Affective speech gating

Ioulia Grichkovtsova[1], *Anne Lacheret*[2], *Michel Morel*[1],
Virginie Beaucousin[3], *Nathalie Tzourio-Mazoyer*[3]

[1]CRISCO, Université de Caen, France
[2]MoDyCO, Université Paris X, Nanterre, France
[3]GIN, CNRS UMR 6194, GIP Cyceron, France
ioulia.grichkovtsova@unicaen.fr

Abstract

This study tested the hypothesis that emotions may be identified earlier than attitudes. The gating paradigm was chosen to investigate if such differentiation between emotions and attitudes was possible. Perception test results included the following variables: the identification point, the isolation point and the confusion matrices. Acoustic analysis was conducted and linked to the perception results. Anger and sadness were separated from the other studied affective states on the basis of the results analysis. Interestingly, happiness followed the identification pattern found for attitudes. The future directions of work are presented.
Index Terms: affective speech, speech perception, gating paradigm.

1. Introduction

The scientific context of the present study is placed in the multi-disciplinary perspective of psycholinguistics, phonetics and neurolinguistics. The tested hypothesis originates from the collaborative study on neural bases of affective speech comprehension between two laboratories: linguistic and neuroimaging. In the frames of the study, two categories of affective states were studied and compared: emotions (anger, sadness and happiness) and attitudes (obviousness, irony and doubt). According to the definitions proposed by Scherer [1], emotions represent a synchronised organismic response to the event of major significance, while attitudes like obviousness, doubt, may be described as affective colouring of interpersonal interaction.

The results of our identification psycholinguistic test, reported in [2], showed no significant differences between the two categories on the bases of response analysis. It was hypothesised that the difficulty to distinguish statistically emotions from attitudes in the results of the identification test may be explained by the fact that subjects were permitted to answer only after listening to the whole utterance. The presence of specific prosodic characteristics, occurring from the very beginning of emotional utterances, may allow early identification. Listeners may be able to recognise emotions before the end of the utterance. The identification of attitudes, requiring an intergrated analysis of lexical, syntactic and prosodic structures of the utterance, may happen later than for emotions.

The gating paradigm was chosen to test the hypothesis that emotions may be identified earlier than attitudes. The gating paradigm was originally developed by Grosjean [3], and it was widely used in spoken word recognition research [4, 5]. More recently, it was applied in intonation research [6, 7]. The gating task is based on the principal that an audio stimulus is presented in segments of increasing duration, and the subjects are asked to identify what was said at the end of each segment or "gate". The gating paradigm allows to understand how much acoustic-phonetic information is needed to identify a stimulus.

2. Method

2.1. Stimuli description

Two utterances were pronounced by an actor and an actress for each studied affective state (anger, sadness, happiness, obviousness, irony and doubt), thus the total of 24 utterances was used in the experiment. The lexical meaning of the utterances was not neutral. They were designed to carry natural lexical meaning, the beginning of the utterances was potentially possible for several affective states. The meaning was disambiguated by prosodic and lexical means by the end of the utterance. Ironic utterances were marked by a deliberate contrast between their apparent and intended meaning. It was achieved in the used utterances by opposing the literal meaning of the words between themselves and reinforced by the tone of voice. Examples of the corpus are given in the Appendix.

2.2. Participants

13 subjects were recruited in University of Caen (5 females and 8 males). They were students and professionals working in the university. Average subject age was 31 years old, with standard deviation of 13.2. All subjects were native speakers of French and none reported having any hearing difficulty.

2.3. Procedure

Special software for psycholinguistic perception tests *Perceval* was used for the experiment design. Perceval was developed in the Laboratoire Parole et Langage, Aix-en-Provence, France, and is free for academic purposes [9]. It allows the programming of stimuli presentation according to the gating paradigm without cutting utterances at the gating points. Increment size of gates was fixed at 200 msec. The duration-blocked presentation format was chosen: first all the stimuli of the particular segment size were presented to the listeners in a randomised order, then all the stimuli of the following segment size, and so on. Each block contained 24 stimuli. The experiment was run on a computer in a quiet laboratory room. The whole experiment was run by the Perceval software, responses and response times were automatically recorded in the data file. In average, the experiment took about 40 minutes.

2.4. Analysed variables

Responses proposed after each gate were analysed, both correct responses and observed patterns of confusion. Two variables have been selected for the analysis of correct responses: identification point and isolation point. *Identification point* refers to the gate of the stimulus where correct identification achieves 50 %. *Isolation point* is the gate of the stimulus where the highest identification is achieved and maintained without any change in response thereafter.

2.5. Results

Results for the identification point are shown in Table 1. Anger stands out of the other affective states, as it is the earliest to get recognised at 50 %. Angry utterances are successfully recognised at the first or maximum at the second gate. Sadness also can be recognised as early as the second gate, but more variability in the identification of sad utterances was observed. Interestingly enough, obviousness may be also identified early. Other affective states, and especially irony, were identified later.

Table 1: *Results for the identification point (50 % of recognition)*

1	Anger	200/400 msec
2	Sadness Obviousness	400/1000 msec 400/1200 msec
3	Doubt	600/1400 msec
4	Happiness	800/1800 msec
5	Irony	1600/1800 msec

Results for the isolation point are presented in Table 2. It shows the range of observed gates for each affective state where the highest identification is achieved and maintained without any change in response thereafter. All the studied affective states have some utterances which may be recognised before the end of the utterance. At the same time, only anger and sadness have ***all*** their utterances recognised before hearing the complete form. This observation allowed to separate the affectives states into two groups. Interestingly, based on the analysis of the identification point, happiness was classified not with emotions, but with attitudes.

Table 2: *Results for the isolation point*

1	Anger Sadness	800/1600 msec 800/1800 msec
2	Doubt Obviousness Happiness Irony	800/complète 800/complète 1000/complète 1600/complète

Confusion matrices were used to study the percentage of correct responses and the pattern of confusions. Confusion matrices were based on the principle to plot encoded affective states against the decoded affective states. Two confusion matrices were calculated in the present study: the first on the whole set of responses - Table 3, and the second on the two subgroups of the response data (before and after the isolation point) - Table 4. The separation of the response data allows to study the evolution of confusion patterns. Responses after the identification point present a particular interest, as they show which confusions occur between the identification and isolation points.

Table 3: *Confusion matrice calculated on the whole response data (in percentage). Anger - A, Doubt - D, Obviousness - O, Happiness - H, Irony - I, Sadness - S.*

	A	D	O	H	I	S
A	**74,9**	2,5	7,7	10,5	3,1	1,1
D	8,4	**57,6**	8,4	6,8	15,3	3,3
O	10,8	12,4	**56,9**	1,3	10,6	7,7
H	11,9	4,9	10,0	**54,1**	15,1	3,8
I	2,8	4,0	15,4	29,2	**39,6**	8,9
S	1,3	5,8	18,3	1,1	3,1	**70,1**

Table 3 shows that confusion patterns are observed for all the affective states, nevertheless anger and sadness have the lowest level of confusion. Obviousness, doubt, happiness, and especially irony have high level of confusion. Table 4 gives an important insight in how the confusion patterns develop before and change after the identification point. The general observation is that emotions show less confusion after the identification point than attitudes, this is true also for happiness. For example, before the identification point, sadness is confused with all the other affective states, but significantly with obviousness and happiness. After the identification point, no confusion of sadness with other emotions is observed, only some insignificant confusion with attitudes.

Table 4: *Confusion matrice calculated separately for responses before and after the identification point IP (in percentage). Anger - A, Doubt - D, Obviousness - O, Happiness - H, Irony - I, Sadness - S.*

		A	D	O	H	I	S
A	Before IP	**42,8**	6,5	16,4	26,3	7,6	0
A	After IP	**86,1**	1,15	4,6	5	1,5	1,5
D	Before IP	15,2	**31,9**	13,3	12,8	20,4	6,2
D	After IP	0,5	**87,3**	2,7	0	9,3	0
O	Before IP	11,8	13,5	**40,4**	2,8	14,1	17,4
O	After IP	10,7	12,3	**66,1**	0,4	8,8	1,5
H	Before IP	20,1	9,2	17,7	**32,25**	13,3	7,2
H	After IP	3,2	0	1,6	**87,3**	7,6	0
I	Before IP	3,9	4,9	16,9	36,06	**26,1**	11,9
I	After IP	0	1,7	11,8	13,0	**71,5**	1,7
S	Before IP	3,8	12,2	38,06	33,2	3,8	**38,7**
S	After IP	0	2,4	7,6	0	2,7	**87,06**

The main objective of the acoustic analysis was to understand the relationship between the level of identification and the used pattern of acoustic correlates. The acoustic analysis was conducted with PRAAT software [10] and linked to the identification and isolation points. An example of such analysis is shown in Figures 1 and 2. A large number of acoustic correlates (voice quality, intensity, initial F0, pitch span, pitch level, peak steepness, speech rate) may be used for the encoding and decoding of affective states.

Some difficulty to generalise the acoustic results was encountered due to the fact that the two speakers used different strategies to express the same affective states. Moreover, variability within the same speaker was also observed. This variability on the production level had its impact on the identification level. Figure 3 shows the level of identification for each utterance from the used corpus. For example, utterances 15 and 16 (see Figures 1 and 2) said by the male speaker with happi-

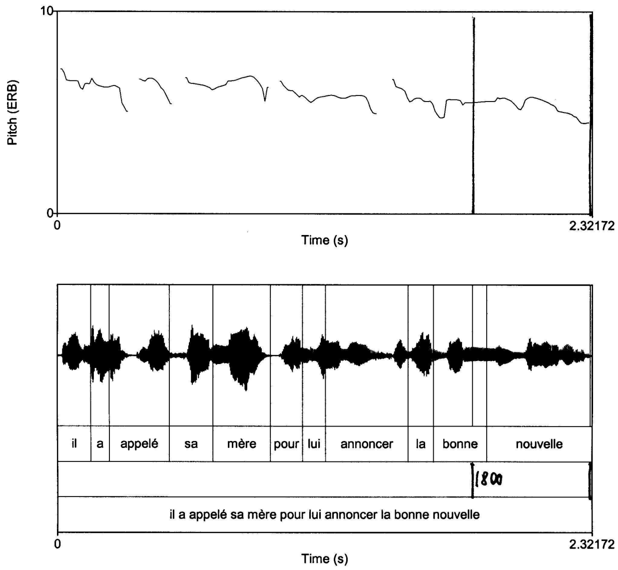

Figure 1: *The graph represents an example of the acoustic analysis realised in PRAAT. Utterance 15 was produced with happiness by the male actor. The identification point is at 1800 msec, and the isolation point is at the end.*

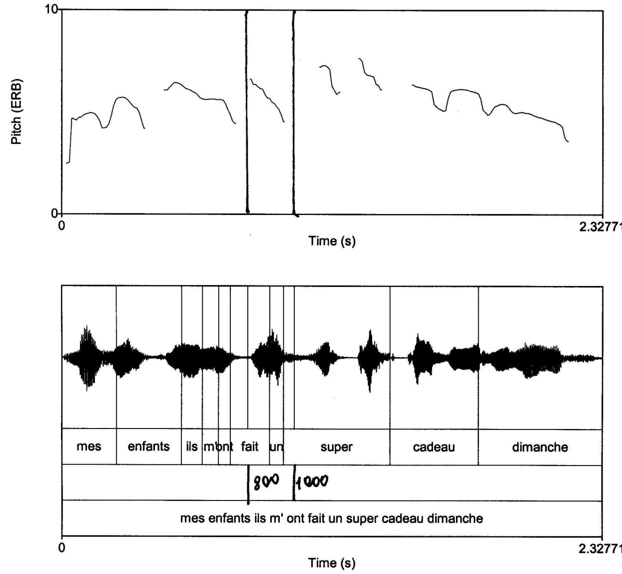

Figure 2: *Utterance 16 was produced with happiness by the male actor. The identification point is at 800 msec, and the isolation point is at 1000 msec.*

Table 5: *Results of acoustic analysis for utterances 15 and 16 (happiness by male actor)*

Measure	Utterance 15	Utterance 16
Initial F_0	6.8 ERB	5.0 ERB
Pitch range	40.7 st	42.6 st
Final F_0	5.1 ERB	4.8 ERB
Speech rate	6.1 syl/sec	5.6 syl/sec
Voice quality	-	smiling

ness differ in that utterance 16 has its isolation point at the 5th gating point (1000 msec), while utterance 15 is identified only at 50 % even in its complete form. Table 5 gives detailed results of the analysed acoustic measures. It is possible to see that utterance 15 is pronounced fast with high pitch level and narrow pitch range and without smiling. On the contrary, utterance 16 has a lower pitch level and a wider pitch range, moreover the actor is smiling. The isolation point for the utterance 16 occurs before the lexical meaning is disambiguated. Due to the lack of space, the detailed analysis of the 24 analysed utterances is not possible to show in this papier.

3. Discussion

The hypothesis that emotions may be identified earlier than attitudes was tested in the present study by the gating paradigm. Based on the results for the recognition point, it is possible to see that anger differs from the other affective states: all the angry utterances were identified very early (1-2 gating point), while other affective states show much more variability. Due to the isolation point, the studied affective states were separated into two groups. The first group comprises anger and sadness, as all angry and sad utterances were recognised before the end of the utterance. Anger and sadness showed much less confused responses than the other affective states. Based on these results, the differentiation of emotions from attitudes was realised. Happiness makes an exception from the other studied emotions, as its recognition follows the pattern identified for attitudes. It is a very interesting observation, it may be explained by a particular communicative role, played by happiness. Happiness may be better controlled by speakers, and it can be used intentionally to colour interpersonal communication, in a way attitudes are used.

Acoustic results showed variability both in the production of the two speakers, and in the production of the same speaker for the studied affective states. It has been acknowledged in the previous studies [10] that affective states are communicated through voice by a variable combination of acoustic parameters. The search for stable associations between a group of particular acoustic parameters and specific affective states has not yet been successful. It has been recently suggested [12] that several different strategies may be successfully used for the expression of the same affective state. The acoustic analysis of the studied corpus does not allow to understand in detail which correlations between the usage of acoustic correlates and their perceptual value are possible, as the corpus has only two speakers, moreover they had different utterances for the same affective state. Nevertheless, it was shown that the chosen pattern of acoustic correlates had influence on the level of recognition, some strategies were more successful than others.

4. Conclusion

The gating paradigm allowed to discover some differences in the identification of emotions and attitudes. The hypothesis that emotions may be identified earlier than attitudes was confirmed for anger and sadness, at the same time it was not possible to separate happiness from attitudes in the perception test results. The study has also shown inter- and intra-speaker variability in the realisation of affective states, and its influence on the perception. Some strategies to encode the same affective state were

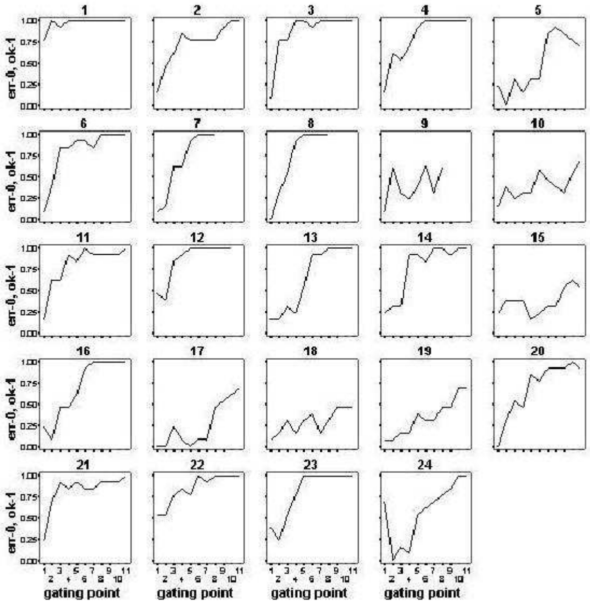

Figure 3: *The graph shows the response results for each individual utterance. Description of utterances: 1-2 (anger female), 3-4 (anger male), 5-6 (doubt female), 7-8 (doubt male), 9-10 (obviousness female), 11-12 (obviousness male), 13-14 (happiness female), 15-16 (happiness male), 17-18 (irony female), 19-20 (irony male), 21-22 (sadness female), 23-24 (sadness male).*

more successful and easier for the identification than others. The used corpus of affective speech included only two speakers, it cannot give deep understanding of possible strategies in the production of affective states. At the present, a new corpus is being developed for the perception tests with the gating paradigm. A large number of speakers will be recorded encoding several affective states on the same neutral utterance. It will help to investigate the width of possible variability on the production level and to search for correlations between the usage of acoustic correlates and their perception value.

5. Appendix

Sadness: J'ai finalement compris que je ne la reverrais plus. (I finally understood that I would never see her again.)
Happiness: Il a appelé sa mère pour lui annoncer la bonne nouvelle. (He called his mother to tell her the good news.)
Doubt: J'ai roulé sans phares en pleine nuit? (I was driving with lights off at night?)
Anger: J'ai encore retrouvé ma voiture neuve toute rayée, c'est inadmissible! (I have found my new car scratched, it is unacceptable!)
Obviousness: Il a mis un manteau pour sortir! (He put on his coat to go out!)
Irony: J'ai réussi ma chute en pleine rue brillament. (I managed to fall in the middle of the street very nicely.)

6. Acknowledgements

This work was supported by a grant of the Basse-Normandie Regional Council. The authors would like to thank all the participants who kindly agreed to do the perception tests.

7. References

[1] Scherer, K. R. (2003) "Vocal communication of emotions: A review of research paradigms." Speech Communication, 40, 227-256.

[2] Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, M. and Tzourio-Mazoyer, N. (2007) "FMRI study of emotional speech comprehension." Cerebral Cortex, 17, 339 - 352.

[3] Grosjean, F. (1996). "Gating." Language and Cognitive Processes, 11 (6), 597-604.

[4] Lickley, R.J., McKelvie, D. and Bard, E.G. (1999) "Comparing human and automatic speech recognition using word gating." Proceedings of the ICPhS Satellite meeting on Disfluency in Spontaneous Speech, UC Berkeley, 2326.

[5] Walley, A., Michela, V. and Wood, D. (1995) "The gating paradigm: Effects of presentation format on spoken word recognition by children and adults." Perception and psychophysics, 57, 343-351.

[6] Aubergé, V., Grépillat, T. and Rilliard, A. (1997) "Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours." Proceedings of the European Conference on Speech Communication and Technology, vol. 2, Rhodes, Greece, 871-874.

[7] Vion, M. and Colas, A. (2006) "Pitch cues for the recognition of Yes-No Questions in French." Journal of Psycholinguistic Research, 35, 427-445.

[8] Gérard, C. and Clément, J. (1998) "The structure and development of French prosodic representation." Language and Speech, 41, 117-142.

[9] Ghio, A., André, C., Teston, B. and Cavé C. (2003) "PERCEVAL: une station automatise de tests de PERCeption et d'EVALuation auditive et visuelle", TIPA, vol. 22, Aix-en-Provence, France, 115-133.

[10] Murray, I.R. and Arnott, J.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, 1097-1108.

[11] Boersma, P. and Weenink, D. (2006) "Praat: doing phonetics by computer." Version 4.5.15 (Computer program). Retrieved December 20, 2006, from http://www.praat.org/.

[12] Grichkovtsova, I. (2007) "A cross-lingustic study of affective prosody production by monolingual and bilingual children: Scottish English and French." PhD thesis, Queen Margaret University.