Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb39. Prosodic correlates of smiled speech

Caroline Emond* and Marty Laforest

***Corresponding author's address: Linguistique, Université du Québec à Montréal, C. P. 8888, Succ. Centre-Ville, Montréal, H3C 3P8, Québec, Canada, caroemond@hotmail.com**

Smiling is a visible expression and an audible one too when it is synchronous with speech. Very few studies have documented the perceptual prosodic cues associated with perceived smiling speech. The first aim of this paper is to study the perception of smiled-speech according to the listeners' gender. The reaction time and the intensity of the perceived smiled-speech were also investigated. The second aim is to identify a combination of prosodic parameters which would allow a phonetic description of smiled-speech. 140 utterances were extracted from spontaneous data (Montréal 1995 corpus) and used as stimuli for a perception test administered to 40 Québec French listeners (20 men, 20 women). Results show that men and women do not perceived smiled-speech in the same way, and women are quicker than men to make their decisions. Moreover, reaction times are faster for utterances perceived as smiling with a high degree of intensity, for both men and women, than those with lower intensity. Perceived prosodic parameters related to pitch height, pitch range, rhythm, and speech rate in relation to smiled-speech and its intensity are also discussed.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4799490]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Smiling is a universal expression that is mostly related to a positive emotional aspect. Smiling as a visual expression or nonverbal behavior has been the subject of many studies (Ekman et al., 1990; Abel, 2002, etc.). Less is known about smiling as an audible expression during speech and findings have been inconsistent. Tartter (1980) and Tartter & Brown (1994) showed that smiling is audible when it occurs with speech that may be as little as one nonsense syllable. Sentences produced by spreading the lips are perceived as being more "smiled" by listeners (Robson & Mackenzie Beck, 1999); mechanical and amused smiles can be distinguished (Schröder et al., 1998; Aubergé & Cathiard, 2003) as well as other types of smiles (Drahota et al., 2008). Some prosodic cues seem to be universal, while others tend to be culture specific (Émond et al., 2007). Gender appears to play a role in the perception of a smile (Émond, 2008). Moreover, laughing, smiling, and crying speech show perceptual similarities; smiling speech is sometimes judged as being sad (Erickson et al., 2009).

It is difficult to compare the findings of previous studies of smiled speech. Smile and laughter were often considered to be the same expression. Many studies of auditory perception took visual modality into account. However, in the global perception of audiovisual stimuli, the visual modality interferes with the auditory modality, as demonstrated by Aubergé and Cathiard (2003). In addition, studies have lacked descriptions of the communication situation and the smile-elicitation method, which affect the type of smiles that are analyzed. These previous studies all used "lab speech" (which could be controlled) as opposed to "real-life, spontaneous data," and, except for the study by Drahota et al. (2008), the studies all used reading tasks. Furthermore, it is important to note that there was a large amount of between-subject and within-subject variability in these studies. If different kinds of smiles are audible in simultaneous speech without visual cues, this suggests that that it may be possible to associate prosodic correlates with smiled speech.

In this paper, "smiled speech" as opposed to "speech laugh" and as defined by Trouvain (2001, 2003) is analyzed. Speech laugh is a reinforcement of the expiratory activity of speech (e.g., a stronger aspiration during unvoiced speech or a tremor during voiced speech segments), and it is perceptible in two syllables on average. Smiled speech, even though it is perceptible in as little as one syllable, is a longer-lasting event, which can be perceived during a whole utterance and where pitch can be raised.

In the current study, the main challenge was to analyze data from spontaneous smiled speech, with all the naturalness this data contains but also with all the limitations it imposes. The objective was to examine how smiled speech was perceived by different genders and to identify prosodic parameter combinations that would provide a phonetic description of smiled speech.

METHOD

Two perceptual auditory experiments were conducted.

Experiment 1

The aim of this experiment was to evaluate the proportion of smiled speech that listeners can hear without visual cues, based on listener gender and the intensity of the perceived smiles. Reaction times were also investigated.

Stimuli. This study used self recordings from Family 2 from a 1995 spoken corpus of four Montreal families (Vincent et al., 1995). 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 conversation was made by the experimenter, a trained listener. Only the 70 smiled utterances produced by the woman were extracted. Because it is impossible with spontaneous data to have exact non-smiled counterparts of smiled utterances, 70 "neutral-condition" utterances were determined, which had an equivalent number of syllables (1 to 17 syllables, mean of 5.65 syllables) and average length (0.49 to 5.15 seconds, mean of 1.51 seconds) compared with the smiled utterances.

Participants and procedure. Forty native Quebec-French–speaking listeners (20 males, 20 females) with no language, speech, or hearing problems were recruited from different universities 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 39. Parsour software (Bastien et al., 2010) was used. Participants were presented with utterances in a random order via headphones. The listeners

were instructed to determine if these utterances were smiling or not smiling by clicking, as quickly as possible, on the appropriate emoticons. If the utterances were perceived as smiling, the listeners were instructed to indicate the intensity of the perceived smiles by using a computer mouse to click on the appropriate place on a visual analog scale (which consisted of a line with a minus sign on the left-hand side and a plus sign on the right-hand-side). A familiarization task preceded the experiment. The test took about 20 minutes.

Experiment 2

The aim of this experiment was to find perceived prosodic parameter combinations that could describe smiled speech phonetically.

Stimuli. A subset of the spoken corpus that was used in Experiment 1 was used for this experiment. Only the utterances perceived as smiled speech by 75% of males or females (n = 58) were initially selected. In the second step, 27 "neutral utterances" were added. These were perceived as non-smiled speech by at least 95% of males or females. Thus the spoken corpus contained a total of 85 utterances.

Participants and procedure. Forty native Quebec-French–speaking listeners (20 males, 20 females, different from the previous experiment) with no language, speech, or hearing problems were recruited from different universities 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 39. Similarly, Parsour software (Bastien et al., 2010) was used and participants were presented with utterances in a random order via headphones. The listeners were instructed to evaluate the intensity of four voice aspects (i.e., the prosodic parameters of pitch height, pitch range, speech rate, and rhythm) for each utterance on a visual analog scale by placing the mouse cursor on the appropriate place on the scale. The scale consisted of a line with a minus sign (–) and the label "pas du tout – aigüe, mélodique, rapide, rythmée" (not at all – high, melodic, fast, rhythmic) on the left side and a plus sign (+) and the label "très – aigüe, mélodique, rapide, rythmée" (very – high, melodic, fast, rhythmic) on the right side. Before the experiment, examples of every extreme of voice aspect were heard by the listeners. A familiarization task preceded the experiment. The test took about 20 minutes.

This paper presents the results of the first auditory experiment (Experiment 1).

RESULTS

Perceived Smiled Speech

The results for the first perceptual auditory experiment are shown in Table 1. Of the 140 utterances, 2 utterances were removed for technical reasons. Only utterances perceived as smiled speech by 75% or more listeners were analyzed. A total of 11 utterances were perceived as smiled speech by all listeners; 13 utterances were perceived as smiled speech by all males; and 18 utterances were perceived as smiled speech by all females. A total of 50 utterances were perceived as smiled speech by 75% or more listeners; 56 utterances were perceived as smiled speech by 75% or more males; and 51 utterances were perceived as smiled speech by 75% or more females. Thus, smiled speech was audible by most listeners without visual cues. Men perceived slightly more utterances as smiled speech than women.

TABLE 1. Number of utterances perceived as smiled or neutral by 100% of listeners or by 75% or more listeners

Listeners	%	☺	😐
ALL	100	11	0
M	100	13	1
F	100	18	9
ALL	75 +	50	40
M	75 +	56	39
F	75 +	51	53

ALL = all listeners; F = females; M = males; ☺ = smiled speech; 😐 = neutral speech

Reaction Times

The reaction time for each utterance was measured in milliseconds. A one-way ANOVA was conducted on the data to assess the influence of gender on decision-making speed. Table 2 shows the mean reaction times for men, women, and all listeners for smiled speech and neutral speech.

TABLE 2. Mean reaction times

Listeners	Reaction times (ms)	
	☺	😐
ALL	1538	1661
M	1657	1820
F	1405	1525

LL = all listeners; F = females; M = males; ms = milliseconds; ☺ = smiled speech; 😐 = neutral speech

Two observations can be made. First, reaction times were faster for the smiling condition than for the neutral condition for all listeners. Second, for both conditions, reaction times were faster for females than for males. This difference was significant, $F(1,38) = 4.43$, $p < .05$.

Intensity of Perceived Smiled Speech

A visual analog scale was used to measure the intensity of each utterance that was perceived as smiled speech. The visual analog scale consisted of a five-point scale going from the lowest intensity on the farthest left to the highest intensity on the farthest right. The five smiled speech categories were "slightly," "somewhat," "moderately," "quite," and "very" smiling. The smiled speech categories were correlated with the mean reaction times. The results are shown in Figure 1. The x-axis shows the five categories for the different intensities of smiled speech. The left-hand y-axis shows the reaction time (in milliseconds) for perceived smiled utterances. The right-hand y-axis shows the number of perceived smiled utterances. The empty rectangles correspond to the mean reaction times for each smiled speech intensity category; the colored rectangles correspond to the number of perceived smiled utterances for each smiled speech intensity category. The blue rectangles represent male listeners and the pink rectangles represent female listeners. Except for the "quite" and "moderately" categories, the number of perceived smiled speech utterances was greater for men than for women for each category. Figure 1 shows that men and women made quicker decisions (had faster reaction times) for utterances that were associated with high-intensity as opposed to low-intensity smiled speech. In other words, there was a link between the intensity of a perceived smile and the speed in which the smile was perceived.

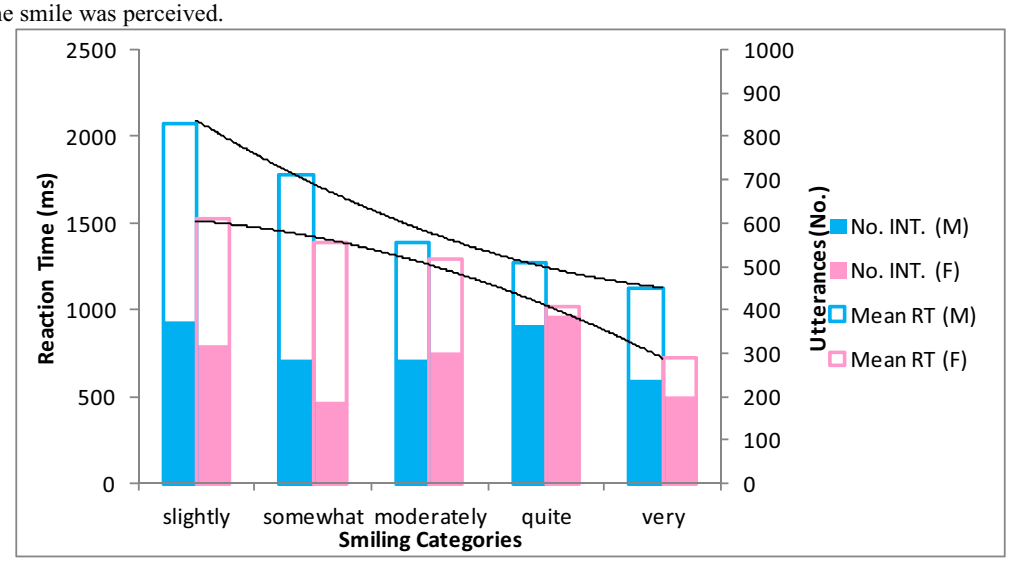

FIGURE 1. Intensity and mean reaction times for perceived smiled speech utterances, for male and female listeners
F = females; M = males; No. INT = number of utterances for each intensity; RT = reaction time

DISCUSSION

As expected, a smile was an audible expression when it occurred with speech. The results showed that most listeners could perceive smiled speech in utterances extracted from spontaneous data, namely a spoken corpus of daily conversation. There was no difference in the number of utterances perceived as smiled by males and females. This finding differs from that of Émond (2008), possibly due to the effect of listening to "spontaneous data" as opposed to listening to data read by a speaker.

The reaction times were also investigated. For the perceived smiling condition, the reaction times were quicker than those for the neutral condition for all listeners. This suggests that the task itself led to quicker reaction times because the participants focused on what they had to find (i.e., the marked condition). For both conditions, reaction times of females were quicker than those of males Males have been shown to be quicker than females in many reaction-time experiments (see Kosinski, 2012 for a review). The current study suggests that mental processes and strategies used to treat linguistic information contained in smiled utterances are different in males and females. Regarding the perception of the smile intensity of smiled utterances, an interesting pattern seems to emerge: The reaction times decreased when smile intensity increased. In other words, if an utterance was perceived to be more "smily," the decision that it was a smiled utterance was made more quickly.

The method used for the second auditory perceptual experiment (Experiment 2) highlights the importance of experiment design when a study aims to establish comparisons with other studies in the field of the prosody of emotional vocal expression, especially smiling. The description of the situation of the communication on the one hand and, the information about the participants (speakers AND listeners) on the other hand are crucial to define all types of smiling speech. Moreover, a tool to analyze prosodic parameters using spontaneous data (e.g., the visual analog scale labels used in Experiment 2) is needed to study speech in its broadest sense. Experiment 2 is currently underway.

ACKNOWLEDGMENTS

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

REFERENCES

Abel, Millicent H. ed. (2002). *An Empirical Reflection on the Smile*, Mellen studies in psychology Series **4** (Lewiston : Edwin Mellen Press) 275 p.

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

Bastien, M., Émond, C. and Ménard, L. (2010). *Parsour Software*.

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

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

Émond, C., Trouvain, J. and Ménard, L. (2007). "Perception of French smiled speech by native vs. non-native listeners: a pilot study," *Proceedings of the Interdisciplinarity Workshop on the Phonetics of Laughter – 16th International Congress of Phonetic Sciences,* edited by J. Trouvain and N. Campbell, 27-30.

Émond, C. (2008). *Les corrélats prosodiques et segmentaux de la parole souriante en français québécois* (MA Thesis, Université du Québec à Montréal, 139 p.).

Erickson D., Menezes, C. and Sakakibara, K. (2009). "Are you laughing, smiling or crying?" *Proceedings of 2009 APSIPA Summit and Conference*, 529-537.

Kosinski, R. J. (2012). "A Litterature Review on Reaction Time," http://biae.clemson.edu/bpc/bp/lab/110.reaction.htm#Mean%20Times (Last viewed 15 jan. 2013).

Robson, J. and Mackenzie Beck, J. (1999). "Hearing smiles – Perceptual, acoustic and production aspects of labial spreading," *Proceedings of 14th International Congress of Phonetic Sciences,* 219-222.

Schröder, M., Aubergé, V. and Cathiard, M.-A. (1998). "Can we hear smile?" *Proceedings of the Conference on Spoken Language Processing* **3**, 559-562.

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

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

Trouvain, J. (2001). "Phonetic Aspects of "Speech-Laughs"," *Proceedings of ORAGE 2001, Orality and Gestuality Conference*, 634-639.

Trouvain, J. (2003). "Segmenting Phonetic Units in Laughter," *Proceedings of 15th International Congress of Phonetic Sciences*, edited by M. J. Solé, D. Recasens et J. Romero, 2793-2796.

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