Acoustic-phonetic properties of smiling revised – measurements on a natural video corpus

Helen Barthel, Hugo Quené

Utrecht Institute of Linguistics OTS, Utrecht University, the Netherlands
h.quene@uu.nl, helen_barthel@yahoo.de

ABSTRACT

Smiling while talking can be perceived not only visually but also audibly. Several acoustic-phonetic properties have been found to cue smiling in the acoustic signal. The aim of this study was to validate properties associated with smiled speech using a natural video corpus. The realisations of monophthongs of the same words spoken with and without a visible smile were compared. The results show a significant increase of intensity (for all words), of F2 (for words with the round vowel /o:/) and of F0 (for all words except the backchannel marker *ja*) in the smiled condition.

Keywords: smiling, spontaneous speech, fundamental frequency, second formant.

1. INTRODUCTION

Smiling contributes to our everyday communication. Smiles are mostly not performed or perceived consciously, but they are nonetheless of great relevance to our social life. Smiling is part of the complex system of nonverbal signals that help interlocutors interpret and understand the smiling speaker. The full range of manifestations and meanings of smiles still has to be explored, although basic smile types and uses are already described in the literature [1]. These classifications and descriptions are focused on visual properties of smiles. However, several studies have shown that smiling is not a solely nonverbal sign, but it is audible as well.

"The acoustic origin of the smile" was first addressed by Ohala [2]. He gives ethological evidence and states that the retraction of the mouth corners was originally used by animals to raise the resonances of the accompanying vocalization. This in turn was to sound smaller and thus convey appeasement and friendliness towards others. A more recent study [3] confirmed Ohala's theory and adds that it is precisely the *effort of sounding smaller* that gives an appeasing impression.

The numerous investigations of smiled speech that have been carried out have used different material and methods and partly gained contradicting results.

Higher fundamental frequency (F0) in smiled speech compared to non-smiled speech was found by Tartter [4] and Lasarcyk and Trouvain [5]. This result was opposed by Drahota et al. [6] and Tartter and Brown [7]. For the latter, whispered speech without F0 was found to be still distinguishable into smiled and non-smiled speech by the listeners, which is a very strong argument against the smile-defining nature of that feature. Drahota et al. [6] found that smiled speech has increased duration and intensity values compared to non-smiled speech. Tartter [4] confirmed an increase of intensity but not of duration.

According to Lasarcyk and Trouvain [5] and Drahota et.al [6] the formants show no considerable differences between smiled and non-smiled speech. However, the latter state that the *perception* of "smileyness" was affected by the formants. The listeners rated a speaker as sounding generally more "smiley" when the difference between first and second formant (F1 and F2) was comparably large and the difference between F2 and F3 comparably small, although no such evidence existed in their acoustical data. In a study by Tartter and Brown [7] the value of F2 increased in most vowels when smiling. Robson and MackenzieBeck [8] confirm the increase of F2 and add higher F3 values to their results. An overall increase of formant dispersion in smile-related words was observed by Quené and Schuerman [9], but only for female speakers. Sex differences appear to play a role for the production of smiles, as smiles are not only used for different expressional purposes [10] but there is also evidence for sex differences in the processing of prosody [11].

The variety of results can partly be explained by the variety of methods and materials, although there are no clear connections between a certain methodology and the results. Some studies used synthetic vowels [3, 5] or re-synthesized materials [12], whereas others used human speech read out [4, 7, 8,] or collected from a corpus including spontaneous speech [9].

However, none of the previous investigations have studied exclusively natural speech material, and this restricts the validity of the phonetic properties that have been reported. The material either consisted of controlled or visually confirmed smiles [3, 5, 6, 7, 8], i.e. read or synthetic speech, or of spontaneous speech recorded only audibly [9] without visual confirmation of smiles. Ideally, materials to investigate the phonetic properties of smiled speech would consist of real-world spontaneous conversations, with video recordings to assess the speakers' visual facial expressions. The aim of this study is to compare the acoustic properties of smiled vs. non-smiled speech in precisely such material.

Based on previous findings, we expect in the smiled speech a relatively high F2 [5, 9] and F3 [9], i.e. a wider dispersion of formants [4], as the main correlates of a visual smile [2]. These expected effects are due to physical changes to the vocal tract, which in turn is caused by the retraction of the corners of the mouth while smiling.

In addition, we expect a relatively high F0, duration and intensity. However, these latter phonetic effects of smiling may be inextricably confounded with those of other communicative functions in spontaneous speech, in particular the effects of emotions on these properties [13, 14].

2. CORPUS AND METHODS

The materials for this study were taken from the IFADV corpus [15]. This corpus was chosen for its spontaneous dialogues, which yield many naturally smiled utterances, and for its high-quality video and audio recordings, allowing both visual annotation and reliable measurement of formants. The corpus consists of 24 dialogues in Dutch between two well-acquainted speakers of a broad age range. The topics of the conversations could freely be chosen by the participants [15]. Speakers show some awareness of the laboratory setting, e.g. glances at the camera or a certain stiffness of the body, but these signs only occur during the first few minutes of each conversation. These first minutes of each recording are not included in the analysis. After that, the participants appear to be relaxed, they stop giggling or just start to gesture more, fumble at their lips and noses, etc. At this point, the conversations can be described as near-natural material.

The recordings last 15 minutes each, and for 20 of them an orthographic annotation is included in the corpus. The participants wore a head-mounted microphone and each were recorded by a camera to their front right [15]. The resulting recordings are of a high quality and the videos show the participant's faces nearly from the front, which is crucial for investigations of facial expressions.

In this investigation we compared smile- and non-smile- realisations of identical monosyllabic word types. Vowel properties were only analyzed for monophthongs, as the measurement of diphthongs entails further issues which cannot be addressed in this study. First, clear stretches of broadly smiled speech and of clearly non-smiled speech were annotated in ELAN (EUDICO Linguistic Annotator, version 4.7.1 [16]). In both cases no other major muscle activity, such as frowning or shrugging, was to be present. For the annotation of a smile the zygomaticus major muscle had to be clearly activated. Slight or fading smiles as well as laughed speech were excluded. Some conversations could not be used for this study, either because there are barely any smiles (dialogues 6 and 10), or because nearly all speech is smiled (dialogues 2 and 4) or because no orthographic transcription was available (dialogues 5, 18, 21, 23). Some speakers also had to be excluded from the analysis because of facial hair (participant W), hoarse voice (participant K) or almost constant smiles or frowns (participants C and X). The smile and non-smile annotations were transferred into the analysis programme Praat (version 5.3.74[17]). The second step of the analysis consisted of a search for orthographic words which were realized by a particular speaker at least once with and once without a smile. As an orthographic annotation already exists for most of the material, this search was conducted automatically. Inflection forms (e.g. <fiets> and <fietsen> for the vowel /iˑ/) were treated as matching material, see [9]. The vowels of the smile and non-smile realisations were measured in terms of their duration and analysed acoustically at their temporal midpoint [9, 18].

The analysis included the measurement of F0, F1 to F5, intensity and formant bandwidths. Eventually, only high-frequent words that occurred at least 10 times for each speaker in each condition were used for the statistical analysis. These words contained the vowels /ɛ, ɪ, iː, ɑ, aː, oː/ and there were overall 1242 items used for statistical analysis.

Each of the resulting acoustic measures of the nuclear vowel (duration in ms, intensity in dB, F0 in semitones relative to 100 Hz, F1 in Bark, F2 in Bark, F3 in Bark) was analysed by means of a separate linear mixed-effects model [19], always with speakers and words as random intercepts, thus capturing the random variation between speakers (due to individual differences in habitual F0 and vocal tract size) and between words (due to vowel phonemes). Facial expression (smiling vs. non-smiling) was included as a fixed factor, and also as random slopes between speakers (i.e., speakers were allowed to vary in their individual effects of smiling). Significance of the facial expression effect was assessed by comparing models using likelihood ratio tests (LRT, α=.05).

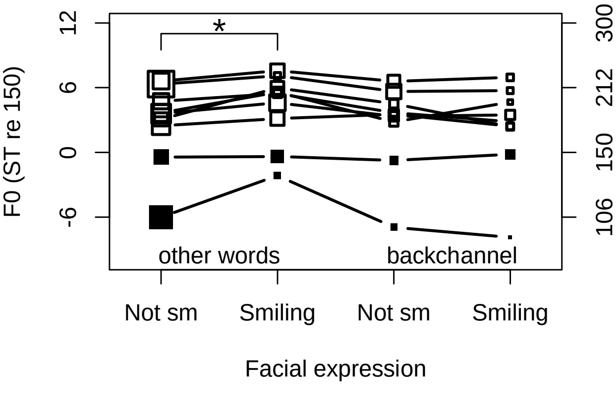
3. RESULTS

For **duration** in ms, the main effect of facial expression was not significant (*β*=2 ms, *t*<1, n.s.; LRT for smiling, *p*=.6439), meaning that the durations of a speaker's nuclear vowel in a word were approximately equal in that speaker's non-smiling and smiling realizations of that word.

For **intensity** in dB, facial expression yielded a significant main effect (*β*=4.3 dB, *t*=7.22, *p*=.0010; LRT for smiling, *p*<.0001). Across speakers and across words, smiling tokens are spoken with larger intensity than non-smiling tokens, by a considerable amount.

For **F0** in semitones, preliminary by-word analyses suggested that smiling increased the F0 for all words, except for the backchannel indicator *ja* ("yes"). This backchannel status (*ja* vs. other words) was therefore included in the LMM. The resulting LMM (LRT for smiling, *p*<.0001) showed that smiling had a significant positive effect for the non-backchannel words (*β*=+1.09 semitones, *t*=2.61, *p*=.0297), but not for the backchannel words (*β*=+0.17 semitone, *t*<1, n.s.), as illustrated in Figure 1.

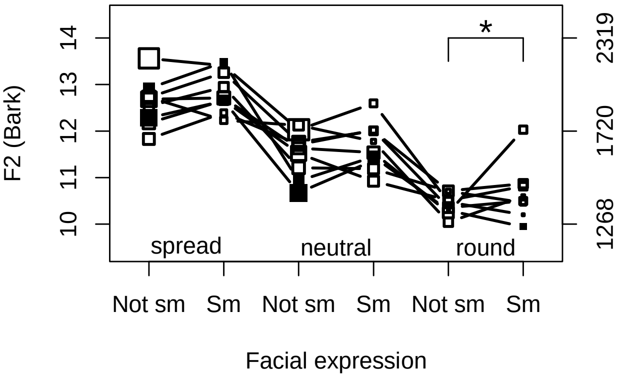
Figure 1: Observed average F0 in semitones (relative to 150 Hz), broken down by speaker sex (open symbols: female, closed symbols: male), speaker, facial expression, and pragmatic word type. Symbol size corresponds to numbers of tokens.



For **F1** in Bark, the main effect of facial expression was not significant (*β*=0.12 Bark, *t*=1.59, *p*=.0935; LRT for smiling, *p*=.1111), meaning that the F1 of a speaker's nuclear vowel in a word were somewhat, but not significantly higher in the smiling realizations than in the non-smiling tokens.

For **F2** in Bark, a preliminary by-word analysis suggested that smiling increased the F2 for words with rounded vowels (*ook*, *zo*) to a larger extent than it did for the other words that have spread or neutral vowels. If this vowel-roundness was included in the model, the resulting LMM (LRT for smiling, *p*=.0033) showed that smiling had a positive but not significant effect on the F2 of neutral vowels (β<0.01 Bark, *t*<1, n.s.) and of spread vowels (β=0.09 Bark, *t*<1, n.s.), whereas smiling yielded a large and significant effect on the F2 of rounded vowels (β=0.50 Bark, *t*=3.04, *p*=.0192), as illustrated in Figure 2.

Figure 2: Observed average F2 in Bark, broken down by speaker sex (open symbols: female, closed symbols: male), speaker, facial expression, and lip rounding of the vowel. Symbol size corresponds to numbers of tokens.



For **F3** in Bark, the vowel roundness was again included in the model, because this roundness is known to influence F3, too [20]. The resulting LMM (LRT for smiling, *p*=.5436) did not show a main effect nor any interaction of facial expression, neither for neutral vowels (β=−0.05 Bark, *t*<1, n.s.), nor for spread vowels (β=−0.02 Bark, *t*<1, n.s.) or round vowels (idem).

4. DISCUSSION

The results show a significant increase of intensity (for all words), of F2 (for words with the round vowel /o:/) and of F0 (for all words except the backchannel marker *ja*) in the smiled condition. The results for F1 and F3 values are not significant.

For the F2, there were no significant results for the spread or neutral vowels /ɛ, ɪ, i:, ɑ, a:/. The reason could be that a smiling gesture does not influence the articulation of these vowels as much, or at least not enough to be acoustically detectable. In agreement with our prediction, however, the articulation of the round vowel /o:/, is considerably affected by the spreading gesture of a concomitant smile, resulting in an F2 which is about 0.5 Bark (or 93 Hz, across speakers) higher in the smiling realizations than in matching non-smiling realizations of the /o:/ vowels. This increase in F2 exceeds the perceptual threshold of 0.3 Bark [21]. Consequently, the pattern of a noticeably higher F2, only for rounded vowels, and absent formant changes for non-round vowels, may well contribute to the perception of smiling in conversational speech [cf. 4, 7, 8, 9].

The results for F0 were significant for all the words except for *ja* ("yes"). This may be explained by the different pragmatic function of *ja* as compared to the other words. The word *ja* served as a backchannel, showing one interlocutor's attention and agreement to what the other was saying. The pragmatic function (backchannel status) was therefore included in the analysis. For the other words, the increase of F0 may be due in part to a somewhat higher position of the larynx while smiling.

The observed increase of F0 and of intensity confirms results of previous studies [4, 5, 6, 9,]. In addition, the joint changes of both phonetic properties in smiled speech indicate that the speakers also vocally express their positive emotions [13, 14] while smiling, and that the vocal expressions of emotions and of smiling may indeed be inextricably confounded in natural conversational speech.

In conclusion, these findings confirm that smiling while talking results in a higher F2 for rounded vowels, and in higher F0 and intensity. Similar findings from controlled studies are validated for natural, conversational speech. The prosodic effects of smiling may also be due to vocal expression of positive emotions. The effect of smiling on F2 is most likely due to interference between lip rounding for rounded vowels, and lip spreading for the smile co-produced with the vowels. In the future, these findings may enable us to assess the occurrence of smiles from phonetic properties of natural speech.

5. ACKNOWLEDGEMENTS

We wish to thank Marcin Włodarczak, for helpful initial discussions, and our thanks for financial support to the Leonardo-Büro Sachsen-Anhalt.

6. REFERENCES

[1] Ekman, P. 2001. Telling lies: clues to deceit in the marketplace, politics, and marriage. New York: W.W. Norton.

[2] Ohala, J. J. 1980. The acoustic origin of the smile. *J. Acoust. Soc. Am.* 68, S33.

[3] Xu, Y., Chuenwattanapranithi, S. 2007. Perceiving anger and joy in speech through the size code. *Proc. 16th Internat. Congress of Phonetic Sciences* Saarbrücken, 2105–2108.

[4] Tartter, V. C. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. *Percept. Psychophys.* 27(1), 24–27.

[5] Lasarcyk, E., Trouvain, J. 2008. Spread lips+ raised larynx+ higher F0= Smiled Speech?-An articulatory synthesis approach. *Proc. ISSP*, 345–348.

[6] Drahota, A., Costall, A., Reddy, V. 2008. The vocal communication of different kinds of smile. *Speech Commun.* 50(4), 278–287.

[7] Tartter, V. C., Braun, D. 1994. Hearing smiles and frowns in normal and whisper registers. *J. Acoust. Soc. Am.* 96(4), 2101-2107.

[8] Robson, J., MackenzieBeck, J. 1999. Hearing smiles-Perceptual, acoustic and production aspects of labial spreading. *Proc. 14th ICPhS* 1, 219–222.

[9] Quené, H., Schuerman, W. 2012. Smile with a smile. *Proc. 13th Annual Conference of the International Speech Communication Association* Portland, 603–606.

[10] Vazire, S., Naumann, L. P., Rentfrow, P. J., Gosling, S. D. 2009. Smiling reflects different emotions in men and women. *Behav. Brain Sci.* 32(5), 403.

[11] Schirmer, A., Kotz, S. A., Friederici, A. D. 2002. Sex differentiates the role of emotional prosody during word processing. *Cogn. Brain Res.* 14(2), 228–233.

[12] Quené, H., Semin, G. R., Foroni, F. 2012. Audible smiles and frowns affect speech comprehension. *Speech Commun.* 54(7), 917–922.

[13] Neuber, B. 2002. Prosodische Formen in Funktion: Leistungen der Suprasegmentalia für das Verstehen, Behalten und die Bedeutungs(re)konstruktion. *Hallesche Schriften zur Sprechwissenschaft und Phonetik* 7. Frankfurt /Main: Peter Lang.

[14] Banse, R., Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70(3), 614-636. doi: 10.1037/0022-3514.70.3.614

[15] van Son, R., Wesseling, W., Sanders, E., van den Heuvel, H. 2008. The IFADV Corpus: a Free Dialog Video Corpus. *LREC*.

[16] ELAN. A professional tool for the creation of complex annotations on video and audio resources. http://www.lat-mpi.eu/tools/elan/.

[17] P. Boersma and D. Weenink. Praat: doing phonetics by computer. http://www.praat.org/.

[18] van Son, R., Pols, L. C. W. 1990. Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *J. Acoust. Soc. Am.* 88(4), 1683–1693.

[19] Quené, H., van den Bergh, H. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *J. Mem. Lang.* 59(4), 413–425.

[20] Pétursson, M., Neppert, J. M. H. 2002. Elementarbuch der Phonetik. Hamburg: Buske.

[21] Kewley-Port, D., Zheng, Y. 1999. Vowel formant discrimination: Towards more ordinary listening conditions. *J. Acoust. Soc. Am.* 106(5), 2945-58.