

IOULIA GRICHKOVTSOVA / MICHEL MOREL / ANNE LACHERET

Perception of Affective Prosody in Natural and Synthesized Speech: Which Methodological Approach?

1. Introduction

Research on affective speech (cf. Johnstone/Scherer 2000; Murray/Arnott 1993) underlines the plural character of prosodic features involved in the expression of vocal emotions and attitudes. These prosodic features include, among others, voice quality, fundamental frequency, speech rate and intensity. Thus, speech prosody has a double function: linguistic and paralinguistic.

According to Ladd (1996), prosody realizes its linguistic and paralinguistic functions through two parallel channels tightly coordinated in time and used simultaneously in the interpretation of utterances with no effect on each other. Paralinguistic features are seen as modifications added to the linguistic features. Nevertheless, the question remains if these two types of prosody are separable for individual analysis of their roles and specific characteristics.

Even if studies into affective speech production have shown differences between emotions in the usage of prosodic parameters, difficulty is observed in determining a significant perceptive value of these parameters in the identification of affective states. Researchers come to an agreement on the importance of voice quality in the production and perception of emotions (d'Alessandro 2006). However, the definition of the perceptive role of other prosodic parameters remains a complex task. These prosodic parameters include pitch, rhythm, loudness, speech rate, and constitute the prosodic contour of an utterance. If voice quality generally realizes paralinguistic functions, prosodic contour plays two roles – linguistic

and paralinguistic. The difficulty to analyze these two functions individually explains the importance of methodological choices in the corpus development of affective speech.

Two main types of corpus are usually used for perceptive analysis of affective speech: utterances with or without lexical content. Perceptive experiment on the basis of utterances with lexical meaning does not only establish the significant role of fundamental frequency in the identification of affective states, but also underlines cross-linguistic differences in the association of fundamental frequency with affective states in English and Dutch (Chen 2004). Studies on the basis of non-lexical utterances used an invented combination of syllables existing in the language (Bänziger/Scherer 2005) or an utterance in a foreign language unknown for the speakers and listeners (Yanushevskaya *et al.* 2006, Gobl/Ní Chasaide 2003). They aimed to neutralize lexical and syntactic constraints during the course of expression and identification of vocal affective states, as well as to facilitate cross-linguistic comparisons. These studies show the importance of the prosodic contour for certain affective states, and cross-linguistic differences for some associations between voice quality, fundamental frequency and affective states. Nevertheless, the absence of lexical meaning in the utterances used can engender bias in the analysis of the role played by intonation. Indeed, utterances devoid of lexical meaning do not permit chunking into syntactic and pragmatic units. We suggest that without distinction between words (function word/content word), without syntactic and information boundaries, gating points of prosodic contours encoding the expression of emotions are affected, and the role of intonation is minimized.

In order to study the role of prosody in the identification of affective states, a series of perception experiments were performed on the basis of two different corpora of affective speech. The first two experiments used a corpus of utterances with affective lexical meaning encoded by two French actors. The third experiment used a multi-speaker corpus of French neutral utterances. This report is organized as follows: the three experiments are described in the next two sections – 1) corpus with affective lexical meaning, 2) multi-speaker corpus with affectively neutral meaning – where research objectives and results are presented. In particular, our methodology

for the corpus design and for the development of perception tests is discussed. Then, follow a general discussion and conclusive remarks.

2. Perception experiments based on the corpus with affective lexical meaning

A corpus of affective speech encoded by two French actors (a male and a female) was recorded in order to address our research questions concerning the role of prosody in the identification of affective states. This corpus contains six affective states: three emotions (anger, sadness, happiness) and three attitudes (obviousness, doubt, irony). We define emotion according to Scherer (2003) as a brief episode of synchronised response of all or most organismic subsystems in response to an external or internal event of major significance. Attitude is defined as affectively coloured stances, preferences or predispositions activated during an interpersonal interaction.

Utterances recorded for the corpus possess affectively charged lexical meaning, identifiable as such. In practice, even if the beginning of each utterance can be associated with different affective states, in the course of its production, little room is gradually left for interpretative manoeuvre. This approach allows to study utterances that are close to natural affective speech, as the expressed affect is in accordance with the lexical meaning. Examples of utterances from the corpus with affectively charged lexical meaning are given in Appendix 1.

Two perception experiments were realized with this corpus. The first experiment aimed to determine if affective states can be identified before the end of the utterances used for their expression. Our hypothesis was that emotions can be identified earlier in the flow of speech than attitudes, as voice quality specific to the expressed emotion is present from the beginning of the utterance. The second experiment was developed to compare the perceptive value of voice quality and prosodic contour for the studied affective states. Our hypothesis was that voice quality is privileged in the emotions, and

prosodic contour in the attitudes. The methodology of the experiments and the obtained results are described in the two following subsections.

2.1. Experiment 1: gating paradigm

In the framework of experiment 1, the possibility to identify affective states before the end of the utterance was tested. According to our hypothesis expressed above, listeners may identify emotions earlier, in the beginning or at least clearly before the end of the utterance. Identification of attitudes requires the analysis of lexical, syntactic and prosodic structures of the utterance, and thus it may be realized later than for emotions.

The gating paradigm was chosen to test the proposed hypothesis. This paradigm allows understanding of how much phonetic information is needed for the optimal identification of affective states. This paradigm is based on the hypothesis that a gating point corresponds to the part of the utterance that triggers the identification of the encoded affective state. In order to localize the gating points in the utterance, our methodology was as follows: each stimulus was presented in segments increased progressively by time steps of 200 ms. Listeners were asked to identify the affective state at the end of each segment. The duration-blocked presentation format was chosen: first, all the stimuli of the particular segment size were presented to the listeners in a random order, then all the stimuli of the following segment size, and so on.

Twenty-four utterances were taken from the recorded corpus: two utterances for each studied affective state (anger, happiness, sadness, obviousness, irony and doubt) pronounced by an actress and two utterances by an actor. Thirteen subjects participated in the test. They were all native French speakers (average age – 31 years old). The Perceval software (Ghio *et al.* 2007), specially developed for psycholinguistic perception tests, was used. The experiment was run for about 40 minutes on a computer in a quiet laboratory room.

Three variables were selected for the analysis of correct responses: identification point, isolation point and final level of identification. Identification point refers to the gate of the stimulus

where correct identification reaches at least 50%. Isolation point is the gate of the stimulus where the highest identification is achieved and maintained without any change in response thereafter. Finally, a general level of identification was calculated on the whole response data in percentage. Results are reported in Table 1.

<i>Affective state</i>	<i>Identification point (ms)</i>	<i>Isolation point (ms)</i>	<i>Identification level (%)</i>
ANGER	200/400	800/1600	74.9
SADNESS	400/1000	800/1600	70.1
OBVIOUSNESS	400/1200	800/complete	56.9
DOUBT	600/1400	800/complete	57.6
HAPPINESS	800/1800	1000/complete	54.1
IRONY	1600/1800	1600/complete	39.6

Table 1. Results for identification point, isolation point and the final level of identification. Two values show the observed range of gating points for each affective state across the four utterances.

The results observed for the three variables allow the differentiation of anger and sadness from the other affective states. For the identification point, we see that utterances encoded with anger are identified at the first or maximum at the second gate. Sadness too can be recognized as early as the second gate, but more variability is observed. The other affective states are identified later. Results for the isolation point also go in the same direction: angry and sad utterances are all successfully recognized before listeners hear the complete form. This is not true for the other expressive modalities.

Based on the gating paradigm, differences were shown in the identification of emotions and attitudes. Our hypothesis that emotions can be identified earlier than attitudes is confirmed for anger and sadness. Happiness follows the identification pattern observed for attitudes in our perception experiment. This observation can be explained by the specific communicative role played by happiness. Speakers may better control happiness; they may use it in interpersonal communication in the same way as attitudes. But it is also often the case with anger, which is identified early. Another explanation may be the difficulty to produce voluntarily specific

'smiling' voice quality by speakers, and thus they might privilege intonation more, like for attitudes.

2.2. Experiment 2: transplantation paradigm

Experiment 2 aimed to compare the perceptive value of voice quality and prosodic contour for six affective states: anger, happiness, sadness, obviousness, irony and doubt. More precisely, our objective was to investigate if intonation and voice quality are equally important for the perception of the studied affective states or whether one of them may be privileged. Our research hypothesis was that voice quality is privileged for marking emotions, while variation of prosodic contour is more used in the expression of attitudes.

Prosody transplantation paradigm (Garcia *et al.* 2006) was used to evaluate the corresponding perceptive roles of voice quality and prosodic contour in the identification of affective states. This method involves extraction and exchange of the prosodic contour between two utterances with the same segmental content: a natural utterance and a synthesized utterance with neutral intonation. Prosody transplantation acts on the fundamental frequency, intensity and temporal parameters, but it does not modify characteristics of voice quality (though degradation is sometimes observed).

Seventy-two utterances were chosen in the recorded corpus: six utterances for each studied affective state (anger, happiness, sadness, obviousness, irony and doubt) pronounced by a female, and six utterances pronounced by a male. The same utterances were synthesized with KALI, a French-speaking text-to-speech diphone synthesis system (Morel/Lacheret 2001). The synthesized utterances did not have any vocal affective meaning, as KALI does not perform this type of treatment. In the process of prosody transplantation, the prosody of Kali was mapped on the utterances encoded by actors, and the prosody of the actor was mapped on the utterance synthesized by Kali. Thus, four versions of each utterance were developed: version 1 – 'natural' (natural prosodic contour and voice quality encoded by the actor), version 2 – 'voice quality' (natural voice quality and prosodic contour from Kali), version 3 – 'prosody' (natural prosodic contour and voice quality from Kali), version 4 – 'lexical' (voice quality and

prosodic contour of Kali without any vocal affective meaning). In total, 288 stimuli were designed for the perception test.

Experiment 2 was run in the same manner as Experiment 1, and it took about 40 minutes. All the stimuli were randomized. Twelve native French speakers (average age – 31 years old) participated in the test. The results of the perception test, displayed in Table 2, show the level of identification for the four versions of audio stimuli in percentage. Results for ‘natural’ stimuli show the best identification of the encoded affective state, as both natural voice quality and prosody of the actor are present. Identification of version 2 ‘voice quality’ and version 3 ‘prosody’ vary with the studied affective state. Version 4 ‘lexical’ has the lowest identification level.

<i>Affective state</i>	<i>Natural (%)</i>	<i>Voice quality (%)</i>	<i>Prosody (%)</i>	<i>Lexical (%)</i>
ANGER	95	59	86	43
HAPPINESS	87	58	59	26
SADNESS	86	84	65	63
DOUBT	85	35	80	15
IRONY	74	46	61	43
OBVIOUSNESS	78	67	65	74

Table 2: Identification of affective speech stimuli. The correct responses are expressed in percentage.

The effect of the lexical meaning is present not only in the ‘lexical’ version of the stimuli, but also in all the other stimuli. Therefore, in order to compensate the influence of the lexical meaning in versions 1-3, a neutralization procedure was applied. The value of the identification level for ‘lexical’ stimuli is assigned to the role of lexical meaning in the identification of the affective states. The identification of the other versions is facilitated by the presence of specific voice quality and/or prosodic contour, which we aim to evaluate. The normalized values are calculated through the extraction of the ‘lexical’ value from the identification values received for ‘natural’, ‘voice quality’ and ‘prosody’ stimuli of the corresponding state. For example, ‘natural’ – ‘lexical’ for anger: $95\% - 43\% = 52\%$. Thus, the received value was judged as the contribution of voice

quality and/or prosody in the affective speech perception. These results are displayed in Figure 1.

The results for anger show that both prosodic contour and voice quality are used in the process of identification. Nevertheless, 'prosody' stimuli are identified more successfully than 'voice quality' stimuli, and this difference is statistically significant. Doubt results show that prosody has the main role in its perception. Obviousness is identified exclusively on the basis of lexical meaning (but maybe it was not sufficiently well realized by the actors?). Voice quality and prosody play an equally important role in the identification of happiness. Lexical meaning is apparently very important in the perception of irony, but prosody also contributes to its better identification. Voice quality is privileged in the identification of sadness.

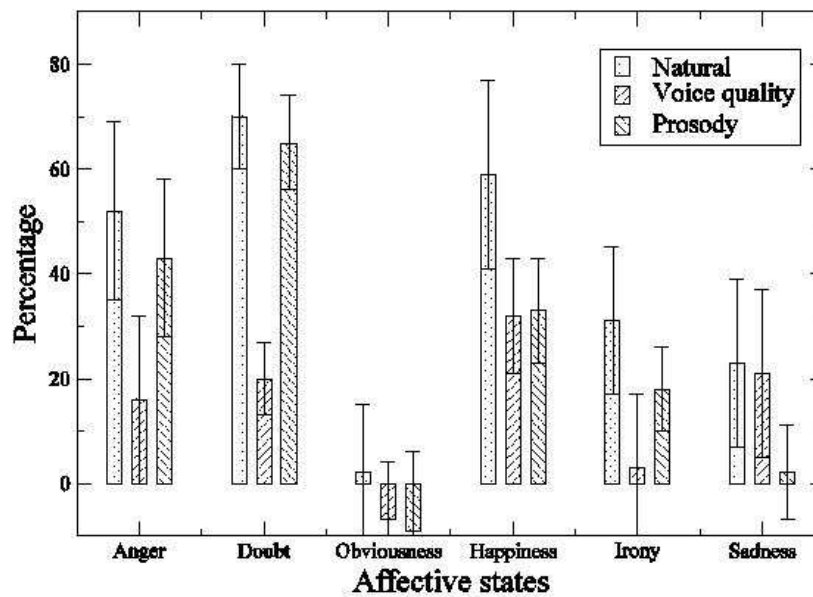


Figure 1: Neutralized values of successful identification.

This experiment investigated the perceptive value of voice quality and prosody for six affective states: anger, happiness, sadness, doubt, obviousness and irony. Previous studies (Yanushevskaya *et al.* 2006)

suggest that voice quality has stronger and more consistent associations with affective meaning than intonation. And yet, these conclusions were drawn from the perception studies with non-sense utterances. Our choice to use natural utterances was motivated by our hypothesis that listeners have difficulty in interpreting the affective meaning of prosodic features without access to chunks, meaningful components of the utterance; as a consequence, results may become biased. Our results show the importance of prosodic contour in the identification of affective states, even if its level of identification is variable. The presence of natural prosody may even be crucial for the successful identification of some affective states, like doubt.

Even if it is not possible to separate completely emotions from attitudes based on the perceptive value of prosody and voice quality, some interesting observations can be made. They show certain distinctions between attitudes and emotions. Concerning attitudes, voice quality was not identified as important in their identification; lexical meaning played an important role for irony and obviousness; prosody was involved in the identification of doubt and irony. The results for emotions show that both prosody and voice quality carry an important perceptive value. While for happiness both prosody and voice quality are equally important, anger privileges prosody. Lexical meaning is generally less important for emotions than for attitudes.

2.3. Discussion of experiments 1 and 2

These two psycholinguistic tests on the perception of attitudes and emotions demonstrated differences between the affective states studied, and the important role played by prosody in the majority of cases. Nevertheless, two important issues were identified: the first concerns the number of speakers – the second, the type of the corpus used.

Our study shows important variability in the realization of affective states by the actors and its influence on their perception. In other words, some strategies may be easier for the identification of affective states than others. In parallel with this observation, it is important to remember that the corpus used for this experiment contains only two speakers. As a result, the question stands about the

generalization of our results and the depth of understanding of the differences observed. In fact, other speakers may use other strategies for the expression of the same affective states; in this case, results of the acoustic and perceptive tests may differ significantly. At present, the level of inter- and intra-speaker variation in the realization of affective states is not known. Even if several multi-speaker speech corpora have appeared (Bänziger/Scherer 2005, Grichkovtsova 2008), the majority of the studies published in this research area use affective speech corpora with a small number of speakers (generally 1-4 subjects). This methodological weakness may explain the contradictory results observed in the studies on affective speech. This understanding of the variable character of affective speech motivated us to develop a multi-speaker corpus for sound and in-depth study in the field.

The neutralization of affective lexical meaning does not fully counterbalance its influence. Accentual prominence is generally realized on the informationally important words of the utterance, precisely those that are involved in the affective identification. By putting emphasis on these words, the prosodic contour facilitates the identification of affects, even in the absence of prosody specific to a particular affective state. It acts like a 'lexical amplifier'. Thus, there is an important risk that our results are overestimated concerning the role of prosodic contour. Our results do not distinguish the two functions realized by prosody: affective expression and lexical amplification. In comparison with experiments based on the corpora without lexical meaning, which may considerably underestimate the role of intonation, our experiments risk going to the opposite extreme. Here comes our motivation to run a new perception experiment on the basis of a corpus with neutral lexical meaning in order to evaluate the specific role of intonation in the identification of affective states without interference with prosodic amplification for affectively charged words.

3. Experiments on the basis of a multi-speaker corpus with neutral lexical meaning

The methodological issues brought up in the above section led us to develop a multi-speaker corpus on the basis of one affectively neutral utterance: *Je vais rentrer à la maison maintenant / I am going home now*. Twenty-two native French speakers (11 males and 11 females) were recorded. The total recording time of the corpus was seven hours. Six emotions (anger, sadness, happiness, disgust, fear and grief) and one neutral statement were selected for this study. Attitudes will make the subject of our future work.

A specific text was designed for each emotion investigated; the same neutral utterance was inserted in the middle of all the texts. Thus, it was possible to record the same lexically neutral utterance encoded in different emotional contexts. The underlying hypothesis was that the expression of emotions inducted from the beginning of the text with emotionally marked lexical meaning will propagate in the production of the neutral utterance. In such a way, the token utterance carries prosodic features of the emotion conveyed in the whole text, and speakers are not consciously aware of that. Each text is read three times by each speaker so that they imagine the context of the situation better and express the desired emotion in the most natural way.¹ The best realization of the token utterance was then extracted from the texts to constitute the corpus for experiment 3.

This corpus of lexically neutral utterances (154 stimuli presented in random order, test running time – 20 minutes) was first validated through an evaluation test with PERCEVAL software in order to filter out badly identified utterances. Ten French listeners (average age 30 years) participated in the test. We fixed the threshold of acceptance at 50% (at least 50% of listeners could identify the expressed emotion). This level of acceptance was chosen in order to exclude those utterances which are identified significantly above chance but which show important confusion with other affective states. We considered that a level of acceptance at 50% was necessary to increase the quality of the selected corpus. The results of this evaluation test are displayed in Table 3. Based on these results, we discarded disgust and fear, which did not give enough utterances at an

¹ Examples of the texts used are given in Appendix 2.

acceptable identification level, and we kept the four following emotions: grief, sadness, anger and happiness.

<i>Emotions</i>	<i>Successful identification (%)</i>
ANGER	73
DISGUST	9
FEAR	14
HAPPINESS	36
GRIEF	32
SADNESS	68

Table 3: Rate of utterances identified by listeners with the level of acceptance at 50%. For example, 36% of happy utterances were identified by at least 50% of listeners.

The multi-speaker corpus was validated for the development of a new perception test (experiment 3) on the basis of two methodological paradigms: gating paradigm and transplantation paradigm. The methodology was the same as that described for experiments 1 and 2. As in experiment 2, prosodic transplantation allowed to develop four versions of each utterance: version 1 – ‘natural’ (natural prosody and voice quality encoded by the actor), version 2 – ‘voice quality’ (natural voice quality and prosody from Kali), version 3 – ‘prosody’ (natural prosodic contour and voice quality from Kali), version 4 – ‘lexical’ (voice quality and prosody of Kali without any vocal affective meaning). Then, according to the gating paradigm, vocal stimuli were cut in fragments increased progressively by steps of one or two syllables until the end of the utterance:

je vais | ren | trer | à la | mai | son | main | tenant |

We did not keep the classic method of durational segmentation largely used in gating studies (segments are increased by fixed steps – 200 ms in our experiment 1) in order to get around speech rate variability depending on individual speakers and affective states.

In total, 1,560 stimuli were designed. Considering the number of stimuli, the test was based on the Latin square method,² which allows reducing the number of stimuli presented to each speaker. Eight versions of the test were prepared based on this method. To have good statistical validity, 16 subjects were recruited (native French speakers, average age 25 years, test running time – 20 minutes). The stimuli were presented in duration-blocked groups of segments through the increasing presentation format. The order of stimuli was randomized in each group. Results for the perceptive value of prosodic contour and voice quality in the affective identification are presented in Figure 2. A more detailed analysis of our results, in particular an acoustic analysis of the stimuli linked to the gating points, will be the subject of our future work.

As five categories of affective states (anger, happiness, sadness, grief and neutral statement) were used for the perception experiment, the identification level of 20% stands for ‘zero’, which represents a value purely depending on chance which will serve as a reference for the significance evaluation of results. Grief and sadness are characterized by a good identification for the version ‘voice quality’. This result is predictable, knowing that these are less controllable emotions, and therefore they modify first physiological parameters not coded in language and mainly influence voice quality. For anger, voice quality plays a less important role than prosodic contour. The importance of intonation for anger highlights a socially and linguistically controlled mechanism, comparable to the one used in attitudinal expressions. Happiness identification in its natural version is superior to the modified versions. Prosody is only just significant; voice quality does not go beyond the ‘zero’ level. The low identification level for the voice quality version is probably due to the prosodic transplantation, which alters the smiling voice in the ‘voice quality’ of version 2 and masks happiness. Even if this result shows limits in the transplantation method, it highlights the role of intonation in the identification of happiness. The identification results for the neutral statement show no difference between version 1 ‘natural’ and

² A Latin square is an $n \times n$ table filled with n different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column.

version 3 'prosody'. Voice quality participation is hardly significant. Thus, a neutral statement is well differentiated from the other emotional modalities, especially by its prosodic contour.

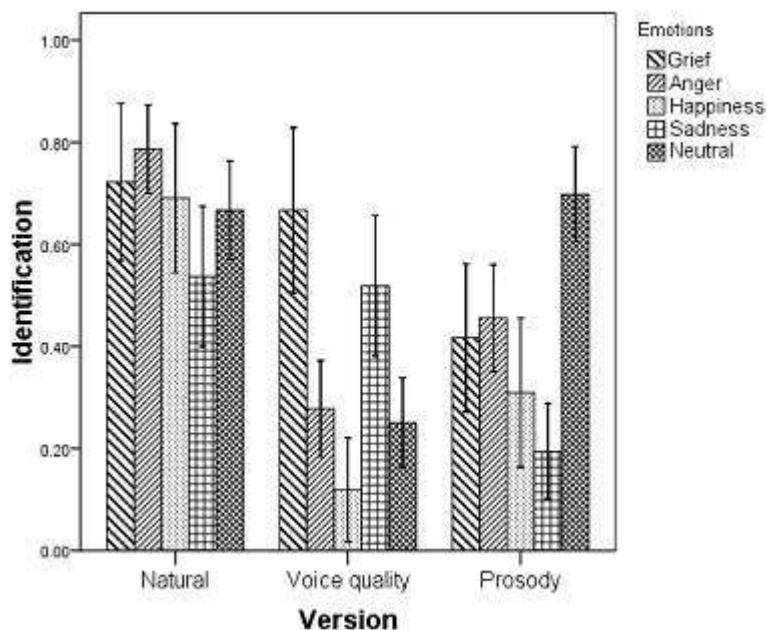


Figure 2: Identification of emotions for 'natural', 'voice quality' and 'prosody' versions (confidence intervals - 95%).

4. Discussion

Testing our main hypothesis according to which usage of non-lexical utterances can be problematic in the study of affective prosody, a series of experiments were conducted. Experiments 1 and 2, which are based on the corpus with affective lexical meaning, show the importance of intonation in the identification of affective states, even if this importance is variable. The presence of natural intonation may even be crucial for the identification of some affective states.

Nevertheless, this approach showed another bias: the complete neutralization of emotionally charged words is not possible. The usage of even 'neutral' prosodic contour still retains the amplification of affectively important words, and thus provides better affective identification on the basis of lexical meaning. Another methodological weakness is often present in experiments on the identification of emotions: insufficient number of speakers (1-4). Even if emotional speech is characterized by significant inter- and intra-speaker variability (Scherer 2003, Grichkovtsova 2008), the question stands if such a small sample of speakers covers possible prosodic strategies for affective encoding in a representative way.

Experiment 3 uses a corpus on the basis of the same utterance with purely referential meaning. This experiment allowed us to determine more reliably the perceptive value of intonation and voice quality by avoiding two biases. The first bias is in the analysis of unnatural utterances without lexical meaning, in which the absence of syntactic structure reduces considerably the perceptive value of intonation. In contrast, the second bias relies on the analysis of utterances already connotated by their verbal material. If experiments 1 and 2 show that the presence of affective lexical meaning on its own is not sufficient for optimal identification of emotions, it is difficult to quantify objectively the contribution of prosodic parameters in the process of identification.

The method used in experiment 3 tests our hypothesis concerning the double status of intonation in the language activity: 1) it amplifies informationally important words in the specific emotional context; 2) it contributes to the identification of emotion by its specific prosodic contour. The usage of a lexically neutral utterance allows to evaluate the role of prosodic contour in the identification of emotions by isolating from its amplification function. It is then possible to characterize its acoustic properties on the qualitative and quantitative levels. The results of experiment 3 concerning the role of prosodic contour in the identification of emotions show that this role is predominant in comparison with its amplification function. This observation is not trivial as it confirms strong interaction between prosodic contours and lexical content. Here comes the interest to use utterances with lexical meaning. Non-lexical utterances can only be used for experiments on voice quality. Some limits were observed for

the technical quality of prosodic transplantation through our experiments, as it can sometimes damage the voice quality of natural voice, while it does not damage the prosodic contour applied to synthesized voice. Thus, it is possible that results for voice quality are underestimated without questioning our results received for prosodic contours.

Another criticism can be made about the usage of only one lexically neutral utterance. We can reasonably suggest that other syntactically and rhythmically different utterances can give slightly different results. Nevertheless, these considerations do not question the methodology chosen. If it were to prove difficult to use such utterances in the same perception test according to our methodology, it would be possible to start the last experiment over with other types of lexically neutral utterances in order to carry out the comparison.

5. Conclusion

As regards methodology, we highlighted and showed how to overcome three major methodological issues existing in affective speech research: (1) the question of lexical meaning of the utterances used (present or not, affectively charged or not), (2) the question of inter- and intra-speaker variability, (3) differentiation of the specific role of prosody in the expression and identification of affective states.

With the help of our experiments, we showed that voice quality is not the only parameter important in the expression and identification of emotions. Our results underline that prosodic contour plays a significant role in the identification of emotions. This result does not only possess some theoretical value: prosodic contour must play an important part in the modeling of emotions in speech synthesis. The usage of natural prosodic contours by speech synthesis gives access, at least partially, to the expression of affective states. Indeed, voice quality sometimes plays a decisive role in identification, but it is difficult to manipulate this parameter in real time. It is much easier to implement prosodic contours (fundamental frequency,

intensity, speech rate) in speech synthesis. The analysis of inter-speaker variability of the corpus and the selection of those strategies which are most adapted for reproduction in speech synthesis constitute the following stage of our project. Detailed acoustic analysis of our corpus in relation with gating points will also be the object of our future work.

References

- d'Alessandro, Christophe 2006. Voice source parameters and prosodic analysis. In Sudhoff, Stefan / Lenertová, Denisa / Meyer, Roland / Pappert, Sandra / Augurzky, Petra / Mleinek, Ina / Richter, Nicole / Schließer, Johannes (eds) *Methods in Empirical Prosody Research*. Berlin, New York: De Gruyter, 63-87.
- Bänziger, Tanya / Scherer, Klaus R. 2005. The role of intonation in emotional expressions. *Speech Communication* 46/3-4, 252-67.
- Chen, Aoju / Gussenhoven, Carlos / Rietveld, Tony 2004. Language-specificity in perception of paralinguistic intonational meaning. *Language and Speech* 47/4, 311-50.
- Garcia, Marie-Neige / d'Alessandro, Christophe / Bailly, Gérard / Boula de Mareüil, Philippe / Morel, Morel 2006. A joint prosody evaluation of French text-to-speech systems. In Calzolari, Nicoletta (ed) *Proceedings of LREC*, Genoa, Italy, 307-10.
- Ghio, Alain / André, Carine / Teston, Bernard / Cavé, Christian 2003. PERCEVAL: une station automatisée de tests de PERception et d'EVALuation auditive et visuelle. *Travaux Interdisciplinaire du Laboratoire Parole et Langage d'Aix-en-Provence* 22, 115-33.
- Gobl, Christer / Ní Chasaide, Ailbhe 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40/1-2, 189-212.

- Grichkovtsova, Ioulia 2008. *A cross-linguistic study of affective prosody production by monolingual and bilingual children: Scottish English and French*. PhD thesis. Edinburgh: Queen Margaret University.
- Grosjean, François 1996. Gating. *Language and Cognitive Processes* 11/6, 597-604.
- Johnstone, Tom / Scherer, Klaus R. 2000. Vocal communication of emotion. In Lewis, Michael / Haviland, Jeannette M. (eds): *Handbook of emotions*. New York: Guilford, 220-35.
- Ladd, D. Robert 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Murray, Iain R. / Arnott, John L. 1993. Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of Acoustic Society of America* 93/2, 1097-108.
- Morel, Michel / Lacheret-Dujour, Anne 2001. Kali, synthèse vocale à partir du texte: de la conception à la mise en oeuvre. *Traitement Automatique des Langues* 42/1, 1-29.
- Scherer, Klaus R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40/1-2, 227-56.
- Yanushevskaya, Irena / Gobl, Christer / Ní Chasaide, Ailbhe, 2006. Mapping Voice to Affect: Japanese listeners. In Hoffmann, Rüdiger / Mixdorff, Hansjörg (eds) *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden: TUDpress, paper OS4-4-265.

Appendices

Appendix 1. Examples of utterances from the corpus with affectively charged lexical meaning

Anger: J'ai encore retrouvé ma voiture neuve toute rayée, c'est inadmissible! (I have found my new car scratched, it is unacceptable.)

Doubt: J'ai roulé sans phares en pleine nuit? (I was driving with lights off at night?)

Happiness: Il a appelé sa mère pour lui annoncer la bonne nouvelle. (He called his mother to tell her the good news.)

Irony: J'ai réussi ma chute en pleine rue brillamment. (I managed to fall in the middle of the street very nicely.)

Obviousness: Il a mis un manteau pour sortir! (He put on his coat to go out!)

Sadness: J'ai finalement compris que je ne la reverrais plus. (I finally understood that I would never see her again.)

Appendix 2. Examples of utterances from the corpus with affectively neutral lexical meaning

Anger: Vous appelez ça une chambre d'hôtel? Regardez un peu ces draps! Ils sont ignobles. Vous ne croyez quand même pas que je vais dormir ici! C'est révoltant! *Je vais rentrer à la maison maintenant!* Ce n'est pas un hôtel ici, c'est un élevage de cafards! (Do you call this a hotel room? Look at these sheets! You don't think that I am going to sleep here! It is disgusting! I am going home now! It is not a hotel here; it is a cockroach farm.)

Grief: Tu sais comme j'aimais mon chien? Hier, quand je suis revenu(e) de voyage, j'ai appris qu'il était mort. Je suis bouleversé(e). *Je vais rentrer à la maison maintenant.* J'vais jamais pouvoir le dire aux enfants. (You know how I loved my dog, don't you? Yesterday when I came back from my trip I found out that she had died. I am overwhelmed. I am going home now. How will I be able to tell the children?)

Happiness: Mon frère reviendra demain! Quelle joie! Je suis si content(e)! *Je vais rentrer à la maison maintenant!* Je vais annoncer cette super nouvelle à ma famille! (My brother is coming tomorrow!

Such a joy! I am so happy! I am going home now! I will tell this excellent news to my family!)

Neutral: J'ai fini de ranger les boites. Elles sont classées et numérotées. *Je vais rentrer à la maison maintenant.* Je reviendrai demain à dix heures. (I have finished putting the projects in order. They are filed and numbered. *I am going home now.* I will be back tomorrow at ten in the morning.)

Sadness: Ce que tu m'as appris m'a fichu le moral à zéro. C'est vraiment déprimant... *Je vais rentrer à la maison maintenant.* J'ai l'impression que c'est une situation sans issue. (Your news makes me feel so low. It is so depressing. I am going home now. I have the impression that there is no way out.)