Clearly, fame isn't everything: Talker familiarity does not augment talker adaptation

Emma R. Hatter[1] · Caleb J. King[1] · Anya E. Shorey[1] · Christian E. Stilp[1]

Abstract

Familiarity with a talker's voice provides numerous benefits to speech perception, including faster responses and improved intelligibility in quiet and in noise. Yet, it is unclear whether familiarity facilitates talker adaptation, or the processing benefit stemming from hearing speech from one talker compared to multiple different talkers. Here, listeners completed a speeded recognition task for words presented in either single-talker or multiple-talker blocks. Talkers were either famous (the last five Presidents of the United States of America) or non-famous (other male politicians of similar ages). Participants either received no information about the talkers before the word recognition task (Experiments 1 and 3) or heard the talkers and saw their names first (Experiment 2). As expected, responses were faster in the single-talker blocks than in the multiple-talker blocks. Famous voices elicited faster responses in Experiment 1, but familiarity effects were extinguished in Experiment 2, possibly by hearing all voices recently before the experiment. When talkers were counterbalanced across single-talker and mixed-talker blocks in Experiment 3, no familiarity effects were observed. Predictions of familiarity facilitating talker adaptation (smaller increase in response times across single- and multiple-talker blocks for famous voices) were not confirmed. Thus, talker familiarity might not augment adaptation to a consistent talker.

Keywords Speech perception · Audition · Spoken word recognition

Introduction

In the long and rich history of research in speech perception, two characteristics are well documented. First, when hearing speech, both what is being said and who is speaking are inter-dependent (e.g., Mullennix & Pisoni, 1990). Given this inter-dependency, attending to the linguistic contents of speech is facilitated by the consistency afforded by hearing the same talker (i.e., talker adaptation, also referred to as talker normalization[1]). This has been demonstrated using various different tasks, such as word identification (Mullennix et al., 1989),

word list recall (Martin et al., 1989), vowel monitoring (Magnuson & Nusbaum, 2007), Japanese mora monitoring (Magnuson et al., 2021), voice classification (Mullennix & Pisoni, 1990), lexical tone categorization (Zhang & Chen, 2016), categorization of isolated phonemes (Assmann et al., 1982), and categorization of phonemes following contextual sounds (Assgari & Stilp, 2015). Across studies, results have repeatedly demonstrated that talker adaptation makes speech perception faster and/or more accurate.

Second, other research has shown that familiarity with a talker's voice can aid speech perception. This benefit has been reported in a wide variety of tasks, such as word recognition in noise (Nygaard et al., 1994), sentence recognition in noise (Souza et al., 2013) or amidst competing voices (Holmes et al., 2018; Johnsrude et al., 2013; Souza et al., 2013), sentence recognition in a dual-task paradigm (Ingvalson & Stoimenoff, 2015), or speeded shadowing (Maibauer et al., 2014). Familiarity with a talker's voice makes it more intelligible, even if it is not immediately recognizable (Holmes et al., 2018). These benefits of talker familiarity have been observed across a wide range of timescales, from recently trained-on voices (Nygaard et al., 1994) to long-term spouses (Johnsrude et al., 2013).

---

[1] Some investigations approach this process by focusing on the processing costs incurred by talker variability (i.e., speech perception becoming slower and/or less accurate when hearing multiple talkers). Here, we focus on the perceptual benefits yielded by talker consistency (i.e., speech perception becoming faster and/or more accurate when hearing one consistent talker) in order to align with other perceptual benefits yielded by talker familiarity, reviewed below.

✉ Christian E. Stilp
christian.stilp@louisville.edu

[1] Department of Psychological and Brain Sciences, University of Louisville, 317 Life Sciences Building, Louisville, KY, USA

Springer

Viewing these two literatures together, talker adaptation and talker familiarity can each benefit speech perception. To date, it is unclear whether these two factors interact with each other. Recent theoretical approaches promote parallel and interactive processing between the acoustic details that signify talkers and words (Pierrehumbert, 2016). Since variability along one dimension challenges perception of the other (Mullennix & Pisoni, 1990; Nygaard et al., 1994; Nygaard & Pisoni, 1998), facilitation along one dimension (via talker familiarity) might also be faciliatory for perception along the other dimension (word recognition in talker adaptation paradigms). Together with the above-cited findings that familiarity can improve speech perception in many different tasks, one might predict that familiarity would promote talker adaptation, improving perception of familiar talkers and eliciting more comparable performance across single-talker and mixed-talker blocks relative to when talkers are unfamiliar.

Magnuson et al. (2021) tested this prediction, with familiar voices for a given participant consisting of speech from their spouse and child (which then served as unfamiliar voices for other participants). In their first experiment, reaction times in a speeded monitoring task were shorter in the single-talker block than in the mixed-talker block, as expected. Contrary to predictions, reaction times for familiar and unfamiliar talkers did not differ. In their third experiment, the same participants completed a transcription task for Japanese morae spoken by the same familiar voices (family members), voices on which they were trained in a separate experiment (trained-on voices), or other voices that were heard in a previous experiment without any additional training (exposed-to voices). While accuracy across conditions was in the predicted order (familiar > trained-on > exposed-to voices), these were not statistically significantly different. Reaction times were not collected in this task. Therefore, it is still unclear if familiarity augments talker adaptation. The null results in Magnuson et al. (2021) could be due to the specific task tested (Japanese mora monitoring or transcription), the talkers used in the experiment, the language being studied, participant characteristics, or any number of other variables.

The present study tested whether talker familiarity promotes talker adaptation in a speeded word recognition task. For familiar (famous) talkers, stimuli consisted of the voices of Presidents of the USA, which were assumed to be recognizable to the general public in the United States. For unfamiliar (non-famous) talkers, stimuli consisted of the voices of less-famous politicians, age-matched to the famous talkers. In a within-subjects design, talker familiarity (famous, non-famous) was fully crossed with talker consistency (single talker, mixed talkers), similar to the general paradigm tested by Magnuson et al. (2021). There were three predictions for the present study. The first was that reaction times for single talkers would be faster than for mixed talkers, as reported in numerous previous studies (e.g., Lim et al., 2019; Magnuson

& Nusbaum, 2007; Magnuson et al., 2021; Nusbaum & Morin, 1992; Stilp & Theodore, 2020). The second prediction was that reaction times for famous talkers would be faster than for non-famous talkers, as seen in Maibauer et al. (2014).[2] Critically, the third prediction was that talker familiarity would promote adaptation to famous talkers in single- and mixed-block formats, making the effects of talker consistency (i.e., the difference in response times across single-talker and mixed-talker blocks) smaller for famous talkers than for non-famous talkers.

Experiment 1

Methods

Participants

Forty-five undergraduate students from the University of Louisville (Louisville, KY, USA) participated in Experiment 1. Participants' mean age was 19.58 years (SD = 2.83). Three additional participants were tested but their results were not included in analyses, as detailed below (see Results). All participants were self-reported native English speakers with no known hearing deficits. Participants received course credit in exchange for their participation.

Stimuli

Stimuli were excised from speeches on americanrhetoric.com, an online speech bank. The words "do" and "to" and their homophones were extracted from speeches using Audacity 3.0.0 software (Audacity Team, 2021). One token of each word from each talker was used in testing. The mean duration of "do" tokens was 177.65 ms, and the mean duration of "to" tokens was 199.57 ms. The mean fundamental frequency (f0) and duration of target words are listed by condition in Table 1. For a post-task questionnaire, a sentence of roughly 2 s duration was also extracted from the same source using Audacity. This duration was selected to maximize potential familiarity with each talker's voice, as talker recognition asymptotes for speech samples beyond this duration (Schweinberger et al., 1997).

For the famous talkers, speech excerpts from the last five Presidents of the USA (Joe Biden, Donald Trump, Barack Obama, George W. Bush, Bill Clinton) were presented. Given that Barack Obama held the office of President the third-most recently at the time of testing (i.e., intermediate

---

[2] In Magnuson et al. (2021), the third experiment also comprised a talker identification task, comparing the accuracy and reaction times with which participants identified familiar versus trained-on voices. Reaction times were provided and were in the predicted directions (familiar < trained-on), but these data were not statistically analyzed.

Table 1 Acoustic measurements of the experimental stimuli ("do" and "to" spoken by each talker) for Experiments 1 and 2. Values are means with the standard deviations in parentheses. Measurements are organized in rows according to the four experimental conditions

Condition	Grand mean f0 (Hz)	Mean stimulus duration (ms)
Single famous (Obama)	143.63 (4.37)	121.25 (31.78)
Mixed famous	154.17 (13.52)	179.60 (43.78)
Single non-famous (Pompeo)	100.16 (2.79)	211.84 (16.42)
Mixed non-famous	126.85 (35.68)	208.65 (61.03)

relative to the other Presidents), he was presented as the single talker; all other talkers were presented in the mixed-talker block. For the non-famous talkers, comparatively lesser-known politicians in roughly the same age range (57–71 years, at the time of testing) as the famous talkers (60–78 years, at the time of testing) were selected: Antony Blinken, Ashton Carter, James Mattis, David Petraeus, and Mike Pompeo. Mike Pompeo was randomly selected and presented as the single talker; all others were presented in the mixed-talker block.

Procedure

All experiments were conducted online using the Gorilla experiment builder (Anwyl-Irvine et al., 2020). After providing informed consent, participants began with a six-trial headphone screen (Woods et al., 2017). On each trial, participants heard three tones and were asked to identify which was the quietest. The correct answer is the tone presented at a level 6 dB lower than the other two tones, which is easily detectable over headphones. Over loudspeakers, a foil tone is heard as being lower in level (due to being presented out-of-phase over speakers, leading to destructive interference). Twenty of the 45 participants failed the headphone screen twice. However, in the statistical analyses reported below, performance did not significantly vary as a function of whether participants passed or failed the headphone screen, so all participants at this stage were retained. The headphone check was utilized as a means of standardizing stimulus presentation in an online experiment, and none of the sounds in the main experiment were dependent on headphone presentation.

In the main experiment, on each trial, participants heard either the word "do" or "to" at a comfortable listening level set by the participant and were asked to press a key corresponding to which word they heard. The experiment consisted of four blocks: a single famous talker, mixed famous talkers, a single non-famous talker, and mixed non-famous talkers. Blocks were presented in random orders across participants, and the 80 trials in each block were tested in random orders as well. After the four blocks were completed, participants

completed a short post-task questionnaire to gauge their familiarity with the talkers. Talkers were presented in random orders, but for each talker, three questions were asked in the same order each time. First, participants heard that talker say "do" and "to" again and were asked to type in the talker's name, or type "x" if they were unsure who the talker was. Second, participants heard one sentence from that talker and were again asked to type in the talker's name or an "x" if they were unsure. Third, participants were shown the name of the talker and were asked to answer "yes" or "no" if they had heard of the talker. Finally, participants were asked their age and to rate their political interest on a scale from 1 (not interested at all) to 10 (very interested); this question was taken from Maibauer et al. (2014).

Results

An inclusion criterion of 80% correct in a given block was implemented for data to be included in statistical analyses (n.b., the same patterns of results are obtained when implementing inclusion criteria at 70% correct or 90% correct). This resulted in the removal of 23 out of 192 blocks, 12 of which comprised all responses from three participants who failed to meet this performance criterion in any of the four experimental blocks. Accordingly, these three participants' responses in the post-task questionnaire were removed from analyses as well.

Accuracy was near ceiling, with block means ranging from 94% to 97% correct. Accuracy was not analyzed; instead, response times were the primary variable of interest (see Fig. 1). As is customary for reaction time analyses in speeded word recognition tasks (e.g., Choi et al., 2018; Stilp & Theodore, 2020), only correct responses were retained (removing 542 trials, or 4.01% of the total data). Additionally, all response times faster than 200 ms were removed, as these responses were too short to reflect the time needed to hear a stimulus and plan a corresponding motor response (e.g., Welford, 1980; removing 13 trials, or 0.10% of the remaining data). Distributions of reaction times were positively skewed, so they were $\log_{10}$-transformed to achieve normality. Finally, mean reaction time was calculated for each participant, and reaction times exceeding three standard deviations from that listener's mean were removed (removing 195 trials, or 0.76% of the remaining data).

Linear mixed-effects modeling was used to predict trial-level reaction times. Fixed effects in this model included target word duration (mean-centered), block (sum-coded, single talker coded as -0.5 and mixed talker coded as +0.5), familiarity (sum-coded; famous coded as -0.5 and non-famous coded as +0.5), and all interactions. The model building process began with a base model with these fixed effects and random intercepts for participants. Random effects were added one at a time and tested via χ^2 goodness-of-fit tests. If the added term explained significantly more variance, it was retained. The final random effects structure was maximal (Barr et al.,
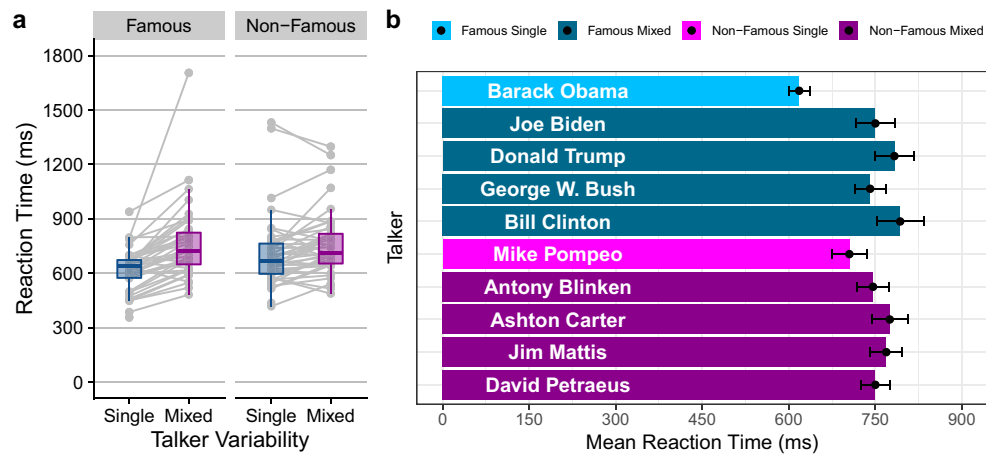
Fig. 1 Results from Experiment 1. **a** Each grey dot connected by a line represents response times for an individual participant; boxplots displaying condition medians are overlaid. **b** Mean response times are listed for each talker, arranged by condition. Bar length and dots indicate means; error bars represent 1 standard error of the mean. (Color online)

2013), including random slopes for block, familiarity, as well as their interaction, with random intercepts for participants, and random intercepts for talkers. Results from the final model are presented in Table 2.

Durations varied across the target words (Table 1), and participants were quicker to respond when the duration was shorter (negative main effect of duration). As predicted, participants responded more quickly in the single talker block (positive main effect of block; estimated mean response times of 592 ms for single talkers and 704 ms for mixed talkers). Also as predicted, participants responded more quickly to famous talkers (positive main effect of familiarity; estimated mean response times of 610 ms for famous talkers and 682 ms for non-famous talkers). The interaction between duration and block was significant. Estimates from the package ggpredict (Lüdecke, 2022) indicated that effects of stimulus duration were stronger in the single-talker block: as stimulus duration decreased, response times decreased more quickly in the single-talker block than in the mixed-talker

block. The interaction between duration and familiarity was also significant. Estimates from ggpredict indicated that as stimulus duration decreased, response times decreased for famous talkers but increased modestly for non-famous talkers. Most importantly, the interaction between block and familiarity was significant. However, contrary to predictions, the sign on this interaction was negative. The talker consistency effect (difference in response times across single-talker and mixed-talker blocks) was predicted to be smaller for famous talkers than for non-famous talkers, but it was larger for famous talkers instead. All means were calculated using the emmeans package (Lenth, 2019).

Results from the post-experiment questionnaire are given in Table 3. For each participant, percent correct recognition was averaged across famous or non-famous talkers and analyzed in a 2 (familiarity: famous, non-famous) × 3 (task: word, sentence, name) repeated-measures ANOVA.[3] There was a significant main effect of familiarity, with more accurate recognition of famous talkers ($M = 52.60\%$, $SE = 2.99$) than non-famous talkers ($M = 5.78\%$, $SE = 1.44$; $F(1, 44) = 437.88$, $p < .0001$; $\eta^2 = 0.69$). There was a significant main effect of task ($F(1.43, 62.94) = 207.72$, $p < .0001$; $\eta^2 = 0.59$). Post hoc paired-samples t-tests with Bonferonni corrections indicated

Table 2 Results from the mixed-effects model analysis of responses in Experiment 1. As described in the text, duration was entered in the model as a continuous factor, centered around the mean. Block and familiarity were sum-coded; the level associated with the -0.5 contrast for each factor is shown in parentheses

	$\widehat{\beta}$	SE	t	p
Intercept	2.810	0.015	181.294	<0.001
Duration	-0.291	0.067	-4.325	<0.001
Block (Single)	0.075	0.016	4.632	0.010
Familiarity (Famous)	0.048	0.017	2.929	0.040
Duration × Block	0.402	0.135	2.981	0.005
Duration × Familiarity	0.480	0.135	3.564	0.001
Block × Familiarity	-0.101	0.034	-2.977	0.034
Duration × Block × Familiarity	-0.007	0.269	-0.027	0.979

---

[3] Responses in the post-experiment questionnaires were analyzed using repeated-measures ANOVA for the sake of simplicity. Accuracy scores were discrete (ranging from 0% to 100% correct in steps of 20%) and distributed differently based on talker familiarity (especially with zero recognition of non-famous talkers from their words or sentences). However, several alternative statistical analyses were not viable. Logistic regressions predicting response accuracy were singular fits given the absence of accuracy (i.e., 1 s in the data) in non-famous talker conditions. While the Friedman test is a nonparametric alternative to a repeated-measures ANOVA, it is not equipped to analyze the present 2 × 3 experimental design. Supplementary analyses employing Wilcoxon matched-pair sign-rank tests with Bonferroni correction for multiple comparisons (dividing α =.05 by 7 planned comparisons = .0071) were conducted and are detailed in the analysis scripts (see Open Practices Statement). Results from the supplementary analyses are all consistent with those detailed in the main text.

Table 3 Results from the post-experiment questionnaires at the end of each experiment, organized by familiarity condition. Values indicate the mean percent correct talker recognition from hearing the "do" and "to" word stimuli again (Experiments 1 and 3; columns 3 and 4), hearing a full sentence (Experiments 1 and 3; columns 5 and 6), or being shown the talker's name and responding whether participants recognized the name of the talker (all experiments; columns 7–9)

Talker	Familiarity	% Words		% Sentence		% Name		
		Exp. 1	Exp. 3	Exp. 1	Exp. 3	Exp. 1	Exp. 2	Exp. 3
Joe Biden	Famous	13.3	4.6	64.4	65.9	95.5	100	95.5
George W. Bush	Famous	4.4	2.3	11.1	15.9	88.8	95.2	84.1
Bill Clinton	Famous	2.2	0.0	2.2	4.55	80	95.2	81.8
Barack Obama	Famous	13.3	13.6	71.1	68.2	93.3	95.2	93.2
Donald Trump	Famous	55.5	34.1	93.3	86.4	100	100	97.7
Antony Blinken	Non-Famous	0	0	0	0	11.1	23.8	9.1
Ashton Carter	Non-Famous	0	0	0	0	11.1	26.2	9.1
James Mattis	Non-Famous	0	0	0	0	15.5	23.8	4.6
David Petraeus	Non-Famous	0	0	0	0	13.3	21.4	6.8
Mike Pompeo	Non-Famous	0	0	0	0	35.5	45.2	25

that all tasks elicited significantly different levels of performance (all $t > 10.00$, all $p < .001$), with lowest recognition from words ($M = 8.89\%$, $SE = 1.55$), better recognition from sentences ($M = 24.20\%$, $SE = 2.96$), and highest recognition when shown the talker's name ($M = 54.40\%$, $SE = 4.49$). Finally, the interaction between familiarity and task was significant ($F(1.51, 66.56) = 83.79$, $p < .001$; $\eta^2 = 0.35$). Post hoc paired-samples t-tests with Bonferonni corrections were computed to assess effects of familiarity in each task. Famous talkers were recognized more accurately from hearing them say "do" and "to" ($M = 17.80\%$, $SE = 2.48$) than were non-famous talkers (mean = 0.00%, $SE = 0$; $t(44) = 7.17$, $p < .001$). Famous talkers were also recognized more accurately from hearing them speak a sentence ($M = 48.40\%$, $SE = 2.95$) than were non-famous talkers ($M = 0.00\%$, $SE = 0$; $t(44) = 16.44$, $p < .001$). Unlike the other tasks, participants sometimes recognized the non-famous talkers when shown their names ($M = 17.30\%$, $SE = 3.79$), but famous talkers were still recognized more regularly ($M = 91.60\%$, $SE = 2.15$; $t(44) = 16.92$, $p < .001$).

Correlations were assessed between self-rated political interest ($M = 4.71$, $SD = 2.37$) and performance in each experimental condition and post-task questionnaire. Political interest was positively correlated with accuracy in the single-talker non-famous condition (Spearman's rho: $\rho = 0.32$, $p = .037$), but was not correlated with performance in any other experimental conditions nor in the post-test survey items.

Discussion

Three predictions were made for Experiment 1. First, reaction times for single talkers were predicted to be faster than for mixed talkers, consistent with a host of studies on talker adaptation (see *Introduction*). Results confirmed this prediction,

with listeners responding an estimated 112 ms faster to the single talker compared to the mixed talkers. Second, reaction times for famous talkers were predicted to be faster than for non-famous talkers (as reported in Maibauer et al., 2014, and as predicted in Magnuson et al., 2021). Results also confirmed this prediction, with listeners responding an estimated 72 ms faster for famous talkers compared to non-famous talkers. Third and most importantly, familiarity was predicted to interact with talker consistency such that the consistency benefits (shorter reaction times to a single talker than to mixed talkers) would be smaller for famous talkers than for non-famous talkers. Contrary to this prediction, talker consistency benefits were larger for famous talkers than the non-famous talkers, indicating responses slowing by a greater degree when hearing variable famous talkers.

Manipulations of talker familiarity in Experiment 1 differed from those in past studies. Many talker familiarity studies used stimuli from voices that were initially unfamiliar but achieved familiarity through training (e.g., Holmes et al., 2021; Magnuson et al., 2021; Nygaard et al., 1994; Nygaard & Pisoni, 1998). The present experiment assumed previous familiarity with familiar talkers, as they are famous figures in the USA (where the study was conducted). This essentially equated fame with familiarity, which is not always an accurate assumption. One could easily envision different levels of personal familiarity with voices that are all justifiably famous (e.g., college undergraduates hearing the voices of TikTok stars of the 2020s vs. movie stars of the 1940s). It is possible that the present stimuli were not sufficiently familiar in order to augment the benefits of listening to a consistent talker. It bears mention that Magnuson and colleagues (2021) also failed to observe consistent benefits of talker familiarity with voices from family members, who are objectively more

familiar to those participants than the politicians' voices tested here. Regarding unfamiliar voices, the non-famous talkers tested here were more familiar to participants than novel talkers used in other experiments; participants achieved at least some recognition of these talkers when showed their names (Table 3). The primary distinction for these stimuli is between famous versus non-famous talkers, which is not the same as familiar versus entirely unfamiliar (i.e., novel) talkers. Recognition of non-famous talkers from audio excerpts was extremely poor overall (Table 3). Thus, the distinction between non-famous and entirely novel appears unlikely to have contributed to the present results.

Stimuli were excised from running speech, prioritizing naturalness and intelligibility without regard to their durations. With such short speech segments (mean duration = 188.61 ms), it can be difficult for participants to recognize the voices of the talkers (Schweinberger et al., 1997). However, the benefits of talker familiarity can still be gained even if the participant does not explicitly recognize the voice (Holmes et al., 2018). In Holmes et al. (2018), talker familiarity increased sentence intelligibility when hearing competing familiar and unfamiliar talkers. From those findings, one would predict that the benefit of talker familiarity might be evident amidst talker variability, even if the participants cannot clearly recognize the familiar talkers they hear. Additionally, while duration varied across stimuli, the statistical analyses of responses controlled for this variability by explicitly testing it as a fixed effect. Responses were faster to shorter stimuli overall, but this partialled stimulus duration information out from assessing the main effects of talker consistency, talker familiarity, and their interaction.

A final consideration for Experiment 1 is that of stimulus recency effects. In Magnuson and colleagues (2021) and other studies measuring talker familiarity, hearing family members and/or trained-on voices gives participants the benefit of possibly hearing those voices recently before, if not the same day as, the experiment. Thus, these stimuli are not only familiar but also recent. Conversely, Experiment 1 had no control over how recently participants heard the talkers prior to the experiment. This makes it difficult to control any recency effect, as participants could have heard the talkers' voices any number of days, months, or years ago. Regardless, however recently participants heard the familiar talkers' voices, neither this experiment nor those reported in Magnuson et al. (2021) observed benefits of talker familiarity that augmented the perceptual benefits from listening to a consistent talker.

Maibauer and colleagues (2014) suggested that giving participants a chance to learn the famous talkers' voices before the experiment begins might maximize familiarity benefits. To this end, Experiment 2 primed the participants on the talkers they would hear before the task began. The post-task questionnaire was moved before the main experiment, allowing participants to hear all talkers' voices and learn their names to maximize the potential benefit of familiarity. This also ensured that all talkers' voices were heard recently before the experiment began. All three predictions from Experiment 1 held in Experiment 2, now with the benefits of talker familiarity predicted to be more evident due to this priming.

Experiment 2

Methods

Participants

Forty-two undergraduate students from the University of Louisville participated in Experiment 2; none had participated in Experiment 1. The mean age of participants was 20.79 years (SD = 5.87). All participants were self-reported native English speakers with no known hearing deficits. Participants received course credit in exchange for their participation.

Stimuli

Experiment 2 used the same stimuli from Experiment 1.

Procedure

Participants gave informed consent and then completed the same headphone screen as in Experiment 1. Seven participants failed the headphone screen, but their data were retained as the pattern of results was highly similar with and without their inclusion. Next, the talker recognition portion of the post-task questionnaire from Experiment 1 was presented to the participants as exposure prior to the main task of Experiment 2. For each talker, participants heard the talkers say "do" and "to" with the talker's name on the screen; no response was required. Then, participants heard the same talker speak a sentence with his name on the screen, again with no response required. Finally, participants saw the talker's name on the screen and replied whether they recognized the name of the talker. After participants heard all ten talkers, they completed the same four blocks as in Experiment 1 in random order. Finally, participants were asked to give their age and rate their political interest.

Results

An inclusion criterion of 80% correct in a given block was again implemented for data to be included in statistical analyses. This resulted in the removal of eight out of 168 blocks. All conclusions reported below hold if alternative inclusionary criteria (70% correct, 90% correct) are employed.

Mean accuracy in each block was again extremely high (M = 96–98%), but response times were still the primary variable of interest (Fig. 2). Following the same protocol as
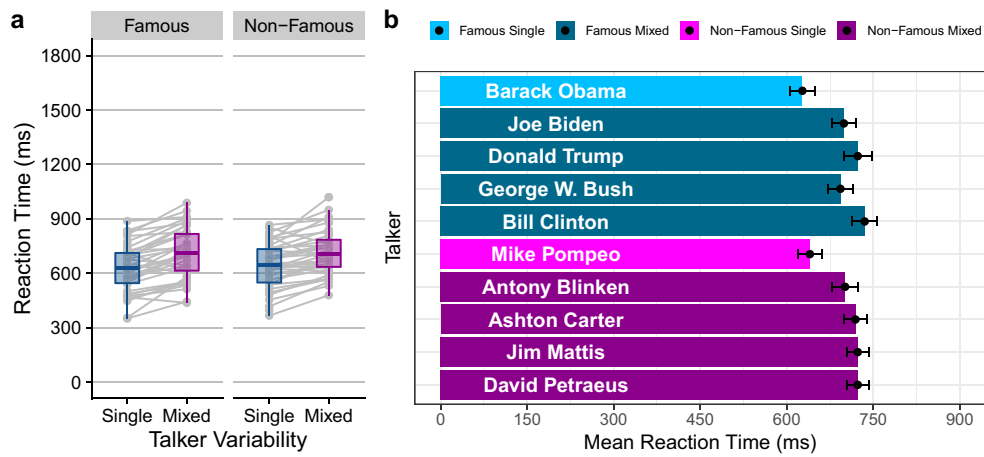
Fig. 2 Results from Experiment 2. **a** Each grey dot connected by a line represents response times for an individual participant; boxplots displaying condition medians are overlaid. **b** Mean response times are listed for each talker, arranged by condition. Bar length and dots indicate means; error bars represent 1 standard error of the mean. (Color online)

Experiment 1, only correct responses were retained (removing 426 trials, or 3.33% of the total data), all response times shorter than 200 ms were removed (removing six trials, or 0.05% of the remaining data), reaction times were $\log_{10}$-transformed to achieve normality, and reaction times exceeding three standard deviations from each listener's mean reaction time were removed (removing 178 trials, or 0.72% of the remaining data).

The approach to model building was exactly the same as that of Experiment 1. The final model architecture matched that reported for Experiment 1: the random effects included random slopes for block, familiarity, as well as their interaction, with random intercepts for participant, and random intercepts for talkers. Results from the final model are presented in Table 4.

Participants were again quicker to respond when stimulus duration was shorter (negative main effect of duration). As predicted, participants were faster to respond in the single-talker block (positive main effect of block; estimated mean response

times of 596 to single talkers and 682 ms to mixed talkers). Contrary to the prediction and the results of Experiment 1, reaction times did not significantly differ for famous (estimated mean response time = 615 ms) and non-famous talkers (estimated mean response time = 661 ms). The interaction between duration and block was significant. Estimates from ggpredict indicated faster response times for shorter stimulus durations in both levels of condition, with the relationship slightly stronger in the multiple-talker block. The interaction between duration and familiarity was also significant. Estimates from ggpredict indicated that shorter word durations from famous talkers led to faster response times, but there was no relationship between word duration and response times for non-famous talkers. Most importantly, contrary to the prediction and the results of Experiment 1, the interaction between block and familiarity was not significant; the change in reaction time across single-talker and mixed-talker blocks did not differ for famous versus non-famous talkers. All means were calculated using the emmeans package.

To compare patterns of performance across experiments, a linear mixed-effects regression was conducted with the fixed main effect of Experiment (sum-coded; Experiment 1 coded as -0.5, Experiment 2 coded as +0.5) and its interactions with other fixed effects added; the random effects structure matched that in other analyses. Results from the final model are presented in Table 5. Interactions including the fixed effect of experiment were of primary interest. The interaction between familiarity and experiment trended toward significance. Its negative coefficient, indicating that familiarity had a modestly smaller effect in Experiment 2, is consistent with familiarity being a significant influence on response times in Experiment 1 (Table 2) but not in Experiment 2 (Table 4), and with the main effect of familiarity being nonsignificant when combining results across experiments (Table 5). For the interaction between block and experiment, the negative coefficient suggests that the effect of single versus mixed talkers

Table 4 Results from the mixed-effects model analyses of responses in Experiment 2. Duration was entered in the model as a continuous factor, centered around the mean. Following Table 2, block and familiarity were sum-coded; the level associated with the -0.5 contrast for each factor is shown in parentheses

	$\widehat{\beta}$	SE	t	p
Intercept	2.805	0.016	171.783	<0.001
Duration	-0.402	0.064	-6.240	<0.001
Block (Single)	0.058	0.020	2.929	0.028
Familiarity (Famous)	0.031	0.019	1.619	0.166
Duration × Block	0.334	0.129	2.590	0.012
Duration × Familiarity	0.276	0.129	2.141	0.037
Block × Familiarity	-0.069	0.039	-1.774	0.135
Duration × Block × Familiarity	0.407	0.258	1.579	0.120

trended towards being smaller in Experiment 2 than Experiment 1 (though still significantly influencing responses in each experiment). Finally, the three-way interaction between block, familiarity, and experiment was not statistically significant. This indicates that the benefits of talker familiarity when listening to a consistent talker did not materially differ across experiments (despite the familiarity-by-block interaction being significant in Experiment 1 but nonsignificant in Experiment 2).

Paired-samples t-tests assessed participants' recognition of famous and non-famous talkers when the talkers' names were displayed. As in Experiment 1, famous talkers' names ($M = 97.1\%$, $SE = 1.09$) were recognized more often than non-famous talkers' names ($M = 28.10\%$, $SE = 4.96$; $t(41) = 12.95$, $p < .001$). Finally, self-rated political interest ($M = 5.15$, $SD = 2.32$) was not correlated with performance in any condition.

Discussion

Experiment 2 exposed participants to the talkers' voices and names prior to the main experiment. This manipulation was suggested by Maibauer et al. (2014) to maximize the benefits of familiarity for the famous talkers' voices, which in turn was predicted here to augment the benefits (i.e., decrease response times) for familiar talkers in the speeded word recognition task. However, by making all talkers' voices recently heard before completing the speeded word recognition task, effects of talker familiarity were extinguished. Response times to famous and non-famous talkers did not significantly differ,

and the interaction between familiarity and block was a non-significant trend. Given that words spoken by famous talkers were recognized more quickly than those spoken by non-famous talkers in Experiment 1, this manipulation appears to have introduced recency effects for all talkers, which diminished the perceptual benefits of familiarity. An alternative option would have been to expose participants to only the famous talkers' voices before the speeded word recognition task, but the main effect of familiarity reported in Experiment 1 without such exposure implied that separating the stimuli in this manner was not necessary.

Across-experiment analyses revealed the different effects of familiarity across participant samples (and concurrently, across the presence vs. absence of the exposure period before the speeded recognition task). Main effects of interest trended toward being weaker in Experiment 2 than in Experiment 1: the benefits of talker consistency (negative block × experiment interaction; $p = .068$) and the benefits of talker familiarity (negative familiarity × experiment interaction; $p = .058$; Table 5). These trends are suggestive of recency effects in the exposure weakening the perceptual benefits of talker consistency and of talker familiarity, both of which were strong in Experiment 1.

The designs of Experiments 1 and 2 were unbalanced, such that participants heard only one talker in each single-talker block (Barack Obama in the famous block, Mike Pompeo in the non-famous block). Those talkers were assigned to the single-talker conditions by experimenter choice (in Obama's case based on a presumed intermediate amount of familiarity; in Pompeo's case based on random assignment). To account

Table 5 Results from mixed-effects model analyses across experiments. Duration was again centered around the mean. Block, familiarity, and experiment were sum-coded; the level associated with the -0.5 contrast for each factor is shown in parentheses

	$\widehat{\beta}$	SE	t	p
Intercept	2.807	0.013	212.328	<0.001
Duration	-0.345	0.053	-6.534	<0.001
Block (Single)	0.066	0.019	3.512	0.016
Familiarity (Famous)	0.039	0.019	2.085	0.090
Experiment (Exp. 1)	-0.006	0.019	-0.314	0.754
Duration × Block	0.371	0.106	3.509	0.001
Duration × Familiarity	0.444	0.106	4.198	<0.001
Block × Familiarity	-0.087	0.038	-2.308	0.066
Duration × Experiment	-0.047	0.067	-0.704	0.481
Block × Experiment	-0.018	0.010	-1.837	0.068
Familiarity × Experiment	-0.017	0.009	-1.911	0.058
Duration × Block × Familiarity	0.332	0.211	1.569	0.121
Duration × Block × Experiment	0.061	0.134	0.452	0.652
Duration × Familiarity × Experiment	-0.247	0.134	-1.841	0.066
Block × Familiarity × Experiment	0.029	0.020	1.470	0.144
Duration × Block × Familiarity × Experiment	0.330	0.268	1.231	0.218

for these talkers (and their tokens of "do" and "to") always appearing in the single-talker condition, statistical analyses included fixed effects of stimulus duration and random intercepts for talkers. Yet, it remains possible that other factors could be influencing the present results. For example, while Barack Obama served as President third-most-recently, participants may have other bases by which his voice is more familiar than other presidents' voices, such as social media presence. Mike Pompeo was randomly assigned to the single non-famous talker, but he was the best recognized (by name) of the non-famous talkers (Table 3), potentially weakening the gap in familiarity between famous and non-famous talkers. Additionally, given the political nature of these talkers, voice familiarity might be influenced by voting patterns. While the state of Kentucky on the whole voted primarily for Republican presidential candidates in the 2016 and 2020 elections (both Donald Trump), Jefferson County (which includes the University of Louisville, where testing took place) voted primarily for Democratic presidential candidates (54.1% of the vote for Hilary Clinton in 2016; 59.1% for Joe Biden in 2020). Participants were not asked to report their political affiliations nor voting patterns, but this might also make voices from one political party more familiar than the other. These and any number of other factors could potentially affect task performance when the experimental design is imbalanced. To account for these external factors, Experiment 3 utilized the same experimental paradigm as Experiment 1 but with a balanced experimental paradigm. Across the participant sample, each famous talker appeared in the single-talker block and the mixed-talker block; the same was true for each non-famous talker.

Experiment 3

Methods

Participants

Forty-four undergraduate students from the University of Louisville participated in Experiment 3. Of the 42 participants who disclosed their age, mean age was 19.31 years ($SD = 1.93$). One additional participant was tested but their results were not included in analyses, as detailed below. All participants were self-reported native English speakers with no known hearing deficits. Participants received course credit in exchange for their participation.

Stimuli

Stimuli were the same as those detailed in previous experiments.

Procedure

The procedure was largely the same as that detailed in Experiment 1. Twelve of the 44 participants failed the headphone screen twice, but all were included in the statistical analyses.[4]

The main experiment was again comprised of two levels of familiarity (famous, non-famous) fully crossed with two levels of talker variability (single, mixed). The primary difference between the procedures of Experiments 1 and 3 was that in Experiment 3, listeners could be assigned to have a different pairing of famous and non-famous talkers serving as the single talkers. Thus, each experiment block from Experiment 1 was quadruplicated in order for each of the five talkers at a given level of familiarity to serve in the single-talker block (and, therefore, be excluded from the related mixed-talker block). To keep the total number of versions of the experiment manageable, each famous talker was yoked with a specific non-famous talker in their respective single-talker blocks. These yokes, determined by random assignment, were as follows: Joe Biden – Jim Mattis, George Bush – Mike Pompeo, Bill Clinton – Antony Blinken, Barack Obama – David Petraeus, Donald Trump – Ashton Carter. Participants were evenly distributed across these five pairings, hearing each of these listed talkers in the single-talker blocks and all remaining famous or non-famous talkers in mixed-talker blocks.

Blocks were presented in random orders, and the 80 trials in each block were again tested in random orders. Participants then completed the same post-task questionnaire as detailed in Experiment 1.

Results

An inclusion criterion of 80% correct in a given block was again implemented for data to be included in statistical analyses. This resulted in the removal of ten out of 180 blocks. All conclusions reported below held if alternative inclusionary criteria (70% correct, 90% correct) were employed.

---

[4] Mixed-effects model analyses revealed differences in performance according to whether listeners passed or failed the headphone screen. Participants who passed the headphone screen ($n = 32$) exhibited a significant intercept, main effect of block, duration-by-block-by-familiarity interaction (all consistent with the full-group analysis in Table 6), and a negative block-by-familiarity interaction (larger difference in response times across single-talker and mixed-talker blocks when talkers were famous). Participants who failed the headphone screen ($n = 12$) exhibited those same significant effects as well as a positive main effect of duration, a negative duration-by-block interaction, and a positive block-by-familiarity interaction (larger difference in response times across single-talker and mixed-talker blocks when talkers were non-famous). However, this analysis necessitated a simpler random effects structure, as singular model fits required the removal of random slopes for the block-by-familiarity interaction by subject. No reason is apparent for these different patterns of results. Rather than reporting only results for those who passed the headphone screen (which would be largely consistent with Table 6), we report results for all listeners and post these supplementary analyses on our Open Science Framework page (see Open Practices Statement for link).

Mean accuracy in each block was again extremely high ($M = 95$–97%), but response times were still the primary variable of interest (Fig. 3). Following the same protocol as previous experiments, incorrect responses were removed (501 trials, or 3.68% of the total data), all response times shorter than 200 ms were removed (removing six trials, or 0.05% of the remaining data), reaction times were log-transformed to achieve normality, and reaction times exceeding three standard deviations from each listener's mean reaction time were removed (removing 202 trials, or 0.78% of the remaining data).

The model building process and final model architecture matched that of the other experiments: the random effects included random slopes for block, familiarity, as well as their interaction, with random intercepts for participant, and random intercepts for talkers. Results from the final model are presented in Table 6.

Unlike previous experiments, participants did not respond more quickly to shorter-duration stimuli (nonsignificant main effect of duration). As predicted, participants were faster to respond in the single-talker block (positive main effect of block; estimated mean response times of 625 to single talkers and 693 ms to mixed talkers). Contrary to predictions, reaction times did not significantly differ for famous (estimated mean response time = 655 ms) versus non-famous talkers (estimated mean response time = 661 ms), nor was the interaction between block and familiarity statistically significant. The only other model term to reach statistical significance was the three-way interaction between duration, block, and familiarity. Estimates from ggpredict indicate slightly faster responses for shorter-duration stimuli in the Single Famous block, and slightly slower responses to shorter-duration stimuli in the Mixed Famous and Single Non-Famous blocks. All means were calculated using the emmeans package.

To compare performance across experimental paradigms, a linear mixed-effects regression was conducted with the fixed main effect of Experiment (sum-coded; Experiment 1 coded as -0.5, Experiment 3 coded as +0.5) and its interactions with other fixed effects added; the random effects structure matched that in other analyses. Results from the final model are presented in Table 7. Interactions including the fixed effect of experiment revealed a host of effects being smaller or absent in Experiment 3 compared to Experiment 1. The negative experiment-by-block and experiment-by-familiarity interactions reveal smaller effects of each in Experiment 3, whereas the positive interaction between experiment, block, and familiarity reflects the change from a significant negative block-by-familiarity interaction in Experiment 1 to a null interaction in Experiment 3. Additionally, contributions of stimulus duration differed markedly across experiments; sensitivity to stimulus duration (i.e., responding more quickly to shorter-duration stimuli) in Experiment 1 was not replicated in Experiment 3.

Results from the post-experiment questionnaire are given in Table 3. As in Experiment 1, for each participant, percent correct recognition was averaged across famous or non-famous talkers and analyzed in a 2 (familiarity: famous, non-famous) × 3 (task: word, sentence, name) repeated measures ANOVA. There was a significant main effect of familiarity, with more accurate recognition of famous talkers ($M = 49.80\%$, $SE = 3.27$) than non-famous talkers ($M = 3.64\%$, $SE = 1.11$; $F(1, 43) = 404.42$, $p < .0001$; $\eta^2 = 0.69$). There was a significant main effect of task ($F(1.69, 72.88) = 248.47$, $p < .0001$; $\eta^2 = 0.59$). Post hoc paired-samples t-tests with Bonferonni corrections indicated that all tasks elicited significantly different levels of performance (all $t > 11.00$, all $p < .001$), with lowest recognition from words ($M = 5.45\%$, $SE = 1.40$), better recognition from sentences ($M = 24.10\%$, $SE = 3.05$), and highest recognition when shown the talker's name ($M = 50.70\%$, $SE = 4.71$). Finally, the interaction between familiarity and task was significant ($F(1.69, 72.64) =$
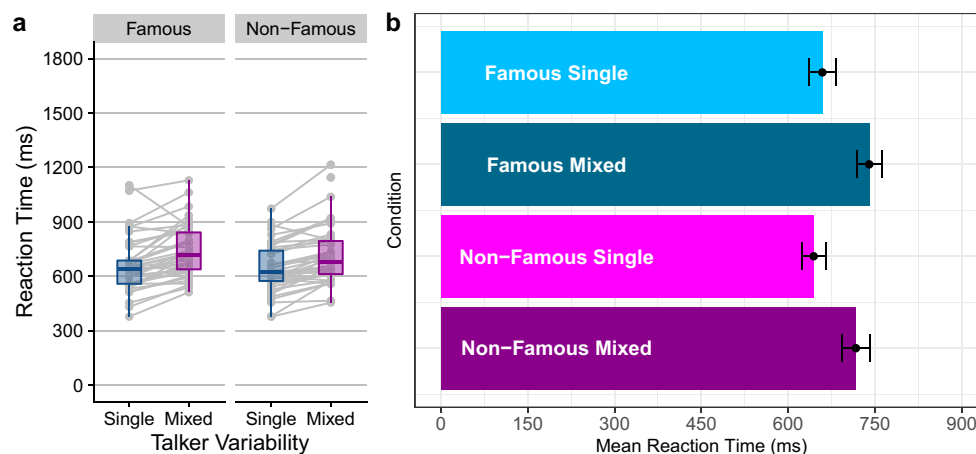

Fig. 3 Results from Experiment 3. **a** Each grey dot connected by a line represents response times for an individual participant; boxplots displaying condition medians are overlaid. **b** Mean response times are listed for each condition. Bar length and dots indicate means; error bars represent 1 standard error of the mean. (Color online)

Table 6 Results from the mixed-effects model analyses of responses in Experiment 3. Duration was entered in the model as a continuous factor, centered around the mean. Block and familiarity were sum-coded; the level associated with the -0.5 contrast for each factor is shown in parentheses

	$\widehat{\beta}$	SE	t	p
Intercept	2.818	0.014	208.460	<0.001
Duration	-0.013	0.050	-0.261	0.794
Block (Single)	0.045	0.006	8.134	<0.001
Familiarity (Famous)	-0.004	0.008	-0.528	0.597
Duration × Block	-0.032	0.072	-0.439	0.661
Duration × Familiarity	-0.077	0.098	-0.781	0.435
Block × Familiarity	-0.009	0.008	-1.104	0.269
Duration × Block × Familiarity	0.428	0.134	3.187	0.001

143.95, $p < .001$; $\eta^2 = 0.45$). Post hoc paired-samples t-tests with Bonferonni corrections were computed to assess effects of familiarity in each task. Famous talkers were recognized more accurately from hearing them say "do" and "to" ($M =$ 10.90%, $SE = 2.56$) than were non-famous talkers (mean = 0.00%, $SE = 0$; $t(43) = 4.27$, $p < .001$). Famous talkers were also recognized more accurately from hearing them speak a sentence ($M = 48.20\%$, $SE = 3.27$) than were non-famous talkers ($M = 0.00\%$, $SE = 0$; $t(43) = 14.73$, $p < .001$). As in Experiment 1, participants sometimes recognized the non-famous talkers when shown their names ($M = 10.90\%$, $SE = 3.08$), but famous talkers were still recognized more regularly ($M = 90.50\%$, $SE = 2.56$; $t(43) = 20.54$, $p < .001$).

Correlations were assessed between self-rated political interest (for the 42 participants who disclosed these ratings, $M =$ 4.04, $SD = 2.00$) and performance in each experimental condition and post-task questionnaire. Political interest was positively correlated with response times to the famous mixed talkers (Spearman's rho: $\rho = 0.37$, $p = .018$) and the non-famous mixed talkers ($\rho = 0.35$, $p = .025$), as well as accuracy for the non-famous mixed talkers ($\rho = 0.38$, $p = .013$). In the post-task questionnaire, political interest was positively correlated with accuracy for naming famous talkers from their words ($\rho = 0.37$, $p = .015$) and from their sentences ($\rho =$ 0.42, $p = .005$), and trended toward significance for recognizing famous talkers by name ($\rho = 0.30$, $p = .056$). Correlations with non-famous talkers were either incalculable (due to no one recognizing them from their words or sentences) or non-significant (recognizing them by name: $\rho = -0.11$, $p = .494$).

Discussion

Experiment 3 tested the same stimuli but an alternative design to that utilized in Experiment 1 by rotating talkers through the conditions so that each one appeared in single-talker and mixed-talker blocks. Unlike Experiment 1, Experiment 3 did not produce a significant benefit of talker familiarity (yielding faster response times for famous talkers compared to non-famous talkers) nor a significant interaction between talker consistency and talker familiarity. In all, Experiment 3 did not produce compelling evidence that talker familiarity augments talker adaptation.

Table 7 Results from mixed-effects model analyses across Experiments 1 and 3. Duration was again centered around the mean. Block, familiarity, and experiment were sum-coded; the level associated with the -0.5 contrast for each factor is shown in parentheses

	$\widehat{\beta}$	SE	t	p
Intercept	2.813	0.010	272.422	<0.001
Duration	-0.154	0.047	-3.275	0.002
Block (Single)	0.061	0.005	12.454	<0.001
Familiarity (Famous)	0.022	0.009	2.321	0.046
Experiment (Exp. 1)	0.007	0.019	0.365	0.716
Duration × Block	0.158	0.072	2.198	0.029
Duration × Familiarity	0.239	0.094	2.549	0.014
Block × Familiarity	-0.050	0.010	-4.956	<0.001
Duration × Experiment	0.266	0.064	4.155	<0.001
Block × Experiment	-0.034	0.010	-3.597	<0.001
Familiarity × Experiment	-0.051	0.009	-5.397	<0.001
Duration × Block × Familiarity	0.167	0.142	1.178	0.240
Duration × Block × Experiment	-0.506	0.129	-3.927	<0.001
Duration × Familiarity × Experiment	-0.358	0.127	-2.817	0.005
Block × Familiarity × Experiment	0.090	0.020	4.574	<0.001
Duration × Block × Familiarity × Experiment	0.598	0.255	2.344	0.019

Two additional factors might have contributed to the differing patterns of results across Experiments 1 and 3. First, different participants were tested in each experiment. While this fact is not surprising, it does have implications for the effectiveness of the familiarity manipulations. Familiarity with the talkers' voices was presumed rather than experimentally controlled. It is possible that the different participant samples (as well as those tested in Experiment 2) had variable amounts of experience hearing these talkers' voices, as suggested by the correlational analyses between self-rated political interest and task performance. While listeners in Experiment 1 rated their political interest as higher (M = 4.71) than listeners in Experiment 3 (M = 4.04), it was the latter group whose ratings were correlated with several aspects of task performance. Higher political interest being positively correlated with higher task accuracy or talker recognition is intuitive. Interestingly, higher political interest was also correlated with slower responses in mixed-talker blocks. This is in the opposite direction of the predicted benefit of talker familiarity (although it bears repetition that no significant overall familiarity benefit was observed in Experiment 3). Participants in Experiment 3 exhibited the strongest relationship between self-related political interest and task performance relative to participants in other experiments reported here or in Maibauer et al. (2014). While the exact reason or mechanism underlying this finding is unclear, it speaks to potentially different roles that latent familiarity with talkers' voices can play in different listener samples.

Second, Experiments 1 and 3 were conducted approximately ten months apart. Because familiarity was presumed and not experimentally controlled, that interval could alter the baseline familiarity (or 'famousness') of the talkers' voices. For example, since Experiment 1 was conducted, Joe Biden served as President of the USA for an additional 10 months. This gave Experiment 3 participants more opportunities to become familiar with his voice relative to Experiment 1 participants. Conversely, Donald Trump (as well as the other Presidents) was 10 months further removed from his term as President, potentially making his voice relatively less familiar for Experiment 3 participants. This also applies to the non-famous talkers who either remained in political office during this interval (Antony Blinken, Ashton Carter) or were further removed from their time in office (Mike Pompeo, David Petraeus, Jim Mattis). While listeners' total experience hearing these talkers was not controlled in either experiment, there might have been unequal opportunities to sample these voices for each listener sample. By leveraging the listeners' inherent familiarity (or lack thereof) with the talkers' voices, the times of testing (Experiment 1: October–November 2021, Experiment 2: January–March 2022, Experiment 3: September 2022) and intervals between testing become particularly salient factors for interpreting task performance.

General discussion

As reviewed in the *Introduction*, talker consistency leads to faster and/or more accurate speech perception. In separate research, talker familiarity can provide a variety of benefits to speech perception. The purpose of this study was to examine if talker familiarity could augment the perceptual benefits yielded by listening to a single talker. Three experiments were conducted using recent Presidents of the USA as famous talkers and less-famous American politicians as non-famous talkers. Response times in a speeded word recognition task were collected in conditions where the talker was consistent (single talker) or varied (mixed talker). The first prediction was confirmed, as participants recognized words more quickly when spoken by a single talker in a block than multiple talkers. The second prediction was partially supported, as responses were faster for famous voices than non-famous voices in Experiment 1 but not in Experiments 2 or 3. Critically, the third prediction, that talker familiarity would interact with talker consistency, was not supported in any experiment. In Experiment 1, responses slowed down more across single- and multiple-talker blocks for famous voices than for non-famous voices, which is the opposite direction of what was predicted. In Experiments 2 and 3, the interaction was not statistically significant, suggesting comparable slowing of responses for famous and non-famous voices. Thus, these data do not support an additional benefit of familiarity to speech perception amidst the benefits from listening to a consistent (single) talker.

While familiarity can result in faster responses to target words (Experiment 1 here; shadowing task of Maibauer et al., 2014), it has yet to specifically augment the perceptual benefits produced by listening to a single talker. Magnuson et al. (2021) also observed talker familiarity failing to interact with (and provide perceptual resilience to) talker variability in their Experiment 1. In their Experiment 3, results trended in the direction of more familiar (family members, trained-on) talkers resulting in more accurate transcription than unfamiliar (exposed-to) talkers, but the interaction between familiarity and variability (i.e., single- vs. mixed-talker blocks) was not statistically significant. When viewing the data in this study alongside those of previous experiments, the null results make it difficult to draw firm conclusions about how the benefit of talker familiarity might augment the benefits from listening to a single talker.

Historically, speech perception amidst talker variability has been interpreted using various viewpoints, some of which are diametrically opposed. On the one hand, extreme versions of talker normalization advocate for "throwing out" extraneous acoustic details of speech in the service of retaining its lexical content; on the other hand, episodic or exemplar models of speech perception propose that all acoustic details are retained (see Magnuson et al., 2021, for a detailed discussion of these approaches). These extreme cases do not appear to be tenable.

In the case of "pure" normalization, data showing sensitivity to acoustic details of the speech signal are not entirely discarded in the service of processing the message (e.g., Assgari et al., 2019; Goldinger, 1996; Stilp & Theodore, 2020). Further, strict normalization accounts appear incompatible with talker familiarity benefitting speech perception at all. In the cases of episodic and exemplar models, unlimited memory or storage is assumed for exemplars (in order to retain detailed tokens from all previously heard talkers), and the roles of forgetting, memory decay, and/or imperfect task performance are often underspecified. Neither of these extreme accounts fully explains perception of talker-specific (the "who") and the linguistic (the "what") contents of the speech signal. Instead, as advocated by Magnuson and colleagues, "speech perception emerges from multiple processes working in parallel on different, but not necessarily independent, aspects of the signal" (2021, p. 1857). It is well established that talker and phonetic content are interdependent (Mullennix & Pisoni, 1990; Nygaard et al., 1994; Nygaard & Pisoni, 1998). To this end, Maibauer et al. (2014) offered interpretations of their results in speeded shadowing tasks both in favor of the time-course hypothesis (that abstract signal details are processed before acoustically detailed representations) and against it. Going forward, it appears that the most productive approaches will allow for both fine-grained acoustic information (which signifies who is speaking) and lexical content (what is being said) to contribute in parallel and in interaction with each other, rather than forcing a strict dichotomy between the two (e.g., Pierrehumbert, 2016).

Speech perception is a complex operation that works on multiple levels of processing concurrently. The speeded word recognition task used here appears to be sensitive to variability in bottom-up acoustic characteristics. Experiments 1 and 2 displayed main effects of duration, with shorter-duration stimuli eliciting faster responses. Additional sensitivity to talker acoustic characteristics was demonstrated by Stilp and Theodore (2020), where response times were graded in accordance with acoustic variability: responses were fastest when hearing a single talker, slowed when hearing four talkers with an intermediate amount of variability in their mean f0s, and slowed further still when hearing four other talkers with much greater variability in their mean f0s. Conversely, talker familiarity appears to be more of a top-down process, as familiar voices are stored in long- or short-term memory (depending on the degree of familiarity). Familiarity can vary greatly, ranging from recently learned voices (e.g., Nygaard et al., 1994) to highly familiar (e.g., family members (e.g., Magnuson et al., 2021) and long-term spouses (e.g., Johnsrude et al., 2013)). The present experiments (as well as those in Magnuson et al., 2021) examined whether and how these two operations interacted and whether the perceptual benefits from talker consistency could be augmented by top-down feedback from memory (i.e., familiarity). Memory

representations might not fully resolve lower-level acoustic differences in the signal, as suggested by Magnuson and colleagues (2021), in which short-term similarity appears to exert a larger influence on perception than long-term similarity. The present results are consistent with this suggestion, as talker familiarity did not augment the benefits of talker consistency (i.e., in a significant interaction in the opposite direction of what was predicted in Experiment 1, and null results in Experiments 2 and 3). Although familiarity might not offer a further perceptual benefit to that of talker consistency, it has many benefits in other domains of speech perception (as discussed in the Introduction).

In conclusion, familiarity with talkers' voices provides a host of benefits to speech perception, but the question of whether talker familiarity augments the perceptual benefits produced by talker consistency remains an open one. It is possible that familiarity simply does not further amplify these benefits of talker consistency, although it is difficult to draw firm conclusions from null results, such as those observed here and in Magnuson et al. (2021). However, repeated observations of null results in future research may indicate a genuine lack of contact between these benefits of talker familiarity and talker consistency for speech perception.

References

Assgari, A. A., & Stilp, C. E. (2015). Talker information influences spectral contrast effects in speech categorization. *Journal of the Acoustical Society of America, 138*(5), 3023–3032.

Assgari, A. A., Theodore, R. M., & Stilp, C. E. (2019). Variability in talkers' fundamental frequencies shapes context effects in speech perception. *Journal of the Acoustical Society of America, 145*(3), 1443–1454.

Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71*(4), 975–989.

Audacity Team. (2021). *Audacity(R): Free Audio Editor and Recorder [Computer application].* (3.0.0). https://audacityteam.org/. Accessed 14 Nov 2022.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 22(5), 1166–1183.

Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychological Science, 29*(10), 1575–1583. https://doi.org/10.1177/0956797618779083

Holmes, E., To, G., & Johnsrude, I. S. (2021). How Long Does It Take for a Voice to Become Familiar? Speech Intelligibility and Voice Recognition Are Differentially Sensitive to Voice Training. *Psychological Science, 32*(6), 903–915.

Ingvalson, E. M., & Stoimenoff, T. M. (2015). Greater benefit for familiar talkers under cognitive load. *Paper presented at the Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science, 24*(10), 1–10.

Lenth, R. (2019). *emmeans: Estimated marginal means, aka least-squares means* (1.7.0). Retrieved from https://CRAN.R-project.org/package=emmeans. Accessed 14 Nov 2022.

Lim, S.-J., Shinn-Cunningham, B. G., & Perrachione, T. K. (2019). Effects of talker continuity and speech rate on auditory working memory. *Attention, Perception, & Psychophysics, 81*, 1167–1177.

Lüdecke, D. (2022). *ggeffects: Create Tidy Data Frames of Marginal Effects for "ggplot" from Model Outputs* (1.1.2). Retrieved from https://cran.r-project.org/web/packages/ggeffects/. Accessed 14 Nov 2022.

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

Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., & Saltzman, D. (2021). Talker familiarity and the accommodation of talker variability. *Attention, Perception, & Psychophysics, 83*(4), 1842–1860. https://doi.org/10.3758/s13414-020-02203-y

Maibauer, A. M., Markis, T. A., Newell, J., & McLennan, C. T. (2014). Famous talker effects in spoken word recognition. *Attention, Perception, & Psychophysics, 76*(1), 11–18.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., Summers, W., & v. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 676–684.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379–390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*(1), 365–378.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech Perception, Speech Production, and Linguistic Structure* (pp. 113–134). OHM.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42–46.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Reviews of Linguistics, 2*, 33–52.

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, and Hearing Research, 40*(2), 453–463. https://doi.org/10.1044/JSLHR.4002.453

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology, 24*(8), 689–700.

Stilp, C. E., & Theodore, R. M. (2020). Talker normalization is mediated by structured indexical information. *Attention, Perception & Psychophysics, 82*, 2237–2243.

Welford, A. (1980). Choice reaction time: Basic concepts. In J. M. T. Brebner & N. Kirby (Eds.), *Reaction Times*. Academic Press.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics, 79*(7), 2064–2072.

Zhang, C., & Chen, S. (2016). Toward an integrative model of talker normalization. *Journal of Experimental Psychology: Human Perception and Performance, 42*(8), 1252–1268.