



Differences in acoustic vowel space and the perception of speech tempo



Melanie Weirich*, Adrian P. Simpson

Friedrich-Schiller-Universität Jena, 07737 Jena, Germany

ARTICLE INFO

Article history:

Received 10 September 2013

Received in revised form

18 December 2013

Accepted 2 January 2014

Available online 24 January 2014

ABSTRACT

Despite various studies describing longer segment durations and slower speaking rates in females than males, there appears to be a stereotype of women speaking faster than men. To investigate the mismatch between empirical evidence and this widespread stereotype, listening experiments were conducted to test whether a relationship between perceived tempo and acoustic vowel space size might exist. If a speaker traverses a larger acoustic vowel space than another speaker within the same time then this speaker might be perceived as speaking faster. To test this, two listening experiments with either exclusively female or male speakers but with varying vowel space sizes were conducted. Listeners were asked to rate the perceived speech tempo of same-sex speaker pairs. The stimuli were manipulated to have the same segment durations and *f0* contour. Results indicate that a positive correlation between acoustic vowel space size and perceived speech tempo exists. Since females exhibit on average a larger acoustic vowel space than males, it is suggested that the stereotype of faster speaking women might arise from this.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

When people are asked who they think speak faster, women or men, the overwhelming general tendency is to name women. This stereotype is quite pervasive and it is spread both by men and women¹ and has been given credence in popular scientific literature such as Brizendine (2006), who makes the unsubstantiated claim women (and girls) speak on average not only more but also faster than males. However, empirical reality is different. Several acoustic studies have found that male speakers show more vowel reduction and elision, shorter sentence and sound durations, and thus actually speak faster than females (e.g. Byrd (1992), Ericsson and Ericsson (2001), Simpson and Ericsson (2003) and Whiteside (1996) for read speech; Simpson (1998) for read and spontaneous speech). In the comprehensive acoustic study of Byrd (1992) using the TIMIT database of read speech, 630 male and female American English speakers are investigated with regard to several temporal parameters. The study revealed longer sentence durations for women, more realizations of stop releases sentence finally and fewer vowel reductions. The results of the smaller investigation by Whiteside (1996) of three male and three female British English speakers are in line with this: she found more pauses, more final lengthening, longer sentence durations, fewer vowel reductions and elisions, and a slower speaking rate for the females. With respect to vowels, Simpson (1998) found in his acoustic investigation of 54 male and female German speakers a systematic sex-specific difference in duration: females produce vowels with longer durations than males. While most studies indeed investigated read speech, the analysis of Simpson (1998) included both read and spontaneous speech. Comparable sex-specific durational differences were found in both speech styles, arguing against the suggestion that women speak slower than men only in lab conditions while reading aloud. Moreover, the studies of Ericsson and Ericsson (2001) and Simpson and Ericsson (2003) revealed that stress plays a role in this sex-specific difference. While longer female durations were found for the stressed vowels, this was not the case in unstressed position. In addition, females were found to produce a greater vowel duration contrast between the same vowel in stressed and unstressed position.

These studies have shown that the question no longer should be, who speaks faster, females or males, but rather why does the common and widespread belief of faster speaking women still exist? What are the reasons for this stereotype to persist in spite of the contradictory results in various studies? However, the aim of this study is not to investigate the underlying stereotype, but to inspect potential spectral measures which influence the perception of speech tempo and therefore might account for the persistence of this stereotype. To answer this question we should look first at the acoustic parameters that have been found to correlate with speech tempo.

Trouvain (2004) lists temporal patterns such as the duration and number of pauses, segment and syllable duration and the number of syllables per second. Lindblom (1963) predicts that there is a relationship between tempo and acoustic and articulatory undershoot. More recent studies have

* Corresponding author. Tel.: +49 3641 94 43 33; fax: +49 3641 94 43 32.

E-mail addresses: melanie.weirich@uni-jena.de (M. Weirich), adrian.simpson@uni-jena.de (A.P. Simpson).

¹ Readers can verify this for themselves by googling a phrase such as “women speak faster than men” and they will typically arrive at links such as the following: http://learnbodylanguage.org/flirting_talk.html, <http://answers.yahoo.com/question/index?qid=20110923143319AAMPevl>.

confirmed the correlation between vowel space size and speaking rate that is also related to speech intelligibility (Turner, Tjaden, & Weismer, 1995). Turner et al. investigated the effect of different speech rates on vowel space area and speech intelligibility in read speech of nine subjects with amyotrophic lateral sclerosis (ALS) and nine age- and gender-matched controls. They found that vowel space size varied with speaking rate (more systematically in the control group) and accounted for 45% of the variance in speech intelligibility.

However, since the stereotypical opinion of faster speaking women than men is based on the *perception* of speakers or genders, the most interesting parameters might be the ones correlating with *perceived* speech tempo independent of or relating in a more abstract fashion to the underlying actual durational patterns. The early study of Lehiste (1976) showed that the perception of duration and tempo is influenced by various parameters, such as the position of a stimulus in a stimulus pair and fundamental frequency (f_0). She found that when two stimuli produced at a monotone and having the same duration, the second is perceived as being shorter. This is explained by the fact that perception is strongly dependent on listeners' (linguistic) expectations. English listeners expect final lengthening. If this is not perceived, the stimulus is rated as shorter than the one before even if it has the same duration. The second factor she found influencing perceived speech tempo was the fundamental frequency contour. An /a/ with a falling or rising f_0 contour was perceived as being longer than an /a/ with a monotonous f_0 contour. More recent studies have confirmed these findings and added further parameters crucial for the perception of tempo (Bond & Feldstein, 1982; Cumming, 2011; Kohler, 1986; Pfitzinger, 2001). Bond and Feldstein (1982) investigated the effect of frequency and intensity on perceived speech rate. Listeners were asked to rate perceived speech rate, pitch, loudness, and perceived duration of electronically altered sets of spontaneous speech with a duration of 20 s each but varying frequency and intensity. Both parameters showed an effect: a higher speech tempo was perceived with an increase in f_0 and (although to a lesser degree) also with an increase in intensity.

Cumming (2011) gives an overview of studies investigating the influence of dynamic f_0 on the perception of duration. Remarkably, results differ in terms of the potential effect. Possible reasons the author names are differences in the listeners' native language and in the nature of the stimuli used. To evaluate the different factors she studied speakers of Swiss German, Swiss French and French. Moreover, the stimuli also included non-linguistic stimuli (i.e. buzzes) and varied in terms of the direction of the f_0 movement. Results revealed a significant lengthening effect of a dynamic f_0 in all language groups and for both linguistic and non-linguistic stimuli, but with the effect being stronger for the words than for the buzzes. Regarding the direction and nature of the contours, results indicated that rises show a weaker lengthening effect than falls and complex contours have a stronger effect than simple rises pointing again to the influence of spectral movement and not only absolute differences in level f_0 . In summary, the results point to a cross-linguistic and general effect not limited to a certain language or only linguistically relevant stimuli. In other words, if more (in terms of spectral information) happens, the listener reasons it has to take longer compared to a less complex event. With respect to perceived speech tempo we can further hypothesize that – other things being equal – the tempo of a more complex spectral event is expected to be greater than a less complex one if the duration remains the same. From these considerations we can hypothesize that spectral events unfolding over time other than a dynamic f_0 might have the same effect. One possible aspect might be the acoustic vowel space a speaker traverses during speech. If a stimulus with a moving f_0 contour is perceived as being longer than the same stimulus with a monotonous f_0 contour, then a speaker traversing a large acoustic space in the same time as a speaker traversing a small acoustic vowel space might be perceived to be speaking faster. Taken one step further, given that the average female vowel space is larger than the average male vowel space (in terms of the polygon spanned by F_1 and F_2), females may be perceived to be speaking faster than males, even if measurable duration patterns are the same or indeed even point in the other direction.

1.1. Acoustic vowel space

It has been found that females on average have a larger acoustic vowel space than males cross-linguistically (see e.g. Diehl, Lindblom, Hoemeke, and Fahey (1996) and Hillenbrand, Getty, Clark, and Wheeler (1995) for American English; Whiteside (2001) for British English; Simpson and Ericsson (2007) for German). Since this difference is crucial for the present study, we also measured the vowel spaces of our participating subjects (see Section 2.1 for more information on the speakers). Fig. 1 shows a plot of the 20 female and 20 male speakers used in the present study. The first and second formants were measured within the stable part in the middle of the vowels /i:/ ϵ a u/ contained in the sentence 'Wie lang hat es denn gedauert?' ('How long did it take?'). As the figure clearly shows the area of the polygon spanned by F_1 and F_2 (in Hz) is much bigger for the females (gray symbols) than for the males (black symbols). Furthermore it is evident that women and men differ especially in F_2 of /i:/ and F_1 of /a/. This is in line with the earlier studies showing the greatest sex-specific differences in vowels with high F_1 and F_2 values (Diehl et al., 1996; Hillenbrand et al., 1995; Simpson & Ericsson, 2007; Whiteside, 2001).

Several reasons have been offered to account for this sex-specific difference including physiological, behavioral and perceptual factors (see Simpson (2009) for a comprehensive overview). A bio-physical reason discussed is the non-uniform difference between females and males in the relation between oral and pharyngeal cavity length (Fant, 1966; Winkler, Fuchs, & Perrier, 2006). In terms of behavioral factors females are claimed to

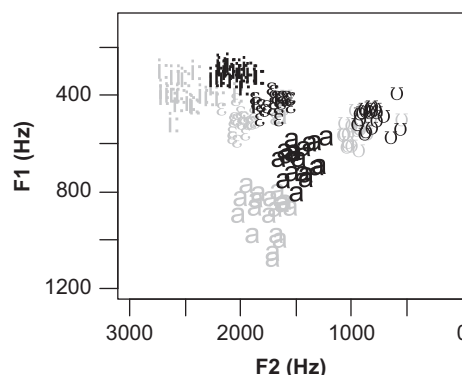


Fig. 1. Vowel spaces from male (black) and female (gray) speakers, measured in the vowels /i: ϵ a u/ within the sentence 'Wie lang hat es denn gedauert'.

Download English Version:

<https://daneshyari.com/en/article/1100759>

Download Persian Version:

<https://daneshyari.com/article/1100759>

[Daneshyari.com](https://daneshyari.com)