Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 75 (2015) 1–13

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Examining visible articulatory features in clear and plain speech

Lisa Y.W. Tang [a], Beverly Hannah [b], Allard Jongman [c,*], Joan Sereno [c], Yue Wang [b], Ghassan Hamarneh [a]

[a] *Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, 9400 TASC1, 8888 University Drive, Burnaby, BC V5A 1S6, Canada*

[b] *Language and Brain Lab, Department of Linguistics, Simon Fraser University, 6203 Robert C. Brown Hall, 8888 University Drive, Burnaby, BC V5A 1S6, Canada*

[c] *KU Phonetics and Psycholinguistics Lab, Department of Linguistics, University of Kansas, 15 Blake Hall, Lawrence, KS 66045-3129, USA*

Received 11 May 2015; received in revised form 13 September 2015; accepted 16 September 2015
Available online 28 September 2015

Abstract

This study investigated the relationship between clearly produced and plain citation form speech styles and motion of visible articulators. Using state-of-the-art computer-vision and image processing techniques, we examined both front and side view videos of speakers' faces while they recited six English words (keyed, kid, cod, cud, cooed, could) containing various vowels differing in visible articulatory features (e.g., lip spreading, lip rounding, jaw displacement), and extracted measurements corresponding to the lip and jaw movements. We compared these measurements in clear and plain speech produced by 18 native English speakers. Based on statistical analyses, we found significant effects of speech style as well as speaker gender and saliency of visual speech cues. Compared to plain speech, we found in clear speech longer duration, greater vertical lip stretch and jaw displacement across vowels, greater horizontal lip stretch for front unrounded vowels, and greater degree of lip rounding and protrusion for rounded vowels. Additionally, greater plain-to-clear speech modifications were found for male speakers than female speakers. These articulatory movement data demonstrate that speakers modify their speech productions in response to communicative needs in different speech contexts. These results also establish the feasibility of utilizing novel computerized facial detection techniques to measure articulatory movements.
© 2015 Elsevier B.V. All rights reserved.

Keywords: Articulation; Clear speech; English vowels; Computational methods; Facial landmark detection

1. Introduction

Previous research has established that the movements of facial articulatory features contribute to the myriad of cues generated during speech (Gagné et al., 2002; Mixdorff et al., 2005; Smith and Burnham, 2012; Tasko and Greilick, 2010). The current study explores how visual cues generated by the visible articulatory movements of the lips and facial muscles are deployed by speakers during production of different speech styles (clearly produced and plain citation form), by utilizing novel computerized facial detection techniques to measure differences in articulatory movements during clear versus plain speech tokens of English tense and lax vowels embedded in /kVd/ contexts.

1.1. Audio-visual speech perception

Research has demonstrated that bimodal (auditory and visual, AV) perception is superior to auditory-only (AO) perception of speech (Massaro, 1987; Sumby and Pollack,

* Corresponding author at: 422 Blake Hall, Lawrence, KS 66045-3129, USA. Tel.: +1 785 864 2384.
E-mail addresses: lisat@sfu.ca (L.Y.W. Tang), beverly_hannah@sfu.ca (B. Hannah), jongman@ku.edu (A. Jongman), sereno@ku.edu (J. Sereno), yuew@sfu.ca (Y. Wang), hamarneh@sfu.ca (G. Hamarneh).

http://dx.doi.org/10.1016/j.specom.2015.09.008
0167-6393/© 2015 Elsevier B.V. All rights reserved.

1954; Summerfield, 1979, 1992). This is presumably due to the additional stream of linguistic information available to the perceiver in the visible articulatory movements of the speaker's lips, teeth, and tongue as useful sources for segmental perception (Kim and Davis, 2014b; Tasko and Greilick, 2010; Traunmüller and Öhrström, 2007). Additionally, visual cues from movements of facial features including the eyebrows, neck, and head may contribute to the perception of prosodic information such as lexical tone, stress, and focus (Chen and Massaro, 2008; Cvejic et al., 2012; Krahmer and Swerts, 2007; Smith and Burnham, 2012).

Further findings reveal that the weight granted to visual cues depends on the relative availability and accessibility of the visual (relative to auditory) information, which is affected by factors such as the visual saliency of articulatory input, the quality of auditory input, and the condition of perceivers. For example, perceivers are found to put more weight on the visual input for rounded vowels than for open vowels, as lip-rounding is more visually salient to uniquely characterize rounded segments than the generic mouth opening gesture (Traunmüller and Öhrström, 2007). Likewise, perceivers are more accurate in identifying speech contrasts with more visible articulatory gestures (e.g., labial/labio-dental /p-f/) compared to those with less visible ones (e.g., alveolar /l-ɹ/) (Hazan et al., 2006). Moreover, research has shown that visual information enhances speech perception when auditory environment is degraded, such as in a noisy environment (Bernstein et al., 2004; Hazan et al., 2010; Sumby and Pollack, 1954; Summerfield, 1979). Visual input has been found to particularly benefit special populations for whom the auditory speech distinctiveness is challenging or unfamiliar, such as hearing-impaired or non-native perceivers (Grant and Seitz, 1998; Sekiyama and Tohkura, 1993; Smith and Burnham, 2012; Wang et al., 2009, 2008). These findings clearly demonstrate that visible articulatory information can provide reliable cues to facilitate speech perception.

1.2. Clear speech

With the goal of increasing their intelligibility, speakers may alter their speech productions in response to the communicative needs of perceivers (Hazan and Baker, 2011; Kim et al., 2011; Smiljanić and Bradlow, 2009; Tasko and Greilick, 2010), such as when speaking in the presence of background noise (Sumby and Pollack, 1954), competing with other talkers (Lu and Cooke, 2008), or communicating with the hearing-impaired or non-native perceivers (Ferguson, 2012; Maniwa et al., 2009; Payton et al., 1994; Picheny et al., 1986). Such accommodations typically involve clear speech, a clarified, hyperarticulated speech style, relative to the natural plain, conversational speech style.

Acoustic measures show that plain-to-clear speech modifications of English vowels may involve increased duration, intensity, fundamental frequency value and range, formant frequency range and distance, and expanded vowel space (Bradlow et al., 1996; Ferguson, 2012; Ferguson and Kewley-Port, 2007, 2002; Ferguson and Quené, 2014; Hazan and Baker, 2011; Lam et al., 2012); as well as more dynamic spectral and temporal changes (Ferguson and Kewley-Port, 2007; Tasko and Greilick, 2010. The clear speech strategies found to be most effective in contributing to intelligibility are the expansion of the vowel space (and corresponding formant changes) and increased duration of vowels (Bond and Moore, 1994; Bradlow, 2002; Ferguson and Kewley-Port, 2007, 2002; Picheny et al., 1986). More specifically, compared to conversational speech, clear speech involves lower second formant for back vowels and higher second formant for front vowels, as well as higher first formant for all vowels, which presumably could be attributed to more extreme articulatory movements and longer articulatory excursions involving a higher degree of mouth opening and jaw lowering (Ferguson and Kewley-Port, 2007, 2002).

Moreover, there is evidence that clear speech vowel characteristics may interact with vowel tensity, with more expanded vowel space and longer duration for tense vowels than for lax vowels in clear speech (Picheny et al., 1986; Smiljanić and Bradlow, 2009). However, such evidence either lacks statistical power (Picheny et al., 1986), or is only restricted to the temporal domain (Smiljanić and Bradlow, 2008). Additionally, despite the fact that both tense and clear vowels bear similar acoustic correlates, the two factors are not cumulative to further enhance intelligibility (Ferguson and Quené, 2014). Further research is needed to examine the extent to which such acoustic effects, if any, are salient in articulation.

1.3. Articulatory features in clear speech

Given that acoustic variations in clear speech may be triggered by alterations in articulatory features, it is conceivable that such articulatory variations are measurable and can be perceived to aid intelligibility. It has been shown that the clear speech strategies that speakers adopt when conversing with normal as well as hearing-impaired persons in noisy settings may further enhance intelligibility when presented in both auditory and visual modalities as compared to audio-only presentation (Gagné et al., 2002; Sumby and Pollack, 1954). Furthermore, research has demonstrated that the benefits accrued from the availability of both visual information and a clear speaking style are complementary and not merely redundant sources of additional information in improving intelligibility over the auditory-only conversational style condition (Helfer, 1997).

The few studies that have performed such kinematic measurements showed positive correlations among articulation, acoustics, and intelligibility measures in clear speech effects (Kim and Davis, 2014a; Kim et al., 2014, 2011; Tasko and Greilick, 2010). Kim et al. (2011) used an Optotrak system to track the articulatory movements of clear speech produced in the presence of background noise

Download English Version:

https://daneshyari.com/en/article/566021

Download Persian Version:

https://daneshyari.com/article/566021

Daneshyari.com