Language Learning 55:2, June 2005, pp. 191–228

The Role of Audiovisual Speech and Orthographic Information in Nonnative Speech Production

V. Doğu Erdener and Denis K. Burnham
University of Western Sydney

Visual information from the face is an integral part of speech perception. Additionally, orthography can play a role in disambiguating the speech signal in nonnative speech. This study investigates the effect of audiovisual speech information and orthography on nonnative speech. Particularly, orthographic depth is of interest. Turkish (transparent) and Australian English (opaque) speakers

---

V. Doğu Erdener and Denis K. Burnham, MARCS Auditory Laboratories.

This article is based on a master's thesis by the first author submitted to the University of Western Sydney (UWS). The preparation of this manuscript was supported by a postgraduate publications incentive granted to the first author by UWS. The authors would like to thank Michael Tyler and Colin Schoknecht from MARCS Auditory Laboratories, Ayhan Koç and Göklem Tekdemir from Boğaziçi University, Istanbul, and Tülin Duru and Nermin Kemikoğlu for their assistance in the research process. Also, invaluable comments from Heather Winskel and four anonymous reviewers are greatly appreciated. We also thank Bruno di Biase and Malcolm Johnston from UWS and Antony Green from Potsdam University, Berlin, for their advice on stimuli. The authors are also grateful, for their assistance in stimulus selection and production, to Tomas de Bhaldraithe and Jorge Segovia. Earlier versions of some of the data obtained in this research were presented at the 28th Annual Australian Experimental Psychology Conference (Melbourne, 2001) and the 7th International Conference on Spoken Language Processing (Denver, 2002).

Correspondence concerning this article should be addressed to Doğu Erdener, MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797, Penrith DC NSW 1797, Australia. Internet: d.erdener@uws.edu.au

were tested for their production of nonwords in Spanish (transparent) and Irish (opaque). We found that transparent orthography enhanced pronunciation and orthographic responses. Results confirm previous findings that visual information enhances speech production and extend them to show the facilitative effects of orthography under certain conditions. Implications are discussed in relation to audiovisual speech perception and orthographic processing and practical considerations such as second language instruction.

Infants' early language-general ability to perceptually discriminate most, if not all, of the world's speech contrasts is well documented. However, by adulthood, these speech perception abilities are reorganized as a result of exposure to the native language. Burnham and his colleagues (Burnham, Tyler, & Horlyck, 2002) indicate four periods over which this organization occurs. The final period, the orthographic period, which occurs between the ages of 6 and 8, is of interest here and appears to be related to the onset of reading, and more specifically, to the effect of orthography on speech perception (Burnham, Earnshaw, & Quinn, 1987; Burnham et al., 2002). In essence, Burnham claims that perception of native and nonnative speech contrasts is sharpened and attenuated, respectively, as a result of experience with phoneme-to-grapheme conversion rules as a product of reading instruction. Consistent with this claim, Burnham (2003) has shown that the degree of attenuation for nonnative speech perception in this period is related to reading ability; children who are good readers for their age show greater attenuation for perception of nonnative speech contrasts. This has been investigated only for English, which has what can be called an opaque orthography, in which phoneme-to-grapheme correspondences are inconsistent compared with those of languages with more transparent orthographies, such as Croatian, Spanish, and Turkish, in which phoneme-to-grapheme correspondences are more consistent. This issue of orthographic depth (transparency vs. opacity) is elaborated in more detail below.

The current study investigates the effect of audiovisual speech cues on the production of nonnative speech sounds by adults. To this end, monolingual speakers of Turkish (transparent orthography) and Australian English (opaque orthography) were tested in four different audiovisual and orthographic conditions on Spanish (transparent orthography) and Irish (opaque orthography) stimuli. Relevant literature on visual speech perception (the effect of visual speech and orthographic information on speech perception) are reviewed below, ahead of further elaboration of the study.

Perceiving Visual Speech

Speech perception is not solely an auditory phenomenon. When available, input from other modalities is also used. In particular, visual information conveyed by lip and face movements has been shown to be an integral part of speech processing. Sumby and Pollack (1954) showed that in noisy conditions, visual input increases the perceived clarity of the auditory signal by a magnitude equivalent to 20 dB. Perhaps the most cited demonstration of the role of visual information in speech perception is that by McGurk and MacDonald (1976). They presented participants with a speaker's lip movements for [ga] dubbed onto the auditory signal [ba]. The resultant percept was either [da] or [ða]. This phenomenon, subsequently termed the *McGurk effect*, has been replicated in languages other than English, such as Finnish (Sams, Manninen, Surakka, Helin, & Kättö, 1998), French (Werker, Frost, & McGurk, 1992), and Japanese (Sekiyama & Tohkura, 1993), and has become a very useful tool in auditory-visual speech research.

Cross-language research shows that there are differences in the perception of audiovisual speech. For example, Sekiyama and Tohkura (1993) found that Japanese speakers attend less to visual speech information than their English-speaking counterparts. In turn, Sekiyama (1997a) found that Mandarin speakers were less prone to the McGurk effect than their Japanese counterparts. According to Sekiyama (1997b; Sekiyama & Tohkura, 1993), one possible reason for this difference is that there may be less need to incorporate visual information in Japanese because in comparison with English, there are relatively few visually differentiable phonemes, no consonant clusters, and only five vowels. Moreover, she infers that the reason for weaker McGurk-proneness of Mandarin than Japanese speakers is because Mandarin is a tonal language with four tones (whereas Japanese is a pitch-accented language, with two pitch-accent values), and lexical tones are not visually discernable, but rather manifested in the auditory dimension. On the other hand, Massaro and his colleagues (Massaro, Cohen, Gesi, Heredia, & Tsuzaki, 1993) tested native English, Spanish, and Japanese speakers on synthetic speech stimuli, which consisted of various combinations of auditory and visual /ba/-/da/ speech continua divided across five steps. Employing two types of response formats, forced-choice and open-ended, they found, in contrast to Sekiyama et al.'s findings, no differences due to language background in the use of visual speech. Massaro et al. (1993) indicate that speakers of Japanese, Spanish, and English were influenced by auditory and visual speech inputs similarly, claiming that the fuzzy logical model of perception, which views perceptual events as a three-stage (evaluation → integration → decision) probabilistic process, is a powerful model for explaining the audiovisual speech phenomenon (see Massaro, 1998, for a detailed description of fuzzy logical model of perception).

A growing number of studies, including some by Sekiyama and her colleagues, have also revealed that in most cases participants give more visually influenced responses when attending to nonnative speech (e.g., Sekiyama, Burnham, Tam, & Erdener, 2003; Sekiyama & Tohkura, 1991, 1993). This occurs not only for Japanese-English comparisons, but also for other language combinations (see Burnham, 1998). Recent studies have shown that visual speech information augments both perception and production when English speakers are exposed to a nonnative language, such as Dutch, German (Reisberg, McLean, & Goldfield, 1987), Korean (Davis & Kim, 1998, 1999), or Spanish (Ortega-Llebaria, Faulkner, & Hazan, 2001). Davis and Kim (1998, 1999) tested native English speakers on the identification and production of Korean phrases. They found that phrases presented in auditory-visual conditions resulted in more accurate productions than in an auditory-only condition. In another study, Ortega-Llebaria et al. (2001) tested English-learning native Spanish speakers on their perception of English consonants, using a computer-based auditory-visual training method. They found that audiovisual speech information reduced consonant errors significantly and limited them to voicing and manner errors. Furthermore, testing Japanese and Korean learners of English, Hardison (1998, 2003) has shown that perceptual training featuring visual information results in earlier word identification than training in auditory-only conditions.

In general, the above studies point to the facilitative aspects of visual speech information in attending to nonnative speech. The current study investigates the effect of visual speech information with and without the orthographic information, an extension of the usual audiovisual speech versus auditory speech comparison. This investigation attempts to chart the unexplored area between the attenuation of the ability to perceive nonnative speech contrasts as a result of reading acquisition (Burnham et al., 2002) and the enhancement of nonnative speech perception and production by the provision of visual speech information (e.g., Davis & Kim, 1998, 1999). Such investigation should provide a novel perspective on the role of orthographic depth in speech and audiovisual speech perception, as this could have implications for applied areas such as foreign language instruction. The role of orthographic representation of speech in speech perception is discussed in further detail below.

The Effect of Visual Speech and Orthographic Information on Speech Perception

In addition to visual face information, speech perception is also facilitated by written input. For example, when spoken words are masked by a noise of the same amplitude, it is reported that the utterances are perceived much more clearly if the printed version of the message is presented at the same time. This suggests that printed words are decoded into an internal speech-like representation; in other words, the perceptual system somehow converts the printed words into internal phonetic structures by establishing a link between the printed words and the auditory input embedded in noise (Frost, Repp, & Katz, 1988). As the same result was found for words and nonwords, Frost and his colleagues (1988) suggest that it is the printed form that is translated into phonetic structures, providing further evidence for the link between reading and phonological processing as well as for the apparent effect of print in disambiguating the auditory input in noise. Comparing the effects of visual speech and print, Massaro, Cohen, and Thompson (1990) investigated participants' separate use of visual face and visual written information. They found that a group that was given auditory speech stimuli along with a written version of those stimuli (the auditory-orthographic group) performed better than a group that was given a talker's face articulating the speech stimuli (the audiovisual group). However, interestingly, the written text presented simultaneously with the auditory signal also led to a clearer perception of the signal, and this was above chance level. A couple of experiments recently reported have also shown that exposure to nonnative speech with orthographic input can result in improvement in perceived accent (Erdener, 2002; Erdener & Burnham, 2002). However, it should be noted that this finding warrants further investigation because of a number of factors such as use of nonword stimuli and the small number of raters of perceived accent in the study (Erdener & Burnham, 2002).

One important dimension of written information is the *orthographic depth*, which varies across the alphabetic writing systems of the world's many languages. Orthographic depth can be defined as the degree to which an alphabetic system deviates from simple one-to-one grapheme-to-phoneme correspondences (Van den Bosch, Content, Daelemans, & De Gelder, 1994) and conceptualized along a transparent-to-opaque continuum. The transparent end of this continuum features languages with unambiguous and simple phoneme-to-grapheme correspondences. The ideal case of this is one in which one phoneme (sound) corresponds to one and only one grapheme (letter or combination of letters). Turkish and Spanish are good examples approaching this end of the continuum. In Turkish, for instance, the writing system is based on the Latin system and has very regular phoneme-to-grapheme correspondences. Each letter corresponds to a single sound, and the phonemic interpretation of a letter does not vary with context (Öney & Durgunoğlu, 1997). Examples of opaque orthographies are English, Hebrew (Van den Bosch et al., 1994), and Irish (King, 2002). Opaque orthographies are characterized by their deviation from relatively consistent phoneme-to-grapheme correspondences.

The effect of orthographic depth on reading acquisition has been documented in a number of studies. Some of these show that children learning to read opaque orthographies (e.g., English) are initially slower in reading-related tasks than children learning to read more transparent orthographies, such as Turkish and German (Frith, Wimmer, & Landerl, 1998; Goswami, Gombert, & Barrera, 1998; Öney & Durgunoğlu, 1997; Öney & Goldman, 1984). This difference holds in the initial stages but is ameliorated later in reading acquisition. Additionally, adult readers of phonologically transparent orthographies (e.g., Croatian; see Lukatela, Popadic, Ognjenovic, & Turvey, 1980) have been found to name words correctly while performing lexical decision tasks, whereas readers of opaque orthographies, such as English, may erroneously read words like *pint* [paɪnt] as [pɪnt] as a result of generalization errors from the pronunciation of words like *hint*.

There is a small number of studies on the effect of orthographic input on speech perception (e.g., Frost et al., 1988), but the effect of orthographic information in combination with visual and auditory information on perception and production has not yet been tested. This is important, as it would provide valuable information on the role of orthographic depth (transparency vs. opaqueness) on the perception, and in turn, the production of nonnative speech.

The present study was conducted in order to investigate whether the inclusion of visual and orthographic information improves the production of nonnative speech. Native Turkish speakers (transparent orthographic background) and native Australian-English speakers (opaque orthographic background) were tested on Spanish (transparent) and Irish (opaque) stimuli across four experimental conditions: auditory-only (Aud-only), auditory-visual (AV), auditory-orthographic (Aud-orth), and auditory-visual-orthographic (AV-orth). The participants were presented with legal nonwords in these four conditions, and the words were scored for phoneme errors.

Predictions

First, in line with previous research findings, a facilitative effect of visual speech information is expected; that is, fewer phoneme errors are predicted for the AV condition than for the Aud-only condition. Secondly, a number of within- and between-group predictions are advanced for the orthographic conditions (AV-orth and Aud-orth). In general, a facilitative effect of transparent orthography (Spanish) is anticipated for speakers of a transparent language (Turkish). In this regard, it was hypothesized that Turkish speakers will make fewer phoneme errors in response to the Spanish stimuli when presented with orthographic input, but their performance for orthographic Irish stimuli will be inhibited because of its opaque structure. Turning to Australian speakers, it is predicted that as they speak a language with an opaque orthography, their responses to both Irish and Spanish stimuli in orthographic conditions will be only marginally different, with perhaps more facilitation in Spanish, as a result of its transparent orthography. Whatever the benefit here, it is expected to be less than for their Turkish counterparts. These orthography-related predictions are summarized in Table 1.

Experimental Design

Four sets of 12 nonword items were prepared for each stimulus language, Irish and Spanish. Presentation of these sets was counterbalanced across the four experimental conditions (i.e., Aud-only, AV, AV-orth and Aud-orth), such that any particular participant was exposed to each stimulus item only once. Experimental items were counterbalanced across conditions and between participants. There were 16 possible experimental condition and stimulus set combinations. Two of the 32 participants in each language group (Australian and Turkish) served in each of the four experimental conditions by four stimulus list conditions. This rolling stimulus design was used in order to eliminate any response bias issue that could have arisen from differential item difficulty across stimulus conditions.

Table 1

Stimulus language and background language (L1) combinations

		Stimulus language	
		Spanish (transparent)	Irish (opaque)
Background language	Turkish (transparent)	Turkish→Spanish *Facilitation*	Turkish→Irish *High inhibition*
	English (opaque)	English→Spanish *Some facilitation*	English→Irish *No effect*

Note. Predictions are given in italics. See text for details.

Figure 1 shows the between- and within-participant factors three dimensionally.

There were two dependent variables in this experiment. The first was the number of phoneme errors made in productions by the participants across the experimental conditions. The second was measured in a *writing task*, from which the orthographic errors were recorded. These are explained in more detail in the Procedure section.

In phoneme error analysis, each utterance was compared with its target nonword, and the number of phonemes that were missing, replaced or added, compared with the modeled nonword, was counted. Each such error was given a score of 1. For example, if *cadu* [kadʊ] was the target, and a participant produced the nonword as [gadʊ], an error score of 1 was assigned. If *cadu* was produced as, say, [gadɔ], then an error score of 2 was assigned. If *cadu* was pronounced as [adʊ] (deletion) or as [kadʊn] (addition), an error score of 1 was assigned. The most frequent error patterns under the four experimental conditions were also noted and collated. This analysis was conducted by the first author, who has extensive previous experience with this type of task.

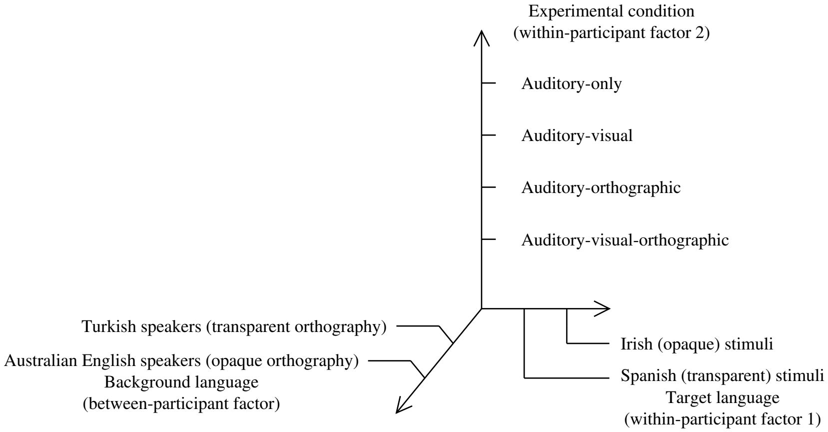

Figure 1. Between- and within-participant factors.

The writing task was used in the orthographic conditions (AV-orth and Aud-orth) and was developed for two reasons: procedural and analytical. Procedurally, this task ensured that participants paid attention to the orthographic input, as well as to the auditory and/or visual signals. Analytically, this task enabled the investigation of patterns in the writing errors that participants made. The total number of writing errors was tallied for each orthographic condition. That is, each missing, replaced, or added letter corresponded to an error score of 1.

Method

Participants

The participants were 32 native speakers of Australian English (22 females and 10 males; $M_{\text{Age}} = 25.66$) and 32 native speakers of Istanbul Turkish (17 females and 15 males; $M_{\text{Age}} = 33.25$). All participants were monolingual speakers of their respective language. A one-way analysis of variance (ANOVA) showed that the Turkish speakers were significantly older than the Australian English speakers, $F(1, 61) = 7.283$, $p < .01$. However, as all participants were monolingual adults, this age difference is not thought to be problematic. In addition, the following protocol was adopted for the recruitment of participants. The participants were (a) not to have been exposed to a foreign language, (b) not to have spent over 3 months in a non-English-speaking (for Australian participants) or non-Turkish-speaking (for Turkish participants) country for any purpose, and (c) to be literate only in their respective native languages.[1]

The Australian speakers were recruited from a pool of 1st-year students enrolled in Psychology 1A and 1B units and from postgraduate students at the Bankstown campus of the University of Western Sydney. The 1st-year students were given credits toward their final course grade in return for their participation. Turkish participants were recruited mainly through word of mouth and the assistance of the Psychology Department at Boğaziçi University in Istanbul. They were given a koala key ring for their participation.

Stimuli

Forty-eight Spanish and 48 Irish nonword stimuli were created based on Spanish and Irish orthographic rules, respectively. The authenticity of the stimuli was confirmed by two linguists who work extensively on Irish and Spanish phonologies and two additional native speakers of each of the stimulus languages. The stimuli consisted of equal numbers of items with a consonant-vowel-consonant (CVC) context and a consonant-vowel-consonant-vowel (CVCV) context. Some vowel components were diphthongs. Even though the CVC context does not occur frequently in Spanish, it is a legal combination in that language. Both CVC and CVCV are legal structures in all of the participant and stimulus languages. They were not compared statistically, as their respective phonological structures are different in Spanish and Irish.

In creating the Spanish (transparent) and Irish (opaque) stimuli, graphemes and diacritics unfamiliar to Australian English and Turkish speakers (e.g., Spanish *baño* and Irish *súil*) were excluded. For both languages, identical protocols were used (see Appendix for the stimulus list.)

Speakers

A native speaker of Chilean Spanish and a native speaker of Irish were recruited to pronounce the stimuli. The Chilean speaker was a 37-year-old male.[2] At the time of recording, he had been living in Australia for 14 years. He uses Spanish mainly to communicate with his family, friends, and relatives and English mainly at home, work, and the university.

The Irish speaker was a 53-year-old male. He is a native speaker of Irish, which he learned at home from his parents.

Although he predominantly uses English for communication at work and home, he also works as a part-time radio broadcaster in Irish language at a multilingual national radio station in Sydney, Australia.

Equipment

A Sony digital video camera (Camcorder DSR-PD100P) and an external microphone (RODE NT 2) attached to the video camera via a digital audiotape (DAT) recorder (used as source of phantom power for the microphone) were used to record the utterances. The experiment was run on a Dell Inspiron 7000 laptop computer equipped with a Pentium II microprocessor, a 12 MB video card and 192 MB RAM. These configurations were sufficient to display the video files continuously without any dropped frames. The participants' oral responses were recorded on digital audiotapes using a TASCAM DA-P1 DAT recorder. Attached to the DAT recorder was a set of headphones (AKGK-270), which enabled the aural presentation of the stimuli, and a head-mounted directional microphone (AKG C-420), which was used to collect the oral responses.

Stimulus Recording and Editing

Recordings were made in a sound-recording booth at MARCS Auditory Laboratories at the University of Western Sydney. For the recordings, the speaker sat in front of a video-camera, and the microphone was placed in the speaker's sagittal plane at approximately 150 cm from the mouth but below camera view. A separate session was conducted for each of the two speakers. Each recording session took approximately 2 hr. Lighting conditions were arranged in a way that speakers' oro-facial movements were clear. The speakers were asked to keep head movements to a minimum while recording and to maintain a neutral facial expression. The speakers were asked to read aloud each stimulus item printed on small index cards. Only one stimulus card was shown at a time to preclude sequential prosodic effects in list reading. Each stimulus was recorded five times, and the best one of the five, on the basis of clarity and accuracy, was selected as the experimental item.

Only the lower part of the speakers' faces was videotaped, from just under eyes and nose level down to the larynx. The laryngeal area was included because it was assumed that the perceivers might use this to identify certain phonemes. One would certainly expect that listeners would pick up cues from orofacial movements: the movements of jaw and lips, and to a certain extent from the muscles around the laryngeal area. The eyes and upper face were occluded because there is evidence that when one is attending to speech, the lower part of the face is used more than the upper part. Despite the evidence that even eyebrow movements provide perceivers with significant paralinguistic cues (Cavé et al., 1996), it appears that lower part of the face disambiguates unfamiliar or nonnative speech input (Davis & Kim, 1998, 1999). Therefore, to direct the listeners' attention to these cues, the upper part of the face was omitted in the recordings.

Raw video recordings of each stimulus were stored on a Sony digital videotape. The images were then captured and converted into MPEG-format video files at 640 × 480 resolution via a Macintosh G3 computer using Adobe Premier software. The average duration of each stimulus was approximately 3000 ms. Each video file was also edited in such a way that the stimulus utterance was preceded and followed by a 250-ms silence. The intensity of audio stimulation was kept at a comfortable listening level, at about 50 dB.

The stimuli were presented in four experimental conditions: Aud-only, AV, AV-orth, and Aud-orth. Stimuli were manipulated in line with these experimental conditions via the DMDX experimental environment (Forster & Forster, 2001), using the appropriate command lines, featuring the auditory, video, and orthographic input channels. Depending on the specific experimental condition, the irrelevant channels were suppressed, except for the auditory input, which was available in all four conditions. For example, in the presentation of Aud-orth condition, the video track was suppressed by reducing the video frame size to nil, and in the Aud-only condition, the orthographic input was deleted from the DMDX command line. The orthographic stimuli were presented simultaneously with the auditory and/or visual input. In each orthographic trial the text was displayed five lines (approximately 5 cm) below the video frame (AV-orth condition) or in the center of the screen (Aud-orth condition), using a 20-point font. These parameters were controlled by appropriate DMDX commands.

Procedure

All participants were tested individually in a quiet room. Australian participants were tested in a sound-attenuated testing room at MARCS Auditory Laboratories at the University of Western Sydney. Turkish participants were tested in a quiet testing room in the Psychology Department at Boğaziçi University. Each participant was seated in front of a laptop computer display unit, about a meter from the screen. Participants wore a head-mounted microphone and a headphone set. They were asked to look at the screen in all four experimental conditions. For the orthographic conditions, they were also asked to read (not aloud) the orthographic version of the stimulus. The experiment featured two main phases: familiarization and testing. In the familiarization phase, participants were trained on the task requirements on 12 practice items, 3 from each experimental condition. This enabled participants to become familiarized with each experimental condition. Before and after this phase, participants were also briefed by the experimenter on the task requirements of the experiment.

Each participant was randomly assigned to 1 of the 16 possible stimulus sets by experimental condition combinations. In each language group, half of the participants began with the Irish stimulus items and the other half with Spanish stimulus items. Each participant was then exposed to each of the four conditions, Aud-only, AV, AV-orth and Aud-orth, for both Irish and Spanish stimuli. The presentation of conditions was counterbalanced across participants using a Latin-square design in order to control for confounding factors such as order effects and fatigue. Prior to each condition, three more practice trials were presented to familiarize participants with the specific condition and clarify any other questions that they might have had regarding the task.

The experiment was self-paced. Participants controlled the presentation of each trial by pressing the space bar on a computer keyboard after responding to the previous trial. The main task across conditions was to repeat each nonword stimulus as quickly as possible. Each trial was preceded by a *Ready* or *Hazir*[3] prompt. The oral responses were captured on digital audiotapes. In every condition, each trial was presented twice randomly, once in each of two separate blocks. The participants were also required to perform a writing task in the orthographic conditions (Aud-orth and AV-orth). In this task, the participants were asked to write down the target item on a response sheet to the best of their memory. When participants finished the writing task for a trial, they were instructed on screen (and orally before the testing session) to press the space bar to continue the experiment with the next item.

The average time of testing for each participant was approximately 1 hr, including the writing task.

Results

Phoneme error analysis was performed to investigate the proficiency with which Australian and Turkish participants produced the Irish and Spanish stimuli in the four experimental conditions. A 2 × (2 × 4) ANOVA with repeated measures on the last factor was conducted with Turkish/Australian as the between-participants factor and target language (Spanish, Irish) and experimental conditions (Aud-only, AV, Aud-orth, AV-orth) as the two within-participant factors. Mauchley's test of sphericity indicated that assumptions for homogeneity of covariance were met for all factors except the experimental-condition factor. Greenhouse-Geisser corrections were made to the degrees of freedom for effects involving this factor (Table 2). Overall results are schematically presented in Figure 2.

Language Background and Stimulus Language

There was no significant difference between the Australian and Turkish participants with respect to their overall phoneme errors, $F(1, 61) = 2.912$, $p > .09$. However, there was a significant interaction of language background and target language, $F(1, 61) = 4.300$, $p < .05$, such that for the Spanish stimuli, Turkish participants made consistently fewer phoneme errors than their Australian counterparts, but for the Irish stimuli, this advantage was attenuated.

The Effect of Orthographic and Visual Information

Overall, there was a significant effect of experimental condition, $F(1, 61) = 24.208$, $p < .01$, and a significant effect of experimental condition with respect to target stimuli, $F(1, 61) = 11.476$, $p < .01$. In addition, there was a significant effect of language background with respect to target stimuli by experimental condition, $F(1, 61) = 4.182$, $p < .007$.

A post hoc Bonferroni analysis performed as a 2 (Turkish vs. Australian speakers) × [2 (Spanish vs. Irish) × 2 (visual information vs. no visual information) × 2 (orthographic information vs. no orthographic information)] ANOVA showed that there was also an overall facilitative effect of orthographic input, $F(1, 61) = 36.788$, $p < .01$, indicating that, in general, participants performed significantly better in conditions featuring orthographic information. Additionally, there was a significant interaction of the visual and orthographic factors, $F(1, 61) = 12.266$, $p < .01$, showing that when orthographic information

Table 2

Mauchley's test of sphericity results for main effects and within-participant comparisons

	Mauchley's *W*	χ^2	*df*	Significance	Epsilon		
Within-participants effect					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Stimulus language	1.000	.000	0		1.000	1.000	1.000
Experimental condition	.816	12.141	5	.033	.871	.929	.333
Stimulus Language × Experimental Condition	.941	3.659	5	.600	.959	1.000	.333

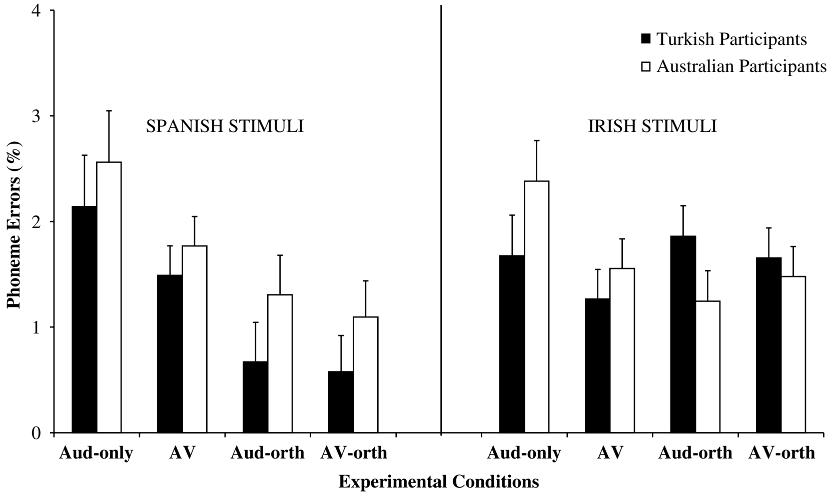

Figure 2. Phoneme errors for Spanish and Irish stimuli (+SE) across speaker groups.

was provided, there was a general reduction in errors across the board, with the advantage due to visual information being ameliorated, whereas when orthographic information was absent, errors increased, and the beneficial effect of visual speech information was evident. Thus these results suggest that orthographic information, when provided, overrides the general facilitative effect of visual information.

As can be seen in Figure 2, Turkish participants performed consistently better than Australian speakers in the nonorthographic conditions. However, group interactions revealed that when orthography was provided, Turkish participants were better than Australian participants for Spanish stimuli but worse for Irish stimuli (see Figure 2). Thus, for speakers of the transparent Turkish, orthography was beneficial for the transparent Spanish, but detrimental for the opaque Irish. For speakers of the opaque Australian English, there was little difference in performance on Spanish and Irish.

Phoneme Error Patterns

In addition to the general error patterns, specific types of phoneme errors were determined. Phoneme error patterns are presented below separately for vowels and consonants.

Vowel data. Table 3 summarizes the three most frequent vowel confusions for language groups for each experimental condition and target language. In general, for vowels, there was a large degree of variance in phoneme replacement errors when orthographic information was not provided.

Turkish and Australian speakers showed similar patterns of errors in their Aud-only and AV responses. The most common Spanish vowel confusion error was [ʊ]-[ɔ] across all four experimental conditions. For AV-orth and Aud-orth responses, there was an interesting finding: Turkish speakers most frequently replaced back vowels with back vowels, and Australian speakers most frequently replaced front vowels with front vowels. For example, whereas the Turkish speakers replaced [ɛ] with [œ] (66.67% of AV-orth and 25.0% of Aud-orth errors), the Australian speakers replaced [ʊ] with [ɔ] (22.22% for AV-orth and 30.43% for Aud-orth; see Table 3).

There was a diversity of errors by Turkish speakers for Irish Aud-only and AV with Irish vowels and for both Turkish and Australian speakers for Irish AV-orth items. Inspection of Table 3 reveals an interesting pattern of errors for Irish [ɪ-ɛ] confusions by Australian participants: The rate of errors for this particular confusion is reduced systematically as visual and orthographic information is provided. For Aud-orth Irish items, there was an interesting pattern in Turkish speakers' responses: The replacement of [ɪ] by [ɪdh] (9.52%) indicated orthographic interference, as they pronounced what was printed, idh [ɪdh], rather than what the print represented, [ɪ]. None of the Australian speakers made an error of this kind.

Consonant data. Consonant errors by Australian and Turkish speakers showed a consistent pattern. The errors can

Table 3

Most frequent vowel confusion error percentages by experimental conditions, target languages, and speaker groups

	Aud-only	AV	Aud-orth	AV-orth
	Irish vowel confustions			
Turkish	[ɪ-a]: 18.18	[ɪ-œ]: 10.82	[a-ɔ]: 16.67	[ɛ-a]: 8.16
	[ɪ-ʊ]: 13.64	[ʊi-ɪ]: 10.82	[ɪ-ɪdh]: 9.52	[ɪ-ɪɔ]: 8.16
	[a-ɛ]: 9.09	[a-ɛ]: 8.11	[ʊɪ-ɪ]: 4.76	[i-a]: 6.12
Australian	[ɪ-ɛ]: 30.24	[ɪ-ɛ]: 19.44	[ɪ-ɛ]: 10.34	[ɪ-ɛ]: 11.11
	[a-æ]: 9.31	[a-æ]: 8.33	[ə-ʊ]: 10.34	[a-ɔ]: 8.33
	[a-ɛ]: 6.98	[ɪ-œ]: 8.33	[ɪɛ-ɛ]: 6.90	[ɛ-aɪ]: 8.33
	Spanish vowel confusions			
Turkish	[ʊ-ɔ]: 31.58	[ʊ-ɔ]: 20.38	[ɛ-œ]: 25.00	[ɛ-œ]: 66.67
	[ɔ-a]: 26.32	[ʊ-ɪ]: 8.33	[ɛ-ɪ]: 25.00	[ɔ-a]: 33.33
	[æ-a]: 10.53	[a-ɛ]: 8.33	[ʊ-ɔ]: 12.50	—
Australian	[ʊ-ɔ]: 16.66	[ʊ-ɔ]: 20.00	[ʊ-ɔ]: 30.43	[ʊ-ɔ]: 22.22
	[ɔ-a]: 13.33	[ɪ-ɛ]: 10.00	[ɔ-a]: 13.04	[ɪ-ɛ]: 11.11
	[ɛ-œ]: 10.00	[a-ɛ]: 10.00	[ɪ-ɛ]: 8.70	[ɔ-œ]: 11.11

Note. Vowels within square brackets indicate the nature of the error: The first vowel is the one that was replaced by the second.

be classified as bilabial confusions, velar confusions, and orthographic interference, and these are discussed in turn.

Overall bilabial confusion scores by speaker groups and target language are presented in Table 4. For Spanish consonants in the Aud-only condition, both Turkish and Australian speakers had higher confusion scores for bilabial stops [b] and [p] (14.49% and 21.0%, respectively) than for any other form of consonant error. A similar pattern was observed for Spanish consonants in the AV condition, with 22.22% and 25.0% errors by Turkish and Australian speakers, respectively. The percentage of Spanish [b]–[p] errors for Turkish speakers was reduced to 4.55% in the

Table 4

Bilabial [b] versus [p] confusion error percentages for Irish, Spanish, and overall stimuli by speaker groups

	Aud-only	AV	Aud-orth	AV-orth
	Irish bilabial [b] versus [p] confusions			
Turkish	6.25	12.20	17.39	17.64
Australian	10.98	17.50	0.00	8.82
	Spanish bilabial [b] versus [p] confusions			
Turkish	14.49	22.22	4.55	0.00
Australian	21.00	25.00	15.00	6.25
	Overall bilabial [b] versus [p] confusions			
Turkish	20.74	34.42	21.94	17.64
Australian	31.98	42.50	15.00	15.07

AV-orth and Aud-orth conditions, but for their Australian counterparts, it remained high at 21.25%.

Fewer bilabial confusions were found in both language groups with Irish stimuli, but bilabial confusions were still more common in the nonorthographic Aud-only and AV conditions: 6.25% and 10.98%, respectively for Turkish speakers and 12.2% and 17.5%, respectively, for Australian speakers. In the Aud-orth and AV-orth conditions, bilabial errors increase, with Turkish speakers making more bilabial errors (35.54%) than Australian speakers (8.82%).

The results of a 2 (background language) × [2 (target language) × 4 (experimental condition)] ANOVA revealed an overall group difference with respect to bilabial confusions, $F(1, 62) = 4.987$, $p < .03$, showing that Turkish speakers made fewer bilabial confusion errors than their Australian counterparts. There was also a significant effect of target language, $F(1, 62) = 13.327$, $p < .001$, such that participants made fewer errors in response to Spanish stimuli than to Irish stimuli. Post hoc analyses in the form of a 2 (Turkish and Australian speakers) × [2 (Spanish versus Irish) × 2 (visual information versus no visual information) × 2 (orthographic information versus no orthographic information)] ANOVA showed similar results, and in addition, it was revealed that there were fewer bilabial phoneme errors when orthographic information was not provided, $F(1, 62) = 12.684, p < .001$. Additionally, there was a significant interaction of visual and orthographic information, $F(1, 62) = 15.354, p < .001$, such that orthographic information was particularly useful in the absence of visual speech information.

The velar confusion scores by speaker groups and target language are presented in Table 5. The second most common consonant confusion was between the velar stops [k] and [g]. Most errors occurred in the AV condition, and this was the case for both Spanish and Irish stimulus sets (19.44% for Turkish and 21.65% for Australian speakers). The results from orthographic conditions were rather interesting. There was a striking drop in the [k]-[g] replacement errors in the AV-orth and Aud-orth conditions, for both Irish and Spanish stimuli, indicating that orthographic input was beneficial in disambiguating the novel lexicon. Whereas the Turkish participants' error rate for velars was 12.68%, for their Australian counterparts, the velar error rate in the orthographic conditions was 8.74%. An ANOVA for velar errors revealed a significant group interaction, $F(1, 62) = 4.892, p < .03$, showing that Australian speakers made fewer velar confusion errors overall than their Turkish counterparts. Results also show that there was a significant effect of experimental condition, $F(1, 62) = 4.934$, $p < .01$. There was no effect of target language with respect to velar errors ($p > .3$), nor was there any significant interaction of target language by speaker groups ($p > .1$) A Bonferroni post hoc analysis showed that there was a significant effect of orthographic information, $F(1, 62) = 10.906$, $p < .001$, and a group interaction with respect to this, $F(1, 62) = 8.299, p < .001$, showing that Australian participants made fewer errors than Turkish

Table 5

Velar [k] versus [g] confusion percentages for Irish, Spanish, and overall stimuli by speaker groups

	Aud-only	AV	Aud-orth	AV-orth
	Irish velar [k] versus [g] confusions			
Turkish	4.17	16.66	4.35	0.00
Australian	7.32	10.00	2.86	5.88
	Spanish velar [k] versus [g] confusions			
Turkish	4.35	2.78	8.33	0.00
Australian	6.32	11.65	0.00	0.00
	Overall velar [k] versus [g] confusions			
Turkish	8.52	19.44	12.68	0.00
Australian	13.64	21.65	2.86	5.88

participants when orthographic information for velars was provided. It was also found that both language groups performed better in conditions with than without orthographic input (AV-orth and Aud-orth), $F(1, 63) = 9.371$, $p < .01$, yet there was no difference between the speaker groups with respect to this ($p > .3$).

Orthographic interference. Overall orthographic interference errors for both types of confusion are summarized in Table 6. There were a number of errors in which the orthographic representation of a phoneme overrode its auditory or visual representation. For Irish the most frequent orthographic interference error was the replacement of [dʒ] by [d], and for Spanish it was the replacement of [x] by [ʒ].

As can be seen in Table 6, there is little difference with respect to Irish [dʒ]-[d] and Spanish [x]-[ʒ] confusions in the Aud-only and AV conditions; in fact, there is no [x]-[ʒ] confusion in Aud-only by either group of speakers. However, there appears to be substantial orthographic interference in Turkish speakers' responses. Indeed the Turkish participants' [x]-[ʒ] confusion

Table 6

Orthographic interference: [dʒ]-[d] confusions in Irish and [x]-[ʒ] confusions in Spanish, expressed in error percentages

	[dʒ]-[d] confusions in Irish				[x]-[ʒ] confusions in Spanish			
	Aud-only	AV	Aud-orth	AV-orth	Aud-only	AV	Aud-orth	AV-orth
Turkish	4.17	0.00	17.65	0.00	0.00	2.78	45.83	22.73
Australian	2.44	2.50	14.29	25.53	0.00	1.56	3.13	0.00

responses to Spanish stimuli increase appreciably when orthographic information is added, from 0 to 45.83% for auditory and from 2.78% to 22.73% for auditory-visual trial types. On the other hand, for the Australian English participants, [dʒ]-[d] confusions increase with the addition of the orthographic information from 2.44% to 14.29% for auditory and from 2.5% to 25.53% for auditory-visual responses.

Writing task errors. The results of the writing task (conducted only for orthographic conditions, Aud-orth and AV-orth) are presented in Figure 3. Participants made significantly fewer written errors for the Spanish than for the Irish stimuli, $F(1, 61) = 59.65$, $p < .003$. This effect interacted significantly with language group, $F(1, 61) = 5.597$, $p < .05$, showing that both Turkish and Australian speakers made very few errors in their Spanish written responses, whereas Australian speakers made far fewer errors than the Turkish speakers in their Irish written responses. This difference is interesting, because it suggests that Australian speakers' relatively good performance on the Irish stimuli may have been due to their experience with reading an opaque orthography. This is similar to the effect of the orthographic background and target language on phoneme errors described earlier (see Figure 2).

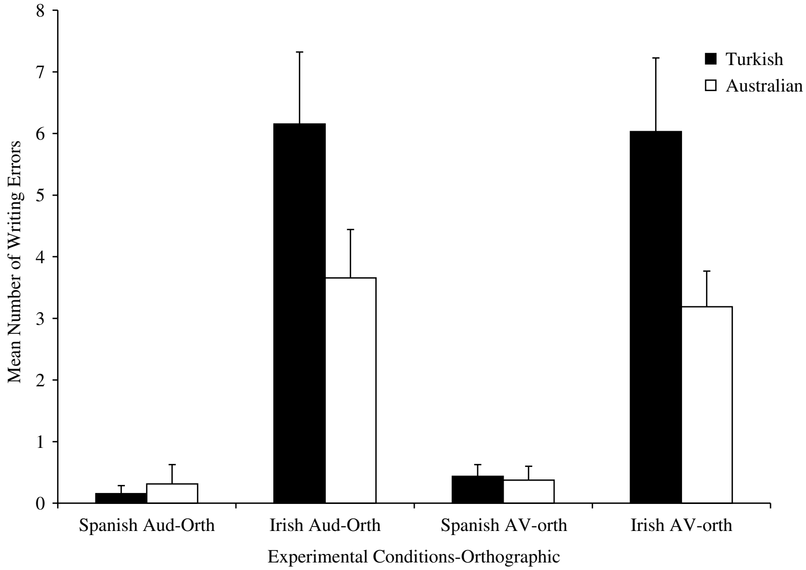

Figure 3. Writing task errors (+SE) in Spanish and Irish across speaker groups.

Discussion

The results showed that there were effects of both visual and orthographic information, and these are discussed in turn.

Visual Information and Nonnative Speech

The phoneme error results confirm previous findings that provision of visual information enhances speech production for nonnative stimuli (Davis & Kim, 1998; Hardison, 1999; Ortega-Llebaria et al., 2001; Reisberg et al., 1987). All participants, irrespective of language background and target language, produced the Spanish and Irish stimuli much more accurately in conditions with visual information than in conditions with no visual information, but this facilitative effect of visual information was apparent only in the absence of orthographic information. There are two possible reasons for this.

First, auditory-visual speech perception would appear to be a natural, ecologically valid process including some degree of redundancy given the common articulatory source of auditory and visual speech information (Summerfield, 1979). Orthographic information, on the other hand, is connected with speech only via learned symbolic representations. Nonetheless, these representations are extremely powerful (Burnham et al., 2002) and appear to affect basic auditory processes. It is quite possible that this overlay or imposition of orthography occurs not only for basic auditory, but also for basic auditory-visual, speech perception. However, this suggestion requires further research.

Second, there may be an effect of working memory as the orthographic conditions would involve cognitive/postperceptual processing of the orthographic information stemming from reading. In the instructions for the orthographic conditions, participants were asked to look at the screen and read (not aloud) the orthographic input. Under such conditions (Aud-orth & AV-orth), it is possible that attention may be an important factor. A recent study (Tyler, 2001) indicates that working-memory consumption is important in the process of comprehending nonnative speech input. This may particularly be the case in the AV-orth condition here; participants might simply have disregarded the auditory and/or visual speech information to some, but perhaps a significant, extent because of the availability of orthographic information. This explanation is consistent with the response pattern of the Turkish participants: In their responses to nonnative speech, with which they had no experience, they appear to have relied more on the orthographic input. Perhaps their specific experience with the transparent Turkish orthography primed them to attend more to orthography than the auditory-visual signals. On the other hand, Australian participants' specific experience with the opaque English orthography perhaps primed them to attend less to the orthography than to the auditory-visual signal. This is consistent with previous studies, which suggest that we pay more attention to face information when attending to nonnative speech (e.g., Fuster-Duran, 1996; Sekiyama & Tohkura, 1993).

Phoneme error analyses revealed different error patterns for vowels and consonants across language groups and target languages. For both speaker groups in Aud-only and AV conditions, there was no clear pattern of vowel errors. Scrutiny of the most frequent vowel errors in both Irish and Spanish showed that in the presence of visual information, there was a reduction in phoneme replacement errors. The phoneme replacement errors also show that overall, the most common errors, bilabial [b-p] and velar [k-g] confusions, were reduced in the presence of visual information, even though it might be thought that visual discrimination of these phonemes should be relatively equivalent, as they share the same places of articulation.

Orthography and Nonnative Speech Production

Overall, in the conditions in which orthographic information was absent (Aud-only & AV), Turkish speakers consistently made fewer errors. However, when orthography was present (Aud-orth and AV-orth conditions), it facilitated the production of nonnative speech stimuli. Most interestingly, when orthographic information was presented and it was transparent (i.e., Spanish), Turkish speakers made many fewer phoneme errors than their Australian counterparts. However, the Turkish perceivers' performance was significantly attenuated when the orthographic information was opaque (i.e., Irish) and was worse than that of their Australian counterparts. On the other hand, in the orthographic conditions the number of errors made by Australian speakers was almost equivalent for Spanish and Irish. These results suggest that Turkish participants are affected by orthographic information more than their Australian counterparts. This view is supported by the results regarding orthographic interference errors. This was also quite noticeable in the Turkish responses to orthographic Spanish stimuli (Aud-orth & AV-orth); Turkish participants consistently made confusion errors between [x] and [ʒ] phonemes. As the Spanish phoneme [x] and Turkish phoneme [ʒ] are represented by the same letter, *j*, this suggests that Turkish participants' productions were affected by orthographic input to a greater extent than those of their Australian counterparts.

The analysis of the writing task provides additional support for the effect of orthography. As predicted, Turkish speakers made fewer spelling errors for Spanish nonwords than for Irish nonwords. The analyses suggest that when Turkish participants encounter new vocabulary in the target languages in this study, they appear to process this input on the basis of the degree to which phonemes and graphemes match consistently. In other words, the Turkish speakers appear to process orthographic information via a grapheme-to-phoneme conversion procedure that assigns individual graphemes to individual phonemes. In the case of Spanish this strategy works well, because of similar orthographic depth for Turkish and Spanish. However, the situation is different for Irish; this phoneme-to-grapheme strategy does not work well because of the opacity of the Irish orthography. On the other hand, Australian speakers were better than their Turkish counterparts in producing Irish stimuli presented with orthographic input, and they also performed better on Irish in the writing task. One reason for Australian speakers' better performance in Irish orthographic conditions might be that speakers of languages with opaque orthographies, like English, develop a whole "picture-orthographic" representation of individual lexical items. While doing this, they may process the auditory and/or visual information more efficiently and in a parallel manner. Perhaps these aspects need to be further investigated with an emphasis on orthographic processing.

The results also suggest that presenting participants with orthographic input is useful in pronunciation, provided that the target language has a transparent orthography. When the target language has an opaque orthography, it seems better not to provide the learners with orthographic input, at least in the initial stages of exposure to a foreign language, and especially if they themselves have experience only with a transparent orthography.

A number of frequent bilabial and velar confusion errors were also found, and there was a clear effect of orthographic information in the reduction of these errors. Turkish participants made fewer bilabial and velar errors overall and in particular when the orthography was transparent. Conversely, Australian speakers made fewer phoneme bilabial and velar errors when the orthography was opaque. One possibility is that Australian participants have greater metalinguistic awareness on the basis of their experience with opaque English orthography, which may allow allocation of more attentional resources to the auditory and/or visual information than to orthographic input.

An alternative explanation regarding the perception of Spanish initial position plosives by Australian speakers can be made on the basis of the relation of the English and Spanish phoneme systems (see the perceptual assimilation model of Best, 1995), in particular for the visually unmarked bilabial [b-p] and velar [k-g] confusions. For example, in Spanish, the initial-position bilabial plosives are realized with a short lag voice onset time (VOT), whereas in English the initial-position plosives have long lag VOT values. The Australian-English speakers might have assimilated these bilabials into their existing native phoneme category on the basis of place of articulation (which has robust visual information, realized by the opening of the lips) and disregarded the VOT information, which probably is not as salient as the visual information.

The above results show that the facilitative effect of orthographic information in nonnative tasks is a function of the degree of native language and nonnative orthographic depth. If these results can be generalized, they may show that although in the early stages of reading acquisition, learning an opaque orthography may have its challenges (Frith et al., 1998; Goswami et al., 1998; Öney & Durgunoğlu, 1997; Öney &

Goldman, 1984), there may be a positive benefit of this later in life, in situations such as processing two types of inconsistent information efficiently in another language (e.g., Irish writing and pronunciation).

Practical Implications for Foreign Language Education

Research in auditory-visual speech processing has significant practical applications. Two major areas of potential application are foreign language teaching and language training of children and adults with hearing impairment.

The current results show that provision of visual information reduces phoneme errors in nonnative speech production. Traditionally, in foreign language teaching settings, 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 orthographic depth (i.e., whether transparent) of the target language, inclusion of orthographic input in foreign language instruction. However, it should be noted that the participants in this study were not learners of Irish and Spanish, so issues such as motivation should be taken into account, and results must be interpreted in terms of speech perception terms. In addition, a possible shortcoming of the present study was the exclusion of an orthographic-only condition (because of earlier theoretical and methodological concerns in the planning of the study). Such a condition would have provided a baseline against which the effect of AV-orth and Aud-orth conditions could be compared. However, based on some of the Turkish responses, it can be speculated that Turkish participants would perhaps have fewer errors in an orthography-only condition than their Australian counterparts as a result of the similar transparency of Spanish and Turkish orthographies in terms of transparency. As for Irish responses, one would expect comparable or better performance by the Australian than the Turkish speakers because of English speakers' relatively greater familiarity with, and exposure to, Irish spellings, such as the names *Siobhan* and *Sean*.

While providing further support for the robustness of visual information in perception and production of unfamiliar nonnative speech stimuli, the current study also provides us with evidence that inclusion of orthographic input in the acquisition of some languages, but not others, may assist learners of those languages. Further research is certainly required using both real-word and nonword stimuli from different languages with varying degrees of orthographic depth.

In this study, providing orthographic information has been shown to be effective in the reduction of phoneme errors in production. Foreign language instruction methods could be amended to render them more efficient and beneficial by including the use of orthographic information. In particular, development of new training methods for the teaching of languages, such as Italian, Spanish, and Turkish, that have transparent orthographies might be developed in order to reinforce auditory and visual inputs. This might include a component of instruction in which students are familiarized with those phoneme-to-grapheme correspondences that are consistent in the target language. Such training could provide an economy in pronunciation teaching and save a considerable amount of time and resources in the learning process. On the other hand, pronunciation components for teaching languages, such as English and Hebrew, that have opaque orthographies (Van den Bosch et al., 1994) might largely emphasize auditory and visual components in earlier stages of teaching.

In summary, the results of this study show that orthographic language background significantly affects the processing of nonnative language at the level of individual words. Studies have yet to be conducted with longitudinal designs and with words in sentences in order to uncover the possible benefits of the use of visual and orthographic information in foreign language pronunciation training.

Given that speech perception is not simply an auditory phenomenon but also uses visual and orthographic input, current models of speech perception, such as Flege's (1999) second language model and Best's (1995) perceptual assimilation model, could profitably be applied and extended in second language acquisition research by including visual speech stimuli. Of interest in this context would be testing the extent to which visual speech categories, as well as phonemes, are assimilated into native phoneme categories. Another intriguing research endeavor would be to investigate the extent to which new prototypes that are formed for novel visual speech categories are confused with similar visual speech categories in the native language.

Revised version accepted 30 June 2004

Notes

[1] Despite the fact that no reading ability screening test was conducted, all participants (except for 1 Turkish participant) had high school educations, and all were rigidly screened through the selection criteria prior to the experiment.

[2] As female speakers were unavailable for Irish, no female Spanish speaker was recruited, either.

[3] The Turkish word for *ready*.

References

Best, C. T. (1995). Learning to perceive the sound pattern of English. In C. Rowe-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 9, pp. 217–304). Norwood, NJ: Ablex.

Burnham, D. K. (1998). Language specificity in the development of auditory-visual speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory visual speech* (pp. 27–60). Hove, UK: Psychology Press/Erlbaum/Taylor & Francis.

Burnham, D. K. (2003). Language specific speech perception and the onset of reading. *Reading and Writing*, *16*, 573–609.

Burnham, D. K., Earnshaw, L. J., & Quinn, M. C. (1987). The development of the categorical identification of speech. In B. E. McKenzie & R. H. Day (Eds.), *Perceptual development in early infancy: Problems and issues. Child psychology* (pp. 237–275). Hillsdale, NJ: Erlbaum.

Burnham, D. K., Tyler, M., & Horlyck, S. (2002). Periods of speech perception development and their vestiges in adulthood. In A. Rohde (Ed.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 281–300). Trier, Germany: Wissenschaftlicher Verlag Trier.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (Eds.). (1996). *About the relationship between eyebrow movements and F0 variations* (Vol. 4). New Castle, DE: Citation Delaware.

Davis, C., & Kim, J. (1998, December). *Repeating and remembering foreign language words: Does seeing help?* Paper presented at the International Conference on Auditory-Visual Speech Processing, Terrigal-Sydney, Australia.

Davis, C., & Kim, J. (1999, September). *Perception of clearly presented foreign language sounds: The effects of visible speech*. Paper presented at the Fourth Annual Auditory-Visual Speech Processing Conference, Santa Cruz, CA.

Erdener, V. D. (2002). *The effect of auditory, visual, and orthographic information on second language acquisition*. Unpublished master's thesis, University of Western Sydney, Sydney, Australia.

Erdener, V. D., & Burnham, D. K. (2002). The effect of auditory-visual information and orthographic background in L2 acquisition. *In Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1929–1932. Bonn, Germany: International Speech Communication Association.

Flege, J. E. (1999). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Forster, J., & Forster, K. I. (2001). DMDX [experimental sotfware]. Tucson: University of Arizona, Department of Psychology. Retrieved April 2, 2005, from http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, *2*(1), 31–54.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory & Language*, *27*(6), 741–755.

Fuster-Duran, A. (1996). Perception of conflicting audio-visual speech: An examination across Spanish and German. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 139–143). Berlin: Springer-Verlag.

Goswami, U., Gombert, J. E., & Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Linguistics*, 19, 19–52.

Hardison, D. M. (1998, December). *Spoken word identification by native and non-native speakers of English: Effects of training, modality, context and phonetic environment*. Paper presented at the Fifth International Conference on Spoken Language Processing, Sydney, Australia.

Hardison, D. M. (1999). Bimodal speech perception by native and non-native speakers of English: Factors influencing McGurk effect. *Language Learning*, *49*(1), 213–283.

Hardison, D. M. (2003). Sources of variability in the perceptual training of /r/ and /l/: Interaction of adjacent vowel, word position, talkers' visual and acoustic cues. *Applied Psycholinguistics*, *24*, 495–522.

King, D. (2002). *Old-Irish spelling and pronunciation*. Retrieved from http://www.smo.uhi.ac.uk/old-irish/labhairt.html

Lukatela, G., Popadic, D., Ognjenovic, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory and Cognition*, *8*(2), 124–132.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, D. W., Cohen, M. M., Gesi, A., Heredia, R., & Tsuzaki, M. (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics*, *21*, 445–478.

Massaro, D. W., Cohen, M. M., & Thompson, L. A. (1990). Visible language in speech perception: Lipreading and reading. *Visible Language*, *22*(1), 8–31.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Öney, B., & Durgunoğlu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. *Applied Psycholinguistics*, *18*, 1–15.

Öney, B., & Goldman, S. R. (1984). Decoding and comprehension skills in Turkish and English: Effects of the regularity of grapheme-phoneme correspondences. *Journal of Educational Psychology*, *76*(4), 557–568.

Ortega-Llebaria, M., Faulkner, A., & Hazan, V. (2001, September). *Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English*. Paper presented at the Auditory-Visual Speech Processing International Conference, Aalborg, Denmark.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). London: Erlbaum.

Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). McGurk effect in Finnish syllables, isolated words and words in sentences: Effects of word meaning and sentence context. *Speech Communication*, *26*, 75–87.

Sekiyama, K. (1997a). Audiovisual speech perception and its inter-language differences. *Japanese Journal of Psychonomic Science*, *15*(2), 122–127.

Sekiyama, K. (1997b). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics*, *59*(1), 73–80.

Sekiyama, K., Burnham, D., Tam, H., & Erdener, D. (2003). Auditory-visual speech perception development in Japanese and English speakers. In J.-L. Schwartz, F. Berthommier, M.-A. Cathiard, & D. Sodoyer (Eds.), *Proceedings of AVSP [Audio Visual Speech Processing] 2003* (pp. 43–47). Grenoble, France: Université Stendhal, Institute de la Communication Parlée.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, *90*(4, part 1), 1797–1805.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(4), 427–444.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, *36*, 314–331.

Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, *51*(2), 257–280.

Van den Bosch, A., Content, A., Daelemans, W., & De Gelder, B. (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, *1*(3), 177–188.

Werker, J. F., Frost, P. E., & McGurk, H. (1992). La langue et les levres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, *46*(4), 551–568.

Appendix A

Spanish Nonwords

Nonword	International Phonetic Alphabet	Nonword	International Phonetic Alphabet
Bema	bɛma	**Naef**	naɛf
Bichu	bɪtʃʊ	**Nir**	nɪr
Bita	bɪta	**Noda**	nɔda
Cadu	kadʊ	**Pean**	pean
Caji	kahɪ	**Pinu**	pɪnʊ
Caut	kaʊt	**Pocu**	pɔkʊ
Depi	dɛpɪ	**Pol**	pɔl
Desu	dɛsʊ	**Revi**	rɛvɪ
Doar	dɔar	**Roas**	rɔas
Feun	fɛʊn	**Rotu**	rɔtʊ
Fis	fɪs	**Saof**	saɔf
Fuir	fʊɪr	**Selo**	sɛlɔ
Fujo	fʊhɔ	**Siar**	sɪar
Gapo	gapɔ	**Sofe**	sɔfɛ
Gito	gɪtɔ	**Soir**	sɔjɪr
Gor	gɔr	**Souf**	sɔʊf
Jego	xɛgɔ	**Sul**	sʊl
Jupi	xʊpɪ	**Taul**	taʊːl
Laer	laɛr	**Tije**	tɪxɛ
Lura	lʊra	**Tuan**	tʊan
Min	mɪn	**Vaem**	vaɛm
Miura	mɪʊra	**Voji**	vɔhɪ
Mogi	mɔgɪ	**Vual**	vʊal
Mur	mʊr	**Yur**	jʊːr

Appendix B

Irish Nonwords

Nonword	International Phonetic Alphabet	Nonword	International Phonetic Alphabet
Baer	ba:ɹ	Liodfaidh	ˈlɪədfɛ
Baghan	ba:in	Meib	mɛb
Biosae	bɪʲose	Muigfidh (Munster dialect)	mʊɪkfɪ
Cail	kiɫ	Muigfidh (Connacht dialect)	mʊɪkɪ
Ceimhin	kɛ:vɪn	Neife	nɪɛfɛ
Ciobfaidh	kɪʲɛpə	Neol	nɪjol
Coib	kiɪb	Niosa	nɪjosʌ
Comhan	kʊ:vən	Nodhar	nijɔl
Damhba	da:vba	Pabh	pav
Deaip	dʒæ:p	Peabhna	peavnʌ
Deimhin	dʒɛ:vɪn	Peain	peajn
Dobh	dʌv	Peibidh	pɛbɪ
Faeic	fʊɛk	Pidfidh	pɪdfɪ
Feot	fɪʲɐt	Pogh	pɑh
Fogh	fʌh	Reibe	rɛbˈɛ
Fuige	fʊigɛ	Riogfaidh	ˈrɪəkə
Fuimhe	fʊɪvæ	Rubh	rʊv
Gadhan	ga:jin	Saghr	sa:r
Gaoin	gijɪn	Seibe	ʃɛbbɛ
Gibfidh	gɪpə	Taibe	tɛbbɛ
Goife	gɪfɛ	Tamhra	tavɹə
Ladhbh	la:v	Taois	tiʲɪʃ
Libfidh (Munster dialect)	lɪpfɪ	Tibh	tɪv
Libfidh (Connacht dialect)	lɪpɪ	Toige	tɔ:ɪgɛ