Acta Psychologica 144 (2013) 583–593

Contents lists available at ScienceDirect

Acta Psychologica

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

Perceiving vocal age and gender: An adaptation approach

Romi Zäske [a,b,*], Verena G. Skuk [a,b], Jürgen M. Kaufmann [a,b], Stefan R. Schweinberger [a,b]

[a] Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University of Jena, Germany
[b] DFG Research Unit Person Perception, Friedrich Schiller University of Jena, Germany

ARTICLE INFO

Article history:
Received 23 January 2013
Received in revised form 20 August 2013
Accepted 20 September 2013
Available online 18 October 2013

Psycinfo codes:
2326 Auditory & Speech Perception;
2340 Cognitive Processes

Keywords:
Auditory aftereffects
Voice perception
Gender
Age

ABSTRACT

Aftereffects of adaptation have revealed both independent and interactive coding of facial signals including identity and expression or gender and age. By contrast, interactive processing of non-linguistic features in *voices* has rarely been investigated. Here we studied bidirectional cross-categorical aftereffects of adaptation to vocal age and gender. Prolonged exposure to young (~20 yrs) or old (~70 yrs) male or female voices biased perception of subsequent test voices away from the adapting age (Exp. 1) and the adapting gender (Exp. 2). Relative to gender-congruent adaptor-test pairings, vocal age aftereffects (VAAEs) were reduced but remained significant when voice gender changed between adaptation and test. This suggests that the VAAE relies on both gender-specific and gender-independent age representations for male and female voices. By contrast, voice gender aftereffects (VGAEs) were not modulated by age-congruency of adaptor and test voices (Exp. 2). Instead, young voice adaptors generally induced larger VGAEs than old voice adaptors. This suggests that young voices are particularly efficient gender adaptors, likely reflecting more pronounced sexual dimorphism in these voices. In sum, our findings demonstrate how high-level processing of vocal age and gender is partially intertwined.

1. Introduction

From auditory cues alone, listeners can recover remarkably detailed information about non-biological sound sources such as the size of an object (e.g., Grassi, Pastore, & Lemaitre, 2013). Human voices are more complex sound sources that convey a multitude of person-related cues in the absence of vision, and even in the absence of intelligible speech. For instance, we usually recognize familiar speakers from their voices, or infer a stranger's approximate age or gender from vocal information alone. However, despite increasing scientific attention to voices as carriers of non-linguistic information, it is largely unclear how these social signals are mentally represented. The present study uses an adaptation approach in order to investigate the perception of vocal age and gender, and their relationships.

Vocal age can be discerned from adult voices with reasonable accuracy (Linville, 1996). It has been repeatedly shown that perceived age correlates positively with chronological age in adult voices (Bruckert, Lienard, Lacroix, Kreutzer, & Leboucher, 2006; Harnsberger, Brown, Shrivastav, & Rothman, 2010; Neiman & Applegate, 1990; Ryan & Burk, 1974; Shipp & Hollien, 1969). Perceptually relevant cues to older age in adult voices include reduced speaking rate (Brown, Morris, & Michel, 1989; Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008) and increased variability of fundamental frequency (f0) (Gorham-Rowan & Laures-Gore, 2006; Linville, 1996; Ramig et al., 2001; Torre & Barlow, 2009). However, while these vocal cues to age may apply to both genders, more gender-specific patterns of vocal aging could relate to changes of f0, the distribution of the first formant (F1), harmonics-to-noise ratio (HNR), and voice onset time (VOT) (Linville, 1996; Stathopoulos, Huber, & Sussman, 2011; Torre & Barlow, 2009). More specifically, f0 and F1 have been reported to fall in women and rise in men after middle age (Linville, 1996; Torre & Barlow, 2009). Moreover, while female HNR follows a non-linear trend across the entire life-span with the lowest amount of noise at childhood and at old age, male HNR increases linearly, indicating less noise with increasing age (Stathopoulos et al., 2011). Moreover, Torre and Barlow (2009) reported a gender by age interaction in VOT, which reflected shorter VOT in voiceless stop consonants for older male voices, as compared both to older female voices and to younger voices of both genders. This suggests that male and female VOT becomes increasingly dissimilar with age.

In contrast to vocal age perception which is more continuous in nature (Zäske & Schweinberger, 2011), the perception of voice gender is more categorical (e.g. Skuk & Schweinberger, in press). Gender perception is usually highly accurate with almost perfect classification performance even for short vocalizations such as isolated vowels (Lass, Hughes, Bowyer, Waters, & Bourne, 1976). The most salient parameters

Abbreviations: ANOVA, analysis of variance; FAAE, facial age aftereffect; FGAE, face gender aftereffect; f0, fundamental frequency; F1, first formant; HNR, harmonics-to-noise-ratio; ML, morph level; RMS, root mean square; SEM, standard error of the mean; VAAE, vocal age aftereffect; VCV, vowel-consonant-vowel; VGAE, voice gender aftereffect; VOT, voice onset time.

* Corresponding author at: Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University of Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany. Tel.: +49 3641 945 935; fax: +49 3641 945 182.
E-mail address: romi.zaeske@uni-jena.de (R. Zäske).

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.09.009

underlying perceived vocal gender are thought to be f0 and formant frequencies (e.g. Coleman, 1976; Gelfer & Mikos, 2005; Ingemann, 1968; Lass et al., 1976; Latinus & Taylor, 2012; Schwartz & Rine, 1968) which are perceived as gender-typical pitch and timbre, respectively. Anatomical differences of the vocal folds and the vocal tract cause women to have higher-pitched voices (Klatt & Klatt, 1990) and higher formant frequencies (Andrews & Schmidt, 1997; Whiteside, 1998). Moreover, female voices are usually breathier due to an incomplete closure of the vocal folds (Henton & Bladon, 1985; Klatt & Klatt, 1990).

While the neuronal basis of vocal age perception is virtually unexplored, neuroimaging research on voice gender perception suggests that specific processing stages relate to distinct brain areas: Early low-level acoustic analysis, e.g. of f0, is associated with superior temporal cortex areas. By contrast, higher-level perceptual evaluation of voices as male or female has been associated with the cingulate cortex and the precuneus (Charest, Pernet, Latinus, Crabbe, & Belin, 2013; Lattner, Meyer, & Friederici, 2005; Sokhi, Hunter, Wilkinson, & Woodruff, 2005). Furthermore, electrophysiological data (Latinus & Taylor, 2012) suggest that the discrimination of male and female voice quality, irrespective of pitch, takes place relatively early (170–230 ms). Despite those intriguing findings, the neuronal underpinnings of voice gender perception are far from fully understood.

Adaptation paradigms have recently been introduced as effective means of investigating the neuronal coding of social signals in voices. Traditionally, adaptation has been studied for low-level stimulus properties such as visual and auditory motion (Anstis, Verstraten, & Mather, 1998; Grantham, 1989), color (McCollough, 1965) or loudness (Reinhardt-Rutland, 1998). In speech research, adaptation has also been shown for linguistic properties such as place of articulation (e.g. Eimas & Corbit, 1973; Holt, 2006). In vision, adaptation to high-level facial information has been extensively studied since the seminal work by Webster and Maclin (1999); for a recent review see Webster and MacLeod (2011). More recently, similar *auditory* aftereffects of voice adaptation have been reported. In a typical adaptation experiment, participants are repeatedly exposed to a certain type of adaptor stimulus, e.g. female or male voices. A prominent aftereffect is then usually observed, such that the subsequent perception of ambiguous test stimuli is biased towards the unadapted category. In the present example, test voices appear more male after female adaptation and vice versa (Mullennix, Johnson, TopcuDurgun, & Farnsworth, 1995; Schweinberger et al., 2008; Zäske, Schweinberger, Kaufmann, & Kawahara, 2009). Similar contrastive aftereffects have been obtained following adaptation to both vocal and facial age (Lai, Oruc, & Barton, 2012; O'Neil & Webster, 2011; Schweinberger et al., 2010; Zäske & Schweinberger, 2011), person identity (Latinus & Belin, 2011; Leopold, O'Toole, Vetter, & Blanz, 2001; Zäske, Schweinberger, & Kawahara, 2010) or affective information in voices and faces (Bestelmeyer, Rouger, DeBruine, & Belin, 2010; Webster, Kaping, Mizokami, & Duhamel, 2004). Such aftereffects are often thought to reflect a selective decrease in neuronal responses after prolonged stimulation.

The notion of voice-induced neuronal adaptation has received support from functional imaging studies (Andics et al., 2010; Belin & Zatorre, 2003; Charest et al., 2013; Latinus, Crabbe, & Belin, 2011) and electrophysiological data (Schweinberger, Walther, Zäske, & Kovacs, 2011; Zäske et al., 2009). For instance, repetition of the same voice irrespective of speech content reduced activity in the right superior temporal sulcus (STS, Belin & Zatorre, 2003), an area earlier identified as voice-selective (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Recently, the right inferior frontal cortex and the left cingulate gyrus have also been associated with the perception of voice identity, as these areas were sensitive to repetitions of perceptually identical voices when physical similarity was controlled (Latinus et al., 2011). Similarly, repetition suppression was reported in bilateral inferior prefrontal cortices, insulae, and the anterior cingulate cortex, for voices that tended to be perceived as belonging to the same gender (Charest et al., 2013). Furthermore, N1, P2, and P3 event-related potentials were reduced when adaptor and test voices were either gender-congruent (Zäske et al., 2009) or originated from the same speaker (Schweinberger et al., 2011). Taken together, these results indicate the existence of specialized neurons coding male and female voice quality as well as idiosyncratic voice properties of individual speakers.

However, while any given vocalization contains multiple social signals, studies on voice perception have typically focused on one dimension at a time. For instance, the ability to judge speaker age from voices was usually investigated for female *or* male voices (e.g. Bruckert et al., 2006; Harnsberger et al., 2008; e.g. Hartman, 1979; Linville, 1987; Shipp & Hollien, 1969; but see Hummert, Mazloff, & Henry, 1999). Conversely, research on voice gender recognition typically considered young adult speakers, thereby largely ignoring possible age-related differences (e.g. Lass et al., 1976; Owren, Berkowitz, & Bachorowski, 2007; Perry, Ohde, & Ashmead, 2001). As listeners are able to extract both vocal age and gender from a given utterance, the question then arises whether these social cues are perceived independently or in a more interactive manner.

Traditional models of face perception (e.g. Bruce & Young, 1986) assume that social cues in faces including identity, emotional expression and facial speech are processed independently and in parallel. More recent models of voice perception propose an analogy to face perception, in both architecture and function (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin, Fecteau, & Bedard, 2004). However, these models largely neglect the processing routes for age and gender information. Moreover, the notion of processing independence for social signals has been repeatedly challenged, at least in the face domain (Young & Bruce, 2011). Specifically, interactive processing has been reported for some social signals, such as identity and expression (Calder & Young, 2005; Fox & Barton, 2007; Kaufmann & Schweinberger, 2004; Martens, Leuthold, & Schweinberger, 2010; Schweinberger & Soukup, 1998), age and identity (Lai et al., 2012), gender and identity (Goshen-Gottstein & Ganel, 2000), gender and expression (Atkinson, Tipples, Burt, & Young, 2005) or gender and age (Barrett & O'Toole, 2009; O'Neil & Webster, 2011; Schweinberger et al., 2010).

Of particular relevance for the present study, facial age aftereffects (FAAEs) were recently established, and these effects were reduced for adaptation-to-test changes in facial gender (Schweinberger et al., 2010, Exp. 1; see also O'Neil & Webster, 2011). While this suggests gender-dependent processing of facial age, a further study found face gender aftereffects (FGAEs) induced by adult faces to be of similar magnitude for test faces of young adults and children (Barrett & O'Toole, 2009). However, smaller FGAEs were observed for adult test faces when children's faces were adaptors, perhaps because children's faces were poor gender adaptors due to smaller sexual dimorphism. These results exemplify that at least two factors need to be considered when interpreting transfer effects of adaptation: (1) the congruency of adaptors and test stimuli with respect to the task-irrelevant dimension, and (2) the saliency of adaptors with respect to the task-relevant dimension. Overall, to establish interdependent processing of two feature dimensions, one needs to study *bidirectional* transfer of aftereffects (Fox & Barton, 2007; Fox, Oruc, & Barton, 2008). Here we pursue an analogous logic for the perception of vocal age and gender.

In early speech perception research, transfer of aftereffects across task-irrelevant dimensions was used to locate the processing of linguistic features at phonetic and/or earlier acoustic processing stages. Adaptation to [ba] or [pa] shifted the category boundary for VOT between [ba] or [pa] towards the adaptor (Eimas & Corbit, 1973). Further research assessed the transfer of aftereffects when changing task-irrelevant stimulus characteristics (e.g., vowel context) between adaptor and test stimuli (Cooper, 1974; Sawusch & Pisoni, 1978). Moreover, studies manipulating the overlap of adaptor and test stimuli with respect to spectral properties or perceived phonetic identity (Roberts & Summerfield, 1981; Sawusch, 1977; Sawusch & Jusczyk, 1981) suggest that the amount of physical adaptor-test similarity appeared to account best for aftereffects, implicating an early acoustic level of adaptation.