ISCA Archive
http://www.isca-speech.org/archive

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

A PERCEPTUAL STUDY OF FRENCH INTONATION

F. Beaugendre†, C. d'Alessandro†, A. Lacheret-Dujour†‡ and J. Terken§

†LIMSI-CNRS, BP 133, F91403 Orsay Cedex, France
‡Department of French Linguistics, University of Caen, Place de la Paix F14000 Caen, France
§Institute for Perception Research/ IPO, P.O. Box 513, NL5600 MB Eindhoven, The Netherlands

Abstract

This paper presents an experimental phonetic study of French intonation. It aims at designing a melodic model for a French Text-To-Speech synthesis system. The methodology of intonation research adopted here is in the tradition of a series of studies, for several European languages, developed at the Institute for Perception Research (IPO, Eindhoven, NL). This work is based on a perceptual description of French intonation, where pitch contours are described in terms of simple pitch movements. The problems addressed in the paper are the definition, standardization and syntactic properties of such movements, starting from a corpus of read text. The melodic model for French was defined in 3 stages: 1. F0-contours stylization; 2. pitch movements standardization; 3. study of a grammar of French intonation. At each stage, perceptual tests were performed in order to check the perceptual quality (equality or acceptability) of the synthetic F0 contours. This work was within the framework of the CEC-ESPRIT-POLYGLOT-2104 project.

1 Introduction

In this paper we describe the development of a melodic model of French intonation. Our model is designed for text-to-speech synthesis. We adopted the perceptual approach initially proposed at the Institute for Perception Research (IPO, Eindhoven, NL) [5]. This methodology has already been applied successfully to several European languages such as Dutch, German, British English and Russian [3]. Nevertheless, no previous studies were reported for French, as far as we know.

The general aim of this work is to describe the melodic properties of the French language, at least for a particular speaker, a particular speaking style and a given pragmatic situation. This description is based on perceptual analyses of acoustic data. Ideally, we shall retain only those acoustic features of actual intonation contours that are perceptually relevant. At a first stage, natural F0 contours were replaced by so-called close-copy stylizations, i.e. a concatenation of simple F0 movements (straight lines in a logarithmic frequency/linear time representation). Close-copy stylizations were proved almost indistinguishable from natural F0 contours, using a first series of perceptual tests. At a second stage, the large amount of F0 movements obtained was analysed and reduced to a small set of 9 standard movements. Standardized F0 contours were proved perceptually acceptable, using a second series of perceptual tests. The third stage was devoted to analyses of the syntactic relationships between standard movements: the aim was to discover the rules governing concatenations of standard F0 movements.

Such a bottom-up methodology (starting at acoustic level up to prosodic level) provides a descriptive basis for intonation synthesis. Nevertheless, the linguistic analyses needed, and the relationships between the linguistic structure, the prosodic structure and F0 contours will not be discussed herein, as the scope of our present work is limited to a perceptual acoustico-phonetic study.

The paper is organized as follows. Section 2 presents the speech material recorded, the stylization methodology, and its perceptual evaluation. Section 3 discusses of pitch movements standardization and classification, and proposes a perceptual assessment of standardization. Section 4 presents the current development of a grammar describing standard movements concatenation. Section 5 concludes.

2 F0 contours stylization

2.1 Speech materials

This study is based on a speech data base containing read sentences. A corpus of 220 sentences was designed, taking into account some syntactic, phonotactic and lexical constraints:

Syntactic constraints: sentence modality (assertive, interrogative, negative or imperative), sentence structure (simple or complex);

Phonotactic constraints: words position in the sentence, number of syllables of sentences and words;

Lexical constraints words structures (simple, derived or composed).

The corpus was read by 3 Parisian speakers (1 female, 2 male). Sentences were pronounced carefully and special attention was paid to the intonation. Nevertheless, the prosody obtained can be qualified of "neutral" (no emotional or emphatic stress for instance). Speech signals were digitally recorded in a soundproof booth (sampling frequency 16 kHz, B&K4165 microphone).

2.2 Stylization procedure

Natural F0 contours are quite complex and this raw materials contain a lot of informations which are difficult to interpret. F0 stylization is based on the hypothesis that unnecessary details of the F0 contours can be ruled out, without any perceptual change. It must be emphasized that no structural assumption is made so far on the nature of such details. For instance some of the pitch's variations related to micromelody (defined here as segmental influences on melody) may be ruled out, if they are not perceived, but other details may be preserved. The only aim of F0 stylization is to obtain reduced contours which must be proved perceptually identical to the original ones. In the process of F0 stylization, natural F0 contours are reduced to a concatenation of straight lines according to a logarithmic frequency scale (semi-tones/second).

It is clear that the data reduction performed provides a better basis for further analyses of F0 contours. A comparison between straight line stylization and other types of interpolations based on melodic target values is reported in [6]. It appeared that these methods were perceptually equivalent. In particular the angular points created at transitions between two straight lines do not have any special effect on the resulting melody, compared to smooth transitions.

The stylization method is an analysis-through-resynthesis technique. The process of stylization is a loop containing three steps:

1. a piece of F0 contour is replaced by a straight line;
2. a new signal is synthesized;
3. this new signal is compared to the original one;

As a general principle, a minimum number of straight lines are used for a given contour. It must be emphasized that this process is not strongly dependant on the person who is performing the stylization.

The analysis/resynthesis technique chosen in this experiment was the classical Linear Predictive Coding (LPC). We developed a software environment allowing for F0 contours stylization [1], using graphic facilities, and close to real time processing.

A first set of 59 sentences have been manually stylized. This process is unfortunately highly time-consuming, but up to now it seemed that the attempts towards automatic close-copy stylization always failed [2].

2.3 Perceptual evaluation of close-copy stylization

Close-copy stylizations must be identical to natural LPC resynthesized original sentences. Perceptual tests were set up in order to check this assumption. For this experiment, 28 sentences were chosen. All these sentences where short enough, ranging from 2 words up to 8 words.

For each sentence, the stimuli were pairs of examples distributed in 4 categories :

category A: two times the resynthesized original;

category B: resynthesized original & close-copy stylization;

category C: close-copy stylization & first alternative version;

category D: close-copy stylization & second alternative version;

In categories A and B, we used LPC resynthesized original sentences instead of original sentences, because of the quality degradation introduced by LPC: the difference of quality between an original sentence and the corresponding LPC stylized version would have been easier to detect than a difference of intonation.

In C and D categories, the alternative versions were based on stylized contours because it was easier to change the slope of a movement or a melodic level on these contours.

The first alternative version contained only light modifications compared to the stylized version. The second alternative version contained more severe modifications. Three types of modifications were chosen: 1. modification of the slope of a straight line movement; 2. modification of the frequency level of a movement; 3. modification of timing (the beginning of a movement is shifted of one or two phonemes). For modifications of types 1 and 2, we referred to perceptual experiments ob the differential thresholds of pitch and pitch changes [7], [4]. The modifications of slopes and levels chosen for category C and D were of the order 1.5 and 3 times the thresholds respectively.

Table 1. summarizes the modifications values for alternative versions.

Table 1: *Standard values to build the alternative contours*

Cat	*Slope (DT)*				*Level (%)*		*Timing*
	Threshold		*Test ref.*		*Thres.*	*Test ref.*	*shift in Nb*
	Rise	*Fall*	*Rise*	*Fall*			*of phonemes*
C	1	1.3	1.5	1.95	5	7.5	1
D	1	1.3	3	3.9	5	15	2

Each pair was played three times at random. The question asked was :*"Are the two melodies of a pair identical?"*. Each pair was presented only one time, and the subjects had to make a forced choice (Yes/No). The signal was presented through high quality headphones (beyer dynamic DT48) at a comfortable sound pressure level.

We got responses from 20 untrained subjects. Table 2 summarizes the results obtained. For each subject, the percentage of answer "Yes" (two identical sentences) were counted. The table shows the mean and standard deviation of these percentages among subjects.

Table 2: *Results of the tests of perceptual equality of close-copy stylization. Cat. : category, NR: number of responses, NY: number of responses "Yes" (all subjects), % Mean NY: mean percentage of responses "Yes" among subjects, NN: number of responses "No", % Mean NN: mean percentage of responses "No" among subjects, % SD: standard deviation of percentages of responses "Yes".*

Cat.	NR	NY	% Mean NY	NN	% Mean NN	% SD
A	1428	1280	89.64	148	10.36	8.92
B	1428	1215	85.08	213	14.92	8.37
C	1428	698	48.88	730	51.12	15.29
D	1428	233	16.32	1195	83.68	10.78

About 90 % of pairs in category A were perceived identical (100 % were identical), and about 85 % of pairs in category B were also perceived identical. The scores were only about 49 % and 17 % in categories C and D. These results are clear enough to confirm the hypothesis that close-copy stylizations are almost indistinguishable from original resynthesized sentences.

3 Pitch movements standardization

3.1 Standardization

At the first stage, natural F0 contours were simplified and replaced by concatenations of straight lines, without noticeable perceptual losses. The second stage concerns F0 movements standardization. The aim is again to perform a data reduction, because of the large amount of different F0 movements obtained at the stylization stage. Standardization in meaning herein classification of the F0 movements using simple acoustic and perceptual criteria. Clearly, F0 contours made of standardized movements are not perceptually identical to stylized contours, but they must be perceptually equivalent. They must have the same prosodic contents (same stresses, same prosodic words, same prosodic patterns etc...) even if they are distinguishable.

Two standard movements are differing at least for one acoustic parameter. The parameters taken into account for classification were :

1. direction of the movement (Rise (R), Fall (F))
2. rate of change (fast or gradual)
3. frequency range of the movement
4. movement duration in number of syllables.
5. timing according to the syllabic boundaries

Acoustic classification and perceptual tests reduced the number of standard movements to 9 movements (5 rises, 3 falls and a flat linking movement). Table 3 summarizes this set of standard movements.

Table 3: *Acoustic parameters of the different classes of movements*

	R1	R2	R3	R4	R5	F1	F2	F3	D
Direction									
rise	+	+	+	+	+				
fall						+	+	+	
Timing									
attack		+			+				
att-nucleus	+		+				+		
pause								+	
Slope									
steep	+	+	+		+		+	+	
gradual				+		+			
flat									+
Frequency range									
full	+	+	+			+	+	+	
half				+	+				

Contrary to other studies [5], the movements are not defined using an absolute size (in ms), and an absolute frequency range, but using the syllabic structure. Syllabic structure means the number of syllables and their composition: attack (at least one consonant) and nucleus (a vowel).

As this study aims at generating F0 contours in a text to speech system, it was necessary to control the actual frequency range of pitch movements. A base line and a top line are computed according to the length of each sentence (see figure 1). These two lines are subjects to declination, with a stronger declination for the top line. This size reduction of the movements towards the end of the sentences was observed also for other languages.

3.2 Perceptual assessment of standardization

The acceptability of standardized contours must be assessed using perceptual tests. These tests are of another nature than the previous ones. The aim is no more to check that standardized sentences are perceptually identical to the natural ones (and they are not!), but that they are perceptually acceptable.

For this experiment 15 sentences were selected, ranging from 6 up to 14 words. For each sentence, four versions were synthesized:

category A: resynthesized original sentence;

category B: standardized sentence;

category C: modified sentence (one more pitch movement, on the first syllable of all lexical words);

category D: modified sentence (one more pitch movement, on the last syllable of all lexical words);

Subjects were asked to rank the natural quality of these sentences, played at random, on a 5 points scale (from very bad to excellent). Responses of 9 subjects are summarized in table 4.

The results are less clear than those obtained for the previous test. Nevertheless, the responses of all the subjects are distributed in two broad categories: on the one hand category A and B (> 3), and on the other hand categories C and D (< 3). These results indicate that subjects were ranking standardized versions closer to natural sentences than to modified sentences.

Table 4: *Perceptual acceptability of standardized contours. AS: all subjects mean. Five points scale (1 .. 5).*

Subjects	A	B	C	D
MG	3.00	2.97	2.23	2.37
ALD	3.67	3.23	2.37	2.63
RS	4.57	4.13	3.40	3.67
CDA	3.87	3.27	2.37	1.70
MJ	3.20	3.10	2.17	2.10
GR	3.73	3.67	2.87	2.93
PHE	3.80	3.67	2.43	2.67
JCM	2.87	2.80	1.93	2.57
SF	3.30	3.53	2.20	2.10
AS	3.55	3.37	2.44	2.52

4 Syntactic constraints on pitch movements

The third stage was, regardless of the linguistic structure, to study the syntactic constraints imposed on the pitch movements. Ideally, we shall define a grammar which generates all and only the possible sequences of pitch movements according to our speech materials. To reach this general goal, it is necessary to introduce some functional assumptions that were not used until now, and are out of the scope of this paper. We shall report only results on co-occurence of movements, without any discussion on the functional aspects of pitch movements concatenation.

A syntactic distribution analysis of the pitch movements was achieved on 59 standardized sentences (about 600 standard movements). This analysis reveals some constraints on the concatenation of movements, as some of the movements never combine with others. Frequencies of co-occurrences can be associated to the pitch movements, taking into account the left and the right context. These results are summarized in table 5.

In order to improve this grammar, it will be necessary to take into account other factors such as the syntactic structure of the sentence, phonotactic constraints, etc.

Figure 1: *From bottom to top: signal waveform, phonemic segmentation, original and stylized pitch contour, movements references, standardized pitch contour.*

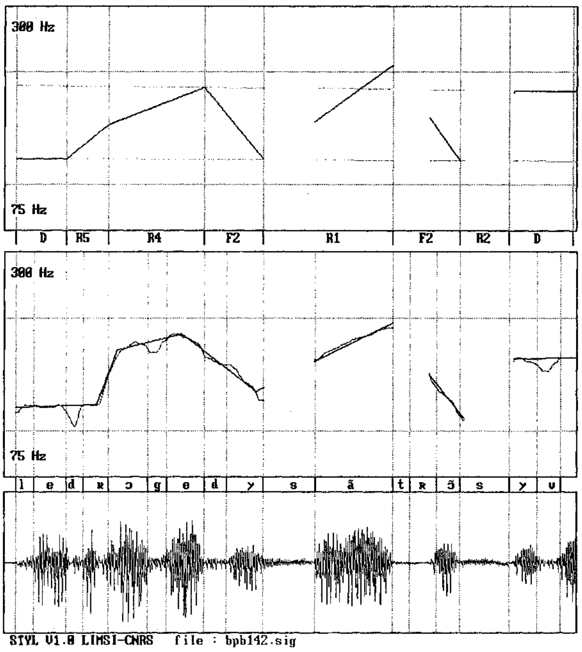

Table 5: *Percentages of Left and Right contexts for each standard movement. LC: left context, RC: right context, SB: sentence boundaries.*

		R1	R2	R3	R4	R5	F1	F2	F3	D	SB
R1	LC	0	0	0	0	0	41	12	0	47	0
	RC	0	0	0	0	0	29	59	12	0	0
R2	LC	0	0	0	0	0	26	28	0	40	6
	RC	0	0	0	0	0	79	0	0	21	0
R3	LC	0	0	0	0	0	14	25	0	45	16
	RC	0	0	0	0	0	57	41	0	2	0
R4	LC	0	0	0	0	100	0	0	0	0	0
	RC	0	0	0	0	0	62	38	0	0	0
R5	LC	0	0	0	0	0	15	9	3	64	9
	RC	0	0	0	100	0	0	0	0	0	0
F1	LC	11	39	27	23	0	0	0	0	0	0
	RC	14	12	6	0	5	0	0	0	19	44
F2	LC	33	0	30	21	0	0	0	0	16	0
	RC	5	24	18	0	4	0	0	0	42	7
F3	LC	100	0	0	0	0	0	0	0	0	0
	RC	0	0	0	0	25	0	0	0	75	0
D	LC	0	11	1	0	0	19	26	3	0	40
	RC	18	21	22	0	24	2	11	0	0	2

5 Conclusion

We presented the development of a melodic model for French which is based on acoustic and perceptual analyses. The main results are:

1. F0 stylization was proved successful in representing natural F0 contours without perceptual degradation.
2. only 9 elementary pitch movements appeared sufficient to represent our speech data, with satisfactory perceptual quality.
3. standard pitch movements obey syntactic constraints, that where analysed here for right and left contexts.

The next steps of our work will be: 1. the development of a more complete grammar, taking into account larger contexts (clauses etc...); 2. the link between this phonetic description and phonotactic, prosodic and syntactic descriptions of the same data; 3. the developement of rules for generating automatically pitch movements from text. This work was within the framework of the CEC-ESPRIT-POLYGLOT-2104 project.

References

[1] F. Beaugendre. *A program for Pitch contour Stylization.*, Notes et documents LIMSI, NDL 92-5, 1992.

[2] L. Boves., and T. Rietveld. "Stylization of pitch contours." *Proceedings of the Institute of Phonetics of the Catholic University of Nijmegen*, pp. 34-40., 1978.

[3] R. Collier. "Multi-language intonation synthesis." *Journal of Phonetics*, vol. 19, pp. 61-73., 1991.

[4] J. 't Hart. "Differential sensitivity to pitch distance, particularly in speech." *J. Acoust. Soc. Am.*, vol. 69, No. 3, pp. 811-821, 1981.

[5] J. 't Hart., R. Collier. and A. Cohen. *A perceptual study of intonation : an experimental-phonetic approach to speech melody.* Cambridge University Press, 1991.

[6] J. 't Hart. "F0 stylization in speech : Straight lines versus parabolas." *J. Acoust. Soc. Am.*, vol. 90, No. 6, pp. 3368-3370, 1991.

[7] A.V. Issachenko., H.-J. Schädlich. *A model of standard German intonation.* The Hague/Paris: Mouton, 1970.