PERCEPTION OF SMILED FRENCH SPEECH BY NATIVE VS. NON-NATIVE LISTENERS: A PILOT STUDY

Caroline Émond[1], Jürgen Trouvain[2] and Lucie Ménard[1]

[1]Université du Québec à Montréal and [2]Universität des Saarlandes
caroemond@hotmail.com, trouvain@coli.uni-sb.de, menard.lucie@uqam.ca

ABSTRACT

Smiling is a visible expression and it has been shown that it is audible too (Tartter [5], Tartter and Braun [6], Schröder et al. [4], Aubergé and Cathiard [1]). The aim of this paper is to investigate the perception of the prosody of smile in 2 different languages. In order to elicitate smiled speech, 6 speakers of Québec French were required to read sentences displayed with or without caricatures. The sentences produced were used as stimuli for a perception test administered to 10 listeners of Québec French and 10 listeners of German who had no knowledge of French. Results suggest that some prosodic cues are universals and others are culture specific.

Keywords: prosody, emotion, smiling, smiled speech.

1. INTRODUCTION

It has been hypothesized by Tartter [5] that because the vocal tract is altered from its neutral position when there is smiling, there should be an audible effect. Even when neutral and smiled speech is produced in controlled and unnatural conditions (mechanically isolated speech samples – one syllable nonsense words), results showed that smiling has audible effects and those effects are automatically associated with positive emotions. Moreover naive listeners can reliably identify smiled speech as such.

Like Tartter [5] and Tartter and Braun [6], Schröder et al. [4] showed in the first perceptual experiment of their study that listeners can accurately discriminate utterances produced with a mechanical smile when presented with their neutral counterparts in audio condition. According to Ekman et al. [2], amusement smiles (so-called Duchenne smiles, i.e. produced with upturned mouth corners, raised cheeks, and crinkling of eyes) and smiles without any relation to amusement are different. Following them, Schröder et al. [4] showed in a second perceptual experiment that spontaneous stimuli (vs. mechanical stimuli) are discriminated just as the amused ones in audio conditions.

In Aubergé and Cathiard [1], the visual and audiovisual conditions were analyzed in relationship with the audio one. They found that the audio modality contained a lot of information in such a visible emotion as amusement. This information is not due only to the change of the vocal tract from its neutral position but there is also "a specific manipulation of the prosody of the speech" (p. 96) in the expression of amusement.

According to those references, it seems universal that smile in speech can be perceived by listeners. The objective of the present study is thus to compare the perception of smiled French speech across 2 languages: Québec French (QC) and German (GE).

2. METHOD

Before going any further, it is important to note that this pilot study is part of a larger study focusing on the prosodic correlates of smiled speech in QC (Émond [3]).

2.1. Corpus, participants and recordings

For the production part of the study, 10 humourous caricatures published in daily newspapers (*La Presse*) were chosen in order to elicitate smiles. 30 fillers were added (20 sentences presented alone or with drawings and the titles of the caricatures without the drawings (10), n = 40).

6 participants ranging in age from 22 to 34 years old (3 men, 3 women) with QC as L1 were recruited in the department of linguistics at the Université du Québec à Montréal. They were not aware of the study's true objective before the recordings. Stimuli were semi-randomized across the speakers, the first ten utterances being neutral.

Speakers were audio-video recorded. The recordings took place in a sound proof room with the following material: an IBM laptop, a Panasonic AG-DVC30 numeric camera, a (DAT) Tascam

numeric recorder and a Shure Beta 58A dynamic microphone.

The participants sat in front of the laptop, with the microphone about 30 cm from their mouth. The instructions were first presented on the screen and a training phase preceded the task. They were asked to read 40 sentences out loud and the test lasted about 10 minutes.

2.2. Selection of the test corpus

The data were digitized with Adobe Premiere and segmented with Goldwave. First, the utterances deemed as spontaneous smiled speech were selected. Audiovisual inspection was done to ensure those utterances were produced with the Duchenne effect. 32 utterances (out of 240 – 40 sentences x 6 speakers) were selected and to these were added 12 fillers i.e. utterances perceived as neutral (n = 44). It is important to note that the preselection of the corpus was only made to select a subset of sentences to be submitted as a perceptual experiment. Even though produced smiled speech corresponds here to the sentences produced with the Duchenne effect, a sentence will be said to be smiled speech only if it is perceived as such by the listeners. Our method is thus clearly listener-oriented, in part because of the origin of our listeners.

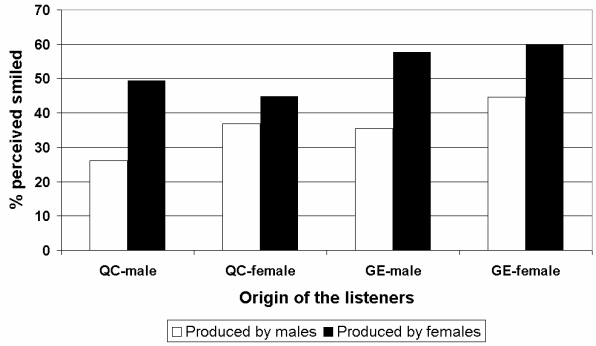
2.3. Perception test

The subset of sentences described above was used for an auditory perceptual experiment. 20 participants aged from 20 to 39 years old participated to the perceptual test: 10 Germans (GE – 5 men, 5 women) and 10 Quebecers (QC – 5 men, 5 women). Stimuli were presented via PC loudspeakers in a random order and mixed condition. The test took place in a quiet room and lasted about 5 minutes. The task consisted of a forced choice between 2 possible answers: smile or not smile (neutral).

3. RESULTS

On the whole, all the listeners behaved the same i.e. the number of utterances perceived as smiled by the GE is proportionally the same as the QC even if only the latter benefit from lexical and semantic access. Indeed, since no delexicalization method was used, QC listeners could have used segmental and prosodic cues, as well as some semantic content. On the contrary, for the GE listeners, the QC sentences did not have any semantic content and can be compared to some kind of "ecological delexicalization." It can thus be hypothesized that GE listeners refered only to prosodic and phonetic parameters. We shall come back to this issue later. Fig. 1 shows the distribution of the utterances perceived as smiled by both linguistic groups. This figure shows that the utterances produced by QC females are perceived more smiled than the ones produced by QC males by all the listeners. However, GE listeners tend to perceive a larger percentage of the sentences as smiled, compared to QC listeners.

Fig. 1: Perception of smiled speech based on the origin of the listeners.



In order to further investigate between-group differences, fig. 2 to 5 represent the perception of smiled speech by linguistic group and gender.

The perception of QC speakers is presented in fig. 2 (male listeners) and fig. 3 (female listeners). Data are grouped according to the origin of the listeners (light bars = QC listeners; dark bars = GE listeners). If we compare for example the perception of the utterance F_Am_28 by male and female listeners (fig. 2 & 3) it can be seen that 3 QC females (out of 5), 3 QC males (out of 5), 1 GE female (out of 5), and 4 GE males (out of 5) perceived this sentence as smiled. Those results show that there is a difference of perception between both linguistic groups but also between gender for the GE listeners.

Concerning utterance F_Lu_35 (fig. 2 & 3), 5 QC females, 3 QC males and all the GE listeners perceived it as smiled. This result suggests the presence of universal prosodic cues but that the perception differs in the members of the same cultural community.

Fig. 2: Recognition rates; female speakers; male listeners.

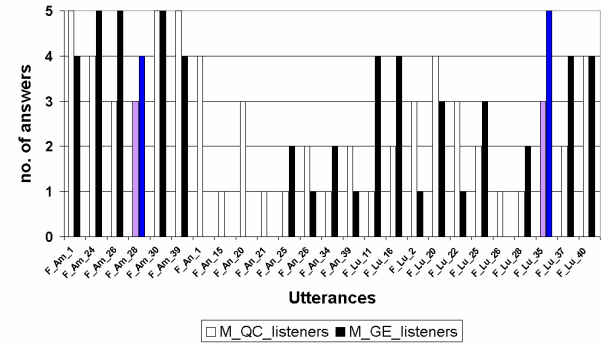


Fig. 3: Recognition rates; female speakers; female listeners.

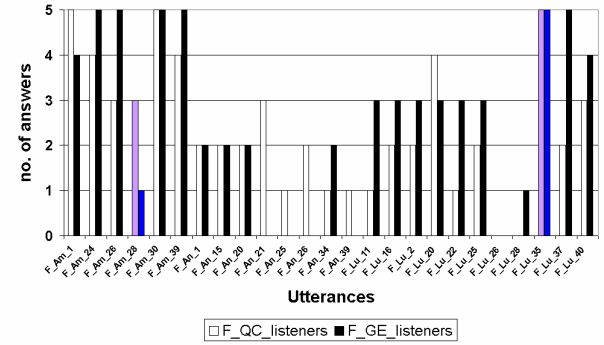


Fig. 4: Recognition rates; male speakers; male listeners.

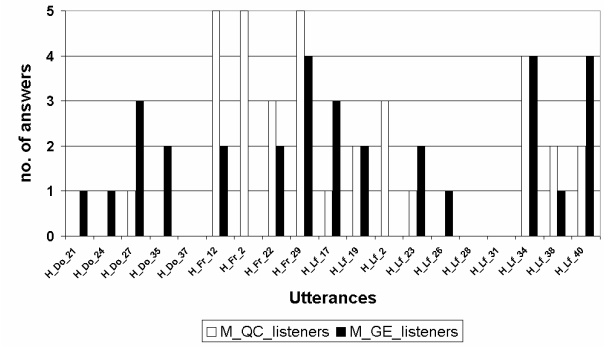
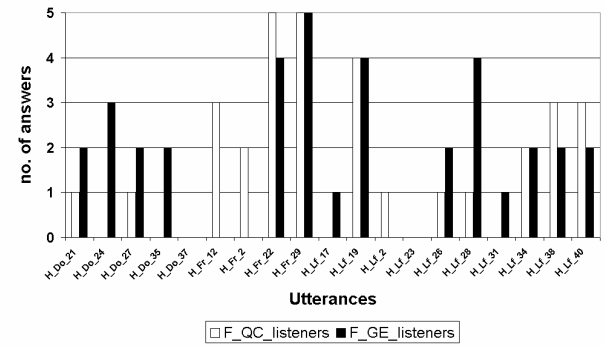
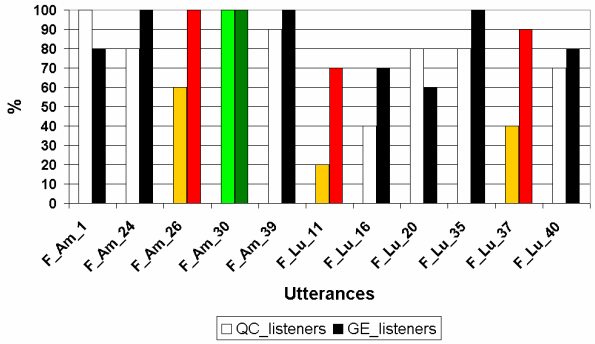


Fig. 5: Recognition rates; male speakers; female listeners.



Because of the important variability in the identification of smiled speech across stimuli, a subset of the corpus was used to study possible prosodic correlates of perceived smile. Utterances perceived as smiled by 70% of the listeners were thus considered. Fig. 6 shows 11 utterances out of 25 for 2 female speakers. There is one clear case where 100% of the listeners agree (F_Am_30). There is disagreement for a couple of utterances (F_Am_26, F_Lu_11, F_Lu_37) where more GE listeners found they were smiled compared to the QC listeners.

Fig. 6: Recognition rates of at least 70% in one of the listener group of 2 QC female speakers.



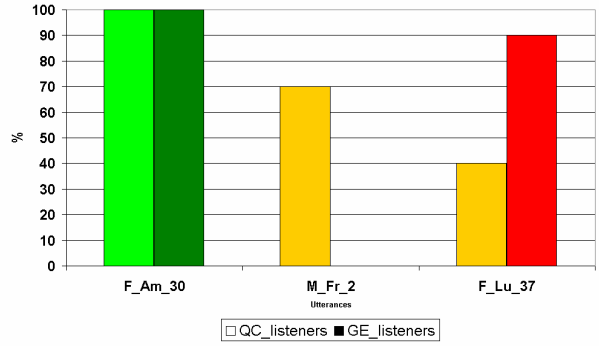
Concerning utterances produced by males, one sentence was perceived as smiled by at least 70% of the listeners.

This suggests that some universal cues may be present in the sentences for which most of the listeners agree, whereas culture-specific prosodic cues are produced in the sentences for which there is disagreement between the listeners. It is interesting to note that only 3 speakers out of 6 seem to be "more smiley." This may indicate that even if the utterances were presented in a random order in a mixed condition, there could be a speaker effect.

An unexpected phenomenon happened during the test: the spreading of the smile from an utterance to another. For that reason the neutral counterpart of the perceived smiled sentences were not present in the test. Maybe adding more fillers between the utterances would have helped. We examined the pitch range (max – min) and shape of the F_0 curves (with the standard software Praat) just to see if there could be an indication in the disagreement of the perception by both linguistic groups, if there could be a path to follow in further investigations even if we do not have those

utterances. Then we had a look at few cases for which all the listeners agreed (F_Am_30) and for which there was strong disagreement between the listeners (M_Fr_2, F_Lu_37). For the sake of clarity, the perception scores are depicted in Fig. 7 for the three sentences. We are aware that no comparison can be made at this stage but we observed a narrower pitch range for M_Fr_2 compared to other speakers which brings the idea that pitch range can be responsible for more subtle and perhaps sometimes culture-specific prosodic cues.

Fig. 7: Three types of recognition of smiled speech: full agreement across both groups of listeners (left), strong disagreement of the GE listeners (mid) and of the QC listeners (right).



4. CONCLUSION

We have seen that basically all the listeners behave the same i.e. the number of utterances perceived as smiled is nearly the same for both linguistic groups. The cases where all the listeners agreed may lend support the hypothesis of the existence of universal prosodic cues for the perception of emotions. However, strong disagreements sometimes arose. We can believe that this hypothesis is also supported when most of GE listeners perceive an utterance as smiled where most of QC listeners did not (e.g. F_Lu_37, fig. 6). In which case, we can suppose that pragmatic and lexical content play a strategic role. Following Thompson and Balkwill [7] our results suggest that "emotions are communicated prosodically through a combination of culturally determined and universal cues" (p. 421). In other words the recognition of smiling in speech is not as universal as expected.

We showed also that there are cases where the difference of the perception is due to listeners' gender. What is suggested here is the perception of emotion differs depending on speakers' and listener's gender. This criterion needs to be taken into account for any research in the field of emotions. For future work, it would necessary to have all the neutral counterparts to the utterances perceived as smiled as well as more participants for the perception test. Listeners from another dialect of the language of the speakers (e.g. French from France vs. French from Québec) would be another interesting variable to study.

Finally, we should not forget that elicitating emotions in this area of research is always harder than expected because obviously we deal with human beings and imponderables are numerous and frequent. The relationship between the experimenter and the participants is crucial. So, the collecting of spontaneous data in an experimental context added to the idiosyncratic aspects of smile and laughter remain at this time a sizeable challenge to researchers, one that should be tackled in future projects.

5. REFERENCES

[1] Aubergé, V., Cathiard M. 2003. Can we hear the prosody of smile? *Speech Communication* 40, 87-97.

[2] Ekman, P., Davidson, R. J., Friesen, W. V. 1990. Duchenne's smile: emotional expression and brain physiology. *Journal of Personality and Social Psychology* 58, 342-353.

[3] Émond, C. 2006. Une analyse prosodique de la parole souriante : une étude préliminaire. *Actes des XXVI^e Journées d'étude sur la parole (JEP)*, Dinard, 147-150.

[4] Schröder, M., Aubergé, V., Cathiard, M. 1998. Can we hear smiles? *Proc. ICSLP Sydney*, 559-562.

[5] Tartter, V. C. 1980. Happy Talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psycho-physics* 27 (1), 24-27.

[6] Tartter, V. C., Braun, D. 1994. Hearing smiles and frowns in normal and whisper registers. *Journal of Acoustical Society of America* 96 (4), 2101-2107.

[7] Thompson, W. F., Balkwill, L.-L. 2006. Decoding speech prosody in five languages. *Semiotica* 158–1/4, 407-424.