Journal of Phonetics 71 (2018) 35–50

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Individual differences in acoustic and articulatory undershoot in a German diphthong – Variation between male and female speakers

Melanie Weirich *, Adrian P. Simpson

Institute for German Linguistics, Friedrich Schiller University Jena, Germany

A R T I C L E   I N F O

Article history:
Received 16 April 2017
Received in revised form 24 May 2018
Accepted 24 July 2018

Keywords:
Vowel undershoot
Diphthong
Inter-speaker variation
Gender-specific differences
Acoustic-articulatory relationship
German
EMA

A B S T R A C T

Individual differences in speech production and, more specifically, in the realization of stress contrasts have been found previously (e.g. de Jong, 1995). This study extends this line of work by investigating potential gender-specific differences in the realization of different accent conditions and more specifically in the degree of undershoot. The reason suggested for these differences is the under-exploitation of the larger male articulatory space during running speech. Differences between male and female speakers in undershoot are investigated (a) by comparing the degree of undershoot in various accent conditions between male and female diphthong productions, and (b) by analyzing the degree of undershoot in relation to a speaker's maximum articulatory vowel space. Articulatory and acoustic data from 11 German speakers (5 males, 6 females) of the diphthong /aɪ/ were analyzed in absolute terms and after normalization for a speaker's maximal articulatory space. In addition to speaker-specific differences in undershoot and in the acoustic-articulatory relationship, results support gender-specific differences, with males exhibiting more undershoot than females in both articulatory and acoustic terms. After normalization with respect to a speaker's maximum articulatory vowel space, females exhibit larger tongue back trajectories than males.

1. Introduction

1.1. Gender-specific differences in speech

Gender-specific[1] differences in speech have been found cross-linguistically both in terms of articulation (Simpson, 2001, 2002, 2003; Weirich & Simpson, 2014a; Weirich, Fuchs, Simpson, Winkler, & Perrier, 2016) and acoustics (Byrd, 1992, 1994; Diehl, Lindblom, Hoemeke, & Fahey, 1996; Ericsdotter & Ericsson, 2001; Henton, 1995; Henton & Bladon, 1985; Hillenbrand, Getty, Clark, & Wheeler, 1995; Simpson & Ericsdotter, 2003, 2007; Titze, 1989; Whiteside, 1996, 2001; Wassink, 1999; Weirich & Simpson, 2014b, 2015a, 2015b). In particular, men and women have been found to differ in acoustic source parameters such as voice quality (Henton & Bladon, 1985; Titze, 1989), average fundamental frequency and variation of fundamental frequency (Boë, Contini, & Rakotofiringa, 1975; Haan & van Heuven, 1999; Stevens, 1998; Takefuta, Jancosek, & Brunt, 1972), as well as aspects of the acoustic filter such as sibilant characteristics (Weirich & Simpson, 2015a) and acoustic vowel space size (Hillenbrand et al., 1995; Peterson & Barney, 1952; Simpson & Ericsdotter, 2007; Whiteside, 2001; Weirich & Simpson, 2014b). For example, a higher average fundamental frequency (e.g. Stevens, 1998), larger f0 excursions (e.g. Haan & van Heuven, 1999), higher Center of Gravity (COG) values in /s/ (e.g. Weirich & Simpson, 2015a) and larger vowel spaces (e.g. Hillenbrand et al., 1995) are found in females than in males. The difference in vowel space size is mainly due to females having higher F1 values for open vowels and higher F2 values for closer front vowels than males (found for several languages including Dutch, English, French, German and Swedish, c.f. Henton, 1995). In addition, temporal parameters such as segment duration or speech rate have revealed gender-specific differences, e.g. for English (Byrd, 1992; Diehl et al., 1996; Hillenbrand et al., 1995; Henton, 1995; Whiteside, 1996), Swedish (Ericsdotter & Ericsson, 2001), Jamaican English (Wassink, 1999), Creek

* Corresponding author at: Institute for German Linguistics, Friedrich Schiller University Jena, Fürstengraben 30, 10117 Jena, Germany. Fax: +49 3641 944332.

E-mail address: melanie.weirich@uni-jena.de (M. Weirich).

[1] We consistently use the term "gender" to cover both gender in its narrower socio-cultural meaning, as well as biological sex. While we are aware that both behavioral and anatomical factors play a role in accounting for differences observed between male and female speakers, it is rarely the case that either factor is the sole cause for a particular pattern.

https://doi.org/10.1016/j.wocn.2018.07.007
0095-4470/

(Johnson & Martin, 2001), and German (Simpson, 1998; Weirich & Simpson, 2014b; Weirich & Simpson, 2015b). These studies have shown that females produce longer vowels than males, more importantly in salient prosodic contexts (Ericsdotter & Ericsson, 2001), as well as in enhancing phonological contrast, e.g. between long and short vowels (Wassink, 1999, Weirich & Simpson, 2015b).

Explaining this gender-specific variability is more complex than simply referring to physiological differences between the sexes. Of course, biological differences do exist and they can explain certain aspects of gender-specific variability (Fant, 1960, 1966; Södersten & Lindestad, 1990; Titze, 1989; Winkler, Fuchs, & Perrier, 2006, Weirich et al., 2016). For example, the cross-sectional area and length of the vocal tract affects the formants (Fant, 1960); the size and histology of the vocal folds affect the average fundamental frequency and voice quality (Södersten & Lindestad, 1990; Titze, 1989). However, differences between languages in the size of gender-specific variability point to cultural, social and learned reasons (Johnson, 2006; Mennen, Schaeffler, & Docherty, 2012; Van Bezooijen, 1995). For example, Van Bezooijen (1995) found a larger difference between the genders in average fundamental frequency in Japanese speakers than in Dutch speakers. Interestingly, this could be related to a positive correlation between perceived attractiveness and fundamental frequency in Japanese women, which was found to be less so in Dutch women.

It is often claimed that humans learn by observing and mimicking (e.g., Social Learning Theory, Bandura, 1977), implying that young children learn the syntactic and prosodic structures, phonological patterns and lexical entries of a language through imitation of the people around them. The child absorbs not only dialectal pronunciation patterns but also sociolinguistic parameters of parents and peers (Chambers, 2003). For instance, gender-specific differences in fundamental frequency and formant values in preschool children cannot be attributed to differences in physical size or vocal tract anatomy (Fitch & Giedd, 1999; Lee, Potamianos, & Narayanan, 1999; Perry, Ohde, & Ashmead, 2001). Li et al. (2016) highlights the impact of an important social-behavioral factor (gender identity) in the development of gender difference in /s/ production. In general, where inter-speaker variability is concerned, reasons are never one-dimensional but manifold and interacting, including biological, behavioral and also perceptual factors. In addition, when it comes to gender, stereotypes, attitudes and expectations (of speakers and listeners) cannot be neglected. For example, even though empirical evidence is different, the cliché of women speaking faster than men is cross-culturally widespread (e.g. Brizendine, 2006). A possible reason is the complex relationship between measured time and *perceived* tempo which has recently been investigated in Weirich & Simpson (2014b). The study showed that a speaker having a larger acoustic vowel space is perceived as speaking faster than a speaker with a smaller acoustic vowel space (within the same gender and controlling for durational and f0 differences), thereby contributing to an explanation why females, who have on average larger vowel spaces than males, might also be perceived as speaking faster.

The size of the acoustic vowel space, and in particular the dispersion of the vowels within that space are spectral correlates of clarity. Also, the slower speaking rate and enhanced phonological contrasts between long and short vowels in females mentioned above have been taken as an argument for the claim that females speak more clearly than males (e.g. Kempe, Puts, & Cárdenas, 2013; Labov, 1990). The reasons for this ascribed clarity in female speech have been explained both in terms of "mumbling" being associated with sounding "macho" (Heffernan, 2010), and "mumbling" being affected by the morphology of the vocal tract that differs between male and female speakers (Weirich et al., 2016).

1.2. Gender and undershoot

A further spectral aspect of clarity is a speaker's dynamic exploitation of the vowel space measured by the degree of undershoot. Undershoot has been the object of investigation in a number of studies (e.g. Cho, 2004; de Jong, Beckman, & Edwards, 1993; de Jong, 1995, 1998; Harrington, Fletcher, & Beckman, 2000; Fowler, 1981; Lindblom, 1963, 1990; Mooshammer & Geng, 2008; Öhman, 1967). According to Lindblom's H&H Theory (1990) a speaker can adapt their speech behavior along a continuum from *hyperspeech*, that is output-oriented maximizing a successful communication process, to *hypospeech,* that is system-oriented and based on a low-cost form of behavior and the principle of economy. One typical feature of hypospeech is the phenomenon of vowel undershoot (Lindblom, 1963), which means that due to temporal constraints a vowel target is not reached by the articulators, resulting in formant undershoot. This vowel reduction can result in a reduced vowel space.

Thus, earlier attempts to explain undershoot described it as an automatic process reflecting a faster rate of motor commands within a shorter period of time (Lindblom, 1963). When time is short, formant displacements are reduced and vowel durations decrease. However, later studies showed that undershoot can also take place at slow rates (Nord, 1986) and does not automatically result from a high speech rate (Kuehn & Moll, 1976). In the revised model of vowel undershoot (Moon & Lindblom, 1994) the factor speech style, i.e. clear (over-articulated) speech vs. normal speech was added. Results from that study showed that over-articulated speech leads to longer and less reduced vowels but also to faster changing formants. Since then the relationship between duration, reduction and velocity has also been investigated in several articulatory studies (e.g. Ostry & Munhall, 1985; Simpson, 2003), showing among other things that the larger an articulatory movement is, the higher its peak velocity, confirming the complex interdependent relationship between articulatory/acoustic distance and duration.

Coarticulation and varying accent conditions have also been proposed as further causes for vowel undershoot. Vowel reduction due to increased coarticulation is explained by target undershoot reflecting the shorter durations in unstressed/unaccented positions. Mooshammer and Geng (2008) examined acoustic and articulatory vowel reduction patterns of tense and lax vowels in German in stressed and unstressed syllables. They found a higher degree of coarticulation in unstressed vowels than in stressed vowels. Thus, a relationship between accent/stress, degree of coarticulation and target undershoot in terms of vowel reduction exists.