Perceived prosodic correlates of smiled speech in spontaneous data

Caroline Émond [1], *Lucie Ménard* [1], *Marty Laforest* [2]

[1] Département de linguistique, Université du Québec à Montréal, Canada
[2] Département de lettres et communication sociale, Université du Québec à Trois-Rivières, Canada
caroemond@hotmail.com, menard.lucie@uqam.ca, marty.laforest@uqtr.ca

Abstract

Smiling is a visible expression and also an audible one when it is synchronized with speech. Very few studies have documented the perceptual prosodic cues associated with perceived smiled speech, and there have been especially few studies using data from spontaneous speech. The aim of this study was to identify a combination of prosodic parameters that would allow a phonetic description of perceived smiled speech. A total of 85 utterances were extracted from spontaneous-speech data (Montréal 1995 corpus) and used as stimuli for a perception test administrated to 40 listeners (20 men, 20 women) of Quebec French. Perceived prosodic parameters of pitch height, pitch range, speech rate, and rhythm related to smiled speech are discussed.

Index Terms: smiled speech, smile perception intonation, prosody

1. Introduction

Smiling as a visual expression or nonverbal behavior has been the subject of many studies (Ekman et al. [1], Abel [2], etc.). From the studies by Tartter [3] and Tartter & Braun [4], we know that smiling is audible when it occurs with speech. Since then, several studies have focused on the production and perception of different kinds of smiles (Schröder et al. [5], Aubergé & Cathiard [6], Drahota et al. [7]) or on the distinction between laughing, smiling, and crying speech (Erickson et al. 8]). Our recent work examined universal versus culture-specific prosodic cues related to smiled speech (Émond et al. [9]) and the role of speaker and listener gender in the perception of smiled speech (Émond [10], Émond & Laforest [11]). Pitch seems to be the most salient parameter to describe smiled speech. Some studies of smiled speech have reported an increase (Tartter [3], Erickson et al. [8]) or an influence (Tartter & Braun [4], Drahota et al. [7]) of the average pitch height as well as a variation of the declination line (Aubergé & Cathiard [6]). In their study on modeling smiled speech with an articulatory speech synthesizer, Lasarcyk & Trouvain [12] found that F0 was the most important parameter involved in the perception of smile.

Most of the previous studies used "lab speech" (as opposed to "real-life," spontaneous data) and reading tasks. Moreover, data were produced in a monological context, involving a speaker who was not interacting with a listener. Even if the use of real-life data leads to several disadvantages when it comes to performing an instrumental analysis, the interactive structure (turn-taking organization, presence of discourse markers, etc.) of real-life data may lead to the discovery of phenomena that cannot be directly observed in lab speech.

In this paper, the prosodic correlates of smiled speech (as defined by Trouvain [13] and [14]) in spontaneous-speech data were analyzed. The objective was to identify, with the help of a perceptual task, combinations of prosodic parameters that could describe smiled speech phonetically.

2. Method

This study was part of a larger study of the perception of smiled speech. The goals of a previous study (Émond & Laforest [11]) were to see how well listeners could auditorily identify smiled speech and how gender impacted the perception of smiled speech. In the present paper, a second perceptual task was conducted, and the results were related to the results from our previous study.

2.1. Stimuli

This study used self recordings from Family 2, from a 1995 spoken corpus of four Montreal families (Vincent et al. [15]). This corpus was collected for sociolinguistic purposes and was thus suitable for our experiment. The 13.38 hours of conversation between the members of Family 2 — a 49-year-old man and a 32-year-old woman — took place in the kitchen during daily mealtimes. A subset of the corpus was selected by the experimenter, a trained listener, for the perceptual experiment. For this experiment, initially only the utterances perceived as smiled speech by 75% and more of males or females (n=58) were selected. "Neutral utterances" perceived as non-smiled utterances by 95% and more of males or females were later added (n=27), so that the spoken corpus contained a total of 85 utterances.

2.2. Participants and procedure

Forty native Quebec-French–speaking listeners (20 males, 20 females) with no language, speech, or hearing problems were recruited in an academic environment for a perceptual auditory experiment. All participants provided written informed consent in accordance with the Board of Ethics of the Université du Québec à Montréal (UQAM). The participants ranged in age from 19 to 35 (mean age 24.5). The procedure described here is similar in some aspects to the one used by Bänziger & Scherer [16]. The Parsour program (Bastien et al., [17]) was used and participants were presented with utterances in a random order via headphones. For each stimulus, the listeners were instructed to evaluate the intensity of four voice aspects: pitch height, pitch range, speech rate, and rhythm. To do so, participants had to position the computer mouse cursor at the appropriate place on a visual analog scale. The scale consisted of a line with a minus sign (–) and the label "pas du tout" ("not at all") on the left side and a plus sign (+) and the label "très" ("very") on the right side. The four visual analog scales were labeled as "aigüe, mélodique, rapide, rythmée" ("high, melodic, fast, rhythmic"). The listeners could listen to each utterance up to three times. The orthographic form of each utterance appeared on top of the voice aspects. Before the actual experiment, listeners heard exemplars of both extreme values ("not at all" and "very") of each of the four prosodic parameters, and they participated in a familiarization task. In the experiment, the stimuli were presented once, in random order. The test took about 30 minutes.

3. Results

For each prosodic parameter—pitch height, pitch range, speech rate, and rhythm—the visual analog scale readings were transposed into 5 categories that were each assigned a numerical value, as follows: not at all (1), slightly (2), moderately (3), quite (4), or very (5). For each stimulus and each prosodic parameter, we calculated the average perceived intensity value for the 20 male participants and the average perceived intensity value for the 20 female participants. Thus, each stimulus resulted in 8 scores: a perceived intensity score for each of the 4 prosodic parameters, for each listener gender. The intensity scores were correlated with the percentage of utterances perceived as smiled speech by the listeners, as determined in our previous experiment (not reported here). The results for each prosodic parameter and each listener gender are shown in Figures 1-4. Linear discriminant analysis was conducted to determine the extent to which prosodic parameters allowed a good clustering between perceived smiled and non-smiled speech.

3.1. Pitch height

In Figure 1, the x-axis shows the average perceived pitch height and the y-axis represents the percentage of perceived smiled speech utterances. Utterances perceived by females are shown as pink circles, whereas utterances perceived by males are represented by blue squares. It can be seen that for utterances perceived as having a moderate pitch height (values ranging from 1 to 3.3 along the x-axis), stimuli were either perceived as non-smiled speech (scores below 20% along the y-axis) or as smiled speech (scores above 50% along the y-axis). Thus, this prosodic parameter did not allow a good clustering between perceived smiled and non-smiled speech in that range. However, utterances perceived as having a very high pitch (values from 4 to 5 along the x-axis) were all perceived as smiled speech (values above 60% along the y-axis). A linear discriminant analysis conducted on the data with the percentage perceived as smiled speech as the grouping factor was significant (F(1,39)=22.02; p<0.01), and resulted in an average correct classification of 80%.

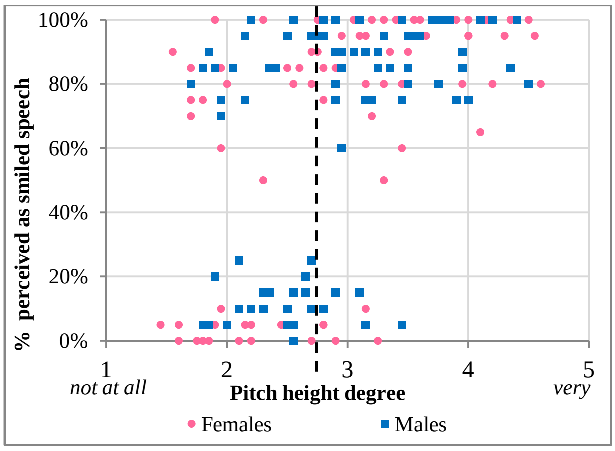

Figure 1: *Degree of the perceived pitch height for perceived smiled speech utterances for female and male listeners.*

3.2. Pitch range

The average intensity values of the perceived pitch range (“melody”) in relation to the percentage of stimuli that were perceived as smiled speech are displayed in Figure 2. A similar pattern to the one shown in Figure 1 for the pitch height parameter was found. Results of the linear discriminant analysis revealed a significant (F(1,39)=23.46; p<0.01) correct classification score of 84%. This slightly higher score compared to the score for the perceived height parameter suggests that perceived pitch range (“melody”) is a slightly better parameter to cluster both groups of stimuli according to their percentage perceived as smiled speech.

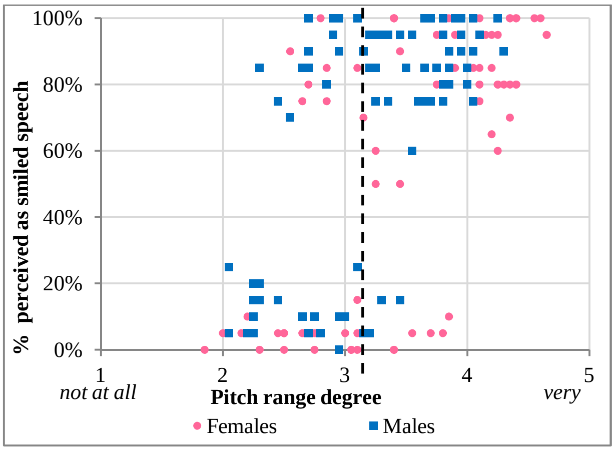

Figure 2: *Degree of the perceived pitch range for perceived smiled speech utterances for female and male listeners.*

3.3. Speech rate

Values of perceived speech rate in relation to the percentage of speech perceived as smiled speech are represented in Figure 3. Speech rate did not seem to be related to the perception of a smiling voice. The perceived intensity of this parameter was very similar for the smiled and the non-smiled speech. A linear discriminant analysis performed on the data did not reveal any significant result.

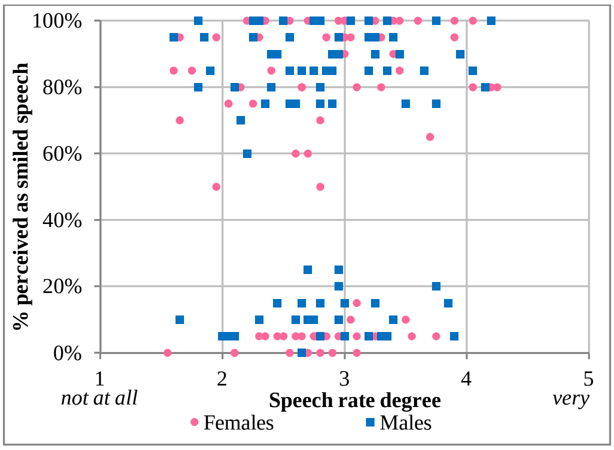

Figure 3: *Degree of the perceived speech rate for perceived smiled speech utterances for female and male listeners.*

3.4. Rhythm

Figure 4 shows the relationship between perceived rhythm and perceived smile. Here again, this prosodic parameter did not seem to be related to the perception of a smiling voice, as confirmed by the lack of a significant result from the linear discriminant analysis.

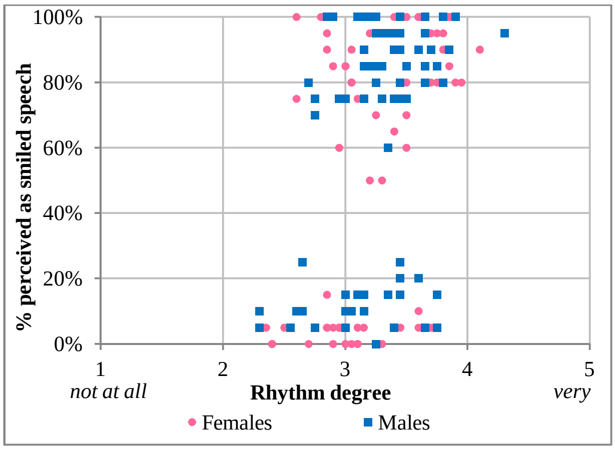

Figure 4: *Degree of the perceived rhythm for perceived smiled speech utterances for female and male listeners.*

To capture the possible relationships between all four prosodic parameters, several linear discriminant analyses were conducted on the data, with *percent perceived as smiled speech* as the grouping variable, and the following dependant variables: perceived prosodic parameters (pitch height, pitch range, speech rate, and rhythm) and listener gender. The combination of prosodic parameters that yielded the highest percentage of correct classification of the data (88%) was perceived height and perceived pitch range. For those parameters, listener gender did not show any significant effect.

4. Discussion

Based on previous work, we knew that it was possible to perceive a smile that occurs with speech. However, we did not know what cues listeners used to identify a human smiling voice. The results presented here show the significant role of intonation. Indeed, the prosodic parameters of pitch height and pitch range (and the combination of both parameters) were the prominent indications of perceived smiled speech. The increase of pitch height in the smiling condition agrees with the findings of Tartter [3] and Erickson et al. [8]. For their part, speech rate and rhythm do not seem to play an important role.

The most important finding in this study concerns the use of spontaneous-speech data. The experimental design developed for this study shows that it is possible to use a corpus made of real-life conversations for the analysis of prosodic parameters. Of course, that kind of data does not permit precise acoustic measurements, because of background noise or conversation overlap inherent to sociolinguistic corpora. But as soon as we have results about the role of some prosodic parameters in a real-life corpus, we can go back to the lab and make recordings related to the purposes of the subject under investigation, in order to make a further step in the comprehension of a given phenomenon. Moreover, methodological issues are of great interest and, as correctly pointed out by Xu [18], need to be enhanced and developed by taking the context, the speakers, and the listeners into account. Keep smiling!

5. Acknowledgements

This work was supported by a grant from the Social Sciences and Humanities Research Council (SSHRC). The authors also wish to thank Marlene Busko for copy-editing the paper.

6. References

[1] Ekman, P., Davidson, R. J. and Friesen, W. V., "The Duchenne Smile: Emotional Expression and Brain Physiology II", *Journal of Personality and Social Psychology*, 58(2):342-353, 1990.

[2] Abel, M. H. [Ed], *An Empirical Reflection of Smile*, Mellen studies in psychology Series 4, Lewiston: Edwin Mellen Press, 275 p., 2002.

[3] Tartter, V. C. "Happy talk: Perceptual and acoustic effects of smiling on speech", *Perception and Psychophysics*, 27(1):24-27, 1980.

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

[5] Schröder, M., Aubergé, V. and Cathiard, M.-A., "Can we hear smile?", *Proc. of the Conference on Spoken Language Processing*, 3:559-562, 1998.

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

[7] Drahota, A., Costall, A. and Reddy, V., "The vocal communication of different kinds of smile", *Speech Communication*, 50:278-287, 2008.

[8] Erickson, D., Menezes, C. and Sakakibara, K., "Are you laughing, smiling or crying?", *Proc. of APSIPA Summit and Conference*, 529-537, 2009.

[9] Émond, C., Trouvain, J. and Ménard, L., "Perception of French smiled speech by native vs. non-native listeners: a pilot study", *Proc. of the Interdisciplinarity Workshop on the Phonetics of Laughter* – 16th ICPhS, 27-30, 2007.

[10] Émond, C., "Les corrélats prosodiques et segmentaux de la parole souriante en français québécois", MA Thesis, Université du Québec à Montréal, 139 p., 2008.

[11] Émond, C. and Laforest, M., "Prosodic correlates of smiled-speech", *Proc. of the 21st International Congress on Acoustics*, to be published.

[12] Lasarcyk, E and J. Trouvain. "Spread Lips + Raised Larynx + Higher F0 = Smiled Speech? – An Articulatory Synthesis Approach." *Proc. of 8th ISSP*, 345-348, 2008.

[13] Trouvain, J., "Phonetic Aspects of 'Speech-Laughs'", *Proc. of ORAGE, Orality and Gestuality Conference*, 634-639, 2001.

[14] Trouvain, J., "Segmenting Phonetic Units in Laughter", *Proc. of the 15th ICPhS*, 2793-2796, 2003.

[15] Vincent, D., Laforest, M. and Martel, G., "Le corpus Montréal 1995. Adaptation de la méthodologie sociolinguistique pour l'analyse conversationnelle", *Dialangue*, 6:29-45, 1995.

[16] Bänziger, T. and K. R. Scherer., "Relations entre caractéristiques vocales perçues et émotions attribuées", *Proc. of Journées Prosodie*, 119-124, 2001.

[17] Bastien, M., Émond, C. and Ménard, L. "Parsour Program", 2010-2012.

[18] Xu, Y. "Speech Prosody: a Methodological Review", *Journal of Speech Sciences*, 1(1):85-115, 2011.