



# The effect of female voice on verbal processing

Laura Smorenburg<sup>a,\*</sup>, Aoju Chen<sup>b</sup>

<sup>a</sup> Leiden University Centre for Linguistics, Van Wijkplaats 4, 2311 BX Leiden, The Netherlands

<sup>b</sup> Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

## ARTICLE INFO

### Keywords:

Gender  
Verbal processing  
Semantic priming effect  
Pitch  
Formants

## ABSTRACT

Previous studies have suggested that female voices may impede verbal processing. For example, words were remembered less well and lexical decision was slower when spoken by a female speaker. The current study tried to replicate this gender effect in an auditory semantic/associative priming task that excluded any effects of speaker variability and extended previous research by examining the role of two voice features important in perceived gender: pitch and formant frequencies. Additionally, listener gender was included in the experimental design. Results show that, contrary to previous findings, there is no evidence that a lexical decision of a target word is slower when spoken by a female speaker than by a male speaker for female and male listeners. Additionally, the semantic/associative priming effect was not affected by speaker gender, neither did female mean pitch or formants predict the semantic/associative priming effect. At the behavioural level, the current study found no evidence for a gender effect in a semantic/associative priming task.

## 1. Introduction

Previous research has shown that female voices impede verbal processing. Specifically, the verbal processing of female voices has been argued to require more cognitive resources than the verbal processing of male voices. This is manifested in slower verbal processing in behavioural findings and increased brain activity in the auditory cortex in neuroimaging studies. Researchers have attributed these findings 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. It is suggested the female voice is more complex: “Although the precise parameters that define the complexity of female voices have not been fully described, the idea is suggested by evidence that female voices, compared to male voices, are more difficult to both recognise (Noyes and Frankish, 1989) and convincingly synthesise (Klatt, 1987) using computer technology” (Sokhi et al., 2005: 577).

However, previous behavioural studies on the effect of female voices on verbal processing have focussed on speaker variability and not on the effect of the female voice in isolation. Also, the specific role of important voice features for gender<sup>1</sup> classification in the processing of voices re-

mains to be examined. In the current study, we aim to test the suggestion in previous findings that female voices impede verbal processing and to extend previous research by examining the role of two voice features, i.e. pitch and formant frequencies.

### 1.1. Acoustic features of the female voice

Listeners can infer gender from voice as male and female voices are differentiable on several acoustic dimensions. The main distinguishing acoustic cue between genders is mean pitch, which is derived from fundamental frequency ( $f_0$ ). Male speakers have a longer vocal tract (Fant, 1970; Simpson, 2009), and longer and thicker vocal cords than female speakers. Male speakers' vocal cords thus vibrate more slowly, given the same amount of air from the lungs (Kahane, 1978), causing lower fundamental frequencies in males relative to females. Studies report a mean pitch of 120 Hz for males and 200 Hz for females in general in American English and in Dutch (Takefuta et al., 1972; Tielen, 1992), although age (Pegoraro-Krook, 1988) and smoking behaviour (Gilbert and Weismer, 1974) may alter these numbers. On its own, mean pitch values can acoustically distinguish speaker gender with 96% accuracy (Hillenbrand and Clark, 2009; Kreiman and Sidtis, 2011: 125). This finding would

(cf. Kreiman and Sidtis, 2011: 142–147). In this paper, we are concerned with both the sex of the speaker and the gender of the voices from the listener's perspective. The term “gender” seems to have a broader connotation than “sex” and is therefore used throughout this paper.

\* Corresponding author.

E-mail addresses: [B.J.L.Smorenburg@hum.leidenuniv.nl](mailto:B.J.L.Smorenburg@hum.leidenuniv.nl) (L. Smorenburg), [Aoju.Chen@uu.nl](mailto:Aoju.Chen@uu.nl) (A. Chen).

<sup>1</sup> Recently, researchers have been trying to distinguish effects of gender and effects of sex in speech, which are two highly correlated, but not synonymous variables. Research in this area has been focussed on the speech of children, individuals with different sexual orientations, and transgendered individuals to tease apart biological and learned factors in speech behaviour

suggest that listeners should be able to utilise mean pitch in isolation to perceive speaker gender. However, superimposing a female mean pitch on a male voice only leads to 34% female perception and superimposing a male mean pitch on a female voice only leads to 19% male perception (Hillenbrand and Clark, 2009). This finding indicates that other acoustic features are also involved in gender perception. Another important acoustic feature that distinguishes between genders in production is vowel formant frequency. Vowel formant frequencies are higher in females than in males due to differences in the shapes and sizes of the vocal tract (Hillenbrand et al., 1995). The combination of the first three formants (F1-3) can acoustically distinguish speaker gender with 92% accuracy (Hillenbrand and 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 and Clark, 2009). Hence, neither mean pitch nor formants in isolation has a decisive role in perceived gender. However, the combination of mean pitch and formants is a reliable cue for gender perception. Superimposing female mean pitch and formants on a male voice leads to 82% female perception and superimposing male mean pitch and formants on a female voice likewise leads to 82% male perception, suggesting that mean pitch and formants make up an important part of gender-related voice characteristics (Hillenbrand and Clark, 2009).

Although perceiving gender with 82% accuracy using only mean 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 and Clark, 2009). Other voice features may also have a small contribution to gender perception. For example, phonation type is known to be correlated with gender in production. Females tend to have breathier voices than males (Klatt and Klatt, 1990), whereas males tend to have creakier and tenser voices than females (Tielen, 1992). Some studies also claim that female speakers tend to speak with a larger pitch range (i.e. the difference between the highest and lowest pitch in an utterance) than male speakers (e.g. Takefuta et al., 1972; Simpson, 2009), or that female speech has a more dynamic pitch and more rising pitch contours than male speech (Kreiman and Sidtis, 2011: 133). The role of phonation type and pitch range and dynamics in perceived gender has not yet been investigated intensively.

In summary, female voices are distinguishable from male voices by their increased values for mean pitch and formants. Phonation type and, possibly, pitch range size, also distinguish female from male voices. The combination of mean pitch and formant information plays a substantial role in gender perception.

### 1.2. Gender effects in verbal processing

Idiosyncratic information such as gender is typically considered extra-linguistic information. Many findings show that listeners store extra-linguistic prosodic information such as talker identity, emotional state and speaking rate into long-term memory (e.g. Bradlow et al., 1999; McMurray and Jongman, 2011; Pisoni, 1993). Moreover, past work suggests less effective verbal processing in the presence of extra-linguistic prosodic information. For example, pitch variations weaken the auditory priming effect (Church and Schacter, 1994), talker variability decreases performance in lexical identification tasks (Mullennix et al., 1989), and the expectation of speaker variability slows down verbal processing (Magnuson and Nusbaum, 2007).

Neuroimaging research on the role of voice gender in verbal processing suggests more brain activation in the right hemisphere – more specifically in the auditory cortex – when listeners listen to female voices than when they listen to male voices. Specifically, in an fMRI study, Sokhi et al. (2005) found more activation in the regions of the auditory cortex that were involved in interpreting prosody in male listeners when listening to female voices than when listening to male voices. As the au-

ditary cortex area is also the area which maps human qualities (e.g. gender) to an acoustic voice signal, Sokhi et al.'s finding would seem to suggest that female voices require more processing than male voices. According to Kreiman and Sidtis (2011: 235), when listening to male voices, there is more activation in brain areas involved in “*what is being said*” and when listening to female voices, there is more activation in brain areas that are involved in processing “*how, and by whom, the message is expressed*”. Note that Sokhi et al. (2005) have provided evidence for the hypothesis that listening to a female voice is a more demanding task than listening to a male voice for male listeners, this finding remains to be replicated with female participants. It can thus not yet be ruled out that listening to female voices only leads to increased brain activation of the auditory cortex in male listeners.

Yang et al. (2013) used a directed forgetting task to examine the role of voice gender and emotional prosody in verbal processing. They found that when one group of participants was directed to forget word list 1 and remember word list 2 and another group was directed to remember both word lists, participants in both groups remembered fewer words from list 1 when the lists were spoken in a female voice than when they were spoken in a male voice. Yang et al. argued that the acoustic salience of female voices drew attention to the voice features and thus impeded verbal processing for female voices. Surprisingly, participants remembered more words from list 1 when the lists were spoken in an angry male voice compared to the neutral female voice, in spite of the fact that the angry male voice had a higher mean pitch than the neutral female voice. This finding suggests 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 and Clark, 2009). Yang et al.'s (2013) results thus provide behavioural evidence that female voices require more processing than male voices, but the exact source of the processing difference for male and female voices remains unclear.

Lee and Zhang (2018, 2015) used a repetition task and a semantic/associative priming task to investigate the role of speaker variability in verbal processing. They found that talker variability affected the access of word meaning. However, talker variability was confounded with gender variability. Results showed that the degree of semantic/associative priming was attenuated when a prime was spoken in a male voice and a target was spoken in a female voice, compared to the condition in which both prime and target were spoken in the same female voice; but no attenuation of the priming effect was observed when a prime was spoken in a female voice and the target was spoken in a male voice, compared to the condition in which both prime and target were spoken in the same male voice. This result indicates that the switch from a male to a female voice affect verbal processing, but not vice versa, i.e. female but not male voices impede verbal processing. For mean reaction times on the other hand, they found that female targets received faster responses, i.e. easier processing of target words spoken by the female than the male speaker, which seems to contradict the finding of an attenuated priming effect for female voices only. Lee and Zhang (2015) suggested that the effect of speaker variability they found was indeed confounded with gender and that the effects might be due to the longer durations of the stimuli spoken by the female speaker relative to the male speaker. Longer duration means that listeners had more time to process the stimuli spoken by the female voice, resulting in faster overall reaction times and possible an attenuated semantic/associative priming effect. Lee and Zhang (2018) replicated this study with different speakers of the same gender so that speaker variability was no longer confounded with speaker gender. They found that there was only an effect of speaker variability in a repetition priming task and not in a semantic/associative priming task. The authors therefore concluded that idiosyncratic information seems to be encoded in the phonological form and that “speaker variability is likely to have been resolved before word meaning is accessed” (Lee and Zhang, 2018,75).

In sum, previous research on the role of speaker gender in verbal processing suggests that female voices require more, and thus slower, 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 semantic priming/facilitation may be attenuated for female voices in a lexical decision task with semantic/associative priming. However, the variable listener gender has not been considered, which means that it is possible that impeded verbal processing of female voices only occurs in male listeners. Additionally, impeded verbal processing of female voices has mostly been observed in a context with talker variability, which may be confounded with difference detection. Pu et al. (2005) have shown that it is very difficult to distinguish the priming effect from difference detection. To rule out confounding difference-detection effects and focus on a gender effect instead of speaker variability, voice features of prime-target pairs may better be manipulated between prime-target pairs, instead of within pairs.

## 2. Research questions and hypotheses

The current study has two goals: (1) examine the suggestion in previous findings that female voices may impede verbal processing; and (2) extend previous research by examining the role of two specific voice features, namely mean pitch and formants.

Regarding our first goal, we hypothesise that a female voice impedes verbal processing in a lexical decision task. This is based on previous research showing impeded verbal processing for female voices relative to male voices (Yang et al., 2013; Zhang and Lee, 2011; Lee and Zhang, 2015). Our predictions are that lexical access speed will be slower and that semantic facilitation will be attenuated in female voice conditions. Lexical access speed is reflected in absolute reaction times to target words. Impediment of verbal processing is reflected in attenuated priming/facilitation. The priming/facilitation is computed by subtracting the reaction time to the target word preceded by a related prime (e.g. queen – king) from the reaction time to the same target word preceded by a semantically unrelated prime (e.g. bell – king). Our data should furthermore show faster reaction times of targets that are preceded by related primes in general, showing that semantically related primes facilitate activation of the target word whereas unrelated primes do not (cf. Spreading activation model: Collins and Loftus, 1975). Secondly, we might expect different results from male and female listeners. Namely, it is possible that only male listeners show impeded verbal processing of female voices.

Regarding our second goal, it has been shown that mean pitch is one of the main voice features for gender perception from voice (Hillenbrand and Clark, 2009). We therefore hypothesise that female mean pitch impedes verbal processing when imposed on the male voice and that male mean pitch facilitates verbal processing when imposed on the female voice. Formants, however, have limited power in changing gender perception from voice (Gelfer and Mikos, 2005; Hillenbrand and Clark, 2009; Poon and Ng, 2011). For formants, we therefore hypothesise that

female formants do not impede verbal processing when imposed on the male voice and male formants do not facilitate verbal processing when imposed on the female voice.

## 3. Method

### 3.1. Materials

#### 3.1.1. Experimental stimuli and fillers

The Dutch materials in this study were adapted from an associative priming study (Geuze et al., 2013) and consisted of words taken from the Leuven Association Database (De Deyne and Storms, 2008). Experimental stimuli consist of 64 unique target words, each of which was grouped with a related prime and an unrelated prime into a triplet:

1) draad	naald	roest	(thread – needle – rust)
2) kloen	fiets	boom	(pseudoword – bicycle – tree)

Each target word was presented either together with the related prime or the unrelated prime to the participant as two separate word pairs. In example 1, target word *draad* ‘thread’ could make an experimental pair with related prime *naald* ‘needle’ and with unrelated prime *roest* ‘rust’. In example 2, pseudoword target *kloen* could make a pair with *fiets* ‘bicycle’ and with *boom* ‘tree’ for our filler trials. Related word pairs in the experimental trials have an association strength of at least 0.1, meaning that participants named the target word following the probe in at least 10% of all cases in the first three responses in a continuous association task (De Deyne and Storms, 2008). An equal number of 64 word sets consisting of a pseudoword target and two primes acted as fillers (see example 2). Word pairs with phonological overlap (initial CV or final CVC) were excluded.

The 64 target words (with two subsequent word pairs each) were divided into four lists of 16 target words matched on word length, word frequency, concreteness, age of acquisition, and neighbourhood size, because it has previously been shown that lexical access speed is mediated by these measures (De Deyne and Storms, 2008; Keuleers et al., 2010; Moor and Brysbaert, 2000). Word frequency was based on the logarithmic frequency of words in the SUBTLEX-NL database (Keuleers et al., 2010), which is a database of Dutch word frequencies based on 44 million words from television and film subtitles. 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 Colheart’s N (the number of words that can be produced by changing a phoneme in a word of the same length). Creating these balanced word lists was accomplished with computer programme Match (Van Casteren and Davis, 2007). Independent sample t-tests on the matched measures showed that there were no significant differences between the target stimuli for each voice condition for any of the matched measures according to independent samples t-tests (all  $t(30) < 0.59$ ,  $p > 0.09$ ). The exact matching statistics can be found in Table 1.

**Table 1**

Descriptive statistics for the matching measures for target words per stimuli list. AoA refers to age of acquisition, PLD30 refers to Phoneme Levenshtein Distance, and Colt\_N refers to Colheart’s N.

Matched variable	N	Original voice		Pitch		Formant		Pitch + Formant	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Freq	16	5667	8948	3444	4121	3900	6579	5175	7471
PhonCnt	16	4.25	1.18	4.44	1.36	4.63	1.20	4.19	1.11
SyllCnt	16	1.31	0.48	1.38	0.50	1.38	0.50	1.38	0.50
Concrete	16	4.34	0.55	4.56	0.39	4.63	0.33	4.55	0.55
AoA	16	5.32	1.01	5.48	0.98	5.75	1.24	5.49	1.24
PLD30	16	1.50	0.30	1.60	0.47	1.65	0.45	1.58	0.51
ColtN	16	12.25	9.43	11.63	9.55	9.63	9.58	10.81	9.33

**Table 2**

Acoustic measurements for the male and female speaker and *t* values, degrees of freedom and *p* values from independent samples *t* tests. Significance levels: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Measure	N	Male		Female		<i>t</i>	<i>df</i>	<i>p</i>
		M	SD	M	SD			
dur [ms]	192	0.46	0.08	0.66	0.12	20.10	382	<0.001***
new-dur [ms]	192	0.46	0.08	0.46	0.28	0.43	382	0.67
pitch [Hz]	192	97.77	16.16	205.34	32.19	40.37	382	<0.001***
F1 [Hz]	192	737.72	165.13	794.67	166.38	3.37	382	<0.001***
F2 [Hz]	192	1720.58	247.77	1810.30	251.42	3.52	382	<0.001***
F3 [Hz]	192	2758.41	202.51	2910.74	193.30	7.37	382	<0.001***

Each of the four word-pair lists occurred in each voice gender (male, female). The four lists of matched word pairs were assigned to the four acoustic manipulation types (no manipulation, mean pitch manipulation, formant manipulation, mean pitch + formant manipulation) by means of a Latin Square design. In other words, the same word pairs were used across voice gender, but not across manipulation type. This was to limit the repetitions of each word pair within the experiment. Participants were presented with 512 trials in total (16 target words  $\times$  2 prime types  $\times$  4 manipulation types  $\times$  2 speaker genders), half of which were fillers.

For the presentation order, experimental items and fillers were randomised with computer programme Mix. A pseudorandom order was generated such that neither the same voice condition (original voice, formants, pitch, formants + pitch), nor the same type (related, unrelated, or non-word filler) were repeated more than two times in a row. Additionally, because target words occurred four times across type and voice condition, the minimal distance between identical target words was set at eight trials.

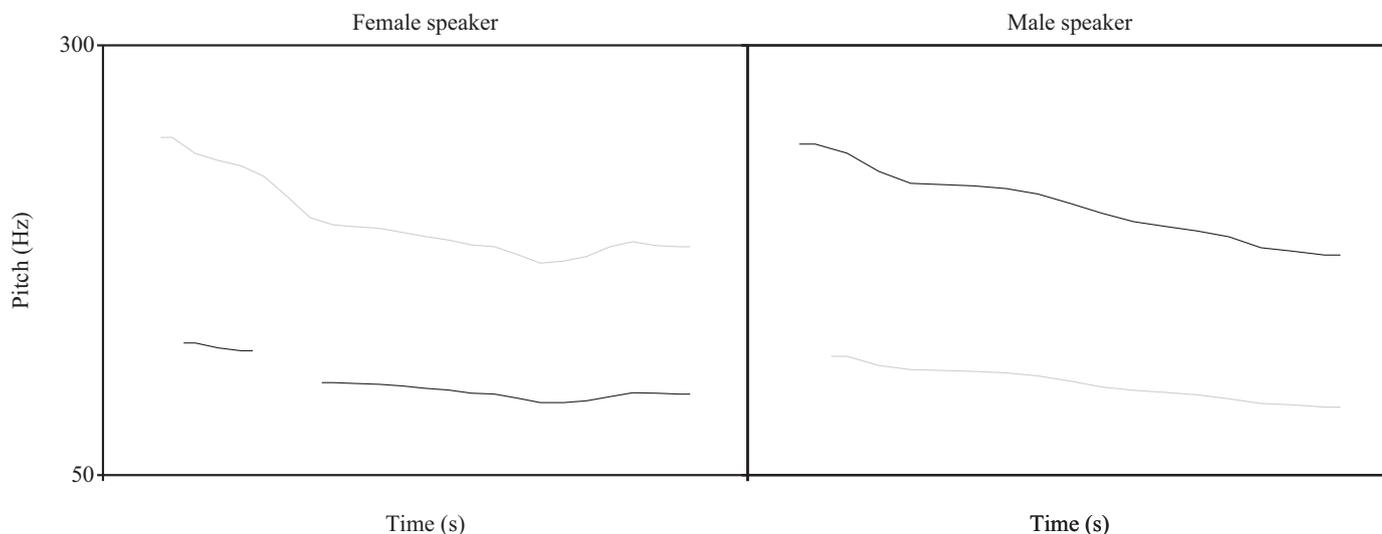
### 3.1.2. Acoustic manipulation of pitch and formants

One male speaker (age = 22) and one female speaker (age = 23) with a Standard Dutch accent who had a typical male and typical female voice respectively were recruited to record the stimuli. They received €5.00 for their contribution to this study. Recordings were made with a Zoom H1 Handy Recorder using a 44,100 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 and Weenink, 2017). All recordings were firstly normalised on amplitude. Secondly,

the recordings were analysed for pitch and formant frequencies (F1–F3) so that averages could be established for both the male and female speaker (see Table 2).

The original word duration (dur) was significantly different between the male and female speaker (see Table 2). Durations of words spoken by the female speaker were compared to the male speaker's pronunciation and adjusted accordingly per item, such that each item had comparable length in the male and female voice. The original and duration-adjusted (new-dur) items were presented to four native speakers of Dutch, who judged whether the original or adjusted duration sounded more natural in a forced-choice decision task. A one sample *t*-test (0 = original sounded more natural 1 = duration adjusted sounded more natural) shows that scores were significantly different from zero ( $t(159) = 23.40$ ,  $p < 0.001$ ). Participants judged the adjusted, sped-up version as more natural sounding in 76.5% of all cases.

Following Hillenbrand and Clark (2009), the female/male ratios for formant values were calculated from the averages in Table 2, 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 shown in Table 2. An example manipulation for the Dutch word "clown" 'clown' with formant and pitch contours can be found in Figs. 1 and 2.



**Fig. 1.** Original-voice pitch contours (grey; female: 188.7 Hz, male: 104.7 Hz) and pitch-shifted contours (black; female: 100.5 Hz, male: 208.1 Hz).

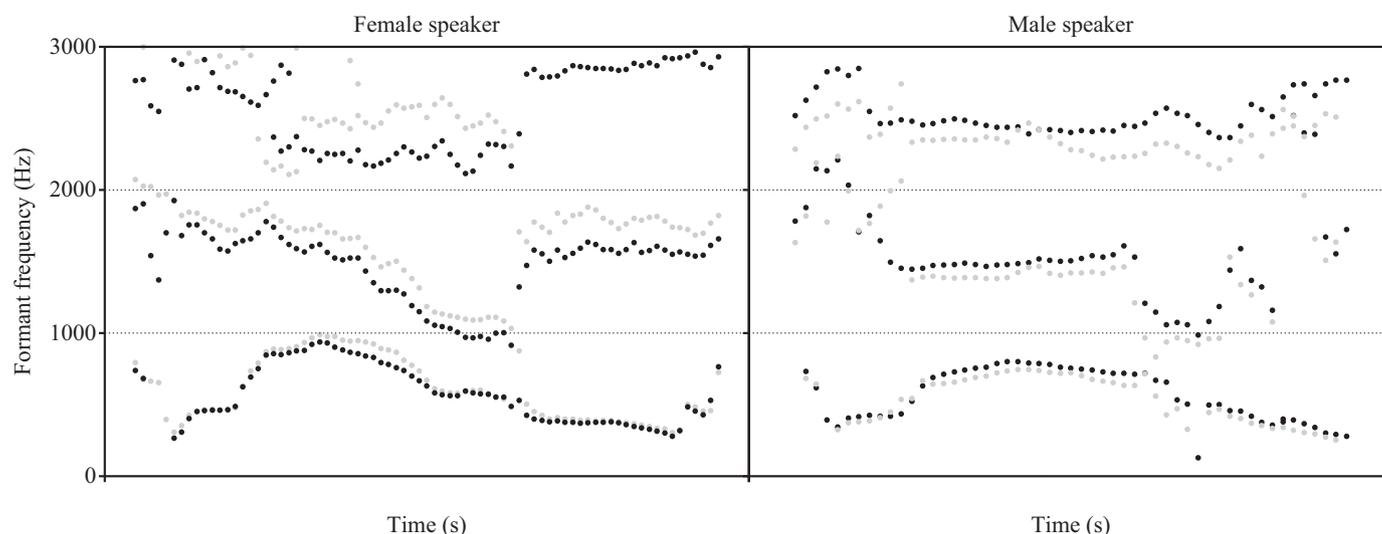


Fig. 2. Original-voice formant contours (grey) and formant-shifted contours (black).

Table 3

Perceived gender and rating certainty scores per source gender and manipulation type.

Manipulation type	Perceived gender				Rating certainty			
	Source female		Source male		Source female		Source male	
	M	SD	M	SD	M	SD	M	SD
None	0.98	(0.14)	0.00	(0.00)	4.66	(0.73)	4.94	(0.25)
Formants	0.99	(0.10)	0.15	(0.36)	4.64	(0.67)	3.24	(1.13)
Pitch	0.56	(0.50)	0.83	(0.38)	2.67	(1.26)	3.31	(1.31)
Pitch+formants	0.46	(0.50)	0.85	(0.36)	2.53	(1.03)	3.40	(1.24)

As the manipulated stimuli may differ in perceived gender, we computed a perceived gender score via a perception 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. They were asked to judge whether the speaker of the experimental target words sounded “male” or “female” and indicate their rating certainty for all experimental voice conditions used in the current study. The perceived gender scores and certainty scores for each condition are shown in Table 3. 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 equal appearing interval scale ranging from ‘very uncertain’ (1) to ‘very certain’ (5). The scores for perceived gender show that, in the voice condition without manipulation, the female speaker was perceived as female and the male speaker was perceived as male. The high certainty scores indicate that listeners were very certain about their gender judgement of the original voices. For the formant manipulation, the perception of the female speaker as female did not change, whereas the male speaker was occasionally perceived as female. For the pitch manipulation, the female speaker was perceived as female about half of the time and the male speaker was mostly perceived as female. When both pitch and formants were manipulated, the change in perceived gender relative to the original voice condition was the largest. These results show that the pitch manipulation and the combined pitch and formant manipulation change the perceived gender. The lower certainty ratings for the voice manipulated conditions indicates that listeners were not very certain about their judgement.

### 3.2. Participants

Forty-three native speakers of Dutch (20 males, 23 females, age:  $M = 25.72$  years,  $SD = 10.56$ ) participated in this study. They were

recruited through the participant database of the Utrecht Institute of Linguistics at Utrecht University. None of the participants reported to have dyslexia or any hearing defects. Four participants reported to speak more than one native language. Prior to participation, the participants were asked to read an information letter and sign a participation approval form. The participants received financial compensation for their participation as per the standards of the Laboratory of the Utrecht Institute of Linguistics where the experiments were conducted. The study was approved by the Ethical Assessment Committee of Linguistics (ETCL) of the Utrecht Institute of Linguistics.

### 3.3. The lexical decision task with auditory priming

The participants were asked to seat themselves in front of a computer screen in a sound attenuated booth located in the laboratory. A button box containing a yes-button and a no-button was placed in front of them. An auditory lexical decision task with auditory priming was run using software programme ZEP (Veenker, 2017). The auditory stimuli were 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 the participants were asked to take a two-minute pause. The participants’ progress was displayed in terms of how many trials out of the total number of trials were completed on the bottom right corner of the computer screen. A visual yes-button and no-button reflecting the button box was also displayed at the end of each auditory stimulus so that no mistakes were made regarding which button on the button box designated a “yes” versus a “no” response. Response accuracy and reaction time were measured from the target onset. The prime-target interval was specified at 250 ms. The inter-trial interval

was specified at 1500 ms and the task was auto-paced. The experiment lasted about 40 min for each participant, including instructions, practice trials, and three two-minute pauses.

#### 4. Statistical analysis

Three types of responses were excluded from further analysis: (1) responses to filler pseudoword targets (50% of all items); (2) 58 missing values (5% of remaining experimental items); (3) 1300 incorrect responses (11.87% of remaining experimental items). The semantic priming effect was calculated by subtracting the reaction time to a target word preceded by a related prime from the reaction time to the identical target word, i.e. the same target word with the same voice source gender and manipulation type, preceded by an unrelated prime. Each priming effect data point, i.e. target word, thus contained two correctness values (one for the unrelated prime word trial and one for the related prime word trial). The data points were excluded when one or both responses were listed as incorrect. This resulted in 4670 data points for absolute reaction time and exactly half that number, i.e. 2335 data points, for the semantic priming effect. 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. As absolute reaction time data displayed right skew, absolute reaction time was log-transformed (base 10).

Linear mixed-effect modelling was used to examine the effects of Trial Type (0 = unrelated 1 = related), Listener Gender (0 = male 1 = female), Perceived Gender (score from 0 to 8 reflecting a scale of male (0) to female (8) voice perception), and Manipulation Type (1 = original voice, 2 = pitch, 3 = formants, 4 = pitch + formants) on both the absolute reaction time and on the semantic priming effect. The predictor variables (i.e. main effects and interactions) were added to the fixed part of the model in a forward, stepwise manner (see Table 4); one additional factor was added at a time and the interaction factors that did not improve a model were removed in the subsequent model. The models' fits were compared by log likelihood estimation. Trial Type was only part of the modelling for absolute reaction time and not the semantic priming effect, as the semantic priming effect was computed as the difference in reaction time between the two types of trials. The random part of the model contained random intercepts for item, i.e. target word, and participant.

#### 5. Results

##### 5.1. Absolute reaction time

Model 1 was a significantly better fit on the data than the null model ( $\chi^2(1) = 455.05, p < 0.001$ ), indicating that there was a significant ef-

**Table 4**

Entry of predictor variables in model-building. Only the added variable is displayed in each model.

Model	Predictor variables
0	(1   Target Word) + (1   Participant)
1	+ Trial type
2	+ Perceived gender
3	+ Trial type : Perceived gender
4	+ Manipulation type
5	+ Trial type : Manipulation type
6	+ Perceived gender : Manipulation type
7	+ Trial type : Perceived gender : Manipulation type
8	+ Listener gender
9	+ Trial type : Listener gender
10	+ Perceived gender : Listener gender
11	+ Manipulation type : Listener gender
12	+ Trial type : Perceived gender : Manipulation type: Listener type

fect of trial type ( $\beta = 0.06, SE = 0.003, t = 21.88$ ). In other words, the reaction times to target words were faster when they were preceded by related primes (log RT = 2.85, SD = 0.12) than when they were preceded by unrelated primes (log RT = 2.92, SD = 0.12). None of the more complex models led to a better fit, indicating that none of the other predictor variables had significant effects on absolute reaction time. Differences in absolute reaction time between the male and female speaker and between voice manipulation types were very small, as shown in Table 5.

##### 5.2. Semantic priming facilitation effect

Model 2 was not a significantly better fit on the data than the null model ( $\chi^2(1) = 0.11, p = 0.75$ ). None of the more complex models led to a better fit. This shows that none of the predictor variables had a significant effect on the semantic priming facilitation effect.

To check the absence of effects for the semantic priming facilitation effect, this analysis was repeated on a subset of the data that showed a positive semantic facilitation effect ( $N = 1623$ ), i.e. faster reaction time to related versus unrelated word pair. Again, Model 1 was not a significantly better fit on the data than the null model ( $\chi^2(1) = 0.79, p = 0.37$ ). Neither did subsequent inclusion of predictor variables lead to better-fitting models to the data.

#### 6. Discussion

Previous research has suggested that female voices are processed more slowly. In the present study, we tried to eliminate possible difference-detection effects in trials, which is hard to distinguish from the priming effect, by manipulating experimental conditions between prime-target pairs, instead of within prime-target pairs. In other words, instead of presenting the prime in a male voice and the target in a female voice, which biases the listener to expect unrelated content, we

**Table 5**

Mean reaction time and prime effect (in ms) per source gender and manipulation type.

Source gender	Manipulation	Related		Unrelated		Prime effect	
		M	SD	M	SD	M	SD
Male	Original	748	(224)	847	(270)	99	(318)
	Pitch	752	(242)	879	(293)	127	(308)
	Formant	736	(228)	857	(264)	121	(300)
	Pitch + formant	759	(239)	861	(275)	101	(291)
	<b>Total</b>	<b>748</b>	<b>(233)</b>	<b>860</b>	<b>(275)</b>	<b>112</b>	<b>(305)</b>
Female	Original	725	(237)	824	(249)	99	(303)
	Pitch	745	(212)	863	(285)	118	(302)
	Formant	743	(231)	859	(284)	116	(323)
	Pitch + formant	724	(241)	831	(297)	106	(322)
	<b>Total</b>	<b>734</b>	<b>(231)</b>	<b>844</b>	<b>(280)</b>	<b>110</b>	<b>(313)</b>

presented both prime and target in the same voice. Unexpectedly, excluding effects of talker variability, we found no effects of voice gender. This is contrary to earlier findings from Lee and Zhang (2015), who reported a reduction of semantic/associative priming when target words were spoken in a female voice. Lee and Zhang (2015) suggested that this effect might have been due to the longer durations of stimuli spoken by the female speaker relative to those spoken by the male speaker. We manipulated the durations of the female stimuli to match the male stimuli and found no reduction of priming. It is therefore likely that the gender effect in Lee and Zhang's study (2015) was indeed a result of the longer duration of targets spoken by the female speaker. Given that speaker variability typically includes variations in duration, especially when both male and female speakers are concerned, it is recommendable to control stimuli duration in this type of research using time-sensitive tasks.

Extending previous literature, we included a predictor variable for listener gender. This was important because it was possible that our data would show that there is not a female voice effect, but rather an opposite gender effect, meaning that only males would show slower lexical decision speed for female voices. This hypothesis was based on neuroimaging research by Sokhi et al. (2005), who found that male participants listening to male voices showed more 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), whereas the same male participants listening to female voices showed more brain activation in the auditory cortex. However, we found no statistically significant evidence for an effect of listener gender but evidence for a semantic facilitation priming effect. That is to say, the participants had shorter reaction times to target words that were preceded by related primes than to target words that were preceded by unrelated primes regardless of experimental conditions and listener gender.

It should be noted that there seemed to be an asymmetry in the effect of the voice manipulations for the male and female speakers. Namely, the manipulations of the male voice had a larger effect on perceived gender than manipulations of the female voice. Asymmetry in perceived gender has been observed before, for example by Owren et al. (2007), who explained this asymmetry as follows: “[while] the presence of critical features of ‘maleness’ virtually guarantees that the talker is an adult male [...], their absence does not unequivocally imply that the talker is an adult female” (931). This would imply that superimposing male pitch and formants on the female voice would have a larger effect on perceived gender than superimposing female pitch and formants on the male voice. However, in sentence manipulations, both the current results and results from Hillenbrand and Clark (2009) showed that upward shifts in mean pitch and formants had a slightly larger effect on perceived gender than downward shifts. This suggests that Owren et al.'s (2007) account might not generalise to voice manipulations, or rather, to voice features such as mean pitch and formants in isolation. In the current study, this means that more tokens were perceived to be spoken by a female speaker than a male speaker. In that sense, our data might not be completely balanced. However, in our statistical analysis, we included perceived gender as a continuous fixed factor, which indicated the perceived ‘femaleness’ on a scale from 0 to 8. We thus do not expect that the asymmetry observed here affects the current findings.

Furthermore, we only used one male speaker and one female speaker to create our stimuli. Although these speakers of Dutch were typical of their genders in speech production, the difference in their formant values was smaller (around 5% on average for F1–F3 measured over full words) than one might expect on the basis of the literature on isolated vowels. For example, for speakers of American English, a ratio of approximately 15% (e.g. Hillenbrand and Clark, 2009; Huber et al., 1999) is common. It may thus be useful to see whether the present findings generalise to multiple speakers and to more extreme pitch and formant manipulations.

To conclude, the current study has yielded no evidence that words spoken by a female voice are processed more slowly than words spoken by a male voice as measured by absolute reaction times and by the semantic priming effect. Additionally, there is no evidence that female pitch or formants slow the processing of words.

To expand our understanding of the role of speaker gender in verbal processing mechanisms, we suggest that future research focus on neuroimaging techniques. These techniques might sometimes reveal qualitative differences in processing that behavioural experiments do not reveal. Even though the present behavioural study has yielded no evidence for impeded verbal processing in female voice features, neuroimaging techniques may still show that the presence of female voice features activate distinct regions in the brain. Alternatively, female voice features might activate distinct brain regions in male listeners only. The first evidence for this prediction has been reported by Sokhi et al. (2005). Replicating 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 gender and listener gender.

#### Author statement

Laura Smorenburg: Conceptualization, Methodology, Investigation, Resources, Data Curation, Formal Analysis, Writing – Original Draft, Visualization.

Aoju Chen: Formal Analysis, Supervision, Writing – Review and Editing.

#### Author's note

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

#### References

- Boersma, P. and 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. *Percept. Psychophys.* 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. *J. Exp. Psychol.* 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. *Psychol. Rev.* 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. *Behav. Res. Methods* 40 (1), 198–205. <https://doi.org/10.3758/BRM.40.1.198>.
- Fant, G., 1970. Acoustic Theory of Speech Production, With Calculations Based On X-Ray Studies of Russian Articulations. De Gruyter Mouton, The Hague <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. *J. 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. *J. Psycholinguist. Res.* 3 (3), 225–231. <https://doi.org/10.1007/BF01069239>.
- Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K., 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97 (5), 3099–3111. <https://doi.org/10.1121/1.411872>.
- Hillenbrand, J.M., Clark, M.J., 2009. The role of f0 and formant frequencies in distinguishing the voices of men and women. *Percept. Psychophys.* 71 (5), 1150–1166. <https://doi.org/10.3758/APP.71.5.1150>.
- Huber, J.E., Stathopoulos, E.T., Curione, G.M., Ash, T.A., Johnson, K., 1999. Formants of children, women, and men: the effects of vocal intensity variation. *J. Acoust. Soc. Am.* 106 (3), 1532–1542.
- Kahane, J.C., 1978. A morphological study of the human prepubertal and pubertal larynx. *Am. J. Anat.* 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. *Behav. Res. Methods* 42 (3), 643–650. <https://doi.org/10.3758/BRM.42.3.643>.

- Klatt, D.H., 1987. Review of text-to-speech conversion for English. *J. Acoust. Soc. Am.* 82 (3), 737–793.
- Klatt, D., Klatt, L., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87 (2), 820–857. <https://doi.org/10.1121/1.398894>.
- Kreiman, J., Sidtis, D., 2011. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. John Wiley & Sons Ltd, Oxford <https://doi.org/10.1002/9781444395068>.
- Lee, C.Y., Zhang, Y., 2015. Processing speaker variability in repetition and semantic/associative priming. *J. Psycholinguist Res.* 44 (3), 237–250. <https://doi.org/10.1007/s10936-014-9307-5>.
- Lee, C.Y., Zhang, Y., 2018. Processing lexical and speaker information in repetition and semantic/associative priming. *J. Psycholinguist Res.* 47 (1), 65–78. <https://doi.org/10.1007/s10936-017-9514-y>.
- Luce, R.D., 1986. *Response Time: Their Role in Inferring Elementary Mental Organization*. Oxford University Press, New York.
- Magnuson, J.S., Nusbaum, H.C., 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol.* 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. *Psychol. Rev.* 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. *Psychol. Res.* 63 (2), 159–162. <https://doi.org/10.1007/PL00008174>.
- Mullennix, J.W., Pisoni, D.B., Martin, C.S., 1989. Some effects of talker variability on spoken word recognition. *J. Acoust. Soc. Am.* 85 (1), 365–378.
- Noyes, J.M., Frankish, C.R., 1989. A review of speech recognition applications in the office. *Behav. Inf. Technol.* 8 (6), 475–486.
- Owren, M.J., Berkowitz, M., Bachorowski, J.A., 2007. Listeners judge talker sex more efficiently from male than from female vowels. *Percept. Psychophys.* 69 (6), 930–941.
- Pegoraro-Krook, M.I., 1988. Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia Phoniatica 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 Commun.* 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*, pp. 1630–1633.
- 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>.
- Simpson, A.P., 2009. Phonetic differences between male and female speech. *Lang Linguist 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.G., Brunt, M., 1972. A statistical analysis of melody curves in the intonation of American English. In: *Proceedings of the 7th International Congress of Phonetic Sciences*, Montreal, Canada, pp. 1035–1039.
- Tielen, M.T.J., 1992. *Male and Female Speech: An Experimental Study of Sex-related Voice and Pronunciation Characteristics*. UvA, Amsterdam.
- Van Casteren, M., Davis, M.H., 2007. Match: a program to assist in matching the conditions of factorial experiments. *Behav. Res. Methods* 39 (4), 973–978. <https://doi.org/10.3758/BF03192992>.
- 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. *Psychol. Rec.* 58 (3), 475. <https://doi.org/10.1007/BF03395630>.
- 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. *J. Acoust. Soc. Am.* 129 (4), 2662–2662. <http://dx.doi.org/10.1121/1.3588905>.