An Exploration of the Accentuation Effect: Errors in Memory for Voice Fundamental Frequency (F0) and Speech Rate

Georgina Gous

Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom

50 Shakespeare Street, Nottingham, NG1 4FQ

georgina.gous2013@my.ntu.ac.uk

Andrew Dunn

Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom

50 Shakespeare Street, Nottingham, NG1 4FQ

andrew.dunn@ntu.ac.uk

Thom Baguely

Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom

50 Shakespeare Street, Nottingham, NG1 4FQ

Thomas.baguley@ntu.ac.uk

Paula Stacey

Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom

50 Shakespeare Street, Nottingham, NG1 4FQ

paula.stacey@ntu.ac.uk

An Exploration of the Accentuation Effect: Errors in Memory for Voice Fundamental Frequency (F0) and Speech Rate

Abstract

The accentuation effect demonstrates how memory often reflects category typical representations rather than the specific features of learned items. The present study investigated the impact of manipulating fundamental frequency (F0) and speech rate (syllables per second) on immediate target matching performance (selecting a voice from a pair to match a previously heard target voice) for a range of synthesised voices. It was predicted that when participants were presented with high or low frequency target voices, voices even higher or lower in frequency would be selected. The same pattern was also predicted for speech rate. Inconsistent with the accentuation account, the results showed a general bias to select voices higher in frequency for high, moderate, and low frequency target voices. For speech rate, listeners selected voices faster in rate for slow rate target voices. Overall it seems doubtful that listeners rely solely on categorical information about voices during recognition.

Keywords: Recognition Memory, Fundamental Frequency (F0), Speech Rate, Accentuation Effect, Voice Disguise

Introduction

Human cognitive processing resources are limited and this presents a challenge in a rapidly changing social environment. Given these limitations, people devise short-cut strategies to simplify the nature of incoming information. One proposed strategy is categorisation in which it is assumed that stimuli are reduced into cognitively simple categories which contain other stimuli that are equivalent/analogous to each other (e.g. same colour, same shape, same tone) and different from other stimuli (Brosch, Pourtois, & Sander, 2010). This process of categorisation means that it becomes less cognitively effortful when an observer encounters a new stimulus. However, the act of placing stimuli into distinct categories can lead to distortions which result in the stereotyping of some distinctive features (Hogg & Vaughan, 2010). For example, when stimuli covary by constant amounts on a given continuum, people are less likely to perceive stimuli within the same category to be different than when stimuli are placed in different categories. In other words, people minimise the perception of differences within a category and maximise the perception of differences across categories. Consequently, when people are asked to recall properties of stimuli within a category, they tend to recall features typical of the category overall, rather than the individual properties of the stimulus. This is known as the *accentuation effect* (Fiske, Gilbert, & Lindzey, 2010; Huart, Corneille, & Becquart, 2005; Sutton & Douglas, 2013).

Accentuation effects have been found to be real and robust and have been observed with both non-social (e.g. Krueger & Clement, 1994; Tajfel & Wilkes, 1963) and social stimuli (e.g. Eiser, 1971; McGarty & Penny, 1988; McGarty & Turner, 1992; Krueger & Rothbart, 1990; Queller, Schell, & Mason, 2006; Haslam & Turner, 1992). Recent work has shown how accentuation effects can also affect perceptions of facial stimuli. For example, adding a featural characteristic of a particular race (such as a Hispanic or African American hairstyle) to a facial composite leads participants to judge faces as more typical of that racial origin compared to

when no modification or labels were used (MacLin & Malpass, 2001). Similar results have been observed in other studies where faces have been given a more white European name (Hilliar & Kemp, 2008), or when the faces have been labelled as 'black' (Levi & Banaji, 2006). Others have shown that categorising faces can lead to errors in memory at the recognition stage. For example, Corneille, Huart, Becquart, and Brédart (2004) examined the impact of categorisation on the recollection of ethnically ambiguous faces. Participants were presented with faces lying at various locations on mixed-race continua (Caucasian-North African and Caucasian-Asian faces were used as images in the morphing program). Recollections of faces towards the middle of continua were distorted; participants reported them to contain more ethnic features typical of the category they were closest towards than they actually contained. Comparable effects have also been found when using gender ambiguous faces (Huart, Corneille, & Becquart, 2005), and ambiguous angry and happy faces (Halberstadat & Niedenthal, 2001).

Surprisingly, very few researchers have considered categorisation or accentuation effects in relation to voices. This is remarkable because variations in the paralinguistic characteristics of the voice (the way it is being said; e.g. rate of utterance, frequency of utterance, loudness etc.) can occur within the same speaker (also known as within-speaker, or intra-speaker variation). Speakers rarely pronounce given words or phrases in an identical way on different occasions, even if the second utterance is produced in close succession (Hollien, 1990). The same speaker can sound different from time-to-time because of factors such as time of day, fatigue, intoxication (from alcohol or drugs), thought distractions, situational demands, mood state, changes in health and physical status, stress, and a speaker's emotional state (Nolan, 2005; Saslove & Yarmey, 1980). Speakers can also modify their own voice by means of disguise. Research has stressed how such changes can introduce great acoustic variation and increase errors in memory for the voice (Endres, Bambach, & Flosser, 1971; Reich, Moll, & Curtis, 1976; Reich & Duke, Zhang, 2012). At this point, it is important to distinguish between,

and include a working definition of, the terms ‘voice’, ‘speech’, and ‘speaker’. ‘Voice’ refers to the sound produced by a person’s vocal equipment, and is uttered through the mouth as speech (Traunmuller, 2000). ‘Speech’ refers to the vocalised form of human communication that conveys information between a speaker and a listener. There are two types of features of the speech signal; spectral features (i.e. frequency based features, including F0, intonation, and prosody) and temporal features (i.e. time domain features, including speech rate and amplitude). ‘Speaker’ refers to a person who produces a speech sample.

Mullenix, Stern, Grounds and Tessmer (2010), in one of the few studies to explore this topic in voices, found evidence for accentuation effects for voice memory. The researchers investigated the effects of manipulating fundamental frequency (F0) and speech rate (using words per minute) on recognition memory for voices. To do this, Mullenix et al. (2010) created a number of versions of a male synthesised target voice; a version that was higher than the original voice and fell within the higher F0 speaking range (which they labelled ‘high F0’), a version that was lower than the original voice and fell within the lower F0 speaking range (labelled ‘low F0’), and the original version of the voice which fell in the moderate F0 speaking range (labelled ‘moderate F0’). Similar manipulations were also applied for the speech rate condition to obtain target voices that were faster in rate (labelled ‘fast rate’), slower in rate (labelled ‘slow rate’), and the original version (labelled ‘moderate rate’). This resulted in six conditions of interest (i.e. high, moderate, and low F0, and fast, moderate, and slow speech rate). Using a two-alternative forced choice (2AFC) voice recognition task, participants were presented with one of the target voices and were then asked to recognise this from a pair of sequentially presented voices. The paired voices included the previously heard target voice and a distractor voice which consisted of a modulated version of the target (which was either higher or lower in F0, or faster or slower in speech rate). The results showed a fairly predictable pattern of memory errors. Listeners selected voices lower in F0 than the low F0 target voice, and voices

higher in F0 than the high F0 target voice. However, there was no difference in the selection of higher or lower F0 distractor voices for moderate F0 target voices. In contrast, for speech rate, listeners selected voices slower in rate than the slow rate target voice. However, there was no difference in the selection of faster and slower rate distractor voices for moderate and fast rate target voices.

According to Mullenix et al. (2010), the effect of increased recognition error in the low and high F0 conditions likely reflects an accentuation effect. They argue that listeners place the higher and lower F0 voices they hear into cognitively simple categories, leading them to recall features most salient to that category (i.e. a higher or lower F0) rather than the individual properties the voices have (i.e. actual F0). A similar pattern of findings has also been found for F0 using both a male and female synthesised voice, where listeners selected voices lower in F0 than the low F0 target voice, and voices higher in F0 than the high F0 target voice (Stern, Corneille, Huart, & Mullenix, 2004). The absence of an effect for speech rate is not unexpected since within-speaker variation in speech rate can be highly variable; sometimes people speak quickly, while other times they speak slowly. Whilst variations in F0 also exist, under normal circumstances F0 is likely to be relatively stable (Mullenix et al., 2010; Stern et al., 2004). Thus, it is likely that listeners are more familiar with experiencing speech rate variability and are hence more robust to variation. As a consequence, different properties of the voice may be more or less susceptible to category-based memory distortions (Mullenix et al., 2010).

The present study examined the impact of variations in F0 and speech rate for a set of unfamiliar synthesised voices in a similar manner to Mullenix et al. (2010), but with a number of important extensions and modifications to the procedure. First, we used a slightly larger set of synthesised voices (two male, two female), which increases the generalisability of the findings. Second, we kept the target and distractor voices within a F0 and speech rate range

that is typical in the population for English speakers. This is important given that it is highly unusual to hear voices outside of the typical male and female range in everyday situations. Third, we also included sex of voice and listener sex as independent variables in our design. This is important given that research has emphasised sex differences in verbal episodic memory tasks, with women often performing at a higher level than men (Herlitz, Nilsson, & Backman,1997; Lewin, Wolgers, & Herlitz, 2001; McGivern, Huston, Byrd, King, Siegle, & Reilly, 1997). Others have also reported an own-gender bias (i.e. better recognition performance for voices of an observers own sex) for unfamiliar voices (Roebuck & Wilding, 1993).

Following Mullenix et al. (2010), we investigated the impact of manipulating overall mean F0 (in Hz) and speech rate (in syllables per second[1]) on immediate target matching performance. We used a 2AFC procedure in which listeners were asked to recognise a target voice from a voice pair that contained the previously heard target voice and a modulated version of the voice. There were six conditions of interest (i.e. high, moderate, and low F0, and fast, moderate, and slow speech rate). In keeping with the terminology used by Mullenix et al. (2010), three versions for each target voice were created for both the F0 and speech rate conditions. For the F0 condition, we created a version that was higher than the original voice and fell within the higher F0 speaking range (labelled 'high F0'), a version that was lower than the original voice and fell within the lower F0 speaking range (labelled 'low F0'), and the original version of the voice which fell in the moderate F0 speaking range (labelled 'moderate F0'). Similarly, for the speech rate condition, we created a version that was faster than the

[1] Calculations using syllables rather than words are often considered as being a more accurate and reliable estimate of the rate of speech (Dlugen, 2012). This is because calculations using words are dependent upon the length of the words spoken in the spoken sentence, and not all words in the English language are equal. It was therefore decided upon to use syllables per second (syll/sec) for all calculations of speech rate.

original voice (labelled ‘fast rate’), a version that was slower than the original voice (labelled ‘slow rate’), and the original version of the voice (labelled ‘moderate rate’). To obtain our distractor voices, we further increased and decreased each target voice in F0 and speech rate.

It was expected that the results would parallel those of Mullenix et al. (2010). For F0, we predicted that there would be a memory bias for high and low F0 target voices but not for moderate F0 target voices. Specifically, we expected to see an increase in the selection of voices higher in F0 when high F0 target voices were presented, and an increase in the selection of voices lower in F0 when low F0 target voices were presented. We were more tentative with our predictions for speech rate since Mullenix et al. (2010) found no memory biases for their speech rate manipulations, but in line with the accentuation effect, we hypothesised that people would be more likely to select distractors that were faster in rate for voices that had a fast speech rate, and to select distractors slower in rate for voices that had a slow speech rate.

Method

Design

The participants were arbitrarily allocated to either the F0 condition or the speech rate condition. For each condition, the experiment employed a 2 x 2 x 3 x 3 x 2 mixed factorial design. The between-subjects factor was listener sex (male or female). The within-subjects factors were sex of voice (male or female), target type (*for F0:* high, moderate or low, *for speech rate:* fast, moderate or slow), magnitude of distractor change (*for F0:* 5%, 7%, or 10%, *for speech rate:* 10%, 12%, or 20%) and direction of manipulation (*for F0:* increase or decrease in F0, *for speech rate:* increase or decrease in rate). The dependent variable measured was mean percentage of errors made (i.e. percentage of time listeners choose the distractor voice instead of the target voice).

Participants

A total of 60 undergraduate students (30 males; 30 females) were recruited from Nottingham Trent University and they received course credit for their participation. The inclusion criteria for the study required individuals to be between 18-30 years of age, have no known hearing deficits, have English as their first language, and not undergone any musical training.

A total of 30 individuals contributed to the F0 condition (15 males; 15 females). The ages of the participants ranged from 18 to 27 years old (M = 21.03 years, SD = 2.09 years). A further 30 individuals contributed to the speech rate condition (15 males; 15 females). The ages of the participants ranged from 18 to 30 years old (M = 21.72 years, SD = 2.62 years).

Stimuli and materials

Natural Reader 12.0 (http://www.naturalreaders.com/index.html) was used to create the four voice samples (four different identities, two male and two female). Natural Reader is a text-to-speech software with realistic and natural sounding synthesised voices, generating speech samples from concatenated pieces of real human speech. Synthetic speech was used because of the need for precisely controlled stimuli that varied in F0 and speech rate. Concatenated speech also gives the advantage of sounding more natural than fully synthesised speech. The target speech samples were created by typing the following phrase *"Spring is the season where flowers appear, summer is the warmest season of the year."*, in Natural Reader. The four original voice samples were then manipulated in F0 and speech rate using Audacity® software (http://www.audacityteam.org/) [2]. Audacity® is a free audio software that can be used

[2] It should be noted that formant values changed freely as a result of manipulations in F0. Changes in formants would occur naturally in real voices when changes in F0 are made. This helped to retain the naturalness of the voices used by limiting any irregularities that might by introduced in the voices via the use of synthesised speech (refer to Appendix A for further details about the formant values of the voices).

to edit sounds and was chosen to manipulate the voices because it allowed us to alter one characteristic (F0 and speech rate) whilst holding the other constant.

In order to select stimuli for the main experiment, we pilot tested a number of speech samples for both the F0 and speech rate conditions. This enabled us to create additional voice samples that were both higher and lower in F0, and faster and slower in speech rate. Using a perceptual discrimination paradigm, 72 participants (36 males and 36 females) were given a 2AFC (same/different key press) voice matching task. Participants responded by indicating whether the two stimuli on each trial were the 'same' or 'different'. The stimuli were presented as within voice pairs with a 1 second inter-stimulus separating them. The original target voice was used as the 'standard' stimulus and presented on all trials. The standard stimulus was paired with either itself or a modulated version (increased/decreased in F0, or increased/decreased in speech rate) and presented in a random order. For F0 the modulated versions were increased and decreased by 5% and 10%, and for speech rate they were increased and decreased by 5%, 10%, 15%, and 20%. This was considered appropriate given that a modification in F0 elicited a greater audible change than it did for speech rate. This resulted in a total of 104 trials (13 trials for each voice, with each trial being counterbalanced and presented twice). For F0, a setting of plus and minus 6.63% was judged as 50% discriminable, and for speech rate this was 11.52%. Smaller manipulations in F0 and speech rate were judged as sounding more similar to the target voice, whereas larger manipulations were judged as sounding less similar to the target voice. The pilot testing also allowed us to determine whether the distractor voices chosen for the main experiment were discriminable from the target voice.

For each of the original synthesised voices, we used the 10% modulated versions to obtain target voices in the higher and lower F0 range, and the 20% modulated versions to obtain target voices in the faster and slower speech rate range. For F0, the typical adult male will have an F0 between 80-180 Hz, and for an adult female this will be between 165-255 Hz (Titze,

1994). For speech rate, the typical range for male and female speech is 3.3 to 5.9 syllables/sec (Arnfield, Roach, Setter, Greasley, & Horton, 1995). It is important to emphasise however that different speaking styles typically entail different speaking rates and therefore absolute values can differ (Brown, 2014). All four original (unedited) voice samples fell within the moderate speaking range for both F0 and speech rate, and thus acted as moderate target voices. This resulted in six experimental conditions of interest; low F0, moderate F0 and high F0, and slow speech rate, moderate speech rate, and fast speech rate.

Based on the findings from the pilot study, we increased and decreased each target voice by a further 5%, 7%, and 10% for F0, and by a further 10%, 12%, and 20% for speech rate, to obtain our distractor speech samples. This resulted in a total of six modulated versions (i.e. distractor voices) for each target voice sample; three increased in F0 or speech rate, and three decreased in F0 or speech rate (refer to Appendix B; Table B1 for F0 values, and Table B2 for speech rate values). All of the voices samples, whether targets or distractors, fell within the typical F0 and speech rate range for normally voiced speech for English speakers. The distractor voices spoke the same phrase as the target voices.

The voice samples were tested to determine how naturalistic (i.e. how realistic, or lifelike) they sounded (refer to Appendix C for further details). This is important because the authors wanted to ensure that the voices used were generalizable to those voices that are heard in a real-world environment. Mean naturalness ratings across all of the voices were 73.46% for F0 manipulations and 72.6% for speech rate manipulations. Whilst we recognise that these are not perfect, these values are nevertheless slightly higher than those identified elsewhere (e.g. 70%) (see Jreige, Patel, & Bunnell, 2009), and are a reasonable indication that the synthesised voices used for experimentation are representative of real voices. It should also be noted that the voice samples contained smooth formant transitions and there were no intonational irregularities or prosodic mismatches across words.

The voice samples were also tested to determine whether the four voices used for experimentation were perceived as being different speakers (i.e. different identities). This was important because the authors wanted to ensure that all of the voices used for experimentation were distinct from each and that they would not be confused with another voice that they had previously heard (refer to Appendix D for further details). The results showed that listeners could correctly determine that the voices were different speakers with almost 100% accuracy. Thus, it can be assumed that the voices used for experimentation were distinct from each other and perceived as being different speakers.

All of the speech samples were saved as separate .wav files and presented binaurally using Sony dynamic stereo headphones (Model No. MDR-V150). The experiment was run on a Sony Vaio laptop computer (Model No. SVF153B1YM) using PsychoPy version 1.7701 (Peirce, 2007) to control the presentation and collect participant responses.

Procedure

The participants were arbitrarily allocated to either the F0 or speech rate condition. For each experimental condition, there were 144 trials in total. Specifically, there were four different voices (two male and two female), each with three target voices (high, moderate, and low F0, or fast, moderate, and slow speech rate). For each target voice, there were 12 trials in total (each of the three target voices were paired with one of the six distractor voices, with each trial being presented twice). In each trial, participants were first presented with one of the target speech samples. After a one second gap, the participants were presented with sequentially paired voices that included the target voice (present in all trials) and one of the six distractor voices (that was either increased or decreased in F0 or speech rate). There was a one second inter-stimulus interval between presentation of each voice. The trials were counterbalanced so that half the time the target voice was presented first, and half the time the target voice was

presented second. The order of the trials were randomised across participants using PsychoPy. Following presentation of each trial the participants were asked 'which voice matched the voice you previously heard, voice one or voice two?'. The participants had to indicate whether the 1st or the 2nd voice in the voice pair matched the target voice by pressing '1' or '2' on the numerical keypad. The voices were presented at the same loudness for all participants. This was at level that was typical of a conversation you would hear in everyday life. Upon completion of the experiment, participants were fully debriefed.

Analyses

The results were analysed using mixed-group Analysis of Variance (ANOVA), one for the F0 manipulations and one for the speech rate manipulations. Owing to the high number of main effects and possible interactions, it was necessary to adjust the *p*-values from the main analysis to account for the familywise error rate. A Hochberg correction was therefore applied to the results of the main ANOVA (Hochberg, 1988). In addition, a Hochberg correction was applied to the simple main effects, which were conducted using pairwise *t*-tests. Furthermore, and for reasons of clarity, we present here only the significant findings of these analyses or where non-significant findings are directly relevant. Full ANOVA tables displaying the degrees of freedom (*df*), F ratios (*F*), effect sizes (generalised eta squared; η_g^2), and adjusted *p* values using the Hochberg correction (*p*) for all the main study variables and associated interactions are provided in Appendix E (*for F0*) and Appendix F (*for speech rate*).

Results

Fundamental Frequency (F0)

Table 1 presents the mean percentage of errors made for each distractor type, listed separately for the three target conditions (high, moderate and low F0), the sex of the target voice, and listener sex.

Table 1.

Mean percentage of errors made by distractor type (magnitude and direction of distractor change), target F0 (high, moderate (mod), or low), sex of target voice (collapsed across male and female target voices), and sex of listener (male or female).

	Male Listener						Female Listener					
	Male Target Voice			Female Target Voice			Male Target Voice			Female Target Voice		
	High	Mod	Low	High	Mod	Low	High	Mod	Low	High	Mod	Low
Distractor												
+10%	**6.67**	**6.67**	**16.67**	**3.33**	**6.67**	**6.67**	*8.33*	**13.33**	**21.67**	**6.67**	**15.00**	**13.33**
	14.84	*11.44*	*22.49*	*8.80*	*14.84*	*11.44*	*15.43*	*20.85*	*20.85*	*14.84*	*18.42*	*12.91*
+7%	**15.00**	**21.67**	**40.00**	**11.67**	**13.33**	**23.33**	**20.00**	**26.67**	**38.33**	**10.00**	**25.00**	**23.33**
	18.42	*20.85*	*28.03*	*18.58*	*12.91*	*22.09*	*23.53*	*22.09*	*24.76*	*18.42*	*29.88*	*14.84*
+5%	**20.00**	**25.00**	**38.33**	**18.33**	**23.33**	**38.33**	**55.00**	**30.00**	**30.00**	**30.00**	**21.67**	**40.00**
	21.55	*25.99*	*20.85*	*19.97*	*17.59*	*26.50*	*28.66*	*28.66*	*19.37*	*21.55*	*16.00*	*18.42*
-5%	**8.33**	**21.67**	**13.33**	**23.33**	**13.33**	**5.00**	**10.00**	**25.00**	**10.00**	**15.00**	**18.33**	**10.00**
	15.43	*18.58*	*18.58*	*22.09*	*16.00*	*10.35*	*15.81*	*23.15*	*15.81*	*12.68*	*17.59*	*18.42*
-7%	**1.67**	**10.00**	**13.33**	**3.33**	**11.67**	**8.33**	**10.00**	**18.33**	**10.00**	**6.67**	**11.67**	**10.00**
	6.46	*18.42*	*12.91*	*8.80*	*16.00*	*15.43*	*18.42*	*17.59*	*22.76*	*11.44*	*22.89*	**15.81**

-10%	**6.67**	**13.33**	**8.33**	**6.67**	**5.00**	**50.00**	**6.67**	**6.67**	**6.67**	**6.67**	**11.67**	**6.67**
	14.84	*18.58*	*18.09*	*14.84*	*10.35*	*14.02*	*11.44*	*14.84*	*14.84*	*14.84*	*16.00*	*11.44*

Note: Means are shown in bold. Standard deviations (SD) are shown in italics.

The mean matching error scores for each listener were entered in a mixed ANOVA for the between subjects factor of listener sex (male or female) and the within subjects factors of sex of voice (male or female), target F0 (high, moderate or low), magnitude of distractor change (5%, 7%, or 10%) and direction of manipulation (increase or decrease in F0). This revealed a significant main effect of direction of manipulation, $F(1, 28) = 94.56$, $p < .03$, $\eta_g^2 = .07$, with significantly more errors being made when distractor voices were higher in F0 ($M = 21.20$, $SD = 7.38$) than when they were lower in F0 ($M = 10.51$, $SD = 5.01$) [3]. There was also a significant main effect of magnitude of distractor change, $F(2, 56) = 50.75$, $p < .03$, $\eta_g^2 = .13$. Significantly more errors were made when distractor voices were manipulated by 5% ($M = 22.64$, $SD = 7.63$) compared to when they were manipulated by 7% ($M = 15.97$, $SD = 7.31$), $t(29) = 4.74$, $p < .001$, $d = 0.37$, and 10% ($M = 8.96$, $SD = 5.66$), $t(29) = 10.10$, $p < .001$, $d = 0.76$ [4]. Significantly more errors were also made when distractor voices were manipulated by 7% ($M = 15.97$, $SD = 7.31$) compared to when they were manipulated by 10% ($M = 8.96$, $SD = 5.66$), $t(29) = 5.63$, $p < .001$, $d = 0.39$. No other main effects were significant or close to significance (adjusted $p > .93$).

In addition to the main effects, there was a significant interaction between target F0 and direction of manipulation, $F(2, 56) = 9.27$, $p < .05$, $\eta_g^2 = .03$ [5]. As can be seen in Figure 1, the

[3] Generalised eta-squared statistics (η_g^2) are reported here in order to facilitate comparison between studies with different designs. Generalised eta-square describes the proportion of sample variance accounted for by an effect in an independent design with no manipulated factors (Olejnik & Algina, 2003).

[4] Cohen's *d* values were determined by calculating the mean difference between the two groups, and then dividing the result by the overall pooled standard deviation from all conditions (for F0 = 17.95, for speech rate = 22.98).

[5] The tests of simple main effects that follow are again adjusted using the Hochberg correction. Note that the Hochberg correction is not conditional on a significant *F* ratio in order to protect the Type 1 error. We corrected

listeners selected higher F0 distractors more often than they selected lower F0 distractors. This effect was strongest for low F0 target voices, with more errors being made when target voices were paired with distractors higher in F0 (M = 27.50, SD = 10.63) than distractors lower in F0 (M = 8.89, SD = 9.11), $t(29) = 8.37$, $p < .001$, $d = 1.04$. A similar pattern of findings was apparent for high F0 target voices, with more errors being made when target voices were paired with distractors higher in F0 (M = 17.08, SD = 12.09) than distractors lower in F0 (M = 8.75, SD = 7.48), $t(29) = 3.73$, $p < .01$, $d = 0.46$. More errors were also made when moderate F0 target voices were paired with distractors higher in F0 (M = 19.03, SD = 13.07) than distractors lower in F0 (M = 13.89, SD = 9.21), $t(29) = 2.40$, $p < .05$, $d = 0.29$.

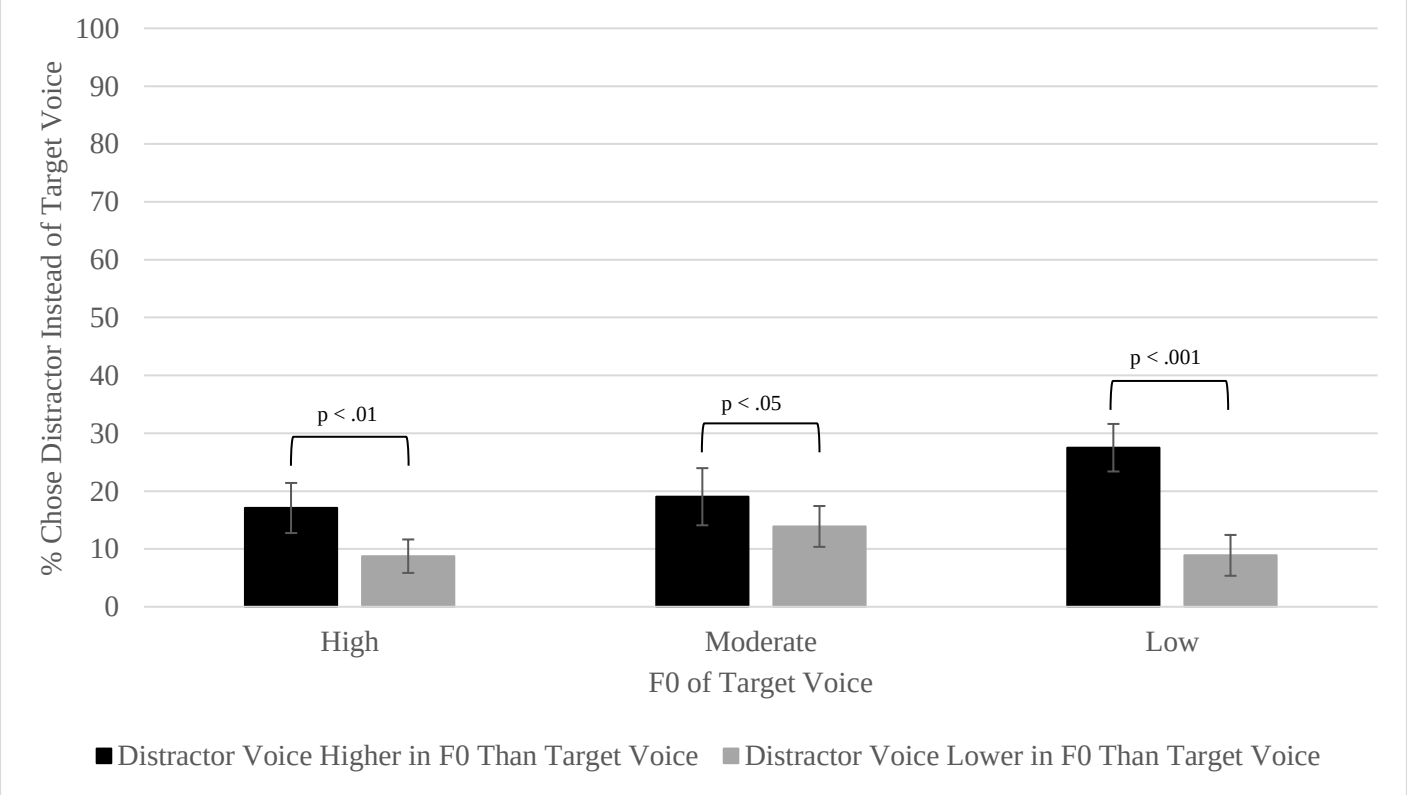

Figure 1. Mean percentage of errors made (i.e. chose distractor voice instead of target voice) for the three F0 target voice conditions. 95% confidence intervals are also shown.

for all six possible simple main effects. However, for reasons of brevity we report here only the three simple main effects that are of direct interest.

There was also a significant interaction between direction of manipulation and magnitude of distractor change, $F(2, 56) = 18.01$, $p < .03$, $\eta_g^2 = .04$ [6]. Figure 2 shows that listeners selected distractor voices higher in F0 more often than they selected distractor voices lower in F0 when identifying target voices. This effect was strongest for distractor voices that sounded more similar in F0 to target voices. Specifically, listeners made more errors for distractor voices higher in F0 ($M = 30.83$, $SD = 11.24$) than distractor voices lower in F0 ($M = 14.44$, $SD = 7.24$) when distractor voices were manipulated by 5%, $t(29) = 8.05$, $p < .001$, $d = 0.91$. Listeners also made more errors for distractor voices higher in F0 ($M = 22.36$, $SD = 10.64$) than distractor voices lower in F0 ($M = 9.58$, $SD = 6.58$) when distractor voices were manipulated by 7%, $t(29) = 7.02$, $p < .001$, $d = 0.72$. A similar pattern of findings was also observed for distractor voices that sounded less similar in F0 to target voices (i.e. manipulated by 10%), with more errors being made for distractor voices higher in F0 ($M = 10.42$, $SD = 6.17$) than distractor voices lower in F0 ($M = 7.50$, $SD = 7.37$), $t(29) = 2.13$, $p < .05$, $d = 0.16$.

[6] We corrected for all nine possible simple main effects. However, for reasons of brevity we report here only the three simple main effects that are of direct interest.

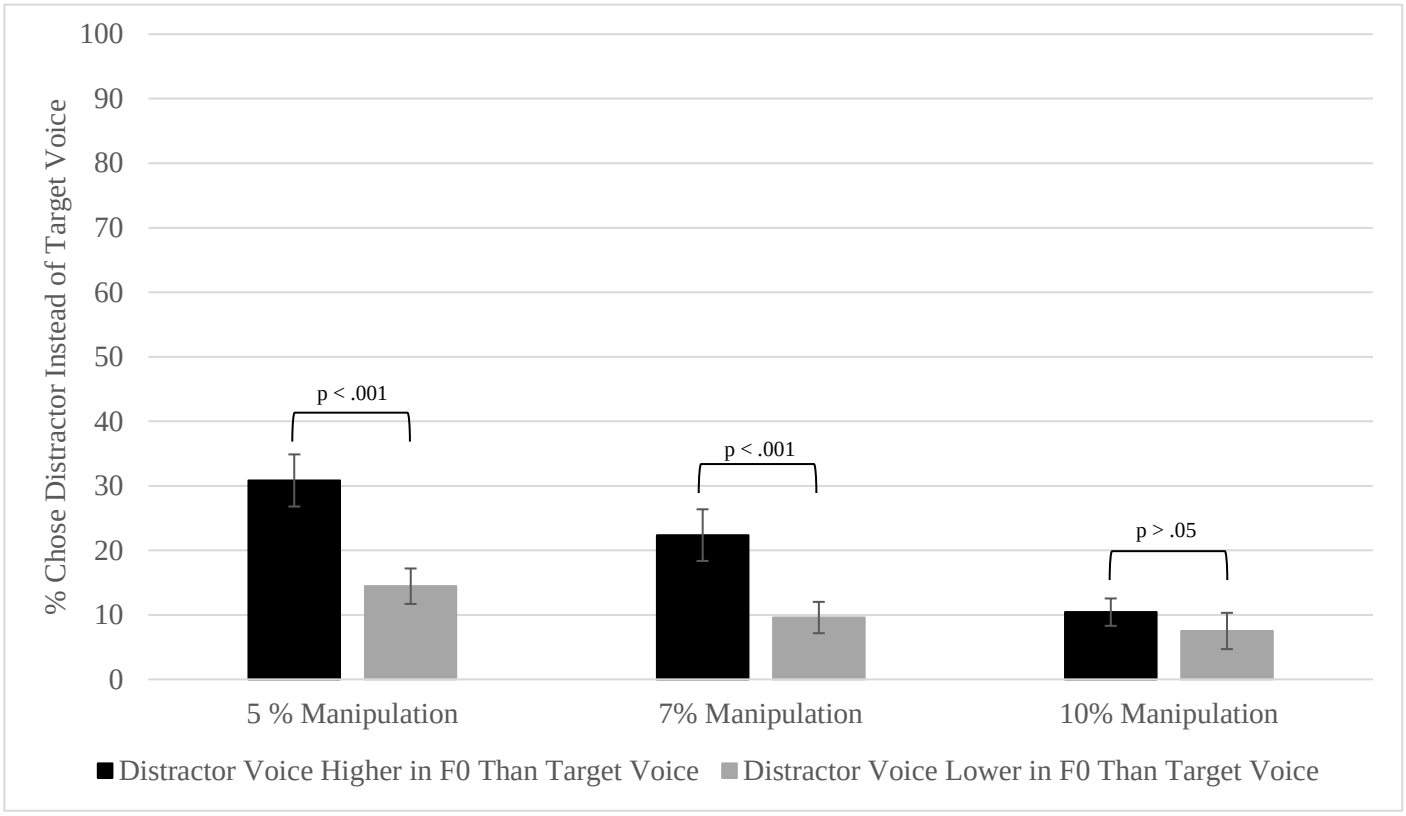

Figure 2. Mean percentage of errors made for F0 (i.e. chose distractor voice instead of target voice) for the 5%, 7%, and 10% distractor manipulations. 95% confidence intervals are also shown.

No other interaction effects were significant or close to significance (adjusted $p > .31$).

Speech Rate

For speech rate, the percentage of mean matching errors made for each distractor type, listed separately for each of the three target conditions (fast, moderate and slow speech rate), the sex of the target voice, and listener sex, are presented in Table 2.

Table 2.

Mean percentage of errors made by distractor type (magnitude and direction of distractor change), target speech rate (fast, moderate (mod), or slow), sex of target voice (collapsed across male and female target voices) and sex of listener (male or female).

Male Listener						Female Listener					
Male Target Voice			Female Target Voice			Male Target Voice			Female Target Voice		
Fast	Mod	Slow	Fast	Mod	Slow	Fast	Mod	Slow	Fast	Mod	Slow

Distractor												
+20%	**20.00**	**11.67**	**35.00**	**30.17**	**10.00**	**33.33**	**15.00**	**18.33**	**30.00**	**25.00**	**18.33**	**30.00**
	21.55	*20.85*	*29.58*	*28.79*	*18.33*	*24.40*	*18.42*	*14.84*	*25.36*	*32.73*	*24.03*	*25.36*
+12%	**40.00**	**26.67**	**25.00**	**30.00**	**30.00**	**48.33**	**35.00**	**33.33**	**26.67**	**30.00**	**25.00**	**45.00**
	26.39	*22.09*	*16.37*	*27.06*	*28.66*	*22.09*	*18.42*	*26.16*	*17.59*	*27.06*	*18.90*	*19.37*
+10%	**33.33**	**43.33**	**41.67**	**30.17**	**30.00**	**51.67**	**38.33**	**35.00**	**40.00**	**30.17**	**26.67**	**45.00**
	26.16	*33.36*	*18.09*	*28.79*	*25.36*	*25.82*	*24.76*	*28.03*	*29.58*	*28.79*	*17.59*	*23.53*
-10%	**33.33**	**43.33**	**20.00**	**26.67**	**33.33**	**21.67**	**41.67**	**28.33**	**23.33**	**40.00**	**25.00**	**30.00**
	27.82	*29.07*	*21.55*	*22.09*	*26.16*	*22.89*	*34.93*	*26.50*	*25.82*	*22.76*	*18.90*	*19.37*
-12%	**38.33**	**28.33**	**16.67**	**35.00**	**20.00**	**36.67**	**35.00**	**23.33**	**23.33**	**38.33**	**20.00**	**35.00**
	20.85	*28.14*	*20.41*	*24.64*	*23.53*	*22.89*	*29.58*	*25.82*	*14.84*	*26.50*	*25.36*	*22.76*
-20%	**21.67**	**13.33**	**6.67**	**26.67**	**13.33**	**15.00**	**23.33**	**15.00**	**10.00**	**23.33**	**10.00**	**6.67**
	22.89	*12.91*	*14.84*	*24.03*	*20.85*	*18.42*	*17.59*	*22.76*	*20.70*	*19.97*	*15.81*	*11.44*

Note: Means are shown in bold. Standard deviations (SD) are shown in italics.

The matching error scores for each listener were entered in a mixed ANOVA for the between subjects factor of listener sex (male or female) and the within subjects factors of sex of voice (male or female), target speech rate (fast, moderate or slow), magnitude of distractor change (10%, 12%, or 20%) and direction of manipulation (increase or decrease in rate). This revealed a significant main effect of direction of manipulation, $F(1, 28) = 12.55$, $p < .05$, $\eta_g^2 = .02$, with significantly more errors being made when the distractor voices were faster in

speech rate (M = 30.56, SD = 7.48) than when they were slower in speech rate (M = 25.05, SD = 8.06). There was also a main effect of magnitude of distractor change, $F(1, 28) = 50.27$, $p < .05$, $\eta_g^2 = .10$. Significantly more errors were made when distractor voices were manipulated by 10% (M = 33.69, SD = 8.35) compared to when they were manipulated by 20% (M = 18.75, SD = 7.95), $t(29) = 9.90$, $p < .001$, $d = 0.65$. Significantly more errors were also made when distractor voices were manipulated by 12% (M = 30.97, SD = 8.29) compared to when they were manipulated by 20% (M = 18.75, SD = 7.95), $t(29) = 9.03$, $p < .001$, $d = 0.53$. However, there were no differences in errors made for distractor voices manipulated by 10% (M = 33.69, SD = 8.35) and 12% (M = 30.97, SD = 8.29), $t(29) = 1.52$, $p > .05$, $d = 0.12$. No other main effects were significant or close to significance (adjusted p >.96).

In addition to the main effects, there was also a significant interaction between target speech rate and direction of manipulation, $F(2, 56) = 15.12$, $p < .05$, $\eta_g^2 = .06$ [7]. Figure 3 shows that for slow speech rate target voices, listeners selected distractors faster in rate (M = 37.64, SD = 11.13) more often than they selected distractors slower in rate (M = 20. 42, SD = 10.68), $t(29) = 6.34$, $p < .001$, $d = 0.75$. However, there was no difference in the selection of distractors faster in rate (M = 28.35, SD = 14.86) and distractors slower in rate (M = 31.94, SD = 13.33) for fast speech rate target voices, $t(29) = -1.22$, p >.05, $d = 0.16$. Furthermore, there was no difference in the selection of distractors faster in rate (M = 25.69, SD = 13.97) and distractors slower in rate (M = 22.78, SD = 12.89) for moderate speech rate target voices, $t(29) = 1.20$, $p > .05$, $d = 0.13$.

[7] We corrected for all nine possible simple main effects. However, for reasons of brevity we report here only the three simple main effects that are of direct interest.

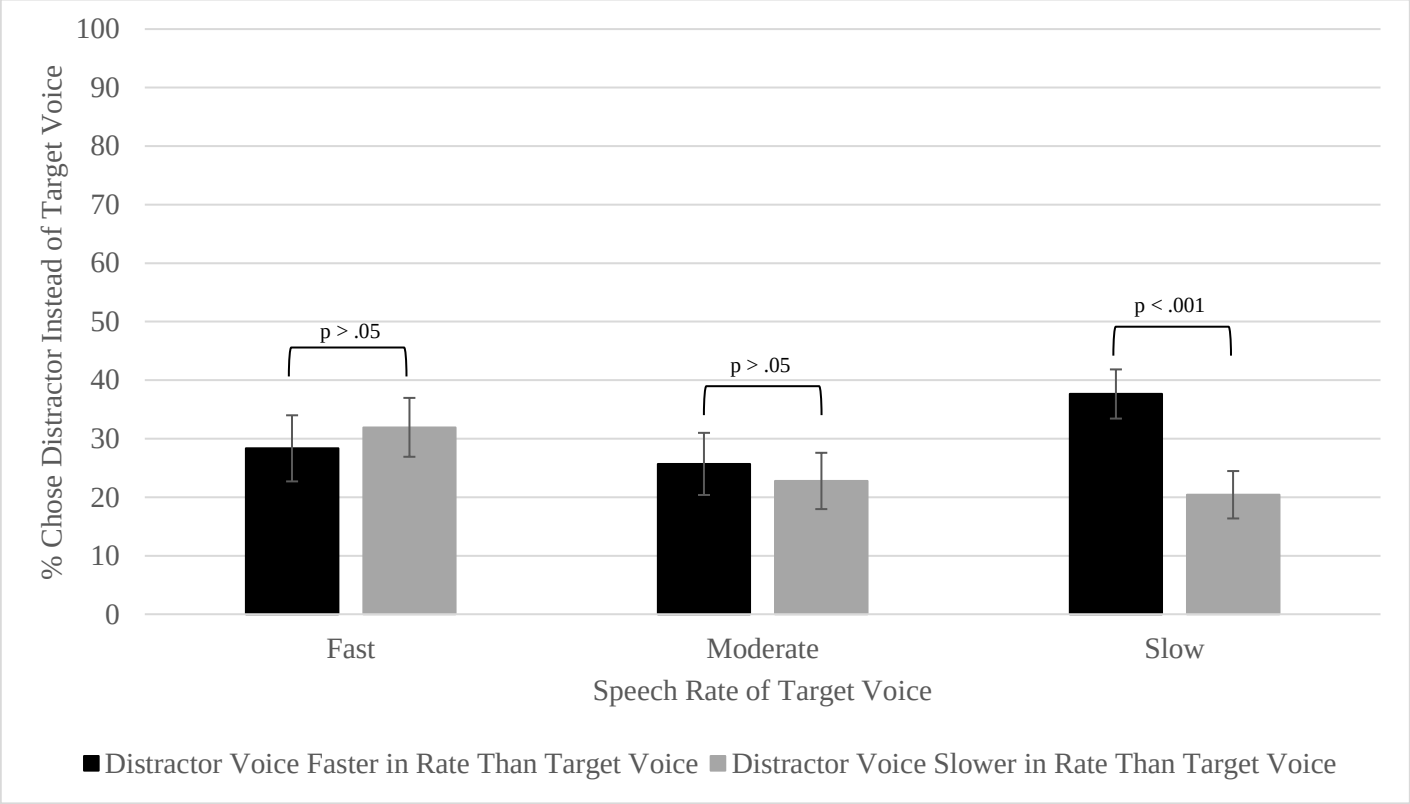

Figure 3. Mean percentage of errors made (i.e. chose distractor voice instead of target voice) for the three speech rate target voice conditions. 95% confidence intervals are also shown.

No other interaction effects were significant or close to significance (adjusted $p > .43$).

Discussion

The current research investigated the impact of manipulations in F0 and speech rate on immediate target matching performance (selecting a voice from a pair to match a previously heard target voice) for a range of unfamiliar synthesised voices. We found that there was an increase in the selection of voices higher in F0 when high, moderate, and low F0 target voices were presented. For speech rate, there was an increase in the selection of voices faster in speech rate when slow speech rate target voices were presented. However, no such effect was detected for fast and moderate speech rate target voices. Therefore, in terms of our original hypotheses, there was no evidence for accentuation effects for voice memory. Furthermore, for both the F0 and speech rate conditions, more errors were made identifying target voices when paired with distractor voices manipulated by a smaller magnitude (i.e. 5% for F0, and 10% for speech rate)

compared to those manipulated by a greater magnitude (i.e. 10% for F0, and 20% for speech rate). This is perhaps unsurprising given that the results from the pilot study suggest that voices manipulated by a smaller magnitude are harder to distinguish between, and sound more similar, to original voices than voices manipulated by a greater magnitude. Thus, more errors are likely to be made identifying target voices when paired with distractor voices manipulated by a smaller magnitude because any differences between the voices are more difficult to detect. There was no effect of either sex of voice or listener sex on errors made identifying target voices.

Fundamental Frequency (F0)

The results presented here do offer some support to those identified by Mullenix et al. (2010) in that errors in memory are likely to occur for voice F0. However, the finding of an increase in the selection of voices higher in F0 is difficult to explain using the accentuation effect alone. We believe that this outcome is not an anomaly in our data set given that the findings are reasonably consistent across all target voices. They are also unlikely to be the result of order effects because we counterbalanced the voices that were presented to listeners in the voice pair. Given that synthesised voices were used for experimentation, we do acknowledge that some of the acoustic properties of the stimuli could explain the observed pattern of findings. However, we believe this is unlikely given that the voices used were rated as sounding natural, formant frequencies changed freely, formant transitions were smooth, and there were no intonational irregularities or prosodic mismatches across words. This alleviates concerns that something uncontrolled and artificial about the stimuli were driving the findings. Rather, we propose that the extensions and modifications made to the study procedure may explain the difference in results. First, we kept the target and distractor voices within a F0 range that is typical in the population (i.e. between 80-180 Hz, and for an adult female this will be between 165-255 Hz (Titze, 1994). In contrast, the manipulations made by Mullenix et al.

(2010) fell considerably outside of this range. Second, we used a set of four synthesised voices, whereas Mullenix et al. (2010) used only a single voice. Therefore, it is quite possible that the findings identified by Mullenix et al. (2010) were due to the peculiarity of the stimuli (i.e. an unusually high or low F0) used in the experiment. Using a more representative and generalizable set of voices, as in the present study (i.e. a slightly larger set of synthesised voices, with manipulations in F0 and speech rate kept within a range that is typical in the population for English speakers), the accentuation bias is no longer found. The data reported here suggest little or no accentuation bias for the memory of voice F0.

Why then do listeners make more errors recognising voice F0 when paired with distractor voices higher in F0 compared to when they are paired with distractor voices lower in F0? It is quite possible that listeners had difficulty discriminating between the frequencies of some of the voice pairs in the experiment. Indeed, research has identified that it is more difficult to discriminate between voices of higher frequencies compared to voices of lower frequencies (Moore, 1995). In the present study, listeners may have made fewer errors identifying target voices when paired with distractor voices lower in F0 because they were more efficient at detecting the changes in frequency than when distractor voices were higher in frequency. This interpretation would account for why there was no effect of listener sex on errors made identifying target voices, because there is no reason to believe that the perceptual capabilities of the listener would differ substantially between male and female listeners. It would also explain why there was no difference in errors made for male and female target voices. Although female voices are higher in F0 than male voices, the findings are based upon a listener's ability to detect any *differences* in the frequencies of the voices in the voice pair, and this is independent of the frequency of the target voice itself.

It is also likely that listeners made more errors identifying target voices when paired with distractor voices higher in F0 compared to when they are paired with distractor voices

lower in F0 because they resemble voices that are typically heard in the general population. Inflection refers to the frequency patterns in a person's speech, where the voice rises and falls, either upwards or downwards in frequency (Fairbanks, 1940). Research has shown that all types of inflections are greater in upward inflection than they are in downward inflection (e.g. Barbaranne, 1981; Fairbanks & Pronovost, 1939). Furthermore, researchers have shown that when people are asked to choose a method of disguise, they are more likely to raise the frequency of their voice rather than lowering it (e.g. Mathur, Choudhary, & Vyas, 2016; Masthoff, 1996). Such evidence suggests that people are more likely to increase, rather than decrease, the frequency of their voice when they speak. Thus, the listeners in the present study may be selecting distractor voices higher in F0 more often than distractor voices lower in F0 because they are more familiar with these types of utterances and it sounds like a more plausible version of the target voice (i.e. an inflected version of the target voice).

The finding that listeners were more likely to select distractor voices higher in F0 compared to distractor voices lower in F0 was particularly prevalent for the low F0 target voice condition. This bias may have arisen because voices higher in F0 are perceived as less threatening than voices lower in F0. Research has shown that both male and female voices lowered in F0 are perceived as more dominant than the same voices raised in F0 (Borkowska & Pawlowski, 2011; Fraccaro, O'Connor, Re, Jones, DeBruine, & Feinberg, 2012; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Puts, Gaulin, & Verdonili, 2006). Furthermore, evidence tends to suggest that people will often exhibit avoidance type behaviour when exposed to aversive stimuli (Corr, 2013). Assuming that voices lower in F0 would be rated as more dominant and threatening than the voices higher in F0 in the present study, listeners may have selected the higher voice of the pair because it sounded less dominant and less threatening. This interpretation would explain why an increase in the selection of higher F0 distractors was particularly prevalent for the low F0 target voice condition; because the voices were decreased

in F0 sufficiently for the higher F0 voices in the pair to be perceived as less threatening to the listener. It would also account for why there was no effect of either sex of voice or listener sex; perceptions of dominance have been found to be equivalent for both male and female voices and male and female listeners (Jones, Feinberg, DeBruine, Little, & Vukovic, 2010). Further work would be required to confirm or disconfirm this explanation to our finding. Another possibility that also deserves equal consideration is that English voices lower in F0 for both males and females tend to co-occur with covariations in voice quality (e.g. Aberton, Howard, & Fourcin, 1989). A bias towards selecting the higher F0 distractor voices could reflect the unnaturalness of the voices lowered in F0 without a concomitant change in voice quality. Whilst the voices were rated as sounding natural, this issue might still remain even if naturally sounding voices were modified to have a lower F0.

Finally, as pointed out by one reviewer, for which we are grateful, it is worth noting that the naturalness ratings for the voices with higher F0 manipulations tended to yield slightly higher naturalness rating scores than those with lower manipulations (refer to Appendix C for further details). One possible interpretation of this is that the listeners preferred the more natural sounding voices (i.e. the higher F0 manipulations) and were thus, more likely to select them. Unfortunately, because the naturalness ratings came from a different population to those in the 2AFC tasks reported here, it was not appropriate to formally test this possibility. Our tuition, given that the voices were generally perceived to be natural sounding across the board, the differences observed between the voices being relatively small, are unlikely to have impacted upon the matching tasks. Thus, whilst we accept it is a possibility that naturalness may have an effect, we are unable to resolve the question here.

Speech Rate

For speech rate, listeners selected voices faster in speech rate when slow speech rate target voices were presented. Thus, the findings presented here cannot be explained using the accentuation effect. Given this, it is possible that the findings could be accounted for by the listener's level of familiarity of the voice heard. In natural speech, a person speaking more slowly is likely to be more hesitant, making more silent pauses or filled pauses (e.g. *um, er*). In the present study, a decreasing speech rate did affect the rate of continuous production but did not lead to increased pauses of any kind. It is therefore unlikely that the speech samples used were an entirely natural rendition of slower speech, at least of a type that listeners most typically hear. It is possible that at the lower margins of the speech rate manipulated samples (i.e. the slowest samples), but not elsewhere, the participants may have selected a faster voice in the pair because it sounded more realistic.

Faster speaking voices might also sound more favourable when compared with slower speaking voices in the slow speech rate pairings. Indeed, research suggests that speech rates can influence a listener's perceptions of a speaker's personality and social skills. For example, faster speaking styles have been shown to be rated more favourably (Stewart & Ryan, 1982), and viewed as more competent and socially attractive than voices spoken at a slower rate (Street, Brady, & Putman, 1983). Slower speaking styles have also been identified as sounding weaker, less truthful, and less empathetic than voices spoken at a faster rate (Apple, Streeter, & Krauss, 1979). It is possible that listeners were more likely to select a faster voice in the pair because they preferred the sound of the voice. However, such selections may have been made only for the slow speech rate condition because these voices were slowed sufficiently for the faster rate voices in the pair to be rated more favourably, and thus selected by the listener. The above explanations would also account for why there was no effect of either sex of voice or listener sex on errors made identifying a target voice, as there is no reason to suggest that the level of

familiarity or preference for faster voices would differ between male and female voices, or for male and female listeners.

Concluding Comments

The results from the present study suggest that, at least for synthesised voices, listeners are susceptible to distortions in memory for certain properties of the voice more so than others. However, the accentuation bias does not account for our findings here. Therefore, it is doubtful that listeners rely solely on the categorical information self-generated about the voice at the time of encoding to aid in recognition of the voice at a later stage. The present study has thus contributed to our understanding of the mechanisms important for accurate voice recognition and such work may prove as a useful conceptual tool in determining the properties of voice that are more or less affected by intra-individual variation. Future work in this field should focus on framing their research with a more applied perspective in mind. For example, it would be particularly valuable to establish whether the results from the present study would also extend to real, rather than synthesised voices. Future work could also be undertaken to determine the impact of longer retention intervals on errors made identifying voices. This is especially interesting given that in a real world criminal situation there is uncertainty over the time period between hearing a voice and being asked to identify the voice at a later date. Such work would undoubtedly advance on the research currently being carried out in this domain and further our understanding of the impact of manipulations of certain characteristics of our voice, whether it be through unintentional or deliberate means.

References

Aberton, E. R. M., Howard, D. M., & Fourcin, A. J. (1989). Laryngographic assessment of normal voice: A tutorial. *Clinical Linguistics and Phonetics,* 281-296. doi: http://dx.doi.org/10.3109/02699208908985291

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37,* 715-727. doi: http://dx.doi.org/10.1037/0022-3514.37.5.715

Arnfield, S., Roach, P., Greasley, P., & Horton, D. (1995). Emotional stress and speech tempo variation. *Processings of ESCA-NATO Tutorial and Research Workshop on Speech Under Stress.* Lisbon, 13-15.

Barker, B.A., & Newman, R. S. (2004). Listen to your mother! The role of familiarity in infant streaming. *Cognition, 94,* B45-B53. doi: http://dx.doi.org/10.1016/j.cognition.2004.06.001

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour, 82,* 55-59. doi: http://dx.doi.org/10.1016/j.anbehav.2011.03.024

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion, 24,* 377-400. doi: http://dx.doi.org/10.1080/02699930902975754

Brown, A. (2014). Pronunciation and phonetics: A practical guide for English language teachers. New York: Routledge. doi: http://dx.doi.org/10.1075/jslp.1.2.07mac

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts toward more typical category exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology, 86,* 236-250. doi: http://dx.doi.org/10.1037/0022-3514.86.2.236

Eiser, J. R. (1971). Enhancement of contrast in the absolute judgement of attitude judgements. *Journal of Personality and Social Psychology, 17,* 1-10. doi: http://dx.doi.org/10.1037/h0030455

Fraccaro, P. J., O'Connor, J. J., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: Deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour, 85,* 127-136. doi: http://dx.doi.org/10.1016/j.anbehav.2012.10.016

Halberstadat, J. B., & Niedenthal, P. M. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology, 81,* 587-598. doi: http://dx.doi.org/10.1037/0022-3514.81.4.587

Endres, W., Bambach, W., & Flosser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *The Journal of The Acoustical Society of America, 49,* 1842-1848. doi: http://dx.doi.org/10.1121/1.1912589

Haslam, S. A., & Turner, J. C. (1992). Context-dependent variation in social stereotyping 2: The relationship between frame of reference, self-categorization and accentuation. *European Journal of Social Psychology, 22,* 251-278. doi: http://dx.doi.org/10.1002/ejsp.2420220305

Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory and Cognition, 25,* 801-811. doi: http://dx.doi.org/10.3758/bf03211324

Hilliar, K. F., & Kemp, R. I. (2008). Barak Obama or Barry Dunham? The appearance of multiracial faces is affected by the names assigned to them. *Perception, 37,* 1605-1608. doi: http://dx.doi.org/10.1068/p6255

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–803. doi: http://dx.doi.org/10.1093/biomet/75.4.800

Hogg, M., & Vaughan, G. (2010). *Essentials of social psychology.* Harlow: Pearson Education Limited.

Fiske, S. T., Gilbert, D. T., & Lindzey, G. (2010). *Handbook of social psychology: Volume one*. New jersey: John Wiley & Sons, Inc. doi: http://dx.doi.org/10.1002/9780470561119

Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology, 41,* 598-608. doi: http://dx.doi.org/10.1016/j.jesp.2004.10.007

Jones, B. C., Feinberg, D., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour, 79,* 57-62. doi: http://dx.doi.org/10.1016/j.anbehav.2009.10.003

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67,* 596-610. doi: http://dx.doi.org/10.1037/0022-3514.67.4.596

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology, 59,* 651-663. doi: http://dx.doi.org/10.1037/0022-3514.59.4.651

Levi, S. V. (2015). Talker familiarity and spoken word recognition in school-age children. *Journal of Child Language, 42,* 843-872. doi: http://dx.doi.org/10.1017/s0305000914000506

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

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135,* 501-512. doi: http://dx.doi.org/10.1037/0096-3445.135.4.501

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favouring women in verbal but not visuospatial episodic memory. *Neuropsychology, 15,* 165-173. doi: http://dx.doi.org/10.1037/0894-4105.15.2.165

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7,* 98-118. doi: http://dx.doi.org/10.1037/1076-8971.7.1.98

McGarty, C., & Penny, R. E. C. (1988). Categorization, accentuation and social judgement. *British Journal of Social Psychology, 27,* 147-157. doi: http://dx.doi.org/10.1111/j.2044-8309.1988.tb00813.x

McGarty, C., & Turner, J. C. (1992). The effects of categorisation on social judgement. *British Journal of Social Psychology, 31,* 253-268. doi: http://dx.doi.org/10.1111/j.2044-8309.1992.tb00971.x

McGivern, R. F., Huston, J. P., Byrd, D., King, T., Siegle, G. J., & Reilly, J. (1997). Sex differences in visual recognition memory: Support for sex-related difference in attention in adults and children. *Brain and Cognition, 34,* 323-336. doi: http://dx.doi.org/10.1006/brcg.1997.0872

Moore, B.C. J. (1995). *Hearing.* San Diego, CA: Academic Press.

Mullenix, J. W., Stern, S. E., Grounds, B., Kalas, R., Flaherty, M., Kowalok, S., May, E., & Tessmer, B. (2010). Earwitness memory: Distortions for voice pitch and speaking rate. *Applied Cognitive Psychology, 24,* 513-526. doi: http://dx.doi.org/10.1002/acp.1566

Newman, R. S., & Evers, S. E. (2007). The role of talker familiarity on stream segregation. *Journal of Phonetics, 35,* 85-103. doi: http://dx.doi.org/10.1016/j.wocn.2005.10.004

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods,* 8, 434-447. doi: http://dx.doi.org/10.1037/1082-989X.8.4.434

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162,* 8-13. doi: http://dx.doi.org/10.1016/j.jneumeth.2006.11.017

Puts, D. A., Gaulin, S. J. C., & Verdonili, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27,* 283-296. doi: http://dx.doi.org/10.1016/j.evolhumbehav.2005.11.003

Queller, S., Schell, T., & Mason, W. (2006). A novel view of between-categories contrast and within-category assimilation. *Journal of Personality and Social Psychology, 91,* 406-422. doi: http://dx.doi.org/10.1037/0022-3514.91.3.406

Reich, A. R., & Duke, J. E. (1979). Effects of selected vocal disguise upon speaker identification by listening. *The Journal of The Acoustical Society of America, 66,* 1023-1028. doi: http://dx.doi.org/10.1121/1.383321

Reich, A. R., Moll, K. L., & Curtis, J. F. (1976). Effects of selected vocal disguises upon spectrographic speaker identification. *The Journal of The Acoustical Society of America, 60,* 919-925. doi: http://dx.doi.org/10.1121/1.2002461

Roebuck, R., & Wilding, J. (1993). Effects of vowel variety and sample length on identification of a speaker in a line-up. *Applied Cognitive Psychology, 7,* 475-481. doi: http://dx.doi.org/10.1002/acp.2350070603

Rose, P., & Duncan, S. (1995). Naïve auditory identification and discrimination of similar voices by familiar listeners'. *Forensic Linguistics, 2,* 1-17. doi: http://dx.doi.org/10.1558/ijsll.v2i1.1

Stern, S. E., Mullenix, J. W., Corneille, O., & Huart, J. (2004). Distortions in the memory of the pitch of speech. *Experimental Psychology, 54,* 148-160. doi: http://dx.doi.org/10.1027/1618-3169.54.2.148

Street, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. *Journal of Language & Social Psychology, 2,* 37-56. doi: http://dx.doi.org/10.1177/0261927x8300200103

Stewart, M. A., & Ryan, E. B. (1982). Attitudes toward younger and older adult speakers: Effects of varying speech rates. *Journal of Language and Social Psychology, 1,* 91-109. doi: http://dx.doi.org/10.1177/0261927x8200100201

Sutton, R., & Douglas, K. (2013). *Social psychology.* Basingstoke : Palgrave Macmillan.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology, 54,* 101-114. doi: http://dx.doi.org/10.1111/j.2044-8295.1963.tb00865.x

Titze, I. R. (1994). Principles of voice production. Englewood Cliffs, NJ: Prentice-Hall. doi: http://dx.doi.org/10.1121/1.424266

Zhang, C. (2012). Acoustic analysis of disguised voices with raised and lowered pitch. *Chinese Spoken Language Processing (ISCSLP),* 353-357. doi: http://dx.doi.org/10.1109/iscslp.2012.64