Lexical and Non-lexical Tone and Prosodic Typology

Daniel Hirst

Lexical and Non-lexical Tone and Prosodic Typology

Daniel Hirst

CNRS UMR 6057 Parole et Langage, Université de Provence
daniel.hirst@lpl.univ-aix.fr

Abstract

Prosodic typology has generally concentrated on those aspects of prosodic representation which are assumed to be represented in the lexicon. It is argued here that non-lexical representation at various levels, underlying phonological, surface phonological and phonetic, can also constitute a basis for prosodic typology. An example is given of a low-level comparison of English and French pitch patterns. A prosodic model integrating these different levels is presented which, it argued can provide a useful tool for the investigation of prosodic typology and for a more robust basis for establishing the more abstract levels including those of lexical representations.

1. Introduction

Prosodic typology is often limited to the classification of languages with respect to their lexical characteristics. In this presentation I argue that utterances can be characterised at several different levels of abstraction and I present a model capable of representing the prosody of utterances in a language-independent manner at these different levels. While languages differ considerably at the most abstract levels, at the more concrete levels these differences are less obvious. I show, however, that even at the *phonetic* level, typological differences can be observed between languages such as French and English. I argue that the application of such a multi-level model can provide a more robust basis for characterising prosodic typologies including those at the higher and more abstract levels.

2. Lexical Prosody

2.1. Lexical prosodic typology

One of the basic functions of prosody, found in almost all languages of the world is to contribute to the identity of lexical items. These lexical characteristics are those which a speaker of the language must know in order to speak and understand the language. A sufficient (but arguably not necessary) condition for claiming that a prosodic characteristic of a language is lexical is the existence of minimal pairs in the language, which differ only in that feature.

These lexical prosodic characteristics constitute a natural and fairly consensual basis for a language typology and it is widely accepted that *accent*, *quantity* and *tone* are the basic categories of this typology.

A fourth category has sometimes been introduced, intermediate between tone and accent and usually referred to as *pitch-accent*.

Pitch accent, as found in Japanese, for example, shares with accent the fact that it is syntagmatic by nature, only one syllable of a lexical item can carry the accent. At the same time it shares with tone the fact that it is paradigmatic, i.e. that it can be present or absent in a given word.

Thus, in a hypothetical tone language with just two tones, high (H) and low (L), we might expect to find a four-way contrast on bisyllabic words.

L L L H H L H H

In an accent language, we would find only two possible patterns with stressed (S) and unstressed (u) syllables:

S u u S

By contrast, in a pitch accent language like Japanese (Abe, 1998) we find three patterns with unaccented (u) and lexically high, accented syllables (H):

u u u H H u

Languages like Swedish are sometimes included in the same category as Japanese.

In Swedish, however, the distinction between acute (A1) and grave (A2) accents only applies to non-final syllables (Bruce, 1977), (Gårding, 1998) so that the following patterns are possible:

A1 u A2 u u A1

I have suggested (Hirst and Di Cristo, 1998a), that rather than introduce a fourth lexical prosodic category, these two languages might better be characterised as combining tone and accent, although each in a different way.

The Japanese system is primarily tonal and only secondarily accentual. Not every word in Japanese has a lexically specified high tone; when it is present, it can only occur on one syllable in the word.

The Swedish system is the opposite, basically an accentual system with tonal characteristics. Every word in Swedish must have a lexically specified accent; in non-final position this accent may be one of two (tonal) types.

We might consequently characterise Japanese as an accentual tone language and Swedish as a tonal accent language.

The way in which lexical characteristics are represented in the lexicon is, of course, extremely theory-dependent. The empirical basis for classifying a given language as having lexical accent, quantity or tone (or any combination of these) is, furthermore, by no means a trivial question.

2.2. Languages without lexical prosody

In a rather small minority of languages, there seems to be no need to specify any lexical prosody at all.

Modern standard French appears to be an example. Note however that this is not true of some dialects of modern French. Some conservative speakers today maintain what was

once a more productive distinction between long and short vowels as in:

"mètre" /mɛtʁ/ (metre)
vs.
"maître"/mɛːtʁ/ (master).

Arguably in dialects like midi French, where, unlike in standard French, the schwa vowels are generally maintained, pairs such as

"boiteux" /bwaˈtø/ (lame)
and
"boîte" /ˈbwatə/ (box)

could be taken as minimal pairs for stress with the /ə/ vowel being analysed as an unstressed allophone of /ø/.

The fact that modern standard French does not require the lexical specification of prosody does not, of course, mean that French has no prosody, but rather that we need to account for this prosody at a more surface level.

French connected speech has contrasts which, on the surface, are rather similar to those found in languages with lexical contrasts.

These include contrasts of of accentuation as in:

J'enlève son verre /ʒɑ̃ˈlɛvsɔ̃ˈvɛʁ/
(I take away his glass)

as against:

Jean lève son verre /ˈʒɑ̃ˈlɛvsɔ̃ˈvɛʁ/
(Jean raises his glass)

constrasts of duration for consonants as in:

Il part tôt /ilpaʁto/ (he leaves early)

as against:

Il partent tôt /ilpaʁtto/ (they leave early)

contrasts for duration of vowels:

"Il va battre l'ennemi" /ilvabatʁlenəmi/
(he is going to beat the enemy)

as against:

"Il va abattre" /ilvaːbatʁlenəmi/
(he is going to cut down the enemy)

as well, of course, as melodic contrasts as in:

Non. Non… Non! Non?

In the case of languages with no lexical prosody, we need to provide an explanation for the way in which utterances which in their underlying (lexical) representation have no specification for prosody are provided with such a specification in the final output.

3. Underlying and surface prosody

One possible explanation is that the observable prosodic characteristics of utterances in all languages are determined by an underlying abstract representation. Just as it is not possible to pronounce an utterance with no fundamental frequency, no intensity and no duration, so, on this account, it is not possible to produce an utterance without some specification of tone, accent and quantity. As we have seen, these are sometimes determined by the lexicon of the language. When they are not, we might assume some form of language-specific prosodic well-formedness constraint to provide an underlying representation with its missing prosodic specifications.

Just as with the lexical representation of prosody, the way in which these specifications are formulated are likely to be very theory-dependent. My colleagues and I have suggested elsewhere (Hirst et al., 2000) that between the lexical representation and the acoustic signal we may usefully distinguish three levels of representation which we refer to as:

- underlying phonological representation
- surface phonological representation
- phonetic representation

The distinction between phonetic and phonological representation which we adopt is basically that of Trubetzkoy (Trubetzkoy, 1939) who distinguished between representations consisting of continuous variables (phonetic) and those containing discrete categories (phonological).

I shall return to phonetic representations below (section 3.1). The distinction between surface and underlying representations is between representations where each phonological symbol corresponds to some observable characteristic of the speech signal and those where it does not necessarily do so. Very often it is possible to simplify the linguistic description by postulating more abstract elements which are not directly observable.

A classic example is that of *downstep* as illustrated in the Ghana language, *Akan*. (Fromkin, 1972).

In this language, there is a three way surface tonal contrast on the final syllable of disyllabic words.

/mɛ́ hɔ̀/ (I will strike)
/mé bo/ (my stone)
/mé bɔ̀/ (my breast)

where /´/ represents a high tone, /`/ a low tone and unmarked vowels correspond to a tonal realisation intermediate between high and low. In the earliest accounts of this and similar languages, this intermediate tone was called a "mid" tone. A number of facts remained mysterious about this mid tone however. First, it could only occur after a high tone. So with bisyllabic words, instead of the nine expected patterns for three tones, only five were attested:

LL LH HL HM HH

Furthermore, this mid tone had an effect on any following high tones which are lowered to the same level as that of the mid tone.

The explanation offered for this strange behaviour was that the word "stone", in other contexts, appears in the form /ɔ `bó/ with a low tone nominal prefix. This prefix is normally elided after a preceding vowel. Since there is, in Akan, as in many other tonal and non-tonal languages, a general effect of *downdrift* which has the effect of lowering the second high tone in a sequence /H L H/, Fromkin observed that if the expression "my stone" were realised with what is presumably its underlying form /mé ò bó/, this would provide precisely the observed pitch height for the final syllable. Instead of representing this tone as a mid tone, then it was proposed to represent it as a *downstepped* high tone,

/mé ꜜbó/

where the downstepping is assumed to be the surface reflex of the underlying low tone.

A more abstract solution to this analysis was later proposed (Clements and Ford, 1979). On this analysis, the apocope of the vowel /ɔ/ affects only the vowel and not the tone so that the tone is left 'floating' in the phonological representation. A *floating tone* is assumed not to surface as a pitch target but to continue to affect the phonetic value of the following high tone. The notion of floating tone has since been used to account for tonal phenomena in a wide range of languages including a number of cases where the only surface trace of a morpheme is its effect on following tones (Goldsmith, 1990).

An account making use of an underlying floating tone is in many cases simpler and more explanatory than an account describing only the observable surface patterns.

The notion of floating tone has also been used in the description of intonation patterns as an underlying representation of a surface downstepped tone (Hirst, 1998). I have further argued that a surface difference between neutral declarative patterns in British English and American English as a downstepping sequence or accents or a sequence of falling pitch accents respectively (Pike, 1945) can be accounted for as being derived from the same underlying representation.

A characterisation of a typological prosodic difference between French and English intonation patterns, based on the notion of a prosodic template for tonal units [LH] and [LH] (Hirst, 1988), (Hirst et al., 2000) also makes use of the notion of underlying phonological representation.

The distinction between surface and underlying phonological representations thus makes it possible to envisage a wider application of the notion of prosodic typology going beyond lexical contrasts to underlying and surface phonological representations, whether these are lexical or derived with respect to prosodic well-formedness constraints or templates. The existence of floating tones in a language for example can only be brought to light by the association of underlying and surface representations.

3.1. Phonetic representations

Between the surface phonological representation and the acoustic signal we assume a level which we refer to as that of *phonetic representation.* Unlike those models of fundamental frequency which are essentially production-oriented (Fujisaki 1988) or perception-oriented (Hart ('t) et al., 1991), (Alessandro (d') and Mertens, 1995), the model we propose as a phonetic representation is an acoustic model of fundamental frequency target points corresponding to positions on the modelled curve where the slope is null (= zero first derivative) linked together by a smooth continuous monotonic transition. We take this level of phonetic representation to be an appropriate interface between constraints on speech production and speech perception.

The *Momel* algorithm developped at the LPL (Hirst and Espesser, 1993, Hirst et al., 2000, Horne, 2000) provides an automatic phonetic representation of a fundamental frequency curve. The algorithm is often referred to as a *stylisation* of fundamental frequency but it should more properly be called a *model* since it consists in factoring the raw fundamental frequency curve into two components without any loss of information. These are a macroprosodic component, consisting of a continuous smooth curve (represented as a quadratic spline function) corresponding to the linguistic function of the contour, and a microprosodic component consisting of deviations from the macroprosodic curve caused by the nature of the phonematic segment (voiced/unvoiced obstruent, sonorant, vowel etc) (cf(Di Cristo and Hirst, 1986). The output of the algorithm is a sequence of target points which are sufficient to define the macroprosodic component of the fundamental frequency when used as input to a quadratic spline function.

The Momel algorithm is currently implemented in various speech-analysis environments including *Mes* (Unix) (Espesser, 1996), *SFS* (Windows) (Huckvale, 2000) as well as in the multi-platform system *Praat* (Boersma and Weenink, 1996-2004) in the form of a script (Auran, 2003) calling Momel as an external C program. The software is freely available for non-commercial non-military research.

A recent evaluation of the algorithm was made (Campione, 2001) using recordings of the continuous passages of the Eurom1 corpus for five languages (English, German, Spanish, French, Italian) in all a total of 5 hours of speech. The evaluation estimated a global precision of 97.6% by comparison with manually corrected target estimation. Compared to the 46982 target points provided by the automatic analysis, 3179 were added manually by the correctors and 1107 removed. The algorithm gave only slightly worse results (93.4% precision) when applied to a corpus of spontaneous spoken French. The majority of these corrections involved systematic errors, in particular before pauses, which an improvement of the algorithm should eliminate.

The output of the algorithm as a sequence of target points is particularly suitable for interpretation as a sequence of tonal segments such as the INTSINT representation described below, but the relatively theory-neutral nature of the modelling, together with its reversibility, has allowed the algorithm to be used as input for other types of annotation including ToBI (Maghbouleh, 1998, Wightman and Campbell, 1995) and the Fujisaki model (Mixdorff, 1999)

3.2. Surface phonological representation

The prosodic annotation alphabet INTSINT was based on the descriptions of the surface patterns of the intonation of twenty languages (Hirst and Di Cristo, 1998b) and was used in that volume for the description of nine languages (British English, Spanish, European Portuguese, Brazilian Portuguese, French, Romanian, Bulgarian, Moroccan Arabic and Japanese).

Intonation patterns are analysed in this framework as consisting of a sequence of tonal segments, defined in one of two ways: either globally with respect to the speaker's pitch range (**Top**, **Mid** or **Bottom**) or locally with respect to the preceding target (**Higher**, **Same** or **Lower**) with an iterative variant of these locally defined targets (**Upstepped**, **Downstepped**) assuming that an iterative tone can be followed by the same tone whereas a non-iterative tone cannot and furthermore that the iterative tones correspond to a smaller pitch interval than the non-iterative ones.

This transcription system, originally designed as a tool for linguists transcribing the intonation of utterances of different languages, was intended to provide at least a first approximation to a prosodic equivalent of the International Phonetic Alphabet. As the authors of the ToBI system themselves insist (Pierrehumbert et al., 2004 (in press)) this is specifically not the case for the ToBI system.

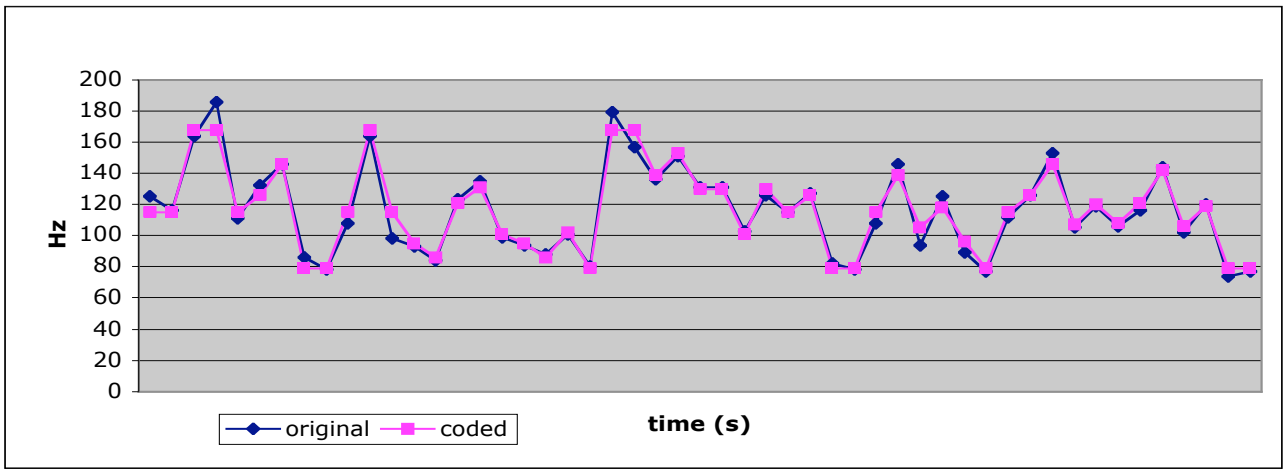

Figure 1. *Coding of the F0 targets from a passage from the Eurom1 corpus showing the original target points(lozenges) estimated by the Momel algorithm and the target points(squares) derived from the optimised INTSINT coding.*

In the case of INTSINT, it was intended from the first that the transcription should be convertible to and from a sequence of target points.

A first version of an algorithm for converting between Momel and INTSINT was described in (Hirst et al., 2000). An extension of the system to annotate duration and tonal alignment has also been proposed (Hirst, 1999).

A simpler and more robust algorithm has since been developed (Hirst, 2001). In this version, target points are coded on the basis of two speaker (and perhaps utterance) dependent parameters: *key* and *range*. Given these, the absolute tones are defined as the limits of the speaker's pitch range (**Top** and **Bottom**) assumed to be symmetrical around the central value (**Mid**). The relative tones are then defined by an interval between the preceding target point ($\mathbf{P_{i-1}}$) and the two extreme values taken as an asymptote for these targets as in the following:

$$P_i = P_{i-1} + c.(A-P_i)$$

where **A** is either **T**, (for **H** and **U**) or **B** (for **L** and **D**) and where **c** is set at 0.5 for the non-iterative targets **H** and **L** and at 0.25 for the iterative targets **U** and **D**.

This algorithm, applied to the targets of the French and English passages of the Eurom1 corpus (Chan et al., 1995), was optimised over the parameter space:

$$key = mean \pm 50 \text{ (in Hz)}$$
$$range \in [0.5, 2.5] \text{ (in octaves)}.$$

Interestingly, the mean optimal range parameter resulting from this analysis was not significantly different from 1.0 octave. It remains to be seen, however, how far this result is due to the nature of the EUROM1 corpus which was analysed (40 passages consisting each of 5 semantically connected sentences) and whether it can be generalised to other speech styles and other (particularly non-European) languages.

The symbolic coding of the F0 target points obviously entails some loss of information with respect to the original data, unlike the Momel analysis which is entirely reversible.

The loss of information is, however, quite small as can be seen from Figure 1 which illustrates the output from the optimised INTSINT coding compared to the original target points for a complete five sentence passage from the Eurom1 corpus.

4. Prosodic paradigms

Work on automatic language identification (Thymé-Gobbel and Hutchins, 1996) has shown that including prosodic parameters derived from measurements of pitch and amplitude contours on a syllable by syllable basis led to an improvement in the performance of a segmental based language identification system when applied to four languages (English, Spanish, Japanese and Chinese) chosen as representatives of different typological groups. Overall features derived from measurements of pitch were found to be the most useful for discrimination. (Cummins et al., 2000) obtained similar results from a recurrent neural network using only delta-F_0 and the band limited amplitude envelope as network inputs.

Phonological comparisons of the intonation patterns of different languages suggest that the analysis of fundamental frequency patterns should reveal significant differences between different languages. In the case of English and French, (Hirst, 1988, Hirst and Di Cristo, 1998a, Jun and Fougeron, 2000) brought to light a distinction between the underlying pitch patterns[1]. Abstracting away from more global intonation patterns, words in English are basically associated with a falling pitch pattern whereas they are associated with a rising pitch pattern in French. This phonological characterisation, however, undergoes a number of local modifications so that the actual observed surface configurations may be quite different from these more abstract underlying patterns.

[1] Details of these analyses of French intonation patterns differ. Jun & Fougeron associate a double rising pattern LHLH directly with words, while Hirst and Di Cristo associate a simple rising pattern LH with a Tonal Unit, of which there may be more than one per word.

Since this algorithm makes it possible to extract fundamental frequency targets from a raw fundamental frequency curve, this opens the possibility of typological analyses of tonal phenomena at the phonetic level. In the next section I present results from an attempt to carry out such a comparative analysis on the fundamental frequency patterns of English and French (Hirst, 2003).

5. A comparison of pitch parameters of English and French.

In the course of the European *SAM* project, a multilingual corpus *Eurom1* was recorded containing a number of different types of read speech including numbers, sentences and continuous 5 sentence passages(Chan et al., 1995). During the *Multext* project (Véronis et al., 1994.), the continuous passages of the *Eurom1* corpus were analysed and annotated with hand-aligned word labels and hand-corrected F0 target point estimation using the *Momel* algorithm (Hirst and Espesser, 1993, Hirst et al., 2000). The resulting prosodic database for 5 languages (English, French, German, Italian, Spanish) is currently distributed by *ELRA* (Campione and Véronis, 1999a).

In this study, pitch parameters derived from the English and French passages were analysed. It has been shown (Hirst et al., 2000) that re-synthesis replacing the original F0 by a quadratic spline function defined by a sequence of target points is virtually indistinguishable from the orginal recording, The following analyses consequently made use only of the target points obtained from the recordings. Seven parameters were calculated from the sequence of target points for each recording of each passage.

- *octave* : the absolute $\log_2$ value of the individual targets
- *interval*: the absolute (octave) difference between successive targets
- *rise*: the octave difference between successive targets calculated only when the second value is greater than the first
- *fall*: the octave difference between successive targets calculated when the first value is greater than the second
- *slope*: the absolute slope in octaves per second between successive targets
- *rise-slope*: the slope between successive targets for rises
- *fall-slope*: the slope between successive targets for falls

For each parameter the mean, standard deviation and coefficient of variation were calculated.

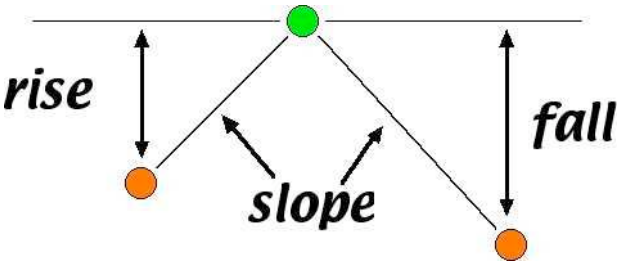

Figure 2. *Illustration of the parameters of interval (in octaves) and slope (in octaves/second) for rising and falling sequences of F0 targets.*

As had already been shown (Campione and Véronis, 1999b), the analysis of these target points reveals significant effects for language and gender of speakers for this corpus.

As expected, male speakers had significantly lower mean values than female speakers with mean values respectively of 136 and 233 Hz ($F(1;246) = 1070$, $p < 0.0001$). There was also, however, a significant difference between French speakers who were significantly higher pitched than English speakers ($F(1;246) = 71$, $p < 0.0001$). The interaction between the two factors was, however, also highly significant. ($F(1;246) = 15$, $p < 0.0001$). The mean values (in Hz) were as follows:

Table 1. *Mean values of target values for English and French male and female speakers.*

	Male	Female
English	131	213
French	142	262

The small number of speakers involved in the study and the large inter-speaker variability, as can be seen in Figure 2, makes it difficult to predict whether this language specific gender effect would be replicated for larger databases.

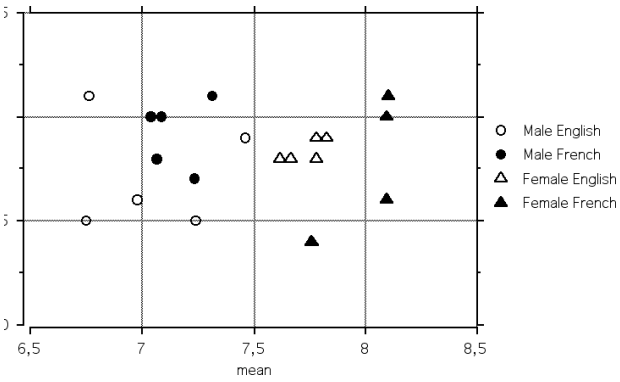

Figure 2. Mean vs. coefficient of variation of F0 targets for male (circles) and female (triangles) speakers of English (empty) and French (filled). Two English female speakers had nearly identical values and are not distinguishable in this figure.

Analysis of variance on the 21 different parameters analysed revealed highly significant ($p < 0.0001$) differences between the English and French recordings for a number of parameters.

Table 2 summarises these parameters ordered by descending degree of significance. Parameters marked * also showed a significant gender effect. Parameters marked ** also showed a significant interaction between the effects of language and gender. and are consequently likely to be less useful for discrimination.

Table 2. *Parameters by descending degree of significance. m = mean, sd = standard deviation, cv = coefficient of variation.*

Parameter	F value (1;246) p < 0.0001
Interval (cv)	83
*Rise interval (m)	77
**Octave (m)	71
**Fall interval (m)	71
*Fall interval (cv)	68
Fall interval (sd)	54
Interval (sd)	43
*Fall slope (cv)	31
*Octave (cv)	30
Absolute slope (m)	25
Fall slope (sd)	20
Octave (sd)	19
**Rise slope (cv)	18
*Rise interval (cv)	16

The 21 parameters were submitted to a discriminant analysis using the *Praat* software (Boersma and Weenink, 1996-2004). On the basis of this, the language was correctly identified for 87.6% of the recordings with the following confusion matrix.

Table 3. *Classification matrix for discriminant analysis*

	Predicted	
	English	French
English	132	18
French	13	87

Five individual parameters each gave over 70% correct discrimination in isolation:

Table 4. *Parameters by decreasing percentage of correct discrimination. m = mean, sd = standard deviation, cv = coefficient of variation.*

Parameter	Percentage correct
Absolute interval (cv)	74.0
Rise interval (m)	72.4
Octave (m)	71.6
Fall interval (m)	71.6
Fall slope (cv)	70.8

Four out of all possible combinations of two parameters gave over 79% correct identification with, in each case, the parameter Fall Interval (sd) combined with either Octave (cv), Fall (sd), Fall (cv) or Fall slope (m). Three out of all possible combinations of three parameters gave each 82.8 correct identification:

- Octave (m) + Interval (sd) + Rise interval (m)
- Interval (cv) + Rise interval (m) + Fall slope (m)
- Interval (cv) + Rise slope (m) + Fall slope (m)

A final statistical test on these parameters was obtained by using a Classification and Regression Tree analysis with the *Cruise* software available from Kim Hyun Joong (Kim and Loh, 2001). Using this program, the passages were divided into a training set of 230 recordings and a test set consisting of twenty recordings (one from each of the twenty speakers). The algorithm was run using its default values which include univariate split type and linear discriminant split method, estimated prior probabilities from the distribution of the training set, equal misclassification costs and pruning by cross validation. The resulting optimised tree contained only 7 terminal nodes and achieved 86.5% correct classification on the training data .

Table 5. *Classification matrix for regression tree using the Cruise algorithm on the training data.*

	Predicted	
	English	French
English	124	16
French	15	75

Applying the tree to the test data gave 90% correct identification.

Table 6. *Classification matrix for regression tree using the Cruise algorithm on the test data.*

	Predicted	
	English	French
English	9	1
French	0	10

Summing the two tables gives a total of 87.2% correct identification which is very close to the 87.6% given by the Discriminant Analysis using all 21 parameters and all of the data.

Figure 3 shows the regression tree output from this algorithm.

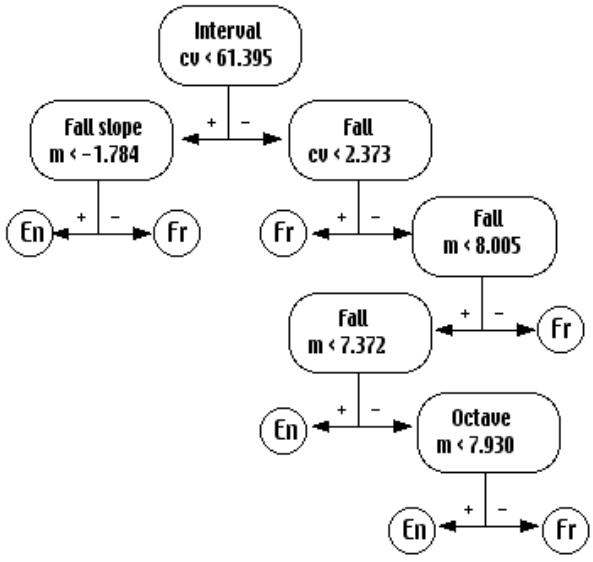

Figure *3. Regression tree from the CRUISE program showing the optimal prediction of language from the parameters (see text). cv= coefficient of variation, m = mean..*

The statistical analysis of the F0 targets from continuous passages recorded by ten English speakers and ten French

speakers confirmed that systematic differences can be found between the values of the target points for the two languages. Both the discriminant analysis and the regression tree analysis showed that parameters which are particularly useful for discriminating the languages are those involving a falling sequence of target points. The general tendancy seemed to be that in English, falls tended to be steeper, smaller and more variable than in French. These results were obtained without any consideration of the distribution of the falls with respect to phonological, lexical or syntactic constituents. It is likely that taking these into account will bring further light on the nature of the prosodic difference between the two languages.

6. Conclusion

In this presentation I have tried to show that while the prosodic characteristics of languages may appear very different when considered from the point of view of lexical prosody, the surface characteristics and phonetic representations of utterances from different languages are more similar than might be expected so that they can in fact be described using a common system of representation at these intermediate levels. I suggest furthermore that the use of such a common descriptive framework could provide a useful tool for establishing the more abstract prosodic characteristics of languages, including lexical characteristics, on a more robust empirical basis, in particular in the light of the application of such tools to large speech corpora (as in Auran et al., 2004) where automatic techniques of investigation are indispensable.

7. References

Abe, Isamu. 1998. Intonation in Japanese. In *Intonation Systems. A Survey of Twenty Languages*, eds. D.J. Hirst and A Di Cristo, 360-375. Cambridge: Cambridge University Press.

Alessandro (d'), C., and Mertens, P. 1995. Automatic intonation stylisation using a model of pitch perception. *Computer Speech and Language*.

Auran, C. 2003. Momel and Intsint package. http://www.lpl.univ-aix.fr/~auran/.

Auran, C., Bouzon, C., and Hirst, D.J. 2004. The Aix-Marsec project. An evolutive database of spoken British English. In *Proceedings of the Second International Conference on Speech Prosody*, eds. Bernard Bel and Isabelle Marlien. Nara.

Boersma, Paul, and Weenink, David. 1996-2004. *Praat, a system for doing phonetics by computer*. University of Amsterdam http://www.fon.hum.uva.nl/praat/.

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence perspective*.: TILL XII.

Campione, E. 2001. *Etiquetage prosodique semi-automatique de corpus oraux : algorithmes et méthodologie*, Université de Provence, Aix-en-Provence: Doctoral thesis.

Campione, E. , and Véronis, J. 1999a. A multilingual prosodic database. Paper presented at *ICSLP*, Sydney.

Campione, E. , and Véronis, J. 1999b. A statistical study of pitch target points in five languages. Paper presented at *ICSLP*, Sydney.

Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Tracoso, I., Veld, C., and Zeiliger, J. 1995. Eurom - a spoken language resource for the EU. Paper presented at *4th European Conference on Speech Technology (Eurospeech '95)*, Madrid.

Clements, N., and Ford, K.C. 1979. Kikuyu Tone Shift and its Synchronic Consequences. *Linguistic Inquiry* 10:179-210.

Cummins, F., Gers, F., and Schmidhuber, J. 2000. Language identification from prosody without explicit features.

Di Cristo, A., and Hirst, D.J. 1986. Modelling French micromelody: analysis and synthesis. *Phonetica* 43:11-30.

Espesser, R. 1996. MES : Un environnement de traitement du signal. *Proceedings XXIe Journées d'Etude sur la Parole* 1996 June 10-14 : Avignon, France:447.

Fromkin, V. 1972. Tone features and tone rules. *Studies in African Linguistics* 3:47-76.

Fujisaki , H. 1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In *Vocal Physiology: Voice Production, Mechanisms and Functions*, ed. O. Fujimura. New York: Raven Press.

Gårding, E. 1998. Intonation in Swedish. In *Intonation Systems. A Survey of Twenty Languages*., eds. D.J. Hirst and A Di Cristo, 112-130. Cambridge: Cambridge University Press.

Goldsmith, John A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

Hart ('t), J., Collier, R., and Cohen, A. 1991. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.

Hirst, D.J. 1988. Tonal units as phonological constituents: the evidence from French and English intonation. In *Autosegmental Studies in Pitch Accent*, eds. H. Van der Hulst and N. Smith, 151-165. Dordrecht: Foris.

Hirst, D.J., and Espesser, R. 1993. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix-en-Provence* 15:75-85.

Hirst, D.J. 1998. Intonation in British English. In *Intonation Systems. A Survey of Twenty Languages.,* eds. D.J. Hirst and A. Di Cristo, 56-77. Cambridge: Cambridge University Press.

Hirst, D.J., and Di Cristo, A. 1998a. A survey of intonation systems. In *Intonation Systems. A Survey of Twenty Languages*., eds. D.J. Hirst and A. Di Cristo, 1-44. Cambridge: Cambridge University Press.

Hirst, D.J., and Di Cristo, A. 1998b. *Intonation System. A Survey of Twenty Languages.* Cambridge: Cambridge University Press.

Hirst, D.J. 1999. The symbolic coding of duration and alignment. An extension to the INTSINT system. Paper presented at *Eurospeech '99*., Budapest.

Hirst, D.J., Di Cristo, A, and Espesser, R. 2000. Levels of representation and levels of analysis for the description of intonation systems. In *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*, ed. M. Horne, 51-87. Dordrecht: Kluwer Academic Publishers.

Hirst, D.J. 2001. Automatic analysis of prosody for multi-lingual speech corpora. In *Improvements in Speech Synthesis*, eds. Eric Keller, Gérard Bailly, Alex Monaghan, Jacques Terken and Mark Huckvale: Wiley.

Hirst, D.J. 2003. Pitch parameters for prosodic typology. A preliminary comparison of English and French. Proceedings *ICPhS*, Barcelona.

Horne, M. 2000. *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publishers.

Huckvale, M. 2000. *Speech Filing System. Tools for speech research*. http://www.phon.ucl.ac.uk/resource/sfs/.

Jun, Sun-Ah, and Fougeron, Cécile. 2000. A phonological model of French intonation. In *Intonation. Analysis, Modelling and Technology.,* ed. Antonis Botinis, 209-242. Dordrecht: Kluwer Academic Publishers.

Kim, H. , and Loh, W.-Y. 2001. Classification trees with unbiased multiway splits,. *Journal of the American Statistical Association* 96. http://www.stat.wisc.edu/~loh/cruise.html

Maghbouleh, A. 1998. ToBI accent type recognition. Paper presented at *International Conference on Spoken Language Processing*.

Mixdorff, H.-J. 1999. A novel approach to the fully automatic extraction of Fujisaki model parameters. Paper presented at *ICASSP 1999*.

Pierrehumbert, Janet, Hirschberg, Julia, and Shattuck-Hufnagel, Stefanie. 2004 (in press). The original ToBI system and the evolution of the ToBI framework. In *Prosodic Models and Transcription.*, ed. Sun-Ah Jun. Oxford: Oxford University Press.

Pike, Kenneth. L. 1945. *The intonation of American English*. Ann Arbor: The University of Michigan Press.

Thymé-Gobbel, A., and Hutchins, S.E. 1996. On using prosodic cues in automatic language identification. In *Proceedings ICSLP '96*, 1768-1771. Phliadelphia, PA.

Trubetzkoy, N.S. 1939. *Principes de Phonologie*. (= trans. by J. Cantineau of *Grundzüge de Phonologie*). Paris: Klincksieck.

Véronis, Jean, Hirst, D.J., Espesser, R., and Ide, Nancy. 1994. NL and speech in the MULTEXT project. Paper presented at *AAAI-94 Workshop on the Integration of Speech and Natural Language Processing.*, Seattle.

Wightman, C. , and Campbell, N. 1995. Improved labeling of prosodic structure. *IEEE Trans. on Speech and Audio Processing*.