Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker

Xin Xie,[a] Kodi Weatherholtz,[b] Larisa Bainton,[c] Emily Rowe,[d] Zachary Burchill, Linda Liu, and T. Florian Jaeger

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14627, USA

(Received 8 August 2017; revised 28 February 2018; accepted 1 March 2018; published online 11 April 2018)

How fast can listeners adapt to unfamiliar foreign accents? Clarke and Garrett [J. Acoust. Soc. Am. **116**, 3647–3658 (2004)] (CG04) reported that native-English listeners adapted to foreign-accented English within a minute, demonstrating improved processing of spoken words. In two web-based experiments that closely follow the design of CG04, the effects of rapid accent adaptation are examined and its generalization is explored across talkers. Experiment 1 replicated the core finding of CG04 that initial perceptual difficulty with foreign-accented speech can be attenuated rapidly by a brief period of exposure to an accented talker. Importantly, listeners showed both faster (replicating CG04) and more accurate (extending CG04) comprehension of this talker. Experiment 2 revealed evidence that such adaptation transferred to a different talker of a same accent. These results highlight the rapidity of short-term accent adaptation and raise new questions about the underlying mechanism. It is suggested that the web-based paradigm provides a useful tool for investigations in speech adaptation. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5027410

[JFL] Pages: 2013–2031

I. INTRODUCTION

Natural speech exhibits within- and cross-talker variability: sound signals carrying the same linguistic content can vary drastically between talkers; even within a talker, no two utterances of the same word are identical. A central question in speech perception is how human listeners handle this variability—what are the mechanisms that underlie this ability and what are its limits? Foreign-accented speech, for example, deviates substantially from the local varieties that listeners experience from a native-accented community and often poses great perceptual difficulty among inexperienced listeners (Flege *et al.*, 1997; Munro and Derwing, 1995). The processing cost of foreign-accented speech compared to native-accented speech manifests in both decreased recognition accuracy and prolonged response times (Adank *et al.*, 2009; Bradlow and Bent, 2008; Floccia *et al.*, 2006; Weil, 2001). Similar processing cost has been found for unfamiliar regions dialects, albeit of smaller magnitudes (Adank *et al.*, 2009; Floccia *et al.*, 2006).

A number of studies have demonstrated that short-term exposure with an unfamiliar foreign accent greatly reduces the initial processing difficulty, resulting in more accurate and faster responses in tasks such as sentence transcription, word detection and so on (Bradlow and Bent, 2008; Weil, 2001). Such findings parallel other reports on listeners' flexibility to adjust to various forms of dialect-accented (Smith *et al.*, 2014), acoustically shifted (Dupoux and Green, 1997), or acoustically degraded speech (e.g., Dahan and Mead, 2010; Davis *et al.*, 2005). A particularly influential study found that native listeners adapt in a surprisingly rapid manner (Clarke and Garrett, 2004; henceforth CG04). In several experiments, native-English listeners heard either a native-English talker, or a foreign-accented talker of moderate proficiency in English. Reaction times were recorded to measure processing difficulty (more details about CG04 are provided below). Initially, foreign-accented speech resulted in much slower reaction times relative to native-accented speech. But these initial delays were found to be attenuated within *a minute* of exposure. Furthermore, after this brief exposure (12–16 sentences), reaction times to the foreign-accented talker were as fast as those to the native English talker. This finding has been widely cited as strong evidence for a highly flexible perceptual system that allows native listeners to rapidly accommodate variation that does not conform to native phonological rules and/or acoustic-phonetic distributions (Google Scholar lists 382 citations to this work, as of 2/21/2018). Notably, the speed and scope of adaptation demonstrated in this study stands in stark contrast to other studies that find persistent accent effects that are not resolved by even more extensive exposure (up to a few days) to a foreign accent (e.g., Floccia *et al.*, 2009; Wade *et al.*, 2007). The discrepant results raise the question whether they reflect differences in task demands or, non-exclusively, inherent limits of perceptual adaptation processes. We know of no other study that has reported similarly rapid adaptation to foreign-accented speech. In addition, since the majority of studies on accent adaptation have examined improvements over longer experiments (Sidaras *et al.*, 2009; Tzeng *et al.*, 2016) or even multiple days (Bradlow and Bent, 2008;

[a]Electronic mail: xxie13@ur.rochester.edu
[b]Present address: Khan Academy, Mountain View, CA 94041, USA.
[c]Present address: Cimpress, 275 Wyman St, Waltham, MA 02451, USA.
[d]Present address: athenahealth, Inc., 311 Arsenal St., Watertown, MA 02472, USA.

Wade *et al.*, 2007; Weil, 2001), it is not known whether extensive exposure is *required* for foreign accent adaptation.

In the present study, we thus conduct a replication of CG04. We closely follow their methods, but address certain problematic choices in their analysis. In a second experiment, we go beyond the original study, and ask whether rapid adaptation to a single foreign-accented talker is transferrable to unfamiliar talkers of the same accent. The answer to this question is critical, as CG04 is sometimes cited as showing evidence of rapid *accent adaptation*—i.e., adaptation to an accent, rather than a specific talker's rendering of the accent. However, in the original CG04 study, participants were only ever exposed to, and tested on, a single foreign-accented talker. Theoretically, adapting to a foreign accent may change perceptual sensitivity to acoustic-phonetically relevant features in the particular accent and may consequentially reshape the encoding of these features as they are mapped onto higher-level linguistic representations. As such, adaptation to the exposure talker potentially benefits subsequent communication with other talkers of the same accent. Alternatively, the rapid adaptation might be highly specific to the particular voice heard during exposure. In this case, it is unlikely that learning will be freely transferrable to a different talker. If listeners can adapt within one or two minutes of accent exposure and this learning can be transferred to a different unfamiliar talker, then it suggests that accent adaptation does not require extensive novel learning experiences. Then an important theoretical question is: what kind of mechanism supports such rapid adaptation? We consider answers to this question in the discussion of our results. We begin by briefly reviewing the methods and results of CG04 and then we provide additional motivations as to why we seek to replicate this study.

A. Review of Clarke and Garrett

Across three experiments, CG04 used performance (measured by accuracy and RT) in a cross-modal word matching task (see Fig. 1 for a schematic illustration) to assess native-English listeners' adaptation to two types of foreign accents (Spanish-accented and Mandarin-accented English). Listeners heard English sentences and saw a visual probe word at the offset of each sentence; their task was to indicate whether the probe word matched the final word of the sentence with a speeded response. In experiment 1, listeners were either exposed to three blocks of Spanish-accented English and tested with the same speaker in block 4 (*Accent* condition), or exposed to native-accented speech in blocks 1–3 and tested with the Spanish-accented speaker in block 4 (*Control* condition), or heard native-accented speech throughout blocks 1–4 (*No accent* condition). Each block consisted of four sentences.

All RTs were adjusted by subtracting the participant's average RT from a baseline block following block 4, in which listeners heard another native-English speaker. These (adjusted) RTs, as well as the error rates within blocks 1–4 were analyzed as indexes of processing difficulty. The *Accent* group had much slower RTs on block 1, but showed an immediate RT decrease after block 1 and continuing decrease through the exposure blocks. On block 4—the test block—RTs in the *Accent* group were significantly lower than RTs in the *Control* group and equal to that in the *No accent* group (RT: *Accent* = *No accent* < *Control*). This result suggests that not only did the *Accent* group gain an advantage over the *Control* group following the Spanish accent exposure, but also the initial processing difficulty was fully attenuated to yield a native-like performance (*Accent* = *No accent*) in block 4. CG04 also analyzed error rates but found no significant differences across blocks 1–4.

In experiments 2 and 3, listeners in the *Accent* group heard Spanish- and Mandarin-accented English, respectively (experiment 3 had six sentences per block). In contrast to experiment 1, the exposure speech heard by the *Control* group was embedded in noise to equate its initial processing difficulty to that of foreign-accented speech. This was done so as to rule out artifacts due to potentially increased attention to the (foreign-accented) speech signal in the *Accent* group. Experiments 2 and 3 found that exposure to noisy speech did not enhance test performance with the foreign accent in block 4. Thus, the *Accent* group's improvements in block 4, as measured by RTs, cannot be attributed to adaptation in effortful listening conditions, but rather reflect learning of the particular foreign accent. As in experiment 1,

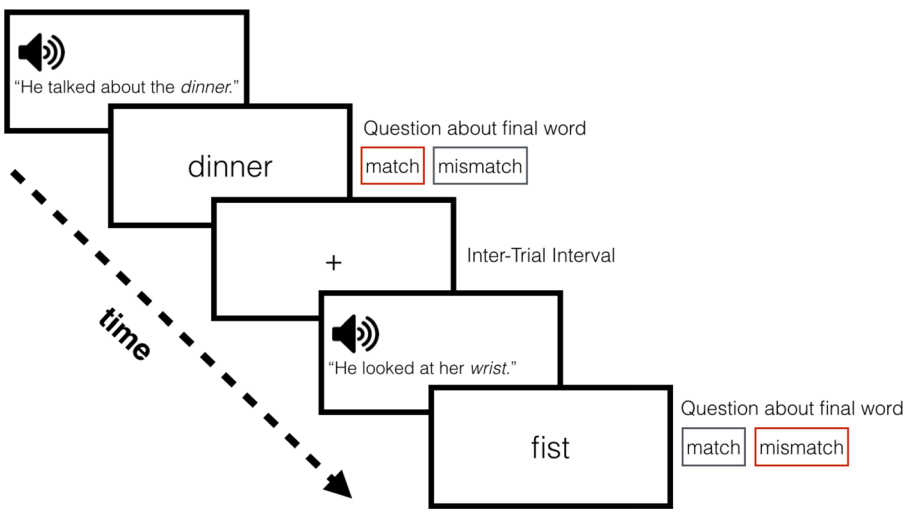

FIG. 1. (Color online) Sequence of two example trials in the cross-modal word matching task. Each trial consists of an audio sentence followed by a visual probe word in print. The visual probe is presented at the offset of the sentence and is terminated by a key press.

experiments 2 and 3 returned no significant differences for error rates across blocks 1–4.

In addition to the theoretical appeal, our decision to replicate and extend CG04 is also methodologically motivated. First, the difference between the *Accent* group and the *Control* group in CG04 cannot be entirely ascribed to accent adaptation, given that the *Control* group experienced a talker change in addition to an accent change from block 3 to block 4, whereas the *Accent* group were listening to the same talker. It is well known that talker switches create additional processing cost in speech perception (e.g., Magnuson and Nusbaum, 2007; Nusbaum and Magnuson, 1997). As also acknowledged in CG04 (p. 3656), it is possible that the poorer performance of the *Control* group during the test block is (at least partially) due to the talker change. By testing both groups with a novel foreign-accented speaker in block 4, we can avoid this confounding issue and at the same time begin to explore the transferability of adaptation.

Second, as described above, CG04 found adaptation in terms of processing speed, but no corresponding improvement in processing accuracy, despite evidence that listeners in the *Accent* group experienced initial processing difficulties in both speed and accuracy. In contrast, other studies have failed to find RT improvements following short-term accent familiarization (e.g., Adank and McQueen, 2007; Floccia *et al*., 2009). And, these differences in improvability of accuracy versus speed have been ascribed theoretical relevance. For example, some proposals hold that short-term exposure does not ameliorate processing difficulty with foreign accents (e.g., Floccia *et al*., 2009). However, caution is needed before one interprets these results as evidence for fundamental limits of rapid accent adaptation. In particular, CG04's ability to detect an accuracy adaptation effect was limited by their use of analysis of variance (ANOVA) to model proportional data (correct vs incorrect responses). This analysis approach is problematic for accuracy data (cf. Jaeger, 2008), particularly when accuracy is near 0 or 1, as was the case for CG04 (with error rates ranging from 0% to 14% across conditions). It is therefore possible that their analysis was not sufficiently sensitive to detect an accuracy adaptation effect. The present study addresses this possibility by employing an alternative analysis approach, which avoids this problem.

A final, non-critical, motivation for the present study is to assess the viability of conducting accent adaptation experiments over the web. Previous research, including work from our lab, has shown that web-based crowdsourcing paradigms are capable of replicating canonical findings regarding language processing, including adaptation and phonetic recalibration effects (e.g., Kleinschmidt and Jaeger, 2012). However, it remains largely an open question whether response times to spoken input, and hence speed-based processing effects, can be reliably assessed via the web.

We present two web-based experiments that examine rapid foreign accent adaptation in a short exposure-and-test paradigm, focusing on the speed and accuracy of adaptation as well as the transfer of adaptation from one talker to another talker. Experiment 1 is a web-based replication of the most critical conditions of CG04. Experiment 2 exploits the same paradigm and tests whether adaptation effects transfer to a different unfamiliar talker. These two experiments represent all of the data collected under this project (i.e., there are no failed unreported experiments). Other work from the same lab has, however, since then replicated the findings reported below in three separate experiments in as of yet unpublished studies using the same procedure but a different accent and different sentence materials. While our experiments and analysis approach were not pre-registered, we planned to, and then did, closely follow the design and analyses of CG04, except where this was either problematic (e.g., for the analyses of error rates) or not possible because critical information was not provided in CG04 (the original authors could not be reached; one left academia and one retired). These exceptions are noted below. We do so in order to minimize our researchers' degree of freedom. In some places we note additional analyses we conducted that go beyond CG04. The results reported below hold under all additional analyses we conducted (if not, we would report so, following good practice, Simmons *et al*., 2011).

II. EXPERIMENT 1

This experiment replicates experiment 3 in CG04 to examine rapid adaptation to a foreign-accented talker. We implemented a web-based version of the cross-modal matching task. Participants listened to spoken sentences and then judged whether a visual probe word, which appeared on screen at sentence offset, matched (or mismatched) the final word of the sentence. Reaction times (RTs) and error rates were measured as indices of processing speed and accuracy, respectively. Like CG04, the experiment used a between-participants design (see Fig. 2). One group of participants, the *Accent* group, heard sentences produced by a Mandarin-accented English speaker during the initial exposure phase, followed by sentences from the same speaker during the subsequent test phase. There were two *Control* groups: *Control in clear* and *Control in noise*. Participants in both *Control* groups were initially exposed to native-American English speech, followed by the same Mandarin-accented speaker during the test phase. The only difference is that speech was presented in noise in the *Control in noise* group that elevated initial perceptual difficulty of native-accented speech to the level of Mandarin-accented speech.

Following CG04, we predict that performance in the *Accent* condition will be initially comparable to that in the *Control in noise* condition but poorer than that in the *Control in clear* condition due to the difficulty of processing foreign-accented speech. Due to task adaptation, performance will improve over the course of the exposure phase, independent of exposure condition. There may also be additional performance improvements in the *Accent* and *Control in noise* conditions because participants are expected to undergo accent adaptation or noise adaptation above and beyond task adaptation. Finally and most critically, processing difficulty with the foreign accent should be attenuated by the exposure, giving the *Accent* condition an advantage over both *Control* conditions in the test phase.

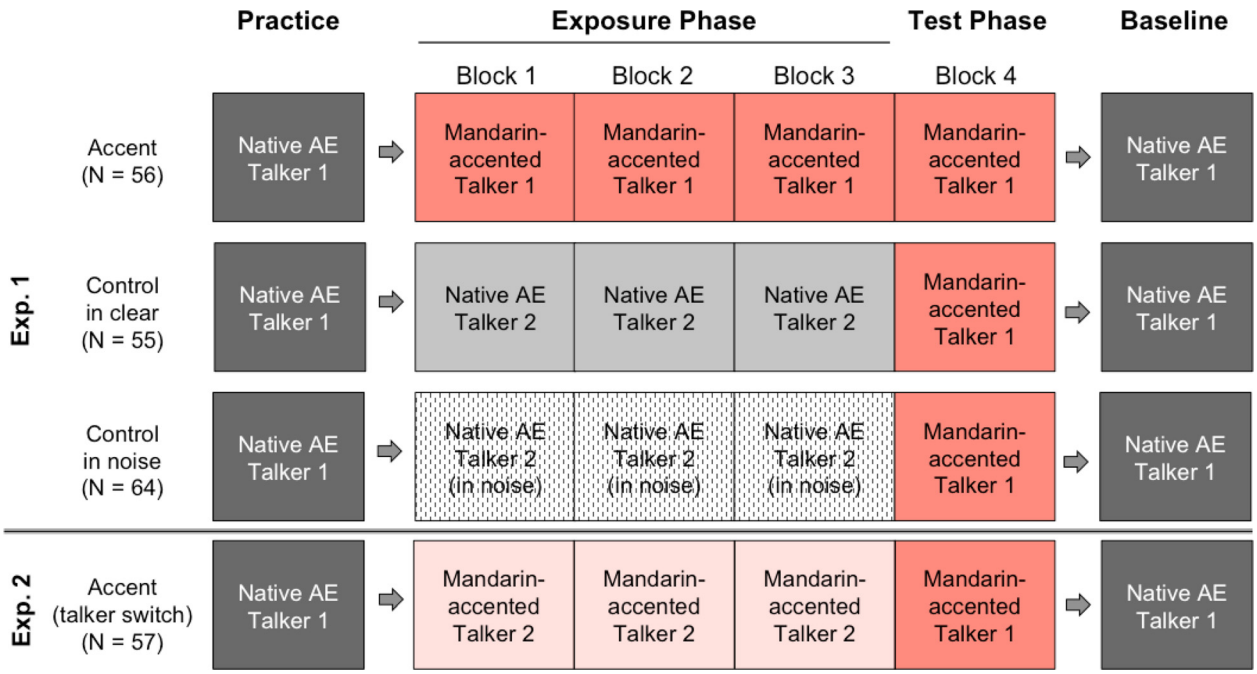

FIG. 2. (Color online) Overview of experimental design for experiments 1 and 2. Each block, including the practice and baseline blocks, comprised six low probability sentences for which the final word was not predictable from the preceding context (e.g., "*Dad pointed at the grass*").

One important limitation to the interpretation of the error rate results obtained by CG04 is that they employed an analysis approach known to be problematic for the analysis of categorical data (Jaeger, 2008). Here we take a different analysis approach that avoids this problem. Given that there is no *a priori* reason to favor either measure, we expect the changes in performance to be demonstrated in either response accuracy or response time. That is, for participants in the *Accent* condition, the initial processing difficulty would result in higher error rates and/or longer response latencies during exposure. During test, however, they are expected to have fewer errors and/or reduced response times than the *Control* groups because *Control* participants hear foreign-accented speech for the first time in this block, whereas participants in the *Accent* condition are already familiar with the foreign accent.

A. Method

1. Participants

CG04 had between 20 and 30 participants per condition in their experiments. Given our aim of a high-powered replication and the potential noisiness of data collected over the web, we doubled the number of participants (i.e., about 50 participants per condition after exclusions). Using Amazon's Mechanical Turk, we recruited a total of 175 participants. Recruitment was limited to monolingual native speakers of American English with normal hearing, an approval rating of 99% on Mechanical Turk, and the ability to use headphones to complete the experiment. The experiment took less than 5 min to complete including instructions. Participants were paid \$0.50 (\$6/h). We excluded participants whose responses on the post-experiment survey did not match our eligibility requirements: i.e., participants who were not monolingual English speakers, who did not use (in-ear or over-ear) headphones, or who reported high familiarity to Mandarin accent (based on self-report of hearing that accent "all the time"). In an effort to identify and exclude participants who were not faithfully performing the experiment (e.g., participants who were multi-tasking or who temporarily disengaged), we excluded participants whose mean RT in any block of the experiment was greater than three standard deviations from the corresponding condition mean (see Table I).

2. Design

Our design was identical to that of CG04. The main portion of the experiment comprised four blocks of six sentences (same as in experiment 3 of CG04): three exposure blocks and one test block, presented with no breaks. In each of these four blocks, half of the trials (n = 3) were followed by a matching visual probe word, and half were followed by a mismatching visual probe word (Fig. 1). The inter-trial

TABLE I. Number of participants recruited and excluded for each condition in experiments 1 and 2. Percentage (out of total recruitment) is reported in parentheses. "Eligibility-excl." represents participants excluded for not meeting our criteria on language background, accent familiarity or audio equipment. "Performance-excl." represents excluded participants due to RT profiles in the experiment (see text for details).

	Condition			
	Experiment 1		Experiment 2	
	Accent	*Control in clear*	*Control in noise*	*Accent*
Recruited	56 (100%)	55 (100%)	64 (100%)	57 (100%)
Eligibility-Excl.	5 (9%)	6 (11%)	3 (5%)	7 (12%)
Performance-Excl.	2 (3.5%)	4 (7%)	6 (9%)	4 (7%)
Remaining	49 (87.5%)	45 (82%)	55 (86%)	46 (81%)

interval was 500 ms. During the three exposure blocks, participants in the *Accent* condition heard a total of 18 sentences produced by a Mandarin-accented speaker. Participants in the *Control in clear* condition heard the same 18 sentences produced by a native speaker of American English. Participants in the *Control in noise* exposure condition heard the same stimuli as in the *Control in clear* condition, but those stimuli were embedded in speech-shaped white noise. During the test block, all participants heard stimuli produced by the Mandarin-accented speaker, allowing us to assess the effect of recent exposure on the processing of foreign-accented speech. Four lists were created to balance the exposure and test blocks in a Latin square design. Two versions of each list were created in reversed orders. This resulted in eight lists per condition for a total of 24 lists. In addition to the main experimental blocks, all participants completed a pre-experiment practice block and a post-experiment baseline block. The practice and baseline blocks were identical across participants: each comprised a fixed set of six novel sentences produced by a different native American English speaker than in the experimental block. The baseline block was designed to assess participants' baseline RTs in the cross-modal matching task to control for individual differences in response speed. To this end, the baseline block occurred at the end of the experiment after participants had adapted to the task and involved stimuli from a native speaker who was equally familiar to all participants but who did not occur during the experimental trials (i.e., the talker from the practice trials). We expected RTs in the baseline block to provide a reasonable measure of participants' baseline response speed in this task, independent of slow downs due to task, accent or talker adaptation (though we note that the transition from the test block to the baseline block involved a change in talker, unlike transitions between exposure blocks; this might be expected to lead to some temporary slow-down, an issue to which we return in Sec. II C).

3. Materials

The set of spoken sentences and corresponding visual probe words used by CG04 are not available. So, we used a novel set of materials, but followed CG04 decisions in designing and selecting experimental materials, including the spoken sentences and visual probes. The full set of stimulus materials is listed in the supplementary material.[1]

a. Sentence recordings. The sentence materials comprised 36 low probability sentences: short declarative sentences in which the final word is not predictable from the preceding context ("Dad pointed at the grass"). These sentences were taken from the Revised Speech Perception In Noise (SPIN-R) test (Kalikow *et al.*, 1977). Of these 36 sentences, 24 served as experimental items, and the remaining 12 were used for practice and baseline blocks. The 12 practice and baseline sentences were recorded in a quiet room by the third author, a female native speaker of American English.

Recordings of the 24 experimental sentences were taken from the Wildcat corpus in the OSCAAR database (Van Engen *et al.*, 2010). We selected a full set of recordings from each of two talkers: a female native American English speaker (Wildcat talker ID: talker 438) and a female Mandarin-accented English speaker with a moderately strong foreign accent based on the subjective impression of the authors (Wildcat talker ID: talker 411). We selected talkers who produced as few hesitations and disfluencies as possible across the set of 24 sentences. Still, some recordings did contain disfluencies such as false starts, strongly affecting the relative length of the recordings. We edited out major hesitations and disfluencies using PRAAT (Boersma and Weenink, 2015). All excisions were made at zero-crossings during silences to avoid introducing audible distortions or clicking. Even after the removal of disfluencies, the two talkers from the OSCAAR database varied in speech rate (see Table II; note that the native talker tended to produce longer carrier phrases but shorter final words, relative to the Mandarin-accented talker, M-Accent 1). Given that we are measuring reaction times in response to stimulus sentences, systematic variability between talkers in overall word and sentence durations could confound our measure of processing speed (cf. discussion in CG04, p. 3650). Thus, we

TABLE II. Properties of experimental stimuli before and after duration normalization. Duration differences summarize the difference between each sentence produced by the native AE talker and the Mandarin-accented talkers (M-accent 1 and M-accent 2 served as the exposure talker in Exp. 1 and Exp. 2, respectively). Shaded columns show percent change in duration of stimuli as a result of duration normalization. Positive numbers indicate stimulus lengthening. Negative numbers indicate shortening.

		Duration (ms) of stimulus tokens		Duration difference (ms) between item tokens (Native AE - M-Accent)		% change in duration due to normalization	
	Talker	mean (sd)	range	mean (sd)	range	mean (sd)	range
Carrier phrase	M-accent 1	984 (256)	[597, 1601]	380 (344)	[−132, 932]	23.0 (22.0)	[−6.5, 57.4]
	M-accent 2	1204 (221)	[765, 1576]	160 (341)	[−490, 745]	9.3 (16.0)	[−17.7, 39.9]
	Native AE	1364 (312)	[888, 2321]			−12.7 (11.1)	[−26.7, 7.5]
Final word	M-accent 1	605 (108)	[381, 803]	−107 (83)	[−314, 29]	−8.2 (6.0)	[−19.6, 3.1]
	M-accent 2	589 (100)	[440, 846]	−91 (72)	[−236, 92]	−7.4 (6.2)	[−19.1, 8.8]
	Native AE	498 (104)	[332, 701]			12.2 (10.3)	[−2.9, 34.7]
Total	M-accent 1	1589 (270)	[1160, 2130]	274 (382)	[−310, 907]	9.9 (12.8)	[−8.8, 30.8]
	M-accent 2	1794 (234)	[1295, 2275]	69 (364)	[−623, 742]	3.3 (10.7)	[−15.0, 20.7]
	Native AE	1862 (352)	[1337, 3018]			−6.1 (9.7)	[−19.0, 10.7]

applied the same duration normalization used by CG04. Specifically, for the 24 experimental stimuli produced by the Mandarin-accented and native-accented talkers, we digitally adjusted the length of the final (target) word to equal the corresponding mean duration of the two original productions. We also equated the length of the preceding sentence context (CG04 did not report taking this procedure).

For the Mandarin-accented talker, the carrier phrase was lengthened by 23% on average across sentences, and the final word was shortened by 8.2% on average (see Table II for corresponding ranges and standard deviations). Correspondingly, the carrier phrase was shortened by 12.7% on average across sentences, and the final word was lengthened by 12.2% on average for the native talker. Critically, however, the normalization procedure introduced minimal, if any, distortion across the experimental stimuli, based on the subjective impression of the authors.

After the duration normalization procedure, all sentences (including the practice and baseline sentences) in each condition were scaled to an average intensity of 65 dB. CG04 did not report normalization of stimulus intensity. Lack of normalization would leave open the possibility that the observed differences in processing speed were influenced (in part) by the ease with which listeners could hear, and hence understand, the native- versus foreign-accented talkers. By normalizing the average intensity of the stimuli, we address this potential confound.

Finally, to create the exposure materials for the *Control in noise* condition, the 24 experimental stimuli produced by the native English speaker were embedded in speech-shaped white noise at a signal-to-noise ratio (SNR) of +2 dB (i.e., speech signal = 65 dB; noise = 63 dB). Recall that the goal of this noise embedding was to create a condition in which listeners heard native speech but nonetheless experienced initial difficulty in terms of processing speed and accuracy. Specifically, we aimed to match the initial processing difficulty in the *Control in noise* condition to the initial difficulty in the *Accent* condition. By matching initial difficulty in these two conditions, we could then assess whether changes in the speed and processing of foreign-accented speech over the course of the experiment were due to accent adaptation, or instead due to increased attention or task engagement due to initial processing difficulty. Following CG04, we first embedded the native speech in noise at an SNR of +1 dB. Results of a pilot study showed that listeners' initial performance at this SNR was worse than that in the *Accent* condition, whereas an SNR of +2 dB provided the desired level of initial processing difficulty (as shown in Sec. II B below).

b. Visual probe words. Visual probes were written words presented on screen immediately following the spoken sentence. The probe words were horizontally and vertically centered on screen. For half of the trials, the visual probe matched the final (target) word of the preceding sentence. For the other half of trials, the visual probe mismatched the target word. These mismatches were created by altering the target word by one phoneme in either the onset, vowel or coda position (e.g., He looked at her wrist. – FIST). The mismatching visual probes comprised an equal number of each phoneme substitution type (onset, vowel, coda). All visual probe words were familiar monosyllabic or bisyllabic English nouns. The mismatching visual probes were selected to closely correspond in frequency to the matching visual probes, in terms of words per million (matching probes: M = 62 wpm, SD = 91; mismatching probes: M = 63 wpm, SD = 162). Note that we measured word frequency using the much larger—and more recent—CELEX English word form database (Baayen *et al.*, 1995), whereas CG04 used the frequency data reported by Kučera and Francis (1967).

4. Procedure

The entire experiment was conducted over the web using Amazon's Mechanical Turk. Participants were randomly assigned to one of the conditions (*Accent*, *Control in clear*, or *Control in noise*) and then randomly assigned to an experimental list (one of eight lists per condition), with list assignment balanced across participants. Participants began the experiment by verifying their eligibility (monolingual native English speaker wearing headphones) and giving informed consent. Participants then completed a short transcription task (two words) to ensure that they could hear and respond to audio stimuli.

After this initial consent and verification procedure, participants were given instructions about the task. The exact instructions from CG04 study are not available; however, they indicated that participants were "instructed to respond quickly and accurately and were warned that at some time during the experiment the voice would change" (p. 3650). We therefore gave participants the following task instructions:

> *You will hear a series of sentences. After each sentence, a word will appear on the screen. Your job is to identify whether the word on the screen matches the last word of the sentence. Please respond as quickly as possible without sacrificing accuracy. The person speaking might change throughout the experiment.*

Participants were instructed to press either "x" or "m" on their keyboard for "match" and "mismatch," with the correspondence between button and response counterbalanced across participants. After the experiment, participants completed a short survey assessing their audio quality, language background, and familiarity the foreign accent they heard during the experiment. The complete post-experiment survey is presented in the supplementary material.[1]

5. Analysis

We followed CG04 analysis procedure, except that we used mixed-effects models fit to the trial-level data instead of ANOVA fit to by-subject block means. Analyzing the data at the trial-level allows us to benefit from all the additional information, and thus power, provided by that data, which would be lost if we aggregated by block. Further, mixed-effects models allow us to simultaneously account for random by-participant and by-item variability. All mixed-effects analyses were performed in R (R Core Team, 2014)

using the *lme4* package (version 1.1-13; Bates *et al.*, 2017). P-values for linear mixed-effects analyses were obtained using the Satterthwaite approximation for degrees of freedom, as implemented in the *lmerTest* package (version 2.0-33, Kuznetsova *et al.*, 2016).

Two dependent measures were analyzed as indices of adaptation: errors as an index of processing accuracy, and response times (RTs) as an index of processing speed. Errors were analyzed using logistic mixed-effects regression. RTs were analyzed using linear mixed-effects regression. RTs were adjusted prior to analysis to account for individual differences in baseline response speed: each participant's mean RT from the post-experiment baseline block was subtracted from their RTs during the exposure and test blocks.[2] We excluded trials with extreme RTs. Based on visual examination of the distribution of RTs, we first excluded trials with RTs less than 200 ms or greater than 8000 ms. We then excluded trials with RTs greater than three standard deviations from the corresponding participant's mean RT. These criteria were looser than that adopted by CG04 (RTs greater than 2000 ms or outside +/− 2SD) in order to reduce data loss, following standard practice in our lab. Our criteria resulted in a total exclusion of 3.1% of trials.

a. Mixed-effects model specification. For all analyses, the full mixed-effects models included fixed effects for the predictors of theoretical interest (i.e., exposure condition, block), as well as fixed effects for the counterbalancing nuisance variables (i.e., list and list order), and all possible interaction terms. Additionally, all models were specified with the maximal random effects structure justified by the experimental design: that is, by-subject and by-item random intercepts, by-subject random slopes for all within-subject design variables (i.e., experimental block), and by-item random slopes for all within-item design variables (i.e., exposure condition). If this analysis failed to converge within 10 000 iterations, the model was systematically simplified in a step-wise fashion until the model converged. For this process, we started by simplifying the random effects structure: removing correlations among random effects, and then dropping random effects term with the least variance. In some cases, noted below, it was necessary to remove additional fixed effects that were inconsequential for the theory being tested: namely, the counterbalancing nuisance variables. This was only the case for analyses of error rates, which were often close to 0% (a floor effect), supporting fewer degrees of freedom for the analysis (see, e.g., references in Jaeger, 2011).

b. Mixed-effects model reporting. For each mixed-effects analysis, we report the full set of fixed effect parameter estimates (including all multi-way interactions with counterbalancing nuisance variables) in the supplementary material.[1] In the majority of these analyses, the counterbalancing variables were non-significant, either as main effects or in interaction with the experimental variables. Therefore, in the interest of readability, we report in the main text only parameter estimates for the experimental predictor variables (exposure condition and experimental block) and refer the reader to the supplementary material[1] whenever the counterbalancing variables reached significance.

B. Results

CG04 reported evidence of accent adaptation in terms of processing speed but not processing accuracy. Therefore, we first analyze RTs to assess whether we replicated CG04 in terms of processing speed effects. We then present the error analyses.[3]

1. Response times

Figure 3 shows baseline-normalized RTs throughout exposure and test as a function of exposure condition and experimental block.

a. Exposure phase. To assess the effect of exposure condition on processing speed during the initial exposure phase, we fit a linear mixed effects model to adjusted RTs on trials that were answered correctly (87.7% of data). Fixed effects were specified for exposure condition [sliding contrast to compare (i) *Accent* vs *Control in clear* and (ii) *Control in noise* vs *Accent*], exposure block [Helmert coded

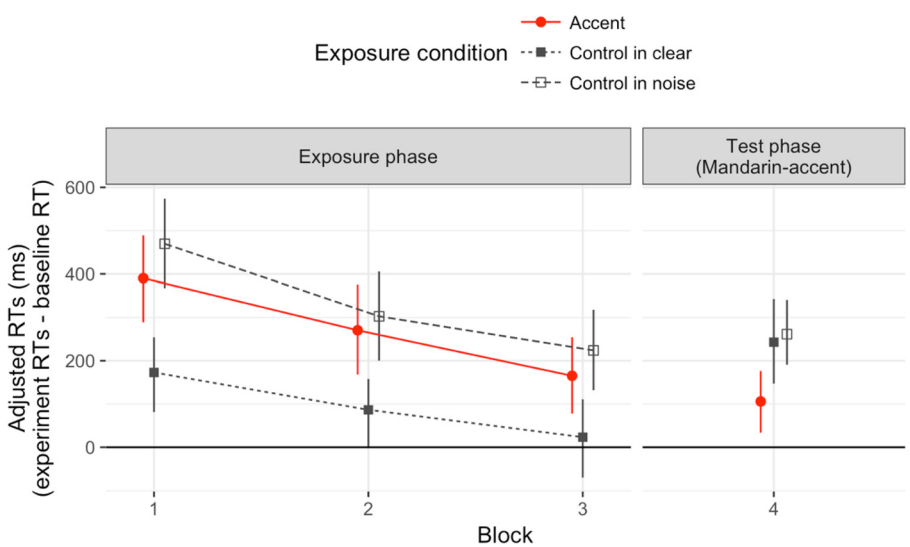

FIG. 3. (Color online) Experiment 1 baseline-normalized RTs by exposure condition and block during the exposure and test phases. Dots indicate block means with corresponding bootstrapped 95% confidence intervals.

TABLE III. Experiment 1, summary of analysis of adjusted RTs during exposure.

Predictors (fixed effects)	Parameter estimates		Test statistic	Satterthwaite approx.	
	Coef $\hat{\beta}$	SE ($\hat{\beta}$)	t	df	p
(Intercept)	244.2	34.3	7.12	52.7	<0.001
CONDITION 1 (= *Accent* vs *Control in clear*)	186.6	62.6	2.98	108.4	<0.01
CONDITION 2 (= *Control in noise* vs *Accent*)	64.0	64.5	0.99	88.9	0.32
BLOCK 1 (= Block 2 vs 1)	−59.5	11.5	−5.20	291.0	<0.001
BLOCK 2 (= Block 3 vs mean of 1 and 2)	−46.8	6.9	−6.78	179.3	<0.001
CONDITION 1: BLOCK 1	−22.9	28.5	−0.80	280.6	0.42
CONDITION 1: BLOCK 1	−24.0	28.1	−0.85	317.6	0.39
CONDITION 2: BLOCK 2	−20.2	17.1	−1.18	172.2	0.24
CONDITION 2: BLOCK 2	−0.3	16.8	−0.02	195.5	0.99

to compare (i) block 2 vs block 1 and (ii) block 3 to the mean of blocks 1 and 2], counterbalancing list (four-level nuisance factor; sum contrast coded), and counterbalancing list order (forward vs reverse; sum contrast coded), and all interactions.

Table III summarizes the effects of interest (the full analysis summary including all counterbalancing nuisance predictors is reported in the supplementary material[1]). There was a significant effect of condition such that participants in the *Accent* condition were slower overall than participants in the *Control in clear* condition, indicating difficulty associated with processing foreign accented speech. There was no significant difference between the *Accent* and *Control in noise* conditions (see Fig. 3). Thus, the difficulty associated with listening to speech in noise was comparable to the difficulty associated with foreign-accented speech, in terms of the overall effect on processing speed. There was also a significant main effect of block: RTs decreased over the course of the exposure phase independent of condition, indicating adaptation to the task.

Consistent with CG04, there was no significant interaction between condition and block: that is, the magnitude of the RT decrease throughout the exposure phase was comparable across conditions. Thus, to the extent that accent adaptation occurred during exposure, the accent adaptation effect is small and indistinguishable from task adaptation, at least when analyzed under the assumption of linear effects across exposure blocks.

b. Test phase. The same analysis as for exposure was repeated for RTs from the test phase, except that block was not included in the analysis (since there was only one test block). The parameter estimates for the experimental variables are summarized in Table IV (the full analysis summary including all counterbalancing nuisance variables is reported in the supplementary material[1]).

The effect of condition was significant: replicating CG04, participants who heard Mandarin-accented English during the initial exposure phase were faster to respond correctly than participants in the control conditions (*Control in clear* and *Control in noise*). Thus, initial exposure to the foreign-accented talker attenuated accent-related processing difficulty, resulting in a behavioral improvement above and beyond the effect of task exposure (*Accent* < *Control in clear*) or improvement due to increased task engagement resulting from increased baseline processing difficulty (*Accent* < *Control in noise*). Indeed, a *post hoc* test revealed that there was no difference between the *Control in noise* and *Control in clear* conditions in the (accented) test phase ($\hat{\beta}_{\text{Control in noise vs Control in clear}} = 23.7$, $t = 0.32$, $p = 0.75$; see Fig. 3), suggesting that increased listening effort due to noise manipulation did not transfer to enhanced comprehension of foreign-accented speech.[4]

TABLE IV. Experiment 1, summary of analysis of adjusted RTs during test.

Predictors (fixed effects)	Parameter estimates		Test statistic	Satterthwaite approx.	
	Coef $\hat{\beta}$	SE ($\hat{\beta}$)	t	df	p
(Intercept)	218.5	43.6	5.01	30.8	<0.001
CONDITION 1 (= *Accent* vs *Control in clear*)	−163.9	58.8	−2.79	117.8	<0.01
CONDITION 2 (= *Control in noise* vs *Accent*)	163.2	57.5	2.84	85.5	<0.01

2. Error rates

Figure 4 shows the proportion of errors throughout exposure and test as a function of exposure condition and block.

a. Exposure phase. To assess the effect of exposure condition on processing accuracy during the exposure phase, we fit a logistic mixed-effects model to errors (incorrect response = 1, correct response = 0). The predictors included in the analysis were identical to those in the analysis of exposure RTs. Table V summarizes the variables of interest (see the supplementary material[1] for a summary of the full converging model with all counterbalancing variables).

There was a significant effect of condition such that participants in the *Accent* condition made more errors overall than participants in the *Control in clear* condition, indicating difficulty associated with processing foreign-accented speech. There was no significant difference in error rates between the *Accent* and *Control in noise* conditions: as also shown in Fig. 4, the error rates for the *Control in noise* condition correspond closely to the error rates for the *Accent* condition across blocks, paralleling the results of the RT analysis above. Thus, the difficulty associated with listening to speech in noise was comparable to the difficulty associated with foreign-accented speech, in terms of both processing speed (as shown in Fig. 3 above) and processing accuracy. No other predictors were significant. Notably, the coefficient estimates for the effect of block indicated a near zero (but numerically negative) change in error rates over the course of the exposure phase. Thus, participants showed little evidence of adaptation in terms of processing accuracy during the course of exposure.

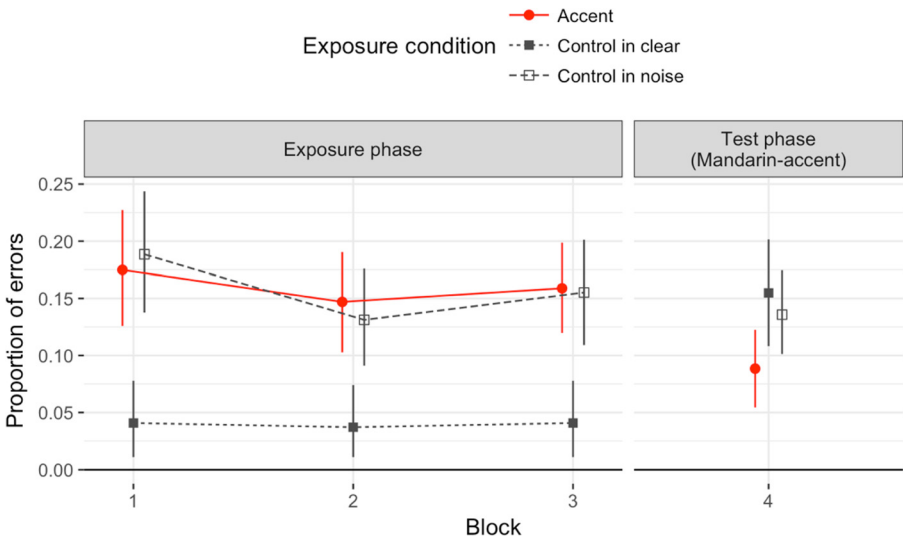

FIG. 4. (Color online) Experiment 1, proportion of errors by exposure condition and block during the exposure and test phases. Dots indicate mean errors per block with corresponding bootstrapped 95% confidence intervals.

b. Test phase. To assess the effect of exposure on processing accuracy of foreign-accented speech, a logistic mixed-effects model was fit to errors during the test phase. The initial analysis included the same predictors as the analysis of test RTs. By-participant and by-item random intercepts were specified. Table VI summarizes the variables of interest (see the supplementary material[1] for a summary of the full converging model with all counterbalancing variables).

Participants in the *Accent* condition made significantly fewer errors than participants in the *Control in clear* condition. Thus, exposure to foreign-accented speech improved processing accuracy of that accented talker relative to task control. Further, participants in the *Accent* condition made marginally fewer errors than participants in the *Control in noise* condition. Thus, the accuracy benefit resulting from exposure to a foreign accent cannot be attributed to greater task engagement due to initial processing difficulty. Indeed, a *post hoc* test revealed that the *Control in noise* exposure condition did not result in a significant reduction in errors for the (accented) test phase, compared to the *Control in clear* condition ($\hat{\beta}_{\text{Control in noise vs Control in clear}} = -0.09$, $z = -0.65$, $p = 0.52$; see Fig. 4).

TABLE V. Experiment 1, summary of analysis of errors during exposure.

	Parameter estimates		Wald's test	
Predictors (fixed effects)	Coef $\hat{\beta}$	SE ($\hat{\beta}$)	Z	Pz
(Intercept)	−2.8	0.22	−12.7	<0.001
CONDITION 1 (= *Accent* vs *Control in clear*)	2.0	0.33	6.3	<0.001
CONDITION 2 (= *Control in noise* vs *Accent*)	−0.1	0.24	−0.3	0.77
BLOCK 1 (= Block 2 vs 1)	−0.2	0.10	−1.6	0.12
BLOCK 2 (= Block 3 vs mean of 1 and 2)	0.0	0.06	0.1	0.90
CONDITION 1: BLOCK 1	0.0	0.27	0.0	0.98
CONDITION 1: BLOCK 1	−0.1	0.17	−0.8	0.43
CONDITION 2: BLOCK 2	0.0	0.16	0.1	0.91
CONDITION 2: BLOCK 2	0.0	0.10	−0.1	0.91

C. Discussion

Experiment 1 provides evidence for adaptation in terms of both speed and accuracy. Specifically, both the RT profile and the error profile in the *Control in noise* condition closely matched performance in the *Accent* condition during exposure. Thus, we succeeded in matching these two conditions in terms of overall processing difficulty. However, performance at test showed that participants in the *Accent* condition were both faster and more accurate than participants in the *Control in noise* condition, and even more so when compared to the *Control in clear* condition. Thus, listening to one form of difficult-to-understand speech during exposure (i.e., speech in noise) did not influence processing accuracy when listening to a different form of difficult-to-understand speech at test (i.e., foreign-accented speech). Taken together, these results suggest that relatively brief exposure to foreign-accented speech influences processing accuracy and, moreover, that this accuracy benefit cannot be attributed to increased attention or task engagement (e.g., general strategy for compensating for initial processing difficulty), given performance in the *Control in noise* condition.

Our results replicate the main findings of CG04 that adaptation to the accented speaker occurs rapidly within this brief exposure paradigm. There are three caveats to this replication. First, while CG04 did not observe any adaptation effects in accuracy measures (error rates) either during exposure or at test, we found that exposure to accented speech engendered a subsequent benefit at test in terms of both processing speed and accuracy, showing a more robust

TABLE VI. Experiment 1, summary of analysis of errors during test.

	Parameter estimates		Wald's test	
Predictors (fixed effects)	Coef $\hat{\beta}$	SE ($\hat{\beta}$)	Z	Pz
(Intercept)	−2.6	0.32	−8.2	<0.001
CONDITION 1 (= *Accent* vs *Control in clear*)	−0.8	0.32	−2.4	<0.05
CONDITION 2 (= *Control in noise* vs *Accent*)	0.6	0.31	1.9	= 0.05

adaptation effect. Given that we used different experimental materials and analysis approaches than CG04, it is not possible to pinpoint the discrepancy to a single factor. Nevertheless, our results suggest that both processing speed (as measured by RTs) and processing accuracy can benefit from accent exposure (cf. Floccia *et al.*, 2009). We return to this point in Sec. IV.

Second, a somewhat surprising result of CG04 is that by the end of exposure (16 sentences), listeners showed no difference in RTs when responding to foreign-accented speech versus native-accented speech (experiment 1, p.3651). This result has sometimes been cited as evidence for a "complete" accent adaptation. We instead found that neither RTs nor errors in the *Accent* condition matched native-like processing at the end of exposure. *Post hoc* tests revealed that the *Accent* condition had significantly more errors ($\hat{\beta}_{\text{Accent vs Control in clear}} = 0.95$, $z = 4.24$, $p < 0.001$; see Fig. 4) and longer response times ($\hat{\beta}_{\text{Accent vs Control in clear}} = 166.5$, $t = 1.92$, $p = 0.06$; see Fig. 3) than the *Control in clear* condition in block 3 (18 sentences). This discrepancy is likely due to differences in the accent strength of the particular talkers selected in our study versus CG04. In this regard, our observations are consistent with other findings that processing difficulty of foreign-accented speech often persists, albeit diminished, after brief exposure.

Third, similar to CG04, we did not observe any decrease in error rates during exposure for any of the three groups, in contrast to a clear RT improvement in all groups. It is possible that perceptual learning, especially at the earliest stages, is non-monotonic. For example, listeners might sometimes temporarily get stuck on "wrong" hypotheses about the representations that underlie unfamiliar input, before further data points allow learners to adjust those hypotheses.

Following CG04, we have so far interpreted the differences in processing speed and accuracy between conditions at test as evidence of accent adaptation: that is, exposure to a foreign-accented talker facilitated processing of that talker's accent, relative to task control. It is possible, however, that the behavioral differences at test reflect, or were at least amplified by, differences in task complexity between conditions. Participants in the *Accent* condition heard the same talker throughout exposure and test, whereas participants in the control conditions experienced a change in talker between the exposure and test blocks. Given that talker switching is often assumed to be associated with attentional costs, as reflected in processing differences in the presence versus absence of talker changes in native speech perception (e.g., Mullennix *et al.*, 1989; Magnuson and Nusbaum, 2007), an important question is whether the processing differences at test indeed reflect a processing advantage in the *Accent* condition, relative to the control conditions, due to preceding accent exposure, or whether these differences instead reflect a processing *dis*advantage in the control conditions, relative to the *Accent* condition, due to the attentional cost of talker switching. This question, which was left open (though acknowledged) by CG04, must be addressed in order to conclude that the observed processing differences at test are in fact evidence of rapid accent adaptation. In experiment 2, we thus changed the Mandarin-accented talker between exposure and test blocks in the *Accent* condition, to control for talker switching effects.

III. EXPERIMENT 2

For experiment 2, we ran a new version of the *Accent* condition, which we then compared against performance in the *Control in clear* condition (we refer to it as the "Control" condition in experiment 2) from experiment 1. For this new *Accent* condition, participants heard Mandarin-accented English throughout the experimental blocks (as in the *Accent* condition in experiment 1), but the Mandarin-accented talker changed between the exposure and test blocks. Thus, participants in both the *Accent* and *Control* conditions encountered an unfamiliar talker at test. Specifically, the accented test talker remained the same in all conditions of experiments 1 and 2, but a new accented exposure talker was used for the *Accent* condition of experiment 2. If participants in the *Accent* condition of experiment 2 continue to show a processing speed and accuracy advantage at test, relative to control, these processing effects cannot be attributed to differences in task complexity or talker familiarity between conditions (this assumes that participants detect the talker switch in the new *Accent* condition—an assumption that we return to in Sec. III C).

Of additional theoretical interest in experiment 2 is that the talker switch allows us to also assess whether the benefits of adaptation are restricted to a particular talker. In order for participants in the *Accent* condition to show a processing advantage at test relative to control, they must transfer from one accented talker (i.e., the exposure talker) to a new talker with the same accent. Early work on foreign accent adaptation found no evidence of generalized perceptual benefits for a different talker following exposure to a single foreign-accented talker (Bradlow and Bent, 2008). More recent work, however, suggests that single-talker exposure can, in fact, be sufficient for cross-talker generalization (Reinisch and Holt, 2014; Xie and Myers, 2017), though the magnitude of the generalization effect can vary considerably across exposure-test talker pairs). This highlights that the nature of the exposure conditions that lead to robust cross-talker generalization is still poorly understood. In terms of the current study, it is therefore possible that initial exposure to one Mandarin-accented talker will not be sufficient for listeners to successfully generalize to a new Mandarin-accented talker at test. In that case, we would expect participants in the *Accent* condition (comparable to control) to have difficulty processing speech from the Mandarin-accented talker at test. Given these caveats about cross-talker generalization, observing a processing advantage in the *Accent* condition at test relative to *Control* would provide strong evidence of transferrable adaptation effects.

Here, our primary interest lies in establishing whether the evidence for adaptation observed in experiment 1 (and in CG04) is in fact due to the attentional cost of talker switching in the *Control* condition. We note that the generalizability of accent adaptation is a separate issue. That is, while improved performance for the unfamiliar test talker speaks for a transfer of adaptation benefits, it does not

necessarily stand for talker-general adaptation: generalization to one specific talker of the same accent does not guarantee generalization to all talkers of the same accent. Rather, we see it as a proof of concept that may inspire further explorations into the mechanisms of generalization in accent adaptation.

A. Method

1. Participants

A total of 57 participants were recruited for the new foreign-accented condition. Recruitment method and payment were the same as in experiment 1. Of these 57 participants, seven were excluded for not meeting eligibility criteria, and four were excluded due to their RT profiles. After exclusions there were a total of 46 participants in the new accented condition.

2. Design

The design was identical to experiment 1, with one exception: participants heard Mandarin-accented English throughout the experimental blocks (consistent with the *Accent* condition in experiment 1), but the talker changed between the exposure and test blocks (see Fig. 2). The Mandarin-accented talker presented during the test phase was the same talker heard during test in experiment 1; the exposure talker was replaced by a different Mandarin-accented talker. This allowed us to use performance in the *Control in clear* (henceforth *Control* in this experiment) condition from experiment 1 as a baseline for comparison.

3. Materials

Speech materials used in the practice, baseline and test blocks were identical to that in experiment 1. Eighteen sentences during the exposure blocks were taken from a different female Mandarin-accented speaker in the WILDCAT corpus (WILDCAT ID: 414). Note that as in experiment 1, productions of this new Mandarin-accented speaker were normalized for duration such that each sentence (and its final word) equated the mean of the new talker's original production and the mean of the native AE speaker's original production (used in the control conditions in experiment 1). Due to production differences between this new Mandarin-accented speaker and the Mandarin-accented speaker used in experiment 1, durations of each exposure item were not identical to those in experiment 1, although the by-item difference for target words was not significant [$t(23) = 0.797$, $p = 0.434$].[5] The procedure for generating and counterbalancing stimuli lists was the same as in experiment 1.

4. Procedure

The experimental procedure was identical to experiment 1, as was the procedure for preparing and analyzing the data. The trial-wise outlier exclusion procedure resulted in a loss of 2.5% of trials.

B. Results

Throughout all analyses below, the model specification was the same as in experiment 1, except that the Mandarin-accented condition from experiment 2 (rather than the *Accent* condition from experiment 1) was compared to the *Control* condition in experiment 1 (sum contrast coding: *Accent* condition = 1; *Control* = −1).

1. Response times

Figure 5 shows baseline-normalized RTs throughout the experiment as a function of exposure condition and experimental block. The Mandarin-accented condition from experiment 2 is plotted alongside the data from experiment 1 for comparison.

a. Exposure phase. A linear mixed effects model was fit to adjusted RTs on correctly answered trials using the same specifications as for the analysis of exposure RTs in experiment 1. Random effects included intercepts for participants and items, along with a by-participant slope for block and a by-item slope for condition. Table VII summarizes the experimental effects of interest (see the supplementary material[1] of a complete summary of fixed effects estimates).

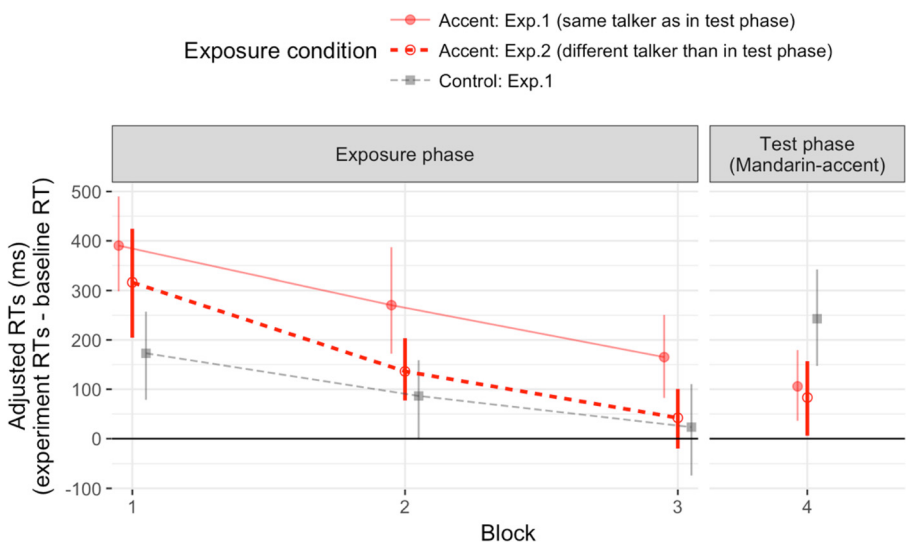

FIG. 5. (Color online) Experiment 2 baseline-normalized RTs by exposure condition and block during the exposure and test phases. Dots indicate block means with corresponding bootstrapped 95% confidence intervals. Open circles indicate data from experiment 2. Data from experiment 1 (denoted by filled symbols and lighter lines) is plotted for comparison. The critical comparison is between the new *Accent* condition from experiment 2 and the *Control* condition from experiment 1.

TABLE VII. Experiment 2, summary of analysis of adjusted RTs during exposure.

Predictors (fixed effects)	Parameter estimates Coef $\hat{\beta}$	SE ($\hat{\beta}$)	Test statistic t	Satterthwaite approx. df	p
(Intercept)	135.8	38.3	3.55	46.0	<0.001
CONDITION (= *Accent vs Control*)	38.7	24.7	1.56	85.5	0.12
BLOCK 1 (= Block 2 vs 1)	−63.5	13.0	−4.87	171.5	<0.001
BLOCK 2 (= Block 3 vs mean of 1 and 2)	−49.3	7.9	−6.23	106.7	<0.001
CONDITION: BLOCK 1	−26.1	13.0	−2.00	171.6	<0.05
CONDITION: BLOCK 2	−16.2	7.9	−2.04	106.6	<0.05

Unlike in experiment 1, the main effect of condition was not significant, even though RTs in the *Accent* condition were numerically larger relative to *Control*. There was a main effect of block: RTs decreased over the course of the exposure phase, as expected due to task adaptation. There was also a significant interaction of condition and block: the change in RTs across blocks was larger in the *Accent* condition than in the *Control* condition. These results provide statistical evidence for accent adaptation above and beyond task adaptation during the exposure phase—unlike in experiment 1.

As shown in Fig. 5, RTs in the *Accent* condition were higher than *Control* in block 1 but rapidly converged against RTs in the *Control* condition over the course of the exposure phase. A *post hoc* analysis confirmed a significant difference between the two conditions at the start of the experiment ($\hat{\beta}_{\text{Accent vs Control}} = 78.5$, $t = 2.36$, $p = 0.02$): during block 1, the *Accent* condition were 177 ms slower on average to make a correct response. This difference between conditions was no longer significant in block 2 ($\hat{\beta}_{\text{Accent vs Control}} = 33.2$, $t = 1.21$, $p = 0.23$). Thus, within 12 trials of exposure, participants in the *Accent* condition overcame the initial slowdown in processing caused by the talker's foreign accent.

b. Test phase. A linear mixed effects model was fit to adjusted RTs on correctly answered trials using the same specifications as for the analysis of test RTs in experiment 1. Random effects included intercepts for participants and items, along with a by-item slope for condition. Table VIII summarizes the experimental effects of interest (see the supplementary material[1] for a full model). The main effect of condition (*Accent* < *Control*) was significant, suggesting that exposure to one Mandarin-accented talker facilitated processing speech from a different Mandarin-accented talker, relative to task control.

TABLE VIII. Experiment 2, full analysis of adjusted RTs during test.

Predictors (fixed effects)	Parameter estimates Coef $\hat{\beta}$	SE ($\hat{\beta}$)	Test statistic t	Satterthwaite approx. df	p
(Intercept)	187.6	47.3	3.96	38.1	<0.001
CONDITION (= *Accent* vs *Control*)	−91.6	30.3	−3.03	90.6	<0.01

2. Error rates

Figure 6 shows the proportion of errors throughout exposure and test as a function of exposure condition and block.

a. Exposure phase. A logistic mixed-effects model was fit with random intercepts for participants and items, but no random slopes. Table IX summarizes the experimental effects of interest (see the supplementary material[1] for a complete model summary). None of the experimental predictors were significant. As shown in Fig. 6, error rates in the *Accent* condition from experiment 2 were low throughout the exposure phase and closely paralleled the error rates in the *Control* condition. This is in line with the RT pattern: the accented exposure talker in experiment 2 seems to have been *a priori* more intelligible for our participants, compared to the accented exposure talker in experiment 1.

b. Test phase. A logistic mixed-effects model was fit with random intercepts for participants and items, along with

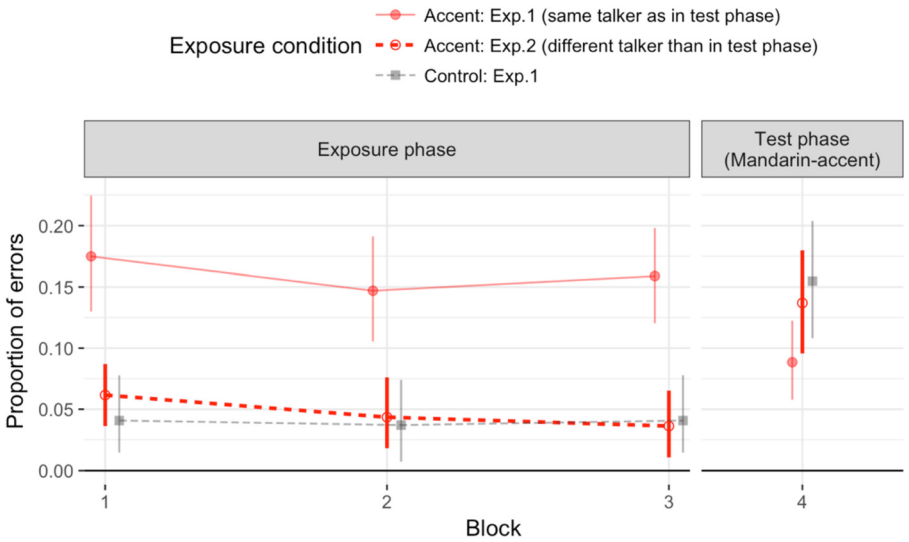

FIG. 6. (Color online) Experiment 2, proportion of errors by exposure condition and block during the exposure and test phases. Dots indicate mean errors per block with corresponding bootstrapped 95% confidence intervals. Open circles indicate data from experiment 2. Data from experiment 1 (denoted by filled symbols and lighter lines) is plotted for comparison.

TABLE IX. Experiment 2, summary of analysis of errors during exposure.

	Parameter estimates		Wald's test	
Predictors (fixed effects)	Coef $\hat{\beta}$	SE ($\hat{\beta}$)	Z	Pz
(Intercept)	−4.6	0.42	−10.9	<0.001
CONDITION (= *Accent* vs *Control*)	0.3	0.24	1.4	0.17
BLOCK 1 (= Block 2 vs 1)	−0.1	0.18	−0.8	0.43
BLOCK 2 (= Block 3 vs mean of 1 and 2)	−0.1	0.10	−0.5	0.60
CONDITION: BLOCK 1	−0.1	0.18	−0.6	0.56
CONDITION: BLOCK 2	−0.1	0.10	−0.9	0.37

a by-item random slope for condition. There was no significant difference in terms of error rates between the *Accent* condition and the *Control* condition (Table X; see the supplementary material[1] for a complete model summary).

C. Discussion

It was clear that accent exposure to one talker facilitated recognition of another talker's accented speech in terms of processing speed, in the absence of corresponding significant patterns in processing accuracy. This result addresses a potential confound of both the original study in CG04 and our experiment 1. In those earlier experiments, the test talker was identical to the exposure talker in the *Accent* condition, but not in the *Control* condition. In those experiments, the slow-down in processing speed observed in the *Control* condition could thus possibly be the attentional costs listeners experience when they hear the different talker in the *Control* condition (Goldinger *et al.*, 1991; Magnuson and Nusbaum, 2007). In experiment 2, the test talker in the *Accent* condition was different from the exposure talker, making this condition more comparable to the *Control* condition than the *Accent* condition in experiment 1, as both conditions now involved a talker switch. The fact that both *Accent* conditions (Exp. 1 and Exp. 2) had equally fast responses during the test block—both being faster than the *Control* condition (also Fig. 5)—suggests that processing slow-downs due to a talker switch are unlikely to explain the results of experiment 1 (or the results of the original CG04 study).

One potential objection to this argument is that participants might have failed to detect the change in talker between the accented exposure and test blocks in experiment 2. On the one hand, the two Mandarin talkers have noticeable differences in their intelligibility, as evidenced by the differences in listeners' performance during the exposure phase in experiments 1 and 2 (see Figs. 5 and 6). This suggests that they might be easily distinguishable. On the other hand, the accented exposure and accented test talkers in experiment 2 were arguably acoustically more similar than the native exposure talker and the accented test talker in experiment 1, and similarity between talkers is known to affect whether listeners notice a change in talker (Magnuson and Nusbaum, 2007; Nygaard *et al.*, 1995). Failure to detect a change from one accented talker to another talker of the same accent (and gender as well as age) might further be exacerbated by participants' lack of familiarity with the accent (Goggin *et al.*, 1991).

TABLE X. Experiment 2, summary of analysis of errors during test.

	Parameter estimates		Wald's test	
Predictors (fixed effects)	Coef $\hat{\beta}$	SE ($\hat{\beta}$)	Z	Pz
(Intercept)	−2.6	0.4	−6.4	<0.001
CONDITION (= *Accent* vs *Control*)	−0.2	0.21	−0.9	0.37

It is therefore possible that listeners were less likely to notice the change in talker in the *Accent* condition in experiment 2, compared to the *Control* condition. This would mean that attentional costs associated with the change in talker in the *Control* condition could theoretically explain the processing slow-down observed for the accented test talker in the *Control* condition. However, several aspects of the present findings argue against this interpretation.

First, the slow-down in processing speed that we observe for the accented test talker in the *Control* condition is too large to be plausibly due to the attentional demands commonly associated with a switch in talker (e.g., McLennan and Luce, 2005; Magnuson and Nusbaum, 2007; Papesh *et al.*, 2016). Previous work has investigated the slow down in processing speed associated with a switch from one native talker to another native talker. These studies have found processing delays of 6%–7% or less associated with a change in talker, compared to the absence of a talker change (e.g., Magnuson and Nusbaum, 2007; McLennan and Luce, 2005; with talker switch costs increasing super-linearly in average RTs). Consider, for example, the word recognition task in Papesh *et al.* (2016). The slow-down they observed for a change in talker was reliably smaller than 100 ms when the average RT was above 1500 ms. Average RTs in the present study were considerably faster at around 1000 ms across conditions (not reported in the text). Yet the slow-down from the exposure to the test block in the *Control* condition was nearly 250 ms (see Fig. 5). That is, although responses were 50% faster in our experiments compared to those of Papesh and colleagues, the slow-down observed when the *Control* condition for the first time encountered an accented talker was 2.5 times *larger* than the talker switch cost observed by Papesh and colleagues. This makes it rather unlikely that the slow-down observed in our *Control* condition (but not our *Accent* condition) is solely driven by attentional costs associated with a talker switch.

Second, whatever attentional costs are associated with a change in talker, they are much smaller compared to practice effects and accent effects. Even for the *Accent* condition in experiment 1, RTs decreased by more than 500 ms from the practice block to exposure block 1, despite the fact that this transition involved a clear change in talker. Similarly, the transition from the test block to the final baseline block resulted in a decrease in RTs (as reflected in positive adjusted mean RTs in the test block, cf. Fig. 5), despite the fact that this transition, too, involved a clear change in talker. Therefore, a change in talker alone does not seem to result in any particularly large slow-down in processing speed (in line

with the previous works cited above). Taken together, this makes it unlikely that the results of experiments 1 and 2 can be reduced to attentional costs associated with a change in talker.

IV. GENERAL DISCUSSION

The current study presents a web-based replication and extension of CG04. We exposed native-English listeners to Mandarin-accented speech and tested their perceptual difficulty with the same talker (experiment 1) or a different talker of the same foreign accent (experiment 2), relative to task control participants. This yielded three major findings.

First, with respect to the primary goal of this study, we replicate the findings of CG04: brief exposure of only a few minutes of foreign-accented speech rapidly attenuated listeners' initial processing difficulty with the non-native accent, as demonstrated by increasingly faster responses in a cross-modal word matching task. Second, in contrast to CG04, which reported no evidence of improvement in processing accuracy, we also found that lab-induced experience with the accent led to fewer errors in the same task within the same short period of exposure. Critically, the enhancements in terms of both processing accuracy and processing speed were promoted by adaptation to the specific accent, above and beyond changes related to task familiarity (*Accent* vs *Control in clear*) or attentional engagement for more effortful listening (*Accent* vs *Control in noise*). Third, this rapid adaptation to accented speech transferred to an unfamiliar talker. Specifically, exposure to one Mandarin-accented talker facilitated the online processing of sentences from another unfamiliar talker. This transfer is of theoretical relevance because it suggests that that the adaptation observed in experiments 1 and 2 is unlikely due to normalization processes driven by low-level auditory properties: normalization (e.g., of pitch range, speaking rates, or spectral energy, Holt and Lotto, 2002; Nearey, 1989; Miller and Volaitis, 1989) tends to operate on even shorter-term time scales, such as a few syllables or words (Reinisch, 2016; Sjerps and Reinisch, 2015; see Weatherholtz and Jaeger, 2016 for a review), and would thus not be expected to transfer to another talker.

Together, our results reinforced CG04's finding that accent adaptation occurs rapidly. Compared to other studies on natural accent adaptation (e.g., Bradlow and Bent, 2008; Sidaras *et al.*, 2009), two important aspects of the present study are that we examined adaptation within a very short time frame, and that we combined offline measures (accuracy) with online RT measures in a single paradigm. We thus discuss the implications of our results in relation to existing findings on accent adaptation. We begin by discussing how our results relate to seemingly conflicting findings (Floccia *et al.*, 2009). Then we relate our findings on rapid accent adaptation to a broader literature that has assessed adaptation across multiple time scales. We point out outstanding questions that remain to be addressed by future studies. These questions are critical for our understanding of the flexibility of human speech perception. In particular, in considering possible mechanisms by which listeners adapt and generalize across talkers, we call attention to a distinction between two theoretically different mechanisms—*model learning* and *model selection*. As we elaborate below, this distinction may guide future experiments to elucidate whether long-term language experience may benefit short-term accent adaptation and whether various forms of adaptation—some occurring within a minute and some unfolding over days or weeks—are really supported by the same mechanism or not.

A. Rapid accent adaptation: The importance of task

To the best of our knowledge, only one other published study has employed a similar paradigm as used by CG04 to investigate the time course of accent adaptation (Floccia *et al.*, 2009). Our results are in conflict with those of Floccia and colleagues. Given the paucity of studies that examine adaptation effects within the few minutes of accent exposure, this conflict deserves attention.

Using a similar length of exposure as in CG04 but a different task, Floccia *et al.* (2009) compared the speed of foreign accent perception (e.g., French-accented English) after exposure to either the same foreign accent (from the same or a different talker), or different varieties of native accented English (both familiar and unfamiliar regional varieties). Participants' response speed to foreign-accented test stimuli was consistently slower than their responses to a familiar accent. Response speed did not improve over time.[6] Thus, unlike the present study, Floccia and colleagues do not find evidence for rapid adaptation to foreign-accented speech. They interpret this null result as substantiating a theoretical distinction between "comprehensibility" and "intelligibility" (Derwing and Munro, 1997). Specifically, Floccia and colleagues argue that "comprehensibility" (as measured by response time, Floccia *et al.*, 2009, p. 381) reflects "pre-lexical processing" and "…seems constantly impaired by the presentation of an unfamiliar accent, as suggested by the long-lasting slowing down of word identification delays," whereas improvements in "intelligibility" (as measured by accuracy, Floccia *et al.*, 2009, p. 380) reflects post-lexical processing that "…can be taught to become more efficient, perhaps by applying a specific phonological accent-filter onto the outcome of lexical activation" (Floccia *et al.*, 2009, p. 402). The latter part of the argument, namely the improvability of intelligibility, is based on evidence from other work that recognition accuracy improves at least over longer (multi-day) exposure to unfamiliar accents (e.g., Bradlow and Bent, 2008; Weil, 2001).

Our results—replicating CG04's—suggest that processing speed is *not* constantly impaired. Rather, we find that the speed of processing foreign-accented speech improves with exposure to such speech. This raises questions about the extent to which the results of Floccia *et al.* (2009) require a theoretical distinction between comprehensibility and intelligibility. We propose that the seemingly inconsistent null result observed by Floccia and colleagues may instead be reconciled if one considers the methodological differences between the studies. Floccia *et al.* (2009) used a lexical decision task. This differs from the cross-modal matching task employed here and in CG04, and this choice of experimental task affects the interpretation of effects on processing speed in at least two ways. First, whether we *expect* accent

adaptation to result in changes in processing speed depends on the task participants were asked to perform. The cross-modal matching task requires more fine-grained phonetic processing of a word than that required by a lexical decision task. For instance, if a production of the word "bed" is confusable with both "bad" and "bid," listeners can still be certain that it is a word and respond promptly in a lexical decision task. For this task, the existence of similar neighbors (bad and bid) thus is expected to *facilitate* processing. Indeed, although we are not aware of studies on accent speech that speak to this prediction, such facilitation has been observed for native speech perception (e.g., Baayen *et al.*, 2006; Balota *et al.*, 2004). In a word matching task, however, the same accented production of bed—which may be easily confusable with both bad and bid—would make the word matching task *harder*, leading to slower RTs and/or lower accuracy. That is, we *a priori* expect the cross-modal word matching paradigm employed in CG04 and the present study to be more sensitive to changes in processing times that are associated with accent adaptation.

Second, the choice of experimental task can affect the *statistical power* to detect an effect—even for the same type of outcome variable. This holds, in particular, for outcome variables that are bounded, like processing speed and accuracy, as statistical power tends to decrease close to those boundaries. For example, the word matching task employed in both the present study and CG04 results in very low error rates, reducing the power to detect effects on accuracy (for relevant power simulations, see Dixon, 2008). Similarly, there is an inherent "soft lower bound" for RTs, reflecting motor planning and other processes that are not affected by accent adaptation. This can make it difficult to detect any effect when processing times are already short. Moreover, increases and decreases in RTs tend to be nonlinear, such that tasks or items that show longer RTs will be inherently more likely to be improved than tasks or items that have shorter RTs (e.g., Wagenmakers *et al.*, 2007). This, too, provides an explanation for the difference in RT patterns between the studies: average RTs in response to foreign-accented speech in the word matching task (here: ~1300 ms across blocks; CG04 did not provide average RTs) were more than 40% longer than that in the lexical decision task (Floccia *et al.*, 2009: 898 ms, p. 397).

We thus tentatively conclude that the null result Floccia and colleagues obtained for the speed of processing foreign-accented speech is a result of the task employed in their study. This also means that neither the present results, nor those of Floccia and colleagues, provide support for the proposed theoretical distinction between comprehensibility and intelligibility. It thus remains an open question whether prelexical processing and lexical processing are differently affected by exposure to a foreign accent, including differences in the malleability of these processes.

B. Accent adaptation at different time scales: The same mechanism?

Behavioral effects of exposure to unusual pronunciations, including accented speech, have been observed at *multiple temporal scales*—from one to two minutes to hours to days. This is illustrated in Fig. 7. Most previous work, however, has focused on adaptation at slower time-scales than the rapid adaptation observed in the present study. This raises a number of questions, which we discuss after providing a brief overview of other lines of research on adaptation at different time scales.

A large body of work has focused on exposure-elicited changes at the level of phonetic categories, which is often referred to as perceptual recalibration. These studies have focused on changes in the perception and categorization of individual sound segments. They have demonstrated that adaptation to talker-specific speech patterns results in altered phonetic boundaries, as typically measured by phonetic categorization responses (e.g., Eisner and McQueen, 2006; Kraljic and Samuel, 2006). For instance, if a talker produces

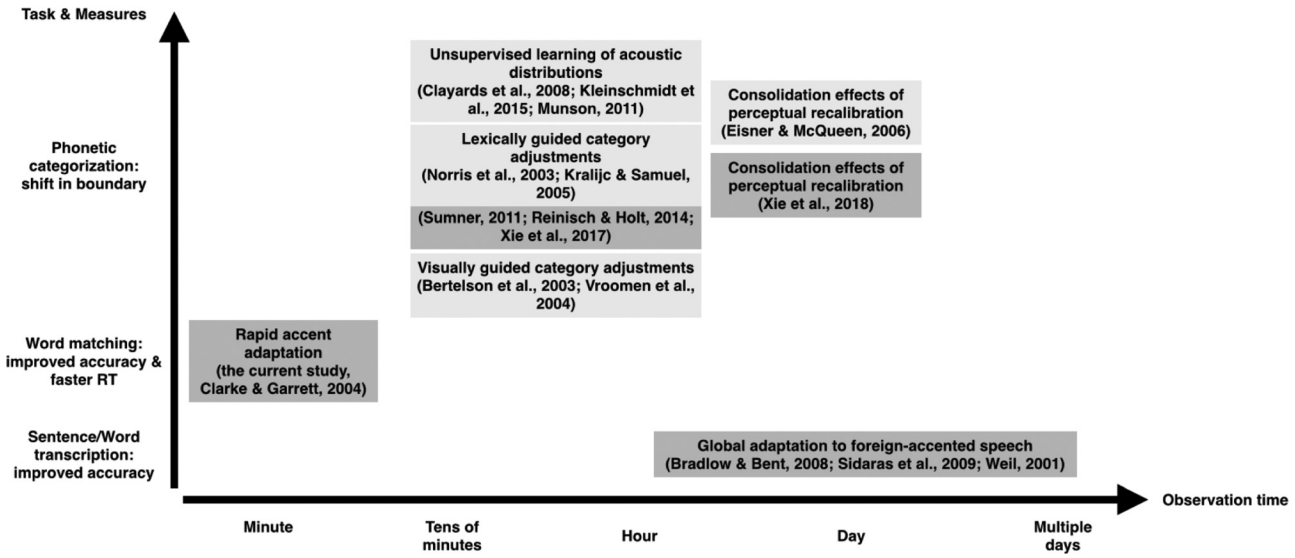

FIG. 7. Illustration of different paradigms and findings demonstrating adaptation to unusual pronunciations, including foreign-accented speech (dark shade) and other types of altered speech (light shade).

perceptually ambiguous /d/s (for instance, due to non-native accents), listeners can use lexical or other contextual information to disambiguate the sound. Critically, listeners also seem to use this information to adjust how they subsequently categorize similarly ambiguous /d/ sounds, even in the absence of disambiguating context. This adjustment results in more accurate recognition of subsequent /d/ pronunciations from the same talker.

Studies in this tradition typically have investigated recalibration after hundreds of exposure trials, lasting dozens of minutes (for a review, Samuel, 2011; though see Kleinschmidt and Jaeger, 2012 and Vroomen *et al.*, 2007 for shorter exposures). Relevant to the present purpose, perceptual recalibration has also been observed when the ambiguous sounds are embedded in foreign-accented speech (Reinisch and Holt, 2014). These and other results have raised the possibility that the mechanisms that underlie boundary shifts are the same that underlie adaptation to foreign accents (Sumner, 2011; Xie *et al.*, 2017). One proposal, for example, holds that both perceptual recalibration and accent adaptation can be understood as a form of distributional learning (Clayards *et al.*, 2008; Kleinschmidt and Jaeger, 2015).

A separate line of work has directly assessed adaptation to globally accented speech, instead of individual segments. Using transcription tasks, they provide evidence that accent adaptation enhances recognition accuracy of a talker's speech globally, not limited to specific sounds or words (e.g., Tzeng *et al.*, 2016; Bradlow and Bent, 2008; Sidaras *et al.*, 2009; Weil, 2001). The line of work has generally examined improvements over longer experiments (Tzeng *et al.*, 2016; Sidaras *et al.*, 2009)—sometimes over multiple days (Bradlow and Bent, 2008; Wade *et al.*, 2007; Weil, 2001)—than employed in the present study.

Compared to these aforementioned studies, our results demonstrate the rapidity of accent adaptation. This raises several questions for future work. First, to what extent does the rapid improvement in the perception of accented speech reflect the same type of changes that has been argued to underlie perceptual adaptation over a longer exposure (such as the studies discussed in the previous paragraphs)? The present results and those of CG04 suggest that rapid adaptation to foreign-accented speech is more than a lowering of decision criterion (to become faster), and more than shifting a boundary between two contrastive categories (since listeners show better recognition for multiple words and sentences overall). But whether the mechanisms that underlie rapid adaptation are the same as those that underlie boundary shifts during perceptual recalibration or increases in transcription accuracy after multi-day exposure is an open question. One way future work can address this question is by testing whether the behavioral improvements in the current word matching task transfer to enhanced performance in other kinds of tasks, such as phonetic boundary shifts, facilitated lexical decisions, and/or increase in transcription accuracy.

A second related question is whether the rapid adaptation effects we observe here would persist over time. The longevity of adaptation effects in general is an understudied topic. Effects of global accent adaptation have been observed across multi-day training sessions (e.g., Bradlow and Bent, 2008; Weil, 2001), but longitudinal studies that extend beyond the training period are lacking. Emerging evidence suggests that the representational changes following perceptual recalibration tend to last for at least a few days without additional exposure to the adapted talker (e.g., Eisner and McQueen, 2006; Xie *et al.*, 2018). Whether the facilitated processing of accented speech observed here after very brief exposure persists over hours and days—as has been observed for perceptual recalibration—or not is an open question.

Another important question is to what extent improvements in recognition are generalizable to other talkers and by what mechanism. Our experiment 2 suggests that rapid accent adaptation *can* transfer to an unfamiliar talker of the same accent, but we do not know the principles by which such transfer occurs. We know of no other published studies on cross-talker generalization following such rapid adaptation to naturally spoken sentences.[7] Previous work has focused on cross-talker generalization at longer time scales, often including exposure over multiple days (e.g., Bradlow and Bent, 2008; Lee *et al.*, 2018; Reinisch and Holt, 2014; Weil, 2001; Xie *et al.*, 2018). It remains to be seen whether rapid adaptation resulting from brief accent exposure is reliably generalizable and whether such generalization has the same empirical signatures as generalization on slower times scales.

C. Mechanisms of talker accent adaptation and generalization

Finally, the rapidity of adaptation observed in the present work raises important questions about the underlying mechanism. Intuitively, cognitive processes that unfold over longer time scales—on the order of minutes or hours or days—are more likely to be taken as evidence of *learning* than effects that elapse within a few seconds (e.g., adjustment for speaking rates). Indeed, the literature on slow adaptive processes has made a number of proposals about the kind of learning mechanism underlying accent adaptation. Some work has asked whether the adaptation reflects the learning of new phonetic representations (e.g., Reinisch *et al.*, 2014; Xie *et al.*, 2017) or just temporary adjustment of listeners' decision criteria (Clarke-Davidson *et al.*, 2008).

Next, we address this question first with regard to the talker-specific adaptation observed in experiment 1 and then with regard to the transfer of adaptation observed in experiment 2. We introduce a novel distinction between *model learning* and *model selection*. As we will argue below, this distinction is of particular relevance when one considers the rapidity of adaptation.

1. Talker-specific adaptation (experiment 1)

Conceptually, we can distinguish two computational mechanisms by which rapid adaptation to a specific accented talker might proceed.

The first possibility, which we will refer to as *model learning*, would be that listeners induce new phonetic representations for that accented talker (for related ideas, see Bradlow and Bent, 2008; Kleinschmidt and Jaeger, 2011; Lancia and Winter, 2013). One way to conceptualize this process of inducing new phonetic representations (or acoustic-to-category

mappings) is as the building of a generative model for a particular talker, a *talker model* (Kleinschmidt and Jaeger, 2015). For the present purpose we adopt this terminology. We note, however, that in other theoretical frameworks, facilitated recognition of the speech from a particular talker may be achieved without assuming an abstract model for a talker. For instance, in exemplar theories, a talker model would be equivalent to a set of exemplars (e.g., Goldinger, 1998; Johnson, 2006). Under either framework, listeners need to learn to relate the input of the previously unfamiliar talker to the newly acquired information about that talker. It is then this type of implicit knowledge about the talker-specific phonetic cue distribution associated with phonological categories and words that enables listeners to achieve more accurate and faster performance during the test phase in experiments like ours.

Another possibility, which we refer to as *model selection*, is that despite our effort to recruit naive listeners, some or all of our participants already had learned a model of Mandarin-accented English or other relevant talker models, based on input previously experienced outside of the laboratory. Following other work on accent adaptation (e.g., Bradlow and Bent, 2008; Clarke and Garrett, 2004), we recruited our participants to be monolingual native-English participants who reported to have no prior experience of Mandarin-accented speech. Despite these self-reported criteria, it is possible, however, that our participants had relevant previous experience, including individual encounters with a Mandarin accent or very similar accents. Rather than to learn a new talker model, the input during the exposure phase of the experiment might then have allowed listeners in the Accent group to *select* the appropriate model (or weighted mixture of models, cf. Kleinschmidt and Jaeger, 2015, pp. 180–181) for the present input.

Model learning and *selection* describe two different (though mutually compatible) possibilities. For example, it is possible that the induction of abstract representations (talker models) requires more time, possibly even specific mechanisms operating during sleep (for related discussion, see Fenn *et al.*, 2013; Tamminen *et al.*, 2012; Xie *et al.*, 2018). This becomes perhaps most apparent when one considers cases in which robust recognition of an accent (or, for that matter, a second language) requires different phonetic features, so that these features need to be *learned*. Indeed, those tend to be the cases that are hard in second language understanding (e.g., /l/-/r/ contrast for Japanese learners; Yamada and Tohkura, 1991; see Bradlow, 2008 for a review) and accent perception (e.g., Arslan and Hansen, 1997), with learning in some cases remaining incomplete even after years of exposure (e.g., Bradlow, 2008; Dufour *et al.*, 2007, 2010). Model selection, on the other hand, intuitively describes a process that could happen over faster time scales. While neither the present results nor previous work allows us to distinguish between model selection and learning, the brevity of exposure in the present experiments and the observation that it is sufficient to elicit adaptation, highlight the need to distinguish between adaptive mechanisms that might be operating at different timescales, jointly contributing to robust speech perception.

2. Generalization to an unfamiliar talker

The distinction between *model learning* and *model selection* also has implications for how we interpret evidence of generalization in experiment 2. The first possibility is the generalization reflects talker-to-talker transfer of learning of specific acoustic-to-category mappings. It implicates that, even with the same exposure talker, the degree to which listeners generalize to a novel test talker depends on the extent to which the test talker resembles the exposure talker in the production of specific sounds, for instance, word-initial /d/s with shorter or longer voice onset times (Reinisch and Holt, 2014; Xie and Myers, 2017).

Another possibility is that the enhanced recognition of the second Mandarin-accented talker is due to improved selection of a talker model, instead of any direct transfer of learning from one talker to another. In other words, hearing the first Mandarin talker may serve to reweight certain previously learned talker models (e.g., re-familiarize participants in the *Accent* condition with the relevant speech properties) such that speech from the second Mandarin talker is more likely to be perceived through the appropriate talker model rather than another model (for example, a non-native talker model instead of a native talker model). In this case, generalization depends on the extent to which an abstract talker model (previously learned outside of the laboratory) is predictive of the idiosyncratic properties of the second talker. Whether listeners rely on individual-to-individual transfer or generalization through an abstract talker-independent model is a question for future research. It is possible that different kinds of generalization may be at play at different stages of accent adaptation as listeners' most recent experience gets integrated with prior experience and as the structure of listeners' prior knowledge changes.

V. CONCLUSION

In sum, the data presented here replicate the core finding of CG04 that initial perceptual difficulty with foreign-accented speech can be attenuated rapidly by a brief period of exposure to an accented talker. It further reveals that both processing accuracy and speed can be enhanced as a result of accent exposure. Finally, the CG04 paradigm provides a valid method to assess talker-specific adaptation and can be adapted to address remaining issues about accent adaptation, including the integration of recent exposure and long-term language experience and the mechanisms of cross-talker generalization.

ACKNOWLEDGMENTS

We thank Ralf Haefner and audiences at *NCUR* 2016 and *LabPhon* 2016 for their valuable feedback. This study would not have been possible without generous financial support by the Office of the Dean for Undergraduate Studies (Richard Feldman) and additional support by NIH R01 Grant No. HD075797 to T.F.J. All views expressed here are those of the authors and do not necessarily reflect those of the funding bodies.

[1]See supplementary material at https://doi.org/10.1121/1.5027410 for (a) the full set of stimulus materials; (b) the complete post-experiment survey; and (c) the full set of parameter estimates for mixed-effects models reported in the text.

[2]Since we used mixed effects models to analyze trial-level data, the adjusted RTs were calculated at the trial level for each participant. This differs from Clarke and Garrett (2004) who analyzed block means for exposure and test and, hence, calculated adjusted RTs at the block level.

[3]One participant had high block-wise error rates ($> 50\%$) across all blocks. We suspect that this participant misused the keys associated with "match" and "mismatch" responses. Excluding this participant did not qualitatively change the statistical results in any of the analyses (including those in the supplementary material[1]) reported in this paper. We thus report the results with this participant included.

[4]All *post hoc* tests reported in this paper were conducted by repeating the mixed effects analysis while re-coding the contrasts for Condition and/or Block to appropriately address the question under discussion. For example, here we re-coded Condition using sum contrast coding with *Control in noise* $= -1$, *Control in clear* $= 1$.

[5]Given that we are comparing exposure effects of this novel Mandarin-accented speaker against that of the native AE speaker in experiment 1, we considered any effect that varying word durations may have on response times. To control for this, we conducted additional analyses by adding the duration of the final word as a predictor of the RTs. There was no effect of word duration on RTs either during exposure or at test (p's > 0.25). Critically, the reported results hold with or without word duration as a predictor. In the main text, we report results without this predictor.

[6]Floccia and colleagues did not report whether the accuracy for the (accented) test talker was affected by exposure condition (presumably because accuracy was at ceiling in all conditions, cf. Floccia *et al.*, 2009, pp. 385).

[7]In her dissertation, Clarke (2003) reported an experiment that is similar to our experiment 2: one Spanish-accented talker appeared in blocks 1–3 and a second Spanish-accented talker appeared in block 4. There was weak evidence of generalization when only the first two trials of block 4 were considered, but no clear generalization to the unfamiliar talker (relative to the control condition) when all six trials of block 4 were considered.

Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (**2009**). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," J. Exp. Psychol. Human Percept. Perform. **35**, 520–529.

Adank, P., and McQueen, J. M. (**2007**). "The effect of an unfamiliar regional accent on spoken word comprehension," Paper presented at the *ICPhS XVI*, Saarbrucken, 6–10 August.

Arslan, L. M., and Hansen, J. H. (**1997**). "A study of temporal features and frequency characteristics in American English foreign accent," J. Acoust. Soc. Am. **102**, 28–40.

Baayen, R. H., Feldman, L. B., and Schreuder, R. (**2006**). "Morphological influences on the recognition of monosyllabic monomorphemic words," J. Mem. Lang. **55**, 290–313.

Baayen, H., Piepenbrock, R., and Gulikers, L. (**1995**). "The CELEX lexical database" [CD-ROM], Linguistic Data Consortium, Philadelphia, PA.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (**2004**). "Visual word recognition of single-syllable words," J. Exp. Psychol. Gen. **133**, 283–316.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck G., and Green, P. (**2017**). "lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1-13)" [computer software], http://CRAN. R-project.org/package_lme4 (Last viewed 8/18/2017).

Boersma, P., and Weenink, D. (**2015**). "Praat: doing phonetics by computer (version 5.4.19) [computer program]," http://www.praat.org/ (Last viewed 9/16/2015).

Bradlow, A. (**2008**). "Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast," in *Phonology and Second Language Acquisition*, edited by J. G. Hansen Edwards and M. L. Zampini (Benjamins, Amsterdam), pp. 287–308.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Clarke, C. M. (**2003**). "Processing time effects of short-term exposure to foreign-accented English," Doctoral dissertation, University of Arizona.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Clarke-Davidson, C. M., Luce, P. A., and Sawusch, J. R. (**2008**). "Does perceptual learning in speech reflect changes in phonetic category representation or decision bias?," Percept. Psychophys. **70**, 604–618.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (**2008**). "Perception of speech reflects optimal use of probabilistic speech cues," Cognition **108**, 804–809.

Dahan, D., and Mead, R. L. (**2010**). "Context-conditioned generalization in adaptation to distorted speech," J. Exp. Psychol. Human Percept. Perform. **36**, 704–728.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Derwing, T. M., and Munro, M. J. (**1997**). "Accent, intelligibility, and comprehensibility: Evidence from four L1s," Stud. Second Lang. Acquis. **19**, 1–16.

Dixon, P. (**2008**). "Models of accuracy in repeated-measures designs," J. Mem. Lang. **59**(4), 447–456.

Dufour, S., Nguyen, N., and Frauenfelder, U. H. (**2007**). "The perception of phonemic contrasts in a non-native dialect," J. Acoust. Soc. Am. **121**, EL131–EL136.

Dufour, S., Nguyen, N., and Frauenfelder, U. H. (**2010**). "Does training on a phonemic contrast absent in the listener's dialect influence word recognition?," J. Acoust. Soc. Am. **128**, EL43–EL48.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. Human Percept. Perform. **23**, 914–927.

Eisner, F., and McQueen, J. M. (**2006**). "Perceptual learning in speech: Stability over time," J. Acoust. Soc. Am. **119**, 1950–1953.

Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (**2013**). "Sleep restores loss of generalized but not rote learning of synthetic speech," Cognition **128**, 280–286.

Flege, J. E., Bohn, O. S., and Jang, S. (**1997**). "Effects of experience on non-native speakers' production and perception of English vowels," J. Phon. **25**, 437–470.

Floccia, C., Butler, J., Goslin, J., and Ellis, L. (**2009**). "Regional and foreign accent processing in English: Can listeners adapt?," J. Psycholing. Res. **38**, 379–412.

Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (**2006**). "Does a regional accent perturb speech processing?," J. Exp. Psychol. Human Percept. Perform. **32**, 1276–1293.

Goggin, J. P., Thompson, C. P., Strube, G., and Simental, L. R. (**1991**). "The role of language familiarity in voice identification," Mem. Cogn. **19**, 448–458.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goldinger, S. D., Pisoni, D. B., and Logan, J. S. (**1991**). "On the nature of talker variability effects on recall of spoken word lists," J. Exp. Psychol. Learn. Mem. Cogn. **17**, 152–162.

Holt, L. L., and Lotto, A. J. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**, 156–169.

Jaeger, T. F. (**2008**). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," J. Mem. Lang. **59**, 434–446.

Jaeger, T. F. (**2011**). "Corpus-based research on language production: Information density and reducible subject relatives," in *Language from a Cognitive Perspective: Grammar, Usage, and Processing. Studies in Honor of Tom Wasow*, edited by E. M. Bender and J. E. Arnold (CSLI Publications, Stanford), pp. 161–197.

Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phon. **34**, 485–499.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Kleinschmidt, D. F., and Jaeger, T. F. (**2012**). "A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 34, No. 34.

Kleinschmidt, D. F., and Jaeger, T. F. (**2015**). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychol. Rev. **122**, 148–203.

Kraljic, T., and Samuel, A. G. (**2006**). "Generalization in perceptual learning for speech," Cogn. Psychon. Bull. Rev. **13**, 262–268.

Kučera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English* (Dartmouth Publishing Group, Sudbury, MA).
Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2016**). "lmerTest: Tests in linear mixed effects models" (R package version 2.0-33) [computer software], retrieved from http://CRAN. R-project.org/package_lmerTest (Last viewed 12/3/2016).
Lancia, L., and Winter, B. (**2013**). "The interaction between competition, learning, and habituation dynamics in speech perception," Lab. Phonol. **4**, 221–257.
Lee, C., Oey, L., Simon, E., Xie, X., and Jaeger, T. F. (**2018**). "How we comprehend foreign-accented speech: Learning to generalize across talkers," Univ. Rochester J. Undergrad. Res., in press.
Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol. Human Percept. Perform. **33**, 391–409.
McLennan, C. T., and Luce, P. A. (**2005**). "Examining the time course of indexical specificity effects in spoken word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **31**, 306–321.
Miller, J. L., and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category," Percept. Psychophys. **46**, 505–512.
Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**(1), 365–378.
Munro, M. J., and Derwing, T. M. (**1995**). "Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech," Lang. Speech **38**, 289–306.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Nusbaum, H. C., and Magnuson, J. (**1997**). "Talker normalization: Phonetic constancy as a cognitive process," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego), pp. 109–132.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1995**). "Effects of stimulus variability on perception and representation of spoken words in memory," Percept. Psychophys. **57**, 989–1001.
Papesh, M. H., Goldinger, S. D., and Hout, M. C. (**2016**). "Eye movements reveal fast, voice-specific priming," J. Exp. Psychol. Gen. **145**, 314–337.
R Core Team (**2014**). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria), http://www.R-project.org/.
Reinisch, E. (**2016**). "Speaker-specific processing and local context information: The case of speaking rate," Appl. Psycholing. **37**, 1397–1415.
Reinisch, E., and Holt, L. L. (**2014**). "Lexically guided phonetic retuning of foreign-accented speech and its generalization," J. Exp. Psychol. Human Percept. Perform. **40**, 539–555.
Reinisch, E., Wozny, D. R., Mitterer, H., and Holt, L. L. (**2014**). "Phonetic category recalibration: What are the categories?," J. Phon. **45**, 91–105.
Samuel, A. G. (**2011**). "Speech perception," Ann. Rev. Psychol. **62**, 49–72.
Sidaras, S. K., Alexander, J. E. D., and Nygaard, L. C. (**2009**). "Perceptual learning of systematic variation in Spanish accented speech," J. Acoust. Soc. Am. **125**, 3306–3316.
Simmons, J. P., Nelson, L. D., and Simonsohn, U. (**2011**). "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant," Psychol. Sci. **22**, 1359–1366.
Sjerps, M. J., and Reinisch, E. (**2015**). "Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception," J. Exp. Psychol. Human Percept. Perform. **41**, 710–722.
Smith, R., Holmes-Elliott, S., Pettinato, M., and Knight, R. A. (**2014**). "Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization," Q. J. Exp. Psychol. **67**(3), 590–608.
Sumner, M. (**2011**). "The role of variation in the perception of accented speech," Cognition **119**, 131–136.
Tamminen, J., Davis, M. H., Merkx, M., and Rastle, K. (**2012**). "The role of memory consolidation in generalisation of new linguistic information," Cognition **125**, 107–112.
Tzeng, C. Y., Alexander, J. E., Sidaras, S. K., and Nygaard, L. C. (**2016**). "The role of training structure in perceptual learning of accented speech," J. Exp. Psychol. Human Percept. Perform. **42**, 1793–1805.
Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., and Bradlow, A. R. (**2010**). "The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles," Lang. Speech **53**, 510–540.
Vroomen, J., van Linden, S., De Gelder, B., and Bertelson, P. (**2007**). "Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses," Neuropsychologia **45**, 572–577.
Wade, T., Jongman, A., and Sereno, J. (**2007**). "Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds," Phonetica **64**, 122–144.
Wagenmakers, E., Brown, S., and Rayner, Keith. (**2007**). "On the linear relation between the mean and the standard deviation of a response time distribution," Psychol. Rev. **114**, 830–841.
Weatherholtz, K., and Jaeger, T. F. (**2016**). "Speech perception and generalization across talkers and accents," in *Eskimo-Aleut* (Oxford Research Encyclopedias, Linguistics).
Weil, S. (**2001**). "Foreign accented speech: Encoding and generalization," J. Acoust. Soc. Am. **109**, 2473(A).
Xie, X., Earle, F. S., and Myers, E. B. (**2018**). "Sleep facilitates generalisation of accent adaptation to a new talker," Lang. Cogn. Neurosci. **33**, 196–210.
Xie, X., and Myers, E. B. (**2017**). "Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers," J. Mem. Lang. **97**, 30–46.
Xie, X., Theodore, R. M., and Myers, E. B. (**2017**). "More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories," J. Exp. Psychol.: Hum. Percept. Perform. **43**, 206–217.
Yamada, R. A., and Tohkura, Y. I. (**1991**). "Age effect on acquisition of non-native phonemes: Perception of English/r/and/l/for native speakers of Japanese," in *Proceedings of the 12th International Congress of Phonetic Sciences*, Vol. 4, pp. 450–453.