Word count: 8,806

Female Voices Impede Verbal Processing:
A Closer Look at the Role of Pitch and Formants

Laura Smorenburg

Utrecht University

Linguistics master's thesis

Dr. Aoju Chen

3 August 2017

Author's note

This thesis paper is a required component of the graduate research programme Linguistics at Utrecht University. The paper has been reviewed and graded by supervisor and first reader dr. Aoju Chen and second reader prof. dr. René Kager.

This study was approved by the Ethical Assessment Committee Linguistics (ETCL) of the Utrecht Institute of Linguistics OTS under ETCL reference number 3843386-01-2017.

This research is supported by a personal grant to supervisor dr. Aoju Chen. Special thanks to dr. Aoju Chen for her support and supervision of this thesis study.

Correspondence regarding this thesis paper should be addressed to B.J.L. (Laura) Smorenburg at: b.j.l.smorenburg@uu.nl.

Abstract

Previous research indicates that female voices are processed differently than male voices. For example, lists of words are remembered less well when spoken in a female voice, increased brain activity is visible in the auditory cortex for female voices in male listeners, and reaction times in lexical decision tasks seem to indicate longer reaction times for processing female voices. The current study investigated whether female voices are processed slower than male voices through a lexical decision task with auditory prime and specifically examined the role of pitch and formants in the supposed gender voice effect. Opposing previous findings, the current study does not find that word meaning access speed is slower in female voices compared to male voices as measured by the semantic priming effect. Additionally, female pitch or formants are not significant predictors for the semantic priming effect size, indicating that female pitch and formants have no role in words meaning access speed as measured by the semantic priming effect.

Keywords: Gender, verbal processing, semantic priming effect, pitch and formants

Table of contents

List of Tables and Figures

1. Introduction

Previous research indicates that female and male voices are processed differently. More specifically, female voices seem to impede verbal processing. Researchers have attributed this finding to the high acoustic salience and complexity of female voices. Typical female voices are characterised by increased values along several acoustic dimensions compared to male voices, including mean pitch, formant frequencies, and breathiness. Importantly, prosodic processing of female voices seems to require more cognitive resources than the prosodic processing of male voices. Behavioural findings report slower verbal processing of female voices compared to male voices and neuroimaging studies show increased brain activity in the auditory cortex for perception of female voices compared to male voices. If it is true that female voices require more cognitive resources for prosodic processing and thereby impede later processes such as semantic processing, implications are that the semantic contents of messages are conveyed less well in female voices. Subsequently, this finding may give rise to reinterpretations of the influence of female voices in communication and gender roles. Previous research investigating the effect of the female voice on verbal processing did not investigate the role of important voice features in gender perception. Therefore, the current study investigates the effect of pitch and formants on verbal processing. In addition, some methodological issues in previous behavioural research will be addressed.

1.1 Female voice features

Listeners can infer gender from voice, which is made possible by the fact that male and female voices are acoustically differentiable on several acoustic dimensions. The main distinguishing acoustic cue between genders is pitch, which is derived from fundamental frequency (f0). Male's vocal tracts are longer[1] (Fant, 1970) and their vocal cords longer and

[1] The average length of the vocal tract is 14.5 cm in females and 17 to 18 cm in males (Simpson, 2009).

thicker than females', causing the vocal cords to vibrate more slowly (Kahane, 1978), which all contributes to lower resonant frequencies in males relative to females. Studies report a mean pitch of 120 Hz for males and 200 Hz for females in American English and in Dutch (Takefuta, Jancosek, & Brunt, 1972; Tielen, 1992), although age (Pegoraro-Krook, 1988) and smoking behaviour (Gilbert & Weismer, 1974) may lower these numbers. On its own, pitch can acoustically distinguish speaker gender with 96% accuracy (Hillenbrand & Clark, 2009), which would suggest that listeners should be able to utilise pitch in isolation to perceive speaker gender. However, superimposing a female pitch on a male voice only leads to 34% female perception and superimposing a male pitch on a female voice only leads to 19% male perception (Hillenbrand & Clark, 2009). This finding indicates that gender perception from voice involves more than mapping pitch values onto gender categories and that more acoustic features are involved. One other important distinguishing voice feature between genders is vowel formant frequency. Vowel formant frequencies are higher in females than in males (Hillenbrand, Getty, Clark, & Wheeler, 1995) and the combination of the first three formants (F1-3) can acoustically distinguish speaker gender with 92% accuracy (Hillenbrand & Clark, 2009). Yet, listeners also do not seem to be able to use formant frequency (F1-3) as the only distinguishing cue in gender perception; superimposing female formants on a male voice only leads to 19% female perception and superimposing male formants on a female voice leads to 12% male perception (Hillenbrand & Clark, 2009). Hence, neither pitch nor formants in isolation has a decisive role in perceived gender. However, the combination of pitch and formants seems to be a reliable cue for gender perception. Superimposing female pitch *and* formants on a male voice leads to 82% female perception and superimposing male pitch *and* formants on a female voice likewise leads to 82% male perception, suggesting that pitch and formants make up an important part of gender-related voice characteristics (Hillenbrand & Clark, 2009).

Although perceiving gender with 82% accuracy using only pitch and formant information could be described as successful gender perception, accuracy is higher in original male and female voices, i.e. 99.6% for both male and female voices (Hillenbrand & Clark, 2009). Other voice features may have a small contribution to gender perception. For example, phonation type is correlated with gender. Females tend to have breathier voices than males (Klatt & Klatt, 1990), whereas males tend to have creakier and tenser voices than females (Tielen, 1992; Van Rie, 1993). However, the role of phonation type in perceived gender has not been investigated. Some studies also claim that females' pitch ranges are wider than males' (e.g. Takefuta et al., 1972). Other studies report that males have wider pitch ranges (cf. Simpson, 2009). For Dutch, no gender differences were found for f0 range or variability. Instead, these measures seem to make up individual differences and cannot be generalised to a gender (Tielen, 1992). Contesting the finding of equal pitch ranges for both genders in declarative speech, e.g. Tielen (1992), Haan & van Heuven (1999) have shown that females make larger local pitch excursions in questions, resulting in wider *local* pitch ranges for females relative to males. Other voice features than pitch and formants may thus distinguish male from female voices, but their impact on perceived gender has not been investigated.

In brief, female voices are distinguishable from male voices by their increased values for mean pitch, formants, breathiness, and size of local pitch excursions. The combination of pitch and formant information has a substantial role in gender perception, but it is currently unclear what role these features play in previously reported verbal processing differences for male and female voices.

1.2 Prosodic processing of extra-linguistic information

The fact that listeners can perceive speaker gender from voice means that listeners process extra-linguistic information. Information such as gender identity, sociocultural identity, affect, and attitude are generally conveyed through prosody by means of intonation, rhythm, speaking rate, and other prosodic measures. There are two main views on processing of extra-linguistic prosodic information. In the traditional abstractionist view, the incoming speech signal is normalised such that all the “noisy” prosodic information is filtered out of the linguistic signal (e.g. Nearey, 1990, 1997), leaving only prototypical, abstract representations consisting of single phonemic, or gestural units. Opposing this theory is the detailed encoding view, which postulates that linguistic information cannot be separated from extra-linguistic prosodic information and that all available information in the rich acoustic signal may be processed and used (e.g. McMurray & Jongman, 2011). Many findings support a detailed encoding view of verbal processing. For example, Pisoni (1993) showed that listeners commit extra-linguistic prosodic information such as talker-identity characteristics, emotional state and speaking rate into long-term memory. Similarly, Bradlow, Nygaard, & Pisoni (1999) found that changes in speaker and speaking rate resulted in more errors in determining whether a word was “old” or “new” in a list of words, again suggesting that extra-linguistic prosodic information is encoded and stored in memory. These findings indicate that extra-linguistic prosodic information is processed along with linguistic information and is not a disposable by-product of speech.

Moreover, past work also suggests processing costs for extra-linguistic prosodic information, i.e. impaired verbal processing in the presence of extra-linguistic prosodic information[2]. For example, older listeners who listened to a story told with neutral prosody remembered more words than those who listened to the same story read with positive or negative prosody (Fairfield, Domenico, Serricchio, Borella, & Mammarella, 2016).

---

[2] Some studies also show impeded verbal processing in the presence of *linguistic*, in this case phonological, information; relative clauses that contain words with phonological overlap (*baker* – *banker*) are read more slowly and processed less accurately than the non-overlapping controls (Acheson & MacDonald, 2011).

Additionally, Magnuson & Nusbaum (2007) showed that the expectation of speaker variability results in slower verbal processing even when there is no speaker variability; listeners who were expecting to hear a different speaker each trial reacted slower to the target than listeners who were expecting to hear the same speaker. The authors suggest that the expectation of speaker variability heightens perceptual sensitivity, i.e. listeners displayed processing cost for heightened sensitivity to extra-linguistic prosodic information. Correspondingly, Church & Schacter (1994) stress the importance of pitch information in the perceptual representation of words and show that auditory priming effects are impaired in implicit auditory identification and stem completion tasks when a prime is followed by a target that differs from the prime in pitch or speaker. Manipulation of decibel level, on the other hand, did not impair priming effects. The presence of extra-linguistic prosodic information may thus influence verbal processing of words.

1.3 Gender and prosodic processing of extra-linguistic information

Some research has investigated the role of speaker gender in verbal processing. There are several findings suggesting that female voices require more processing than male voices. Firstly, an experiment by Yang, Yang, & Park (2013) on directed forgetting found that when one group of participants is directed to forget word list 1 and remember word list 2 and another group is directed to remember both word lists, participants in both groups remember fewer words from list 1 when the lists were spoken in a female voice. The authors argue that the acoustic salience and complexity of female voices draws attention to the voice features and therefore poses more processing cost on extra-linguistic prosodic elements, resulting in impeded semantic processing and word recall for female voices. Surprisingly, this same experiment finds no main effect for angry prosody, which the authors characterise as higher pitched and more intense than the voice in the neutral condition. In fact, in this experiment the male speech in the angry condition has a higher mean f0 than the female speech in the neutral condition[3], yet more words from list 1 are recalled when the lists are spoken in an angry male voice compared to the neutral female voice (Yang et al., 2013). It thus seems that f0 is not responsible for the main effect of gender on directed forgetting such that increased f0 values would cause impeded word recall. One possible explanation for the gender effect that Yang et al. (2013) found on directed forgetting is that directed forgetting may not be correlated with pitch, but with perceived gender. Pitch in isolation has a limited role in perceived gender. When only pitch is increased in a male voice, as is the case in the male angry prosody, the perceived gender generally does not change from male to female (cf. Hillenbrand & Clark, 2009). Yang et al.'s (2013) results thus indicate that female voices require more processing than male voices, but the exact source of the processing difference for male and female voices cannot be derived from these results.

In support of Yang et al.'s (2013) finding that female voices require more processing than male voices, fMRI research with male listeners shows that male and female voices activate distinct regions in the male brain. More specifically, regions of the auditory cortex that are involved in interpreting prosody[4] are more activated by perception of female voices than by perception of male voices in male listeners, whereas brain areas involved in mental imagery are more activated by perception of male voices in male listeners (Sokhi, Hunter, Wilkinson, & Woodruff, 2005). The authors note that the observation that female voices trigger increased brain activation of the auditory cortex area which maps human qualities (e.g.

[3] Female neutral speech has a mean f0 of 213 Hz and male angry speech has a mean f0 of 268 Hz (Yang, Yang, & Park, 2013).

[4] In male listeners, the right anterior superior temporal gyrus is more activated in perception of female voices than male voices; male voices, on the other hand, elicit more brain activation in the mesio-parietal precuneus area, an area also referred to as “the mind’s I” and is involved in episodic memory and imagining of sounds (Sokhi, Hunter, Wilkinson, & Woodruff, 2005).

gender) to an acoustic voice signal is in line with the hypothesis that female voices are more acoustically complex and require more prosodic processing than male voices. However, this finding has not been replicated with female participants and Sokhi et al's (2005) results might be interpreted differently if the opposite effect is found for female participants, i.e. female participants might show increased brain activity for male voices compared to female voices.

Zhang & Lee (2011) also show differences in word meaning access speed for male and female voices. They investigated the role of speaker-gender variability in verbal processing through a lexical decision task with auditory priming. The priming effect is taken to reflect semantic processing and consists of the difference in reaction time to the target word between semantically related word pairs (e.g. *king – queen*) and reaction time to the target word in unrelated word pairs (e.g. *bell – queen*). Generally, it is expected that semantically related primes facilitate activation of the target word whereas unrelated primes do not, resulting in faster reaction times (i.e. faster processing) of targets that are preceded by related primes (cf. Spreading activation model: Collins & Loftus, 1975). In turn, the prime or target may be acoustically manipulated to investigate which voice features mediate the priming effect and thus influence semantic processing. Zhang & Lee (2011) manipulated speaker gender and report that when a prime is spoken in a male voice and a target is spoken in a female voice, the priming effect was attenuated compared to the condition in which both prime and target were spoken in the same female voice, i.e. processing is facilitated when the speaker of the prime and target are the same. However, when a prime is spoken in a female voice and the target is spoken in a male voice, no attenuation of the priming effect was observed compared to the condition in which both prime and target were spoken in the same male voice. Additionally, although Zhang & Lee (2011) do not explicitly compare the same-speaker conditions in their paper, results indicate that the priming effect is larger in the female same-speaker condition than in the male same-speaker condition. At first glance, this finding seems to contradict the hypothesis that female voices are processed more slowly than male voices. However, the raw reaction times show much longer overall reaction times for female voices than for male voices, indicating that, although the priming facilitation effect is greater in the female same-speaker condition, lexical access speed is faster for the male voice in this experiment (Zhang & Lee, 2011). It is, however, hard to interpret the results obtained by Zhang & Lee (2011) because of the confounding variable listener gender. That is, Zhang & Lee (2011) did not use a balanced participant group, which consisted of 45 females and 15 females, and moreover failed to include the variable listener gender in the analysis. An ERP study by Wirth et al. (2007) shows that female listeners perform deeper semantic processing[5] than males in passive reading of words. The authors state that this deeper semantic processing results in "faster processing of related words in the active neural networks" (2007, p. 1987) in females, which implies that one might expect a larger priming effect in female listeners than in male listeners. It is currently unclear if this listener gender effect interacts with speaker gender as none of the previous research has included both of these variables. Additionally, although the priming effect can be a useful tool to investigate mediating variables in semantic processing, the most typical use of priming tasks seems inherently flawed. More specifically, by posing variability within prime-target pair, i.e. manipulating a voice feature in the prime but not the target such that the word pairs that participants hear differ from one another on this voice feature, participants could just be responding to the unexpected voice difference between prime and target. For example, in the word pair *king – queen*, Zhang & Lee (2011) only manipulated the speaker gender of the prime word *king* in half of the trials, such that half of the trials contained prime-target pairs spoken in the same voice and half of the trials contained prime-target pairs in which the prime was spoken in a male voice and the target in a

[5] Females showed earlier and longer lasting N400 effects as well as longer lasting activation of temporal networks relative to males (Wirth et al., 2007).

female voice. The opposite presentation order (prime in female voice and target in male voice) also occurred. Pu et al. (2005) have shown that it is very difficult to distinguish the priming effect from difference detection. As a result, Zhang & Lee (2011) can only make claims about whether hearing the same voice in the prime-target pair facilitates semantic priming and whether a different voice (in this case a different gender) inhibits semantic priming, in which case difference detection theoretically equals the semantic priming effect. This problem may be solved by experimentally manipulating prime and target between pairs, instead of within pairs. In other words, the voice features of *king* and *queen* are identical such that the reaction time to the target *queen* is not influenced by difference detection but rather by the experimental manipulation only.

In sum, previous research on the role of voice gender in verbal processing indicates that female voices require more prosodic processing than male voices. For example, fewer words are recalled from lists spoken by female speakers compared to male speakers, more brain activity is visible in the auditory cortex for female voices compared to male voices, and overall reaction times are slower and semantic priming effects larger for female voices in a lexical decision task with semantic priming.

2. Research questions and hypotheses

As discussed in section 1, previous research indicates that female voices seem to impede verbal processing. This finding has been explained by the high acoustic salience and complexity of female voices relative to male voices. The typical female voice can be characterized by increased values along several acoustic dimensions including pitch, formants, breathiness, and local pitch excursions. The implication from previous research is that the relatively high acoustic salience and complexity of female voices requires more cognitive resources for prosodic processing, consequently delaying or impeding higher processes in verbal processing, such as semantic processing. However, it is currently not clear what in the female voice, i.e. which acoustic voice features, are responsible for this gender effect in verbal processing. Moreover, some methodological flaws have been observed in previous behavioural studies, namely the exclusion of listener gender as an independent variable and the possible interfering effect of difference detection between prime and target in priming experiments. The current study expands on previous work by investigating the role of two specific voice features in verbal processing, by including listener gender as a variable, and by eliminating possible difference detection effects within each trial. The current study will test three hypotheses to investigate whether female voice attributes interfere with semantic processing.

Based on previous research that seems to show impeded and delayed verbal processing for female voices relative to male voices (Yang et al., 2013; Zhang & Lee, 2011), it is hypothesized that a female voice impedes verbal processing. More specifically, lexical access speed is expected to be slower and semantic facilitation is expected to be larger in female voice conditions as measured by the semantic priming effect size (difference in reaction time between related and unrelated word pairs). This prediction for the semantic priming effect is based on the assumption that female voices are implied to impede semantic processing because of increased prosodic processing cost in female voices. Prosodic processing is an earlier process in verbal processing and is shown to take more time and show more brain activation when listening to female voices compared to male voices; preactivation of targets through related primes should thus speed semantic processing relatively more for female voices than male voices as semantic word access in male voices already seems “optimal”, i.e. access to word meaning is not hindered by prosodic processing load to the same extent in male voices as in female voices. Secondarily, previous findings which indicate that female listeners conduct deeper semantic analysis than male listeners suggest that we might expect a main effect of listener gender. Female listeners are predicted to produce larger priming effects than male listeners.

Secondly, based on previous research indicating that pitch is one of the main voice features for gender perception from voice (Hillenbrand & Clark, 2009), we hypothesise that female pitch impedes verbal processing and that male pitch facilitates verbal processing. Lastly, as previous research shows that formants have limited power in changing gender perception from voice (Gelfer & Mikos, 2005; Hillenbrand & Clark, 2009; Poon & Ng, 2011), we hypothesise that female formants do not impede verbal processing and male formants do not facilitate verbal processing. Having different predictions for pitch and formant manipulation depending on perceived gender, it is expected that there is a significant interaction between manipulation type and perceived gender.

3. Method

3.1 Experimental design

To investigate the role of voice gender, perceived gender, listener gender, pitch, and formants on verbal processing, three steps were taken. First, stimuli and fillers were recorded from a male and female speaker and a subset of materials acoustically manipulated to create a male voice that carries female pitch or formants as well as a female voice that carries male pitch or formants. Secondly, each target word was rated on perceived gender in a rating task. Thirdly, verbal processing in the different voice conditions was tested in a lexical decision task with auditory priming. The resulting data was then analysed in a 2×2×2 design: voice manipulation (pitch, formants) × voice source gender (female, male) × and listener gender (female, male). Additionally, the continuous independent variable perceived gender was included in the analysis for the manipulated voice conditions. The dependent variable is the priming effect size, which is derived from the difference in reaction time to target words between related and unrelated word pairs in the lexical decision task.

3.2 Participants

Thirty-eight native speakers of Dutch (15 males, 23 females, age: M = 26.62, SD= 10.92) were recruited through the UiL OTS participant database to partake in this study. None of the participants reported to have dyslexia or any hearing defects. Four participants reported to have more than one native language. Prior to participation, participants were asked to read an information letter and sign a participation approval form. Participants received financial compensation of €7 for their participation as per the standards of the UiL OTS research lab where the experiments were conducted. This study design was approved by the Ethical Assessment Committee Linguistics (ETCL).

3.3 Materials

3.3.1 Experimental stimuli and fillers

The Dutch materials in this study were adapted from an associative priming study (Geuze, Gerven, Farquhar, & Desain, 2013) and consisted of words taken from the Leuven Association Database (De Deyne & Storms, 2008). Experimental stimuli consist of 48 unique word sets that contained a target, a related prime[6] and an unrelated prime (see example 1a), which were presented to the participant as two separate word pairs (e.g. *naald – draad* and *roest – draad*). An equal number of 48 word sets consisting of a pseudoword target and two

[6] Related word pairs have an association strength of at least 0.1, meaning that participants named the target word following the probe in at least ten per cent of all cases in the first three responses in a continuous association task (De Deyne & Storms, 2008).

primes acted as fillers (see example 1b). Word pairs with phonological overlap (initial CV or final CVC) were excluded to prevent interfering phonological priming effects in the output data.

(1) Stimuli examples

a.	unrelated prime	related prime	target	
	roest	naald	draad	(*rust – needle – thread*)
b.	unrelated prime	unrelated prime	target	
	boom	fiets	kloen	(*tree – bicycle –* pseudoword)

Partly following the stimuli design of Hillenbrand & Clark (2009), test stimuli and fillers occur in six experimental voice conditions in a 3×2 design: voice manipulation type (no manipulation, formants, pitch), and source voice gender (female, male). Each word pair occurred twice, once in the male source voice and once in the female source voice, meaning that each target word appeared four times to the participant. Both test stimuli and fillers consisted of 48 stimuli sets each (see Appendix A), which were subdivided into three balanced lists of 16 word sets for the three different experimental voice conditions per source voice gender. Participants were presented with 384 trials in total, half of which were fillers. The stimuli sets across experimental voice conditions were matched on word length, word frequency[7], concreteness, age of acquisition, and neighbourhood size[8], because it has

[7] Word frequency was based on the logarithmic frequency of words in the SUBTLEX-NL database (Keuleers, Brysbaert, & New, 2010), which is a database of Dutch word frequencies based on 44 million words from television and film subtitles.

previously been shown that lexical access speed is mediated by these measures (De Deyne & Storms, 2008; Keuleers, Brysbaert, & New, 2010; Moor & Brysbaert, 2000). Creating these balanced word lists was accomplished with computer programme Match (Van Casteren & Davis, 2007). Independent sample t-tests on the matched measures show that there are no significant differences between the target stimuli for each voice condition for any of the matched measures. The exact matching statistics can be found in Appendix B.

For the presentation order, experimental items and fillers were ordered with computer programme Mix, which generated three different pseudorandom orders such that neither the same voice condition (original voice, formants, pitch), nor the same type (related, unrelated, or nonword) were repeated more than two times in a row. Additionally, because target words occur four times across type and voice condition, the minimal distance between identical target words was set at eight trials.

3.3.2 Acoustic manipulation of pitch and formants

One male and one female speaker with Standard Dutch accents and typical male and typical female voices were recruited to record the stimuli for this study and received €5.00 for their contribution to this study. Recordings were made with a Zoom H1 Handy Recorder using a 44100 Hz sampling frequency (16 bit accuracy rate) in a sound attenuated booth. The speakers were asked to speak clearly at a normal volume, with clear pauses between words, and with falling intonation for each word. Acoustic manipulation of stimuli sets was done in computer programme Praat (Boersma & Weenink, 2017). All recordings were firstly

---

[8] Neighbourhood size was balanced across voice conditions on the following measures: Phoneme Levenshtein Distance (minimum number of substitutions, insertions, or deletions required to turn one word into another), and Coltheart's N (the number of words that can be produced by changing a phoneme in a word of the same length).

normalised on amplitude. Secondly, recordings were analysed on pitch and formant frequencies (F1-F3) so that averages could be established for both the male and female speaker (see Table 1).

Table 1

Acoustic measurements male and female voices

Measure	N	Male	Female	*t*	*df*	*P*
dur	192	.46 (.08)	.66 (.12)	-20.10	382	<.001***
new_dur	192	-	.46 (.28)	-.43	382	.67
pitch	192	97.77 (16.16)	205.34 (32.19)	-40.38	382	<.001***
F1	192	737.72 (165.13)	794.67 (166.38)	-3.37	382	<.001***
F2	192	1720.58 (247.77)	1810.30 (251.42)	-3.52	382	<.001***
F3	192	2758.41 (202.51)	2910.74 (193.30)	-7.37	382	<.001***

Note. Duration is displayed in seconds, pitch measurements in Hertz. The original word duration was significantly different between the male and female speaker. Durations of words spoken by the female speaker were compared to the male pronunciation and adjusted accordingly per item, such that each item had comparable length in the male and female voice. The original and duration-adjusted items were presented to four native speakers of Dutch, who judged whether the original or adjusted duration sounded more natural. A one sample t-test (0 = original sounded more natural, 1 = duration adjusted sounded more natural) shows that scores were significantly different from zero; $t(159) = 23.40$, $p < .001$. Participants judged the adjusted, sped-up version as more natural sounding in 76.50% of all cases.

Following Hillenbrand & Clark (2009), the female/male ratios for formant values were calculated from the averages in Table 1, such that acoustic manipulations of formants could be based on these ratios. Formant shift ratios and new absolute pitch median values were then used in the internal Praat function “change gender”, through which the formant frequencies can be shifted by ratios and the pitch median can be assigned a new absolute value. This Praat function changes pitch or formants of a sound through TD-PSOLA overlap-add synthesis. To superimpose male formants on the original female voice in this study, formants had to be shifted by a ratio of 0.95. To superimpose female formants on the original male voice, the inverted ratio was used. The new pitch median corresponded to the mean pitch for the intended gender manipulation as found in Table 1. An example manipulation with formant and pitch contours can be found in Appendix C.

3.3.3 Perceived gender in manipulated stimuli

To check if the acoustic manipulations have the expected effect on perceived gender before running the lexical decision task experiment, three male and five female native speakers of Dutch (age: M = 27.16, SD = 8.96) were recruited to participate in a rating task in which they were asked to judge whether the speaker of the experimental target words sounded “male” or “female” for all experimental voice conditions used in the current study. Since raters are asked to judge natural male and female voices as well as acoustically manipulated voices that might be perceived as unnatural, this forced binary choice was accompanied by a 5-point Likert scale on which raters were asked to indicate how certain they were of their choice. Descriptive statistics for perceived gender rating scores and rating certainty scores per manipulation and voice source gender type can be found in Table 2.

Table 2

Perceived gender and rating certainty scores

Manipulation type	Perceived gender		Rating certainty	
	Source female	Source male	Source female	Source male
None (original voice)	.98 (.14)	.00 (.00)	4.66 (.73)	4.94 (.25)
Formants	.99 (.10)	.15 (.36)	4.64 (.67)	3.24 (1.13)
Pitch	.56 (.50)	.83 (.38)	2.67 (1.26)	3.31 (1.31)

Note. Perceived gender scores represent the percentage of ‘female’ ratings, i.e. a score of 1 represents 100% ‘female’ ratings, a score of 0 represents 100% ‘male’ ratings. Rating certainty scores represent scores on a 5-point Likert scale ranging from ‘very uncertain’ (1) to ‘very certain’ (5). The first row of means represents the unmanipulated voice conditions, i.e. the original male and female voices.

A three-way mixed ANOVA was done to ascertain whether there were effects of voice manipulation type (no manipulation, formants, pitch), voice source gender (female, male) and rater gender (female, male) on the outcome measures ‘perceived gender’ and ‘certainty score’. Perceived gender ratings showed a significant main effect of manipulation type; $F(2, 12) = 7.04$, $p < .01$. Post hoc pairwise comparisons show a significant difference ($p = .02$) between the original voice (M = .49) and pitch manipulation (M = .68), but no significant difference (p

= .24) between the original voice (M = .49) and formant manipulation (M = .57). Perceived gender ratings also showed a significant main effect of voice source gender; $F(1, 6) = 280.38$, $p < .001$. Post hoc pairwise comparisons indicate a significant difference ($p < .001$) between the male voice source (M = .32) and female voice source (M = .84). Additionally, there was a significant interaction between manipulation type × voice source gender; $F(2, 12) = 31.98$, $p < .001$. This shows that the main effect of voice manipulation type significantly differed between the female and male voice source (see Figure 1). Perceived gender ratings showed no significant main effect of rater gender; $F(1, 6) = .21$, $p = .66$. Rater gender did not significantly interact with source gender ($F(1, 6) = .78$, $p = .41$) or with manipulation type ($F(2, 12) = .26$, $p = .78$). In other words, female and male raters did not differ in their perceived gender ratings.

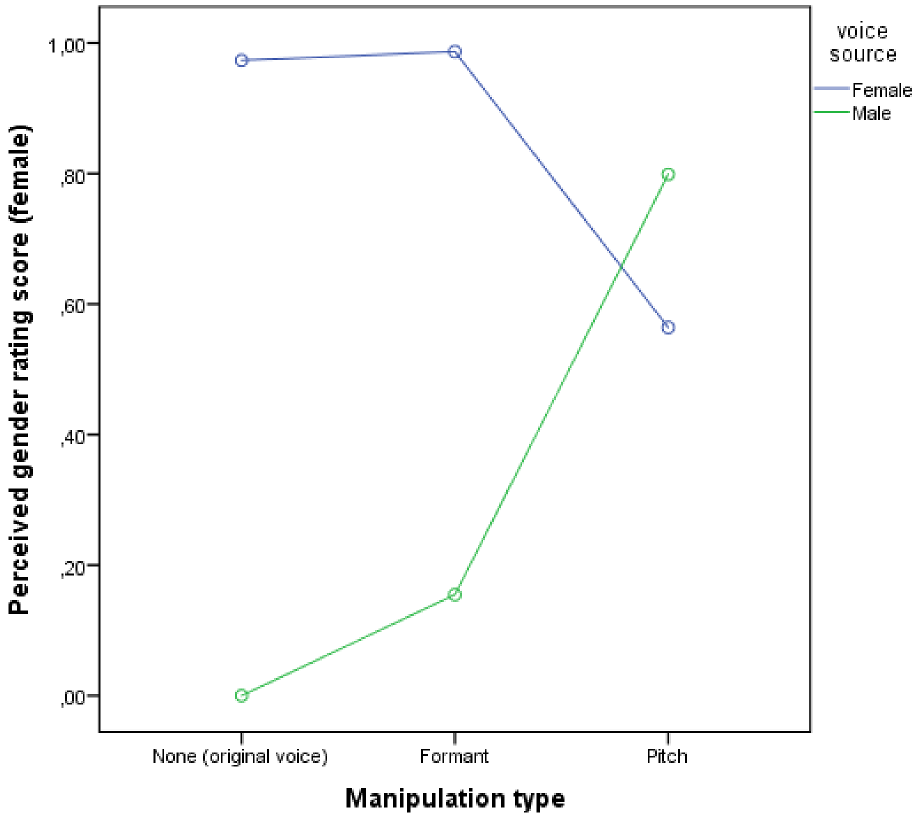

Figure 1. Perceived gender rating scores per manipulation type and voice source gender

The perceived gender rating task also included a certainty rating. Rating certainty scores showed a significant main effect of manipulation type on certainty ratings; $F(2, 12) = 22.35$, $p < .001$. Post hoc pairwise comparisons indicate that rating certainty scores for the original voice ($M = 4.78$) significantly differ from both the formant manipulation ($M = 3.91$, $p < .01$) and the pitch manipulation ($M = 3.02$, $p < .01$). There was no significant difference in rating certainty scores between the formant and the pitch manipulation ($p = .10$). There was no significant main effect of voice source gender; $F(1, 6) = .55$, $p = .49$. However, there was an interaction between manipulation type and voice source gender; $F(2, 12) = 12.23$, $p < .001$. This indicates that the main effect of voice source gender on rating certainty significantly differed between manipulation types (see Figure 2). Lastly, rating certainty scores showed no main effect of rater gender; $F(1, 6) = .00$, $p = .99$).

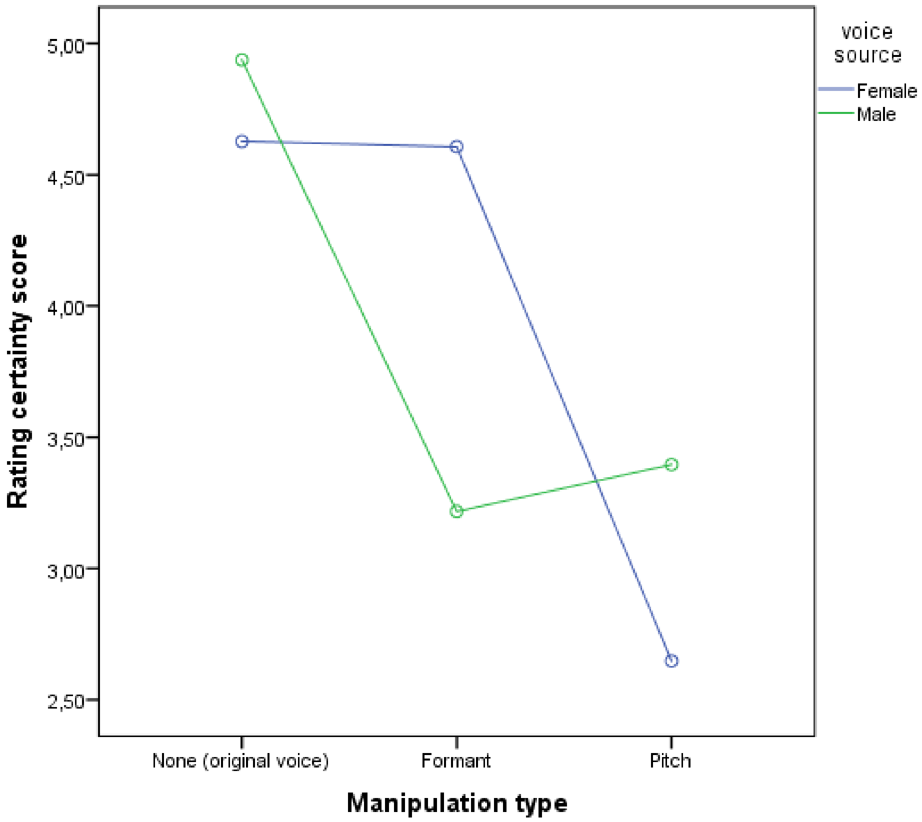

Figure 2. Rating certainty scores per manipulation type and voice source gender

Overall our results show that the effect of acoustic manipulations on perceived gender in the current study conform to expectations from the literature. As expected, the original male and female voices were perceived correctly in almost all cases. Corroborating findings by Hillenbrand & Clark (2009), the current study finds that pitch manipulation affects perceived gender. However, opposed to Hillenbrand & Clark (2009), the current study does not find that formant manipulation significantly affects perceived gender, i.e. there were no significant differences between the original voice condition and the formant manipulated voice condition. Other previous research is consistent with the current finding that formant manipulation has little to no effect on perceived gender (e.g. Poon & Ng, 2011). Certainty ratings were higher for the original male and female voice conditions compared to the manipulated voice conditions, showing a general effect of voice manipulation on certainty ratings. In line with findings by Hillenbrand & Clark (2009), participants were less certain in their perceived gender ratings of manipulated voices. This is most likely a consequence of the manipulated voice conditions sounding less natural than the original voice conditions.

3.4 Procedure

For the lexical decision task with auditory priming, the participants were asked to seat themselves behind a computer screen in a sound attenuated booth located at the UiL OTS lab in Utrecht. In front of them a button box containing a yes-button and a no-button was placed. An auditory lexical decision with auditory priming task was run in software programme ZEP (Veenker, 2017). Sound was played over BeyerDynamic DT770 headphones. The participants were asked to respond to auditory targets that were preceded by primes and classify the targets as existing words of Dutch or pseudowords/nonwords. The experimental trials were presented in four blocks of 96 trials, each of which took around eight minutes to complete. After each block participants were asked to take a two-minute pause. The experiment screen displayed participant's progress in the experiment block and a visual yes-button and no-button reflecting the button box so that no mistakes were made regarding which button on the button box designated a "yes" versus a "no" response. In line with previous research (e.g. Zhang & Lee, 2011), response accuracy and reaction time were measured from the target onset. The interval between prime and target was specified at 250 ms. The intertrial interval was specified at 1500 ms. Including instructions, practice trials, and three two-minute pauses, the task took around 40 minutes for participants to complete.

4. Results

Responses to filler pseudo-word targets were excluded from analysis. As the lexical decision task was auto-paced and the response window was set at the target duration plus 1500 ms, there are 43 missing values resulting from responses outside of this response window. The missing responses comprise 0.01% of all items. A total number of 428 incorrect responses were excluded from analysis, which comprised 11.87%[9] of all items. Additionally, Luce (1986) has shown that valid reaction times are minimally 100 ms long and a minimum cut-off point between 100 and 200 ms is generally used to trim reaction time data (Whelan, 2008). However, our data did not include data points below 200 ms, so no minimum cut-off point was used. No general agreements exist about maximum reaction times cut-off points, so no maximum cut-off point was used. The analysis of the lexical decision data was performed in three separate analyses.

In the first analysis, a paired samples t-test was performed to verify the assumption that reaction times to unrelated target words were significantly longer than reaction times to related targets. Additionally, correlations were calculated to see if there was a relationship

[9] Each priming effect data point contained two correctness values (one for the unrelated target word trial and one for the related target word trial). Data points were excluded when one or both were listed as incorrect.

between priming effect size and absolute reaction time measures to test if a larger priming effect reflects slower verbal processing and greater semantic priming facilitation.

Secondly, to test our first hypothesis on voice gender and verbal processing, a two-way mixed ANOVA was performed to test for differences in our dependent variable priming effect size for between-subject variable listener gender and within-subject variable original voice gender in a 2×2 design: listener gender (female, male) × original voice gender (female, male). Only the data for the original female and original male voice conditions were included in this analysis.

Thirdly, to test our second and third hypotheses on the roles of pitch and formants on verbal processing, a multiple regression analysis was performed on the data for the manipulated voice conditions only in a 2×2×2 design consisting of the following predictor variables: manipulation type (pitch, formants) × voice source gender (female, male) × listener gender (female, male). Additionally, continuous independent variable perceived gender was included as a predictor variable.

4.1 ANALYSIS I

4.1.1 Semantic priming facilitation effect

A paired samples t-test was conducted to compare absolute reaction times to target words in unrelated prime trials and related prime trials. There was a significant difference in the reaction times for targets in the unrelated prime trials (M = 840.89, SD = 267.19) and the related prime trials (M = 728.57, SD = 224.86); $t(3176) = -20.87$, $p < .001$. These results show that reaction times were faster in related prime trials compared to unrelated prime trials, i.e. there is a semantic priming facilitation effect.

4.1.2 Correlations priming effect and absolute reaction times

Pearson product-moment correlation coefficients were computed to assess the relationship between the priming effect size and absolute reaction times. There was a positive correlation between priming effect size and absolute reaction times in the unrelated prime trials ($r = .66$, $n = 1198$, $p < .001$). There was a negative correlation between priming effect size and absolute reaction times in the related prime trials ($r = -.61$, $n = 1198$, $p < .001$). Increases in priming effect size were thus correlated with increases in absolute reaction times for unrelated prime trials, but were correlated with decreases in absolute reaction times for related prime trials.

4.2 ANALYSIS II

A two-way mixed ANOVA including only the data for the original voice conditions was performed to test for differences in priming effect size for between-subject variable listener gender and within-subject variable original voice gender. Priming effect size did not show a significant main effect of original voice gender; $F(1, 36) = .03$, $p = .87$. This shows that the original female voice does not trigger larger priming effect sizes than the original male voice. This suggests that original female voices are not responded to more slowly than male voices. Listener gender did not show a significant effect for priming effect size either; $F(1, 36) = .00$, $p = .99$. There was no significant interaction between original voice gender and listener gender; $F(1, 36) = .40$, $p = .53$. The results do seem to indicate that there might be a slight trend towards an interaction between original voice gender and listener gender (see Figure 3).

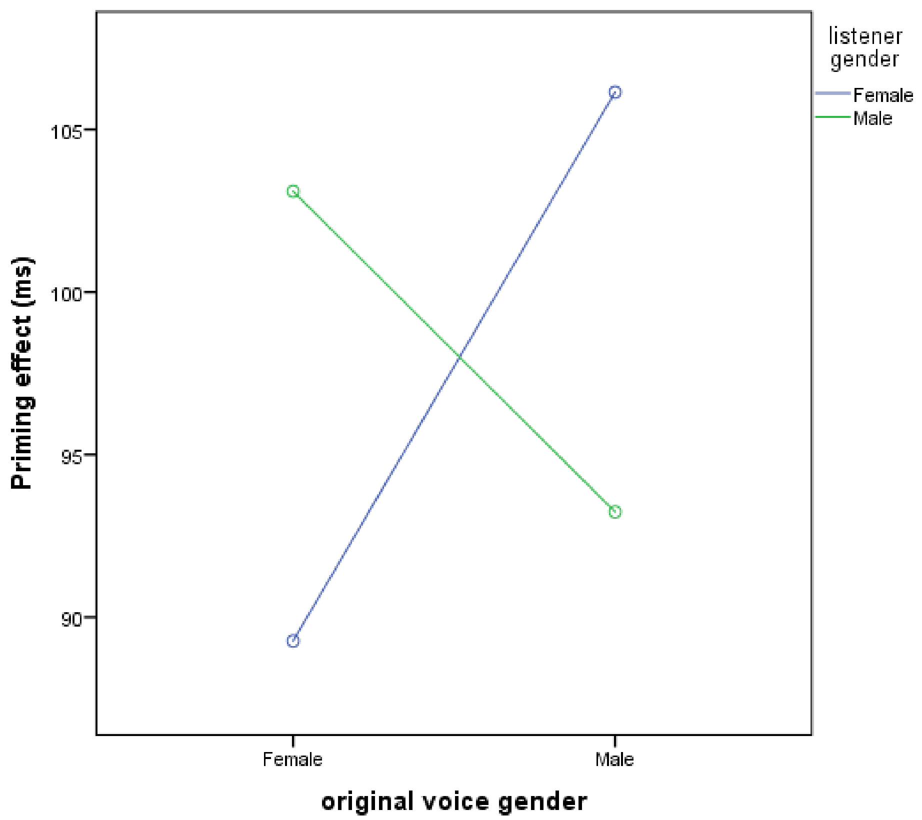

Figure 3. Priming effect size per original voice gender and listener gender

4.3 ANALYSIS III

A linear multiple regression analysis was performed to investigate the roles of pitch and formants in verbal processing through the semantic priming effect. Independent variables manipulation type, voice source gender, perceived gender and listener gender are included in the multiple regression model as predictors. Following general practices (Shieh, 2011), continuous independent variable perceived gender was mean-centered, i.e. each value was subtracted with the mean to mitigate multicollinearity between our predictor variables and the interaction terms, Our dichotomous gender predictor variables were coded in the following way: ‘male’ was coded by ‘0’ and ‘female’ was coded by ‘1’. This means that the regression model will take the dependent variable scores for the ‘male’ category as the reference category to compare the scores for the 'female' category against. Coefficients for voice source gender and listener gender will thus represent the relation between gender and priming effect size for the 'female' category. For dichotomous variable manipulation type, 'formant manipulation' was coded by '0' and 'pitch manipulation' by '1'. The coefficient for predictor variable manipulation type will thus represent the relation between manipulation type and priming effect size for the 'pitch manipulation' category, taking the 'formant manipulation' category as the reference category.

4.3.1 Building the regression model

Predictor variables are added to the model through block wise entry in a hierarchical fashion, first adding the predictor variables and then adding the interaction terms.

Table 3

Hierarchical block wise entry of predictor variables

Model	
1	Manipulation type
	Source gender
	Listener gender
	Perceived gender
2	Manipulation × source gender
	Manipulation × perceived gender
	Manipulation × listener gender
	Source × perceived gender
	Source × listener gender
	Perceived gender × listener gender

4.3.2 Interpreting the regression model

A multiple linear regression was calculated to predict the semantic priming effect size based on manipulation type, voice source gender, perceived gender and listener gender. For model

1, no significant regression equation was found ($F(4, 2055) = .18$, $p = .95$), with an R^2 of .00. Neither was there a significant regression equation found for model 2 ($F(10, 2049) = .90$, $p = .53$), with an R^2 of .00. This means that .00% of the variance in priming effect size was predicted by our predictor variables in both model 1 and model 2. Coefficients per predictor variable and interaction terms are presented in Table 4.

Table 4

Coefficients regression models

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	124.34	14.45		8.60	< .000
	manipulation type	6.13	13.47	.01	.46	.65
	source gender	-5.81	14.94	-.01	-.39	.70
	perceived gender	-.42	2.65	.00	-.16	.87
	listener gender	-6.53	13.49	-.01	-.48	.63
2	(Constant)	154.99	54.67		2.84	< .000
	manipulation type	-64.78	67.57	-.11	-.96	.34
	source gender	37.25	107.76	.06	.35	.73
	perceived gender	11.78	13.76	.11	.86	.39
	listener gender	12.01	24.60	.02	.49	.63
	manipulation × source gender	-4.77	98.49	-.01	-.05	.96
	manipulation × perceived gender	5.15	22.59	.02	.23	.82
	manipulation × listener gender	9.26	27.41	.01	.34	.74
	source gender × perceived gender	-3.37	2.41	-.17	-1.49	.14
	source gender × listener gender	-44.16	3.38	-.07	-1.45	.15
	perceived gender × listener gender	-.79	5.37	-.01	-.15	.88

The unstandardized coefficients in Table 4 represent the predicted value of the priming effect for each model. In other words, in model 1 participants' predicted priming effect size equals 124.34 + 6.13 (manipulation type) + 5.81 (source gender) + .42 (perceived gender) - 6.53 (listener gender). However, none of the predictor variables or interaction variables are significant predictors of priming effect size in either model.

5. Discussion

5.1 Semantic priming effect and absolute reaction times

The literature on linguistic priming effects has shown that lexical decision reaction times are faster when targets are preceded by semantically related primes compared to when they are preceded by semantically unrelated primes. The current study has replicated this result.

Correlations between the priming effect size and the absolute reaction time measures from the unrelated prime and related prime trials show that an increase in priming effect size is correlated with slower reaction times in unrelated trials and faster reaction times in related trials. In other words, priming effect size reflects slower word meaning access in unrelated trials and greater facilitation effect in related trials. The priming effect size measure can thus be used to make statements about word meaning access speed (in unrelated prime trials) and facilitation of word meaning access (in related prime trials).

5.2 Original voice gender and the priming effect

Contrary to expectations from the literature suggesting that female voices are processed more slowly than male voices, our second analysis including the original voice conditions showed no significant differences in priming effect size for our female and male voice conditions. Lexical access speed is thus not slower and semantic facilitation is not larger in female voice conditions compared to the male voice conditions. In other words, a female voice does not seem to impede verbal processing of words. Hence, our first hypothesis based on previous findings is not supported by the results of this study. However, results do seem to indicate a slight trend for an interaction between original voice gender and listener gender (note that this interaction was not significant). Female listeners seem to produce smaller priming effect sizes for the original female voice condition relative to the original male voice condition and male listeners seem to produce smaller priming effect sizes for the original male voice condition relative to the female voice condition (see Figure 3 in section 4.2). Although more data and analysis is needed to verify this new hypothesis, it is possible that the previously reported effects for the female voice might in actuality be an effect of ‘the voice of the opposite gender’. This hypothesis has not been checked in previous research in which the participant groups were either all male (Sokhi et al., 2005), or in which listener gender was not included as an independent variable in the analyses (Zhang & Lee, 2011).

5.3 Pitch and formants and the priming effect

Given the substantial role for pitch but limited role for formants in perceived gender, we hypothesized that the interaction term for manipulation type × perceived gender would be a significant predictor for priming effect size. In other words, the effect of perceived gender on priming effect size would be different between manipulation types because, in general, pitch manipulation changes perceived gender whereas formant manipulation does not. The results of the regression analysis do not support our second and third hypothesis on the role of formants and pitch in verbal processing as measured by the semantic priming effect, as there was no significant interaction term coefficient for manipulation type × perceived gender. In fact, none of the predictor variables and interaction terms significantly predicted priming effect size and the models could explain .00% of the variance in priming effect size. This means that word meaning access speed is not slower and semantic facilitation is not larger in voice conditions in which a male voice has female formants or pitch compared to voice conditions in which a female voice has male formants or pitch. In other words, female formants and pitch do not seem to impede verbal processing of words.

6. Conclusion

Contrary to indications in previous research suggesting slower verbal processing for the female voice, the current study finds no indications that words spoken by a female voice are processed more slowly than words spoken by a male voice as measured by the semantic priming effect. Female pitch or formants also do not seem to slow processing of words. We want to suggest that previous findings might have been due to some methodological flaws and the exclusion of the independent variable listener gender. Previous priming research (e.g. Zhang & Lee, 2011) has imposed voice gender variation within prime-target trial, meaning that the measured reaction time in these studies might be due to difference detection between the switch in voice genders. When difference detection to this voice gender switch is eliminated by posing voice gender variation between prime-target trials instead of within trials, no effect of the female voice is found, i.e. the independent variables original/source voice gender and manipulation type are not predictors for the priming effect size.

In the future, it might be interesting to study whether the tentatively suggested ‘voice of the opposite gender’-effect may be verified. Future research may use neuroimaging techniques to investigate this hypothesis. Although it has been shown that male listeners show increased brain activity in the auditory cortex for female voices compared to male voices (Sokhi et al., 2005), this study has not been replicated with female participants. Sokhi et al. (2005) furthermore indicate that male participants listening to male voices showed brain activation in the mesio-parietal precuneus area, which is an area involved with the imagining of sounds and is also sometimes referred to as “the mind’s ear” (p. 577). Reduplicating this neuroimaging research with female participants may indicate whether activation in this area referred to as “the mind’s ear” is associated with similarity of speaker voice gender and listener voice gender and whether increased activation in the auditory cortex is associated with dissimilarity between speaker voice gender and listener voice gender.

References

Acheson, D. J., & MacDonald, M. C. (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language*, *65*(2), 193–207. https://doi.org/10.1016/j.jml.2011.04.006

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.26). Retrieved from http://www.praat.org/

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219. https://doi.org/10.3758/BF03206883

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521–533. https://doi.org/10.1037/0278-7393.20.3.521

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

De Deyne, S., & Storms, G. (2008). Word associations: norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198–205.

Fairfield, B., Domenico, A. D., Serricchio, S., Borella, E., & Mammarella, N. (2016). Emotional prosody effects on verbal memory in older and younger adults. *Aging, Neuropsychology, and Cognition*, *0*(0), 1–10. https://doi.org/10.1080/13825585.2016.1219690

Fant, G. (1970). *Acoustic Theory of Speech Production, With Calculations based on X-Ray Studies of Russian Articulations*. The Hague: De Gruyter Mouton. https://doi.org/10.1515/9783110873429

Gelfer, M. P., & Mikos, V. A. (2005). The Relative Contributions of Speaking Fundamental Frequency and Formant Frequencies to Gender Identification Based on Isolated Vowels. *Journal of Voice*, *19*(4), 544–554. https://doi.org/10.1016/j.jvoice.2004.10.006

Geuze, J., Gerven, M. A. J. van, Farquhar, J., & Desain, P. (2013). Detecting Semantic Priming at the Single-Trial Level. *PLOS ONE*, *8*(4), e60377. https://doi.org/10.1371/journal.pone.0060377

Gilbert, H. R., & Weismer, G. G. (1974). The effects of smoking on the speaking fundamental frequency of adult women. *Journal of Psycholinguistic Research*, *3*(3), 225–231. https://doi.org/10.1007/BF01069239

Haan, J., & van Heuven, V. J. (1999). Male vs. female pitch range in Dutch questions. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1581–1584). Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1581.pdf

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166.

Kahane, J. C. (1978). A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy*, *151*(1), 11–19. https://doi.org/10.1002/aja.1001510103

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

Klatt, D., & Klatt, L. (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

Luce, R. D. (1986). *Response time: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 391–409. https://doi.org/10.1037/0096-1523.33.2.391

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219–246. https://doi.org/10.1037/a0022325

Moor, W. D., & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research*, *63*(2), 159–162. https://doi.org/10.1007/PL00008174

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, *18*, 347–373.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, *101*(6), 3241–3254.

Pegoraro-Krook, M. I. (1988). Speaking Fundamental Frequency Characteristics of Normal Swedish Subjects Obtained by Glottal Frequency Analysis. *Folia Phoniatrica et Logopaedica*, *40*(2), 82–90. https://doi.org/10.1159/000265888

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, *13*(1–2), 109–125.

Poon, S., & Ng, M. (2011). Contribution of voice fundamental frequency and formants to the identification of speaker's gender. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong.

Pu, J., Peng, D., Demaree, H. A., Song, Y., Wei, J., & Xu, L. (2005). The recognition potential: Semantic processing or the detection of differences between stimuli? *Cognitive Brain Research*, *25*(1), 273–282. https://doi.org/10.1016/j.cogbrainres.2005.06.001

Shieh, G. (2011). Clarifying the role of mean centring in multicollinearity of interaction effects. *British Journal of Mathematical and Statistical Psychology*, *64*(3), 462–477.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Sokhi, D. S., Hunter, M. D., Wilkinson, I. D., & Woodruff, P. W. R. (2005). Male and female voices activate distinct regions in the male brain. *NeuroImage*, *27*(3), 572–578. https://doi.org/10.1016/j.neuroimage.2005.04.023

Takefuta, Y., Jancosek, E., & Brunt, M. (1972). A Statistical Analysis of Melody Curves in the Intonation of American English. Presented at the Seventh International Congress of Phonetic Sciences, The Hague: Mouton. Retrieved from https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/AD765389.xhtml

Tielen, M. T. J. (1992). *Male and Female Speech: An Experimental Study of Sex-related Voice and Pronunciation Characteristics*. Univ.

Van Casteren, M., & Davis, M. H. (2007). Match: a program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, *39*(4), 973–978.

Van Rie, J. (1993). Voice quality description of speakers of standard Dutch. Unpublished report, Nijmegen University.

Veenker, T. J. G. (2017). The Zep Experiment Control Application (Version 1.10). Beexy Behavioral Experiment Software. Retrieved from http://www.beexy.org/zep/

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475.

Wirth, M., Horn, H., Koenig, T., Stein, M., Federspiel, A., Meier, B., … Strik, W. (2007). Sex Differences in Semantic Processing: Event-Related Brain Potentials Distinguish between Lower and Higher Order Semantic Analysis during Word Reading. *Cerebral Cortex*, *17*(9), 1987–1997. https://doi.org/10.1093/cercor/bhl121

Yang, H., Yang, S., & Park, G. (2013). Her Voice Lingers on and Her Memory Is Strategic: Effects of Gender on Directed Forgetting. *PLOS ONE*, *8*(5), e64030. https://doi.org/10.1371/journal.pone.0064030

Zhang, Y., & Lee, C.-Y. (2011). Talker variability in lexical access: Evidence from semantic priming. *The Journal of the Acoustical Society of America*, *129*(4), 2662–2662.

Appendices

Appendix A

Table 5

Experimental stimuli

Condition	Item id	Target	Related prime	Unrelated prime
Original voice	e1	clown	circus	pil
Original voice	e2	glad	ijzel	toekomst
Original voice	e3	nacht	donker	pan
Original voice	e4	bos	bomen	verhaal
Original voice	e5	snoep	zoet	eindpunt
Original voice	e6	vogel	merel	fornuis
Original voice	e7	zwart	kraai	saai
Original voice	e8	boog	pijl	zwaan
Original voice	e9	kaas	rasp	struik
Original voice	e10	groente	bloemkool	cent
Original voice	e11	pasta	pizza	goot
Original voice	e12	duif	vrede	bout
Original voice	e13	groot	reus	forel
Original voice	e14	sprookje	fabel	loper
Original voice	e15	hamer	spijker	wolk
Original voice	e16	dik	vet	winst
Pitch manipulation	e17	vork	bestek	bus
Pitch manipulation	e18	grijs	bewolkt	procent
Pitch manipulation	e19	bult	kameel	zaag
Pitch manipulation	e20	muziek	geluid	blauw
Pitch manipulation	e21	fles	kurk	aap
Pitch manipulation	e22	zout	chips	balg
Pitch manipulation	e23	pad	kikker	veter
Pitch manipulation	e24	slaap	bed	album
Pitch manipulation	e25	wortel	konijn	ontzag
Pitch manipulation	e26	portret	foto	piste
Pitch manipulation	e27	voetbal	elftal	begrip
Pitch manipulation	e28	dief	inbraak	drankje
Pitch manipulation	e29	trap	leuning	kool
Pitch manipulation	e30	vlinder	rups	pub
Pitch manipulation	e31	lip	mond	specht
Pitch manipulation	e32	kassa	winkel	worst
Formant manipulation	e33	fruit	mango	nieuw
Formant manipulation	e34	draad	naald	wolf

Formant manipulation	e35	droog	dorst	strijd
Formant manipulation	e36	bloed	vampier	plint
Formant manipulation	e37	schaap	wol	vlieger
Formant manipulation	e38	vinger	nagel	pap
Formant manipulation	e39	tennis	racket	antiek
Formant manipulation	e40	brand	rook	contract
Formant manipulation	e41	rat	muis	dol
Formant manipulation	e42	taart	vlaai	dak
Formant manipulation	e43	arend	vogel	juf
Formant manipulation	e44	biljart	keu	wekker
Formant manipulation	e45	fabriek	werk	jurist
Formant manipulation	e46	groen	gras	bruis
Formant manipulation	e47	venster	raam	gek
Formant manipulation	e48	geit	bok	spuit

Table 6

Filler stimuli

Condition	Item id	Pseudoword	Prime 1	Prime 2
Original voice	f1	pems	rugzak	akker
Original voice	f2	kloen	kort	training
Original voice	f3	spopt	dieet	bijl
Original voice	f4	buster	naam	paal
Original voice	f5	prel	dolk	baan
Original voice	f6	nit	tafel	hobby
Original voice	f7	visken	trui	toga
Original voice	f8	blorukt	hamster	stof
Original voice	f9	duk	plant	grond
Original voice	f10	staas	kokos	glijbaan
Original voice	f11	stus	jacht	deurknop
Original voice	f12	snaat	rog	machine
Original voice	f13	flindes	arena	halte
Original voice	f14	kie	haan	jurk
Original voice	f15	kloor	meer	pony
Original voice	f16	zoe	tang	bromfiets
Pitch manipulation	f17	miek	kraan	geld
Pitch manipulation	f18	stesser	blok	zolder
Pitch manipulation	f19	niel	pakket	lijn
Pitch manipulation	f20	pleem	magneet	kanon
Pitch manipulation	f21	conklect	ham	zwempak
Pitch manipulation	f22	spomget	schaal	schaaf

Pitch manipulation	f23	pabat	roos	plaat
Pitch manipulation	f24	slulp	piloot	ouder
Pitch manipulation	f25	dem	pit	vleugel
Pitch manipulation	f26	vuut	haring	beschuit
Pitch manipulation	f27	hawooi	sloot	zuster
Pitch manipulation	f28	megel	wiel	neus
Pitch manipulation	f29	spinsen	hoorn	saus
Pitch manipulation	f30	sastaal	eenhoorn	trauma
Pitch manipulation	f31	bawo	poef	reptiel
Pitch manipulation	f32	lepen	vezel	ketel
Formant manipulation	f33	wokker	stok	paling
Formant manipulation	f34	spoon	actie	mot
Formant manipulation	f35	fimf	lama	honing
Formant manipulation	f36	ganen	ballon	datum
Formant manipulation	f37	oener	borstel	slof
Formant manipulation	f38	sora	verf	knie
Formant manipulation	f39	elber	micro	plezier
Formant manipulation	f40	trij	lood	zanger
Formant manipulation	f41	spebel	slee	lolly
Formant manipulation	f42	schien	winnaar	erwt
Formant manipulation	f43	muin	hart	geluid
Formant manipulation	f44	pazoor	roest	honger
Formant manipulation	f45	lenk	beker	stop
Formant manipulation	f46	speven	boek	kiwi
Formant manipulation	f47	broen	anker	hoef
Formant manipulation	f48	lossa	paard	mus

Appendix B

Table 7

Descriptive statistics: Stimuli matching

Matched variable	N	Original voice Mean	Original voice SD	Pitch Mean	Pitch SD	Formant Mean	Formant SD
Freq	16	5666.81	8947.81	3444.12	4121.39	3899.75	6578.83
PhonCnt	16	4.25	1.18	4.44	1.36	4.63	1.20
SylCnt	16	1.31	0.48	1.38	0.50	1.38	0.50
Concreteness	16	4.34	.55	4.56	.39	4.63	.33
AoA	16	5.32	1.01	5.48	.982	5.75	1.24
PLD30	16	1.5	.30	1.6	.472	1.65	.45
Cont_Nphon	16	12.25	9.43	11.63	9.55	9.63	9.58

Note. Values for all measures were taken from the SUBTLEX-NL database (Keuleers et al., 2010). AoA refers to age of acquisition, PLD30 refers to Phoneme Levenshtein Distance (the minimum number of substitutions, insertions, or deletions required to turn one word into another), and Colt_Nphon refers to Coltheart's N (the number of words that can be produced by changing a phoneme in a word of the same length.

Table 8

Stimuli matching: Independent samples t-test results

Matched variable	Original voice × Pitch *t*	*df*	*p*	Original voice × Formant *t*	*df*	*p*	Pitch × Formant *t*	*df*	*p*
Freq	.90	30	.37	.64	30	.53	-.23	30	.82
PhonCnt	-.42	30	.68	-.89	30	.38	-.41	30	.68
SylCnt	-.36	30	.72	-.36	30	.72	.00	30	1.00
Concreteness	-1.33	30	.19	-1.78	30	.09	-.48	30	.64
AoA	-.46	30	.65	-1.08	30	.29	-.68	30	.50
PLD30	-.74	30	.46	-1.12	30	.27	-.28	30	.78
Cont_Nphon	.19	30	.85	.78	30	.44	.59	30	.56

Appendix C

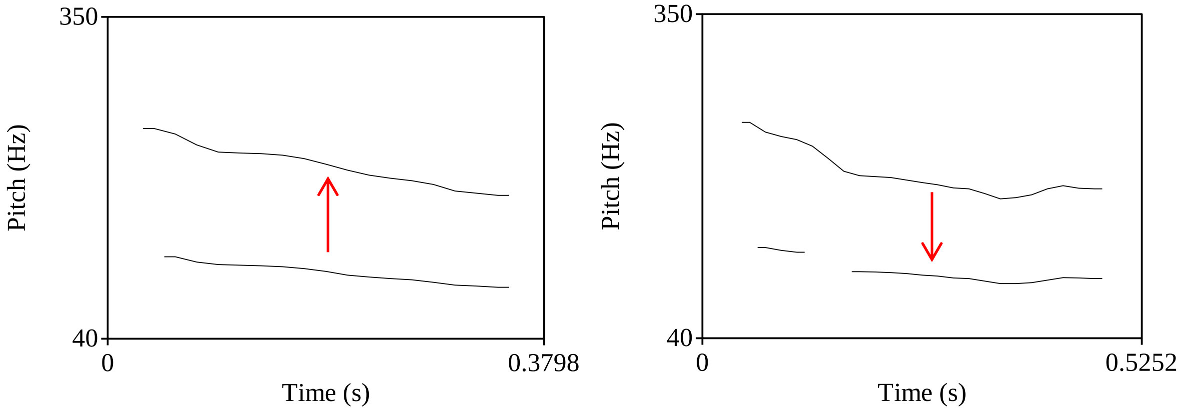

Figure 4. Example pitch manipulation of Dutch word "clown".
Note. Left: original male pitch (104.7 Hz) and shifted up pitch (208.1 Hz). Right: original female pitch (188.7 Hz) and shifted down pitch (100.5 Hz).

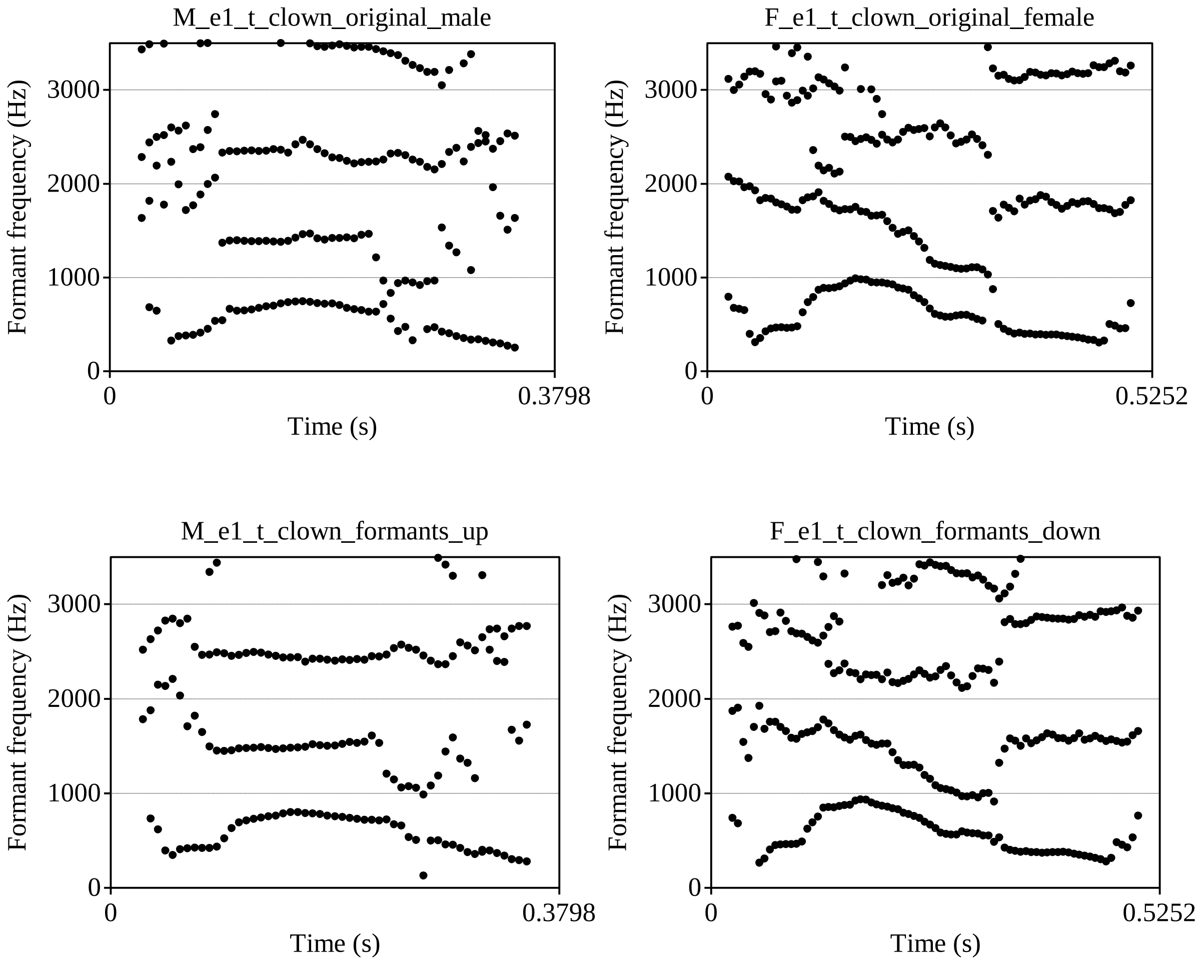

Figure 5. Example formant manipulation of Dutch word "clown".