Auditory-Visual Speech Perception in School and Preschool Children

V. Doğu Erdener & Denis K. Burnham

MARCS Auditory Laboratories

University of Western Sydney, Australia

http://marcs.uws.edu.au ; d.erdener@uws.edu.au

Abstract

The development of auditory-visual speech perception was investigated in pre-school children, school children and adults. Results show a link between auditory-visual speech perception and language-specific speech perception in school children. In addition it was found that speechreading ability in early childhood was predicted by cognitive abilities. Additionally, adults' and preschool children's level of visual speech influence was predictable from auditory-only native language perception. Results are discussed in relation to language-specific developmental challenges that might necessitate support information in the form of visual speech information.

1. Introduction

Speech perception is not only an auditory event. When they are available, perceivers use various other sources of non-auditory information when processing speech, such as tactile (Plant, Gnosspelius, & Levitt, 2000), orthographic (Erdener & Burnham, 2005), and visual. This study investigated visual speech perception. When the auditory signal is degraded, visual speech information (face and lip movements) enhances the perception of speech signal (Sumby & Pollack, 1954). Even in clear listening conditions speech perception is influenced by visual speech information as evidenced by the *McGurk effect* (McGurk & MacDonald, 1976), in which auditory [ba] dubbed onto visual [ga] is perceived as "da" or "tha" (Figure 1). Thus the McGurk effect can be used as an index of visual speech influence in auditory-visual speech perception research.
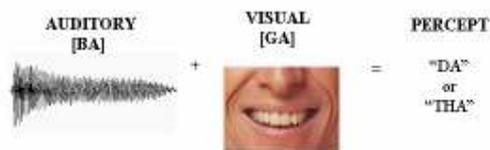

Figure 1: The McGurk effect.

One of the aims of auditory-visual speech research is to understand how auditory and visual (AV) speech information are integrated. An approach to investigating this how question is to find out when AV speech perception occurs; i.e., use developmental data (Bernstein, Burnham, & Schwartz, 2002). This can be obtained via two methods: The *Ontogenetic* method which allows for the testing of individuals with same language background at different ages and the *differential* method in which individuals at the same ages but with different language experience are tested. A series of differential studies have shown that native Japanese speakers are less susceptible to the McGurk effect than native English speakers (Sekiyama & Tohkura, 1993), and Japanese speakers are more influenced by visual speech than Mandarin speakers (Sekiyama, 1997). In addition to cross-language differences,

age-related differences in auditory-visual speech perception have been found in a number of ontogenetic studies , indicating that adults are more prone to visual speech influence than children (Massaro, Thompson, Barron, & Laren, 1986; McGurk & MacDonald, 1976). Combining the differential and ontogenetic methods, Sekiyama and Burnham (2004) tested Japanese- and English-speaking 6-, 8-, 11-year-olds and adults using the McGurk paradigm in three experimental conditions (auditory-only [AO], auditory-visual [AV], and visual-only [VO]). They found that for English-speakers there was minimal use of visual speech information at 6 years, but a significant increase from age 6 to 8, which remained stable at 11 years through to adulthood. The use of visual speech information by the Japanese-speaking 6-year-olds was equivalent to that for the English-speaking 6-year-olds; however, the influence of visual speech remained at this level across all Japanese age groups. These results pose the question of what causes the increase in visual speech influence in English-speaking children aged between 6 and 8 years. Sekiyama and Burnham (2004) offered several reasons for differences between English and Japanese speakers. First, in Japanese and various cultures, looking at the talker's face is considered inappropriate. Second, in Japanese there are fewer visually identifiable speech elements (e.g. lack of labiodentals) and less extensive mouth movements. In addition, the use of two pitch accents to signify meaning differences in Japanese may also mean less need for visual speech information. However, these explanations do not address the reason for the increase of the influence of visual speech in English speakers aged 6 to 8 years. The onset of schooling and reading acquisition occur between these ages, warranting an investigation of the role of new linguistic skills in auditory-visual speech perception. Two factors seem relevant. First, the level of reading skill has been shown to be related to children's language-specific speech perception (LSSP) (Burnham, 2003). LSSP is measured by the extent to which perception of native speech affects the perception of non-native speech. Second, articulatory abilities have been shown to be related to auditory-visual speech perception with children (Desjardins, Rogers, & Werker, 1997).

The following three experiments the relation of these factors to the development of AV speech perception. In Experiments 1A

and 1B four groups of English-speaking school children and a group of English-speaking adults were given tests of reading, articulation, and LSSP,. In Experiment 2 preschool children aged 3 and 4 years tested on LSSP, executive functions and receptive vocabulary knowledge again the relationship with AV speech perception was determined.

2. Experiment 1A: school children

2.1. Method

2.1.1. Participants

Ninety-six monolingual Australian English-speaking children in four age groups were tested: 5-year-olds (n=24, M_{age}=5.52), 6-year-olds (*n*=24, M_{age}=6.80), 7-year-olds (*n*=24, M_{age}=7.59), and 8-year-olds (*n*=24, M_{age}=8.56).

2.1.2. Stimuli & dependent variables

Auditory-visual speech perception test. The stimuli consisted of /ba/, /da/, and /ga/ syllables produced by 2 English and 2 Japanese speakers. The utterances were videotaped and edited to create auditory-only (AO), visual-only (VO), and auditory-visual (AV) stimuli. There were equal numbers of congruent (AV+) and incongruent (AV-) stimuli. Three kinds of incongruent AV stimuli were created: Aud-/ba/ + Vis-/ga/; Aud-/da/ + Vis-/ba/; Aud-/ga/ + Vis-/ba/). In total there were 48 stimuli: 12 AO (3 consonants x 4 talkers), 12 VO (3 consonants x 4 talkers), and 24 AV (3 consonants x 4 talkers x 2 congruence types) stimuli. AO and AV stimuli were also presented with wide-band noise at a signal-to-noise ratio of +4dB. *The dependent variable* was the number of auditory-correct responses. From these a visual speech index (VSI) score was computed by subtracting the auditory-correct responses to the AV- items from auditory-correct responses to the AV+ items. This represents the extent to which there is visual influence in the AV+ items over the AO items, plus that in the AV-items compared with AO items.

Language specific speech perception stimuli. LSSP stimuli consisted of three syllabic items: the voiced bilabial stop [ba], the voiceless bilabial stop [pha], and the voiceless unaspirated stop [pa], spoken by a female native Thai speaker. Three exemplars were used for each syllable and two sets of 36 speech contrasts (18 native, 18 non-native in each version) were created. The native (English) contrast was [pa] vs. [pha] (perceived as /ba/ and /pa/ by English language speakers), and the non-native contrast was [ba] vs. [pa] (both perceived as /ba/ by English language speakers). *The dependent variable* was the difference between the discrimination index (DI) scores for native (N-DI) vs. non-native (NN-DI) contrasts. The DI was calculated as the difference between the number of 'same' responses on (AA) trials (hits) and the number of 'same' responses on different (AB) trials (misses) divided by the total number of trials.

Reading test. The reading subtest of Wide Range Activities Test (WRAT-3) (Wilkinson, 1993) was used. It consists of 15 letters and 42 words presented in order of increasing orthographic complexity (e.g. *in* to *terpsichorean*). *The dependent variable* was the percent correct words read aloud.

Articulation test. The Queensland Articulation Test (QAT) (Kilminster & Laird, 1978) was used. QAT is a 64-item picture-naming test based on the consonants of Australian English presented in initial, medial, and final positions, mindful of phonotactic restrictions (e.g. /ŋ/ is tested in medial

and final (e.g. swi**ng**ing, swi**ng**) but not initial position). The dependent variable was percent correct responses.

2.1.3. Procedure

Children were tested individually at schools. Order of tasks and conditions was counterbalanced. *Auditory-visual* stimuli were presented with a program written in visual basic. Conditions (AO, VO, & AV) were blocked and block order was counterbalanced, with random presentation of stimuli within blocks. The auditory component of stimuli was set at 65 dB, and each block was also repeated with band noise. There were 96 trials in total. Using an identification task, participants were asked to indicate their responses by pressing one of the buttons labelled 'BA', 'DA', 'GA' on a gamepad. Five-year-olds indicated their responses orally before pressing a key. In the *LSSP* test participants were presented with 36 speech contrasts using an AX discrimination task. Stimuli were presented on a computer via DMDX (Forster & Forster, 2003), and native and non-native trials were blocked separately. Responses were collected via a keyboard with shift keys labelled "Same" and "Different". The keys were also labelled with two green squares and a red circle and a green square to make the task easier for 5-year-olds. *Reading* stimuli were presented using the standard cards of the test publisher. The task was to read aloud each letter and word when prompted. The test ended after 5 consecutive errors. For *articulation*, the task was to name photo objects randomly presented on a computer screen. To obtain responses to items that could not be photographed, (e.g. *nothing)* the experimenter prompted children to say the target words.

2.2. Results

2.2.1. Results: analyses of variance

Auditory-visual speech perception test. VSI scores were subjected to a 4 x (2 x 2) (age x noise/clear x stimulus language) ANOVA with repeated measures on the last two factors. There was no significant group difference on VSI scores [F (1, 92) = 1.103, p> .05]. However, the analyses revealed an age-related cubic trend in the clear VSI scores [F (1, 92) = 5.946, p < .02 (Figure 2). In addition, there was an effect of stimulus language: VSI scores for non-native were greater than for native stimuli [F (1, 92) = 7.342, p< .01], and an effect of background noise [F (1, 92) = 62.700, p<.001], with higher VSI scores in noise. Results of a 4 x 2 (age x stimulus language) ANOVA showed an increase over age in the VO scores [F (1, 92) = 8.225, p<.01], and higher VSIs for native stimuli [F (1, 92) = 5.871, p<.05].
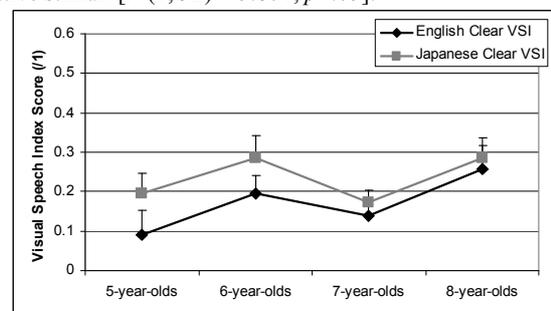

Figure 2: Clear VSI scores. Error bars show standard errors.

LSSP test. A 4 x 2 (age x native/non-native contrast) ANOVA of N-DI and NN-DI scores revealed no age-related differences

for N-DI and N-NN DI scores, but did for NN-DI scores [F (1,92)=6.086, $p<.05$]. The N-DI scores were greater than NN-DI scores [F (1, 92) =148.231, $p<.001$] (Figure 3).
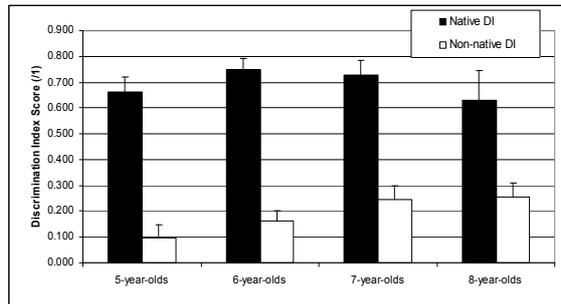

*Figure 3.*The N-DI and NN-DI scores. Error bars show standard errors.

Reading & articulation tests. ANOVA results showed that reading scores increased over age [F (1, 92) = 295.93, $p< .0001$], the largest between 5 and 6 years [$F_{Quadratic}$ (1, 92) = 5.028, $p< .05$]. The ANOVA of articulation scores also revealed an increase with age [F (1, 92) = 93.328, $p< .0001$], levelling out at 7 [$F_{Quadratic}$ (1, 92) = 13.573, $p< .001$].

2.2.2. Results: regression analyses

A sequential multiple regression analysis was run with VSI scores as the criterion and six predictors in order of age, and scores in AO, VO (speechreading), N-NN DI, articulation, and reading. Correlation coefficients are given in Table 1.

Table 1. Correlation coefficients in Experiment 1A.

	VSI	Age	AO	VO	N-NN	Artic.	Read.
VSI	-						
Age	.14	-					
AO	.07	.43**	-				
VO	.33**	.32**	.26**	-			
N-NN DI	.21*	-.15	-.00	.05	-		
Articulation	.07	.69**	.30**	.25**	-.15	-	
Reading	.13	.86**	.46**	.35**	-.06	.68**	-

* Sig. at α=.05, **Sig at α= .01

Of the six predictors in the equation, VO and N-NN DI scores reliably predicted VSI scores: adding VO (speechreading) to the equation in the third step ($R=.34$, $R^2=.11$, F (1, 90) =9.50, $p<.01$), and N-NN DI in the fourth steps ($R=.40$, $R^2=.16$, F (1, 89) =4.60, $p<.05$) reliably increased R^2 (Table 2).

Table 2. Multiple regression of age, AO,VO, N-NN DI, articulation and reading scores as predictors of VSI.

Step	Variables	B at Step	β at Step	R² at Step	Final B (at Step 6)	Final β (at Step 6)
1	**Age**	.020	.140	.020	.026	.182
2	**AO**	.014	.012	.000	-.050	-.042
3	**VO**	.314	.326	**.094****	.298	**.310****
4	**N-NN DI**	.089	.213	**.044***	.089	**.215***
5	**Articulation**	-.184	-.058	.002	-.147	-.046
6	**Reading**	-.054	-.068	.001	-.054	-.068

* Sig. at α=.05, **Sig. at α= .01

A second regression analysis was performed in which N-DI and NN-DI scores were entered as separate predictors. Results showed that the addition of VO (speechreading) to the equation in the third ($R=.34$, $R^2=.11$, F (1, 90) =9.50, $p<.01$), and the N-DI in the fourth steps ($R=.39$, $R^2=.15$, F (1, 89) =3.725, $p<.05$) reliably predict VSI scores.

2.3. Discussion

Except for N-NN DI scores all variables showed an age-related increase. In LSSP test, native speech perception was better than for non-native speech perception in all four age groups. The analysis of VSI scores in the clear condition showed that visual speech influence increased over age as found earlier (Massaro et al., 1986), but with a cubic trend: an increase from age 5 to 6, then a drop at 7 and an increase again at 8. The exact reason for this is not clear, but some possibilities are discussed below. In addition, speechreading ability (VO) also increased over age. Analyses show that LSSP and speechreading predict the influence of visual speech between 5 and 8 years. *Why should greater relative attention to native than non-native speech contrasts predict visual influence in speech perception?* Burnham (2003) found that the onset of reading instruction is related to LSSP in English-speaking children. Reading is a task with a high cognitive demand, especially in English due to the complexity of its grapheme-phoneme correspondences, which makes the English reading acquisition a challenging task (Öney & Goldman, 1984). Given this, English-speaking children may seek all *extra* information available, including visual speech input to help them develop stable phoneme-to-grapheme linkages. Moreover, it appears that children who are good at attending to native speech contrasts and ignoring non-native speech ones and who have better reading ability are just those children who are good at using auditory-visual speech information. The temporary decline in visual speech influence at around 7 years of age may also be explained by the process of reading acquisition. At age 7 reading may start to be an automatic skill and so visual speech information is not needed as back-up resource. As the data show, the 8-year-olds' performance jumps back to that of 6-year-olds' while their reading scores increase substantially. These results show that visual speech input might help perceivers manage challenges of language development, e.g. reading. In addition, speechreading ability seems to develop as a separate construct that is related to articulation and reading. The correlation between articulation and speechreading can be explained by a two-way feedback mechanism in which visual-only speech information helps children articulate speech sounds accurately and the ability to articulate speech sounds correctly leads to better speechreading (Desjardins et al., 1997). It is also possible that the correlation between reading and speechreading might partly be explained by the *way* reading is taught at schools. Teachers usually employ a specific speech style in classrooms called *teacherese,* a distinct hyperarticulated speech style (Håkansson, 1987). Hyperarticulated speech is visually robust with exaggerated articulatory movements, and improve visual speech detection (Lees & Burnham, 2005). In summary, the results of Experiment 1A show that LSSP predicts the degree of auditory-visual speech integration in school children. Experiment 1B investigated whether these findings held for adults.

3. Experiment 1B: Adults

3.1. Method

3.1.1. Participants, stimuli & procedure

Forty-eight native speakers of English (M_{age}= 21.77, SD=6.16) were tested. They were first year psychology students at the University of Western Sydney, Australia. Stimuli, apparatus, dependent variables and procedure were identical to Experiment 1A.

3.2. Results

Auditory-visual speech perception test. A 2 x 2 (stimulus language x background noise) ANOVA revealed no stimulus language effect [$F_{(1, 46)}$ =0.61, $p>$ .05], but an effect of background noise [$F_{(1, 44)}$ =42.06, $p<$. 0001] (Figure 4). A *t*-test analysis of VO data revealed a better performance for non-native than for native stimuli [$t_{(47)}$ =-3.066]. *T*-test analysis of AO scores showed an advantage of native stimuli [$F_{(1, 46)}$ =11.32, $p<$ .005] and background noise [$F_{(1, 46)}$ =39.04, $p<$ .0001].

Speech Perception Test. A *t*-test analysis revealed that N-DI scores were greater than NN-DI scores, $t_{(47)}$ =7.653, $p<$.0001.

Regression Analyses .A sequential multiple regression analysis was performed with VSI scores as the criterion and five predictors in the order of AO, VO, N-NN DI, articulation, and reading. The analysis revealed that only AO scores significantly predict VSI scores and improve R^2 reliably, R=-.394, R^2=.155, $F_{(1, 46)}$ =8.457, $p<$.01.

3.3. Discussion

Experiment 1B revealed no predictive link between LSSP and auditory-visual speech perception. It may be that such a link is active during linguistically challenging periods of language development and becomes redundant when not needed. In other words auditory-visual speech perception may support the process by which speech categories become language-specific and phonologically-based, and this in turn prepares perceivers for the acquisition of cognitively high demand language-specific skills (e.g. reading). In adulthood when reading is an automatic skill, such support may not be required. The next question addressed was whether the link between LSSP and auditory-visual speech perception occurs during preschool years and the use of auditory-visual speech information is related to cognitive skills and an age-specific language-specific challenge, vocabulary acquisition.

4. Experiment 2: preschool children

Research suggests that speech perception organisation during infancy prepares for later language development, e.g. vocabulary development (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005). The emergence of representational thought occurs simultaneously with other facets of cognitive and language development, such as object representation (Lifter & Bloom, 1989) and vocabulary spurt (Nazzi & Bertoncini, 2003). No study has investigated the link between cognitive abilities and auditory-visual speech perception both of which require multimodal processing. The current study investigated whether auditory-visual speech perception is linked to LSSP, cognitive ability and receptive vocabulary knowledge in children aged 3 and 4 years. It was hypothesised that auditory-visual speech perception ability should be predicted from LSSP, vocabulary knowledge and cognitive abilities if these are developmentally related; plus the two groups of children should differ on these measures due to maturation and amount of experience.

4.1. Method

4.1.1. Participants

Forty-eight Australian English monolingual 3- (n=24, M_{age}=3.083, SD=0.168) and 4-year-old (n=24, M_{age}=4.207, SD=0.163) preschool children were tested.

4.1.2. Stimuli, dependent variables & procedure

Auditory-visual speech perception. There were 60 (12 AO, 12 VO & 36 AV) speech contrast stimuli. An AX discrimination task was used. To ensure that children understood the 'same' and 'different' concepts a practice session was run using pairs of pictures. In each trial perceivers were presented with a contrast of two auditory-visual speech stimuli and asked whether they were same or different. The first items were produced by a male speaker and the second ones by a female. Children were told that the man was teaching the lady some words and asked to indicate "if what the lady said was correct". A discrimination task visual speech index score (VSI-AX) was calculated using only 'different' AV incongruent speech contrasts, which differed either on both auditory and visual components or on visual component only. The 'same' trials were not included in VSI-AX calculation as any response choice to these items would not have revealed on what component a given response was based, e.g., if perceivers responded 'same' to the auditory-visual speech contrast Aud-[ba]+Vis-[ga] vs. Aud-[ga]+Vis-[ga], then this trial was given a VSI-AX of '1'. The resultant VSI-AX score was converted to a proportion score.

Language specific speech perception. The dependent variable for LSSP (N-NN DI) was N-DI *minus* NN-DI scores. The stimuli and procedure were identical to the first two experiments except that every correct response was rewarded with a 5-second cartoon clip to keep the task enjoyable.

Executive Function Test. Two components of executive function was measured using the Flexible Item Selection Test (FIST) (Jacques & Zelazo, 2001): rule abstraction and cognitive flexibility. In each trial three pictures which related to one another in one of the three dimensions were presented: colour, size, and shape. One of the items always matched the other two dimensions. The task was to find the first match (e.g. blue boat & yellow boat-shape dimension) which measured rule abstraction, and then the second match (blue boat and blue shoe-colour dimension), which measured cognitive flexibility. The test reveals a FIST score based on rule abstraction and cognitive flexibility scores.

Receptive Vocabulary Test. Vocabulary knowledge was measured with Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1997). The test consists of blocks of picture items which children are asked to identify out of four options in each trial e.g. the child is shown four pictures of a frog, fish, cat, and a dog and asked to point at the picture of a frog. The test progresses until child makes 8 errors in a block. The test yields a standard PPVT score.

4.2. Results

4.2.1. Auditory-visual speech perception test

A *t*-test on the VSI-DX scores revealed a significant difference between 3-, and 4-year-olds, t (46) =-3.927, $p<$ .000, (Figure 4). A *t*-test analysis of the AO data showed that the 4-year-olds (M=0.79, SD=0.15) performed better than 3-year-olds (M=0.59, SD=0.21) [t (46) =-4.171, $p<$.000], however the they did not differ in VO condition [t (46) =-1.860, $p>$.05].
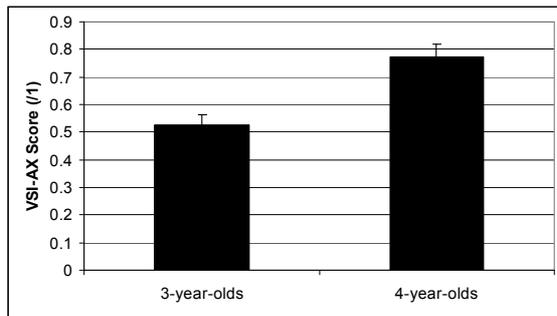

Figure 4. Mean VSI scores. Error bars show standard error.

4.2.2. LSSP test

A 2 x 2 (age x native/non-native) ANOVA of the N-DI and NN-DI scores revealed a significant difference between the groups [F (1, 46) = 7.709, $p<$.01], but no interaction of age and native/non-native scores [F (1, 46) = 1.095, $p>$.05] (Figure 5).
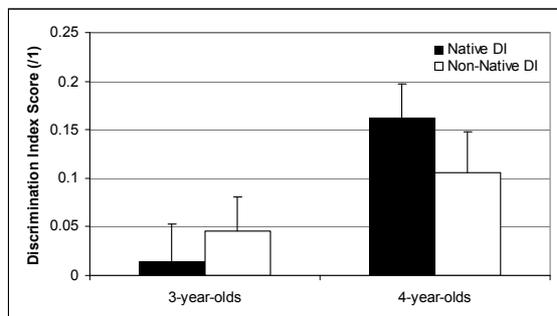

Figure 5. N-DI and NN-DI scores. Error bars show standard errors.

4.2.3. Executive function & receptive vocabulary tests

A 2 x 2 (age x rule abstraction / cognitive flexibility) ANOVA of the FIST data showed a significant difference between the groups [F (1, 46) = 10.078, $p<$.001] and an interaction with rule abstraction/ cognitive flexibility scores [F (1, 46) = 131.324, $p<$.0001]. The *t*-test analysis of PPVT scores revealed a significant difference between 3- (M=67.6, SD=13.0) and 4-year-olds (M=43.3, SD=14.3), t (46) =-6.145, $p<$.0001.

4.2.4. Regression analyses

A sequential multiple regression was performed with VSI-AX scores as the criterion and four predictors in the order of AO, VO, N-NN DI, FIST and PPVT scores. Overall correlation coefficients are presented in Table 3. The introduction of AO (R=.537, R^2=.288, F (1, 46) =18.629, $p<$.000) in the first step and the introduction of FIST (R=.646, R^2=.094, F (1, 46) =6.957, $p<$.05) increased the R^2 reliably, whereas other variables did not increase the R^2 (Table 4).

Table 3. Correlation coefficients in Experiment 2.

	VSI-AX	AO	VO	N-NN	FIST.	PPVT
VSI-AX	-					
AO	.54**	-				
VO	.21	.37**	-			
N-NN DI	.30*	.21	.10	-		
FIST	.51**	.34**	.35**	.41**	-	
PPVT	.45**	.56**	.27*	.41**	.54**	-

* Sig. at α=.05, **Sig at α= .01

Table 4. Multiple regression of AO, VO, N-NN DI, FIST and PPVT as predictors of VSI-AX.

Step	Variables	B at Step	β at Step	R² at Step	Final B (at Step 5)	Final β (at Step 5)
1	AO	.624	.537	**.288****	.491	**.422****
2	VO	.021	.013	.000	-.134	-.082
3	N-NN DI	.158	.192	.035	.052	.063
4	FIST	.012	.364	**.094***	.012	**.359***
5	PPVT	.000	.016	.000	.000	.016

* Sig. at α=.05, **Sig at α= .01

4.3. Discussion

The two groups differed on all skills measured but speechreading ability. Results showed that cognitive abilities and native auditory speech ability predicted amount of visual speech influence. In addition, auditory-visual speech perception was correlated highly with all other variables, except with speechreading which was also found to be predicted from basic cognitive abilities, indicating that not only auditory but also auditory-visual aspects of speech are relevant to cognitive functions. This suggests that the way auditory-visual speech code is processed *perceptually* might be similar to the processing of two or more dimensions of a event, similar to the concept of conservation in preoperational children (Piaget, 1986). In other words, the ability to process multimodal speech information might be related to the cognitive ability to process multidimensional events, and these abilities may co-develop and interact as a result of both maturation and language experience.

5. Conclusion

Experiments here showed that visual speech influence assists perceivers when they face language-specific challenges such as reading acquisition and exposure to new speech styles. In addition we found a link between auditory-visual speech perception and LSSP for school children but not for adults and preschool children. Despite this, the preschool children's and adults' auditory-visual speech integration were predictable from the AO scores, indicating a link between native speech processing and auditory-visual speech perception. In addition, we found that preschoolers' level of visual speech influence was predictable from their basic

cognitive abilities, suggesting a link between auditory-visual speech perception as a multimodal event and multimodal processing of cognitive events. One possible target population for the investigation of the role of auditory-visual speech perception as a back-up source in linguistically challenging situations could be adults who are learning a second language. Another target population to study the link between visual speech information and LSSP could be children with a phonological speech disorder.

6. Acknowledgements

7. References

Bernstein, L. E., Burnham, D., & Schwartz, J.-L. (2002). Special session: Issues in audiovisual spoken language processing (when, where, and how?). In J.H.L. Hansen & B. Pellom (Eds.), *Proceedings of International Conference on Spoken Language Processing 2002,* (Vol.3, pp. 1445-1448), Denver, Colorado.

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

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

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

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in non-native speech production. *Language Learning, 55*(2), 191-228.

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

Håkansson, G. (1987). *Teacher talk. How teachers modify their speech when addressing learners of Swedish as a second language*. Lund, Sweden: Lund University Press.

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): a measure of executive function in preschoolers. *Developmental Neuropsychology, 20*(3), 573-591.

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

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the "critical period". *Language Learning and Development, 1*(3 & 4), 237-264.

Lees, N., & Burnham, D. (2005). Facilitating speech detection in style!: the effect of visual speaking style on the detection of speech in noise. In E. Vatikiotis-Bateson, D. Burnham & S. Fels (Eds.), *Proceedings of Auditory-Visual Speech Processing International Conference* (pp. 23-28). Vancouver Island, BC, Canada.

Lifter, K., & Bloom, L. (1989). Object knowledge and the emergence of language. *Infant Behavior and Development, 12*(4), 395-423.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contribution to speech perception. *Journal of Experimental Child Psychology, 41*, 93-113.

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? *Developmental Science, 6*(2), 136-142.

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.

Piaget, J. (1986). T*he construction of reality in the child* (M. Cook, Trans. 3 ed.). New York, NY: Ballantine Books.

Plant, G., Gnosspelius, J., & Levitt, H. (2000). the use of tactile supplements in lipreading swedish and english: a single-subject study. *Journal of Speech and Hearing Research, 43*, 172-183.

Sekiyama, K. (1997). 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. (2004). Issues in the development of auditory-visual speech perception: adults, infants, and children. In S. H. Kim & D. H. Youn (Eds.), *Proceedings of International Conference on Spoken Language Processing 2004,* (Vol.2, pp.1137-1140), Jeju Island, South Korea.

Sekiyama, K., & Tohkura, Y. I. (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.

Wilkinson, G. S. (1993). *WRAT3: Wide Range Achievement Test Administration Manual*. Wilmington, DE: Psychological Assessment Resources Ltd.