ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Smile with a smile

Hugo Quené, Will Schuerman

Utrecht institute of Linguistics OTS, Utrecht University, Utrecht, the Netherlands
h.quene@uu.nl, w.l.schuerman@students.uu.nl

Abstract

Smiling during talking yields speech with higher formants, and hence larger formant dispersion. Previous studies have shown that motor resonance during perception of words related to smiling can activate muscles responsible for the smiling action. If word perception causes smiling activation for such smile-related words, then this motor resonance may occur also during production, resulting in larger formant dispersion in these smile-related words. This paper reports on formant measurements from tokens of the Corpus of Spoken Dutch. Formant values of smile-related word tokens were compared to semantically different but phonetically similar word tokens. Results suggest that formant dispersion is indeed larger in smile-related words than in control words, although the predicted difference was observed only for female speakers. These findings suggest that motor resonance originating from a word's meaning may affect the articulatory and acoustic realization of affective spoken words. Female speakers tend to produce the word *smile* with a smile.

Index Terms: smiling; affect; motor resonance; formants; formant dispersion;

10.21437/Interspeech.2012-183

1. Introduction

Smiling during talking effectively reduces the length of the vocal tract [1]. Speech produced with a smile has a higher pitch, as well as higher formants and larger dispersion between formants [2, 3], relative to speech produced with a neutral facial expression. Listeners can hear whether speech is spoken with a smile [2, 3], basing their judgements on F0 and on formant dispersion [4].

This study is grounded in the hypothesis that linguistic and affective processing are not separate but interacting components of spoken language processing. With regard to speech perception, it has been argued that perceiving speech is perceiving articulatory gestures [5, 6]. In addition, language comprehension involves motor resonance of the actions being described in the (linguistic) utterance, such as turning or pushing or pulling [7, 8, 9].

With regard to affect perception, it has similarly been argued that perceiving affective gestures requires motor simulation of these gestures [10, 11, 12, 13, 14, 15]. Reading printed affective words (e.g. *honest, dangerous*) results in motor activation of the muscles involved in the corresponding facial affective expressions, viz. smiles and frowns [8]. Moreover, spoken words are perceived slower if the word's phonetic form (formant dispersion, indicating smile or frown) is inconsistent with the word's affective valence [4], which again indicates that affective resonance contributes to comprehension of spoken language.

If these affective and motor resonance processes do indeed contribute to comprehension, as the evidence discussed above suggests, then the same resonance processes are also expected to contribute to speech production, where motor activation is essential. Thus one might expect that affective words tend to be produced with a consistent facial expression, because the articulatory and affective properties of a word are integrated during speech production. Specifically, the word *smile* may tend to be produced with a congruent smile, the word *frown* with a frown, etc.

Of course, the affective resonance hypothesis does not predict that speakers are forced to mimic these affective facial gestures whenever they produce an affective word. The facial gesture may be blocked by voluntary control (e.g. if the speaker wants to keep a neutral facial expression), and/or by semantic distance between the speaker and the agent, e.g. by third-person perspective and/or negation, as in the utterance *he did not smile*. Nevertheless, given a sufficiently large speech corpus, we predict that smile-related words will tend to have larger formant dispersion, relative to phonetically similar control words, due to (micro-)activation of the muscles involved in smiling gestures [8], and the consequent acoustic effects of lip spreading [1, 2, 3]. This activation presumably results from motor resonance between the smile-related word and the affective smiling gesture.

2. Methods

2.1. Materials

Tokens of spoken words were selected from the Corpus of Spoken Dutch [16], which is available with an orthographic transcription tier for the entire corpus. Tokens of smile-related words were found by searching for the string `glimlach*` (meaning "smile") in the orthography tier. Because of possible differences in vowel quality between the variants of Dutch spoken in the Netherlands and Belgium (Flanders), only the Netherlands Dutch part of the corpus was used. The search yielded 167 word tokens with a meaning related to smiling, both nouns and verbs, in various inflection forms (58 tokens of *glimlach* /xlɪmlɑx/ "smile", 58 *glimlachte*, 16 *glimlachen*, 15 *glimlacht*, and 20 tokens of 6 other forms). Lexical stress falls on the initial syllable in all these word forms.

For comparison, words were also selected that have a similar phoneme sequence /C lɪm/ in their stressed syllable, but whose meaning is unrelated to smiling. Relevant control words were *glim* /xlɪm/ "glimmer, gleam", *klim* /klɪm/ "climb", *glimp* /xlɪmp/ "glimpse", and their derivates. In total 62 word tokens were selected in this word category from the same sub-corpus (41 tokens of 5 forms of *glim*, 3 tokens of *glimp*, 14 tokens of 4 forms of *klim*, 2 tokens of the proper name *Glimmerveen*, and 2 hapax forms).

The 229 selected word tokens had been realized by 136 different speakers (73 female, 63 male). Only 13 speakers (9 female, 4 male) contributed tokens of both smile-related and control words, and this involved only 36 tokens. Table 1 shows how the numbers of tokens are distributed over speakers. For example, there were 2 female and 2 male speakers who each contributed exactly 1 control token and 1 smile-related token to the sample; there were also 2 other male speakers who each contributed 3 control tokens and no smile-related tokens, etc.

Of the 167 smile-related tokens, 100 were from female speakers and only 67 were from male speakers, but this difference was not significant [frequency-weighted goodness-of-fit, $\chi^2(1) = 2.604, p = .107$].

Table 1: *Numbers of speakers, broken down by sex and by numbers of contributed control tokens (rows) and smile-related tokens (columns). Boldface numbers indicate speakers contributing tokens in both categories.*

sex	nr of control tokens	nr of smile-related tokens 0	1	2	3	4	8
F	0		31	7	4	5	1
	1	13	**2**	**4**	**1**		
	2	3	**2**				
	3						
M	0		22	10	4	2	
	1	17	**2**	**1**			
	2	2	**1**				
	3	2					

Each selected word token was excerpted from the corpus and stored as a separate audio file. Speaker identifier code and sex were also retrieved (from the corpus metadata) for each selected word token.

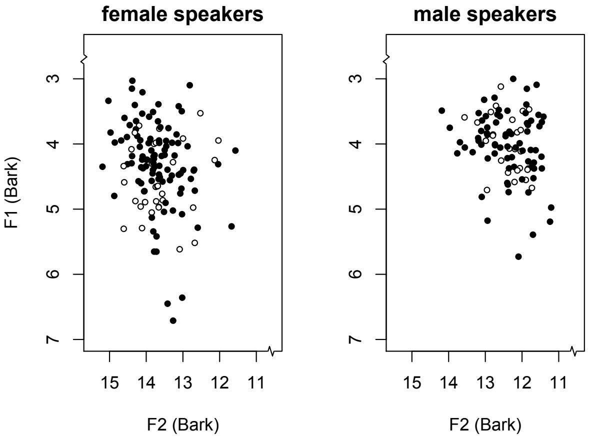

Figure 1: *Formant frequencies of F_2 and F_1 in Bark units, in /ɪ/ vowel tokens, broken down by speaker sex and by semantic category (filled symbols: smile-related, open symbols: control).*

2.2. Formant analysis

Formants were measured in the nuclear vowel /ɪ/ in each word token, using Praat 4.6.36 [17] (Burg method). Because of the large variation in the corpus (e.g. overlap with speech from other speakers), it was not possible to use the same time point for all measurements. The preferred time point of measurement within the vowel was (in decreasing order of preference) (a) the temporal midpoint of the vowel, (b) the point where F_1 is highest, (c) the point of maximal intensity, after [18]. For each vowel the frequencies of formants F_1, F_2 and F_3 were measured and converted to Bark units ([19] formula 6).

3. Results

The observed values for F_1 and F_2 are displayed in Figure 1, broken down by speaker's sex and by the semantic category of each word token. Formant frequencies of the /ɪ/ vowel show considerable overlap between the smile-related words (filled symbols) and control words (open symbols), in particular for male speakers.

The primary dependent measure however was the amount of formant **dispersion**, or difference $F_2 - F_1$ in Bark units, measured in the nuclear /ɪ/ vowel of each word token. The data were analyzed by means of mixed-effects regression, with speakers and tokens-within-speakers as nested random effects [20]. Fixed effects were the semantic category of each word (smile-related vs. control), and the sex of the speaker. The resulting optimal model is summarized in Table 2. (This optimal model did not perform worse than "random slopes" models in which the fixed effects were also allowed to vary between speakers and/or between tokens, according to likelihood ratio tests.) The resulting model confirms that (in control word tokens) the formant dis-

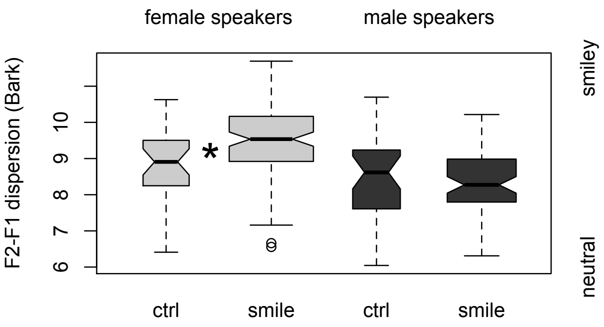

Figure 2: *Boxplots of formant dispersion $F_2 - F_1$ in /ɪ/ vowel tokens, broken down by speaker sex (F: lighter boxes, M: darker boxes) and by semantic category (control words vs smile-related words). Box width corresponds with number of tokens; notches indicate approximate 95% confidence intervals of the box's median.*

persion is smaller for males than for females ($p = .057$). This confirms the expected effects of differences in vocal tract length between females and males. The resulting model also confirms that formant dispersion is indeed larger for smile-related words than for control words, as predicted, but only for female speakers. For male speakers, by contrast, formant dispersion is approximately equal in smile-related words and in control words, yielding a significant interaction effect (see Figure 2).

If the effects of semantic category of the word token are to be compared within speakers only, then the analysis is limited to the 13 speakers who contributed tokens to both semantic categories ($N = 36$ tokens, see Table 1). This sample is so small that none of the effects are remotely significant (all $t < 1$, n.s.), not even the difference in formant dispersion between the sexes.

4. Discussion

A priori, formant dispersion is expected to be larger for (adult) male speakers than for (adult) female speakers, due to sexual differences in vocal tract length. This difference is indeed observed, both for the control words and for the smile-related words. The sex difference is only marginally significant, however, probably due to the large variability between speakers and the very low number of tokens per speaker (see Table 1). The Netherlands part of the Corpus of Spoken Dutch may be insufficiently large for the present study, and it is recommended to replicate this study using larger corpora yielding not only more tokens, but preferably also more tokens per speaker.

For female speakers, the predicted main effect of semantic category is indeed observed. The magnitude of this effect is somewhat smaller than the sex effect, if individual differences between speakers are taken into account (as is done in the analysis summarized in Table 2). Female speakers show a larger dispersion than male speakers in control word tokens, and female speakers also have a larger dispersion in smile-related word tokens than in their control word tokens. If formant dispersion is indeed monotonically related to the amount of smiling, then the larger dispersion for smile-related words suggests that female speakers smile more while saying *smile* than while saying control words. This confirms the main hypothesis in this study, viz. that linguistic and affective processes are interacting during spoken language processing. As the processes resonate in a non-directional fashion, the affective resonance occurs both in speech production, as shown in the present study, and in speech perception [4].

Table 2: *Estimated parameters of mixed-effects model of formant dispersion $F_2 - F_1$ in /ɪ/ vowel tokens. Estimates of fixed parameters are given in Bark units with standard errors and with significance levels based on MCMC simulations. The intercept refers to control words by female speakers. Estimates of random parameters are given in standard deviations of Bark units, with 95% confidence interval of the estimate. $N = 229$; intra-speaker ICC $= 0.48$.*

fixed part			
effect	*estimate*	*s.e.*	*p*
(intercept)	8.99	0.17	.0001
Sex.Male	-0.52	0.25	.0566
Categ.Smile	0.36	0.19	.0032
(interaction)	-0.50	0.28	.0264
random part			
effect	*estimate*	*95% C.I.*	*N*
speaker	0.651	(0.00, 0.277)	136
token	0.680	(0.84, 1.030)	229

The observed phonetic effect of smiling is remarkable, given that formant dispersion was measured in the /ɪ/ vowel only. In Dutch this vowel is typically produced without lip rounding [23], i.e. with (moderate) spreading. In case of the control word tokens, the female speakers seem to have used only a moderate amount of lip spreading; in the case of smile-related word tokens, the female speakers increased the lip spreading somewhat further, yielding larger formant dispersion for smile-related word tokens.

The origin of smiling as an affective gesture may be understood from the so-called Frequency Code [1, 21, 22]. Smiling (i.e. lip spreading) results in higher formants and larger dispersion between formants [2, 3]. For the listener, these acoustic properties suggest that the speaker has a relatively short vocal tract; this seemingly shorter vocal tract in turn suggests a smaller body size, corresponding with a less threatening and more friendly attitude. Speakers seem to be aware that lip spreading, and consequent larger formant dispersion, conveys a smaller body size. When a sample of speakers was instructed to sound "feminine", about half of the female and half of the male speakers reported that they spread their lips to do so [24].

In human social interaction, however, smiles may carry multiple meanings. A smile can be an expression of a friendly attitude, but it can also express joy and/or confidence and/or dominance [15]. Interestingly, males and females seem to differ in the communicative function of the produced smile: according to [25], female smiles tend to convey social warmth and friendliness, whereas male smiles tend to convey self-confidence and lack of distress. This sex difference might explain the observed interaction pattern, where female speakers tend to produce (the Dutch equivalent of) the word *smile* with a smile, whereas there is no such effect for male speakers. In other words, female speakers tend to also execute the affective gesture that they verbally describe, whereas male speakers do not exhibit such affective motor resonance. Thus the resonance patterns seem to differ between male and female speakers. Obviously, further research is necessary to confirm these possible sexual differences in affective resonance between verbal meanings and affective gestures.

In the future we plan to expand this study with other affective verbal expressions for which phonetic effects might be expected, such as lip rounding co-occurring with saying *frown*, and we plan to expand it to other languages for which large speech corpora are available.

5. Conclusion

In conclusion, the present results indicate that for female speakers, the verbal expression of an affective action, i.e. saying the word *smile*, results in larger formant dispersion, which indicates a corresponding smiling action during speech production. Female speakers, but not male speakers, tend to say the word *smile* with a smile, presumably because of affective motor resonance between the smiling gesture being described and being executed.

6. Acknowledgements

The authors wish to express their sincere thanks to Bob Port, Gün Semin, Francesco Foroni and Karin Wagenaar for ideas and discussions, to Sander van der Harst for technical assistance, and to Dagmar van der Voordt for corpus metadata analyses.

7. References

[1] J.J. Ohala, "The acoustic origin of the smile," *J. Acoust. Soc. America*, 1980. Available: http://linguistics.berkeley.edu/˜ohala/papers/smile.pdf

[2] V.C. Tartter and D. Braun, "Hearing smiles and frowns in normal and whisper registers," *J. Acoust. Soc. America*, 96:2101–2107, 1994.

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

[4] H. Quené, G.R. Semin and F. Foroni, "Audible smiles and frowns affect speech comprehension," *Speech Commun.*, 54:917-922, 2012.

[5] A.M. Liberman and I.G. Mattingly, "The motor theory of speech perception revised," *Cognition*, 21:1–36, 1985.

[6] B. Galantucci, C.A. Fowler, and M.T. Turvey, "The motor theory of speech perception reviewed," *Psychonomic Bull. and Rev.*, 13:361–377, 2006.

[7] R.A. Zwaan and L.J. Taylor, "Seeing, acting, understanding: Motor resonance in language comprehension," *J. Exper. Psychology: General*, 135:1–11, 2006.

[8] F. Foroni and G.R. Semin, "Language that puts you in touch with your bodily feelings: The multimodal responsiveness of affective expressions," *Psychol. Science*, 20:974–980, 2009.

[9] B. Tomasino, P.H. Weiss and G.R. Fink, "To move or not to move: imperatives modulate action-related verb processing in the motor system," *Neuroscience*, 169:246–258, 2010.

[10] J.K. Hietanen, V. Surakka, V., and I. Linnankoski, "Facial electromyographic responses to vocal affect expressions," *Psychophysiology*, 35:530–536, 1998.

[11] U. Dimberg, M. Thunberg, and K. Elmehed, "Unconscious facial reactions to emotional facial expressions," *Psychol. Science*, 11(1):86–89, 2000.

[12] V. Gallese, "The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity," *Psychopathology*, 36:171–180, 2003.

[13] P.M. Niedenthal, "Embodying emotion," *Science*, 316(5827):1002–1005, 2007.

[14] V. Gallese, "Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood," *Psychol. Research*, 73:486–498, 2009.

[15] P.M. Niedenthal, M. Mermillod, M. Maringer and U. Hess, "The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression," *Behav. and Brain Sc.*, 33:417–433, 2010.

[16] N. Oostdijk, "The Spoken Dutch Corpus: Overview and first evaluation". In *Proc. LREC-2000, II:887–894*. Available: http://www.lrec-conf.org/proceedings/lrec2000/pdf/110.pdf

[17] Boersma, P., and Weenink, D., Praat: Doing phonetics by computer, 2007. Online: http://www.praat.org

[18] Van Son, R.J.J.H., and Pols, L.C.W., "Formant frequencies of Dutch vowels in a text, read at normal and fast rate", *J. Acoust. Soc. America*, 88:1683–1693, 1990.

[19] Traunmüller, H., "Analytical expressions for the tonotopic sensory scale", *J. Acoust. Soc. America*, 88:97–100, 1990.

[20] H. Quené and H. van den Bergh, "On multi-level modeling of data from repeated measures designs: A tutorial," *Speech Commun.*, 43:103–121, 2004.

[21] J.J. Ohala, "Cross-language use of pitch: an ethological view," *Phonetica*, 40:1–18, 1983.

[22] Y. Xu and A. Kelly, "Perception of anger and happiness from resynthesized speech with size-related manipulations," presented at Speech Prosody 2010, Chicago, 2010. Available: http://www.isca-speech.org/archive/sp2010/papers/sp10_027.pdf

[23] G. Booij, *The Phonology of Dutch*. Oxford: Oxford Univ. Press, 1995.

[24] V. Cartei, H.W. Cowles, and D. Reby, "Spontaneous voice gender imitation abilities in adult speakers," *PLoS ONE*, 7:e31353, 2012.

[25] S. Vazire, L.P. Naumann, P.J. Rentfrow and S.D. Gosling, "Smiling reflects different emotions in men and women," *Behav. and Brain Sc.*, 32:403–405, 2009.