Turkish Journal of Psychology, December 2015, *30* (76), 28-31

Summary
The McGurk Illusion in Turkish

Doğu Erdener
Middle East Technical University
Northern Cyprus Campus

The research in the past four decades has firmly established that speech perception is not solely an auditory process but an auditory-visual one, where auditory and visual speech information in the form of face and lip movements are integrated to form a resultant percept. Originally, the role of visual speech information had been thought to be significant source of speech information only in degraded listening conditions (Sumby & Pollack, 1954); however, as demonstrated by an illusory effect known as the *McGurk Effect*, the visual speech information does affect the resultant percept in clear listening conditions, too (McGurk & MacDonald, 1976). In a typical demonstration of McGurk Effect, when subjects hear an auditory /ba/ dubbed onto the face of a talker saying /ga/, they tend to perceive a /da/, a percept that is present neither in auditory nor in visual form. A reverse arrangement apparently results in a /bga/ or a /gba/ response. McGurk effect has been demonstrated beyond the syllable context, e.g., word and sentence (e.g., Sams, Manninen, Surakka, Helin, & Kättö, 1998). Although the direct ecological validity of McGurk Effect can be questionable, it has so far yielded to be a commonly used tool to measure the extent to which visual speech information affects the resultant speech percept.

McGurk effect has been studied developmentally in various language environments. Studies with infants have so far shown that infants *do* integrate the two sources of speech information (Burnham & Dodd, 2004; Desjardins & Werker, 2004; Kuhl & Meltzoff, 1982; Rosenblum, Schmuckler, & Johnson, 1997). Coincidentally, the original McGurk effect study was a developmental research as well. In that study it was found that adult participants were more prone to the McGurk effect compared to their two younger counterpart groups. A similar pattern of results were also obtained elsewhere (Desjardins, Rogers, & Werker, 1997; Erdener & Burnham, 2013; Massaro, Thompson, Barron, & Laren, 1986).

Combining ontogenetic and differential methods, Sekiyama and Burnham (2008) tested 6-, 8-, and 11-year-old children and adults from English and Japanese language backgrounds on McGurk type stimuli. They found comparable use of visual speech information by English and Japanese 6-year-olds; however, between 6 and 8 years of age a sharp increase in visual speech influence for English-, but not Japanese-speaking children, was observed and remained at the same low level at 6, 8, and 11 years through to adults. For the English language speakers a *significant* increase was observed at 8 years was maintained afterwards. The authors attributed this developmental difference to both language-specific and developmental differences: English-speaking adults and older children were faster in visual speech domain than in auditory speech domain *and* than their Japanese counterparts, and Japanese perceivers were faster in auditory-only modality than their English-speaking counterparts. A subsequent study by Erdener and Burnham (2013) showed that the developmental differences in the English-speaking context were mainly attributable to the development of native speech perception relative to non-native speech perception. Such findings point out to the fact that an ambient language environment seems to impose different requirements onto its native speakers. In fact, a number of other cross-language studies consistently point out to inter-language as well as neural-based individual differences (Nath & Beauchamp, 2012) with respect to the level and amount of integration of auditory and visual speech information as outlined in the following section.

McGurk effect does not seem to be a uniformly perceived illusory effect and in fact the existing literature suggests cross-language differences. For instance, in a study using a McGurk effect identification task, Chen and Hazan (2009) tested Mandarin- and English-speaking children aged at 8 and 9 years as well as adults and found that, irrespective of age, the effect of visual speech

Address for Correspondence: Yrd. Doç. Dr. Doğu Erdener, ODTÜ Kuzey Kıbrıs Kampusu Psikoloji Programı Kalkanlı Güzelyurt KKTC via Mersin 10 Türkiye
E-mail: d.erdener@gmail.com

information was greater when participants observed a non-native speaker and that this effect increased with age. Additionally, Sekiyama and Tohkura (1993) found that native Japanese speakers are less prone to McGurk effect, thus visual speech information, considerably less than their American counterparts in their respective native languages but more than native Mandarin speakers (Sekiyama, 1997). Sekiyama suggested that, compared to Japanese, there may be less need to integrate visual speech in Mandarin but more in Japanese, in relative terms, and even more in English. In fact, there are, presumably, fewer visually discernable phonemes and fewer vowels in Japanese than in English. The even weaker visual speech effect in Mandarin (a tone-based language) may be due to the observation that tonal information, where structured pitch variations indicate meaning differences, is not visually manifested (Burnham, Ciocca, Lauw, Lau, & Stokes, 2000). However, irrespective of native language, subjects' use of visual speech information considerably increased when they watched speakers talking a non-native language. This finding was later supported in other language contexts such as Dutch, German (Reisberg, McLean, & Goldfield, 1987), Korean (Davis & Kim, 2001) and Spanish (Ortega-Llebaria, Faulkner, & Hazan, 2001). In a more recent study, Wang, Behne, & Jiang (2009) presented native Korean, Mandarin and English speakers with speech stimuli made up of labiodentals (e.g., /f/ as in *f*light, non-Korean), interdentals (e.g., /θ/, as in *th*ick, non-Korean and non-Mandarin) and alveolars (/s/ as in *s*till) in auditory-visual, auditory-only and visual-only listening conditions. Despite the fact that native English speakers performed better than the other two groups on the identification of these phonemes, both Korean and Mandarin perceivers showed native-like performance for labiodentals, which have a relatively higher degree of visibility than interdentals and alveolars, for which these groups showed poorer performance.

These results from above studies suggest that the degree to which we use visual speech information depends on a number of factors such as the native phonemic repertoire (Sekiyama, 1997; Sekiyama & Tohkura, 1993), degree of visual discernibility of visemes (Burnham et al., 2000; Wang et al., 2009) and age (McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008). Such inter-language differences warrant further investigation into the nature of auditory-visual speech perception in other languages that have yet to be studied. And Turkish is one such language. In the only auditory-visual speech perception study with Turkish participants, Erdener and Burnham (2005) tested native speakers of Australian English and İstanbul Turkish on the perception and production of non-word stimuli presented in auditory-only, auditory-visual, auditory-orthographic and auditory-visual-orthographic conditions and showed greater reliance by Turkish speakers on the orthographic input when attending to non-native stimuli whether or not visual speech information was available. Although that study revealed some evidence regarding the use of visual speech information, there is still the need to illuminate the extent to which Turkish perceivers use visual speech information and document responses to McGurk stimuli in this specific language context. Thus the aim of this study is to investigate, for the first time, the McGurk effect in native Turkish speakers using both Turkish-native and non-native (American English) stimuli. Thus on the basis of Turkish phonology and phonotactical requirements presented above, two predictions were advanced: (1) native Turkish speakers should yield McGurk-type responses given that Turkish language has presumably no known phonological elements (e.g., lexical tone) that may not necessitate the use of visual speech information in McGurk-type stimuli; (2) similar to other studies, Turkish speakers should yield more McGurk type responses to non-native stimuli than to native stimuli. As there is only a single modality involved, no significant difference is expected between English and Turkish visual-only (lipreading) conditions.

Method

Participants

Forty-five English language preparatory students were recruited from the School of Foreign Languages at Middle East Technical University - Northern Cyprus Campus (METU-NCC). Each student was paid 20TL (around US$9) for participation. Of these participants, 37 were females and eight were males with an overall average age of 20.9 years (*SD* = 2.09 years). The students were enrolled in a beginner level English language program and had very little or no previous experience with any foreign language. All students were native Turkish speakers from both Cyprus and Turkey.

Stimuli and Materials

The stimuli were produced by four talkers, two native speakers of Turkish and American English each with one male and one female for each stimulus language. The McGurk stimuli were recorded in a quiet room using a digital video camera. The talkers maintained a neutral facial expression during the recording of each stimulus. The stimuli were presented on a 16" monitor laptop computer with suitable hardware specifications to display video files conveniently using a small software program developed at the Psychology Program at METU-NCC.

There were two types of stimuli created in both English and Turkish: auditory-visual (AV) and visual-

only (VO; also lipreading). The AV stimuli consisted of incongruent auditory and visual components. In a pilot study, it was ensured that the fused responses of these components resulted in the perception of real and pseudo words in the respective stimulus language. The auditory and visual components of the stimuli were made up of CVC-, VCV-, CVCV- and other multiple complex CV-context combinations. Originally, 66 and 72 stimulus AV files were created for English and Turkish, respectively. Then for each language 24 best ones were chosen to be used in the experiment. The VO stimuli were randomly chosen by the stimulus presentation software from among the same pool of video stimuli in which the auditory components were deleted.

Procedure

Each participant was tested individually in a quiet testing room. The stimuli were presented in blocks on the basis modality (AV and VO) with the order of stimulus presentation randomized. The order of Turkish and English stimuli was also counterbalanced across participants. Both AV and VO stimuli were presented using software prepared by a research assistant. A simple identification task was implemented in the procedure. In each trial a video file was shown in which the AV and VO stimuli were played. At the end of each stimulus presentation the question "What was spoken?" and "Konuşmacı ne dedi?" appeared for English and Turkish stimuli, respectively. Participants had 25 seconds to respond by typing their responses in a box that appeared below the question. Each participant went through a task familiarization phase in which they completed ten trials. In the experimental phase, 24 stimuli were presented randomly in each language block. Each stimulus was presented twice, yielding a total of 48 trials in each block. The whole session took about 30 minutes on average.

Results

Responses revealing visual influences were deemed as visually influenced responses. The data was subjected to a series of t-tests. First, the results yielded that for both English, t_{44} = -3.266, $p < .005$, and Turkish stimuli, t_{44} = -9.570, $p < .001$, participants gave visually based responses above chance levels, set at 50% visually-based response rate, revealing that perceivers use visual speech information when tested with McGurk stimuli. Second, as per the prediction, it was revealed that perceivers are more prone to McGurk effect in English than in Turkish, $t_{44} = 6.623$, $p < .001$. Finally, the VO data from English and Turkish stimuli were compared. The results here showed no differences between the two target language data sets, t_{44} = -1.497, $p > .05$.

Discussion

As predicted, participants gave visually based responses to McGurk stimuli in both English and Turkish stimuli well above chance levels showing that the visual speech information is used in the way it is used in other languages which were tested in other McGurk studies, such as Italian (Bovo, Ciorba, Prosser, & Martini, 2009) and Spanish (Ortega-Llebaria et al., 2001). McGurk effect was observed at a greater magnitude for non-native stimulus language (English) than for native language stimuli for the Turkish speakers tested here. This finding lends clear support to earlier findings that the effect is more robust with foreign, unfamiliar languages (e.g., Sekiyama, 1997) presumably due to the fact that perceivers rely more on non-auditory information in order to back up and clarify the incoming speech input. As found earlier, and also here, when perceivers attend to non-native speech, they presumably use more visual information along with other sources, if available, for reasons summarized above and elsewhere. On the other hand, what seem to be important in determining the use of visual speech information are the phonological and phonotactic requirements of a given language. Cross-language studies conducted so far shed some light on this issue and reveal a somehow hierarchical and relative use of visual speech information within first languages, e.g., Japanese and English (Sekiyama, 1997; Sekiyama & Tohkura, 1993). Sekiyama found that while English speakers use more visual speech in their L1, their Japanese counterparts do this to a lesser extent. However, relatively speaking, Japanese speakers use this information more than their Mandarin-speaking counterparts. Sekiyama explained this primarily on the basis of the fact that English has a plethora of visually discernable consonant clusters (e.g., /kr/ as in crack) and a larger repertoire of vowels (around 15), whereas Japanese has no visually detectable consonant clusters and a very small number of vowels (around five) thus necessitating less need for visual speech information in Japanese than in English. Comparing Japanese and Mandarin, Sekiyama (1997) suggested that Mandarin is a tonal language where syllable-based pitch variations denote semantic differences wherein same syllables with different tones mean different things, e.g., a "ma" can mean five different things in Mandarin: mother (mā), to bother (má), horse (mǎ), to scold (mà), and an interrogative particle (ma) (Learn NC Editions, 2013). However, Japanese has only two such syllable-based pitch variations and thus in relative terms rendering Mandarin speakers in less need for visual speech information as, most probably, lexical tones are not as visually distinguishable as consonants and vowels (but see Burnham, Lau, Tam, & Schoknecht, 2001).

The aim of this study was to preliminarily test McGurk effect among native Turkish speakers and establish its status in the context of this language. As discussed above, the degree to which visual speech information is used in a given native language seems to be related to the phonological and phonotactic requirements. In this respect a further prediction can be advanced regarding the role of visual speech information in Turkish as an L1. Turkish is an agglutinated language with a complex morphological set of rules. For instance, what can be expressed with a few lexical items in other languages, say English, can be expressed with much less number of items in Turkish with subjects and objects embedded in the same lexical unit. For instance, the phrase "I can achieve" is expressed with a single lexical item in Turkish: "başarabilirim" (*başar*: achieve + *a* + *bilir*: be able to + *-im*: affix denoting first person). Thus further research in auditory-visual speech perception in the context of Turkish can focus on its relatively unique morphological structure shared with few other languages (e.g., Hungarian and Finnish). Such research can serve at least two purposes. First, this would provide researchers with an opportunity to explore, for the first time, the relationship between auditory-visual speech information and morphological formations. With its rich morphological structure, Turkish language is a very suitable context for this endeavor. Second, in an applied research sense, studying auditory-visual speech perception with native Turkish speakers should help us gain insight into the dynamics of the role of visual speech information in the context of second language (L2) acquisition. This has a very important practical implication because of a large number of Turkish students studying higher education institutes in Turkey and Cyprus that use English as a medium of instruction. Practical implementation of the findings in auditory-visual speech perception research in forming L2 curricula will definitely improve the teaching techniques in such educational settings (Erdener, 2012) as experimentally evidenced in earlier studies on L2 teaching (Davis & Kim, 2001; Hardison, 2003; Hazan, Sennema, Faulkner, & Ortega-Llebaria, 2006; Hazan, Sennema, Iba, & Faulkner, 2005; Ortega-Llebaria et al., 2001). Currently, we are analyzing data from native Turkish-speaking learners of English where we are looking into the role of visual speech information in L2 acquisition in the context of Turkish and the relationship between the amount of visual speech influence and the length of lexical units (Erdener & Pehlivan, in prep.).