Sounds, words, and a woman's voice: Phonological effects on perceived voice gender categorization

Almut Naja Jebens

(s3596567)

University of Groningen

E-mail: a.n.jebens@student.rug.nl

Date: July 21, 2021

1st examiner: Prof. dr. ir. Deniz Başkent

2nd examiner and daily supervisor: Dr. Laura Rachman

Abstract

When listening to the human voice, listeners are able to perceive speaker-related information, for example the speaker's gender. Previous research has revealed that the perception of voice gender is determined by two anatomically related vocal characteristics that vary with speaker size and hormone levels: the average fundamental frequency (F0), related to the glottal pulse rate, and perceived as the vocal pitch, and the formant frequencies, related to vocal tract length (VTL), described as the voice timbre. It has been shown that speaker identification and discrimination are influenced by linguistic processing, and especially the familiarity with the spoken language facilitates voice perception. However, if this effect arises at the phoneme or word level is unclear, as well as how this influences the perception and use of certain vocal parameters, such as F0 and VTL, for speaker discrimination or identification across listening conditions. Here, we studied the effects of lexical and phonological processing on the weighting of F0 and VTL on perceived voice gender categorization in normal-hearing adult listeners by manipulating the lexical status and recording direction, and F0 and VTL properties of female reference voices. Listeners gave significantly more weight on F0 and VTL when listening to words and nonwords compared to time-reversed nonwords. This indicates that phonological processing enhances the perceptual weighting of F0 and VTL for perceived voice gender categorisation. The interplay between linguistic and perceptual processes are discussed as well as methodological considerations when phonological processing are impaired and perception F0 and VTL are limited.

Keywords: voice perception, language familiarity effect, Fundamental frequency, formant frequencies, voice gender

Introduction

The human voice carries acoustic cues that help listeners to infer important information about the listener. This can be the speaker’s age, (Dilley et al., 2013; Smith & Patterson, 2005) or the sex/gender (Klatt & Klatt, 1990; Leung et al., 2018) which are also referred to as indexical information. Next to that, humans and other mammals express emotions vocally (Briefer, 2012). The listener perceives this auditory information (Cutler et al., 1997 for a literature overview), and integrates it with the visual information coming from facial expression and gesture (Young et al., 2020). In speech, and more specifically in prosody, the clause, if the speaker makes a statement or asks a question, and the speaker’s tone, if the speaker asks or states that ironically for example, are expressed vocally via intonation (???). The indexical information, emotions, and prosodic information is inferred by using specific information in the speech signal as voice cues. These relate to the anatomy and physiology of the speaker (Titze, 1989), but also vary with speaking style and can thus be affected by societal, cultural and the linguistic context. Consider voice pitch as an example which differs between men and female speakers (Titze, 1989), and used differently by men and women (Loveday, 1981) and depends on the degree of formality of the conversational context (Idemaru et al., 2020). The crucial contribution of these voice cues in social interactions and in conversations is manifold: By using voice cues, familiar voices are recognized, and indexical information in unfamiliar voices is classified, make listeners aware of emotions of the vis-à-vis are; in so-called cocktail-party listening conditions, that are especially challenging for hearing-impaired and the elderly population (Noble & Gatehouse, 2004), voice cues enable listeners to track the target voice which is the prerequisite for processing the speech signal linguistically. This way, voice cues could potentially even aid the acquisition of the native and a foreign language as a bootstrapping mechanism (Höhle, 2009), helping children to segment the speech signal to infer the phonemes, words and syntactical structure of their native language.

Because of this interplay between language processing and voice perception, and their relevance for daily communication when hearing is normal or impaired, or when a language is acquired, previous research investigated effects on both cognitive functions. Especially, the familiarity of the voice or the native language seems to have a facilitatory effect: n the one hand, a “familiar talker advantage” (Case et al., 2018; Levi, 2015; Levi et al., 2011; Nygaard & Pisoni, 1998) for language processing has been reported, the “language familiarity effect” (Fleming et al., 2014; Goggin et al., 1991) for voice perception has been found as well, that is, higher speaker discrimination or identification on the other hand.

The language familiarity effect seems to be a robust effect and has been investigated across the life span, in the typical and atypical development and across different tasks and linguistic materials (for an overview see Levi (2019)). It is still unclear how the linguistic processes at different linguistic levels alter the use of specific voice cues for categorizing voices for indexical information, for example for the gender/sex of the speaker. In this study, we explore the effects of lexico-semantic and phonological processing on the perception of voice cues for categorizing speakers into perceived male or female characteristics.

In the literature, the perception of indexical features such as speaker’s sex/gender is described as a bricolage of vocal and articulatory features (Klatt & Klatt, 1990). These can vary with speaking style (Whiteside, 1999; Zimman, 2017) being limited by the the speaker’s anatomy. In previous studies in which voice characteristics were manipulated in a reference voice, two anatomically related vocal parameters have been identified to alter a speaker’s perceived gender systematically (Fuller et al., 2014; Nagels et al., 2020; Skuk & Schweinberger, 2014). These are the fundamental frequency (F0), mainly determined by the glottal pulse rate (GPR), which is perceived as speaker’s voice pitch and the speaker’s vocal tract length (VTL), closely related to the speaker’s size, and which shapes the distribution of the formant frequencies in the speech signal, adding to the speaker-characteristic timbre (Klatt & Klatt, 1990; Skuk & Schweinberger, 2014). The GPR is determined by the mass and length of the vocal cords (Titze, 1989) and their elasticity (Ayache et al., 2002). For intonation in speech, but also for singing, pitch is modulated by contracting and releasing the laryngeal muscles (Hoh, 2010; Unteregger et al., 2017). In speech, the range and variation of pitch differs by language (for an overview see (Traunmüller & Eriksson, 1995)), and that has been shown for languages that fall into the same class (Mennen et al., 2012) and into different (Keating & Kuo, 2012) classes in terms of stress-based and tonal characteristics. Other factors that have been identified to affect pitch is if the language is the speaker's native or foreign (Zimmerer et al., 2014), and how formal the conversation is (Idemaru et al., 2020). This results in GPR variations within speakers of 3.7 semitones as a standard variation in natural speech (Kania et al., 2006), and listeners accept variations of 3.8 semitones within one speaker before detecting speaker change (Gaudrain et al., 2009). As a consequence, the GPR of female and male speakers come in great variations and overlap (Eriksson, 1995), making pitch a less distinctive indicator for speaker's sex. However, averaged over time, GPR is described stable over time as more stable indicator, referred to as mean or averag F0, and therefore used for speaker identification. As described in the source-filter theory (Titze & Martin, 1998), F0 is resonating in the supra-laryngeal vocal tract of the speaker while its length (VTL), width and shape determine the formant frequencies of the voice, characterizing the voice's timbre (von Kriegstein et al., 2006): The longer the vocal tract, the more prominent the lower frequencies in the speech signal (Fant, 1970). Within one speaker, VTL can be modulated be extended by moving the articulators such as lips and velum farther away from the source of the speech signal, or by retracting the larynx (Briefer, 2012), but these mechanisms of modulations are more limited than the ones for GPR, and in listeners attribute a smaller range of VTL, namely of 2.2 semitones, to one speaker (Gaudrain et al., 2009). The individual ranges of F0 and VTL probably play a role when voice cues are see associated with indexical features that are linked to anatomical parameters.

In humans, the anatomical underpinnings of F0 and VTL are sexually dimorphous, varying with speaker's size and hormone levels: The vocal cords in males are thicker and longer, so on average, men's average F0 is 12 semitones (st) lower than that of women (Klatt & Klatt, 1990; Titze, 1989). Within male, but not female speakers, F0 is associated with testosterone levels (Dabbs & Mallinger, 1999). The vocal tract in male speakers is on average

0.23 longer than in female speakers (Titze, 1989), resulting in formant frequencies that are 3.6 st lower in male speakers. This difference in VTL can be linked to overall-size differences between males and females (Roser et al., 2013); VTL correlates with overall speaker size and weight in humans and non-human primates (Fitch, 1997; Pisanski et al., 2014; Rendall et al., 2005). Adding to that, the male vocal tract lengthens during puberty due to hormonal changes (Markova et al., 2016). Listeners use both F0 and VTL to estimate speaker's size, age, and sex or gender (Smith & Patterson, 2005), while the latter is an important voice cue from an evolutionary (Hodges-Simeon et al., 2015) and social perspective (???), and will be the focus of our study. The evidence that F0 and VTL are crucial for categorising the speaker's perceived sex/gender comes from studies where these were systematically manipulated according to the anatomical differences between the sexes and created the sensation of listening to the opposite sex's/gender's voice. The effective use of these voice cues develops through childhood and the language development, and it takes many years for children to reach adult-like levels of use of the cues for perceived vocal gender categorization (Nagels et al., 2020). In individuals with impaired hearing that are provided with a cochlear implant, their use deviates from normal hearing. Differing from use of both F0 and VTL, implant users rely mostly on F0. The differing weighting of voice cues than normal hearing potentially hinders identification of a speaker's gender also in everyday listening conditions or force to make use of other gender-specific voice cues (Fuller et al., 2014). The potential relationship between the acquisition of the native language and F0 and VTL as voice cues and the potential explanation for the difficulties in discriminating and identifying speakers in cochlear implantees highlights the relevance of these voice cues to further the understanding of voice perception and how this is associated with linguistic processes.

Voice cues, such as F0 and VTL, are not only relevant in social interactions when it comes to identifying features such sex/gender of the vis-à-vis, but also for language processing. For example, F0 and VTL serve discriminating and tracking one speaker to understand what is said in a conversation and especially in multiple talker listening conditions, such as a which are described as the “cocktail-party” problem (Cherry, 1953). In the laboratory, these situations are usually investigated in speech-on-speech listening tasks. Differences in F0 and VTL have been shown to increase attending one speaker when listening to multiple talkers (Darwin et al., 2003) and increase intelligibility of the linguistic content which has been demonstrated in a repetition task conducted by (Başkent & Gaudrain, 2016). The facilitatory effect of F0 and VTL for speech segregation and intelligibility can already be observed in normal-hearing children from the year of 4 on (Flaherty et al., 2019; Nagels et al., 2021; Zaltz et al., 2020). The differences in F0 and VTL between male and female voices could potentially explain the observation that sex/gender differences between speakers enhance the intelligibility in dichotic listening tasks or competing speech (Brungart et al., 2001; Festen & Plomp, 1990), when these features were not manipulated, but listeners were exposed with speakers of different sexes. Moreover, the use of specific voice cues such as F0 and VTL could also explain one phenomenon that is described as the “familiar talker advantage”: Listeners adapt to speaker-related characteristics such as F0 and VTL or articulatory properties and consequently, show enhanced linguistic processing. This has been observed for sentence recognition (Goggin et al., 1991) and word identification which has been studied in adults (Nygaard & Pisoni, 1998) and in school-aged children (Levi, 2015) with an even stronger effect in elderly listeners (Yonan & Sommers, 2000); when shadowing one voice in stream segregation (Newman & Evers, 2007); and for attending and tracking, but also for suppressing voices in competing speech (Johnsrude et al., 2013). Besides tracking one speaker, the perception of VTL is also crucial for phonological processing, and thus for language understanding, because these phonemes are meaningful: VTL is critical in determining the frequencies of the formants of the phonemes, more specifically of the vowels in speech of the individual (Irino & Patterson, 2002). F0 and VTL as voice cues are thus crucial in terms of language understanding when it comes to tracking single voices, might explain the observed adaptation effects for speakers and are crucial in language processing at the phoneme level.

The counterpart of the familiar talker advantage is the “language familiarity effect”, first described by (Fleming et al., 2014; Goggin et al., 1991), which refers to the enhanced ability to discriminate and to identify voices when one is familiar with the language, in most cases the native language. As described above, F0 and VTL are used for speaker discrimination and identification. Therefore, it could be suspected that these language effects could also influence the weighting of these cues for the identification of the speaker’s sex/gender. With some exceptions of a few electrophysiological (Conde et al., 2018) and imaging studies (Hu et al., 2017; Zäske et al., 2017), these language (familiarity) effects have mainly been investigated by using behavioural designs across different populations and by using different materials. The outcome of these studies suggests different levels of linguistic processing that interfere with voice perception that result in differences in performance in the ability to discriminate and identify voices. This suggests that the language familiarity is a robust effect. In between-subject designs, studies compared the performance in speaker discrimination and identification in listeners with different native languages (Bregman & Creel, 2014; Brungart et al., 2001; Drozdova et al., 2017; Fleming et al., 2014; Goggin et al., 1991; Hu et al., 2017; Kadam et al., 2016; Levi, 2018; McLaughlin et al., 2015, 2019; Perrachione et al., 2011, 2019; Sharma et al., 2020), across life span from early infancy on (Fecher et al., 2019; Fecher & Johnson, 2018a, 2018b, 2019, 2021), and in the monolingual and bilingual and typical and atypical language development (Bregman & Creel, 2014; Case et al., 2018; Theodore & Flanagan, 2020). However, from these studies it is hard to define which linguistic processes interfere with the perceptual processes and resulting in these so-called language-familiarity effect. A foreign language differs from the native language in many aspects due to their phonotactics, and words and sentences are not meaningful to the speaker so they cannot be processed lexical-semantically. Another approach to define the linguistic level at which this language familiarity effect was to assess voice perception in conditions in which phonological abilities play are assumed to be reduced or impaired, for example in developmental dyslexia (Perrachione et al., 2011; Kadam et al., 2016). Studies with between-subject designs thus seem problematic to define the crucial linguistic parameters that led to the language familiarity effect, calling for studies that focussed on the linguistic material.

Studies with within-subject-designs manipulated a range of linguistic parameters that could indicate which linguistic processes interfere with the processing of voice cues for speaker discrimination and identification of indexical information and both lexico-semantic and phonological processes have been suggested. In an identification task, (Goggin et al., 1991) manipulated the lexical status of the stimuli and found an advantage for meaningful words compared to nonwords in a speaker identification task without training; and (Perrachione et al., 2019) confirmed this advantage of words when including an additional training for speaker identification. Phonological processes were investigated in terms of complexity and phonotactics. The complexity effect means that the availability of phonemes is beneficial for voice identification which suggests that phonological processes are somehow intertwined with the voice perception. (Mary Zarate et al., 2015) showed this effect by comparing speaker identification when presenting vocal sounds and disyllabic words. In an EEG-study, there was an effect of phonological complexity in the attention and towards vocal stimuli (Conde et al., 2018). Phonotactics has been investigated by comparing the effect of a native and foreign accents that either adhere to the phonotactic rules of the native language or by manipulating recording of speech. In a voice identification task, phonetic variations in familiar words from a familiar to a foreign accent impeded the accuracy in speaker identification (Ganugapati & Theodore, 2019). According to the authors, the benefit of the familiarity with the words was no longer present, when phonetics was unfamiliar to the listener. Another common manipulation is the recording direction (Levi, 2019). In reversed speech, some characteristics of the language might remain intact, while other phonotactic parameters are violated, for example the voice on-set times of consonants and coarticulation cues (Levi, 2019). In a voice dissimilarity rating, listeners perceived voices as more dissimilar in reversed speech compared to forward speech, in the native as well as in a foreign language, which suggests that this effect is independent of lexical content and lexical-semantic processes (Perrachione et al., 2019). In some cases, the distinction between lexical-semantic and phonological processes might not be that clear: In a study by Koelewijn and colleagues (submitted), just-noticeable differences of F0 and VTL, a measure of sensitivity, were smaller when comparing forward to reversed played monosyllabic CVC words; However, it is not clear if it is the playback direction solely, resulting in untypical coarticulation and voice onset times, or because the reversed stimuli are not meaningful to the listener because the unintelligibility interrupts lexical-semantic processing. For sex/gender categorisation in speech, this could mean that the weighting of F0 and VTL to make a judgement for perceived gender of a speaker could be altered by recording playback directions, since listeners show to be more sensitive to them in this condition. In contrary to this evidence, supporting the idea that phonological and lexical-semantic processes modulate voice perception, and probably also perceived voice gender, voice dissimilarity ratings which are likely associated with the sensitivity of voice cues were found to be influenced heavily by voice characteristics such as F0 and VTL, and less by linguistic characteristics (Perrachione et al., 2019).

The present study

In the current study, we assess if linguistic processing interferes with the weighting of F0 and VTL for a categorization task of perceived voice gender and we investigate these effects across multiple (three) speakers. This study is part of a larger project PICKA (Perception of indexical cues in kids and adults) on the perception of voice and speech. The current experiment was designed according to previous studies conducted by (Fuller et al., 2014) and (Nagels et al., 2020) in which F0 and VTL of a female reference voice were manipulated creating the sensation of a male voice by voice synthesis and using intermediate steps for F0 and VTL values. We extended the experimental design in terms of linguistic conditions and speaker variability to enhance generalisability. In this study, we test the same 9 voice conditions as used by (Nagels et al., 2020) to cover the range of voices for varying F0 and VTL values, and a 2 alternative forced choice paradigm that forces the listener to categorize the voice gender they perceived into one of the two categories of gender (Fuller et al., 2014). To investigate the lexical-semantic and phonological processes, lexical status and recording direction is manipulated which results in three linguistic conditions, words, non-words and reversed non-words. By using these three conditions, influences from lexical-semantic and phonological influences can be disentangled. Non-words adhere to the phonotactic rules of the language, but do not transport any meaning. Reversing nonwords affects certain phonetic features, such as voice onset times, consonant clusters, coarticulation, and the stress of the vowel within unintelligible stimuli. Next to this linguistic variability, we test perceived voice gender categorization across three speakers, to enhance the generalizability of potential effects. Therefore, three speakers were selected from the VariaNTS corpus (Arts et al., 2021), a tool that was developed to represent a wide range of speakers, and accents and natural variations in speech in high-quality recordings in which the pre-processing of the stimuli was reduced to a minimum. We will test these effects in normal-hearing adults with Dutch as a native language.

Hypotheses

First, we expect to confirm the findings of (Fuller et al., 2014) and (Nagels et al., 2020) and generalize them in the three reference voices that F0 and VTL alter voice gender categorization. In terms of our research question, we hypothesize that linguistic processes interfere with the perception and the use of F0 and VTL for gender categorization. In accordance with the findings of Koelewijn and colleagues (submitted) of higher sensitivity to differences in F0 and VTL in forward compared to reversed words, we expect a higher weighting of F0 and VTL for categorizing gender in forward words compared to reversed nonwords. If these differences in the cue weighting arises at the lexical-semantic level, we expect to see a higher weighting of F0 and VTL in words compared to nonwords, but no difference between nonwords and reversed words. If arising at the phonological level, we expect to see higher cue weighting in words and nonwords compared to reversed nonwords.

Alternatively, according to the findings of (Perrachione et al., 2019), it could also be expected that the perception and the weighting of F0 and VTL for categorizing the speaker's sex/gender remains unaffected by linguistic manipulations and is stable across linguistic manipulation; neither the lexical status of the stimulus (word or non-word) or its phonotactic properties such as recording direction (forward versus time-reversed) would then impact the weighting of F0 and VTL. In the earlier-mentioned study conducted by Perrachione et al., 2019, speaker-dissimilarity ratings were more dependent on the physical properties of the voice, most important the fundamental and the formant frequencies, while linguistic variations only had a small impact on such ratings. This could mean that in the voice gender categorization task, the weighting of fundamental frequencies remains stable across the linguistic conditions.

Methods

Ethical approval

Ethical approval for the study was given by the Medical Ethical Review Committee of the University of Groningen (METc 2018/427) for all experiments that were conducted as part of the earlier mentioned PICKA project. All experiments and methods are performed in accordance with the relevant guidelines and regulations. Participants were provided with detailed information about the study and gave their written consent before performing the experiment.

Participants

Twenty adults (mean age: 26,75, median: 24, range: 18–49 years; 2 women, 18 men) were recruited and reimbursed via the online testing-platform (*https://prolific.co/*). The demographic information was collected via the PICKA questionnaire. All participants are native speakers of Dutch and reported to have no history of language or reading impairments. Five of the twenty participants were raised multilingually. Participants self-reported having no hearing problems and underwent a speech-in-noise task (Smits et al., 2004) prior to the experiment, and 19 of the 20 participants' normal hearing was confirmed via this test. Participants' demographic information is given in Appendix 1.

Stimuli and apparatus

According to the testing conditions, testing items are controlled for linguistic features, produced by three different speakers, and the vocal conditions are manipulated. The items consist of 8 forward words and 8 nonwords that are presented in forward and in time-reversed recording direction. The words and nonwords are spoken by three female speakers and taken from the VariaNTS corpus (Arts et al., 2021) the same voice conditions were manipulated later. Words and nonwords are controlled for phonological and words for morphological and lexical-semantic and parameters. In terms of lexical-semantic features, words are monomorphemic nouns, rated as highly on a scale from 1 (unfamiliar) to 7 (highly familiar) by Dutch native speakers (Arts et al., 2021). They are classified as high-frequent by (Arts et al., 2021) based on two corpora, ranging from 21 to 515 per million according to the CELEX database (Baayen et al., 1996) and from 24 to 274 per million according to the SUBTELEX database (Keuleers et al., 2010). In the CELEX database frequencies are based on written language, namely such as in books and journals, while the SUBTELEX database used subtitles of movies. Since this study tests the perception of spoken language or speech, the latter might fulfil the purpose of this study better. Apart from that, the SUBTELEX was developed more recently compared to the CELEX database, and closer to the testing point of this study; therefore, the frequency values are assumed to be more reliable. In terms of their phonological features, all word and nonword stimuli have a low neighbourhood-density (Marian et al., 2012): number of neighbours is defined by the words within the language that result from deleting, substituting, or adding phonemes of the respecting item. Nonwords are having a high phonotactic probability which was derived from biphone frequencies of their phonemes (Arts et al., 2021). This value refers to how frequent two subsequent phonemes in real Dutch words are that are used in the nonwords based on the CLEARPOND database (Marian et al., 2012). Additionally, the nonwords were rated by native Dutch listeners as highly probable according to their sound structure on a scale from 1 (lowest probability) to 7 (highest probability; Arts et al., 2021), referred to as “mean probability”. Words and nonwords are controlled for the position of the consonant cluster, either appearing in the beginning or end of the stimulus and balanced across words and nonwords. Due to a possible interaction between speaker’s VTL and the formants (Irino & Patterson, 2002), words and nonwords are matched for their vowel. Controlled parameters for words are given in table 1 and for nonwords in table 2.

Table 1

Practice and testing words

no	vowel	block	item	Phon. structure	Frequency per million		Neighbourhood density	Mean familiarity rating
					CELEX	SUB-TELEX		
1	/a:/	test	smaak	CCVC	75	29	8	7
2	/e/	test	berg	CCVC	21	31	4	7
3	/e:/	test	bril	CCVC	32	24	5	7
4	/ ɛ/	test	hoofd	CVCC	515	274	6	7
5	/i:/	test	steen	CVCC	46	36	8	6,9
6	/I/	test	bron	CVCC	42	29	7	7
7	/o:/	test	kamp	CVCC	31	40	7	7
8	/ʊ/	test	fiets	CVCC	74	46	7	7
		practice	stuur	CCVC	25	16	8	7
		practice	vuist	CVCC	23	14	4	7

Overview about the words and nonwords used for the testing and practice blocks. The table gives the number of the item, the vowel, the block (test or practice phase), the item, the phonological structure in terms of consonants (C) and vowels (V), the frequency values per million according to CELEX (Baayen et al., 1996) and SUB-TELEX (1996; Keuleers et al., 2010), the neighbourhood density (Marian et al., 2012), and the familiarity rating (Arts et al, 2021).

Table 2

Non-words for testing and practice

no	vowel	block	item	Phon. structure	Neighbourhood density	Phonological probability	Mean probability
1	/a:/	test	prien	CCVC	0	0.0182	0.0182
2	/e/	test	dreer	CCVC	1	0.0129	0.0129
3	/e:/	test	jorf	CVCC	0	0.0124	0.0124
4	/ ɛ/	test	saark	CVCC	0	0.0269	0.0269
5	/i:/	test	frool	CCVC	0	0.0128	0.0128
6	/I/	test	frag	CCVC	0	0.0132	0.0132
7	/o:/	test	sirs	CVCC	0	0.02	0.02
8	/ʊ/	test	selm	CVCC	1	0.0125	0.0125
		practice	speif	CCVC	0	0.013	0.013
		practice	praum	CCVC	0	0.0129	0.0129

Table 2. Table 1 gives an overview about the words and nonwords used for the testing and practice blocks. The table gives the number of the item, the vowel, the block (test or practice phase), the item, the phonological structure in terms of consonants (C) and vowels (V), and neighbourhood density (NHD; Marian et al., 2012), and the phonological (Marian et al, 2012) and mean probability ratings (Arts et al., 2021.

The recordings of words and nonwords, produced by 3 female speakers, were taken from the VariaNTS corpus (Arts et al., 2021). The speakers' F0 and their height and weight, which are correlated with their VTL (Fitch & Giedd, 1999), are given in table 2. The three speakers were selected from the 8 female speakers after applying of high-pass filter with a cut-off at 80Hz (Butterworth filter, 12th order) and 5ms of silence in the beginning in Adobe audition (*Software voor het opnemen en bewerken van audio | Adobe Audition*) because these voices sounded the most natural and recordings the clearest in terms of artefacts by the author of this thesis. The VariaNTS corpus was prepared to represent variations of realistic speaking styles and therefore the original recordings were in high quality, but not manipulated heavily in post-processing. However, for the current study, it was necessary to synthesize the voices and some small artifacts that come from realistic recordings may be amplified with your experimental manipulations resulting in artefacts that might influence performance of the participants. Therefore, we had to do the selection from existing recordings.

Recording direction of the nonwords were manipulated in MATLAB. The three female reference voices were then manipulated with the same parameters as Nagels (2002), resulting in the same 9 voice conditions: F0 is decreased in steps of 0.0, 6.0 and 12.0 st and VTL in steps of 0.0, 1.8 and 3.6 st. These manipulations were done by using the PyWorld wrapper (*GitHub - JeremyCCHsu/Python-Wrapper-for-World-Vocoder*) and applied according to Gaudrain & Başkent (2015). Taking all linguistic and vocal manipulations into account, this resulted in a total of 648 stimuli for each participant ((3 F0 values * 3 VTL values) * (8 words + words + 8 time-reversed words) * 3 speakers).

Table 3

Speaker information

speaker	gender	age (years)	Height (cm)	weight (kg)	Mean F0 (Hz)
2	female	20	171	59	214.36
12	female	22	175	78	191.83
15	female	21	176	65	199.38

Table 3 gives the speakers' gender, age, height, weight and mean F0 used in the experiment.

Procedure

The experiment was completed remotely and written in the JavaScript framework "JsPsych" (de Leeuw, 2015). Due to online-testing, audiometric testing could not be done beforehand. Instead, participants were asked to complete a digit-in-noise test (Smits et al., 2004). The outcome of this test was not used as an exclusion criterion. Before testing, participants were exposed to one practice item and asked to adjust the volume at a comfortable level and to keep this volume over testing.

To investigate the weighting of F0 and VTL for categorizing voice gender across the 3 linguistic conditions, a 2-alternative forced choice procedure was used. Before starting the actual experiment, the participants underwent 6 practice trials, consisting of different words, nonwords and reversed nonwords containing different vowels/formants and were spoken by a different female speaker than the three included speakers in the main experiment to prevent any adaptation effects from the voices or stimuli. Practice stimuli also stemmed from the VariaNTS corpus (Arts et al., 2021) and were controlled for the same linguistic variables as the testing stimuli. In the experiment, stimuli are presented in 9 blocks. Linguistic conditions were presented in separate blocks, but speakers and voice conditions were randomized within blocks. to speakers and linguistic conditions and the order of stimuli was s randomized in each block. The whole experiment took participants about 50 minutes. Appendix 2 gives the timeline of the experiment including instructions.

In each trial, participants fixated to the screen (100ms) and listened to a stimulus that was randomly selected according to the linguistic condition of the block. Listeners were forced to click the response buttons with "man" or "vrouw" (i.e., "man" or "woman" in Dutch). After the response, there was a gap of 1000 ??? ms and the next trial started. The trial procedure is depicted in figure 1.

Figure 1

Trial procedure

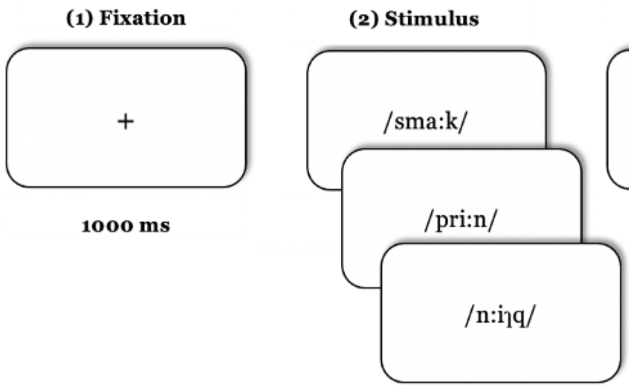


Figure 1 illustrates the trial procedure for 3 linguistic conditions. In the block-design, one of the three linguistic conditions appeared (words, nonwords, reversed nonwords) which are here represented by the 3 example items "smaak" for words, "prien" for nonwords and "prien" in a reversed recording direction.

Data Analysis

A mixed-effects logistic regression model with an interaction between voice cue (F0 and VTL), linguistic condition (word, nonword, reversed nonwords) and a random intercept was applied to investigate the linguistic effects on cue weighting of F0 and VTL for perceived voice gender categorization. The data were analysed in R (Version 1.4.1717) using the lme4 package (Bates et al., 2015). As a first step, F0 and VTL differences within each speaker were normalised in relation to the reference voice of the respective speaker: F0 was defined as $\delta F0 = -\Delta F0/12 - 0.5$ and VTL as and $\delta VTL = \Delta VTL/3.6 - 0.5$. Within this model, the manipulated voices were assigned normalized values reflecting the voice condition instead of their frequencies of F0 and VTL in Hz: the female reference voice (F0: 0.0 st and VTL: 0.0 st) received a value of -0.5 for δF0 and -0.5 for δVTL, while the voice with the extreme manipulations (F0: -12.0 st and VTL: +3.6 st) that should evoke the sensation of listening to a male voice according to earlier studies conducted by (Fuller et al., 2014; Nagels et al., 2021) received a value of +0.5 for δF0 and +0.5 for δVTL. As a second step, we extracted the coefficients of the participants' responses in each linguistic condition (words, nonwords and reversed nonwords) in a mixed-effects logistic regression model with random intercepts and for δF0 and δVTL for each participant and *female* as an outcome variable: this model predicts the values on a logit scale in relation to the normalized δF0 and δVTL values, ranging from -0.5 to +0.5. In the lme syntax, is this notated as: female ~ (δF0+δVTL|participant). The responses' coefficients for δF0 and δVTL for every linguistic condition were converted into "Berkson" (Bk) units for each st: One Bk per st equals the double of the categorizing the stimulus as "male". The Berkson units per participant are fitted in a generalized linear mixed-effects model with random intercepts per participant.

As a third step, we compared models in a backward stepwise model selection with ANOVA Chi-Square test and based on their significance (p=.05), factors were kept in the model. We started with the full factorial model and a two-way interaction between the fixed effects of *voice cue* (F0 and VTL) and *linguistic condition* (words, nonwords, reversed nonwords) and a random intercept per participant and *cue weight* in bk/st as an outcome variable: cue weight ~ voice cue * linguistic condition + (1|participant). Lastly, in a post-hoc analysis, we estimated the mean of the cue weight according to the best-fitting model in pairwise analyses for voice cue and linguistic conditions with Bonferroni correction. For this, we used the emmeans()-function from the emmeans package (Lenth et al., 2021).

Results

Figure 2 shows the average of the perceived gender categorization for each voice condition resulting from both F0 and VTL manipulations for each linguistic condition: words, nonwords, reversed nonwords. Each matrix shows the relative responses of "perceived as woman" (yellow) and "perceived as man" responses (violet) for the voice conditions.

Figure 2

Average responses for each voice condition in words, nonwords, time-reversed nonwords

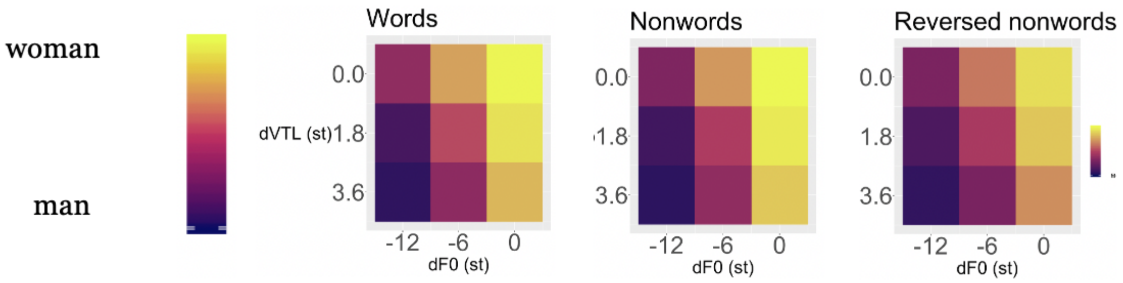


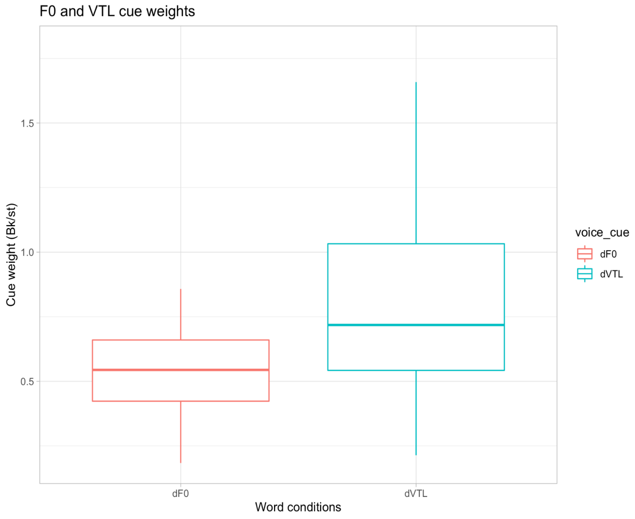
Figure 2 shows average perceived voice gender categorisation judgement as a function of differences in F0 (x-axis) and VTL (y-axis) in st, shown for words, nonwords and reversed nonwords (from left to right). Red corresponds to 100% "perceived as man" responses and yellow corresponds to 100% "perceived as woman" responses.

Model comparison showed that the full (factorial) model with random intercepts per participant fitted significantly better than the full factorial model without random intercepts per subject [χ^2 (1) = 18.08, p<.001]. Backward stepwise selection revealed that the best fitting and most parsimonious model was the full model without interaction and *voice cue* and *linguistic condition* as main effects. This is notated as *cue weight ~ voice cue * linguistic condition + (1|participant)* in the lme4 syntax. This model did not differ significantly in its fit that the full factorial model with interaction between *voice cue* and *linguistic condition* [χ^2 (2) =2.29, *p*=.32]. The full model showed a significantly better fit than the model with only *voice cue* [χ^2(2) = 13,97, *P* < .001] or only *linguistic condition* as a main effect [χ^2(2) = 25.20, *P*= 5.2e-07].

The significant effect/main effect of voice cue showed that listeners put more weight on VTL compared to F0 across linguistic conditions [*Estimate*=-0.247, t=-5.272, p<.001]. Figure 3 depicts the cue weights in bk/st for F0 and VTL. The significant effect/main effect of linguistic condition shows that listeners weighted significantly more cue weights on F0 and VTL for words compared to reversed nonwords [*Estimate*=0.1995, t=3.473, p=0.0023] and nonwords compared to reversed nonwords [*Estimate*=0.1780, t=3.099, p=.0076], while there was no significant difference in cue weighting between words and nonwords [*Estimate*=0.0215, t=0.274, p=1.0]. When listening to reversed nonwords compared to words, listeners gave 77% of the weight given to F0 and 72% of the weight given to VTL. When comparing reversed nonwords to forward nonwords, listeners gave 70% of the weight given to F0 and 81% of the weight given to VTL. Figure 3.2 depicts the total cue weights (F0 and VTL) in bk/st in the linguistic conditions for words, nonwords, and reversed nonwords.

Figure 3

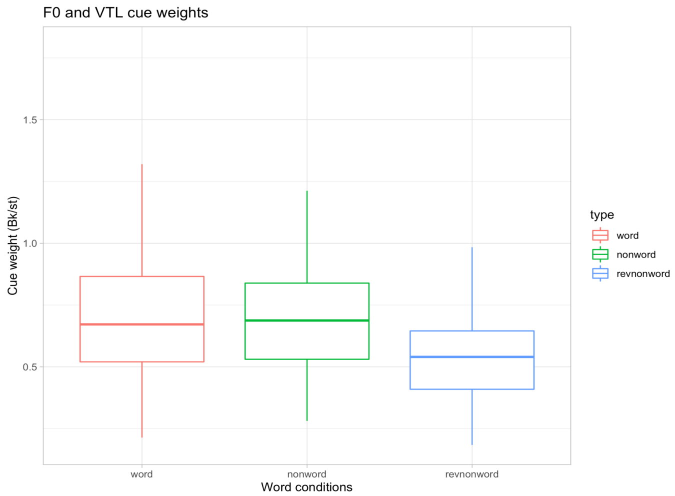
Median cue weights of F0 and VTL



The boxplots show the median cue weights of F0 and VTL in bk/st across linguistic conditions.

Figure 4

Median cue weights for words, nonwords, time-reversed nonwords



T box shows the cue weights of both voice cues F0 and VTL in the word, nonword and reversed nonword conditions. The boxes show the lower and upper quartiles, and the whiskers show the lowest and highest data points within plus or minus 1.5 times the interquartile range.

Discussion

The aim of the current study was to investigate if linguistic processing alters the use of voice cues, F0 and VTL, for voice gender categorization, and if this language effect can be explained by lexical-semantic or rather phonological processes. To that end, adults with self-reported normal hearing underwent a perceived voice gender categorization task in which we both manipulated voice cues and linguistic parameters: We synthesized a range of voices that varied from perceived as man to perceived as woman and with in-between categorization of perceived gender. To disentangle lexical-semantic and phonological effects, we manipulated the lexical status and phonotactic features of the stimuli, by contrasting words with nonwords, and the nonwords in a forward and a time-reversed recording direction. Then we calculated the perceptual weights of F0 and VTL and compared them across the three linguistic conditions. First, we found a main effect for both *voice cue* and *linguistic condition* on cue weighting. Second, pairwise comparisons showed that perceptual cue weights were significantly higher for VTL than for F0 across linguistic conditions. Cue weights for F0 and VTL were higher in words and nonwords which both adhere to the phonotactic rules of the language, compared to reversed nonwords in which these rules are violated. Because we found no effect of lexical status, but one of recording direction, our results suggest that the perceptual weighting of F0 and VTL is altered by phonological, but not by lexical processes/these changes in cue weights arise at the phoneme, and not at the lexical level.

The outcome of this study, that phonological, but not lexical processing enhances the perceptual weighting of F0 and VTL for perceived voice gender comes with some implications for the interplay between linguistic and perceptual processes when weighting these voice cues in this specific, but also beyond the perceived voice gender categorisation task and for the theoretical accounts that explain this interplay at the linguistic/cognitive level. Further, our results are also relevant when studying the perception and use of F0 and VTL for voice identification when these phonological processes are hampered or impaired such as in developmental or acquired language and reading disorders, or when the use of voice cues such as VTL for such identification tasks is limited as it is the case in CI users (Fuller et al., 2014).

One explanation for the effect of recording direction, that we interpret as phonological processes, could be that also the perception of non-linguistic information such as the voice cues F0 and VTL is facilitated when the speech input can be segmented in the corresponding phonemes of the native language. That is even the case when the linguistic information is not relevant to the task. In cognitive models that describe the auditory processing of (monomorphemic) words, the processing of phonemes occurs at an earlier stage than the processing of words and their meaning are processed. An example for such a model is the dual-route model (Morton, 1969). That the perceptual processes of speech sounds is altered by phonemic representations, and thus associated with meaning, has also been shown in an ERP study by Näätänen and colleagues (1997), mismatch negativity was increased when two presented sounds were presented as different phonemes in the native language. We investigated the effects of phonological processing by time-reversing speech which affects speech at the segmental as well as the suprasegmental level and impedes the segmentation into phonemes. At the segmental level, or phoneme level, voice-onset times are violated that distinguish phonemes with the same placement of articulation. At the supra-segmental level, or syllable level, sound combinations emerge that do not adhere to the phonotactic rules of the language, and the intonation pattern, that is decreasing in Dutch (Domahs et al., 2014), is reversed. That means that in the reversed nonwords, listeners cannot rely on voice-onset times, sound combinations and intonation patterns to segment the input into meaningful sounds or phonemes of their native language.

Our results suggest that the perceptual weighting of F0 and VTL interacts with the linguistic processing at the phoneme level, however, it needs to be determined at which processing state this interaction effect emerged. To be more specific, we need to explore if this effect can be related to the early perceptual stage, the discrimination ability or becomes relevant when F0 and VTL need to be perceptually weighted, for example in a perceived gender categorisation task that we used. One theoretical account explains the interplay of VTL , one of the voice cues that we manipulated, and vowel perception at the early perceptual level, and might point that the interaction effect between phonological processing and voice cue weighting might occur at a later, and not at the early processing state: For correct vowel categorization, VTL has to be estimated by the listeners, because the distribution of the formants/vowels depend on the speaker's individual VTL. Irino & Patterson (2002) propose that extraction of VTL and phonemes, more specific vowels, can be described by a stabilised wavelet-mellin transform based on their temporal scaling/distribution. There is evidence that phonological processes already affect earlier the perceptual processes that play a role for discrimination tasks, for example. Koelewijn and colleagues (submitted) found that sensitivity to F0 and VTL was higher in forward monosyllabic Dutch words compared to when they were time-reversed which resulted in smaller just-noticeable differences, measured in semitones. The authors of the study interpreted this top-down effect of lexical-content. Because we found only an effect of recording direction within nonwords and no differences in cue-weighting between words and nonwords in our categorisation task, it is very likely that the effect found in the discrimination study, investigating the same voice cues as we did, is also due to recording direction, and probably not due to lexical-semantic processes. However, this would need to be confirmed by using the same testing conditions as we did, to compare words and nonwords, and recording the direction of phonological processes within nonwords that do not transport any lexical information. If confirmed, this would give evidence that phonological processes enhance the perception of F0 and VTL that also affect the perceptual weighting of these cues, for example in a perceived voice gender categorization task.

Our finding that perceptual cue weights for VTL are higher than for F0 across linguistic conditions is in line with the findings of previous findings of studies that studied perceived voice gender categorization. Fuller and colleagues (2014) reported that when only manipulating one of the two voice cues, the extreme manipulations of -12.00 st in F0 and +3.6 st for VTL resulted in 10% and 30% of "man" categorizations, respectively. Nagels and colleagues (2020) found this overall higher weighting of VTL compared to F0, and this discrepancy was already present at the age of 6 years. Nagels et al., (2020) provide different explanations for the observed discrepancy between the two voice cues. For example, listeners might rely more on VTL than on F0 because this cue is more stable over time in speech. They support this explanation by referring to a categorization task with novel sounds (Mirman et al., 2004). In this categorisation task, the authors found a benefit for steady-state acoustic cues compared to rapidly changing acoustic cues. Another explanation could be the variability of F0 and VTL within female compared to female speakers and the variability of these vocal features that listeners attribute to single speakers and might play a role when listeners form abstract categories of speakers such as "man" or "woman: GPR, determining the F0 of the speaker, varies by 4.5 st within one speaker (Kania et al., 2006), while VTL varies by 1 st (Chuenwattanapranithi et al., 2009). Listeners thus accept greater variability of F0 than for VTL before they detected speaker change, namely 3.8 and 2.2 st (Gaudrain et al., 2009). Like some developmental theories that postulate that concrete forms are acquired before abstract categories are formed, it could be suggested that the differences in the intraindividual variability in F0 and VTL in male and female speakers drive/constrain the formation of the category of man's or a woman's voice. That could explain and why listeners accept less variability in VTL than in F0 when categorizing the perceived gender of the speaker which in turn results in heavier cue weights on VTL compared to F0.

Further, our finding also becomes relevant when studying the perception and use of F0 and VTL for voice identification when these phonological processes are still developing or hampered or impaired such as in developmental or acquired language and reading disorders, but also when hearing is impaired and listeners are provided with hearing aids/cochlear implants, or when the use of voice cues such as VTL for such identification tasks is limited as it is the case in CI users (Fuller et al., 2014). CI users are also often diagnosed to have impaired phonological abilities. If phonological processing enhances the sensitivity to and the perceptual weighting of F0 and VTL, this implies that these perceptual processes are restricted when phonological processes are not fully developed yet as in children, or impaired as it is the case in developmental or acquired language and reading impairments. Nagels and colleagues (2020) investigated the development of the sensitivity to F0 and VTL and their perceptual weighting in 4- to 12-year-old children to adult normal-hearing listeners. One of their findings was that their perceptual weighting of F0 and VTL became adult-like from the age of 8 to 12 years. Given the association of phonological processes and perceptual weighting of F0 and VTL that we found, this change in performance could alternatively be explained by the maturation of phonological abilities which are still developing in the age groups that differed significantly from the perceptual weighting. One finding that supports this threshold would be that, for example, children simplify final consonant clusters in productions until the age of 7 years which indicates that the phonological skills are still developing (Haaften et al., 2020). This relationship between the development of phonological processing and the use of F0 and VTL for voice recognition could be supported by studies in which these phonological abilities are impaired. The condition of developmental dyslexia has often been associated with an underlying phonological impairment (Snowling, 1998). This proposed association has led researchers to investigate a potential relationship between phonological abilities and voice recognition in this population and in voice recognition tasks in which listeners potentially also make use of voice cues such as F0 and VTL. Perrachione and colleagues (2011) found worse performance in speaker discrimination in individuals with dyslexia and impaired phonological memory and phonological awareness, and Kadam and colleagues (2016) found the same effect in participants with reduced, but not impaired reading abilities. However, in another study, conducted by Hazan and colleagues (2013), differences between unimpaired and impaired readers in speech-in-noise recognition was rather small so the authors concluded that the phonological representations in impaired readers are intact. To conclude, the mentioned studies investigated individuals whose phonological abilities are assumed to be developed incompletely or impaired, but in the studies that used the between-subject design, might be hard to confirm. Developmental dyslexia, for example, has been found to be associated with impairments in other cognitive domains, for example time perception and attention (Gooch et al., 2011) which makes the proposed association less plausible. The relationship of perceptual weighting of F0 and VTL and phonological processing in normal-hearing adults without a developmental language or reading impairment that we found in this experiment can give rise to other complications in speech-related tasks that are caused by an underlying phonological impairment. These might be even more relevant in children since they pick up linguistic information from the speech input to further their linguistic development.

CI-users are also found to deviate in their cue weighting from normal-hearing listeners, next to attributed phonological impairments. That means that the phonological effect we found in our study implies some methodological considerations when assessing/investigating/testing these individuals in speech-related tasks in which listeners rely on the voice cues F0 and VTL. In terms of the control linguistic test material for phonological features, such as phonemes, syllables, and syllable structure, or recording recording direction, while lexical parameters, for example word frequency, familiarity, word class and so on could be neglected. Fuller and colleagues (2014), who used the same testing procedure as we did, compared the cue weights of F0 and VTL in Ci users and in normal hearing adult listeners in vocoded and nonvocoded speech and showed that CI users mainly rely on F0. First, this could imply that CI users fail to distinguish a man's from a woman's voice when they are required to make this distinction in real life. Since F0 and VTL are voice cues that are used beyond the categorisation of women's and men's voices, the bias on F0 could possibly explain difficulties in other speech-related tasks, such as speech-on-speech perception, and probably also the recognition of familiar voices. While there is no evidence to date how this deviating weighting of voice cues could be further affected by phonological manipulations of the material, phonological abilities in CI users, especially in children, were researched extensively: The phonological development in CI using children differs from normal hearing children according to hearing age (Kral et al., 2014) and are associated with implantation age. Phonological abilities are also related to language abilities on other modalities, for example, d with receptive vocabulary (Lee et al., 2012) and with written word recognition (Bouton et al., 2015), and associated with implantation age (Johnson & Goswami, 2010). Future studies need to determine how these deviating phonological abilities would interact with phonological manipulations in the material. For example, if the phonological processes play a smaller role in CI users than in normal hearing listeners and thus their cue weighting of F0 and VTL is less affected by recording direction or by other phonological manipulations of the material.

Limitations

In this study, in which we aimed to investigate phonological effects on the perceptual weighting of F0 and VTL for voice gender categorisation can be related to the single manipulation namely by time-reversing the nonword which should interrupt phonological processing or the segmentation of the speech signal into its phonemes. This method is convenient because it retains other acoustic properties of the speech signal, for example the duration of the stimuli or the voice cues we manipulated such as F0 and VTL. However, this manipulation is distorting many phonological features at the suprasegmental and segmental level, such as voice 0nset times or intonation, and we cannot trace down which of them is crucial in the segmentation process that ultimately hampers perceptual cue weightings. Next to that, to preserve the Dutch phonotactics, we matched nonwords to words in terms of their phonological features, for example, phonological structure and number of syllables. This relates to another effect that has been described in the literature, the effect of phonological complexity on voice perception which has been associated with the attention towards linguistic stimuli (Conde et al., 2018) or the ability to recognize voices (Zarate et al., 2015). By using only monosyllabic words and nonwords in our study, we cannot exclude the possibility that the effect of complexity would rule out the effect of phonological processes or phonotactics in perceptual weighting of F0 and VTL.

The participants for this study were recruited via the online platform *Prolific*. They reported to have normal hearing and underwent a digit-in-noise task to test their hearing abilities. For 19 out of the 20 participants normal-hearing was confirmed, but this form of testing comes with uncertainties in terms of connection or different sound equipment and speakers. More objective testing, for example, audiometric testing would be preferable. In our study, we covered a wide range of ages, from 18 to 49 years. Nagels and colleagues (2020) considered if there was own-age bias for voice recognition that has been found for face recognition (Anastasi & Rhodes, 2005) that could lead to a bias in the categorisations. Further, we did not balance participants for gender which led to 2 women and 18 men taking part in the study. Male and female listeners have shown to perceive voice onset times differently (Kim, 2019). These voice onset times are violated in the condition in which we reversed the recording direction of the nonwords, and thus, this could have led to a bias.

Due to the online testing situation, the equipment to play the sounds could not be controlled and kept constant between participants and we could not measure how sound quality affected the voice manipulations. Therefore, participants were instructed to keep their volume and constant through the testing session as well as using the same equipment from the start to the beginning, but there are still interindividual differences in the equipment and how this affected the quality of the recordings could not be controlled.

Future directions

Our study showed that linguistic processes affect the perceptual weighting of F0 and VTL, two (important) voice cues which we investigated in a perceived voice gender categorisation task. In the wider literature in voice perception, these language effects have been studied intensively and often labelled as an “language familiarity” or “language ability” effect. To our knowledge, this is the first study that examined language effects specifically at the lexical and phonological level, and how these affect the perceptual weighting of two voice cues/F0 and VTL.

In our study, we found a phonological effect, but no lexical effect. However, with this categorization task, it is still unclear if these effects arise at the perceptual level, or only the perceptual weighting of F0 and VTL. When investigating perceived voice gender categorisation, Nagels and colleagues (2020) tested both discrimination and cue weighting and tested if their cue weights correlated in both tasks. In their study, only 4- to 6-year-old children showed a correlation of these abilities, and this is in line with the tasks effects that have been found in the wider voice perception literature, when investigating the language familiarity effect in discrimination as well as in identification tasks (Levi, 2019). Testing the same linguistic conditions in the discrimination task and in the categorisation task could reveal at which processing level these linguistic effects arise.

The comparison of words, nonwords and time-reversed words enabled us to disentangle lexical-semantic effects, and to control for lexical-semantic, phonological and acoustic features. We assumed that time-reversing recording direction would interrupt phonological processes, because time-reversing speech affects the segmental as well as the supra-segmental level, but it remains open what exactly is happening at the phonological level, if these are the violations of voice onset times, the consonant clusters that do not occur in the native language, or the reversed intonation or prosodic pattern that results from reversing vowel of the stimulus. That could be investigated by manipulating these phonetic features in particular. Another alternative would be to use longer stimuli to consider the effect of (phonological) complexity on voice perception, and to extend the material in terms of number of syllables or using anomalous sentences. By this, it could be examined if complexity or these longer stimuli would rule out the effect of phonological processing/the phonotactic parameters of the material.

Finally, considering the clinical application of this research, and to develop valid assessments and effective treatment methods, the phonological effect we found needs to be investigated when the development of phonological processes is impaired and voice cues are used and weighted differently, for example when weighting these cues for speaker identification, or to track them in speech on speech listening conditions, like it is the case in CI users. For CI users, it is generally assumed that these rely heavier on top-down processes, such as lexical-semantic and phonological effects, while being attributed with phonological impairments at the same time. By manipulating specific linguistic features of the material, this could reveal which potential cognitive processes alter the perceptual use of voice cues such as F0 and VTL which is associated with the listening difficulties in everyday life.

Conclusion

In conclusion, our study showed that the perception of F0 and VTL is altered by phonological processing, while we found no lexical effects by studying these cue weightings in a perceived voice gender categorisation task. The phonological effect we found shows that voice perception and language processing are intertwined, and that the perception of F0 and VTL, voices cues that are used beyond the simple categorisation of a man’s and woman’s voice, is enhanced when the speech input can be segmented into the phonemes of the language, probably the native language. Future efforts should be directed to the question where at the phoneme level these effects occur and how these phonological effects alter when phonological processing is impaired or the perception or use of specific voice cues is limited, as it is the case in CI users.

References

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, *12*(6), 1043–1047. https://doi.org/10.3758/BF03206441

Arts, F., Başkent, D., & Tamati, T. N. (2021). Development and structure of the VariaNTS corpus: A spoken Dutch corpus containing talker and linguistic variability. *Speech Communication*, *127*, 64–72. https://doi.org/10.1016/j.specom.2020.12.006

Ayache, S., Fernandes, M., Ouaknine, M., & Giovanni, A. (2002). [Function of the laryngeal muscles in the control of the fundamental frequency of voice]. *Annales D'oto-Laryngologie Et De Chirurgie Cervico Faciale: Bulletin De La Societe D'oto-Laryngologie Des Hopitaux De Paris*, *119*(4), 243–251.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The CELEX Lexical Database (CD-ROM)*. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2339741

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*(3), EL51–EL56. https://doi.org/10.1121/1.4942628

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bishop, D. V. M. (2006). What Causes Specific Language Impairment in Children? *Current Directions in Psychological Science*, *15*(5), 217–221. https://doi.org/10.1111/j.1467-8721.2006.00439.x

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, *130*(1), 85–95. https://doi.org/10.1016/j.cognition.2013.09.010

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence: Vocal communication of emotions. *Journal of Zoology*, *288*(1), 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–2538. https://doi.org/10.1121/1.1408946

Case, J., Seyfarth, S., & Levi, S. V. (2018). Short-term implicit voice-learning leads to a Familiar Talker Advantage: The role of encoding specificity. *The Journal of the Acoustical Society of America*, *144*(6), EL497–EL502. https://doi.org/10.1121/1.5081469

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. *International Journal of Psychophysiology*, *133*, 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Croquelois, A., & Bogousslavsky, J. (2011). Stroke Aphasia: 1,500 Consecutive Cases. *Cerebrovascular Diseases*, *31*(4), 392–399. https://doi.org/10.1159/000323217

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences*, *27*(4), 801–804. https://doi.org/10.1016/S0191-8869(98)00272-4

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *114*(5), 2913–2922. https://doi.org/10.1121/1.1616924

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dilley, L. C., Wieland, E. A., Gamache, J. L., McAuley, J. D., Redford, M. A., Kreiman, J., & Jacewicz, E. (2013). Age-Related Changes to Spectral Voice Characteristics Affect Judgments of Prosodic, Segmental, and Talker Attributes for Child and Adult Speech. *Journal of Speech, Language & Hearing Research*, *56*(1), 159–177. https://doi.org/10.1044/1092-4388(2012/11-0199)

Drozdova, P., van Hout, R., & Scharenborg, O. (2017). L2 voice recognition: The role of speaker-, listener-, and stimulus-related factors. *The Journal of the Acoustical Society of America*, *142*(5), 3058–3068. https://doi.org/10.1121/1.5010169

Eriksson, A. (1995). *The frequency range of the voice fundamental in the speech of male and female adults*. *Unpublished manuscript*, 11.

Fant, G. (1970). *Acoustic Theory of Speech Production*. Walter de Gruyter.

Fecher, N., & Johnson, E. K. (2018a). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America*, *143*(4), 2409–2418. https://doi.org/10.1121/1.5032199

Fecher, N., & Johnson, E. K. (2018b). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1911–1920. https://doi.org/10.1037/xlm0000555

Fecher, N., & Johnson, E. K. (2019). By 4.5 Months, Linguistic Experience Already Affects Infants' Talker Processing Abilities. *Child Development*, *90*(5), 1535–1543. https://doi.org/10.1111/cdev.13280

Fecher, N., & Johnson, E. K. (2021). Developmental improvements in talker recognition are specific to the native language. *Journal of Experimental Child Psychology*, *202*, 104991. https://doi.org/10.1016/j.jecp.2020.104991

Fecher, N., Paquette-Smith, M., & Johnson, E. K. (2019). Resolving the (Apparent) Talker Recognition Paradox in Developmental Speech Perception. *Infancy*, *24*(4), 570–588. https://doi.org/10.1111/infa.12290

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736. https://doi.org/10.1121/1.400247

Fitch, W. T. (1997). *Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques*. 11.

Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental Effects in Children's Ability to Benefit From F0 Differences Between Target and Masker Speech. *Ear and Hearing*, *40*(4), 927–937. https://doi.org/10.1097/AUD.0000000000000673

Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, *111*(38), 13795–13798. https://doi.org/10.1073/pnas.1401383111

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender Categorization Is Abnormal in Cochlear Implant Users. *Journal of the Association for Research in Otolaryngology*, *15*(6), 1037–1048. https://doi.org/10.1007/s10162-014-0483-7

Ganugapati, D., & Theodore, R. M. (2019). Structured phonetic variation facilitates talker identificationa). *J. Acoust. Soc. Am.*, 8.

Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America*, *137*(3), 1298–1308. https://doi.org/10.1121/1.4908235

Gaudrain, E., Li, S., Shen Ban, V., & Patterson, R. D. (2009, September). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. *Interspeech 2009*. https://hal.archives-ouvertes.fr/hal-02144510

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458. https://doi.org/10.3758/BF03199567

Gooch, D., Snowling, M., & Hulme, C. (2011). Time perception, phonological skills and executive function in children with dyslexia and/or ADHD symptoms. *Journal of Child Psychology and Psychiatry*, *52*(2), 195–203. https://doi.org/10.1111/j.1469-7610.2010.02312.x

Hazan, V., Messaoud-Galusi, S., Rosena, S., Schlauch, R., & Wright, B. (2013). The Effect of Talker and Intonation Variability on Speech Perception in Noise in Children with Dyslexia. *Journal of Speech, Language & Hearing Research*, *56*(1), 44–62. https://doi.org/10.1044/1092-4388(2012/10-0107)

Hodges-Simeon, C. R., Gurven, M., & Gaulin, S. J. C. (2015). The low male voice is a costly signal of phenotypic quality among Bolivian adolescents. *Evolution and Human Behavior*, *36*(4), 294–302. https://doi.org/10.1016/j.evolhumbehav.2015.01.002

Hoh, J. F. Y. (2010). Chapter 2.1—Laryngeal muscles as highly specialized organs in airway protection, respiration and phonation. In S. M. Brudzynski (Ed.), *Handbook of Behavioral Neuroscience* (Vol. 19, pp. 13–21). Elsevier. https://doi.org/10.1016/B978-0-12-374593-4.00002-4

Höhle, B. (2009). *Bootstrapping mechanisms in first language acquisition*. *47*(2), 359–382. https://doi.org/10.1515/LING.2009.013

Hu, X., Wang, X., Gu, Y., Luo, P., Yin, S., Wang, L., Fu, C., Qiao, L., Du, Y., & Chen, A. (2017). Phonological experience modulates voice discrimination: Evidence from functional brain networks analysis. *Brain and Language*, *173*, 67–75. https://doi.org/10.1016/j.bandl.2017.06.001

Idemaru, K., Winter, B., Brown, L., & Oh, G. E. (2020). Loudness Trumps Pitch in Politeness Judgments: Evidence from Korean Deferential Speech. *Language and Speech*, *63*(1), 123–148. https://doi.org/10.1177/0023830918824344

Irino, T., & Patterson, R. D. (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. *Speech Communication*, *36*(3), 181–203. https://doi.org/10.1016/S0167-6393(00)00085-6

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, *24*(10), 1995–2004. https://doi.org/10.1177/0956797613482467

Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America*, *139*(1), EL6–EL12. https://doi.org/10.1121/1.4937488

Kania, R. E., Hartl, D. M., Hans, S., Maeda, S., Vaissiere, J., & Brasnu, D. F. (2006). Fundamental Frequency Histograms Measured by Electroglottography During Speech: A Pilot Study for Standardization. *Journal of Voice*, *20*(1), 18–24. https://doi.org/10.1016/j.jvoice.2005.01.004

Kawahara, H., Masuda-Katsuse, I., & de Cheveign, A. (1999). Restructuring speech representations using a pitch-adaptive time±frequency smoothing and an instantaneous-frequency- based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 21.

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, *132*(2), 1050–1060. https://doi.org/10.1121/1.4730893

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Kim, S. K. (2019). Gender differences in voice onset time perception. *The Journal of the Acoustical Society of America*, *146*(4), 3053–3053. https://doi.org/10.1121/1.5137586

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87*(2), 820–857. https://doi.org/10.1121/1.398894

Leung, Y., Oates, J., & Chan, S. P. (2018). Voice, Articulation, and Prosody Contribute to Listener Perceptions of Speaker Gender: A Systematic Review and Meta-Analysis.

Journal of Speech, Language, and Hearing Research, *61*(2), 266–297. https://doi.org/10.1044/2017_JSLHR-S-17-0067

Levi, S. V. (2015). Talker familiarity and spoken word recognition in school-age children*. *Journal of Child Language*, *42*(4), 843–872. https://doi.org/10.1017/S0305000914000506

Levi, S. V. (2018). Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition*, *21*(3), 523–536. https://doi.org/10.1017/S1366728917000153

Levi, S. V. (2019). Methodological considerations for interpreting the Language Familiarity Effect in talker processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(2), e1483. https://doi.org/10.1002/wcs.1483

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *The Journal of the Acoustical Society of America*, *130*(6), 4053–4062. https://doi.org/10.1121/1.3651816

Loveday, L. (1981). Pitch, Politeness and Sexual Role: An Exploratory Investigation into the Pitch Correlates of English and Japanese Politeness Formulae. *Language and Speech*, *24*(1), 71–89. https://doi.org/10.1177/002383098102400105

Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., Pike, G. B., Veillette, S., Chakravarty, M. M., Pausova, Z., & Paus, T. (2016). Age- and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. *Hormones and Behavior*, *81*, 84–96. https://doi.org/10.1016/j.yhbeh.2016.03.001

Zarate, J.M., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*, *5*(1), 11475. https://doi.org/10.1038/srep11475

McLaughlin, D. E., Carter, Y. D., Cheng, C. C., & Perrachione, T. K. (2019). Hierarchical contributions of linguistic knowledge to talker identification: Phonological versus lexical familiarity. *Attention, Perception, & Psychophysics*, *81*(4), 1088–1107. https://doi.org/10.3758/s13414-019-01778-5

McLaughlin, D. E., Dougherty, S. C., Lember, R. A., & Perrachione, T. K. (2015). *Episodic Memory for Words Enhances the Language Familiarity Effect in Talker Identification*. 4.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, *131*(3), 2249–2260. https://doi.org/10.1121/1.3681950

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, *10*(1), 5074. https://doi.org/10.1038/s41598-020-61732-6

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, *149*(5), 3328–3344. https://doi.org/10.1121/10.0004791

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, *35*(1), 85–103. https://doi.org/10.1016/j.wocn.2005.10.004

Noble, W., & Gatehouse, S. (2004). Interaural asymmetry of hearing loss, Speech, Spatial and Qualities of Hearing Scale (SSQ) disabilities, and handicap. *International Journal of Audiology*, *43*(2), 100–114. https://doi.org/10.1080/14992020400050015

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human Voice Recognition Depends on Language Ability. *Science*, *333*(6042), 595–595. https://doi.org/10.1126/science.1207327

Perrachione, T. K., Furbeck, K. T., & Thurston, E. J. (2019). Acoustic and linguistic factors affecting perceptual dissimilarity judgments of voices. *The Journal of the Acoustical Society of America*, *146*(5), 3384–3399. https://doi.org/10.1121/1.5126697

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C., & Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour*, *95*, 89–99. https://doi.org/10.1016/j.anbehav.2014.06.011

Prolific | Online participant recruitment for surveys and market research. (n.d.). Retrieved 27 June 2021, from https://www.prolific.co/

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America*, *117*(2), 944–955. https://doi.org/10.1121/1.1848011

Roser, M., Appel, C., & Ritchie, H. (2013). Human Height. *Our World in Data*. https://ourworldindata.org/human-height

Sharma, N. K., Krishnamohan, V., Ganapathy, S., Gangopadhayay, A., & Fink, L. (2020). Acoustic and linguistic features influence talker change detection. *The Journal of the Acoustical Society of America*, *148*(5), EL414–EL419. https://doi.org/10.1121/10.0002462

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of Fundamental Frequency, Formant Frequencies, Aperiodicity, and Spectrum Level on the Perception of Voice Gender.

Journal of Speech, Language, and Hearing Research, *57*(1), 285–296. https://doi.org/10.1044/1092-4388(2013/12-0314)

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

Snowling, M. (1998). Dyslexia as a Phonological Deficit: Evidence and Implications. *Child Psychology*, *3*(1), 8.

Software voor het opnemen en bewerken van audio | Adobe Audition. (n.d.). Retrieved 5 June 2021, from https://www.adobe.com/nl/products/audition.html

Theodore, R. M., & Flanagan, E. G. (2020). Determinants of voice recognition in monolingual and bilingual listeners. *Bilingualism: Language and Cognition*, *23*(1), 158–170. https://doi.org/10.1017/S1366728919000075

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*(4), 1699–1707. https://doi.org/10.1121/1.397959

Titze, I. R., & Martin, D. W. (1998). *Principles of Voice Production*. *The Journal of the Acoustical Society of America*, *104*(3), 1148–1148. https://doi.org/10.1121/1.424266

Traunmüller, H., & Eriksson, A. (1995). *The frequency range of the voice fundamental in the speech of male and female adults*.

Unteregger, F., Honegger, F., Potthast, S., Zwicky, S., Schiwowa, J., & Storck, C. (2017). 3D analysis of the movements of the laryngeal cartilages during singing. *The Laryngoscope*, *127*(7), 1639–1643. https://doi.org/10.1002/lary.26430

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A Dissociation Between Familiar and Unfamiliar Voices. *Cortex*, *24*(2), 195–209. https://doi.org/10.1016/S0010-9452(88)80029-7 von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., & Griffiths, T. D. (2006). Processing the acoustic effect of size in speech sounds. *NeuroImage*, *32*(1), 368–375. https://doi.org/10.1016/j.neuroimage.2006.02.045

Whiteside, S. P. (1999.). A Comment on Women's Speech and its Synthe*sis. Perceptual and Motor skills, 88,* 110–112. https://doi.org/10.2466/pms.1999.88.1.110

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, *15*(1), 88–99. https://doi.org/10.1037/0882-7974.15.1.88

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and Voice Perception: Understanding Commonalities and Differences. *Trends in Cognitive Sciences*, *24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

Zaltz, Y., Goldsworthy, R. L., Eisenberg, L. S., & Kishon-Rabin, L. (2020). Children With Normal Hearing Are Efficient Users of Fundamental Frequency and Vocal Tract Length Cues for Voice Discrimination. *Ear & Hearing*, *41*(1), 182–193. https://doi.org/10.1097/AUD.0000000000000743

Zäske, R., Awwad Shiekh Hasan, B., & Belin, P. (2017). It doesn't matter what you say: FMRI correlates of voice learning and recognition independent of speech content. *Cortex*, *94*, 100–112. https://doi.org/10.1016/j.cortex.2017.06.005

Zimman, L. (2017). Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. *Language in Society*, *46*(3), 339–370. https://doi.org/10.1017/S0047404517000070

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. *7th International Conference on Speech Prosody 2014*, 1037–1041. https://doi.org/10.21437/SpeechProsody.2014-196

Appendix

Appendix 1

Participants' information

no	age (years)	gender	Native language	Normal-hearing confirmed in digit-in-noise test
1	23	Vrouw	Dutch, French	no
2	28	Man	Dutch	yes
3	36	Man	Dutch	yes
4	20	Vrouw	Dutch	yes
5	24	Man	Dutch	yes
6	20	Man	Dutch	yes
7	20	Man	Dutch, English	yes
8	30	Vrouw	Dutch, English German	yes
9	24	Man	Dutch	yes
10	22	Man	Dutch	yes
11	21	Man	Dutch	yes
12	38	Man	Dutch	yes
13	19	Man	Dutch	yes
14	26	Man	Dutch	yes
15	49	Man	Dutch, French, English	yes
16	18	Man	Dutch, English	yes
17	30	Man	Dutch	yes
18	28	Man	Dutch	yes
19	34	Man	Dutch	yes
20	25	Man	Dutch, English	yes

Appendix 1 gives the demographic information about participants collected from the responses of the PICKA questionnaire, the number of the participants, their age in years, their gender, native languages, if normal hearing was confirmed via the digit-in-noise test (Smits et al., 2004)