Influence of prior auditory and visual information on speech perception: Evidence from Japanese singleton and geminate words

Yizhen Zhou¹, Yosuke Nakamura²*, Ryoko Mugitani³,⁴ and Junji Watanabe⁴,⁷

¹Department of Information and Communications Engineering, Tokyo Institute of Technology, 2–12–1 Ookayama, Meguro-ku, Tokyo, 152–8550 Japan
²Department of Information Processing, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, 226–8502 Japan
³The Faculty of Integrated Arts and Social Sciences, Japan Women’s University, 1–1–1 Nishiikuta, Kawasaki, 214–8565 Japan
⁴NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3–1 Morinosato Wakamiya, Atsugi, 243–0198 Japan

(Received 17 June 2020, Accepted for publication 2 October 2020)

Abstract: The goal of this study was to demonstrate the influence of prior auditory and visual information on speech perception, using a priming paradigm to investigate the shift in the perceptual boundary of geminate consonants. Although previous research has shown visual information such as photographs influences the perception of spoken words, the effects of auditory and visual (written or illustrated) information have not been directly compared. In the present study, native Japanese speakers judged whether or not a spoken word was a geminate word after hearing/seeing a prime word/pseudoword that contained either singleton or geminate feature. The results indicate the spoken words, written words and even illustrations presented prior to the target sounds, can guide boundary shift for Japanese geminate perception. Significantly, the influence of auditory information is independent of the lexical status of the primes, that is, both word and pseudoword auditory primes with geminate sound features induced a significant bias. On the other hand, visual primes induced the bias only when the primes coincided lexically with the targets, indicating the influence of visual information on geminate perception is different from auditory information.

Keywords: Speech perception, Geminate perception, Spoken word processing, Duration

1. INTRODUCTION
1.1. Multisensory Nature of Speech

It has been argued that speech perception has a multimodal nature wherein listeners may use speech information from visual modality. For instance, it has been well established that visual information regarding facial detail influences speech perception and helps listeners distinguish sounds degraded with noise [1–4]. The famous McGurk effect and its subsequent studies have demonstrated that speech perception is not only an auditory process, as the perception of internal phonetic category is also sensitive to visual information [5–7]. Moreover, visual information of orthographic representations also influences speech perception. There is increasing evidence from cross-modal studies that visually presented words can influence spoken word processing [8,9].

Many cross-modal studies have used priming paradigm to investigate the multisensory nature of speech. In the priming paradigm, a prime and a target are presented in close temporal succession and participants perform a task (e.g., lexical decision or categorization) on the target. Some interesting empirical findings have been yielded in cross-modal priming using written word primes. For instance, Kouider and Dupoux [8] showed that written words primed identical auditory target words only when the primes were consciously processed, i.e., the priming effects were observed with primes having longer exposure durations (67 ms), but absent when prime durations were at 33 or 50 ms. This result can be explained by bimodal interactive activation model [9,10]. In this model, there are lexical and sublexical links between phonological and
orthographic representations (see Fig. 1). It is suggested that orthographic lexical representations are activated when a masked visual prime is presented for 33 ms, but not phonological lexical representations; no cross-modal priming effect is induced. However, when the masked prime duration is extended longer (67 ms), phonological lexical representations are also activated, allowing the visual identical prime to activate its auditory target; this induces the cross-modal priming effect.

Although previous research has revealed the influence of visual information, e.g. written words, on spoken word processing using repetition priming, uncertainty remains regarding how visual information is accessed and organized under multimodal interaction. Particularly, the effect of non-orthographic information (e.g., illustrations or photographs) has not been compared with the equivalent spoken or written words. The illustrations may serve as alternative representations of the identical words. They can evoke semantic processing as it is activated in processing of the texts [11]. Moreover, it has been demonstrated that a sequence of photographs primed target spoken words in an identification task when the context and target stimuli were identical [12]. Therefore, it is possible that the consciously perceived visual information in the form of illustration influences the spoken word processing, in a similar way as the identical written words do. However, whether the illustrations induce the priming effect has not yet been investigated.

In the present study, we further investigated the influence of visual information on speech perception by investigating the effects of orthographic (written words) and non-orthographic visual primes (illustrations), and also comparing them with the effect of auditory primes (spoken words) to examine the modality differences. In addition, our study verified whether pseudoword primes produce the priming effect. According to the bimodal interactive activation model, visual pseudoword primes may also induce a cross-modal priming effect because there is a sublexical interface between phonological and orthographic representations.

Moreover, most previous studies have typically used lexical decision task. In the lexical decision task, participants judge whether a target stimulus is a word or not, and the reaction times and error rates are measured. The task is specifically effective for testing the particular process engaged in evaluating the lexical status of stimuli. However, given the assumption that visual information plays a role during spoken word processing, one would expect priming effects to be observed even in perceptual tasks such as categorization that reflects a direct perceptual change in target processing other than reaction time measures, but few studies have focused on this point. Given that, we used a categorization task in this study. The effect size is measured using shift size in perceptual boundary of Japanese geminates.

1.2. Geminate Categorization

Japanese, like many other languages, exhibits a duration contrast in consonants [13]. The long consonants, i.e., geminates (known as “sokuon” in Japanese), are substantially longer than short consonants (singletons). Many studies of geminates have examined acoustic correlates of geminates (e.g., [14]) and found that duration is the primary correlate of singleton/geminate distinction. In the Japanese language, there are two acoustical types of geminates: silent (e.g., /tt/, /kk/) and fricative geminates (e.g., /ss/). Native Japanese speakers have a specific tendency when listening to geminates: they share an abstract representation of geminates similar to a silence [15,16]. This tendency is likely influenced by Japanese phonology and orthography, and the native speakers’ linguistic knowledge is heavily influenced by orthography more than non-native speakers [17]. From the viewpoint of phonology, Japanese has long been asserted as a mora-timed language, and hence the geminates in Japanese are moraic. The closure duration in a silent geminate or sustained frication in a fricative geminate aligns with the duration of a mora (i.e., a perceptual unit of timing) [18,19]. For example, a word containing a geminate like /batta:/ (“batter”) has four moras, while /bata:/ (“butter”) containing no geminate is a three-mora word. With regards to orthography, the geminates are represented with “ッ” (small “tsu”) regardless of the acoustical types. In Japanese, the occurrence of silent geminates is much more frequent than fricative geminates. The percentage of silent geminates (76 percent) was much greater than fricative geminates (24 percent) among 2,476 geminate consonants appeared in the 5th edition of Iwanami Japanese dictionary [20]. The present study therefore focuses on the perception
of silent geminates (referred to as geminates in the rest of this paper).

The most important perceptual cue used by Japanese speakers to distinguish between singletons and geminates is the closure duration [21]. The longer the closure duration, the more likely the target stimulus to be perceived as a geminate word [22–25]. The perceptual studies of geminates have usually determined the perceptual boundary between singletons and geminates, i.e., the durational value that corresponds to the categorical shift in perception. A shift in the perceptual boundary of geminates can indicate an influence on geminate perception. For instance, it has been demonstrated that geminate perception is influenced by preceding vowel duration [23] or speech rate at sentence level [26]. Similarly, we assessed the influence of auditory and visual primes by determining the shift in the perceptual boundary of target words preceded by auditory and visual primes. The target words used in this study were a minimal pair of words that has the same phonemes but different closure durations: the singleton word /bata:/ ("butter") and the geminate word /batta:/ ("batter").

In the present study, we evaluated the effects of identical primes (experiment 1) and pseudoword primes (experiment 2) presented in auditory and visual modality on the perception of spoken words. The pseudowords primes were also a minimal pair (/tepa/ and /teppa/). To directly investigate the perceptual change caused by primes, we adopted the priming paradigm in a categorization task of geminates: participants judged whether a target word was a geminate or not after perceiving a prime consciously. The present study extends previous studies of cross-modal priming in spoken word processing by comparing the effects of auditory and visually (written/illustrated) presented primes directly.

2. METHOD

2.1. Participants

Twenty-four paid participants (17 females) took part in experiment 1. In experiment 2, 22 participants (16 females) including 17 people from experiment 1 participated. The age range of the participants was 20–40 years. All the participants were native Japanese speakers, and they had normal or corrected-to-normal vision and no hearing impairments. The experiments were approved by the Ethics Committee of NTT Communication Science Laboratories and were conducted in accordance with the ethical standards set down in the 2013 Declaration of Helsinki. Written informed consent was obtained from all the participants in advance.

2.2. Apparatus

The experiments were conducted in NTT Communication Science Laboratories, Japan. The stimulus presentation and response collection were controlled with a personal computer (Mac Pro) running MATLAB 8.5 with PsychToolbox 3. The participants were seated in front of a 13.3’ Mac Pro computer display with a resolution of 2,560 × 1,600 pixels and a refresh rate of 60 Hz. The built-in speaker of the Mac Pro presented all the auditory stimuli, and the average sound levels were normalized. The participants responded by using the computer keyboard (‘←’ and ‘→’ keys). The experiment was conducted in a dark and quiet room.

2.3. Stimuli

The sound stimuli (/bata/, /batta/; /tepa/, /teppa/) were recorded by a professional native Japanese narrator in a soundproof room at a sampling frequency of 44.1 kHz. We extracted the intervals /ba/ and /ta/ from the base stimulus of singleton /bata/ and then inserted the silent intervals between each of the /ba/ and the following /ta/. The closure durations between /ba/ and /ta/ were edited to range from 130 to 210 ms in 10-ms steps to create a /ba/-/ta/ continuum of nine auditory stimuli (see Fig. 2 for the waveforms). Before the experiment, we asked 27 participants to identify the stimuli without presenting any prime. Each stimulus was randomly repeated 25 times. The proportion of judging 130 ms and 210 ms as geminate was 0.296 and 99.1 percent, respectively. This means that participants could surely perceive the stimulus containing the closure duration of 130 ms as /bata/, and 210 ms as /batta/. All nine auditory stimuli were used as the target

Fig. 2 Waveforms of auditory targets /bata/ and /batta/.

38
stimuli. For prime stimuli, the spoken word primes were either /bata:/ (closure duration of 130 ms between /ba/ and /ta:/) or /batta:/ (210-ms closure duration) in experiment 1, and either /tepa/ (130 ms closure duration between /te/ and /pa/) or /teppa/ (210-ms closure duration) in experiment 2. Figures 3(a) and 3(b) show the written word and illustration primes used in experiment 1. The written word primes used in experiments 2 are shown in Fig. 3(c). The stimuli of all the written words/pseudowords were shown as 490 × 490 pixels images (the size of the characters “バター”, “バッター”, “テパ” and “テッパ” was 302 × 105, 382 × 105, 202 × 105, and 288 × 105 pixels, respectively). The size of the illustrations used in experiment 1 was 710 × 709 pixels.

2.4. Procedure

In experiment 1, the participants performed a categorization task to identify the target auditory stimuli (one of nine closure durations) as either singleton word /bata:/ or geminate word /batta:/ . The participants were seated in front of a computer and instructed to fixate the center of the screen and respond to the target stimuli as accurately as possible by pressing one of two response keys (the “singleton” and “geminate” keys) on the computer keyboard. Three experimental conditions were created: (1) a spoken-word prime condition, (2) written-word prime condition, and (3) illustration prime condition. Each experimental condition involved 450 trials (see Fig. 4 for the trial structure), and these trials were separated into five blocks (90 trials per block). In each block, all the prime-target pairs (2 categories (singleton or geminate) of primes × 9 types of targets) were randomly repeated five times. Therefore, to complete all three conditions, each participant took part in a total of 1,350 trials (90 trials × 5 blocks × 3 conditions). Under the first condition, the spoken word primes (either /bata:/ with a 130-ms closure duration or /batta:/ with a 210-ms closure duration) were presented prior to the presentation of the target stimuli. The durations of the word /bata:/ and /batta:/ were 660 and 740 ms, respectively. Each trial began with a white noise.
sound (in a 20 Hz to 20 kHz frequency range) to remove the effect of the previous trial. Under the second condition, the prime was either the written word /bata:/ or /batta:/. The written word primes were presented randomly for 700 ms (i.e., the average duration of the spoken word primes) prior to the presentation of the target stimuli. To eliminate the effect of the previous trial, each trial began with a noise image (gray mosaic). Under the third condition, the primes were changed to the illustration /bata:/ (butter) or /batta:/ (batter). The procedure was exactly the same as for the written word primes. The order in which each condition (5 blocks in succession) was presented was counterbalanced across participants. Participants took a 25-minute rest between each condition.

In experiment 2, using pseudowords as the primes, the participants performed the same categorization task as in experiment 1. Two conditions were created, and each condition involved 450 trials. Under the first condition, targets were preceded by auditory stimuli consisting of either /tepa/ (130-ms closure duration) or /teppa/ (210-ms closure duration). The durations of the pseudowords /tepa/ and /teppa/ were 660 and 740 ms, respectively. Under the second condition, the primes consisted of the written pseudoword /tepa/ or /teppa/ and they were presented randomly for 700 ms. The order in which conditions 1 and 2 were first presented was counterbalanced across all the participants.

2.5. Data Analysis

The participants’ judgments of the target stimuli were used in the analysis. The proportions of the target stimuli judged to be the geminate word /batta:/ were calculated for each prime category (singleton and geminate) to fit a psychometric function. Each psychometric function was fitted by a cumulative logistic distribution function to the data with the maximum likelihood method to estimate the 50% response point, i.e., the point of subjective equality (PSE). Therefore, the closure duration at this PSE is the perceptual boundary of the singleton word /bata:/ and the geminate word /batta:/. Figure 5 shows a representative example of the priming effect. The solid and dashed lines in the figure are the psychometric functions for the spoken-word prime condition obtained for participant M.Y in experiment 1. Circles and triangles show the raw data; solid and dashed lines show the fitted psychometric curves.

2.5. Data Analysis

The participants’ judgments of the target stimuli were used in the analysis. The proportions of the target stimuli judged to be the geminate word /batta:/ were calculated for each prime category (singleton and geminate) to fit a psychometric function. Each psychometric function was fitted by a cumulative logistic distribution function to the data with the maximum likelihood method to estimate the 50% response point, i.e., the point of subjective equality (PSE). Therefore, the closure duration at this PSE is the perceptual boundary of the singleton word /bata:/ and the geminate word /batta:/.

Figure 6 shows the results indicating the group means of the closure durations at the PSE for the three conditions. The prime stimuli in the two categories are indicated by the bar color, with light and dark bars representing the prime stimuli of singleton /bata:/ and geminate /batta:/, respectively. The error bars show the standard error of the mean. * indicates the statistical significance of $p < 0.05$, and **** indicates the statistical significance of $p < 0.0001$ (three-way repeated measures ANOVA). $N = 24$.
variable, we conducted a two-way repeated measures analysis of variance (ANOVA) using SPSS Statistics software. The within participants factors were the category of primes (singleton word /bata:/; geminate word /batta:/) and conditions (spoken word primes, written word primes, illustration primes). The main effect of the prime category was significant \( F(1,23) = 20.622, p = 0.000, \eta^2 = 0.473 \). The main effect was also observed between conditions \( F(1,4.45,33.228) = 7.528, p = 0.005, \eta^2 = 0.247 \). No significant interaction was observed between prime category and condition \( F(1.429,32.876) = 3.064, p = 0.076, \eta^2 = 0.118 \). The results indicate that when the prime stimuli of /batta:/ were presented, the participants were more likely to judge the target stimulus as the geminate word /batta:/; and vice versa when /bata:/ was presented. Bonferroni’s method for adjusted pairwise comparison revealed that the perceptual boundary of the auditory (spoken word) prime condition was significantly larger than that of visual (written word/illustration) conditions [for the auditory and visual (written word) prime condition, \( p = 0.024 \); for the auditory and visual (illustration) prime condition, \( p = 0.021 \); for the visual (written word) and visual (illustration) prime condition, \( p = 1.000 \)]. The ANOVA results were further supported by a Wilcoxon signed-rank test (geminate primes induced a smaller perceptual boundary than singleton primes among 18, 19 and 15 out of 24 participants under the spoken-word prime, written-word prime and illustration prime condition, respectively. For the spoken-word prime condition: \( Z = -3.021, p = 0.003, \eta^2 = 0.617 \); for the written-word prime condition: \( Z = -2.623, p = 0.008, \eta^2 = 0.535 \); for the illustration condition: \( Z = -2.721, p = 0.007, \eta^2 = 0.555 \). In other words, a target stimulus with the same closure duration was more likely to be judged as the singleton word /bata:/ in the auditory condition than in the visual condition, as the perceptual boundary in the auditory condition was larger than that for the visual condition.

The slopes of the fitted functions obtained from the 24 participants’ averaged responses were the following: (1) singleton prime: 0.0325, geminate prime: 0.0307; (2) singleton prime: 0.029, geminate prime: 0.0288; and (3) singleton prime: 0.0275, geminate prime: 0.0285, for the spoken-word prime, written-word prime, and illustration prime condition, respectively.

### 3.2. Experiment 2

The average closure durations at the PSE for the two conditions are shown in Fig. 7. Using the closure duration boundary as the dependent variable, we conducted a two-way repeated measures ANOVA with the category of primes (singleton pseudoword /tepa/; geminate pseudoword /teppa/) and conditions (spoken pseudoword primes; written pseudoword primes) as the within participants factors. We found a significant main effect of the prime category \( F(1, 21) = 39.096, p = 0.000, \eta^2 = 0.651 \). The main effect of the condition was not significant \( F(1, 21) = 2.380, p = 0.138, \eta^2 = 0.102 \). Furthermore, there was a significant interaction between prime category and condition \( F(1, 21) = 15.830, p = 0.001, \eta^2 = 0.430 \). We tested for simple main effects between prime category under the spoken-pseudoword prime and written-pseudoword prime conditions to understand the nature of the interaction [for the spoken-pseudoword prime condition, \( F(1, 21) = 45.568, p = 0.000, \eta^2 = 0.685 \); for the written-pseudoword prime condition, \( F(1, 21) = 2.000, p = 0.172, \eta^2 = 0.087 \)]. A Wilcoxon signed-rank test confirmed the results of the ANOVA (21 out of 22 participants had a smaller perceptual boundary when they perceived geminate primes than singleton primes under the spoken pseudoword condition: \( Z = -4.575, p = 0.000, \eta^2 = 0.975 \); 9 out of 22 participants had a smaller perceptual boundary when geminate primes were perceived under the written pseudoword condition, \( Z = -0.928, p = 0.354, \eta^2 = 0.198 \). These results indicate that for auditory conditions alone, the participants were more likely to judge the target stimuli as geminate /batta:/ when the prime stimuli of the geminate pseudoword /teppa/ were presented, and they were more likely to judge the target stimuli as the singleton /bata:/ when the singleton pseudoword /tepa/ was presented. That is, the perceptual boundary of geminate/singleton words differed only as regards the spoken pseudoword primes.

The slopes of the fitted functions obtained from the 22 participants’ averaged responses were the following:
For experiment 1, we constructed a full model (with fixed effects of prime category and modality), and compared the reduced models against it to specifically examine the effects of prime category and modality. Likelihood ratio tests confirmed that reduced model lacking the effect of prime category and modality (likelihood ratio test comparing the full model and the reduced model without the interaction: $\chi^2 = 3.435, p = 0.064$). In experiment 2, there was a significant interaction between prime modality and modality (likelihood ratio test comparing the full model and the reduced model lacking the interaction: $\chi^2 = 4.217, p = 0.040$). Post-hoc comparisons (Bonferroni-correction) indicate that only primes presented in auditory modality induced the priming effect ($\beta = 0.05030, p = 0.0027$). Similar to the ANOVA results, the primes presented in auditory modality influenced the judgments of singleton/geminate targets differently compared to primes presented in visual modality. In experiment 1, although the participants were more likely to perceive the targets as geminates when geminate primes were presented both in auditory and visual modality, a target was more likely to be perceived as a geminate word if the word primes were presented in visual modality. When pseudoword primes were used in experiment 2, only primes presented in auditory modality induced the priming effect. The full model results are summarized in Tables 1 and 2.

### Table 1 The results from the linear mixed effect model (Exp. 1).

| Fixed effects   | B    | Std. Error | T     | p    |
|-----------------|------|------------|-------|------|
| (Intercept)     | 0.50704 | 0.13248     | 3.827 | 0.00367 |
| Modality        | 0.02519 | 0.01274     | 1.976 | 0.04834 |
| Prime category  | -0.06250 | 0.01472     | -4.254 | 0.00002 |
| Modality:       | 0.03343 | 0.03343     | 1.855 | 0.06387 |
| Random effects  | SL | Variance | Std. Dev. |
| Participant ID  | (Intercept) | 0.00889 | 0.09427 |
| Target stimuli  | (Intercept) | 0.15366 | 0.39199 |
| Residual        | 0.02339 | 0.15292    |
| Model           | AIC | BIC | logLik | deviance | df. resid |
|                 | -1.043.1 | -1.006.9 | 528.6 | -1.057.1 | 1.289 |

### Table 2 The results from the linear mixed effect model (Exp. 2).

| Fixed effects   | B    | Std. Error | T     | p    |
|-----------------|------|------------|-------|------|
| (Intercept)     | 0.51393 | 0.13443     | 3.823 | 0.00360 |
| Modality        | -0.00222 | 0.01424    | -0.156 | 0.87604 |
| Prime category  | -0.05030 | 0.01424     | -3.532 | 0.00044 |
| Modality:       | 0.04141 | 0.02014     | 2.056 | 0.04009 |
| Random effects  | SL | Variance | Std. Dev. |
| Participant ID  | (Intercept) | 0.01162 | 0.1078 |
| Target stimuli  | (Intercept) | 0.15698 | 0.3962 |
| Residual        | 0.02008 | 0.1417     |
| Model           | AIC | BIC | logLik | deviance | df. resid |
|                 | -710.1 | -677.3 | 362.0 | -724.1 | 785 |

(1) singleton prime: 0.0271, geminate prime: 0.0288; and (2) singleton prime: 0.0313, geminate prime: 0.0312, for the spoken-pseudoword prime and written-pseudoword prime condition, respectively.

### 3.3. Linear Mixed Effect Model Results

The linear mixed model analysis was conducted to assess the fixed effects of prime category and modality. For experiment 1, we constructed a full model (with fixed effects of prime category and modality), and compared the reduced models against it to specifically examine the effects of the prime category and modality. Likelihood ratio tests confirmed that reduced model lacking the effect of prime category significantly differed from the full model ($\chi^2 = 25.688, p = 0.000$). The effect of modality also had a clear impact on responses of geminates ($\chi^2 = 24.813, p = 0.000$). There was no significant interaction between geminate category and modality (likelihood ratio test comparing the full model and the reduced model without the interaction: $\chi^2 = 3.435, p = 0.064$). In experiment 2, there was a significant interaction between prime modality and modality (likelihood ratio test comparing the full model and the reduced model lacking the interaction: $\chi^2 = 4.217, p = 0.040$). Post-hoc comparisons (Bonferroni-correction) indicate that only primes presented in auditory modality induced the priming effect ($\beta = 0.05030, p = 0.0027$). Similar to the ANOVA results, the primes presented in auditory modality influenced the judgments of singleton/geminate targets differently compared to primes presented in visual modality. In experiment 1, although the participants were more likely to perceive the targets as geminates when geminate primes were presented both in auditory and visual modality, a target was more likely to be perceived as a geminate word if the word primes were presented in visual modality. When pseudoword primes were used in experiment 2, only primes presented in auditory modality induced the priming effect. The full model results are summarized in Tables 1 and 2.

### 4. DISCUSSION

The present study investigated the influence of auditory and visual primes on Japanese geminate perception. We designed a categorization task for Japanese singleton/geminate words with a priming methodology and estimated the changes in the perceptual boundary of the singleton/geminate words. The results of experiment 1 demonstrate that singleton and geminate primes presented in both auditory and visual modality, even in the form of illustration, induced a perceptual bias in geminate perception. Singleton word primes caused an increase in perceptual boundary of closure duration, and the geminate word primes shortened the perceptual boundary. In other words, participants were more likely to judge the target word as a geminate word when they heard a geminate-featured prime. However, the influence of prior auditory and visual information on geminate perception may differ. The perceptual bias induced by auditory primes was different compared with written and illustrated visual primes in experiment 1. Moreover, the perceptual boundary of the target words changed significantly no matter if the prime was identical word (experiment 1) or pseudoword (experiment 2) with auditory primes, whereas only the real word primes led to the priming effect with visual (written word) primes. The information of the closure duration presented in the visual modality also influences the geminate perception, but it is dependent on the lexical status of the visual primes.

The difference between the influence of visual and auditory primes was also supported by the psychometric
functions obtained for singleton and geminate primes from the averaged responses of all participants. In experiment 1, the fitted functions shifted horizontally for all three conditions. In contrast, only the function of spoken-pseudoword prime condition shifted in experiment 2. For written-pseudoword prime condition, the singleton function coincided with the geminate function. Because the slopes of the fitted functions were similar, they indicate the degree of consistency of categorization and the precision of the perceptual boundary between singleton and geminate (i.e., how rapidly the perception changes) were similar within each condition.

Previous cross-modal studies investigating the influence of visual information on spoken word processing have focused on repetition priming using written word primes (e.g., [8,9]). In contrast, our results show that visual information in the form of illustration also influences geminate perception when their meanings are matched. It provides evidence that illustration may evoke the related phonological representation, and hence listeners use it to inform the perception of subsequent auditory targets. In addition, our study not only revealed illustrations directly influence the perception of spoken words just as photographs do [12], but also demonstrated illustration primes generated significant bias just as written word primes did. The phonological representations may be activated by written words and illustrations in a similar way, and their influence on spoken word perception was a result of a match in phonological representations between primes and targets. The bias induced across modalities is different than within auditory modality, and it may be because there is a larger number of paths of mappings involved across modalities (see Fig. 1). Alternative to the assumption that illustrations evoked phonological representations, another possible interpretation of the priming effect found in illustrations is that it may be induced by the shared semantic information between the primes and the targets. A strong priming effect has been demonstrated in the recognition of auditory targets (e.g., salt) preceded by semantically-related visual primes (e.g., pepper) [28]. It indicates that semantically-related words can prime target words across modalities, although in our study the primes and targets are more than semantically “related”: they contain identical meanings.

For auditory primes, a significant bias was observed in perception of closure duration induced by both words and pseudowords. It is consistent with previous research showing that responses to the target words are facilitated when the auditory primes and targets were phonologically similar [29–31]. The results also indicate the boundary shift in geminate perception is independent of the lexical status of the auditory primes. In other words, the priming effect induced by auditory primes appears to have a prelexical locus. The target word processing is thus biased because its recognition involves the use of prelexical units that were already activated during the auditory prime processing.

In recent years, several studies have reported that orthography influences spoken word processing. These studies have demonstrated orthographic effects in lexical decision and shadowing tasks (e.g., [32–34]). These results are reconciled with a theoretical framework (e.g., the bimodal interactive activation model) that predicts orthographic effects in spoken word processing, as the connections between phonology and orthography is bidirectional at both lexical and sublexical levels. The bimodal interactive activation model predicts that orthography influences the perception of spoken words in “real-time” when a word is heard regardless of its lexicality. Given that, we should have observed a boundary shift even with the written pseudoword primes in experiment 2. However, our result was not as expected; the influence of orthographic information presented by visual primes seems relatively less important compared to phonological information. This is in good agreement with previous work that there is an overall primacy of phonological information in spoken word processing [35]. In that study, it has been demonstrated that phonological relatedness dominates in spoken word recognition and produces substantial priming effects. Orthographic relatedness alone, however, did not show significant priming effect in auditory lexical decision or naming task.

Our finding is also supported by the study which has compared orthographic effects in different spoken word recognition tasks. Ziegler, Ferrand and Montant [36] found that orthographic effects are stronger in lexical decision than in other tasks such as rime detection task, and that the shadowing task has the weakest effect. In fact, there are also certain shadowing tasks for which no orthographic effect has been reported [37,38]. These results showed a decrease in orthographic effects when the task relies less on accessing the lexical representations. Taken together, the disappearance of visual effect in experiment 2 indicates that the influence of orthography may have been embodied during the literacy acquisition (i.e., learning to read and spell), and thus partial phonological representations may have been shaped for the real words that speakers have learnt. As a result, participants easily succeeded to form and access the phonological representations of words in experiment 1 but they failed to do so for the pseudowords in experiment 2. In general, the influence of orthography on spoken word processing is a possible account for the present results although they are not easily accommodated in the bimodal interactive activation model.

Alternative to the influence of orthography, there is another possible account for the effects we observed, that is, the phonological activation in visual modality may
occur at different speed in perception depending on the lexicality of the word. The orthographic information acquired from visual primes may be converted into a phonological code before presentation of the auditory target. Thus, orthography may not be activated during target processing that only phonological processing is involved in the judgement of singleton/geminates. Therefore, the results that significant priming effect was found in experiment 1 using written word primes, but not in experiment 2 using written pseudoword primes may result from faster phonological encoding of the prime when corresponding to words than when corresponding to pseudowords. The repeated exposure to words may enhance the strength of distributed and interactive connections between phonological, orthographic and semantic representations [39], and hence induces a lexicality effect in spelling-sound conversion. To verify the orthographic influence in spoken word processing, the time course of converting orthographic representations into phonological codes by written words in the task should be investigated.

Clearly, more work is needed to determine the influence of orthography using perceptual tasks. For example, the influence of the degree of the stimuli’s neighborhood should be addressed in future research. The loanwords in Japanese or readable written representations of geminates may also be used to examine the orthographic effect on geminate perception. In addition, the robustness of the priming effect can be assessed by placing carrier sentences before and after the target stimuli to regulate the speech rate within the participants.

To summarize, the present experiments were designed to determine the influence of prior auditory and visual information on geminate perception. As observed here, the closure duration presented in the form of auditory signal, visual orthography (written word) and even non-orthographic visual information (illustration), can guide boundary shift for Japanese geminate perception. In addition, our results indicate that the phonological information presented in auditory and visual modalities are processed differently: the prelexical phonological information is activated when mapping auditory signal and mental lexicon but the visual orthography may only activate lexical phonological information; the auditory information is critical for inducing a perceptual change in geminates, and the visual information appears to have a different impact that lexicality is required for information presented in the visual modality.

CONFLICTS OF INTEREST

Author JW is an employee of NTT Communication Science Laboratories, which is the basic-science research section of Nippon Telegraph and Telephone Corporation. Author RM was an employee of NTT Communication Science Laboratories. Author YN is employed by Yahoo Japan Cooperation. There are no patents, products in development or marketed products to declare. The other author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

[1] A. B. Buchwald, S. J. Winters and D. B. Pisoni, “Visual speech primes open-set recognition of spoken words,” Lang. Cogn. Process., 24, 580–610 (2009).
[2] N. P. Erber, “Interaction of audition and vision in the recognition of oral speech stimuli,” J. Speech Lang. Hear. Res., 12, 423–425 (1969).
[3] W. H. Sumby and I. Pollack, “Visual contribution to speech intelligibility in noise,” J. Acoust. Soc. Am., 26, 212–215 (1954).
[4] A. Q. Summerfield, “Use of visual information in phonetic perception,” Phonetica, 36, 314–331 (1979).
[5] L. Brancacio, J. L. Miller and M. A. Paré, “Visual influences on the internal structure of phonetic categories,” Percept. Psychophys., 65, 591–601 (2003).
[6] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” Nature, 264, 746–748 (1976).
[7] K. Tippana, “What is the McGurk effect?” Front. Psychol., 5, 725 (2014).
[8] S. Kouider and E. Dupoux, “A functional disconnection between spoken and visual word recognition: Evidence from unconscious priming,” Cognition, 82, B35–B49 (2001).
[9] J. Grainger, K. Diependaele, E. Spinelli, L. Ferrand and F. Farioli, “Masked repetition and phonological priming within and across modalities,” J. Exp. Psychol. Learn. Mem. Cogn., 29, 1256–1269 (2003).
[10] J. Grainger and L. Ferrand, “Phonology and orthography in visual word recognition: Effects of masked homophone primes,” J. Mem. Lang., 33, 218–233 (1994).
[11] Y. Zhou, H. N. Ho and J. Watanabe, “Perceptual-semantic congruency facilitates semantic discrimination of thermal qualities,” Front. Psychol., 8, 2113 (2017).
[12] S. Shigeno, “Effects of auditory and visual priming on identification of spoken words,” Percept. Mot. Skills, 124, 529–623 (2017).
[13] P. Ladefoged and I. Maddieson, The Sounds of the World’s Languages (Blackwell, London, 1996).
[14] K. Idemaru and S. Guion, “Acoustic covariants of length contrast in Japanese stops,” J. Int. Phon. Assoc., 38, 167–186 (2008).
[15] M. Sadakata, M. Shingai, A. Brandmeyer and K. Sekiyama, “Perception of the moraic obstruent /Q/: A cross-linguistic study,” Proc. 13th Annu. Conf. Int. Speech Commun. Assoc., Sep 9–13, Portland, OR, pp. 891–894 (2012).
[16] M. Sadakata, M. Shingai, S. Sulpizio, A. Brandmeyer and K. Sekiyama, “Language specific listening of Japanese geminate consonants: A cross-linguistic study,” Front. Psychol., 11(5), 1422 (2014).
[17] A. Nagano and M. Shimada, “Morphological theory and orthography: A representation of lexemes,” J. Linguist., 50, 323–364 (2014).
[18] T. J. Vance, An Introduction to Japanese Phonology (State University of New York Press, Albany, NY, 1987).
[19] M. Han, “Acoustic manifestations of mora timing in Japanese,” J. Acoust. Soc. Am., 96, 73–82 (1994).
[20] W. Takahashi, “Nihongo-no sokuon /Q/ nitsuite (Japanese
Y. ZHOU et al.: AUDIO-VISUAL EFFECTS ON GEMINATE PERCEPTION

moraic obstruent /Q/),” J. Fac. Educ. Shinshu Univ., 95, 59–68 (1998).
[21] S. Kawahara, “The phonetics of sokuon, obstruent geminates,”
The Handbook of Japanese Language and Linguistics: Phonetics and Phonology,
in H. Kubozono, Ed. (Mouton De Gruyter, Berlin, 2015), pp. 43–73.
[22] S. Amano and Y. Hirata, “Perception and production boundaries between single and geminate stops in Japanese,” J. Acoust. Soc. Am., 128, 2049–2058 (2010).
[23] J. Kingston, S. Kawahara, D. Chambless, D. Mash and E. Brenner-Alsop, “Contextual effects on the perception of duration,” J. Phon., 37, 297–320 (2009).
[24] R. Oba, A. Braun and J. Handke, “The perception of Japanese geminates by native and non-native listeners,” Phonetician, 92, 9–29 (2005).
[25] S. Watanabe and N. Hirato, “The relation between the perceptual boundary of voiceless plosives and their moraic counterparts and the duration of the preceding vowels,” Onsei Gengo, 1, 1–8 (1985).
[26] Y. Hirata and J. Whiton, “Effects of speech rate on the singleton/geminate distinction in Japanese,” J. Acoust. Soc. Am., 118, 1647–1660 (2005).
[27] A. Kuznetsova, P. B. Brockhoff and R. H. B. Christensen, “lmerTest Package: Tests in linear mixed effects models,” J. Stat. Softw., 82, 1–26 (2017).
[28] J. E. Anderson and P. J. Holcomb, “Auditory and visual semantic priming using different stimulus onset asynchronies: An event related brain potential study,” Psychophysiology, 32, 177–190 (1995).
[29] S. D. Goldinger, P. A. Luce, D. B. Pisoni and J. K. Marcario, “Form-based priming in spoken word recognition: The roles of competition and bias,” J. Exp. Psychol. Learn. Mem. Cogn., 18, 1211–1238 (1992).
[30] M. Radeau, J. Morais and J. Segui, “Phonological priming between monosyllabic spoken words,” J. Exp. Psychol. Hum. Percept. Perform., 21, 1297–1311 (1995).
[31] L. M. Slowiaczek, H. C. Nusbaum and D. B. Pisoni, “Phonological priming in auditory word recognition,” J. Exp. Psychol. Learn. Mem. Cogn., 13, 64–75 (1987).
[32] C. Chéreau, M. G. Gaskell and N. Dumay, “Reading spoken words: Orthographic effects in auditory priming,” Cognition, 102, 341–360 (2007).
[33] M. Taft, A. Castles, C. Davis, G. Lazendic and M. Nguyen-Hoan, “Automatic activation of orthography in spoken word recognition: Pseudohomograph priming,” J. Mem. Lang., 58, 366–379 (2008).
[34] J. C. Ziegler, M. Muneaux and J. Grainger, “Neighborhood effect in auditory word recognition: Phonological competition and orthographic facilitation,” J. Mem. Lang., 48, 779–793 (2003).
[35] M. Burton, A. Jongman and J. Sereno, “Phonological and orthographic priming effects in auditory and visual word recognition,” Work. Pap. Cornell Phon. Lab., 11, 17–41 (1996).
[36] J. C. Ziegler, L. Ferrand and M. Montant, “Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks,” Mem. Cognit., 32, 732–741 (2004).
[37] C. Pattamadilok, J. Morais, P. Ventura and R. Kolinsky, “The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French,” Lang. Cogn. Process., 22, 700–726 (2007).
[38] P. Ventura, J. Morais, C. Pattamadilok and R. Kolinsky, “The locus of the orthographic consistency effect in auditory word recognition,” Lang. Cogn. Process., 19, 57–95 (2004).
[39] M. S. Seidenberg and J. L. McClelland, “A distributed, developmental model of word recognition and naming,” Psychol. Rev., 96, 523–568 (1989).