# Pitch parameters for prosodic typology. A preliminary comparison of English and French

**Daniel Hirst**

Laboratoire Parole et Langage
CNRS & Université de Provence, Aix-en-Provence
daniel.hirst@lpl.univ-aix.fr

## Abstract

The search for objective paradigms for establishing prosodic typologies among languages is a major challenge for speech science. Recent work in the area of speech rhythm has shown that an appropriate choice of parameters can provide revealing evidence for traditional typological classifications. In the area of pitch there has been less activity. This paper presents a preliminary comparison of pitch parameters for English and French continuous passages from the multilingual corpus Eurom1. Pitch targets derived from the F0 curves obtained from the recordings using the Momel algorithm were compared without taking into account their distribution with respect to phonological, lexical or syntactic constituents. Despite this, the parameters analysed using both discriminant and regression tree analysis gave over 80% correct discrimination between the two languages and pointed to the characteristics of falling pitch movements as being particularly distinctive.

## 1. Introduction

Despite the vast quantity of research in the field of prosody carried out during the last century, there is actually rather little we really know today about the ways in which the prosody of different languages, dialects, individual speakers or speaking styles differ from one another. Much of the knowledge which we do have is in the form of rather abstract characterisations such as phonological descriptions. Attempts to translate these into the form of quantitative data have usually proved surprisingly unsuccessful.

This lack of quantitative characterisations is rather surprising. It seems obvious that such knowledge would be of considerable use in a number of important areas. Speech technology, including speech synthesis and automatic speech recognition, is one area where such knowledge seems crucial if we are not to start again from scratch each time we wish to describe another language, dialect, speaker or speaking style. Second language teaching is another area where objective quantitative evaluation procedures might be hoped to provide a precious assistance to both learners and teachers. The same thing applies to the field of speech pathology, where practitioners are often at a loss to describe, far less to quantify, the sometimes quite striking prosodic characteristics of their patients' speech.

A major challenge for research in the field of speech prosody for the twenty-first century will consequently be the search for prosodic paradigms, objective quantitative procedures which will give different results for different languages. Ideally such procedures would also provide a means of quantifying deviations from a norm so that they could be used directly as a difference metric for the characterisation of dialectal variants, non-native speakers, speech synthesis or speech pathology.

## 2. paradigms for prosodic typology

### 2.1. acoustic paradigms

Naïve speakers often have the impression that languages other than their native language are spoken faster or with higher pitch, and that utterances in these languages are louder, more musical, more rhythmical etc.

In fact, there is very little published evidence attesting any such simple acoustic differences between pairs of languages. Roach [22] noted that while some languages appear to be spoken faster than others when measured in terms of syllables per second, this difference disappears when the rate is measured in phones per second, and that observed differences in tempo are more likely to be due to differences in speaking style than to the language.

### 2.2. rhythm

Linguists have for many years attempted to classify languages following a typological distinction between stress timed languages, syllable timed languages and mora-timed languages [19, 1, 17].

Typical examples of each of these, it has been claimed, are English, French (but see [25]) and Japanese, respectively. Attempts to establish empirical foundations for such classification, however, for many years proved almost embarrassingly unsuccessful [21, 7]. It appeared that there were no significant differences between the mean durations of either syllables or stress-groups across the traditional language categories. In more recent work, however, [8, 20, 18] it has been shown that a more careful choice of parameters could provide an effective empirical basis for the traditional prosodic classification.


ISBN 1-876346-48-5 

### 2.3. pitch

Work on automatic language identification [23] has shown that including prosodic parameters derived from measurements of pitch and amplitude contours on a syllable by syllable basis can lead 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 [6] 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, the two languages with which we are concerned in this paper, [10, 11, 14] brought to light a distinction between the underlying pitch patterns[1]. Abstracting away from more global intonation patterns, accent groups 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, is subject to a number of local modifications so that the actual observed surface configurations may be quite different from these more abstract underlying patterns.

## 3. 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[5]. During the *Multext* project [24], the continuous passages of the *Eurom1* corpus were analysed and annotated with hand-aligned word labels and hand-corrected quadratic spline modellling of fundamental frequency curves using the *Momel* algorithm [13, 12]. The resulting prosodic database for 5 languages (English, French, German, Italian, Spanish) is now distributed by *ELRA* [3].

In this study, pitch parameters derived from the English and French passages were analysed. It has been shown [12] 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.

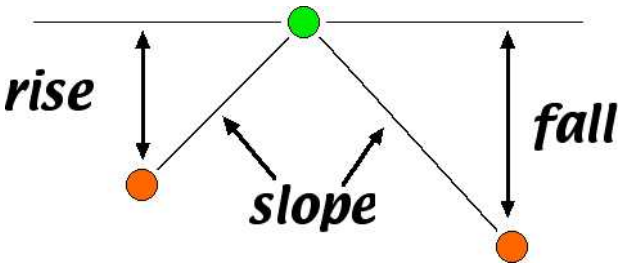


*Figure 1.* Illustration of the parameters of interval (in octaves) and slope (in octaves/second) for rising and falling sequences of F0 targets.

As had already been shown [4], 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.

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

[1] Details of these analyses 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.

for a number of parameters. Table 2 summarises these parameters ordered by descending degree of significance. m = mean, sd = standard deviation, cv = coefficient of variation. Paremeters 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.

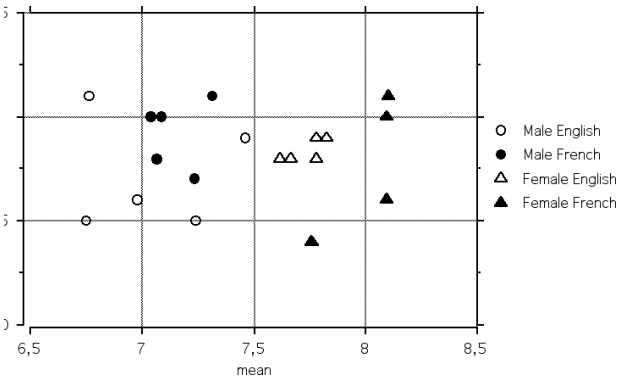


*Figure 2.* Mean vs. coefficient of variation of F0 targets for male (circles) and female (triangles) speakers of English (empty) and French (filled). Two English female speakers had nearly identical values and are not distinguishable in this figure.

*Table 2.* Parameters by descending degree of significance

| 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 [2]. 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 discrminant 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.

| 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 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 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 HyunJoong [16, 15]. 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 95% 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.


ISBN 1-876346-48-5 

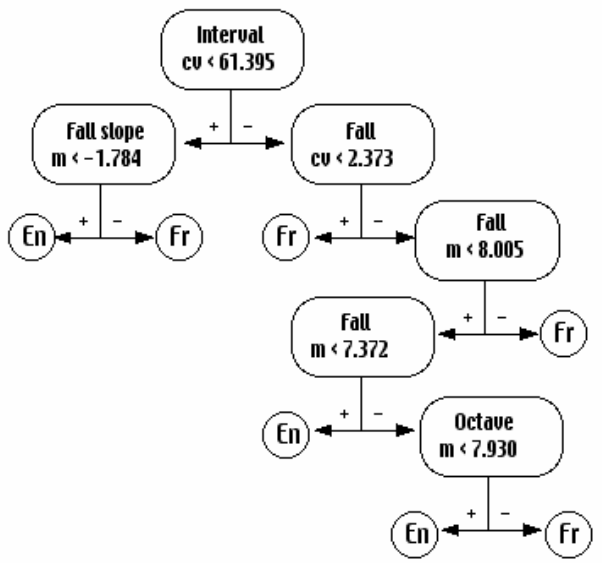


*Figure 3.* Regression tree analysis from the Cruise algorithm.

## 4. Conclusion

The statistical analysis of the F0 targets obtained from recordings of continuous passages 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.

## 5. References


[1] Abercrombie, D., "Syllable quantity and enclitics in English", in *In Honour of Daniel Jones. Papers contributed on the occasion of his eightieth birthday, 12 September 1961*., D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, and J.L.M. Trim, (eds.). 1964, Longmans, London. p. 216-222.

[2] Boersma, P. and Weenink, D., "Praat, a system for doing phonetics by computer". 1996-2003, University of Amsterdam.

[3] Campione, E. and Véronis, J. "A multilingual prosodic database". in *Proceedings ICSLP*, Sydney, 1999.

[4] Campione, E. and Véronis, J. "A statistical study of pitch target points in five languages". in *Proceedings ICSLP*, Sydney, 1999.

[5] Chan, D. *et al.* "Eurom - a spoken language resource for the EU". in *Proceedings Eurospeech '95*, Madrid, 867-870, 1995.

[6] Cummins, F., Gers, F., and Schmidhuber, J. "Language identification from prosody without explicit features". in,, 2000.

[7] Dauer, R., "Stress-timing and syllable timing reanalyzed". *Journal of Phonetics*, 11 1983. 51-62.

[8] Eriksson, A., *Aspects of Swedish speech rhythm*, Department of Linguistics. 1991, University of Göteborg, Göteborg. p. xii+234.

[9] Farinas, J. and Pellegrino, F. "Automatic rhythm modelling for language identification." *Proc. Eurospeech 2001*, 2539-2542.

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

[11] Hirst, D.J. and Di Cristo, A., "A survey of intonation systems.", in *Intonation Systems. A Survey of Twenty Languages*., D.J. Hirst and A. Di Cristo, (eds.). 1998, Cambridge University Press, Cambridge. p. 1-44.

[12] Hirst, D.J., Di Cristo, A., and Espesser, R., "Levels of representation and levels of analysis for the description of intonation systems.", in *Prosody: Theory and Experiment*, M. Horne, (ed.). 2000, Kluwer Academic Publishers, Dordrecht. p. 51-87.

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

[14] Jun, S.-A. and Fougeron, C., "A phonological model of French intonation", in *Intonation. Analysis, Modelling and Technology*., A. Botinis, (ed.). 2000, Kluwer Academic Publishers, Dordrecht. p. 209-242.

[15] Kim, H. and Loh, W.-Y., "Classification trees with unbiased multiway splits,". *Journal of the American Statistical Association*, 96 2001.

[16] Kim, H. and Loh, W.-Y., "CRUISE User Manual". 1998, Department of Statistics, University of Wisconsin, Madison.

[17] Ladefoged, P., *A course in phonetics.*, Harcourt, Brace, Jovanovich, New York, 1975.

[18] Low, E.L. and Grabe, E., "Quantitative characterisations of speech rhythm. Syllable timing in Singapore English.". *Language and Speech*, in press.

[19] Pike, K.L., *The intonation of American English*., The University of Michigan Press, Ann Arbor, 1945.

[20] Ramus, F., Nespor, M., and Mehler, J., "Correlates of linguistic rhythm in the speech signal". Cognition, 73(3), 1999. 265-292.

[21] Roach, P., "On the distinction between "stress-timed" and "syllable-timed" languages.", in *Linguistic controversies*., D. Crystal, (ed.). 1982, Edward Arnold, London.

[22] Roach, P., "Some language are spoken more quickly than others.", in *Language Myths,* Bauer and Trudgill, (eds.). 1998, Penguin, Harmondsworth.

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

[24] Véronis, J., Hirst, D.J., Espesser, R., and Ide, N. "NL and speech in the MULTEXT project.". in *Proceedings AAAI-94 Workshop on the Integration of Speech and Natural Language Processing*., Seattle, 1994.

[25] Wenk, B.J. and Wioland, F., "Is French really syllable timed?". *Journal of Phonetics*, 10 1982. 193-216.