

# EXPERIMENTAL STUDY OF THE EXPRESSION OF EMOTIONS AND ATTITUDES IN FOUR LANGUAGES

Olivier Piot

University of Paris III, Paris, FRANCE

## ABSTRACT

This paper is a study of the prosodical means people with different cultural backgrounds use in order to express emotions, or convey attitudes in speech. An experiment was carried out in four languages (American English, Korean, Finnish, Vietnamese), in which a recording of a common corpus by two men and two women was made. It included two sentences, one of them being one word length (the first name «Natacha», or a translation of it, in a vocative context), the other one including a displacement of the speaker («OK, now, I'm going to the station to get that packet»). Thirteen attitudes were chosen, on the basis of our thesis work on the expression of attitudes in speech. The parameters measured were: amplitude, mean and final contour for  $F_0$  and intensity, as well as mean syllable duration.

## 1. INTRODUCTION

The study of the expression of emotions in speech has become for a few years a major topic of research in phonetics. This recent interest comes in part from the fields of speech recognition and synthesis, but the comprehension of the phenomena involved remains a challenge for psychological theories [7]. Moreover, taking attitudes (taken here as including the expression of emotions) into consideration raises some other difficulties, linked for example to prosodic patterns or cultural constraints [2, 5]. Emotions are then embodied with the issues of individual choice and linguistic norm, which must be taken into account as it is the general case. In this experiment, described in section 2, the expression of thirteen attitudes in four different languages was recorded. These attitudes were chosen in order to confront theoretical hypothesis and claims that can be found on this subject area.

## 2. EXPERIMENTAL PROCEDURE

### 2.1. Description of the corpus

The attitudes tested included:

- Threatening (1) and domination relationship (12) (13) [4];
- Big (2) (3) (4) and small (5) (8) (9) (10) arousal [7];
- Basic emotions (2) (7) (8) (9) (10);
- A contradiction between pragmatico-semantic content and attitude (6) (11);

The common corpus we used was translated in each language from French. It included two sentences, one of them being one word length (the first name «Natacha», in a vocative context), the other one involving a displacement of the speaker («OK, now I'm going to the station to get that packet»). We chose this latter one because we believed its context was likely to enhance rhythm differences, without implying any particular emotion. When «Natacha» was not an available name in a language, we asked our translator to find another one, if possible trisyllabic and making use of no other vowel than /a/ (so that intrinsic pitch phenomena could be neglected). The second sentence was simply translated, resulting in a 10 to 16 syllables version. Each attitude was described by a single sentence, so

that the task was easy to understand and unambiguous (see table 1. below).

Nbr.	Sentence	Attitude
1	"Natacha"	To your child who has just disobeyed you, warning before acting severely
2	"Natacha"	To your child who has just disobeyed you for the fifth time, reaction
3	"OK, now I'm ..."	To a friend, you have to go fast (but you would prefer to stay)
4	"OK, now I'm ..."	To a friend, you have to go fast (very valuable packet)
5	"OK, now I'm ..."	To a friend, you've got time (but you would prefer to stay)
6	"OK, now I'm ..."	To a friend, you've got time (very valuable packet)
7	"Natacha"	To your child who has just done something well (and you're used to it)
8	"Natacha"	To your child who has just disobeyed you for the 80 <sup>th</sup> time, resignation
9	"OK, now I'm ..."	To a friend, after a deception, "sad"
10	"OK, now I'm ..."	To a friend, after having been bored all day long, "boredom"
11	"OK, now I'm ..."	To a friend, but thinking about something else
12	"Natacha"	Calling of a colleague who is very much superior to you on the hierarchy level
13	"Natacha"	Calling of a colleague who is very much inferior to you on the hierarchy level

Table 1. English text of the common corpus.

### 2.2. A brief linguistic presentation

Following the introduction it is useful to summarise briefly the prosodic constraints of the four languages under study here. American English is a "stress" language, the accented syllable being mostly the first or second within the word (intensity and duration increase): this accent is "free" according to the terminology of Paul Garde [3]. Finnish has a "fixed" accent on the first syllable of the word (intensity and duration increase) which is independent of vowel quantity, and a lexico-syntactic marking of several modalities (e. g. interrogation). Vietnamese is a tone language, its standard dialect (the one recorded here) has six tones. It also has a lexico-syntactic marking of several modalities and attitudes (for instance it distinguishes between standard and surprised interrogation), therefore we may expect less variation for this language.

### 2.3. Recording procedure

For each language, we recorded two women and two men of similar ages from the same speech community. They were given

the text of the corpus a few days in advance, so that they could prepare for the recording. This took place in an anechoic room, with high quality recording tools. The experimentalist was also sitting in the room but, as he kept watching the recording level in order to make sure that there was no saturation, the subject could not be disturbed by any interactional process. We made a pause in the recording whenever the subject asked for it.

### 3. RESULTS AND DISCUSSION

#### 3.1. Sentence 1

For the analysis we used an analogic device called "prozodik", which performs an analogic calculation of both  $F_0$  and intensity, drawn together with the speech signal in real time. After setting its parameters in a way adapted to each speaker, it provides measurements of high accuracy. We then calculated, for each performance, mean syllabic duration  $D$  (ms), mean value  $M$  and amplitude  $A$  of both  $F_0$  (in Quarter Tones) and intensity ( $I$ , in dB). Mean values as well as amplitudes were calculated over all voiced portions for  $F_0$ , and over all syllable peak values for intensity.  $F_0$  and intensity patterns were considered whenever it appeared to be useful to the discussion (see below).

First we compared the 16 performances of (12) and (13), and found significative differences for  $D$ , resp. 213 and 183 ( $p < 0,05$ ), and  $MF_0$ , +3.69 for (12) ( $p < 0.01$ ). If we view the speaker as self-centred,  $MF_0$  difference is in accordance with Ohala's "frequency code" as well as with our theoretical work [4, 5]: lower  $F_0$  is used to convey higher domination. The difference in rhythm is also interesting, and may be explained by the avoidance of too long a process (and state) of sympathy with Natacha in (13). Each language showed the same tendencies.

The comparison of (1) and (2) showed no significant differences, which also is quite in accordance with our predictions. What made the difference was the  $F_0$  pattern on the last syllable, which was recto tono or increasing for (2), and decreasing for (1). This was surprisingly similar for each performance in each language; it may be explained in our theoretical framework as a warning, with sympathy in (1), but with no more (and avoidance of) sympathy in (2).

(7) and (8) show a quite similar rising-falling pattern, even if in (7) it tends to spread over the whole word, while it is often mono- or disyllabic in (8). But here again statistic parameters make the difference:  $D$  amounts to 230 and 338 resp. ( $p < 0.001$ ), which we believe is due to the difference in motivation between the two attitudes: satisfaction constitutes a temporary source of motivation in our theory, while resignation is believed to consist in a withdrawal of involvement, hence of motivation.

#### 3.2. Sentence 2

Between (3) and (4) as well as between (5) and (6), we expected to find greater  $AF_0$  and  $AI$  for the latter than for the former, but (with  $p=0.071$ , and  $p=0.222$  resp.) this tendency, although observed on the mean values, didn't reach significance.

Between (3) and (5) as well as between (4) and (6), we expected to find greater  $D$  for the latter than for the former, which was significant ( $p < 0.05$ ) only in the first case. It must be noticed that  $D$  was significantly shorter for (4) than for (5) ( $p=0.0001$ ), which suggests that some of the extra motivation

implied by the high value of the packet was invested in speech rate, which then explains why neither  $AF_0$  nor  $AI$  differences reached significance. It seems also worthy to mention that  $MI$  difference between (4) and (6) nearly reached significance ( $p=0.06$ ), which suggests that extra motivation expected to invest in speech rate somewhat deviated towards phonatory tension.

In order to evaluate over which parameters the decrease in arousal took place for performances (9) to (11), we compared them to (6) which we believed was the nearest to the most typical, or "standard" situation. Actually no significant difference was found between these last three contours. Mean syllable duration  $D$  was found to be significantly greater for (9) (10) and (11) ( $p < 0.01$ ,  $p < 0.05$ ,  $p < 0.05$  resp.). This tendency is already well known, as well as the fact that sadness has the slowest speech rate of all commonly studied emotions [7]. Sadness was expressed also through a decrease of  $AF_0$ , and boredom by a decrease of  $MI$ . What we suggest in the first case is a decrease of motivation, caused by the vanishing of ways towards satisfaction [6]. This explanation also predicts a decrease of  $MF_0$ , which was observed although not significant ( $p=0.11$ ). If we were to view boredom as involving mainly a decrease of focalization without any decrease of motivation, because of the rarity of situations interesting for one's satisfaction, then this phenomenon could be explained [6], but this hypothesis remains to be justified. The presence of both phenomena in (11) would then be explained by the co-occurrence of these two factors, which doesn't seem unreasonable.

### 4. CONCLUSION

Predictions coming from studies on the expression of emotion in speech, as well as our own theoretical work on the expression of attitudes by prosody, are quite satisfyingly confirmed by this experimental work. Despite their very different prosodic constraints, these four languages showed remarkable similarities in the expression of both emotions and attitudes, which we believe deserves further investigation. Further experimental as well as theoretical work is therefore needed.

### ACKNOWLEDGMENTS

We thank the informants who gracefully translated this corpus from French, and contacted native speakers for our experiment.

### REFERENCES

- [1] Fonagy, I. and Berard, E. 1972. Il est huit heures: contribution a l'analyse semantique de la vive voix. *Phonetica*, 26: 157-192.
- [2] Fonagy, I., Berard, E., Fonagy, J. 1983. Cliches melodiques, *Folia Linguistica* 17: 153-185.
- [3] Garde, P. *L'accent*, PUF, coll. SUP "Le Linguiste".
- [4] Ohala, J., J. 1984. An ethological Perspective on common cross-language utilization of  $F_0$  of voice, *Phonetica* 41: 1-16.
- [5] Piot, O. 1997. Une approche morphogenetique des "cliches melodiques" du francais standard, *Proceedings of the XVI<sup>th</sup> International Congress of Linguists*, Paris.
- [6] Piot, O. 1998. Vers la prediction des formes prosodiques, et de leur role communicatif en francais standard, *Proceedings from the congress Orage98, Besancon*.
- [7] Scherer, K., R. 1986. Vocal affect expression: a review and a model for future research, *Psychol. Bulletin*, 99: 141-165 (1986).