Observing Lip and Vertical Larynx Movements During Smiled Speech (and Laughter)

Sascha Fagel[1], Jürgen Trouvain[2], and Eva Lasarcyk[2]
[1]Berlin Institute of Technology, Department for Language and Communication, Germany
[2] Institute of Phonetics, Saarland University, Saarbrücken, Germany
sascha.fagel@tu-berlin.de, {trouvain|evaly}@coli.uni-saarland.de

Abstract

This paper reports on work in progress investigating lip and vertical larynx movements that occur in smiled speech and laughter. A controlled speech corpus in terms of vowels spoken at various degrees of retraction of lip corners is recorded as well as a corpus of induced naturally smiled speech. A combined motion capture and image processing technique is applied to track marked points on the lips and vertical larynx position synchronously while simultaneously recording the audio speech signal. Data will be analyzed to investigate correlations between lip shape and larynx position in smiled speech and to find effects of smiling on audio speech both differentiating between voluntarily spread lips and induced smiles each at various levels of intensity. Some data of spontaneous laughter is included, too.

Introduction

Smiled speech is unlike laughter not an autonomous vocalization but is conveyed by a change of the properties of speech. Several experimental studies showed that listeners are able to distinguish speech with synchronous smiling from speech without smiling (Tartter & Braun 1994; Schröder et al. 1998; Robson & McKenzieBeck 1999; Drahota et al. 2008). Acoustically, smiled speech differs from non-smiled speech with respect to a higher fundamental frequency and higher formant frequencies (F1 to F3). From the speech production perspective, the higher values for F1 to F3 can be explained by lip spreading which results in a shortened vocal tract. This 'i-face' is – in terms of Ohala's (1983) frequency code – a signal for being submissive where the 'o-face' with an enlarged vocal tract signals dominance. However, there is a conflict between reaching the *linguistic* requirements for the articulation of rounded vowels like /u:/ and the *para-linguistic* signaling of smiling while articulating (Lasarcyk & Trouvain 2008).

Effects of smile on speech also depend on the type of smiling. Schröder et al. (1998) have shown that listeners can differentiate between spontaneous and acted smile on the basis of audio-only presentations. In a recent study by Drahota et al. (2008), listeners could discriminate Duchenne smiles (with retracted lip corners and raised cheeks, i.e. "smiling eyes") from non-Duchenne smiles (with retracted lip corners only) on audio only basis. Note that Duchenne smiles are often regarded as "genuine" or "felt" smiles where non-Duchenne smiles as "feigned" smiles (Surakka & Hietanen 1998). The latter type of smiles may partly serve social functions (Ekman 1985) or, more specifically, manage hierarchical social relationships (Mehu & Dunbar 2008). However, there is also evidence that both types of smiles might belong to only one category but with different degrees of positiveness (Messinger et al. 2001, observed in infants from 1 to 6 months of age).

There are reasons to assume that the change of the vocal tract length for smiled speech is performed not only by retracting the mouth corners. The vocal tract may also be shortened by raising the larynx. With an articulatory speech synthesizer (Birkholz 2005) Lasarcyk & Trouvain (2008) have shown that speech with raised larynx is perceived as smiled even when the lips are not retracted. Instrumental methods for observing the activities of lips and larynx comprise video (Hoole & Kroos 1998), opto-electronics (Kakita & Hiki 1976), mechanics (Gandour & Maddieson 1976), X-ray videos (Perkell 1969) and MRI (Honda et al. 1999). Where recording X-ray is connected to an irresponsible health risk and MRI to date does not allow for reasonable frame rates, video recordings offer the opportunity to acquire the vertical larynx position non-invasively. However, this approach is limited to the acquisition of the vertical larynx position in subjects with well visible Adam's apple. The authors develop a measurement system comparable to that of Hoole & Kroos (1998) with the extension of calibrating the method by the use of ground truth data derived from existing X-ray videos.

Method

To our knowledge there is no study to date that investigates both mechanisms (presumably) active for smiled speech – lip corner retraction and vertical larynx height – at the same time. We combine optical motion capture and image processing to measure both activities simultaneously. Blue beads are attached to the lips, cheeks, forehead and décolleté of the subject. Three cameras (DragonflyExpress, Point Grey Research) with a spatial resolution of 640 x 480 pixels at 60 full frames per second record the speaking subject from different perspectives enabling the automatic computation of the 3D positions of the beads in each frame. A fourth camera of the same type records the neck region in a profile view against a black background to facilitate the extraction of the contour of the neck and the automatic computation of the vertical larynx position relative to the collarbone. Audio is recorded simultaneously.

Two kinds of material are used. For the controlled material the subject reproduced the sustained vowels [i: y: a: u:] as a sequence of isolated vowels in each of three conditions: i) neutral, ii) with slightly retracted mouth corners, and iii) with maximally retracted mouth corners. The last two conditions are supposed to result in mechanical smiles of different intensity. Controlled material also included swallowing and inhaling, i.e. activities with more extreme larynx movements. Inhalation noises possibly play a role for laughter. In addition to the controlled material spontaneous laughter was recorded as well (elicited by a cheerful atmosphere).

The preliminary results based on eyeballing show two main problems with the larynx observations: i) Though the male subject has a visible adam's apple in general, the larynx-induced bulge in the neck contour was not sufficiently visible during vowel articulation, ii) larynx movement was frequently combined with fronting and tilting of the head as well as raising of the thorax making contour distinctions of the laryngeal bulge nearly impossible.

For the phonologically rounded vowels [y: u:] we could observe considerably less protrusion and more mouth corner retraction during mechanical smiles than for neutral articulation which is in partial contrast to the perceptual results in Lasarcyk & Trouvain (2008). Another problem is that the beads at the outer border of the vermillion – at the mouth corners and on the upper and lower lips – can only limitedly predict the inner lip contour. But it is the lip opening area surrounded by the inner lip contour that is acoustically relevant. Furthermore, the protrusion itself seems to manifest differently in a setting with pre-retracted lip corners compared to a neutral setting, resulting in varying shape and thickness of the lips.

References

Birkholz, P. 2005. 3D-Artikulatorische Sprachsynthese. Ph.D. thesis, University of Rostock.

Drahota A, Costall A. & Reddy V. 2008. The vocal communication of different kinds of smile. Speech Communication, 51(4):278–287.

Ekman, P. 1985. Telling Lies. W.W. Norton, New York.

Gandour, J. & Maddieson, I. 1976. Measuring larynx height in Standard Thai using the cricothyrometer. Phonetica 33:241-267.

Honda, K., Hirai, H., Masaki, S. & Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. Lang Speech 42(4):401-11.

Hoole, P. & Kroos, C. 1998. Control of larynx height in vowel production". Proc. ICSLP, 531-534.

Kakita, Y. & Hiki, S. 1976. Investigation of laryngeal control in speech by use of thyrometer. JASA 59: 669-74.

Lasarcyk E, Trouvain J, 2008. Spread Lips + Raised Larynx + Higher F0 = Smiled Speech? - An Articulatory Synthesis Approach. Proc. ISSP, Strasbourg.

Mehu, M. & Dunbar, R. 2008. Relationship between Smiling and Laughter in Humans (Homo sapiens): Testing the Power Asymmetry Hypothesis. Folia Primatol 79:269-280.

Messinger, D.S., Fogel, A., & Dickson, K.L. 2001. All Smiles Are Positive, but Some Smiles Are More Positive than Others. Developmental Psychology 37(5):642-653.

Ohala, J.J. 1983. An ethological perspective on common cross-language utilization of F0 in voice. Phonetica, 41:1–16.

Perkell, J.S 1969. Physiology of Speech Production: Results and Implications of a Cineradiographic Study. Cambridge, Massachusetts: MIT Press.

Robson, J. & MackenzieBeck, J. 1999. Hearing smiles – Perceptual, acoustic and production aspects of labial spreading. Proc. ICPhS San Francisco, 219–222.

Schröder, M. & Aubergé V. & Cathiard, M.-A. 1998. Can we hear smiles? Proc. ICSLP, Sydney.

Surakka, V. & Hietanen, J. K. 1998. Facial and emotional reactions to Duchenne and non Duchenne smiles. International Journal of Psychophysiology, 29(1): 23-33.

Tartter V.C. & Brown D. 1994. Hearing smiles and frowns in normal and whisper registers. JASA 96: 2101–2107.