Providing linguists with better tools: Daniel Hirst's contribution to prosodic annotation.

Céline De Looze[1], Na Zhi[2], Cyril Auran[3], Hyong-Sil Cho[4], Sophie Herment[5], Irina Nesterenko[6]

[1] Speech Communication Lab, Trinity College Dublin, Ireland.
[2]Laboratorio di Linguistica, Scuola Normale Superiore di Pisa, Italy.
[3] Laboratoire Savoirs, Textes, Langage, CNRS & Universités Lille 3 & Lille 1, France.
[4] Microsoft Language Development Center, Portugal.
[5] Laboratoire Parole et Langage, Aix-Marseille Université, France.
[6]Institut National des Langues et Civilisations Orientales (INALCO), France

deloozec@tcd.ie, zhina83@gmail.com, cyril.auran@univ-lille3.fr, t-hych@microsoft.com, sophie.herment@univ-amu.fr, irina.nesterenko@inalco.fr

Abstract

Among Daniel Hirst's contributions to speech prosody, this article addresses those concerned with the development of automatic tools for prosodic annotation. One of Daniel's concerns has been to facilitate the task of phoneticians and linguists involved in developing efficient rules for the analysis and synthesis of speech prosody. This presentation describes some of his innovative and fruitful contributions for the analysis of the prosody of natural languages.

Index Terms: prosodic annotation, tools, form~function, analysis by synthesis, MOMEL-INTSINT, IF, ProZed.

1. Introduction

Daniel Hirst has always insisted that the criteria of pure science should be applied to linguistics. He often likes to quote the French physicist and Nobel laureate Jean-Baptiste Perrin (1870-1942) who said that science consists in explaining "*visible complexity by invisible simplicity*". Scientists are confronted with huge quantities of data. If they want to bring knowledge to the data, they have to reduce its complex description by some simpler principle.

Beyond Daniel's seminal contributions to speech prosody, one of his important legacies can be found in his scientific method. He once noted, "*Science is supposed to be cumulative, explicit, predictive and empirically testable*" [23]. These are the properties that he suggests should apply to linguistic knowledge and which are clearly embodied in his investigations of the prosody of languages. For him, it is the only way for linguists "*to accumulate knowledge rather than to speculate on the nature and relative elegance of abstract models*".

It is clear from forty years of research that Daniel Hirst's work has been driven by these principles and has been of a great benefit to the community of speech researchers, in particular by providing linguists tools for the annotation of speech prosody according to the scientific criteria he has always found essential for the description of natural languages. It is this clarity of method that was necessary for bridging the gap between engineers and linguists.

PhD students, both those under his immediate supervision and those to whom he gave advice, have always found it a great pleasure working with Daniel. Under his tutelage they gained valuable insight, knowledge and experience, made all the more enjoyable by his great sense of humour and benevolence.

This presentation is from six of his former students and collaborators, all of whom worked with him during the preparation of their PhDs. Several more of his former students, who are not able to take part in this Speech Prosody conference, are cited in the references to this paper.

Among his contributions to speech prosody, we have selected, for discussion in this paper, those related to the development of automatic tools for prosodic annotation.

2. Theoretical framework

2.1 The form ~ function interface

A central issue in the area of speech prosody concerns the analysis and modeling of the ways prosody contributes to meaning. However, after many years of research, the mapping between prosodic forms and prosodic functions is still a poorly understood process. One of the reasons for this is that there is no general consensus for either of these two levels of representation. In many cases, the representation of prosody conflates both form and function. Hirst [21], on the contrary, has always argued for a systematic distinction between these two levels of representation in order to avoid circularity in their respective description and mapping. The argument is supported by the assumption that in all languages, a limited number of prosodic functions is expressed by a limited number of prosodic forms but with language-specific associations. In that context, he developed two distinct coding schemes for the representation of prosodic form and function: respectively, the INTSINT coding scheme and the IF-annotation system (3.1.1 and 3.1.2 below), which have already been used for the description of several languages.

Regarding the annotation of prosodic forms and functions, Hirst [21] further suggests that "*manual transcription should be reserved for those aspects of prosody that refer directly to the listener's interpretation of the utterance*", in other words, to its functions. For him, the annotation of a syllable in terms of pitch height, duration and intensity can be left to automatic systems, whereas manual annotation should be reserved for what the utterance means rather than what it sounds like. He argues that a functional annotation is much more relevant as "*the transcriber is required to perform a task of linguistic interpretation rather than a meta-linguistic task of phonetic analysis*". And, as he notes, "*it is well known in psycholinguistic studies [42], meta-linguistic tasks performed by untrained subjects entail considerable problems of interpretation*".

2.2 Levels of representation

Since he postulates a multi-level organization of the form ~ function interface, Daniel Hirst argues at the same time that the formal representation of prosody should be directly related to the acoustics of the speech signal and therefore should not make reference to functional categories. In [26], he proposes that the representation of prosodic form should include four different levels: the *physical* level, the *phonetic* level, the *surface phonological* level and the *underlying phonological* level. Here, the physical level corresponds to the level of acoustic data. The phonetic representation consists in quantitative values derived from the acoustic signal and is taken as the interface between abstract cognitive representations and their physical manifestations. The surface phonological level codes the prosodic form as a sequence of discrete symbols. Finally, the underlying phonological representation of prosodic form is conceived of as the interface between the representation of phonological form and syntactic/semantic interpretation. Each level of representation must comply with an interpretability constraint, which stipulates that each level has to be interpretable in terms of adjacent levels of representation. Such a model of an intonation system is conceived of as an attempt to provide mechanisms which allow us to derive representations at each level from representations at other levels.

2.3 The Analysis-by-synthesis paradigm

In [26], it is noted, "*There have been a number of different implementations of phonological/phonetic models of intonation designed to derive an acoustic output (F0 curve) from a symbolic input (…). As in all fields of speech analysis, however, it is the inverse problem which is really the most challenging. Given a F0 curve, how can we recover a symbolic representation? Even if we are able to perform such symbolic coding automatically, how should we validate the output of such a program?*".

One of Daniel Hirst's contributions to this issue was the idea of an ***analysis-by-synthesis paradigm***. The idea is to use the symbolic representation derived from the raw acoustic data as input to a synthesis system. The acoustic output of such a system can then be compared to the original f0 curve in order to evaluate the efficiency of the model (see *Figure 1*). This goes in hand with his view that the criteria *cumulative, explicit, predictive* and *empirically testable* should be applied to linguistic knowledge - a model that is predictive and empirically testable must also be reversible.

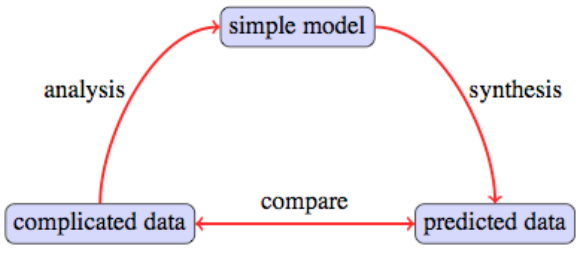

Figure 1: *The analysis by synthesis paradigm [23].*

3. Annotation tools

Daniel Hirst's innovative contributions to prosodic annotations are clearly embodied in the tools he has developed for the automatic analysis and coding of speech melody (MOMEL-INTSINT), speech rhythm (PROZED rhythm) and of prosodic functions (IF). Their particular interest lies in their "theory-neutral" nature, which enabled the comparison of several languages - British English, Spanish, European Portuguese, Brazilian Portuguese, French, Romanian, Russian, Moroccan Arabic and Japanese ([25]). These tools, with their "user-friendly" interface, have without doubt facilitated the task of phoneticians and linguists in developing efficient rules for the analysis and synthesis of speech prosody for a wide variety of languages and dialects. It has to be highlighted that all of them are freely distributed and easily available, which is, in our view, the most important contribution for accumulating linguistic knowledge. In the remaining part of this paper, we give an overview of those tools together with a presentation of some of our own applications of these tools.

3.1 MOMEL and INTSINT algorithms

3.1.1 MOMEL

The MOMEL (MOdelling MELody) algorithm was developed by Daniel Hirst and his colleague Robert Espesser, to provide a phonetic representation of intonation patterns [27]. Here, the fundamental frequency curve is assumed to be the product of two independent components: a global *macroprosodic* component, corresponding approximately to the underlying intonation pattern of the utterance, and a local *microprosodic* component, representing the deviations from the macroprosodic curve which are caused by the segmental content of an utterance. The discontinuity observed in the raw fundamental frequency curve is modelled by the microprosodic component, while the underlying macroprosodic component is modelled as a continuous and smooth curve, using a quadratic spline function (see *Figure 2*). The algorithm thus takes as input a raw f0 curve and gives as output a corresponding sequence of target points for the quadratic interpolation [27].

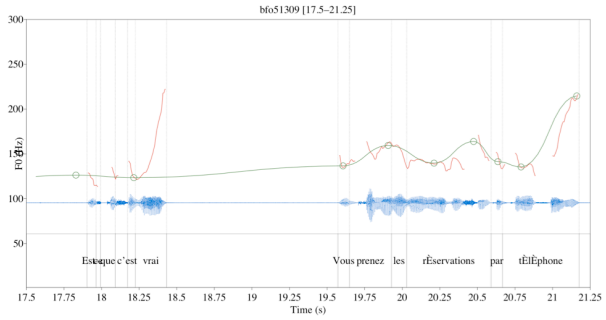

Figure 2: *Automatic output given by the MOMEL algorithm for the utterance "Est-ce que c'est vrai? Vous prenez les reservations par telephone" [22].*

The 'theory-neutral' nature of the algorithm has made it compatible with a number of different theoretical approaches to the description of speech melody and it has been used as a first step with the Fujisaki model [33], ToBI for English [32][45], K-ToBI for Korean [8][9], and INTSINT (see below).

3.1.2 INTSINT

The **INTSINT** system (**In**ternational **T**ranscription **S**ystem for **INT**onation) was developed to provide a surface phonological representation of intonation patterns. It was designed as a potential interface between a purely phonological symbolic representation of intonation and a continuous phonetic representation. The system, based on published descriptions of intonation patterns [19], basically describes an intonation

contour as a sequence of tonal segments, which are labeled using an alphabet of 8 symbols. The tonal segments are assumed to be of three types: (1) **Absolute tones** (*Top, Mid, Bottom*), assumed to refer to the corresponding position of the speaker's current pitch range (defined by the two parameters *key* and *span*); (2) **Relative tones** (*Higher, Same, Lower*), assumed to be defined with respect to the preceding tonal segments; and (3) **Iterative relative tones** (*Upstepped, Downstepped*), also defined relative to the preceding tonal segment but generally involving smaller pitch changes and often occuring in a sequence of steps either upwards and downwards. The relative interpretation of each tone can be illustrated as in Figure 3.

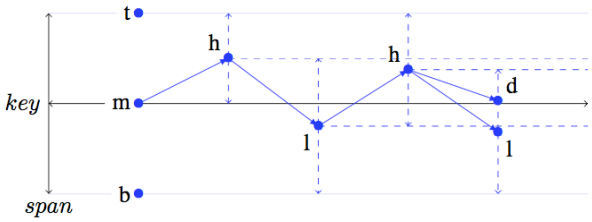

Figure 3*: Graphic illustration of the mapping from INTSINT to MOMEL defined by the two parameters* key *and* span *[22].*

At the surface phonological level, INTSINT is entirely concerned with the representation of prosodic form. The marking of the prosodic structure (*e.g.* boundary, prominence), speaker identification, turn-taking regulation as well as expression of emotions and affects is reserved to another scheme (the IF system, described in 3.3). In this sense, INTSINT can be conceived of as a prosodic equivalent of a narrow IPA transcription system for segmental transcriptions.

3.1.3 Mapping MOMEL to INTSINT

The MOMEL and INTSINT algorithms were both developed to facilitate a mapping between the output of the MOMEL algorithm (pitch targets) and the surface phonological structure provided by INTSINT (tonal symbols). Following the idea of the analysis-by-synthesis paradigm, the mapping from the phonetic to the surface phonological levels can be carried out in the other direction, that of synthesis.

The MOMEL and INTSINT algorithms have been used for the analysis of several languages including (English [2][19], French [39][3][41], Italian [16], Catalan [14], Brazilian Portuguese [15], Venezuelan Spanish [35], Russian [37], Arabic [36] and IsiZulu [30], Korean [29] [8] and Chinese [48,49]) as well as the intonation of L2 speakers (French learners of English [43], and English learners of French, with comparisons made possible between native speakers and learners [17]). A tool which combines the MOMEL-INTSINT algorithms is now freely available as a Praat plugin [22].

3.2 ProZed, a Multilingual Prosody Editor for speech synthesis

The MOMEL-INTSINT algorithms have also been implemented in ProZed, a Multilingual Prosody Editor for speech synthesis, which interfaces with the *MBROLA* [13] *SPPAS* [4] and *Praat* [5] programs. This tool, still under development [24], has been designed for testing models of the prosodic organisation of speech. The prosody of utterances can be manipulated by directly controlling the symbolic representation of prosodic form, which provides an immediate interactive assessment of the prosody as determined by the model; the resulting synthesized stimuli can be used to provide linguists with acoustic evidence for evaluating the variants of coding and alignment derived from data analyses.

Specifically, the rhythmic and tonal aspects of speech can be modelled and manipulated by defining three specific tiers in addition to the phoneme tier: the rhythm unit (RU) tier, the tonal unit (TU) tier and the intonation unit (IU) tier. No assumptions are made about which specific phonological entities should correspond to these units. RU and TU are consequently *defined* as the domain of shorter-term variability in prosody, and IU as the domain of longer-term variability in both duration and pitch.

The prosodic manipulation is determined by defining segmental duration on the RU, pitch on the TU, and long-term parameters on the IU. The pitch contour of speech is represented symbolically by the INTSINT alphabet as a sequence of target points, and the timing precision of target points aligned within the TU is determined via the use of "dummy" targets. The shorter-term definition of pitch targets is further modified by a longer-term parameter on the IU tier, which corresponds to the speaker/utterance's pitch range, defined by *key* and *span*. Similarly, the values determined on the RU level are modified at the IU level according to the speaker/utterance's speech rate.

The prosodic study of standard Chinese under the analysis-by-synthesis paradigm is in progress [49]. By annotating the inventory of prosodic functions, specified by lexical tones and intonation, a high-quality re-synthesis has already been obtained of the melodic features in Chinese spontaneous and read speech.

Such a multi-layered system is, in our view, necessary for the description of the melodic and rhythmic patterns of a language as it allows for a clear distinction between shorter-term and longer-term prosodic variations. A major drawback of most scalar systems for representing intonation patterns is the difficulty in separating global pitch changes (determined by variations in key and span) from local pitch characteristics (determined by changes in the phonological representation of intonation). How to distinguish, for example, a high fall in a narrow pitch range from a low fall in a wide pitch range? To answer this, changes in the f0 domain are accounted for by admitting two tone levels, as assumed in the AM theory [40], or more, as in INTSINT [21]. However, while these models appear adequate for the analysis of short read sentences (as often employed in laboratory speech), the fact that they implicitly assume that a speaker's key and span remain unchanged makes their use fragile for the analysis of spontaneous speech in which variations in pitch range are numerous and may convey different prosodic functions. Similarly, the overlap between short-term features such as segmental duration and longer-term ones such as tempo makes the analysis and modelling of the temporal organisation of speech difficult. How to distinguish for example a short phoneme in a slow tempo from a long phoneme in a fast tempo?

For the study of intonation patterns, Hirst [23] suggested that INTSINT coding can be made on smaller segments of speech, such as interpausal groups, with the assumption that changes in key and span tend to occur between interpausal groups rather than within them. The detection and measurement of pitch range variations can be made upstream with other tools such as those developed by De Looze [12]. In [12], a clustering algorithm, ADoReVA, is used for the automatic detection of pitch range variations. The algorithm calculates the difference between two consecutive units (e.g. interpausal groups) according to their pitch range. It generates a binary tree structure in the form of a layered icicle diagram from which pitch range variations can be graphically represented and automatically estimated (see *Figure 4*).

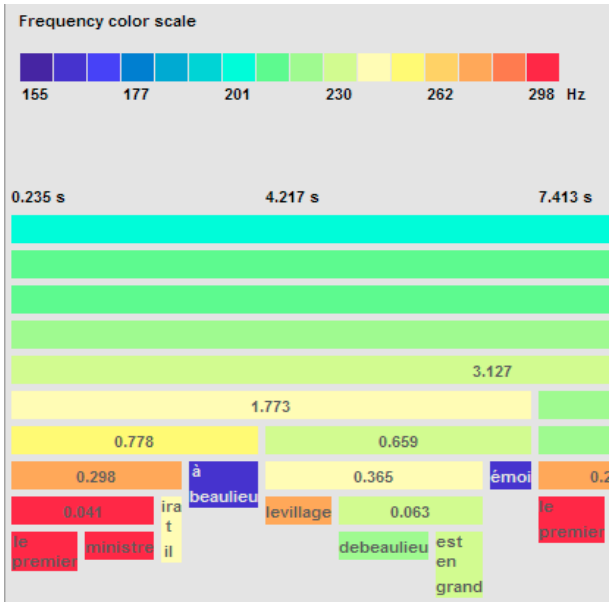

Figure 4*: Extract of a layered icicle diagram representation as obtained with the algorithm ADoReVA. Speech units are grouped together according to their key. The distance between the leaf nodes indicates differences in key between units. The bigger the break in the tree structure, the greater the difference. The colour scale indicates each unit's key. The warmer the colour, the higher the key.*

In [10], it was shown that taking into account those variations in the intonation patterns of French and English spontaneous speech clearly improves the automatic coding of intonation as provided by INTSINT. This method could be applied to other existing intonation systems to improve their efficiency. Similarly, the analysis of the temporal organisation of speech could make a clear distinction between short-term and longer-term variations, the latter being detected upstream, using for instance ADoTeVA [12], a tool developed for the automatic detection of speech-rate variations.

3.3 The OMe-scale: a natural scale for speech melody

In the context of providing tools for the automatic analysis of speech prosody, De Looze and Hirst [11] recently argued that a more natural scale for the analysis of speech melody is the OMe (Octave-Median) scale, using the octave (o) as the basic unit, centred on the median of a speaker's pitch range.

Fundamental frequency, the primary acoustic correlate of speech melody, is generally analysed and displayed using a linear scale (Hertz) or a logarithmic one (usually semitones), offset to an arbitrary reference level. Other scales have also been proposed, including the Mel, Bark and ERB scales. The relevance of these scales, however, remains debatable.

In studies of prosody, the physical scale in Hertz is very often transformed to the semitone scale, with a reference value (called C0) arbitrarily set at 16.3516 Hz [47]. The semitone is, however, in no sense, a natural unit of measurement. It is, in fact, the product of a complex history of Western classical music culture, corresponding to the division of the octave into 12 equal intervals. Recent studies [6, 7, 30, 46] however have shown that it is the octave, not the semi-tone, which appears clearly as the basic unit for the natural perception of the pitch of speech sounds and music. The use of semi-tones has paradoxically had the negative effect of masking the importance of the octave as the fundamental unit in speech perception and production. Re-reading a number of studies on pitch range with this in mind reveals a very large number of cases where authors report an interval close to an octave (=12 sts) or half-octave (6sts) without drawing attention to this fact.

De Looze and Hirst [11] report results based on 4 corpora (85 speakers), in English and in French, which showed that, in the production of natural speech, the lower range of fundamental frequency corresponds to half an octave below the median pitch of a speaker's voice, and the upper range to up to an octave above the median (see *Figure 5*).

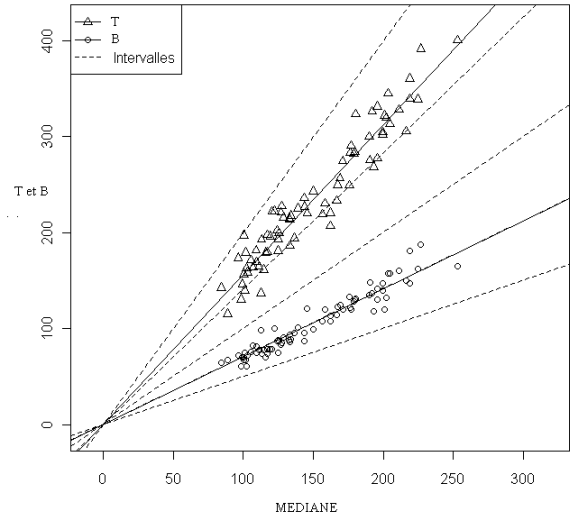

Figure 5*: Graphical representation of Mean-B and Mean-T (as obtained for the 85 speakers with the INTSINT algorithm) with respect to the median. Linear regressions corresponding to Mean-B and Mean-T are traced in continuous lines and the dotted lines represent, from top to bottom, the intervals +octave, +half-octave, unison, -half-octave and –octave with respect to the median. The regression line for Mean-B is indistiguishable from the -0.5 octave line.*

This suggests that the average of high tones and the average of the low tones, i.e. the limits of the range of a speaker, at least for unemphatic speech, usually correspond to about an octave centered on the speaker's median. These musical intervals, defined relative to the median, could therefore be used to estimate the range of a speaker and to offer a natural scale for the analysis and visualization of the melody of speech. This scale is obtained by the formula:

$$\log_2(\text{Hz}/\text{median}) \qquad (1)$$

Figure 6 illustrates a reading of the sentence "Last week my friend had to go to the doctor's to have some injections" by one male speaker. The diameter of the yellow circles corresponds to the duration of each syllable and the dashed blue line corresponds to the Momel curve. The horizontal dotted lines correspond to the speaker's median (middle line) and a half octave above and below the median, delimiting the speaker's unemphatic pitch range corresponding to the median-centred octave. The values of the Median and the Top and Bottom of the central octave are given both in Hz and as musical notes with respect to concert pitch at 440 Hz. The visualization of this recording was obtained automatically from the F0 curve using the Praat plugin ProZed [24]. With this technique, the optimal parameters for the analysis of the fundamental frequency of the speaker are automatically determined from the median pitch.

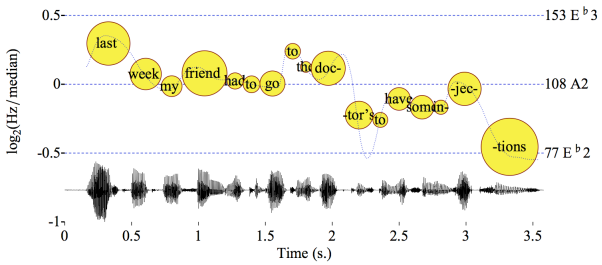

Figure 6*: Graphical representation of a reading of the sentence "Last week my friend had to go to the doctor's to have some injections." displayed using the Octave-Median scale. The diameter of the circles corresponds to the duration of the syllables. The horizontal blue lines delimit the central octave surrounding the speaker's median pitch.*

3.4 IF-system

3.4.1 IF annotation

While there exist a number of different solutions for the automatic analysis of prosodic forms, the automatic analysis of prosodic functions is a far more formidable task. Daniel Hirst proposed a functional annotation system called *IF annotation* - from *Intonative Features,* the title of the published version of his PhD thesis [18], (alternatively glossed as *Intonation Functions*). The system accounts for those aspects of prosodic representation that contribute directly or indirectly to the syntactic interpretation of an utterance. It consists in annotating the prosodic boundaries of an utterance, using two levels (terminal and non-terminal) as well as its relative prominence, using four levels (unaccented, accented, nuclear and emphatic) [21]. Other annotations of prosodic expression including discourse structure, speech act and expression of affect can then be added to augment the linguistic role of prosody encoded with IF.

3.4.2 IF-INTSINT articulation

The preliminary IF-functional system, together with the MOMEL-INTSINT algorithms, was particularly developed in an analysis-by-synthesis perspective to provide a means of establishing an enriched functional annotation system. In this approach, the IF annotation scheme can be used to generate an INTSINT representation, which in turn can be compared to the annotation derived from the observed data. The revealed systematic differences can then be used to either correct the mapping rules between the prosodic form and function representations or to extend the inventory functions currently accounted for by the IF annotation system.

Such a method has already proved its efficiency in the mapping of the formal and functional levels in Finnish [44], Russian [38], English [1], Korean [8] and Chinese [49]. In [37] and in [8] for instance, it has been proposed that the regularities observed in the INTSINT tonal patterns can be captured in relation to IF functions by combining methods of pattern-extraction and prediction using probabilistic grammars, thus adding a probabilistic dimension to the models and allowing for variability of resynthesised prosody.

In [9], the authors investigated phonetic pitch movements in minor prosodic units (accentual phrases, APs) in a Korean read corpus and in [38], minor prosodic constituents (prosodic words) and pitch contours associated with them as a function of their prosodic prominence were studied in a corpus of Russian spontaneous speech. In both studies, it was shown that, to a certain degree, prosodic constituents can be identified via phonetic pitch movements associated with them and that taking into account probabilistic models in the mapping INTSINT-IF contributes to the information organisation in the system. This could be of interest not only for speech technology research and applications but also for psycholinguistic research on speech parsing.

4. Daniel Hirst and Co.

Since 1990 when he became Directeur de Recherche (CNRS) in the LPL, Daniel Hirst has supervised 14 theses and has been a committee member of 65 other theses in France and abroad, and he currently still supervises the work of two PhD students. He has taken part in several international research projects (including Brazil, Venezuela, Finland, China, Algeria, South Korea and United Kingdom) and was the organiser of the 1st Speech Prosody Conference, held in France in 2002 and then every two years in Japan (2004), Germany (2006), Brazil (2008) and the USA (2010) and this year in China.

This 6-page synthesis of Daniel Hirst's contribution to prosodic annotation is unfortunately not sufficient to give a full account of his influence in supervising PhD students' work, nor to relate all the advice he has given, nor his many collaborations worldwide. Those reported here represent just a small part of all his scientific productions but hopefully will give insights of his legacies to the study of speech prosody.

Daniel was required to retire on the 29th of September 2011. However he does not easily give up the stage and was recently appointed professor at Tongji University, Shanghai - he also still continues his research in the CNRS in France as Emeritus Director of Research at the LPL for another 5 years (renewable!) so we can hope to work with him and benefit from his experience and benevolence for many years to come.

5. Conclusion

Daniel Hirst's tools for the automatic analysis and annotation of speech prosody have been developed under the idea that the task of phoneticians and linguists involved in establishing efficient rules for the analysis and synthesis of speech prosody has to be facilitated. One of his main concerns has always been to provide "theory-independent", "user-friendly" and freely-distributed tools which clearly participate in augmenting current linguistic knowledge.

In [22] he writes: "*Providing linguists with better tools will surely result in the availability of more and better data on a wide variety of languages, and such data will necessarily be of considerable interest not only to linguistics as a potentially cumulative science but also to engineers working with speech technology. I sincerely hope that the ProZed algorithm which I describe here will make a modest contribution to this development*". "Modest" is probably an adjective that defines Daniel Hirst well, but certainly not the tools he has developed nor the contribution he has made to speech sciences.

6. References

[1] Ali S, Hirst, D.J., "Developing an automatic functional annotation system for British English intonation". Proceedings Interspeech X, Brighton, 2009.

[2] Auran, C., "Prosodie et anaphore dans le discours en anglais et en français : cohésion et attribution référentielle''. Doctoral thesis, Université de Provence, 2004.

[3] Bertrand, R., "De l'hétérogénéité de la parole: analyse énonciative de phénomènes prosodiques et kinésiques dans l'interaction interindividuelle". Doctoral thesis, Université de Provence, 1999.

[4] Bigi, B., and Hirst, Daniel., “SPPAS: a tool for the automatic analysis of speech prosody”. 6th International conference on Speech Prosody, Shanghai, PRC. [This volume].

[5] Boersma, P. and Weenink, D., “Praat: doing phonetics by computer” [computer program], 2012.

[6] Braun, M., Chaloupka, V., “Carbamazepine induced pitch shift and octave space representation”. Hear. Res., 210, 85-92, 2005.

[7] Braun, M., “A retrospective study of the spectral probability of spontaneous otoacoustic emissions: Rise of octave shifted second mode after infancy”. Hear. Res. 215, 39-46, 2006.

[8] Cho, H., “Etude des propriétés acoustiques de la structure prosodique du coréen“. Doctoral thesis: Université de Provence, 2009.

[9] Cho, H. and Rauzy, S., “Phonetic pitch movements of accentual phrases in Korean read speech”. Proceedings of Speech Prosody 2008, Campinas, Brazil, 2008.

[10] De Looze, C. and Hirst, D.J., “Integrating changes of register into automatic intonation analysis”. Proceedings of the Speech Prosody 2010 Conference, Chicago, United States, 2010.

[11] De Looze, C. and Hirst, D.J., “L'échelle OME (Octave-MEdiane) : une échelle naturelle pour la mélodie de la parole”. Proceedings of the XXVIIIème Journées d’Etude sur la Parole (JEP 2010), Mons, Belgium, 2010.

[12] De Looze, C., “Analyse et interprétation de l’empan temporal des variations prosodiques en français et en anglais“. Doctoral thesis, 2010.

[13] Dutoit, T., “An introduction to Text-to-Speech synthesis”. Kluwer Academic Press, Dordrecht, 1997.

[14] Estruch, M., “Évaluation de l'algorithme de stylisation mélodique MOMEL et du système de codage symbolique INTSINT avec un corpus de passages en Catalan“. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence, 19: 45-61, 2000.

[15] Fernandez-Cruz, R., “L'analyse phonologique et acoustique du portugais parlé par des communautés noires de l'Amazonie“. Doctoral thesis, Université de Provence, 2000.

[16] Giordano, R., “Analisi prosodica e transcrizione intonativa in INTSINT”. Leoni and Giordano [Eds] 2005. Italiano parlato : analisi di un dialogo. (Liguori editore, Naples), 231-256, 2005.

[17] Herment, S., Loukina, A., Tortel, A., Hirst, D.J. and Bigi, B., AixOx, “a multi-layered learners corpus: automatic annotation”. Proceedings of the 4th International Conference on Corpus Linguistics, Jaèn, Spain, 2012.

[18] Hirst, D.J., “Intonative Features. A Syntactic Approach to English Intonation”. Mouton, The Hague, 1977.

[19] Hirst, D.J., “La représentation linguistique des systèmes prosodiques: une approche cognitive“. Habilitation Thesis, Université de Provence, 1987.

[20] Hirst, D.J., “Intonation in British English. D. Hirst, and A. Di Cristo [Eds], Intonation Systems. A Survey of Twenty Languages,” 56-77. Cambridge: Cambridge University Press, 1998.

[21] Hirst, D.J., “Form and function in the representation of speech prosody”. K.Hirose, D.J.Hirst, and Y.Sagisaka [Eds] Quantitative prosody modeling for natural speech description and generation (=Speech Communication 46 (3-4)), 334-347, 2005.

[22] Hirst, D.J., “A Praat plugin for MOMEL and INTSINT with improved algorithms for modelling and coding intonation.” Proceedings of the 16th International Congress of Phonetic, 2007.

[23] Hirst, D.J., “The analysis by synthesis of speech melody: from data to models”, Journal of speech Sciences 1 (1):55-83, 2011.

[24] Hirst, D.J., “ProZed: A speech prosody analysis-by-synthesis tool for linguists”. Proceedings of the 6th International Conference on Speech Prosody., Shanghai, 2012 [this volume].

[25] Hirst, D.J. and Di Cristo, A., “Intonation Systems. A Survey of Twenty Languages”. Cambridge: Cambridge University Press, 1998.

[26] Hirst, D.J., Di Cristo, A. and Espesser, R., “Levels of representation and levels of analysis for the description of intonational systems”. Horne M. [Ed] Prosody: Theory and Experiment. Dordrecht: Kluwer Academic Press, 51-88, 2000.

[27] Hirst, D.J. and Espesser, R., “Automatic modelling of fundamental frequency using a quadratic spline function”. Travaux de l’Institut de Phontique d’Aix 15: 71-85, 1993.

[28] Hirst, D.J., Nicolas, P. and Espesser, R., “Coding the F0 of a continuous text in French: an Experimental Approach”. ICPhS 12 (Aix en Provence), Vol. 5: 234-237, 1991.

[29] Kim, S., Hirst, D.J., Cho, H., Lee, H. and Chung M., “Korean MULTEXT: A Korean Prosody Corpus”. Proceedings of Speech Prosody 2008, Campinas, Brazil, 2008.

[30] Liu, J., Wang, N., Li, J., Shi, B., and Wang, H., “Frequency distribution of synchronized spontaneous otoacoustic emissions showing sex-dependent differences and asymmetry between ears in 2- to 4-day-old neonates”. Int J Pediatr Otorhinolaryngol. May, 73(5):731-6, 2009.

[31] Louw, J.A. and Barnard, E., “Automatic modeling with INTSINT”. Proceedings of the 15th Annual Symposium of the Pattern Recognition Association of South Africa, Grabouw, 107-111, 2004.

[32] Maghbouleh, A., “ToBI accent type recognition”, Proceedings ICSLP, 1998.

[33] Mixdorff, H., “A novel approach to the fully automatic extraction of Fujisaki model parameters”. ICASSP 2000, vol. 3: 1281-1284, 2000.

[34] Monaghan, A., “Generating synthetic prosody: means and ends “. Pré-publication des Actes du Séminaire Prosodie, (Aix, octobre 1992) 9-24, 1992.

[35] Mora-Gallardo, E., “Caractérisation prosodique de la variation dialectale de l'espagnol parlé au Venezuela“. Doctoral thesis, Université de Provence, 1996.

[36] Najim, Z., “Prosodie de l'arabe standard parlé au Maroc : analyse historique, sociolinguistique et expérimentale“. Doctoral thesis, Université de Provence, 1995.

[37] Nesterenko, I., “Analyse formelle et implémentation phonétique de l'intonation du parler russe spontané en vue d'une application à la synthèse vocale“. Doctoral thesis, Université de Provence, 2006.

[38] Nesterenko, I. and Rauzy, S., “On the use of probabilistic grammars in speech annotation and segmentation tasks”. Proceedings of SPECOM 2007, Moscow, Russia, 516-522, 2007.

[39] Nicolas, P., “Contribution de la prosodie à l'amélioration de la parole de synthèse: cas du texte lu en français“. Doctoral thesis, Université de Provence, 1995.

[40] Pierrehumbert, J., “The Phonology and Phonetics of English Intonation”. PhD Dissertation, Cambridge, Mass., MIT, 1980.

[41] Portes, C., “Prosodie et économie du discours : spécificité phonétique, écologie discursive et portée pragmatique du patron d'implication“. Doctoral thesis, Université de Provence, 2004.

[42] Scarna, A. and Ellis, A.W., “On the assessment of grammatical gender knowledge in aphasia: the danger of relying on explicit, metalinguistic tasks”. Language Cognitive Processes 17 (2), 185–201, 2002.

[43] Tortel, A., “Evaluation qualitative de la prosodie d'apprenants français: apports de paramétrisations prosodiques“. Doctoral thesis, Aix-Marseille Université, 2009.

[44] Vainio, M., Hirst, D.J., Suni, A. and De Looze, C., “Using functional prosodic annotation for high quality multilingual, multidialectal and multistyle speech synthesis”. 13th International Conference on Speech and Computer, St. Petersburg, Russie, 164-169, 2009.

[45] Wightman, C. W. and Campbell, W. N., “Improved Labeling of Prosodic Structure”. IEEE Trans. on Speech and Audio Processing. 1995.

[46] Wright, A., Rivera, J., Hulse, S., Shyan, M. and Neiworth, “Music perception and octave generalization in rhesus monkeys”. J Exp Psychol Gen, Sep, Vol 129 No 3, 291-307, 2000.

[47] Young, R.W., “Terminology for logarithmic frequency units”. The Journal of the Acoustical Society of America, 11: 134, 1939.

[48] Zhi, N., Hirst, D.J. and Bertinetto, P.M., “Automatic analysis of the intonation of a tone language. Applying the Momel algorithm to spontaneous standard Chinese (Beijing)”. Proceedings of Interspeech XI, 2010.

[49] Zhi, N. “The music of Beijing Chinese speech. On the interactions of tones and intonations in read and spontaneous Beijing speech”. Doctoral thesis, Scuola Normale Superiore di Pisa, in progress.