THE EFFECT OF AUDITORY-VISUAL INFORMATION AND ORTHOGRAPHIC BACKGROUND IN L2 ACQUISITION

V. Doğu Erdener & Denis K. Burnham
MARCS Auditory Laboratories
University of Western Sydney, Australia
d.erdener@uws.edu.au

ABSTRACT
Visual information from the lips and face is an integral part of speech perception. In addition, orthography can play a role in disambiguating the speech signal in foreign/second language (L2) perception and production. The current study investigates the effect of auditory and visual speech information and orthographic depth, the degree to which a language is transparent (high phoneme-grapheme correspondence), or opaque (low phoneme-grapheme correspondence) on L2 acquisition. Speakers of Turkish and Australian English (transparent and opaque orthographies, respectively) were tested for their production of legal non-words in Spanish and Irish (transparent and opaque orthographies, respectively). Transparent orthographic input (Spanish) enhanced pronunciation in L2, and orthographic reproduction. Native speaker ratings of the participants’ productions also revealed that orthographic input improves accent. Overall results confirm previous findings that visual information enhances speech perception and production, and extend previous results to show the facilitative effects of orthographic input in L2 acquisition under certain conditions.

1. INTRODUCTION
It is now well established that speech perception is not solely an auditory phenomenon. In particular visual information by means of lip and face movements has been shown to be an integral part of speech processing, when available. As early as 1954, it was shown that in noisy conditions visual information is used to disambiguate incoming speech signals [1]. However, the most noted demonstration of the role of vision in speech perception is the McGurk Effect [2], in which dubbing of a speaker’s lip movements for [ga] onto the auditory input [ba] results in a fused percept [da]. This original study by McGurk and MacDonald [2] also showed that adult listeners experience a greater McGurk effect, i.e., more visual influence in speech perception than do young children, suggesting that auditory-visual (AV) speech perception is affected by linguistic experience. However, subsequent evidence suggests that infants as young as 5 months also perceive the McGurk effect, suggesting that experience augments an existing ability [3].

The possible facilitative effect of visual information in the perception and production of non-native speech (L2) has been investigated in a number of studies. In 1970, it was shown that explicit articulatory training in the production of non-native sounds, which involves a certain degree of “visuality”, reveals better performance than auditory training alone [4]. Visual information has also been shown to be effective for speakers of one language, e.g., English, processing unfamiliar languages, such as French, German [5], and Korean [6,7].

While researchers have been working on the effect of visual information in L2 processing, the effect of orthographic background in L2 acquisition has been relatively unexplored. One study by Massaro and colleagues [8] showed that text versions of speech stimuli were effective in facilitating perception of non-native speech sounds. An unpublished study by Barton, Cohen and Massaro [9] yielded further evidence that orthographic information has a facilitative effect on auditory perception. Frost and his colleagues also found that when auditory input is masked by noise, unfamiliar utterances were perceived more clearly when orthographic representations are provided [10].

The current study aims to investigate the effect of both visual information and orthography on the production of L2 words. Participants are presented with four stimulus conditions, auditory only, auditory plus orthography, auditory-visual, and auditory-visual plus orthography. Orthographic depth of the participants’ background language and of the target stimuli were manipulated orthogonally as shown in Table 1. An opaque orthography is one in which phoneme-to-grapheme correspondences are obscure and inconsistent, while in a transparent orthography there are one-to-one phoneme-grapheme correspondences and consistency is the rule. Here both the orthographic experience of the participants (opaque-English, transparent-Turkish) and the orthography to be learned (opaque-Irish, transparent-Spanish) were manipulated. As shown in Table 1, it is hypothesised that when there is orthographic input Turkish speakers will produce Spanish stimuli more accurately than their Australian counterparts, and Australian speakers will produce Irish stimuli more accurately than their Turkish counterparts.

Table 1: Background and Target Languages, Orthographic Depth, and Resultant Hypotheses for the Effect of Orthography on Speech Perception.
2. THE EXPERIMENT: L2 NON-WORD NAMING AND WRITING

2.1. Method

Participants and Design. A total of sixty-four participants were tested, 32 monolingual speakers of Istanbul Turkish (17 females and 15 males, M_age=33.25), and 32 Australian English (22 females and 10 males, M_age=25.66). Monolinguality was required in order to control for and avoid any confounding effect of previous exposure to an L2. Participants were required (a) not to have been exposed to an L2 for any longer than a single six-month period and that in any such period there was no interaction with the native speakers and/or no immersion; (b) not to have spent over three months in non-English-speaking (Australian speakers) and non-Turkish-speaking (Turkish speakers) country; and (c) to be literate only in their respective L1. The Australian speakers were Introductory Psychology and postgraduate students at the University of Western Sydney. The Turkish speakers were obtained through the first author’s contacts and from the staff of the Department of Psychology at Boğaziçi University in Istanbul.

Stimuli and Apparatus. Forty-eight Spanish and 48 Irish non-words were created in accord with the phonotactic and orthographic rules of the languages. The legality of the non-words was verified by two linguists with extensive expertise in Irish and Spanish phonologies, and two native speakers. Non-words were used in order that phonological characteristics could be controlled, and to prevent any previous accidental encounter with the words. The stimuli comprised of an equal number of CVC- and CVCV-context non-words. Spanish items made up the transparent and Irish items the opaque stimuli. Exceptional characters and diacritics such as ñ and ú were excluded.

The stimuli were recorded in a sound-attenuated room at MARCS Laboratories. One male native speaker of Chilian Spanish and one male native speaker of Irish recorded the stimuli. The speaker sat 1.5 metres in front of a video camera (Sony Camcorder DSR-PD100P) against a blue background. The sound was captured with a separate microphone (RODE NT 2) attached to the video camera via a DAT recorder (TASCAM DA-P1), which was used as a phantom power source. The microphone was placed in the speakers’ sagittal plane, but below the visual field of the camera.

The speakers were asked to maintain a neutral facial expression and keep head movements to a minimum. Only the bottom part of each speaker’s face (the laryngeal area, lips, jaws and up to the nose level, excluding the eyes) was recorded to ensure that participants paid attention to these specific areas where orofacial movements are clearly visible. Previous studies of foreign word repetition have found that the lower provides more information than the upper face [7]. Video recordings of each stimulus were converted to individual Quick Time format video files, using Adobe Premier, and transferred to a Dell computer, equipped with an Intel Pentium II microprocessor and 192 MB RAM, on which the experiment was programmed and run.

Procedure. Participants were individually tested in a quiet room. All Australian participants were tested in a sound-attenuated booth at the MARCS Laboratories in Sydney, and the Turkish participants were tested in an experimental room in the Psychology Department at Boğaziçi University in Istanbul. Each participant was seated in front of a computer display unit and asked to wear a head-mounted microphone (AKG C-420) and a set of headphones (AKG K270). They were instructed to attend to the monitor throughout the experiment.

The main task for the participants was to name the stimuli presented in each of four conditions: Auditory-only (Aud-only), auditory-visual (AV), auditory-orthographic (Aud-Orth), and auditory-visual-orthographic (AV-Orth). These four conditions were prepared and manipulated using DMDX software [11]. Each experimental session was preceded by a training session, in which the participants were familiarised with the task via the presentation of three example trials appropriate for the condition. All participants completed each of the four conditions Aud-only, AV, Aud-Orth, and AV-Orth, and were randomly allocated to one of 16 possible stimulus sets to control order and stimulus effects via a rolling design method. In each condition, each item was presented twice, once in each of two separate blocks and participants were asked to pronounce each word as accurately as possible. In addition, in the Aud-Orth and AV-Orth conditions, the orthographic stimulus was presented simultaneously with auditory and/or visual components. In these conditions, participants were asked to write down the non-words to the best of their ability on a response sheet. This was done to ensure that participants were paying attention to the orthographic representations.

2.2. Non-Word Naming and Writing Results

Phoneme Errors. Phoneme errors were calculated by counting phoneme addition, deletion and replacement errors, and mean percent phoneme errors are shown in Figure 1. The data were analysed via a 2 x 2 x 4 (Orthographic experience, x target language x experimental condition) ANOVA. Results reveal that visual information reduced phoneme errors [F (1,61) = 18.894; MS_t=2.731, p < .01], irrespective of background or target language. There was no overall group effect for orthography, yet Turkish speakers made significantly fewer errors with Spanish stimuli than their Australian counterparts [F (1,61) = 4.3; MS_t=3.888, p < .05].

![Figure 1: Mean phoneme errors (+SE) by Turkish (n=32) and Australian (n=32) participants across the experimental conditions and target languages, Spanish and Irish](image-url)
Writing Task. Mean percent orthographic errors are presented in Figure 2. Errors were measured by comparing the phonemes of non-words produced by the participants with the original orthographic presentation, and counting the deletion, addition and replacement errors. There was a significant difference between the responses to Spanish and Irish stimuli \[F (1,62)=59.646; \text{MS}_E=21.129; p< .003\], showing that fewer errors were made in response to Spanish words. There was also a difference between the Turkish and Australian speakers with respect to their responses to Spanish stimuli \[F (1,62)= 5.597; \text{MS}_E=21.129; p < .05\], showing that, as for the phoneme analysis, Turkish speakers made fewer errors than their Australian counterparts on Spanish items. On the other hand, Australian participants made fewer errors than Turkish speakers did on Irish stimuli, showing evidence for a matching effect for depth of native orthography, and depth of the orthography of the target language.

Figure 2: Mean number of orthographic errors (+SE) made in the writing task by Turkish (n=32) and Australian (n=32) speakers in the AV-Orth and Aud-Orth conditions.

3. NATIVE SPEAKER RATING PROCEDURE AND RESULTS

The native speaker rating (NSR) procedure was designed as an additional measure to investigate the extent to which visual and orthographic input affects the spoken accent of participants learning words in new languages.

3.1. Method

Raters and Design. The original speakers of the Irish and Spanish stimuli were recruited to conduct the ratings. (See 2.1; Apparatus and Design). The original talkers were recruited because the stimuli to be rated were all non-words, and these two speakers had had extensive exposure to and experience with these stimulus items in the course of selecting and recording the experimental stimuli. The task of the raters was to score each utterance on a 9-point scale by comparing it to its original.

Stimuli and Apparatus. Stimuli were the recordings obtained from Australian and Turkish participants in the main experiment. Each non-word utterance was saved as a separate sound (wave format) file to be played back on a rating program written using DMDX. The stimuli were presented on a PC equipped with an Intel Celeron™ 450 MHz microprocessor and 191 MB RAM. The raters were required to wear a set of headphones (Sennheizer HD 450) and they used the number pad on the keyboard to respond.

Procedure. For each non-word a separate DMDX program was written. Each program featured the utterances by all of the Australian and Turkish speakers from the main experiment. Raters were instructed to score each utterance from 1(poorly pronounced, definitely non-native) to 9 (perfectly pronounced, definitely native utterances), and to use the complete range of the scale. In each trial the original stimulus as pronounced by the rater preceded the subjects’ utterances and following that the 9-point scale followed on the screen. The items were presented in random order and the raters were not informed as to the language background of the speakers.

3.2 Native Speaker Rating Results

Overall native speaker rating scores averaged across conditions and speakers are summarised in Figure 3. There was a main effect of language background group \[F (1,60)= 9.634; \text{MS}_E=0.769; p < .003\], showing that Australian speakers’ accents were rated more highly than their Turkish counterparts. There was a main effect of target language in that accents in Irish were rated to be more native-like than in Spanish \[F (1,60)= 11.876; \text{MS}_E=0.202; p < .001\]. There was no significant effect of visual information, (p>.5) nor were any group interactions with respect to visual information (p > .6). However, there was a significant interaction of orthographic information \[F (1,60)= 15.875; \text{MS}_E=0.390; p < .001\] and a language background \[F (1,60)= 4.806; p <. 05\], revealing that productions by Australian speakers in orthographic conditions (Aud-Orth & AV-Orth) attracted higher native speaker ratings than productions by Turkish speakers.

Figure 3: Mean Native Speaker Ratings (+SE) for accent for Turkish (n=32) and Australian (n=32) speakers.
4. DISCUSSION & CONCLUSION

The results support previous findings that the provision of visual information enhances speech production in L2. In general, the results naming and writing tasks show that Turkish speakers make fewer phoneme errors in the production of Spanish non-words. As predicted, when orthographic input was provided, Turkish speakers made fewer phoneme errors in the production and writing of Spanish than Irish items. On the other hand, Australian speakers were relatively better at naming and writing Irish stimuli than were their Turkish counterparts. This lends support to the notion that specific experience with an orthography of a certain depth facilitates processing of words from an unfamiliar language with a similar depth. It is possible that speakers of languages with opaque orthographies, like English, develop a "picture-orthographic" representation of word structures. Thus the results suggest that orthographic input is useful in teaching pronunciation, provided that the target orthography is transparent. When the target orthography is opaque it seems better not to provide the learners with orthographic input, at least in the initial stages of L2 acquisition, and especially if they have experience only with a transparent orthography.

It is rather puzzling that Australian speakers were rated to have a more native-like accent than their Turkish counterparts. No effect of visual information was found in orthographic conditions. Australian participants were perceived to have more native-like accents than Turkish speakers for both Irish and Spanish stimuli. It is possible that the raters' immersion in Australian English (both had resided in Australia for over ten years) may have biased them towards favouring the Australian speakers as more native-like in their ratings. There is evidence that the degree of L1 use affects perceived accent in L2 [12]. The dynamics of this interaction may be bi-directional and immersion in L2 may also affect perceived accent in L1. In future studies, it would be advisable to control for this factor, and to include a larger number of native language raters from L1 environments.

Over and above these problems, an interesting finding in these studies was that visual information reduced phoneme errors, but not rated accent. Thus it appears that native-like accent is more than just producing correct phonemes. More research is required to identify the vital features of accent.

In summary, the results of this study show that orthographic background affects the processing of spoken language input in L2 acquisition, at least in the very early stages and at the level of individual words. Further longitudinal studies are required with words in sentences. The results lend support to previous findings that visual information is integrally involved in speech perception, and also assists the production of L2 speech.

The current results have implications for issues in second language acquisition and auditory-visual speech processing. Traditionally in L2 teaching there is extensive reliance upon text and auditory training. In practical terms, the results of this study pinpoint the importance of visual information and depending on the nature of the target language, orthographic input in L2 acquisition. There is now growing evidence that visual information enhances L2 perception and production [4,5,6,7]. The current study provides evidence that inclusion of orthographic input in the acquisition of some languages, but not others, may assist learners to pronounce new vocabulary items. Further research is definitely required using both real and non-word stimuli from different languages with varying degrees of orthographic depth.

5. REFERENCES