languages

MDPI

Article

Sounds over Symbols? The Role of Auditory Cues in Orthographically-Correlated Speech Behavior

Shannon Grippando

Department of Linguistics, University of Arizona, Tucson, AZ 85721, USA; sgrippando@email.arizona.edu

Received: 16 April 2019; Accepted: 8 September 2019; Published: 11 September 2019

Abstract: A recent series of studies found a correlation between orthographic length and speech duration: The more orthographic units in a written form, the longer the speech duration of that word, all else being equal. Modular and encapsulated speech production models argue that orthography should not contribute to articulation when it is not directly and explicitly relevant to speech. Such models demand that other factors such as auditory cues must be contributing to the development of this behavior. If auditory cues are being used in the development of these speech patterns, individuals would be expected to be sensitive to these differences. The current study uses an ABX task to determine whether participants are sensitive to durational differences at lengths similar to those observed in the previously found orthographically-correlated speech behavior. The current results showed no sensitivity to the critical levels of speech duration. Participants only began to show sensitivity at four times the length of the lower-bound durational lengths previously observed in individual's speech patterns. These results call into question whether audio cues are playing a significant role in the development of this speech behavior and strengthen the claim that orthography may be influencing speech in an interactive fashion.

Keywords: orthography; speech duration; psycholinguistics; phonetics; experimental; ABX task; speech perception

1. Introduction

The role of orthography in speech has been a point of uncertainty and contention in psycholinguistics. Encapsulated, modular models of speech production argue that information which is not directly or explicitly relevant to a task will not influence a process (e.g., the WEAVER++ model (Levelt et al. 1999)). In the case of orthography, if written forms are not explicitly present (e.g., an individual is not reading), these types of models predict that orthography will not influence speech behavior since the absence of orthographic forms makes orthography irrelevant to the process of speaking. Alternatively, interactive models of speech allow subsets of language to influence and provide feedback to one another through spreading activation (e.g., Dell 1988; Van Orden and Goldinger 1994). Even if certain types of information are not explicitly relevant to a task, the connections among subsets of information can theoretically influence speech behavior. For example, if the phonological form of a word is activated during speech production, the related orthographic form might consequently be activated. In turn, this activated orthographic information could influence articulation. While interactive models theoretically allow for orthography to influence speech production in this way, even some strongly interactive models do not explicitly detail how orthography might influence speech (e.g., Smith and Goffman 2004).

Some researchers have presented results from reaction time data that they argue is evidence for orthography influencing speech interactively (Damian and Bowers 2003; Rastle et al. 2011). Others contest that they fail to find such results or that the purported orthographic effects are actually a consequence of task or training effects (Alario et al. 2007; Bi et al. 2009; Chen et al. 2002; Mitterer and

Reinisch 2015; Roelofs 2006; Zhang and Damian 2012). However, several recent studies have turned their attention away from using reaction times to determine orthographic influence on the acoustic signal itself. Brewer (2008) observed a relationship between the number of letters in an orthographic form in English and speech duration. For example, given the words *tic*, *click*, and *clique*, the coda is represented by 1, 2, and 3 letters, respectively. Brewer found that as the number of letters in a word-final coda increased, so too did the speech duration of the coda, about 9–29 ms increase per letter. Additionally, an independent whole-word duration effect was observed beyond the durational increase in the coda, about 14–36 ms increase per letter. These effects were found to be independent of word frequency and spelling norm frequency. This behavior was observed in both laboratory experiments in which participants read lists of words off a page and also found in a spoken corpus analysis of spontaneous speech. A similar effect was observed in Japanese: In homophone pairs, the word spelled with more characters was produced significantly longer than the word spelled with fewer characters (Grippando 2018). In another of Brewer (2008)'s experiments, participants read novel English words from a list, such as *vip* and *vipp*. Initially when participants read these unknown words off a page, no difference was found for these pairs of words in either coda or whole-word duration. However, once participants were trained on semantic associations for the novel words with a picture of a novel object, a whole-word duration effect appeared in a picture-naming task: Words more letters in the coda representation (*vipp*) were produced longer overall than words with fewer letters (*vip*). The previously observed coda-level effect was no longer significant. It is important to note that participants received no auditory exemplars for how to pronounce the novel words. The whole-word differences manifested after exposure to differing orthographic forms alone (and semantic associations through training). Brewer argued that orthographic forms must be tied to a stored mental representation for this effect to manifest, and this behavior was not the result of a simple reading effect.

The fact that these durational differences were present in a spoken corpus analysis is strong evidence that this effect is not limited to a reading effect. Additionally, these spoken corpus results are a response to criticisms like those from Mitterer and Reinisch (2015) who have argued that purported orthographic effects only appear in careful, unnatural speech acts. Brewer's novel-word-learning data is also evidence that under certain circumstances, even in the absence of audio cues, orthography alone may influence speech production.

Nonetheless the exact role of orthography and the degree to which it may contribute to these speech patterns still remains in question. In Brewer's novel word learning task, whole-word durations significantly differed but coda durations did not. In her other experiments, these two effects were independent but they always appeared simultaneously. Why would the coda effect not appear in a novel word learning task? One explanation could be that participants were simply not given enough time to properly development strong orthographic representations of the novel stimuli. Indeed, Brewer trained participants in a single session. However, other novel word learning studies have found that other orthographically-driven effects appear immediately after the first day of orthographic training (Rastle et al. 2011). Another answer could be the role of audio cues. Rather than orthography being the sole contributing factor to these durational differences, exposure to auditory exemplars may be a significant contributing factor. While the whole-word duration effect emerged after exposure to differing orthographic forms alone in a novel word learning task, orthographic information alone could be insufficient for the development of the coda-level effect.

There is evidence that seemingly orthographically-related behavior can develop from auditory exposure. Zamuner and Ohala (1999) observed that preliterate children produced consonants significantly longer in words spelled with two medial consonants, like *ballet*, than in words with one medial consonant, like *balance*. Likewise, the consonant tended to be produced as a geminate in words spelled with two medial consonants as opposed to words spelled with one medial consonant. The fact that these children were preliterate necessitates that their behavior must not have originated from orthographic knowledge but must have come from some other source. A likely answer is that they were mimicking adults' speech patterns. Could the speech patterns observed in Brewer (2008)

and Grippando (2018) be developed similarly? Leveltian non-interactive speech production models would necessitate that these durational effects come from relevant and related information, such as auditory cues. While Brewer's whole-word duration effects in her novel experiment are evidence against orthography playing no role, auditory cues could potentially still be a significant factor in the development of this behavior. If this is the case, the lack of auditory exposure for novel words in Brewer's experiments may explain why a coda-length effect did not emerge.

The goal of the current study is to begin to investigate the role of audio in the purported orthographic durational effects observed in studies such as Brewer (2008) and Grippando (2018). For audio to play a significant role in the development of this behavior, individuals would be expected to be sensitive to durational differences at the levels observed in these speech patterns. Humans can distinguish incredibly small differences in auditory duration. A subarea of psychological research called just noticeable differences is devoted to investigating thresholds at which humans perceive changes in stimuli. For just noticeable differences in language, a seminal study by Huggins (1972) found that participants began to report sentences as having unnatural timing once a phonetic segment length is altered by 20 ms. Fujisaki et al. (1975) found just noticeable differences for vowels, plosives, nasals, and fricatives in Japanese as short as 10 ms. Adults have shown sensitivity to voice onset time differences as short as 20 ms and non-linguistic tone duration differences as short as 8 ms (Pakarinen et al. 2007; Ylinen et al. 2006). The durational differences in Brewer were reported to be as short as 9 ms. While 9 ms is in the realm of shortest just noticeable differences described above, there are several key differences between the previous studies and the effect in question. Most importantly, the English segment length differences observed in Brewer involve non-phonemically contrastive segments. Though just noticeable differences were observed in Japanese as short as 10 ms by Fujisaki et al., vowel and consonant length is phonemically contrastive in Japanese which could predispose individuals to finer sensitivity.

The current study seeks to determine whether English-speaking individuals are sensitive to durational differences at the levels observed in the speech behavior reported in Brewer (2008). If participants are sensitive to these differences, audio could be playing a significant role in the development of these purported orthographic speech effects. If this is the case, this may explain why Brewer did not observe a coda level effect in her novel word learning experiment. However, if participants are not sensitive to differences at these levels, this could call into question audio's role in the development of this speech behavior and strengthen claims that orthography may be a dominant contributing factor. An auditory ABX task is used to determine if participants are sensitive to these durational differences.

2. Materials and Methods

60 English-speaking participants were recruited from the University of Arizona's linguistics subject pool. Participants were compensated with 1 experiment credit which was applicable for course credit or extra credit in their University of Arizona classes. Forty-six participants were female, and 9 participants were left handed. Average participant age was 21 (minimum 18, maximum 55). Research with this population was approved by University of Arizona HSPP IRB.

Materials were created from a list of 10 English words and 10 English non-words (see Appendix A for a full list). Brewer (2008) observed durational effects for both words and novel words. However, in a novel-word-learning task, a whole-word duration effect was significant but a coda duration effect was insignificant. To determine whether the lexical status of a word influences sensitivity, both words and non-words were included as items.

All words/non-words were monosyllabic with a CVC structure. All words/non-words began and ended with a nasal (/n/, /m/) or a fricative (/s/, /ʃ/). Brewer (2008) investigated coda-length durations of several categories of sounds, including stops and fricatives. Durational effects were significant for both fricatives and stops. Fricative codas were chosen to represent items with similar phone qualities as those observed in Brewer (2008). Fricatives were chosen over stops because manipulating the duration

of fricatives (described in detail below) involves manipulating a continuous sound. Transformation stops would involve manipulation of relative silence. Additionally, fricatives were chosen over stops to avoid altering over aspects of a phone that might influence perception (such as voice onset time in stops. Words with nasal codas were also included to compare results between sounds Brewer investigated and those she did not. Brewer did not exhaustively investigate durational effects across all possible codas in English. Nasals were used in the current study because they represent a class of sounds that Brewer did not investigate and, like fricatives, are produced continuously.

For each word/non-word, a computer-generated audio base form was created using the *say* command in the Mac OSX terminal with the default male voice (Alex). Computer-generated stimuli were chosen over human-generated stimuli because of quality issues encountered while manipulating human-produced speech. When the duration of human speech was manipulated using PSOLA transformations (described in more detail below), some of the manipulated stimuli had a "tinny" quality or included pops and clicks. In a pilot study, participants reported being able to tell the difference between stimuli pairs because of these non-durational qualities. Manipulating computer-generated stimuli resulted in better quality item pairs that did not have the same issues as the human speech stimuli. The author and several consultants judged that the computer-generated pairs of items did not have any noticeable non-durational differences between them.

The size of durational differences ranged across Brewer (2008)'s experiments and corpus analyses. Across the results, the average effect ranged from about 9–17 ms on average. One analysis showed an effect of 29 ms. Whole-word effects ranged from about 14–36 ms. The effect sizes for whole-word duration and coda duration varied slightly, but effects were generally found around approximately 15 ms and 30 ms for both effects. For consistency, durational manipulations were kept consistent for both whole-word and coda manipulations. The levels of 15, 30, 60, and 250 ms were used for both whole-word and coda manipulations. The 15 and 30 ms conditions were chosen as the critical levels. Fifteen ms was chosen to represent a lower-bound level of effect and 30 ms was chosen to represent an upper bound level of effect consistently found in Brewer's results. If auditory cues are playing a significant role in the development of this behavior, participants should be sensitive to durational differences at these levels. Sixty ms manipulations double the "high" level of effect and were included in a scenario in which participants are not sensitive to the 30 ms condition. Sensitivity at this level would suggest that participants are engaging in the task but not sensitive to the lower critical levels. 250 ms differences were also included. If participants perceive this ABX task to be too difficult or even impossible, they could possibly disengage from the task with a consequential decrease in performance. With the ABX task used in the current experiment, participants only need to perform at significantly above chance to display sensitivity. Even if participants perceive the task to be difficult and do not consciously perceive a correct choice in a trial, subconscious sensitivity can influence results. Thus, it is important to keep participants engaged with the experiment and make them feel like they are capable of performing at least a portion of the tasks. The 250 ms level was included to provide participants with this sense of engagement and act as a check to confirm that participants were performing the experiment adequately.

8 PSOLA transformations were applied to the base form of each word/non-word in Praat (Boersma 2001). A PSOLA transformation is a technique used to manipulate duration or pitch. PSOLA divides a sound into segments and duplicates or deletes segments to achieve the desired length (Moulines and Charpentier 1990). In all cases for the current study, segments were added, not deleted, resulting in the base audio form being manipulated into longer versions. 4 PSOLA transformations were applied to the coda only in increments of 15 ms, 30 ms, 60 ms, and 250 ms. 4 PSOLA transformations were applied to the entire length of a word/non-word in increments of 15 ms, 30 ms, 60 ms, and 250 ms. In addition to the coda-length effect, Brewer (2008) found an independent word-length effect. For example, if a word had a coda-level effect, the whole-word duration also increased beyond what was observed in the coda alone; there was something like a spreading effect across the entire word duration. Brewer did not perform follow-up analyses to determine which parts

of a word increased in these cases. Likewise, Grippando (2018) found whole-word duration effects in Japanese but did not perform follow-up analyses to determine if this effect was localized on certain word positions or phones. Therefore, because of the lack of data about this effect, it was decided to apply a PSOLA transformation equally across the entire word. This type of manipulation may or may not accurately represent the speech behavior observed in Brewer (2008), but it was chosen as a starting point to begin to investigate this effect.

Base forms and manipulated forms were used to create item triplets. In an ABX task, participants are predisposed to choosing the second word (B) over the first word (A) because of recency effects. Therefore, it is important to include all iterations of word pairs AB along with their correct answers X for balancing purposes. In other words, for each word/non-word, there are the items ABA, ABB, BAA, BAB versions for each level of each condition. From the pool of 20 words/non-words at the conditions of focus of manipulation (coda/whole-word) with manipulations at lengths of 15, 30, 60, and 250 ms, this resulted in 640 triplets.

From these stimuli, two lists were created. One list included only manipulations made on codas and one list included only manipulations made across an entire word. Participants performed only one of these two lists. Making this condition between-subjects loses some statistical power but also makes the experiment more manageable for participants. The full list of 640 trials could take participants 1.5–2 h to complete, with potentially diminishing returns from participant performance as the experiment progressed. With coda/whole-word manipulations as a between-subjects condition, participants responded to 320 trials.

Before the experiment began, participants were verbally informed about the task and given instructions. Participants were tested one at a time and wore on-ear headphones. The experiment was run on DMDX software (Forster and Forster 2003). The experiment began with written instructions that reviewed the verbal instructions. 8 practice items preceded the experimental trials. Practice items were composed of words/non-words that were not included in the critical trials. Participants were given a break in the middle of the experiment. The experiment took less than an hour to complete. At the end of the experiment, participants completed a survey with demographic information.

Participants performed an auditory ABX task. Each trial was preceded by ###### displayed in the middle of the screen for 1 s to signal that a new trial was about to begin. Next a triplet of auditory stimuli was played, with each word separated by 500 ms. Each stimulus in a triplet had the same phonological form, but the first two stimuli differed in length (A, B) and the third stimuli (X) was either the first stimulus (A) or second stimulus (B) repeated. After the three stimuli were played another visual cue +++++ was displayed in the middle of the screen indicating that all auditory stimuli for a trial had been presented and that participants may respond. Participants chose whether they thought the third stimulus (X) was either the first (A) or the second (B). They indicated their answer by pressing the left shift key to indicate the first word (A) or the right shift key to indicate the second word (B). Participants were instructed to give their responses as accurately but also as quickly as possible, even if they were not completely sure about which choice was correct. If a participant did not provide a response after 4 s, the trial timed out and the next trial began.

3. Results

Data Analysis

Data from 3 participants was excluded because they did not follow response requirements correctly. This resulted in data from 30 participants in the whole-word manipulation condition and 27 participants in the coda manipulation condition.

R (R Core Team 2018) and lme4 (Bates et al. 2015) were used to analyze the data. For all analyses, alpha was $p < 0.05$. First, a binomial generalized linear mixed effects model was performed to determine which conditions (if any) significantly influenced error rates. Error rate was the dependent variable. Fixed effects included: Length of increased duration (15 ms/30 ms/60 ms/250 ms),

focus of manipulation (whole-word/coda), lexicality (word/nonword), and coda phone (fricative/nasal). A model was initially created with the full random effects structure recommended by Barr et al. (2013). This included by-subjects and by-items random intercepts as well as by-subjects slopes for lexicality, coda phone, and length of duration and by-items slopes for length of duration and focus of manipulation. However, this model failed to converge, which can be a common occurrence with "maximal" random slopes structures. Random slopes were removed from the model one at a time, first by-items and then by-subjects, and the model was tested again. This was continued until the model converged. The final random effects structure included by-subjects and by-items random intercepts and by-subjects random slopes for length of duration.

p-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Lexicality, coda phone, and focus of manipulation did not significantly influence error rates (lexicality: $\chi 2\ (1) = 0.8838$, $p = 0.3472$; coda phone: $\chi 2\ (1) = 0.6401$, $p = 0.4237$; focus of manipulation: $\chi 2\ (1) = 0.008$, $p = 0.9287$). Length of duration significantly influenced error rates ($\chi 2\ (3) = 88.548$, $p < 0.0001$). There were no significant two-way, three-way, or four-way interactions (all $p > 0.1$). Note that the reduced models testing for the significance of lexicality and coda phones failed to converge. In these cases, a separate full model was used that was identical to the full model described above but removed by-subject random slopes for length of duration. All the above analyses were performed again with this model that had no random slopes and no change was observed in significance/insignificance.

The levels for length of duration were examined through inspection of the full model. *Z*-values are given along with estimated *p*-values. Error rates for 15 ms did not significantly differ from 30 ms ($z = -0.004$, $p = 0.997$), but did significantly differ with 60 ms and 250 ms (respectively: $z = -5.709$, $p < 0.0001$; $z = -13.625$, $p < 0.0001$). Error rates for 30 ms did not significantly differ from 15 ms ($z = 0.004$, $p = 0.997$) but did significantly differ from 60 ms and 250 ms (respectively: $z = -5.786$, $p < 0.0001$; $z = -13.760$, $p < 0.0001$). Error rates for 60 ms significantly differed from all levels (15 ms: $z = 5.709$, $p < 0.0001$; 30 ms: $z = 5.787$, $p < 0.0001$; 250 ms: $z = -13.115$, $p < 0.0001$). Error rates for 250 ms significantly different from all levels (15 ms: $z = 13.637$, $p < 0.0001$; 30 ms: 13.783, $p < 0.0001$; 60 ms: $z = 13.129$, $p < 0.0001$).

The results of the above analysis suggest that there is no significant difference in error rates among any of the conditions other than length of duration. To determine whether error rates significantly different from chance (50%), the data was separated by each level of length of duration (15 ms/30 ms/60 ms/250 ms) and analyzed. A binomial generalized linear mixed effect model with no fixed effects (1) was used to analyze the data. The random effects structure included by-subjects and by-items random intercepts, as well was by-subject random slopes for lexicality and coda phone. These results are summarized in Table 1. In summary, the 15 ms and 30 ms levels did not significantly differ from chance. The 60 ms and 250 ms levels did significantly differ from chance.

Table 1. Error rates (%) for length of duration levels. *Z*-scores and estimated *p*-values included below error rates. Error rates that significantly differ from chance marked with * and shaded green.

	15 MS	**30 MS**	**60 MS**	**250 MS**
Error Rate	49 $z = 0.691$ $p = 0.49$	49 $z = -0.521$ $p = 0.602$	40 * $z = -7.134$ $p < 0.0001$	12 * $z = -13.74$ $p < 0.0001$

The above analysis found no significant effect for factors other than length of duration. However, one might reasonably expect that a duration manipulation made on only the coda would have a more pronounced effect than one of a similar duration applied across an entire word, given that a coda manipulation is concentrated to a single phone rather than spread across three phones (in the case of this experiment). Thus, several follow-up analyses were conducted to confirm if error rates significantly differed from chance (50%) at each experimental level. In other words, one analysis for

error rate only used data from 15 ms + word + coda + fricatives items. Another separate analysis only used data from 15 ms + word + coda + nasals items. This was done for all combinations of items. Similar to the previous analysis, a binomial generalized linear mixed effect model with no fixed effects (1) and by-subjects and by-items random intercepts was used to analyze the data. Barr et al. (2013) suggest using random slopes for within-unit conditions. In this analysis, each set of data is broken down to its smallest grouping, factoring out within-unit conditions. Thus, random slopes are not used in this analysis. These results are summarized in Table 2. In summary, across all conditions, error rates for 15 ms and 30 ms did not significantly differ from chance. In all conditions, error rates for 250 ms did significantly differ from chance. For 60 ms, all conditions were significantly different from chance other than whole-word manipulation for nonwords with fricative codas and coda manipulation for real-words with fricative codas.

Table 2. Error rates (%) for length of duration condition separated by experimental conditions. Z-scores and estimated *p*-values included below error rates. Error rates that significantly differ from chance marked with * and shaded green.

			15 MS	**30 MS**	**60 MS**	**250 MS**
Whole Word	**Word**	*Fricative*	52 $z = 0.691$ $p = 0.49$	46 $z = -1.28$ $p = 0.201$	43 * $z = -2.522$ $p = 0.0117$	13 * $z = -7.325$ $p < 0.0001$
		Nasal	48 $z = -1.126$ $p = 0.26$	51 $z = 0.421$ $p = 0.673$	48 * $z = -5.727$ $p < 0.0001$	13 * $z = -9.739$ $p < 0.0001$
	Nonword	*Fricative*	47 $z = -0.87$ $p = 0.384$	49 $z = -0.527$ $p = 0.598$	46 $z = -1.544$ $p = 0.123$	13 * $z = -6.574$ $p < 0.0001$
		Nasal	50 $z = -0.141$ $p = 0.888$	49 $z = -0.326$ $p = 0.744$	41 * $z = -3.847$ $p = 0.00012$	11 * $z = -8.998$ $p < 0.0001$
Coda	**Word**	*Fricative*	48 $z = -0.576$ $p = 0.564$	48 $z = -0.547$ $p = 0.584$	42 $z = -1.725$ $p = 0.0845$	10 * $z = -5.885$ $p < 0.0001$
		Nasal	47 $z = -1.599$ $p = 0.11$	50 $z = 0.064$ $p = 0.949$	38 * $z = -4.813$ $p < 0.0001$	11 * $z = -7.903$ $p < 0.0001$
	Nonword	*Fricative*	52 $z = 0.721$ $p = 0.471$	53 $z = 0.77$ $p = 0.441$	41 * $z = -2.632$ $p = 0.00849$	9 * $z = -5.455$ $p < 0.0001$
		Nasal	52 $z = 0.743$ $p = 0.458$	48 $z = -0.764$ $p = 0.445$	35 * $z = -4.821$ $p < 0.0001$	15 * $z = -6.174$ $p < 0.0001$

In summary, lexicality and focus of manipulation were observed to have no effect on error rates. In all cases, participants were not better than chance for words manipulated by 15 ms and 30 ms. Additionally, participants were consistently better than chance for items manipulated at 250 ms. Participants were better than chance for items manipulated at 60 ms in all cases except for nonwords with fricative codas manipulated at the whole-word level and real-words with fricative codas manipulated at the coda level. In all cases, there is no evidence of sensitivity at the critical levels of 15 ms and 30 ms.

4. Discussion

Brewer (2008) and Grippando (2018) observed durational differences in participants' speech ranging from about 15–30 ms that correlated with orthographic length. The current study found no evidence to support participant sensitivity in sets of items that differed by 15 ms or 30 ms. However, participants were sensitive to manipulations of 60 ms, which is twice the duration of the upper range of the effects observed by Brewer and four times the duration as the lower range.

This result calls into whether audio cues play a significant role at all in the development of these durational speech patterns that have been argued to be correlated with orthography. An ABX task presents a best-case scenario for participants to detect differences between words: Words are presented successively, so participants are able to compare instances back-to-back-to-back. For natural speech in everyday life, a person would rarely be given the opportunity to compare instances of words in immediate and successive fashion like this, especially some words that Brewer used such as *click* and *clique.* Additionally, participants were not tasked with determining which stimulus was longer and which was shorter. Participants simply needed to match the third stimulus with either the first or second. Yet even in this environment and with this simplified task, participants were still not sensitive to the critical levels of 15 ms and 30 ms.

Even at 60 ms, error rates for words with coda fricatives were not significantly different than chance. This is noteworthy because fricatives were one of the categories of phones specifically investigated by Brewer. Coda nasals were included in the current study alongside coda fricatives to investigate whether category of phone affected sensitivity, especially phones that Brewer had not explicitly investigated. Surprisingly, sensitivity for nasals was found in more conditions than fricatives. Granted, p-values for these insignificant fricative conditions were lower than other insignificant conditions, but they were still ultimately insignificant. This points to a threshold of sensitivity to durational differences that is fragile even at 60 ms. Even with manipulations that are four times longer than the lower bound of durational differences that Brewer observed in individuals' speech, participants were still not sensitive to some conditions.

No difference was observed between words and non-words. Brewer (2008) found no durational effects in participants' speech when they read novel words off a page. However, once participants were trained on semantic associations for these novels words, a whole-word effect appeared. She argued that a word must be stored in the mental lexicon for the orthographically-correlated duration effects to appear. However, the current study finds no evidence for lexicality influencing sensitivity to duration.

No difference was found between coda manipulations and whole-word manipulations. One might expect that participants might be more sensitive to a 30 ms manipulation on a coda than a 30 ms manipulation spread across an entire 3-phone word. However, if participants are not sensitive to 15 ms or 30 ms duration manipulations on a coda alone, then it might be reasonable to expect that participants would not be able to detect manipulations made at or below these levels, regardless of whether multiple manipulations at these levels are applied on each phone of a word. If a 30 ms manipulation is applied equally across a 3-phone word, this would result in a 10 ms increase per phone. This is under the 60 ms threshold for sensitivity found in the current experiment. However, phones were not manipulated equally in this fashion in the current study. PSOLA transformations were applied indiscriminately across an entire word for whole-word manipulations, which increased phone length based on ratio of individual phones to entire word duration rather than increase an individual phone a set number of milliseconds. Consequently, a relatively longer phone in a base form was increased more than a relatively shorter phone in the same base form. This was done partially because Brewer (2008) and Grippando (2018) did not conduct segment-level analyses to determine if certain segments of a word were targeted unequally in whole-word duration effects (other than codas by Brewer). A follow-up analysis of the items used in the current study would be beneficial to: (1) Determine the number of milliseconds each phone increased after a whole-word PSOLA transformation was applied; and (2) determine whether there is a threshold for manipulations made on individual phones for whole-word manipulations that correlates with sensitivity. Especially in the longer conditions of 60 ms and 250 ms,

it would be beneficial to determine if sensitivity to items was based on an individual phone crossing a specific durational threshold. Other studies have observed sensitivity at lower levels than the 60 ms threshold in the current experiment. For example, Pakarinen et al. (2007) observed sensitivity to differences as short as 20 ms and Fujisaki et al. (1975) as short at 10 ms. However, both of these studies involved items with phonemically-contrastive durational differences, while the none of the manipulations in the current study are phonemically contrastive. This could explain participants' lower threshold of sensitivity in these studies compared to the current study. In addition, Pakarinen used electroencephalogram (EEG) which is a much more sensitive methodology than the behavioral responses collected in the current study.

It should be noted that the results of the current study should not necessarily be overextended as evidence against all acoustically driven interpretations of orthographically-correlated speech behavior, including the previously discussed results from Zamuner and Ohala (1999). Zamuner & Ohala found that preliterate children's speech mimicked spelling in speech duration and syllabication patterns for word-medial consonants. The distinct populations between their study and the current study of preliterate children and literate adults make comparisons problematic. The current study's primary goal was to investigate speech duration effects observed in Brewer (2008), who only examined adult behavior. Thus, the current study also only focused on an adult population. In addition, Brewer and the current study were concerned with a speech duration contrasts found across the entire length of words and on word-final codas. Zamuner & Ohala focused on a slightly different effect on or across word-medial syllable boundaries. This location may be more salient with a consequently lower threshold of sensitivity. Future research could attempt to replicate the current study's results with a preliterate child population to confirm whether this pattern of sensitivity is limited to adults. Additionally, word-medial consonant durations could be manipulated to determine whether contrasts at this location have a lower threshold of sensitivity than those investigated in the current study.

In conclusion, the current study finds no evidence for sensitivity to durational differences at the levels observed in the speech behavior in Brewer (2008). These results call into question the role of auditory cues in the development of these speech patterns. Modular and encapsulated speech production models argue that spoken language should not be influenced by orthography, especially if orthography is not relevant to a specific task (e.g., Levelt et al. 1999). Thus, such models would necessitate that purported orthographic speech effects are driven by another factor. Sensitivity to auditory cues might be a reasonable explanation for this behavior: An individual hears another person speaking in a certain way and mimics their speech, consciously or unconsciously. In such a case, audio is the predominant contributing factor to this behavior. A weaker version of this audio-driven hypothesis compatible with interactive speech models (e.g., Dell 1988; Van Orden and Goldinger 1994) would allow influence from orthography in the development of these speech patterns but accompanying audio cues might also significantly contribute to development. This was presented as a potential explanation for why Brewer observed a whole-word duration effect but not a coda-level effect in her novel-word-learning experiment: Participants were not exposed to auditory exemplars of the novel words, and thus they could not develop the coda-level effect which relies on the contribution of auditory exemplars. However, the current study finds no evidence to support either the strong or weak versions of this hypothesis given that participants were not sensitive to durational differences at the critical levels of 15 ms and 30 ms. It remains unknown why Brewer did not observe a coda-level effect in her novel-word-learning task. This might be due to participants only receiving a single session's exposure to the novel materials. Future studies should investigate whether a longer training period would result in a coda-level speech duration effect.

On a surface level, future studies could expand upon the current study by determining a more specific threshold of sensitivity. The number of levels for durational differences that could be investigated in the current experiment were limited by the inclusion of other experimental conditions, ABX item balancing requirements, and participant time constraints. Future studies could explore the range between 30 ms and 60 ms to determine if sensitivity exists below

60 ms. Additionally, the current study builds a foundation with behavioral data that could be expanded with brain-imaging data using electroencephalogram (EEG). EEG monitors electrical activity in the brain and uses event-related potentials (ERPs) to measure responses to stimuli. The mismatched negativity (MMN) ERP is associated with sensitivity to auditory changes, including duration (Näätänen et al. 1978; Näätänen and Alho 1997). Eliciting an MMN is extremely robust, with participants not even necessarily needing to be paying explicit attention to the auditory stimuli for a response to be registered. EEG and MMN could be used as a more accurate means to determine if participants are not sensitive to the lower durational differences of 15 ms and 30 ms differences.

Funding: This research received no external funding.

Acknowledgments: The author would like to thank Thomas Bever and the University of Arizona Language and Cognition Lab for providing the space and resources to conduct this study.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

List of words and non-words used to create items in the current experiment.
Words
/mæs/, /mɛʃ/, /mam/, /mun/, /nɛIm/, naIn/, /nun/, /sæʃ/, /sʌm/, /sʌn/,
Non-words
/mis/, /mIm/, /muʃ/, /mʌn/, /nin/, /nɛm/, /nɛn/, /nam/, /sæn/, /sIʃ/.

References

Alario, F.-Xavier, Laetitia Perre, Caroline Castel, and Johannes C. Ziegler. 2007. The role of orthography in speech production revisited. *Cognition* 102: 464–75. [CrossRef] [PubMed]

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef] [PubMed]

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Bi, Yanchao, Tao Wei, Niels Janssen, and Zaizhu Han. 2009. The contribution of orthography to spoken word production: Evidence from Mandarin Chinese. *Psychonomic Bulletin & Review* 16: 555–60.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5: 341–45.

Brewer, Jordan B. 2008. *Phonetic Reflexes of Orthographic Characteristics in Lexical Representation.*. Ann Arbor: ProQuest.

Chen, Jenn-Yeu, Train-Min Chen, and Gary S. Dell. 2002. Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language* 46: 751–81. [CrossRef]

Damian, Markus. F., and Jeffrey. S. Bowers. 2003. Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review* 10: 111–17.

Dell, Gary. S. 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27: 124–42. [CrossRef]

Forster, Kenneth I., and Jonathan C. Forster. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods* 35: 116–24. [CrossRef]

Fujisaki, Hiroya, Kimie Nakamura, and Toshiaki Imoto. 1975. Auditory perception of duration of speech and non-speech stimuli. In *Auditory Analysis and Perception of Speech*. Edited by Gunnar Fant. London: Academic Press, pp. 197–219.

Grippando, Shannon. 2018. More characters, longer speech: Effects from orthographic complexity in Japanese. *Proceedings from Texas Linguistics Society Conference* 17: 27–38.

Huggins, A. William. F. 1972. Just noticeable differences for segment duration in natural speech. *The Journal of the Acoustical Society of America* 51: 1270–78. [CrossRef] [PubMed]

Levelt, Willem J., Ardi Roelofs, and Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1–38. [CrossRef] [PubMed]

Mitterer, Holger, and Eva Reinisch. 2015. Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language* 85: 116–34. [CrossRef]

Moulines, Eric, and Francis Charpentier. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9: 453–67. [CrossRef]

Näätänen, Risto, and Kimmo Alho. 1997. Mismatch negativity–the measure for central sound representation accuracy. *Audiology and Neurotology* 2: 341–53. [CrossRef] [PubMed]

Näätänen, Risto, Anthony W. Gaillard, and Sirkka Mäntysalo. 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* 42: 313–29. [CrossRef]

Pakarinen, Satu, Rika Takegata, Teemu Rinne, Minna Huotilainen, and Risto Näätänen. 2007. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology* 118: 177–85. [CrossRef] [PubMed]

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 10 June 2018).

Rastle, Kathleen, Samantha F. McCormick, Linda Bayliss, and Colin J. Davis. 2011. Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 1588.

Roelofs, Ardi. 2006. The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Rreview* 13: 33–37.

Smith, Anne, and Lisa Goffman. 2004. Interaction of motor and language factors in the development of speech production. *Speech Motor Control in Normal and Disordered Speech* 45: 227–52.

Van Orden, Guy C., and Stephen D. Goldinger. 1994. Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performanc* 20: 1269. [CrossRef]

Ylinen, Sari, Anna Shestakova, Minna Huotilainen, Paavo Alku, and Risto Näätänen. 2006. Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research* 1072: 175–85. [CrossRef] [PubMed]

Zamuner, Tania S., and Diane K. Ohala. 1999. Preliterate children's syllabification of intervocalic consonants. In *Proceedings of the 23rd Annual Boston Conference on Language Development*. Somerville: Cascadilla Press, pp. 753–63.

Zhang, Qingfang, and Markus F. Damian. 2012. Effects of orthography on speech production in Chinese. *Journal of Psycholinguistic Research* 41: 267–83. [CrossRef] [PubMed]