Hearing Research 296 (2013) 131–140

Contents lists available at SciVerse ScienceDirect

Hearing Research

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

Research paper

Gender differences in familiar voice identification

Verena G. Skuk [a,b,*], Stefan R. Schweinberger [a,b]

[a] Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University of Jena, Am Steiger 3/1, D-07743 Jena, Germany
[b] DFG Research Unit Person Perception, Institute of Psychology, Friedrich Schiller University of Jena, Am Steiger 3/1, D-07743 Jena, Germany

ARTICLE INFO

Article history:
Received 30 July 2012
Received in revised form
23 October 2012
Accepted 5 November 2012
Available online 17 November 2012

ABSTRACT

We investigated gender differences in the identification of personally familiar voices in a gender-balanced sample of 40 listeners. From various types of utterances, listeners had to identify by name 20 speakers (10 female) among a set of 70 possible classmates who were all 12th grade pupils from the same local secondary school. Mean identification rates were 67% from sentences, and around 35% for an isolated /Hello/ or a VCV syllable. Even from non-verbal harrumphs, speakers were identified with an accuracy of 18%, i.e. highly above chance levels. Substantial individual differences were observed between listeners. Importantly, superior overall performance of female listeners was qualified by an interaction between voice gender and listener gender. Male listeners exhibited an own-gender bias (i.e. better identification for male than female voices), whereas female listeners identified voices of both genders at similar levels. Individual own-gender identification biases were correlated with differences in reported contact to a speaker's voice and voice distinctiveness. Overall, the present study establishes a number of factors that account for substantial individual differences in personal voice identification.
© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The human voice is not only the carrier of speech, but also provides manifold social information about speakers including a person's identity. However, it is well-known that speaker identification from the voice can be compromised by certain types of brain damage (Van Lancker et al., 1989), or in cochlear implant users, in whom speaker recognition was found to be disproportionately poor compared to vowel recognition (Vongphoe and Zeng, 2005). By comparison, there is as yet little information about individual differences and gender differences in familiar speaker identification from the voice in healthy listeners. The present study aims to fill this gap, by investigating speaker identification in a relatively large and gender-balanced sample of young adult listeners who identified classmates' voices from various types of utterances.

Relatively slow progress in understanding speaker identification from the voice (henceforth voice identification) may be attributed to the fact that researchers used very different tasks and paradigms. Voice discrimination tasks require same/different judgment for two unfamiliar voice samples, which are typically presented in immediate succession. These tasks may involve relatively low-level featural processing of voice samples, with little demand on establishing stable long-term representations. Voice discrimination can be empirically dissociated from familiar voice recognition (Van Lancker and Kreiman, 1987). This is likely because familiarization causes substantial changes in the representations used to identify people, making identification less dependent on specific low-level stimulus features, and more robust to transformations (for similar arguments in the case of face recognition, see Hancock et al., 2000). An intermediate situation with practical relevance is represented by investigations of the reliability of voice identification in earwitnesses, in the context of the criminal justice system (e.g. Goldstein, 1977). While earwitness testimony typically involves the recognition of a once-heard unfamiliar voice, retention intervals are considerable and performance is typically low. Overall, a direct comparison between empirical studies is often difficult because the outcome strongly depends on experimental design, nature and size of the set of voice samples used, and potentially large variability between both speakers and listeners.

In the present study, we focus on individual differences in the identification of personally familiar voices, with a particular

Abbreviations: VG, voice gender; LG, listener gender; U, utterance; S, sentence; H, harrumph; VCV, vowel–consonant–vowel; f_0, fundamental frequency; f_0SD, fundamental frequency standard deviation; *d*, duration; ANOVA, analyses of variance; SEM, standard error of the mean; ANCOVA, analysis of covariance.

* Corresponding author. DFG Research Unit Person Perception, Friedrich Schiller University of Jena, Leutragraben 1, 07743 Jena, Germany. Tel.: +49 3641 945 941; fax: +49 3641 945 182.
E-mail address: verena.skuk@uni-jena.de (V.G. Skuk).

0378-5955/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2012.11.004

emphasis on the role of speaker and listener gender. While many previous studies on familiar voice identification used famous voices (e.g. Hanley et al., 1998; Schweinberger et al., 1997a,b; Van Lancker et al., 1985a,b; Whittle et al., 2011), these studies suffer from the fact that experimenters have little or no control over the content and quality of voice recordings. Nevertheless, one remarkable finding was that while backward presentation of speech massively compromised the identification of some famous speakers, other celebrities were similarly well identified from backward and forward speech. This implies that voice identification involves different acoustic cues for different speakers (Van Lancker et al., 1985a).

Several other studies used personally familiar voices, for which recording can be standardized in principle (e.g. Barsics and Bredart, 2011; Bredart et al., 2009; Bricker and Pruzansky, 1966; Compton, 1963; Ladefoged and Ladefoged, 1980; Lavner et al., 2001). Moreover, personally familiar voices may enjoy stronger representations than famous voices, in analogy to similar findings for faces (e.g. Herzmann et al., 2004). However, because it is hard to identify substantial groups of listeners who are familiar with the same speakers (Yarmey et al., 2001), previous studies on personally familiar voices tend to be limited by very small numbers of speakers and/or listeners (e.g. Ladefoged and Ladefoged, 1980; Zäske et al., 2010). With respect to gender effects, the concern then remains that differences in an individual study do not necessarily reflect systematic effects, but could result from sample characteristics as well. To overcome these limitations, the present study collected data from a comparatively large and gender-balanced sample of 40 12th grade pupils from the same local school, who were all personally familiar with the same set of 20 speakers (classmates).

Person identification from the voice is often thought to follow the same general functional architecture as face identification (for a recent review, see Belin et al., 2011), and thus voice research is often inspired by findings from face perception. Of relevance for present purposes, an *own-gender bias* for faces (i.e. better recognition performance for faces of an observer's own gender) was found by Wright and Sladden (2003). However, several other studies reported an own-gender bias for female observers only, along with the finding that females outperform males in face recognition overall (Lewin and Herlitz, 2002; Rehnman and Herlitz, 2007). With respect to voice recognition, findings appear even less consistent. Roebuck and Wilding (1993) report an own-gender bias in an unfamiliar voice line-up task, but such an effect was not observed by several other studies (Thompson, 1985; Yarmey and Matthys, 1992). Similarly, although it is sometimes claimed that female listeners outperform males in voice recognition, such a difference was also not consistently observed in the above studies. Finally, none of the available studies on personally familiar voices provided clear evidence for either an own-gender bias or an advantage of female listeners (e.g. Bartholomäus, 1973; Yarmey et al., 2001) and such effects apparently were not always tested (e.g. Schmidt-Nielsen and Stern, 1985).

In the present study, we tested a homogeneous group of adolescent listeners aged between 17 and 18 years. We considered (a) that voice recognition should be fully developed at that age (Mann et al., 1979), and (b) that high levels of personal familiarity should result from sharing the past six years of secondary school together. Listeners identified classmates in an open-set identification task from different types of utterances, which included non-verbal harrumphs, the single word /Hello/, two VCV-syllables, and a standardized full sentence. We also considered several other potentially important variables: Specifically, we assessed reported frequency with which individual listeners heard a particular voice (contact), self-rated global competence in familiar voice identification, and confidence of identification in single trials. Finally, *distinctiveness* is often equated with the ease with which a person can be spotted in a crowd, or with the degree of deviation from a prototype. It is well established that distinctive faces are easier to recognize than typical faces, and this holds consistently for both familiar and pre-experimentally unfamiliar faces (Valentine, 1991; Valentine and Bruce, 1986). By analogy, we expected that distinctive voices should be more effectively recognized than typical voices. Although this issue has not been investigated in great detail as yet, distinctive unfamiliar voices have been reported to be recognized more easily in one study (Mullennix et al., 2009), whereas a similar effect could not be found in another study for personally familiar voices (Schmidt-Nielsen and Stern, 1985).

2. Experiment 1 – Familiar voice identification

2.1. Method

2.1.1. Speakers

Twenty speakers (10 female), all between the ages of 17 and 18, were recruited from a total of 70 pupils (33 female) from the 12th grade of a local secondary school (Angergymnasium, Jena, Germany, http://www.angergymnasium.jena.de). Native language was German (17 speakers); three speakers (fKJ, fML, mFS)[1] reported Russian to be the first native language, but as they grew up multilingually, with German as a second language, all spoke German fluently and error free. Their accent was judged to be nondistinctive by the authors and so they remained included in our speaker set. No speaker reported speaking or hearing disorders. All speakers signed a letter of accordance which allowed further usage of the recordings for scientific purposes. For speakers under age, parents also signed that letter. Recordings were obtained between January and March, 2011.

Four additional female voices were recorded from those pupils who were involved in a scientific school project on this research. These voices were used for practice purposes only (see below).

2.1.2. Listeners

Forty-four listeners, also pupils of the same school and year, participated in the hearing experiment. None was genetically related with any of the speakers; however, 17 (9 female) listeners had also served as speakers and thus heard recordings of their own voices during the experiment. No listener reported speaking or hearing disorders. Data from three listeners were excluded due to hardware malfunction, and one speaker was excluded due to reported lack of familiarity with several speakers. Thus, data from 40 listeners (20 female), all between the ages of 17 and 18, were analyzed.

2.1.3. Stimuli

Voices were recorded by means of a Beyerdynamic™ MC-930 condenser microphone with pop protection and a Zoom H4n audio interface (44 kHz, 16 bit). Among a set of utterances, the relevant ones were a salutation (/Hallo/), cut from the sentence 'Hallo, mein Name ist …' ('Hello, my name is …'), the sentence 'Keine Antwort ist auch eine Antwort' ('No answer is an answer as well'), two vowel-consonant-vowel (VCV) syllables (/aba/, /igi/), and a harrumph as a non-verbal vocalization. To homogenize the recording procedure, we prepared 'example' audio files for each utterance (except for harrumphs and /Hallo/). These repeated a target utterance six times, with silence between repetitions, during which speakers were recorded. Example audio files were

[1] Individuals are referred to by their initials, with a leading *f* or *m* to denote female or male participants, respectively.

Download English Version:

https://daneshyari.com/en/article/6287598

Download Persian Version:

https://daneshyari.com/article/6287598

Daneshyari.com