Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

The Development of Auditory-Visual Speech Perception across Languages and Age

Doğu Erdener (1), Kaoru Sekiyama (2) and Denis Burnham (3)

(1) Psychology, Middle East Technical University, Northern Cyprus Campus, North Cyprus
(2) Cognitive Psychology Division, Kumamoto University, Kumamoto, Japan
(3) MARCS Labs, University of Western Sydney, Sydney, Australia

PACS: 43.71Ft; 43.71Hw; 43.71Es

ABSTRACT

To understand the now well-established auditory-visual nature of speech perception, it is necessary to understand how it develops. We know that young infants perceive speech auditory-visually by the fact that they perceive the auditory-visual illusion known as the McGurk effect; that visual information use increases over age in English-language children; and that Japanese-language adults use less visual information than do their English-language counterparts. Here we complete the developmental scene and probe the processes involved. In Experiment, with 6-, 8-, and 11-year-old and adult Japanese- and English-language participants tested on a McGurk task, while 6-year-olds from both language groups were equivalently influenced by visual speech information, there was a significant jump in auditory-visual speech perception between 6 and 8 years in English- but not Japanese-language participants. To investigate this further, in Experiment 2 we gave English-speaking 5-, 6-, 7- and 8-year-olds and adults a McGurk effect task as well as a language-specific speech perception (LSPP) test with native- and non-native speech sounds, and reading and articulation tests. For children, but not adults, visual-only speech perception (lipreading) ability and LSSP predicted McGurk performance – children with good auditory-visual speech perception tended to be those who focussed more on native than non-native speech sounds. In Experiment 3, with 3- and 4-year-olds tested for McGurk effect, LSSP, receptive vocabulary, and cognitive skill, regression analyses showed that auditory-only speech perception and cognitive skill, but not LSSP, predicted auditory-visual speech performance. Together the results show that there is an increase in auditory-visual speech perception between 6 and 8 years in English- but not Japanese-language children, and in English-language children this is related to language specific speech perception processes specifically around that age (5, 6, 7, 8 years) and not before (3, 4 years) or after (adults). It is suggested that LSSP is most variable and most predictive of visual influence in speech perception in the presence of significant linguistic challenges, such as those at the onset of reading instruction.

INTRODUCTION

The now widely acknowledged auditory-*visual* nature of speech perception (e.g., Campbell, Dodd & Burnham, 1998; Dodd & Campbell, 1987; Massaro, 1987, 1998; Stork & Hennecke, 1996) is evident both when visual lip and facial movements information of a talker compensate for degraded acoustic information (Sumby & Pollack, 1954), and also in the 'McGurk effect' (McGurk & MacDonald, 1976) in which auditory [ba] dubbed onto the face movements for [ga] are perceived as "da" or "tha". The McGurk effect shows that speech perception is an auditory-visual phenomenon even in undegraded conditions, and is a useful tool for investigating various processes of auditory-visual speech processing.

In the three studies reported here, the McGurk effect is used to address one overriding question: how does auditory-visual speech perception develop over the lifespan? Before enunciating the specific research questions literature regarding (a) developmental changes and (b) cross-language differences in auditory-visual speech perception will be considered.

With respect to development, one of the most important issues is to establish whether auditory-visual speech integration occurs in infancy. In this respect studies by Burnham and Dodd (2004), Desjardins and Werker (2004) and Rosenblum, Schmuckler and Johnson (1997) converge to show that infants from 4 to 5 months perceive the McGurk effect. This evidence might be taken to suggest that there is little environmental/linguistic influence on the development of auditory-visual integration. However, there is also evidence that auditory-visual speech perception (AVSP) improves over age.

In the original McGurk effect report both adults and children were tested and it was shown that 3- to 5-, and 7- to 8-year-olds have less visual influence in their perception of the McGurk effect than do adults (McGurk & MacDonald, 1976). This reduced visual influence in children is confirmed in later studies with 4- to 6-year-old children, compared with adults (Massaro, 1984; Massaro, Thompson, Barron & Laren, 1986); and also in a gradual developmental increase in visual influence across childhood in 5-, 7-, 9-, and 11-year-olds to adults (Hockley & Polka, 1994). This developmental increase appears to be related to speech articulation experience; preschool children who make articulation substitution errors have, compared to non-substituter children, poorer visual-only speech perception or lipreading, and a lower degree of visual influence, despite equivalent auditory-only speech perception (Desjardins, Rogers & Werker, 1997). In a related study, it has been shown that cerebral palsied adults, who lack experience of normal speech production, tend to show less visual influence in speech perception under some conditions than non-impaired adults (Siva, Stevens, Kuhl & Meltzoff, 1995). Together these results show that visual influence in speech perception is affected by the amount of both *general* experience (older children show greater visual influence), and *specific* experience, namely articulation.

Over and above the amount of general and specific experience, studies with adults suggest that the *type* is also important in visual influence; native speakers of Japanese are less subject to visual influence in the McGurk effect than native speakers of English (Kuhl, Tsuzaki, Tohkura & Meltzoff, 1994; Sekiyama & Tohkura, 1991, 1993; Massaro, Tsuzaki, Cohen, Gesi & Heredia, 1993). Moreover, Japanese more frequently notice the incompatibility between auditory and visual cues in McGurk-type stimuli than do English speakers (Sekiyama, 1994), suggesting that the Japanese tend to process the two sources of information separately.

Together these developmental and cross-language results provide an intriguing picture: on the one hand we know that while infants integrate auditory and visual speech information, the use of visual information increases over age, and on the other that there is less visual influence for adult speakers of Japanese than for adult speakers of English. Here, in Experiment 1 we combined the use of two approaches, the developmental and the cross-linguistic (Burnham & Sekiyama, in press), and tested children (6-, 8-, and 11-year-olds) and adults from Japanese and Australian English language backgrounds in order to pinpoint the age of emergence of differences in AVSP between Japanese- and English-language participants. Experiments 2 and 3 then investigate factors that might be involved in English-language school- and pre-school-children's auditory-visual speech perception.

EXPERIMENT 1: JAPANESE- AND ENGLISH-LANGUAGE CHILDREN AND ADULTS

Method

Participants: 16 English and 16 Japanese monolingual children were tested at each of three ages: 6 years (English mean age = 6.66 years, *SD* = 0.28 years; Japanese, 6.55 years (0.34years)), 8 years (English, 8.40 years (0.27 years); Japanese, 8.54 (0.35)), and 11 years (English, 11.63 years (0.20 years); Japanese 11.63 (0.20)). Forty-eight monolingual adults (24 English, 24 Japanese speakers, range = 18 to 29 years) also participated. All participants reported normal hearing and normal or corrected-to-normal vision. English language participants were tested at MARCS Auditory Laboratories (University of Western Sydney, Australia) or at nearby elementary schools, and the Japanese speakers at Future University, Hakodate, Japan or a nearby elementary school. At an early stage of the experiment, a bilingual experimenter confirmed that the equipment, procedure, and instructions were equivalent in the two countries.

Stimuli: As non-native stimuli induce more visual influence than native stimuli (de Gelder, Bertelson, Vroomen, & Chen, 1995; Fuster-Duran, 1996; Grassegger, 1995; Kuhl et al., 1994; Sekiyama & Tohkura, 1993); and to reduce talker differences that are often observed even within a language (e.g., Gagné, Masterson, Munhall, Bilida, & Querengesser, 1994; Sekiyama, Braida, Nishino, Hayashi & Tsuyo, 1995), speech materials were prepared with 2 speakers from each of the 2 languages and each participant was tested on all 4 speakers.

The stimuli consisted of [ba], [da], [ga] uttered by four talkers (2 English language and 2 Japanese language talkers, 1 male and 1 female in each language), selected for equivalent intelligibility of auditory and visual speech between the two languages. The utterances were videotaped, digitized, and edited on computer to produce audio-only (AO), video-only (VO), and audiovisual (AV) stimuli. Video digitizing was done at 29.97 frames/s in 640 x 480 pixels, and audio digitizing at 32 kHz in 16 bit. Each stimulus was created as a 2.3 s movie of a monosyllabic utterance. The duration of acoustic speech signals in each movie was approximately 330 ms. The movie file was edited with frame unit accuracy (33.3.ms), but the sound portion was additionally edited with 1 ms accuracy (for more details, see Sekiyama, Kanno, Miura, & Sugita, 2003). Half the AV stimuli were matching (e.g., auditory [ba], visual [ba]) and the other half McGurk-type mismatching (e.g., auditory [ba], visual [ga]). Three kinds of mismatching AV stimuli were created by combining within-talker auditory and visual components (auditory [ba] with visual [ga], auditory [da] or [ga] with visual [ba]). VO stimuli, one each for [ba], [da], and [ga], were created by cutting out the audio track. In the AO stimuli, one each for [ba], [da], and [ga], the video of the talking face was replaced by a still face of the talker with the mouth neutrally closed. In total, there were 12 auditory (3 consonants x 4 talkers), 12 visual, and 24 audiovisual stimuli (3 auditory consonants x 2 congruity types x 4 talkers). To obtain a wide range of data, we introduced 4 levels of auditory intelligibility by adding band noise (300 Hz – 12000 Hz) with signal-to-noise (SN) ratios of -4, +4, and +12 dB, together with a no-noise condition.

Procedure: The stimuli were presented from computer (Sharp MJ730R) onto a 17-in CRT monitor (Sony 17GS) and through a loudspeaker (Aiwa SC-B10). Experimental conditions were blocked depending on the modality (AO, VO, AV) and the SN ratio of the auditory stimuli (-4, +4, +12 dB, and Clear), and there were 2 repetitions of each stimulus in a block. Each participant was given the AV condition first. Half the subjects were presented with the stimuli in an AV, AO, VO order, and the other half in an AV, VO, AO order. In the AV and A conditions, speech was presented at 65 dB and the SN ratios, -4, +4, and +12 dB, were determined by the intensity of the added band noise. In the 'Clear' condition, no noise was added. S/N ratios varied across blocks in an increasing manner for half the subjects, and in a decreasing manner for the remainder.. Within each block, stimuli were presented in random order. Participants were asked to watch and listen to each stimulus, decide what they perceived, and press 1 of 3 buttons for a "ba," "da," or "ga" response accurately and without delay.

After each movie file was played, the last frame remained on the screen until one of the 3 buttons was pressed. Onset of the next stimulus was 1.5 s after the button press. Responses were made on a game controller, which input to and were stored on the computer. Before starting 6 practice trials were given to ensure participants understood each (AV, AO, VO) task. The experiment took around 25 minutes for adults and 40 to 60 minutes for children.

Results

Each participant's responses were averaged across syllables and the 2 talkers within each stimulus language. Figure 1 shows the percent correct responses in the matching (AV+), mismatching (AV-), and AO conditions as a function of SN ratio combined across stimulus languages (native and non-native). The size of visual influence was examined via the difference in percent auditorally correct responses between the AV+ and the AV- conditions. As can be seen in Figure 1, the difference between AV+ and AV- conditions is small across all ages for Japanese language participants, correspondingly small for English-language 6-year-olds, but much larger in English-language 8-year-olds, 11-year-olds, and adults. This can be more clearly seen in Figure 2 in which data are collapsed over S/N levels and presented for native (Japanese speakers for the Japanese language participants, and English for the English) and non-native speakers (Japanese for the English, and English for the Japanese).

Analysis of variance (ANOVA) of the degree of visual influence (2 language groups x 4 age groups x 2 stimulus languages x 4 SN ratios) showed a main effect of language group, F (1, 136)=46.991, $p<0.001$, indicating greater visual influence for English language participants. Significant age-related main effects and interactions with language group were found for (i) children vs. adults, F (1, 136)=9.303, $p<0.01$, and F (1, 136)=5.728, $p<0.05$, respectively, and (ii) 6 years vs. 8 and 11 years, F (1, 136)=10.813, $p<0.01$ and F (1, 136)=5.178, $p<0.05$, respectively, indicating that the degree of visual influence increased between 6 and 8 years, but only for the English language children.

There was a significant main effect of stimulus language, F (1, 136=6.905, $p<0.01$), and a significant stimulus language x 11 years vs. adults interaction, F (1, 136)=4.164, $p<0.05$, indicating that over and above a larger visual influence for non-native stimuli, the non-native visual influence advantage increased between 11 years and adults.

Finally, regarding the SN ratio factor, there were significant main effects of noise vs. clear, F (1, 136)=219.728, $p<0.001$, and the linear and quadratic trends of SN ratio, F (1, 136)=364.849, $p<0.001$; and F (1, 136)=15.385, $p<0.001$, respectively, indicating that the size of visual influence was larger at lower SN ratios.

In the AO condition (see Figure 3), accuracy of auditory speech perception generally increased over age in both language groups. ANOVA (2 language groups x 4 age groups x 2 stimulus languages x 4 SN ratios) showed significant age-related effects for children vs. adults, F (1, 136)=38.803, $p<0.001$, and 11 years vs. adults, F (1, 136)=13.449, $p<0.001$] indicating developmental improvement especially between 11 years and adulthood. In younger children, however, a language group x 6 vs. 8 and 11years interaction, F (1, 132)=4.377, $p<0.05$, indicated better AO performance in Japanese than English language children at 6 years, but not later ages. This early auditory superiority in the Japanese may be related to the greater dependence on auditory cues and less dependence on visual cues in AVSP at older ages.

There was no main effect of stimulus language in the AO condition, F (1, 136)=0.767, but there was a significant language group x stimulus language interaction, F (1, 136)=4.113, $p<0.05$, indicating an advantage for native AO stimuli only for the Japanese participants. Auditory speech perception was poorer at lower SN ratios (Fig. 7), as indicated by significant effects of noise vs. clear, F (1, 136)=987.795, $p<0.001$, linear and quadratic SN ratio trends, F (1, 136)=771.207, $p<0.001$; F (1, 136)=38.174, $p<0.001$, respectively.

In the visual-only (VO) condition (see Figure 4), lipreading scores increased over age in both language groups. ANOVA (2 language groups x 4 age groups x 2 stimulus languages) revealed significant age-related effects for children vs. adults, F (1, 136)=32.781, $p<0.001$, 6 vs. 8 and 11 years, F (1, 136)=11.315, $p<0.01$, and 8 vs. 11 years, F (1, 136)=9.842,

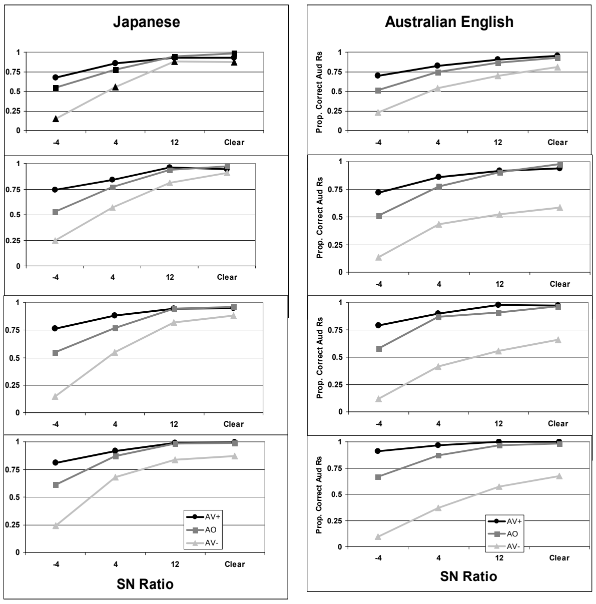

Figure 1. Proportion correct auditory responses in matching (AV+) & mismatching (AV-) auditory-visual, & auditory-only (AO) presentations for Japanese- & Australian English-language 6-, 8-, & 11-year-old children, & adults.

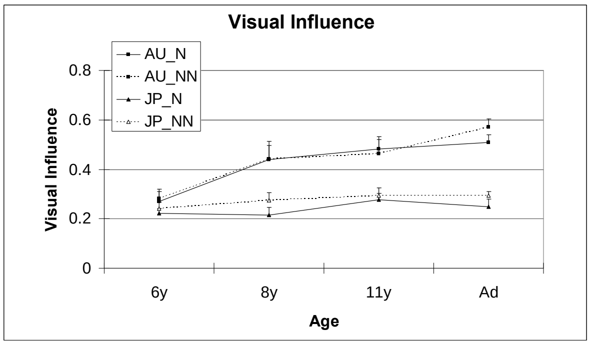

Figure 2. Proportion visual influence (AV+ minus AV-) for Japanese (JP) and Australian English (AU) language adults and 4-, 6-,

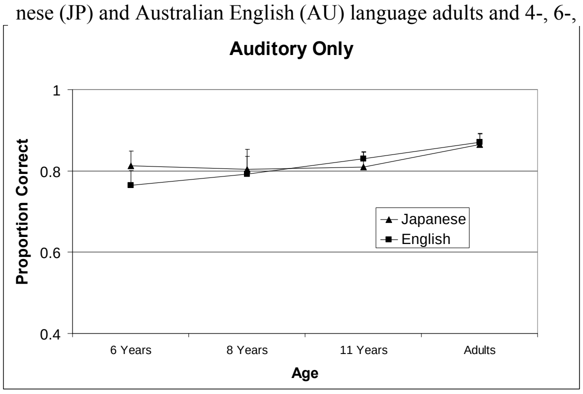

Figure 3. Proportion correct auditory only (AO) for Japanese & Australian English language adults & 4-, 6-, & 8-year-old children.

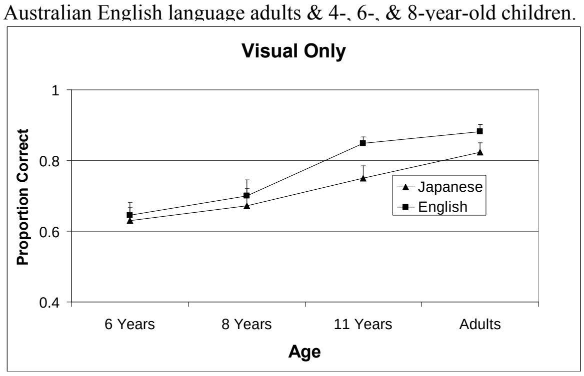

Figure 4. Proportion correct visual only (VO) for Japanese & Australian English language adults & 4-, 6-, & 8-year-old children.

$p<0.01$, indicating an increase in lipreading performance up to 11 years. There was a significant main effect of language group, F (1, 136)=4.441, $p<0.05$, indicating that the lipreading performance was generally better in English than in Japanese language participants. To investigate this further, individual tests at each age were conducted in separate 2 language group x 2 stimulus language ANOVAs, revealing that the main effect of language group was significant only at 11 years [for 6, 8, and 11 years, F (1, 30)=0.011; F (1, 30)=0.308; F (1, 30)=6.217, $p<0.05$, respectively]. Thus, it seems that inter-language differences in lipreading accuracy, if any, do not emerge until 11 years.

VO performance did not differ overall between native and non-native stimuli, F (1, 136)=3.088, $p<0.10$, but a group x stimulus language interaction, F (1, 136)=20.689, $p<0.001$, indicated that native stimuli were better lipread by Japanese participants whereas non-native stimuli were better lipread by English language participants. This suggests that the Japanese talker stimuli were visibly more intelligible than those presented by the English talkers, presumably due to some individual talker differences. Significant age x stimulus language interactions were found only for children vs. adults x stimulus languages, and 11y vs. adults x stimulus language, F (1, 136)=7.493, $p<0.01$; F (1, 136)=6.620, $p<0.05$, respectively, indicating that intelligibility differences are more reliably detectable at later ages.

Discussion

For AO stimulus presentations as well as a general improvement over age, Japanese-language participants outperformed their English-language counterparts at 6 years but at no other age. On the other hand, for VO stimulus presentations, as well as a general improvement over age, English-language participants outperformed their Japanese-language counterparts but only after 11 years of age. Thus there is an initial AO advantage for Japanese-language participants that dissipate at 8 years; and a VO advantage for English-language participants that emerges only after 8 years.

The degree of visual influence is equivalent in Japanese- and English-language 6-year-olds, but between 6 and 8 years there is a significant jump in English, such that the degree of visual influence in 8-year-olds, 11-year-olds and adults is greater in English- than Japanese -language participants.

It appears that there is something about the English language (i) that necessitates relatively greater use of visual speech information, and (ii) that becomes critical in auditory-visual speech perception after but not before the age of 6 years.

With regard to the first issue, the relatively *greater use of visual speech information*, why the Japanese 6-year-olds are more accurate than their English counterparts in AO perception is unknown, but it may be due to the less crowded phoneme space in Japanese language. Japanese syllable identification may be less difficult due to the smaller number of vowels (5 vs. around 14 in English); the lack of some consonant contrasts that occur in English (e.g., /r/ vs. /l/; /b/ vs. /v/; and /s/ vs. /θ/); simpler more regular syllable structure than in English; and no word-initial or word-final consonant clusters (compared to at least 31 in English). English syllable identification may thus be more difficult via auditory information alone, which could result in greater susceptibility to augmentation of speech perception by visual cues. And it just so happens that compared to Japanese, English incorporates visually- but not so auditorally-distinct consonant contrasts, e.g., labiodental-interdental-alveolar, as in ‘four-thaw-saw’, or ‘vat-that-sat’, which do not exist in Japanese. The phonological complexity of English could provide pressure to seek extra sources of information, with the visual distinctiveness of English phonology providing an effective source of such information. In accord with this argument, English-language 8- and 11-year-old children and adults show more visual influence in auditory-visual speech perception than their Japanese counterparts; and English-language were generally more accurate than the Japanese-language participants in VO perception. This difference was significant only at 11 years, a relatively late visual superiority that may be a consequence of relatively more visually-tuned speech perception in young English language participants.

With respect to the second issue, the *timing* of the emergence of greater visual influence in English-language participants, could it be that there is something about the onset of schooling that brings on this change? One significant language-related element of early schooling is reading instruction. The possible involvement of reading and related elements in English-language children’s greater use of visual information between 6 and 8 years is investigated in Experiment 2.

EXPERIMENT 2: AUDITORY-VISUAL- & LANGUAGE-SPECIFIC SPEECH-PERCEPTION IN ENGLISH-LANGUAGE SCHOOL-AGE CHILDREN

This study was designed to examine the basis of increased visual influence in auditory-visual speech perception in English-language children between 6 and 8 years found in Experiment 1. As this occurs in a period in which children begin school and, more specifically, early reading instruction, it was hypothesised that the increase in visual influence might have something to do with reading and language abilities.

5-, 6-, 7-, and 8-year-old children and adults were tested on an auditory-visual speech perception task (similar to that in Experiment 1), reading ability, articulation skill, and a language specific speech perception task. Reading ability was included as this is a significant new and complex ability that children are exposed in their early school years. Articulation was included for, as shown above in the introduction, articulation proficiency (Desjardins et al., 1997) and experience (Siva, Stevens, Kuhl & Meltzoff, 1995) are related to visual influence in speech perception. With respect to language specific speech perception (LSSP), Burnham (2003) has shown with English-language 4-, 6-, and 8-year olds that there is a significant attenuation of perceptual ability with non-native speech contrasts compared to native speech contrasts at 6 years compared to their non-native speech perception ability both before (4 years) and after (8 years). Moreover, it was found that degree of LSSP (perceptual performance for native *minus* non-native speech contrasts) significantly predicted reading ability; good readers were those who were better able to filter out non-native speech sounds.

Method

Participants: 48 first year psychology students participated. 96 monolingual English-speaking children were tested, 24 (12 males, 12 females) in each of 4 age groups: 5 years (mean = 5.4 years), 6 years (6.67), 7 years (7.59), and 8 years (8.52). All participants had normal vision and hearing. 2 additional 5-year-olds withdrew from the experiment and their data were excluded from analyses.

Design: A 2-factor design, age (5 levels: 4 child groups and adults) by task (4 levels: AVSP, LSSP, reading, articulation), was employed with repeated measures on the second factor.

Stimulus Materials and Procedure: *Visual Influence in Auditory-Visual Speech Perception:* The auditory-visual McGurk-

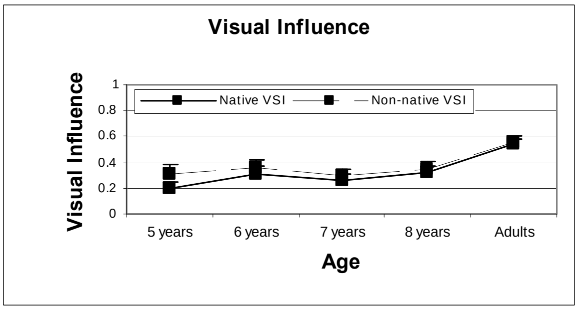

Figure 5. Degree of visual influence (AV+ minus AV-) across age in response to Native & Non-native speakers.

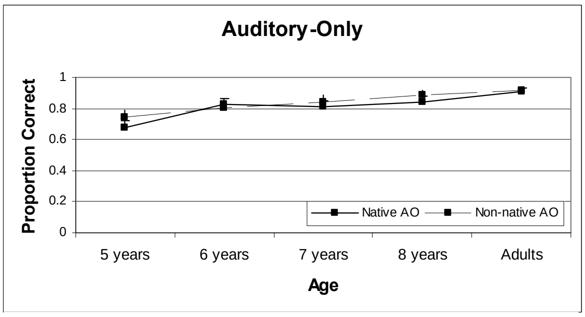

Figure 6. Proportion correct Auditory-only across age in response to Native & Non-native speakers.

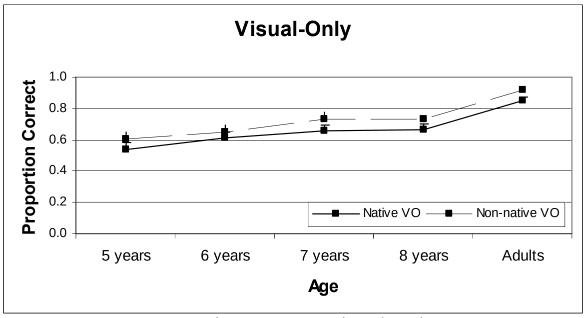

Figure 7. Proportion correct Visual-only across age.

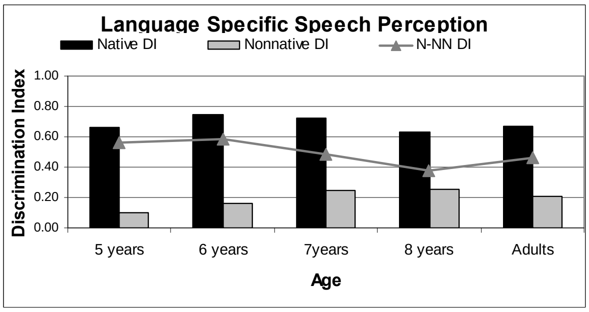

Figure 8. Native & Non-native Speech perception over age

style stimuli were from the same pool of 3 stimuli ([ba], [da] and [ga]) and 4 talkers (2 monolingual Japanese, 2 monolingual English, I male, 1 female of each) as those in Experiment 1. There were 12 AO (3 consonants x 4 talkers), 12 VO (3 consonants x 4 talkers), and 24 AV stimuli (3 auditory consonants x 2 congruity types x 4 talkers). The signal level of the AV and AO stimuli was set at 65 dB, and for AV and AO stimuli there were 2 noise conditions; clear and S/N ratio = +4dB. The testing protocol was as in Experiment 1.

Language-specific Speech Perception (LSSP): The LSSP test consisted of combinations of 3 Thai speech contrasts: native speech contrasts for English language listeners [ba]-[p^ha], and [pa]-[p^ha], and the non-native speech contrast, [ba]-[pa]. The stimuli were presented such that there were 18 native and 18 non-native contrasts using DMDX software for a typical AX discrimination task paradigm (Forster & Forster, 2003). Native and non-native speech contrasts were presented in 2 separate blocks with order of stimuli and blocks randomised.

Reading: The reading subtest of Wide Range Activities Test (WRAT-3) was used (Wilkinson, 1993), involving 2 parts: letter (15 items) and word (42 items) reading.

Articulation: The Queensland Articulation Test (QAT) was used (Kilminster & Laird, 1978). A picture-naming test, it consists of 64 items, which includes all Australian English consonants presented in initial, medial and final positions, consistent with the phonotactic constraints of English.

Results

ANOVAs are presented in turn for each dependent variable ahead of regression analyses.

Visual Influence As in Experiment 1 visual influence scores were derived by subtracting mismatching AV- percent correct scores from the matching AV+ scores. Visual Influence scores for the two sets of stimulus speakers English (Native) and Japanese (Non-native) are shown in Figure 5 collapsed over noise level for each of the ages (5, 6, 7, 8, years, adults). These scores were subjected to a 5 x (2 x 2) (age x [noise/clear x stimulus language]) ANOVA) with repeated measures on the last 2 factors. Overall, adults showed more visual influence in speech perception than children, $F(1,139)= 47.49$, $p<.001$, but there were no significant effects across child age. There was greater visual influence in the noise than the clear condition, $F(1,139)= 109.81$, $p<.001$, with greater effect of noise for adults than for children, $F(1,139)= 5.99$, $p<.05$. There was also a main effect of stimulus language, $F(1,139)= 8.94$, $p<.01$ – showing greater visual influence when the English language participants viewed non-native (Japanese language) speakers. Unexpectedly, despite a slight improvement, there was no significant improvement over child age.

Auditory-only data are shown in Figure 6. ANOVA revealed that adults performed better than children, $F(1,139)=23.10$, $p<.001$, and across child age there was a linear improvement, $F(1,139)=17.38$, $p<.01$, with greater improvement when listening to native than non-native language speakers between 5 and 6 years, $F(1,139)=4.58$, $p<.05$. Additionally, AO scores were higher in clear than noise, $F(1,139)=74.16$, $p<.001$, and they also increased with age more in the clear than in the noise condition, $F(1,139)=6.34$, $p<.05$.

VO scores are shown in Figure 7. ANOVA revealed that adults performed better than children, $F(1,139)= 80.3$, $p<.001$, and that across child ages there was a linear increase, $F(1,139)= 11.11$, $p<.01$. There was better VO performance with foreign than familiar talkers, $F(1,139)= 66.270$, $p<.001$.

LSSP: For LSSP the dependent variable was the discrimination index (DI) score, given by the difference between the number of correct "different" responses on different (AB) trials (hits) and the number of incorrect "different" responses on same (AA) trials (false positives) divided by the total number of trials, for native (N-DI) minus non-native (NN-DI) speech contrasts. These data were subjected to a 5 x 2 (age x native/non-native) ANOVA. Figure 8 shows the overall speech perception data in five age groups for native and non-native language stimuli. An overall effect of native language was observed, $F(1,139)=194.90$, $p<.001$ with no difference between children and adults ($p>.05$).

Reading and Articulation: For reading and for articulation the dependent variables were the proportion of items correct, shown in Table 1. For reading, ANOVA showed a significant difference between adults and children, $F(1,139) = 598.3$, $p<$

.001, as well as a significant linear improvement of reading over age, F (1,139) = 370.85, p< .001. For articulation ANOVA showed adults' articulation was better than that of the children, F (1, 139) = 7.86, p< .05, and there was significant improvement over child age with the greatest increase occurring between 5 and 6 years, F_{linear} (1, 139) = 99.51, p< .001; $F_{quadratic}$ (1, 139) = 14.47, p< .01.

Table 1. Reading & Articulation proportion correct over age

	5 years	6 years	7 years	8 years	Adults
Reading	0.17 *(.12)*	0.40 *(.10)*	0.53 *(.09)*	0.67 *(.09)*	0.84 *(.07)*
Articulation	0.87 *(.05)*	0.93 *(.03)*	0.96 *(.03)*	0.97 *(.02)*	0.95 *(.03)*

Regression Analyses: Sequential multiple regression analyses were conducted separately for the child and the adult data.

For the child data Visual Influence scores were set as the criterion and 6 variables entered as predictors in order of the developmental variable (age) and then the 5 linguistic variables: AO, VO, N-NN DI, Articulation, and Reading. To meet the assumptions 2 outliers with a standardized residual > 3 standard deviations were omitted from the analysis. Results show that among all variables VO in the third step (R=.337, R^2=.113, F(1,90) = 9.50, p<.01), and N-NN DI in the fourth step (R=.396, R^2=.157, F(1,89) = 4.60, p<.05) reliably predicted Visual Influence scores (see Table 2).

Table 2. Regression of age, AO, VO, N-NN, articulation & reading as predictors of Visual Influence Scores in children.

Step	Variables	β at Step	ΔR^2 at Step	F-value	Final β(Step 6)
1	Age	.020	.020	1.834	.182
2	AO	.014	.000	.011	-.042
3	VO	.314	.094**	9.502	.310**
4	N-NN	.089	.044*	4.596	.215*
5	Artic.	-.184	.002	.184	-.046
6	Reading	-.054	.001	.113	-.068

* significant at .05; ** significant at .01

For the adult data the same sequential multiple regression analysis was performed. Inspection of Table 3 reveals that only AO scores at the first step significantly predict Visual Influence scores and reliably improve R^2 (R=.394, R^2=.155, F(1,46)=8.46, p<.01).

Table 3. Regression of age, AO, VO, N-NN, articulation & reading as predictors of Visual Influence Scores in adults.

Step	Variables	β at Step	ΔR^2 at Step	F-value	Final β(Step 6)
1	AO	.394	.155**	8.457	.323**
2	VO	.179	.030	1.630	.184
3	N-NN	-.077	.006	.314	-.069
4	Artic.	.052	.003	.142	.049
5	Reading	.019	.000	.015	.019

* significant at .05; ** significant at .01

Discussion

Both the degree of visual influence and lipreading (VO) ability by English language participants increases from childhood (5, 6, 7, 8 years) to adulthood lending support to previous results (Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008). In addition, VO scores increased over child age though, unexpectedly, visual influence in speech perception did not. Most importantly, it was found that LSSP and lipreading reliably predict visual speech influence in children, but that auditory speech perception alone predicts visual speech influence in adults. These results show (i) that for children, auditory-visual speech perception is not only a reflection of lipreading, but is also related to the degree to which their speech perception is tuned to the native language, and (ii) that the determinants of auditory-visual speech perception differ between children and adults.

Why does greater relative attention to native over non-native speech contrasts predict visual influence in children's speech perception? Burnham (2003) showed that reading ability at the onset of reading instruction is related to language specific speech perception in English-speaking children and reasoned that this is because learning to read entails high cognitive demand. Indeed, one primary task in learning to read is the two-way phoneme←→grapheme mapping, which rests upon (i) categorisation of groups of phones into native language phoneme categories, and so entails attending to phonemic differences between and ignoring phonetic differences within phoneme categories; and (ii) matching all the allophonic variations within these phoneme classes to grapheme-defined labels. This is especially difficult in English because of its opaque orthography with one-to-many grapheme-to-phoneme correspondences (van den Bosch, Content, Daelemans, & De Gelder, 1994), and its phonological complexity, so it may be that learning these labels actually intensifies LSSP and the assimilation of non-native sounds into native language phoneme classes (Burnham, Tyler, & Horlyck, 2002).

Given such challenges, children may well seek any extra information they can to establish phoneme-grapheme links, and it would appear that auditory-visual speech information might just be one of those extra sources of information. Indeed, as pointed out in Experiment 1, on the other side of the phonological complexity coin, English is characterised by a relatively high number of vowels and complex consonant clusters with many visually-, but not auditorally-distinct sounds, e.g., /θ/ vs. /f/ vs. /s/, as in 'thin' vs. 'fin' vs. 'sin'.

In this light the childhood relationship between attention to native (phonemic) over non-native (phonetic) speech distinctions and visual influence in speech perception makes intuitive sense and is consistent with research suggesting a link between reading and lipreading skills (de Gelder & Vroomen, 1998), and showing that dyslexic children use visual speech information less than their normal-reading counterparts (Cavé, Stroumza, & Bastien-Toniazzo, 2007). However, the adult data show a different pattern. Why?

In contrast to the child data, there is no effect of lipreading or LSSP on visual influence in adults; rather auditory-only speech perception is the sole predictor of individual differences in visual influence. This may be because lipreading ability and LSSP only predict visual influence in the presence of significant linguistic challenges, such as those at the onset of reading instruction (Burnham, 2003). Nevertheless, this does not mean that visual speech information is insignificant for adults; on the contrary, adults show greater visual influence than children. By adulthood both reading and native vs. non-native speech perception are usually well established. Accordingly, there are few significant linguistic challenges for monolingual adults and so any subtle individual differences in visual influence in speech perception may no longer be predicted by LSSP or lipreading. They are best predicted by auditory-only speech perception and in this regard Sekiyama and Burnham (2008) found that Japanese children's AO speech perception is better than that of English language children, and that Japanese-language adults have faster reaction times for AO speech judgments than do English-language adults, whereas English-language 11-year-olds have better visual-only speech perception. Subtle vestiges of such differences may underlie those showing up in adulthood here, but may be masked by lipreading and language specific speech perception in times of linguistic challenge, such as when reading instruction begins.

EXPERIMENT 3: AUDITORY-VISUAL- AND LANGUAGE-SPECIFIC-SPEECH-PERCEPTION IN ENGLISH-LANGUAGE PRE-SCHOOLERS

The third study was designed to extend the investigation in Experiment 2 by conducting a similar test with pre-school pre-reading children. Three- and 4-year-old children were tested. Again measures of visual influence in speech perception, and language specific speech perception (LSSP) were taken. In addition, children were tested for age-appropriate cognitive and linguistic ability. For cognitive function, a test of rule abstraction and cognitive flexibility was used to tap children's cognitive flexibility. Auditory-visual speech perception requires perception and integration of information from multiple sources, so it is possible that cognitive flexibility is related to developing AVSP and lipreading in this 3- to 4-year period (e.g., Desjardins et al., 1997). Linguistic ability was tested via receptive vocabulary, a particularly apt diagnostic at this age, as studies have shown that speech perception development is linked to vocabulary development (Nazzi & Bertoncini, 2003; Stager & Werker, 1997).

Method

Participants: Twenty-four 3-year-old (M_{age}=3.08 years, *sd*=0.17 years) and 24 4-year-old (M_{age}=4.21 years, *sd*=0.16 years) preschool children were tested. All were from monolingual English-speaking families and had normal hearing and vision.

Design: A two-factor, age (2 levels: 3-, and 4-year-olds) by task (4 levels: AVSP, LSSP [N-NN], executive function, and receptive vocabulary knowledge) design, with repeated measures on the task factor was applied.

Stimuli, Apparatus and Procedure: All children were tested individually in the presence of a parent. Children were given four tasks in counterbalanced order: a McGurk-like visual influence test, an LSSP task, a flexible item selection test (FIST), and the Peabody Picture Vocabulary Test.

Visual Influence: Visual influence was measured using an AX discrimination task in which children were asked to indicate whether two AV stimuli were "same" or "different". All auditory-visual speech stimuli were produced by the same native English speakers as in Experiments 1 and 2. There were 12 AO, 12 VO and 36 auditory-visual speech stimuli, a total of 60. The speech stimuli were utterances of [ba], [da], and [ga], and comprised the AV (congruent and incongruent), AO, and VO experimental stimuli. Children were presented with as many practice items as required. On each trial children were presented with 2 AV speech combinations and asked whether they were the same or different. The task took between 20 and 40 minutes. A visual speech index (VSI) score was calculated using only 'different' auditory-visual incongruent (AV^-) speech pairs, which differed either on both auditory and visual components or on the visual component alone. For example, a 'same' response to Auditory-[ba]+Visual[ga] vs. Auditory-[ga]+Visual[ga] was deemed a visually-based response, and given a score of '1'. The resultant total number of visually-based responses to 'different' trials was divided by the total number of mismatched different trials (9), to give a proportion VSI score.

Language Specific Speech Perception: For LSSP, the same Thai speech contrasts as in Experiment 2 were used. Stimuli were presented on laptop computer, using *PsyScript* software (Bates & D'Oliveiro, 2003) via a Go-NoGo *category change* paradigm. In this, in *change* trials one sound was presented from 2 to up to 6 times (randomly varying across trials) after which a second (change) sound was presented and played from 4 up to 8 times. In *no-change* (or same) trials the same sound was played 10 times. Correct responses (hits in change, correct rejections in no change trials) were rewarded with a 5-second excerpt from a cartoon story.
Cognitive skill: The *Flexible Item Selection Task (FIST)* (Jacques & Zelazo, 2001) was used to measure two executive function constructs: *rule abstraction* and *cognitive flexibility*. 18 trials were presented in 3 sections: *item identification task*, *favourite item selection task*, and *flexible item selection task*. The test produces 2 scores: a rule abstraction score (max.15), and a cognitive flexibility score (max.15).

Vocabulary: Receptive vocabulary knowledge was measured with the Peabody Picture Vocabulary Test (PPVT-III) (Dunn & Dunn, 1997). Responses are extracted via 12 blocks of a picture naming task that progress in order of difficulty. Each child began with the block of items appropriate for their age group. In each trial children were shown 4 pictures of animals or objects and asked to point to the target word named by the experimenter. Testing progressed until the child made 8 errors in a block or until all 12 blocks were completed.

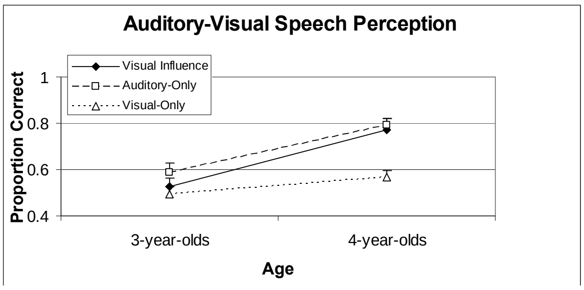

Figure 9: Discrimation Index (max = 1) for Visual Influence and Proportion Correct for Auditory-only & Visual-only trials for 3 - & 4-year-olds

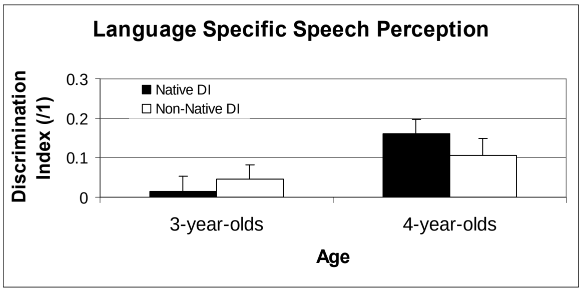

Figure 10: Discrimination Index (max = 1) for perception of Native & Non-Native speech sounds in 3 - & 4-year-

Results

Between-subject comparisons of the two ages using *t*-test and ANOVA are presented ahead of the results of a set of regression analyses to predict visual influence.

The VSI, AO, and VO data are presented in Figure 9. The 4-year-olds scored significantly better on VSI, $t(46)=3.927$, $p<.001$, and on AO, $t(46)=4.171$, $p<.001$, and marginally better on VO, $t(46)=1.860$, $p=.07$ than the 3-year-olds.

The N-NN DI scores are presented in Figure 10. A 2 x (2) (age x native/non-native) ANOVA revealed significantly greater scores for 4- than 3-year-olds, $F(1,46)= 7.709$, $p<.01$, but no native vs. non-native effect, nor any interactions.

For the FIST data (see Table 4) a 2 x (2) (age x rule abstraction/cognitive flexibility) ANOVA revealed significant improvement between 3 and 4 years of age, $F(1,46)= 10.078$, $p<.001$, a main effect of flexibility over rule abstraction, $F(1,46)= 131.324$, $p<.001$, but no interactions.

Finally, the 4-year-olds' vocabulary (see Table 4) was significantly better than the 3-year-olds', $t(46)=-6.145$, $p<.001$.
Regression Analyses: VSI scores were entered as the dependent variable and developmental (age) and speech (AO, VO, N-NN DI), cognitive (executive function) and linguistic (vocabulary) variables entered as predictors. Of the 5 predictors entered into the equation, only AO at Step 2 and FIST scores at Step 6 reliably predicted VSI and increased the R^2. whereas age at Step 1, VO scores at Step 3, N-NN DI at Step 4, and vocabulary scores at Step 5 did not reliably increase R^2 or predict VSI scores (Table 5).

Table 4: Percent and percentile scores for cognitive skills and vocabulary in 3- and 4-year-olds

	% Cognitive Flexibility	% Cognitive Rules	Vocabulary (percentile)
3-yos	80.87	31.93	43.29
4-yos	98.33	54.17	67.58

Table 5. Sequential multiple regression: AO, VO, N-NN DI, FIST & PPVT scores as predictors of VSI.

Step	Variables	β at Step	Δ R^2 at Step	F-value	Final β (at Step 5)
1	Age	.180	.445	.198	11.390
2	AO	.493	.424	.346**	10.176
3	VO	-.003	-.002	.346	.000
4	N-NN	.155	.189	.380	2.359
5	PPVT	-.001	-.041	.381	.045
6	FIST	.099	.340	.451*	5.196

* significant at .05; ** significant at .01

Discussion

The results show that visual influence in speech perception and visual-only speech perception (lipreading) improve between 3 and 4 years in line with previous studies (Massaro et al., 1986; Burnham & Sekiyama, 2008; Sekiyama et al., 2003; Experiment 2 here). In addition auditory-only speech perception improves over age, as does cognitive and vocabulary skills. Both native and non-native speech perception improve over age, but it should be noted that scores at both ages were very low and standard errors high, suggesting that there may have been some task-related difficulty which was less intrusive at the older age.

The regression analyses reveal that the degree of visual influence in AVSP was predicted by AO scores and overall executive function. LSSP did not reliably predict visual influence in speech perception by 3- and 4-year-old preschool children. This is similar to the results obtained in Experiment 2 with adults, i.e., for both groups, adults and preschool children, AO scores predict visual influence in speech perception. There are two possible reasons for this, one methodological and the other theoretical. First, the possibility that there is a floor effect for LSSP scores must be seriously entertained, given the low means and high variability of scores. It could be argued that the variability does in fact reflect individual differences and so these should have power to predict visual influence if there is indeed a relationship, but the low means make this unlikely. The other reason, which we will entertain until further studies can be conducted in which higher LSSP scores can be obtained, is considered below in the Conclusions in the light of the findings from this experiment and those of Experiment 2.

CONCLUSIONS

Three facts about AVSP development have been found.

First, we have established the developmental origin of the greater use of visual speech information by English-language over Japanese-language adults – between 6 and 8 years English-language children begin to use more visual information in their speech perception and this increased use continues on until adulthood. We posit that the reason for this increase in English- but not Japanese-language children is due to the nature of the two languages. Japanese has a simple, even elegant, phonetic structure whereas English has a relative surfeit of vowels, and consonant clusters. Fortunately English also has many visually distinct phonemes and phoneme combinations and children appear to learn to use these around reading onset when such cues would be useful in disambiguating the often opaque phoneme-to-grapheme conversion process. Indeed, it could be suggested that it is the difficulty of such mapping in learning to read English that causes the intensification of both LSSP and increased use of visual speech information. The second finding provides some support for this notion.

Second, we have established that the increased use of visual speech information in English-language children is related to language specific speech perception – those children who are good at filtering out irrelevant non-native speech sounds when perceiving speech are just those children who are good at using visual information to augment their speech perception. Whether there is a causal relationship between LSSP and the AVSP abilities in one direction or the other cannot be determined here. Indeed the relationship could be due to a third (unspecified) variable, a speech perception proficiency factor that facilitates focus on the most useful speech perception information (and possibly in other learning contexts, where the focus involves both adding sources of relevant and filtering sources of irrelevant information). The third finding may elaborate the causal connection issue a little.

Third, we have found that the positive relationship between LSSP and AVSP holds for school age children, but not adults or preschoolers. Why? And why does AO performance predict AVSP in adults and preschool but not school children? A preliminary explanation is given below, though more research is required to elaborate this issue more precisely.

It is reasonable to posit an underlying relationship between LSSP and visual influence in AVSP; to posit that those children who are good at filtering out irrelevant speech information are just those children who are good at attending to additional information that is useful in both perceiving speech and facilitating phoneme-to-grapheme mapping in reading. Following on from this, it could be suggested that the new linguistic challenges at school such as reading, and hearing and interpreting new speech styles, dialects and accents, are met if the child has well-developed LSSP. Indeed there is an accentuation of LSSP at this age and this is related to reading ability (Burnham, 2003). Such accentuation may potentiate any individual differences in LSSP and thus highlight the underlying relationship between LSSP and visual influence in AVSP. However, by adulthood, such linguistic challenges are not present; reading has become automatic, and is usually no longer linguistically challenging, hence the underlying LSSP-AVSP relationship is not so apparent. The same argument could be made for 3 – and 4-year-olds, but further information is required before definitive conclusions can be made in this regard.

At a more general level, the three findings here – the origin of cross-language AVSP differences in childhood, the relationship between LSSP and AVSP around reading onset, but not at other times – suggest that auditory-visual *speech* perception is a special case of auditory-visual perception. It remains for future research to investigate further issues in auditory-visual *speech* perception. For example, it remains to be seen how auditory-visual speech perception develops in the face of phonological delay. In addition, it would be of interest to investigate LSSP and auditory-visual speech perception at times when *adults* are linguistically challenged such as when learning a second language. For now though, we have filled in the developmental landscape for auditory-visual speech perception quite considerably, and have provided the empirical and a tentative theoretical basis for future research

REFERENCES

Bates, T. C., & D'Oliveiro, L. (2003). PsyScript: A Macintosh application for scripting experiments. *Behav. Research Methods, Instruments, & Computers*, **35**, 565-576.

Burnham, D., & Sekiyama, K. (in press). Investigating auditory-visual speech perception development using the ontogenetic and differential language methods. In E. Vatikiotis-Bateson, G. Bailly & Perrier, P. (Eds.) *Advances in auditory-visual speech processing*. Cambridge, MA: MIT Press.

Burnham, D., Tyler, M., & Horlyck, S. (2002). Periods of speech perception development and their vestiges in adulthood. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode.* (pp 281-300). Trier, Germany: Wissenschaftlicher Verlag Trier.

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

Campbell, R., Dodd, B., & Burnham, D. (1998). *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech*. Hove: Psychology Press.

de Gelder, B., Bertelson, P., Vroomen, J., & Chen, H.C. (1995). Interlanguage differences in the McGurk effect for Dutch and Cantonese listeners. *Proceedings, 4th European Conference on Speech Communication & Technology*, 1699-1702.

de Gelder, B., & Vroomen, J. (1998). Impaired speech perception in poor readers: evidence from hearing and speech reading. *Brain & Language,* **64**, 269-281.

Desjardins. R.N. and Werker, J.F. (2004) Is the integration of heard and seen speech mandatory for infants? *Dev. Psychobiology*, **45**, 187-203.

Desjardins, R.N., Rogers, J., and Werker, J.F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *J. Exp. Child Psych,* **66**, 85-110.

Dodd, B., & Campbell, R. (1987). *Hearing by eye: The psychology of lip-reading*. London: Lawrence Erlbaum Asoociates.

Dodd, B., McIntosh, B., Erdener, D., & Burnham, D. (2008). Perception of the auditory-visual illusion in speech perception by children with phonological disorder. *Clinical Phonetics & Linguistics*, **22**, 69-82.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody Picture Vocabulary Test-Third Edition* Circle Pines, MN: American Guidance Service.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behav. Research Methods, Instruments and Computers,* **35**, 116-124.

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.135-143). New York: Springer-Verlag.

Gagné, J.P., Masterson, V., Munhall, K.G., and Bilida, N., & Querengesser, C. (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *J. Academy of Rehab.Audiology*, **27**, 135-158.

Grassegger, H. (1995). McGurk effect in German and Hungarian listeners. *Proc.International Congress of Phonetic Sciences, Stockholm, Vol. 4* (pp. 210-213). Stockholm: Congress organizers at KTH and Stockholm University.

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A Measure of Executive Function in Preschoolers. *Dev. Neuropsychology*, **20**(3), 573-591.

Kilminster, M. G. E., & Laird, E. M. (1978). Articulation development in children aged three to nine years. Australian *J. Human Communication Disorders*, **6**(1), 23-30.

Kuhl, P.K., Tsuzaki, M., Tohkura, Y., & Meltzoff, A.N. (1994). Human processing of auditory-visual information in speech perception: Potential for multimodal human-machine interfaces. In the Acoustical Society of Japan (Ed.), *Proc.International Conference of Spoken Language Processing* (pp. 539-542). Tokyo: Acoustical Society of Japan.

MacDonald, J., & McGurk, H. (1978). Visual influence on speech perception processing. *Perc.& Psychophys*, **24**, 253-257.

Massaro, D.W. (1984). Children's perception of visual and auditory speech. *Child Development*, **55**, 1777-1788.

Massaro, D.W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D.W., Thompson, L.A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *J. Exp.Child Psych.,* **41**, 93-113.

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

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

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Dev. Sci.,* **6**, 136-142.

Sekiyama, K. & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Dev. Sci.,*, **11**, 306-320.

Sekiyama, K. (1994). Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *J. Acoustical Society of Japan (E)*, **15**, 143-158.

Sekiyama, K., Braida, L.D., Nishino, K., Hayashi, M., & Tsuyo, M. (1995). The McGurk effect in Japanese and American perceivers. *Proc. 13th International Congress of Phonetic Sciences*, *Vol. 3* (pp. 214-217). Stockholm: Congress organizers at KTH and Stockholm University.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, **47**, 277-287.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *JASA,* **90**, 1797-1805.

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

Siva, N, Stevens, E.B., Kuhl, P.K, & Meltzoff, A.N. (1995). A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions. *JASA,* **98**, 2983.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature,* **388**, 381-382.

Stork, D.G., & Hennecke, M.E. (1996). *Speechreading by humans and machines: Models, systems, and applications.* New York: Springer.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *JASA,* **26**, 212-215.

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

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. & Development*, **7**(1), 49-63.

Wilkinson, G. S. (1993). WRAT-3: *Wide Range Achievement Test Administration Manual*. Wilmington, DE.