning behind the variable Fb approach is twofold. Firstly, we want to obtain precise measurements for phrase and accent commands as they occur in the original f0 contours. By doing so, the obtained phrase and accent command measurements will relate more directly to declination and tonal events. Secondly, we believe that the base frequency is a parameter that truly varies in certain linguistic and paralinguistic contexts such as sentence type, paragraph structure, or presence/absence of focus. Whether we will find cross-dialect, or indeed within-dialect differentiation for those categories remains an open question.

3.2. Phrase command and declination

The phrase command is understood as the phrase-level component whose amplitude can be viewed as a measure related to declination. On the face of it, the rate of decay for the phrase command increases with the growing value of Ap. Eventually, as the length of the utterance can influence the rate of declination it needs to be taken into account in any calculation. Therefore, our eventual comparative measure for declination for within and cross-dialect data will be based on a combination of these factors.

As the objective is to quantify the intonational events in parallel with the phonological description available, the number of phrase commands must correspond to the actual number of IPs. For instance, all one-IP sentences are modelled with one phrase command only (Figure 3a-e). It needs to be pointed out that we do not deal with cases where the number of IPs is ambiguous in the auditory analysis. This way the global component of the Fujisaki model will depict the true characteristics of f0 contours at the level of phrasing.

The Alpha parameter: following the interpretation suggested in the literature [7–10], alpha is kept constant to ensure a uniform timing and amplitude response for the phrase command. As the analysis showed that the dialects of Irish Gaelic exhibit a considerable degree of declination, our approach uses alpha = 3.0 as in [7–9]. The results for this value are satisfactory and changing the default value is typically not required. The advantage of using 3.0 for alpha is that in the contours where more unstressed material follows the final accent the phrase command models the final fall more accurately than if a lower alpha value were used. On the whole, keeping alpha constant at 3.0 yields satisfactory contour approximations and reduces the number of variable parameters in the model.

3.2.1. Phrase command timing and amplitude

Once the phrase control mechanism, alpha, is set to a constant, the timing and amplitude of the phrase command have to be found. Our strategy adopts that of [9], where the phrase command is fixed at such a time-point that the maximum amplitude of the phrase command coincides with the utterance onset (Figure 3a-e). In this setting the phrase command onset (T0) for alpha = 3.0 is thus located approximately at 330 ms before the voiced onset of the

utterance [9]. This approach has a strong advantage: it ensures uniform decomposition of f0 contours into phrase and accent commands.

The main point of reference for setting Ap is the contour baseline, i.e. the falling part of the phrase command goes through the f0 minima. Our second consideration for setting Ap concerns what prosodic event occurs at the beginning of the utterance, as we fix the time-point of the maximum amplitude of the phrase command Ap to the utterance onset. Where unaccented syllables are in this position, and are labelled as a %0 boundary, the f0 value found at this time-point coincides with the phrase command amplitude peak (Figure 3c and 3e). Where a high boundary is present (%H), Ap is adjusted to the f0 contour by taking into account the accent command that the boundary receives (Figure 3a). The same is done for IP-initial pitch accents which also by default are assigned an accent command.

3.3. Accent commands and tonal events

Accent commands can be basically interpreted as the local commands that capture the amplitude and timing characteristics of the f0 excursions (H and L) associated with the accented syllables of the utterance. They are also used for specific pitch excursions associated with the onset or offset of an IP, i.e. the phrase-initial and phrase-final boundaries.

It is most important that the number of accent commands should correspond to the number of pitch accents and boundary tones identified in an IP in the prior linguistic analysis. When the modelling is carried out automatically, there are often additional accent commands associated with syllables that are clearly not accented. Furthermore, when manually fine-tuning the model, one can often get a more exact match to the original contour by adding in unwarranted accent commands. Clearly, this would skew our results and the measurements would not be relatable to the linguistic events present in the contour, and cross-dialect comparison would be difficult..

By harmonising the Fujisaki analysis with the phonological structure of the melodic contours we may have to compromise on some of the detail of the original contour. This loss of information is unavoidable at this stage, but it is envisaged that we keep track of the difference between the modelled and original contours, so that we can afterwards identify where there are systematic differences which can be further investigated.

Notwithstanding the fact that we constrain the analysis so as not to violate the prior linguistic analysis, there are some occasions where the Fujisaki modelling leads us to reconsider and revise the linguistic interpretation. A prime example concerns the frequent occurrence of an initial high boundary tone in the declaratives of Donegal Irish (see example in Figure 3a). While modelling these utterances we became aware of the fact that these initial high boundaries are more frequent than the prior phonological analysis had implied. The tendency in this dialect to use initial high boundary tones is not very surprising in one way, as the initial accent is a low rising tone (L*H) and the high boundary helps to enhance the low tonal target. The point we would make here is that while the prior linguistic analysis guides the Fujisaki modelling, there is also a two-way process at work, whereby the Fujisaki modelling sometimes leads to a linguistic re-analysis.

The Beta parameter: the rate of rise/decay of an accent command is controlled by the beta parameter, while the ceiling parameter, gamma, determines the cut-off of the accent command amplitude. In accordance with the values suggested in the majority of the works employing the Fujisaki model we use the default values of 20.0 for beta, and 0.9 for gamma [7–10]. In Fujisaki's work beta was kept constant in any given utterance, and generally the differences in beta from one utterance to another are minute [7, 8]. Based on the work carried out so far, we can assume that generally constant beta does not impair the modelling of the accents. This is especially true for high-falling accents.

3.3.1. Accent command polarity

In most European languages only positive accent commands are used. An occasional negative accent command is used for modelling grammatical or paralinguistic uses of low tones in non-tonal languages such as English and Hindi [8]. Negative accent commands are normally used for tone languages (Chinese) and languages with contrastive accent types (Accents 1 and 2 in Swedish) [8].

For the analysis of all but the Donegal dialects only positive accent commands have been required. The situation is more complex in the case of Donegal Irish. Prosodically, this dialect is unique among the Irish dialects, as it uses predominantly rising tones (L*H), while the other varieties mainly employ falling (H*L) tones or high (H*) tones [3]. Typical declarative contours are shown for all the dialects in Figure 3.

One could argue that some rises in the Donegal dialect (with the exception of the phrase-initial rise) could be approximated better with a small negative command followed by a large positive command. There are disadvantages to such an approach. Firstly, there are too many degrees of freedom in how the combined negative and positive commands can be set to match the contour, and this does not inspire confidence that the ensuing measures can be interpreted in a consistent or meaningful way. Secondly, the combined positive/negative accents could not be compared as readily with the positive-only commands used for other dialects. As it transpires that for the data analysed to date, it has been possible to use a positive-only accent command to capture the Donegal contours, we have opted for this solution.

3.3.2. Accent command amplitude and timing

In our analysis, accent command amplitude (Aa) is understood as a measure of the accent height, or prominence. Therefore it is set in accordance with the highest f0 value occurring in an accent. Generally, the aim is to reproduce the height of the original accent with Aa as far as it is possible without negatively affecting the accent command timing.

The accent command onset (T1) is associated with the beginning of the accent; the accent command offset (T2) is associated with the point where f0 starts falling off. In this interpretation T1 is viewed as the time-point of the beginning of a rise pertaining to an accent, and T2 is viewed as the time-point of the beginning of the f0 fall. In sharply pronounced falls (H*L) T2 can be interpreted as a measure of peak timing (e.g., Figure 3d).

As mentioned above, it has proved possible to model the L*H accents of Donegal with only positive accent commands. Consequently, the major difference between the Donegal and the other, more southern dialects where H* or H*L accents dominate, emerges as a difference in the timing of T1. This can be observed in Figure 3: note that the accented syllables are shown with their phonological labels, and that the onset and offset of the accented syllable is shown by the dotted vertical lines to either side of the phonological label. In Donegal Irish, T1 occurs very late relative to the onset of the

accented syllable, and T2 is also very late. T1 typically occurs about half-way into the accented syllable.

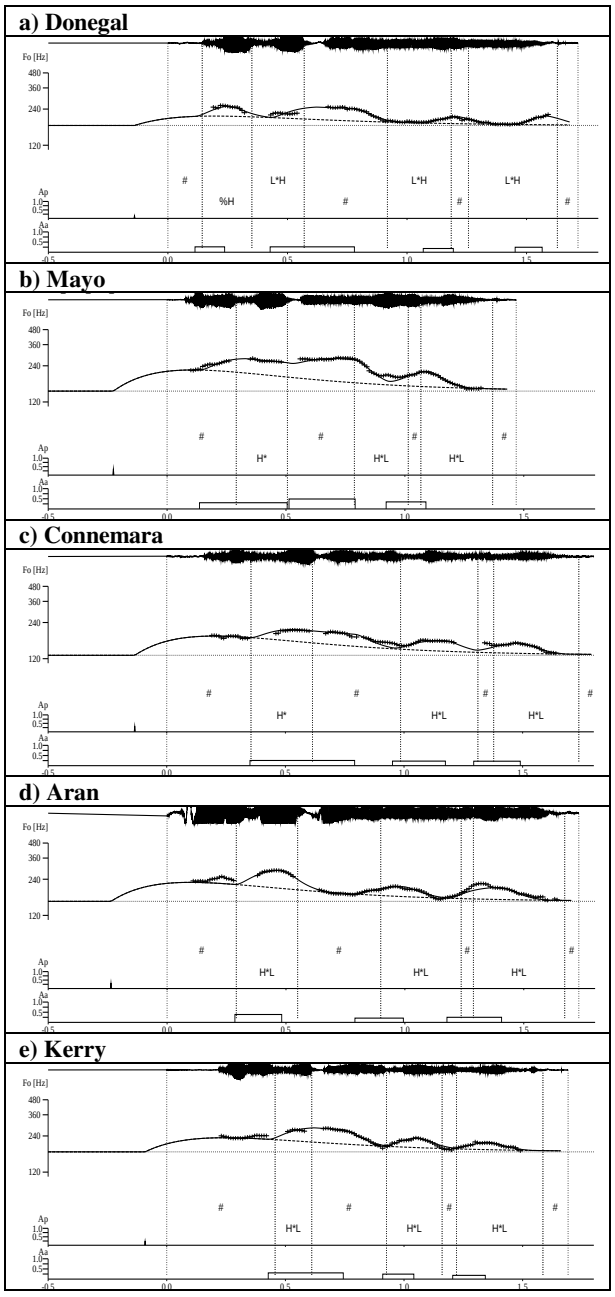

Figure 3: *Parameter settings for the sentence "Tá na héadaigh ina luí i moll" ('The clothes are lying in a heap) for five dialects of Irish.*

All the other dialects have in common a much earlier onset of T1. Even within the southern dialects there are some striking differences. For Connemara, Aran and Kerry, T1 tends to be located at or before the onset of the accented syllable, and there is a further tendency for T1 of the initial accent to be phased slightly later than in the following accents. The Mayo dialect, however, seems to have a much earlier T1, with an anticipatory rise towards the high tonal target (early T1 relative to the beginning of the accented syllable), and a frequently extremely early T2 (near the beginning of the accented syllable in the two falling accents of Figure 3b).

4. Conclusions

Clearly timing measures are going to be important for cross-dialect differentiation, and this is something that the past studies on peak timing in certain utterance types have also shown [2, 3]. We also expect that the quantitative approach adopted will show up other aspects of dialect differentiation, by providing a rich coverage of many features whereby dialects may differ. And while we come with certain expectations about where interesting cross-dialect differences might arise, we are nonetheless believe that the present approach will help us focus where we might not have thought to look. This hopefully will help overcome the limitations of an analysis that is too tied to the phonological, structural level, and explain how people readily hear prosodic differences that the linguist's analysis does not capture.

The linguistically-constrained approach we are adopting entails that it is not always possible to capture every detail of specific melodic contours. The discrepancies between the original and matched contours will need to be investigated to ensure that perceptually important detail is not lost.

All in all, this approach should not only tell us more about the differences among the Irish dialects, but it also has the potential to enable the investigation of cross-language differences: the comparison of Irish and varieties Irish-English is of considerable current interest to us. Eventually, we also hope that this work will shed more light on the nature of prosody.

5. Acknowledgements

This research was supported by the projects *Prosody of Irish Dialects* funded by the IRCHSS and Cabóigín funded by Foras na Gaeilge. We are also grateful to Hansjörg Mixdorff for his assistance.

6. References

[1] <http://www.tcd.ie/CLCS/phonetics/projects/prosody.html>

[2] Ní Chasaide, A. and Dalton, M., "Dialect Alignment Signatures", *Proc. 3rd Speech Prosody*, Dresden, 2006.

[3] Dalton, M. and Ní Chasaide, A., "Melodic alignment and micro-dialect variation in Connaught Irish", in C. Gussenhoven, T. Riad [Eds.], *Tones and tunes: Studies in word and sentence prosody, Vol. 2*, Berlin: Mouton de Gruyter, 293–315, 2007.

[4] Dalton, M. and Ní Chasaide, A., "Nuclear accents in four Irish (Gaelic) dialects", *Proc. XVIth ICPhS*, Saarbrücken, 965–968, 2007.

[5] <http://www.phon.ox.ac.uk/IViE/guide.html>

[6] <http://www.tcd.ie/slscs/clcs/phonetics/caboigin.php>

[7] Fujisaki, H., "Modeling the process of fundamental frequency contour generation", in Y. Tokhura, E. Vatikiotis-Bateson, Y. Sagisaka [Eds.], *Speech perception, production and linguistic structure*, Amsterdam: IOS Press, 313–326, 1992.

[8] Fujisaki, H., Ohno, S., Wang, C., "A command-response model for F0 contour generation in multilingual speech synthesis", *Proc. 3rd ESCA/COCOSDA International Workshop on Speech Synthesis*, 299–304, 1998.

[9] Möbius, B., "Ein quantitatives Modell der deutschen Intonation", Tübingen: Max Niemeyer Verlag, 1993.

[10] Mixdorff, H., "Modelling patterns of German – model-based quantitative analysis and synthesis of F0 contours", unpublished PhD thesis, Technische Universität Dresden, 1997.