CellPress

Current Biology
Magazine

Correspondence

Auditory smiles trigger unconscious facial imitation

Pablo Arias[1,*], Pascal Belin[2,3,4,*], and Jean-Julien Aucouturier[1,*]

Smiles, produced by the bilateral contraction of the zygomatic major muscles, are one of the most powerful expressions of positive affect and affiliation and also one of the earliest to develop [1]. The perception–action loop responsible for the fast and spontaneous imitation of a smile is considered a core component of social cognition [2]. In humans, social interaction is overwhelmingly vocal, and the visual cues of a smiling face co-occur with audible articulatory changes on the speaking voice [3]. Yet remarkably little is known about how such 'auditory smiles' are processed and reacted to. We have developed a voice transformation technique that selectively simulates the spectral signature of phonation with stretched lips and report here how we have used this technique to study facial reactions to smiled and non-smiled spoken sentences, finding that listeners' zygomatic muscles tracked auditory smile gestures even when they did not consciously detect them.

To build a smiling voice effect, we recorded a corpus of French phonemes spoken by actors with and without smiling (see corpus analysis in the Supplemental Information), and used state-of-the-art audio transformation techniques to model the changes observed in the spectral envelope of these sounds. The algorithm increased the frequency of the first two local resonances (or formants) of the spectral envelope, and increased the amplitude of the third formant. The transformation was implemented using a phase-vocoder architecture [4], which allowed us to manipulate just these smile-specific cues, leaving unchanged other characteristics of the voice such as its pitch, content, speed and gender (Figure 1A; see algorithm design section in the Supplemental Information).

We validated the efficacy and selectivity of the effect in three experiments. First, we asked N = 8 participants to evaluate 20 transformed and non-transformed sentences, and found the manipulation significantly affected their impression of speaker's smiliness (see validation 1 in the Supplemental Information). Second, we asked N = 35 participants to overtly imitate phonemes transformed with the effect, and found they did so with congruent mouth shapes (see validation 2 in the Supplemental Information). Finally, an additional N = 20 participants evaluated the effect on a wider series of emotions/attitudes whose prototypical facial expressions involved a variety of facial action units in the mouth region. The smile effect affected ratings of expressions involving stretched/contracted lips regardless of their emotional valence, for example increasing ratings for both joy and irony, and had no effect on emotions involving other mouth shapes, such as anger or surprise (Figure 1C; see validation 3 in the Supplemental Information).

Using the effect, we then implemented a signal-detection procedure in which N = 35 participants were asked to rate the smiliness of a set of 60 manipulated and non-manipulated spoken sentences, while we monitored their zygomaticus major (involved in smiling) and corrugator supercilii (involved in frowning) muscles with facial electromyography (EMG). We considered as 'hits' trials that consisted of smile-transformed sentences for which smiliness was rated above the sample's median score (median = 4.9). As predicted, participants judged speech with manipulated formants to be more smiling (χ^2 (12) = 66.3, p = 3.9 e^{-15}, Figure 1B), with a hit rate of 63%, and spontaneously responded to manipulated speech with both increased zygomatic (cluster permutation test: t = 1.1–1.9 sec; p = 0.001; d = 0.52) and decreased corrugator activity (t = 0.8–1.6 sec; p = 0.008; d = –0.41; Figure 1D).

Importantly, patterns of EMG activity clearly differed across response categories: zygomatic activity was greater for missed trials than correct rejections, and lower for false alarms than hits (main effect of signal: χ^2 (11) = 6.0, p = 0.01; no effect of response: (χ^2 (11)=2.5, p=0.1). In contrast, corrugator activity was most deactivated for hits and false alarms, and least deactivated for misses and correct rejections (main effect of response: χ^2 (11) = 14.7, p = 0.0001; no effect of signal: χ^2 (11) = 1.4, p = 0.2). In sum, while corrugator activity was entirely explained by participants' judgements, their zygomatic activity continued to track the presence of smile-like spectral cues in speech even when smiles were not consciously recognized (Figure 1E). Two alternative analyses using General Linear Mixed Models (GLMMs) with continuous ratings and Causal Mediation Analysis (CMA) supported the same conclusion (see Supplemental Information for details).

Mimicry, the predisposition to mirror a social partner's facial expression and a plausible basis for the human capacity for empathy, has been almost exclusively studied as a visual-motor process [2]. Using expressive speech, it had been so far difficult to rule out that such reactions, when observed, do not simply follow participants' appraisal of the social or emotional significance of the stimuli [5]. Here, we have introduced a uniquely selective technique to control smile-related cues in running speech, and show that these cues trigger a motor reaction even when smiles are not consciously recognized. These results significantly extend earlier work in vision showing that conscious awareness of a stimulus is not necessary for facial reactions [6] by establishing that, even when stimuli are presented consciously and evaluated explicitly, important aspects of auditory social cognition can still operate on an unconscious level.

Processes underlying these unconscious reactions may include automatic motor-articulatory systems also active for lexical comprehension [7], premotor systems preparing for responsive facial gestures [8], and/or the expressive read-out of emotional appraisal operating implicitly and somehow not reaching conscious awareness [9]. In addition, whether such facial imitation of auditory smiles can also occur spontaneously or automatically, for example in the absence of an explicit evaluation task, is left unresolved here (see Supplemental Information). Even so, these results show that the cognition of smiles is not as deeply rooted in visual processing as previously believed. Beyond smiles, they highlight that the oro-labial characteristics of facial expressions [10] have an important and neglected role in shaping how emotions are signaled vocally.

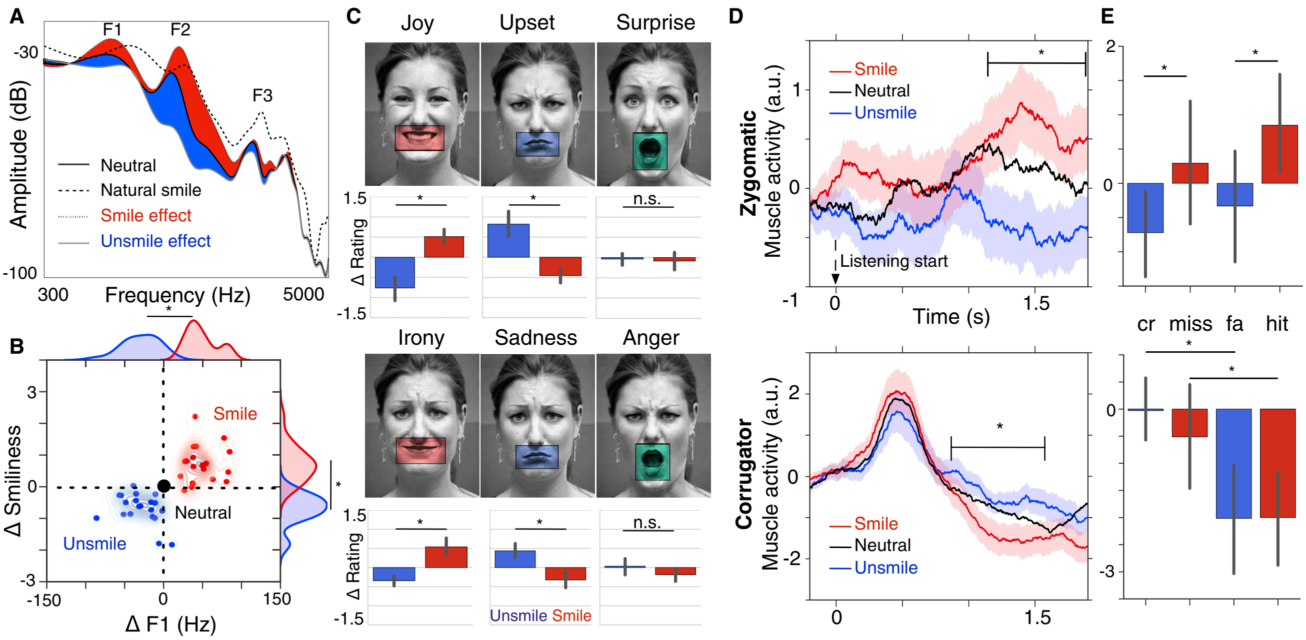

Figure 1. Acoustic, behavioral and electrophysiological consequences of auditory smiles.
(A) Spectral envelopes of neutral (solid bold) and smiled (dotted bold) recordings of a single phoneme, compared with computer-transformed smiled (dotted light) and 'unsmiled' (inverse acoustic transformation, solid light; see algorithm design section in the Supplemental Information) versions of the neutral recording. Red/blue area represent the spectral energy added/suppressed from the spectral envelope with the voice transformation technique. (B) Mean rating of speaker smiliness for smile and unsmile transformations, displayed as a function of computer-generated changes of first formant frequency (all values normalized by corresponding neutral stimuli). Asterisks indicate statistically significant differences. (C) Mean rating of speaker emotion along six emotional/attitudinal dimensions (joy, irony, upset, sadness, surprise, anger) for smile- (red) and unsmile-transformed (blue) versions of 20 sentence stimuli. Ratings normalized by the corresponding non-modified stimuli. Asterisks indicate statistically significant differences; error bars are 95% confidence intervals on the mean. (D) Participants' corrugator and zygomatic EMG activity while rating speaker smiliness for neutral (black), smile- (red) and unsmile-transformed (blue) versions of 20 sentence stimuli, displayed as a function of time. Asterisks indicate time clusters showing statistically significant differences between smile and unsmile conditions; shaded areas represent the standard error on the mean. (E) Mean zygomatic and corrugator EMG activity while rating speaker smiliness, grouped by signal-detection categories: cr, correct rejections; fa, false alarms; smile-transformed categories are in red, unsmile-transformed categories in blue; asterisks indicate significant difference between response categories; error bars are 95% confidence intervals on the mean.

SUPPLEMENTAL INFORMATION

Supplemental Information includes details about algorithm design and validation, experimental procedures and analysis, two figures, and one audio file. It can be found with this article online at https://doi.org/10.1016/j.cub.2018.05.084.

ACKNOWLEDGMENTS

The authors thank Mael Garnotel and Louise Vasa for collecting data, as well as Axel Roebel, Emmanuel Ponsot and Louise Goupil for inputs in algorithm design and data analysis. Experimental data collected at INSEAD/Sorbonne University Center for Behavioural Science. Work funded by ERC StG CREAM 335536 and ANR 2017 REFLETS (to J.-J.A.) and grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI), the Excellence Initiative of Aix-Marseille University (A*MIDEX), and the Institute for Language, Communication and the Brain (to P.B.).

DECLARATION OF INTERESTS

The smile algorithm is patented by the CNRS (patent EP2018/053433).

REFERENCES

1. Darwin, C. (1872). The Expression of the Emotions in Man and Animals. Oxford University Press, USA (1998).
2. Niedenthal, P.M., Mermillod, M., Maringer, M., and Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. Behav. Brain Sci. *33*, 417–433.
3. Drahota, A., Costall, A., and Reddy, V. (2008). The vocal communication of different kinds of smile. Speech Commun. *50*, 278–287.
4. Liuni, M., and Roebel, A. (2013). Phase vocoder and beyond. Musica Tecnologia *7*, 73–120.
5. Hawk, S.T., Fischer, A.H., and Van Kleef, G.A. (2012). Face the noise: embodied responses to nonverbal vocalizations of discrete emotions. J. Pers. Soc. Psychol. *102*, 796.
6. Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. Psychol. Sci. *11*, 86–89.
7. Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. Nat. Rev. Neurosci. *8*, 393–402.
8. Warren, J.E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M.A., Wise, R.J., Rosen, S., and Scott, S.K. (2006). Positive emotions preferentially engage an auditory–motor "mirror" system. J. Neurosci. *26*, 13067–13075.
9. Whalen, P.J., Rauch, S. L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J. Neurosci. *18*, 411–418.
10. Chong, C. S., Kim, J., and Davis, C. (2017). Disgust expressive speech: the acoustic consequences of the facial expression of emotion. Speech Comm. in press.

[1]Sciences et Technologies de la Musique et du Son, UMR 9912, Institut de Recherche et Coordination Acoustique/Musique, CNRS and Sorbonne Université, 75004 Paris, France. [2]Institut de Neurosciences de la Timone, UMR 7289, CNRS and Aix-Marseille Université, 13007 Marseille, France. [3]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QQ, United Kingdom. [4]Département de Psychologie, Université de Montréal, Montréal, QC H3T 1J4, Canada.
*E-mail: arias@ircam.fr (P.A.), pascal.belin@univ-amu.fr (P.B.), aucouturier@gmail.com (J.-J.A.)