The Effect of Accurate Speech Production Experience on the Development of Auditory-Visual Speech Perception in Children

Doğu Erdener[1], Denis Burnham[1], Beth McIntosh[2] & Barbara Dodd[2]

[1]MARCS Auditory Laboratories-University of Western Sydney
Locked Bag 1797, Penrith South DC NSW 1797 Australia

[2]Perinatal Research Centre, Speech & Language Group
Royal Brisbane & Women's Hospital, Herston QLD 4029 Australia

http://marcs.uws.edu.au / http://www.som.uq.edu.au/prc/

Abstract. *This study investigated the development of auditory-visual speech perception (AVSP) in a group of children with and without inconsistent phonological disorder, a speech production disorder. Children were given tests of AVSP, language specific speech perception (LSSP), and executive functions. Results provide evidence of a link between speech production and auditory-visual speech perception, indicating that failure in speech perception as a result of phonological misrepresentation might partly be due to in efficient speechreading and auditory-visual speech integration abilities.*

1. Introduction

Speech processing is an auditory-visual event (Erber, 1969). A classic demonstration of the role of visual information in speech is the McGurk effect (McGurk & MacDonald, 1976). In McGurk effect, auditory syllable [ba] dubbed on visual syllable [ga] is perceived as "da" or "tha". The McGurk effect has been used as an index of visual speech influence as it is also used as such here.

An important question in auditory-visual speech perception (AVSP) research is to understand the linguistic level at which the auditory and visual speech components are integrated: amodal / phonetic (Burnham & Dodd, 2004; Rosenblum & Saldaña, 1996) or phonological (Massaro, Cohen, & Smeele, 1995). To answer this question developmental data are needed (Bernstein, Burnham, & Schwartz, 2002). As a part of this quest, Sekiyama and Burnham (2004) tested Japanese- and English-speaking 6-, 8-, 11-year-olds and adults using the McGurk paradigm. They found that while the visual speech influence increases for English speakers with age this was not the case for Japanese speakers. One explanation was related to speech production in Japanese: the lack of certain classes of visually identifiable speech elements (e.g. labiodentals) and less extensive mouth movements. One of the potential contributing factors to the development of AVSP is language specific speech perception (LSSP), which can be defined as the relative strength of native speech perception compared with non-native speech perception, and is measured by the difference between native versus non-native speech perception. Recently Burnham (2003) found that LSSP is particularly related to reading abilities in children.

A second potentially important factor in the development of AVSP is the *articulation* ability. An influential theory that focuses the link between speech perception and production is the motor theory of speech perception (Liberman & Mattingly, 1985), which suggests that

speech perception involves the perception of a pattern of articulatory movements. While many studies have investigated the relationship between speech perception and production only two have investigated the link between auditory-visual speech perception and speech production. Desjardins, Rogers and Werker (1997) tested two groups of children in a McGurk task: *Substituters*, who made phoneme substitution errors (e.g. /ɵ/ is substituted by /t/ in "thick"), and *nonsubstituters*. They found that substituters were less influenced by visual speech than nonsubstituters, and that experience with correctly articulating speech sounds was related to the degree of visual speech influence. In addition, it was found that cerebral palsied adults with motor speech problems were less prone to visual speech influence than control perceivers (Siva, Stevens, Kuhl, & Meltzoff, 1995). Together, these two studies show a positive relationship between the experience of correctly articulating speech sounds and degree of visual influence in speech perception.

Recently Erdener and Burnham (2005) investigated the role of LSSP, reading, and articulation in auditory-visual speech perception in school children (5 to 8 years) and adults. Results showed increases in all variables tested with age and most importantly the degree of visual speech influence was predicted by LSSP, but not by age, reading or articulation, the latter seemingly in contrast to previous studies (Desjardins et al., 1997; Siva et al., 1995). However, unlike the other two studies, Erdener and Burnham (2005) tested perceivers who did not have speech disorders. The current study was designed to investigate the link between AVSP and LSSP in both speech disordered and non-disordered children.

The speech disorder studied here is inconsistent phonological speech disorder. In this, the speech production problem does not stem from a motor deficit or anatomic deformity but it is due to the misrepresentation of speech sound categories as a result of a failure in phonological processing (Dodd, 1995). The speech is characterised by indiscriminate use of the members of a phonological category (Fox & Dodd, 2001), e.g. substituting a given fricative target with /f/, /ɵ/, or /v/ inconsistently. Testing children with inconsistent phonological speech disorder provides an opportunity to study the link between speech perception and production from an auditory-visual speech perspective. For this purpose children with (SD) and without (no-SD) inconsistent phonological speech disorder were given AVSP, LSSP tests and were also tested on executive functions to control for and test the role of basic cognitive abilities in AVSP development. Executive functions are sets of high cognitive functions enabling to plan, initiate, and execute goal-directed behaviour (Oates & Greyson, 2004).

In this study the following predictions were advanced: (a) if AVSP LSSP are related (Erdener & Burnham, 2005), then no-SD children should show more visual influence in speech perception and stronger LSSP than SD children; (b) if LSSP and AVSP are related, then there should not be any age-related differences either in LSSP nor AVSP in SD children; (c) If LSSP predicts AVSP in linguistically challenging situations (Erdener & Burnham, 2005) then this may be the case for SD children.

2. Method

2.1. Participants

Thirty-nine SD children (M_{age}= 4.34, *sd*=0.69) and 18 no-SD children with normal speech development (M_{age}= 4.12, *sd*=0.46) as assessed by speech pathologists were recruited. All children had normal or corrected vision. To have comparable groups, 18 of the SD children

(M_{age}= 4.06, *sd*=0.49) were gender and age matched to the no-SD children (M_{age}= 4.12, *sd*=0.49). In addition, the full sample of 39 SD children was divided into three age groups to be able to test age effects (3-year-olds: M_{age}= 3.56& *sd*=0.33; 4-year-olds: M_{age}= 4.34 & *sd*=0.18; 5-year-olds: M_{age}= 5.13& *sd*=0.28) within this group. An analysis of variance (ANOVA) revealed that these three SD age groups were significantly different from each other [$F(1,36)=218.14$, $p<.0001$).

2.2. Stimuli, Dependent Variables & Procedure

Auditory-Visual Speech Perception. There were 60 (12 Auditory-Only [AO], 12 Visual-Only [VO] & 36 Auditory-Visual [AV]) speech contrast stimuli. An AX discrimination task was used. In each trial perceivers were presented with a contrast of two auditory-visual speech stimuli. 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 were asked to say "if what the lady said was correct". A 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 just on visual component, e.g. if perceivers responded 'same' to the auditory-visual speech contrast Aud-[ba]+Vis-[ga] vs. Aud-[ga]+Vis-[ga], then this showed a visual influence. The resultant VSI-AX score (max. 9 & min.0) was converted to a proportion score. The 'same' trials were not included in VSI-AX calculation as any response choice to these items would not have revealed on what component (auditory, visual or both) a response was based.

LSSP Test. The LSSP stimuli consisted of three syllabic items: a voiced bilabial stop [ba], a voiceless bilabial stop [p^ha], and a voiceless unaspirated stop [pa] 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. [p^ha] (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 number of 'different' responses on different trials (hits) minus the number of 'different' responses on same trials (false positives) divided by the total number of trials.

Executive Function Test. Two elements of executive function were measured via 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 three dimensions were presented: colour, size, and shape. One of the items always matched the other two items in one dimension. The task was to find the first match (e.g. blue boat & yellow boat-shape dimension), a measure of rule abstraction, and a second match (blue boat & blue shoe-colour dimension), a measure of cognitive flexibility. A standard FIST score based on separate rule abstraction and cognitive flexibility scores was calculated.

3. Results

3.1. Speech Disordered vs. Non-Speech Disordered Children

AVSP Test. The VSI-AX, AO and VO scores for SD and no-SD children were subjected to *t*-tests. Results showed that two groups did not differ on the VSI-AX [$t(1,34) = .550$, $p>.05$] and AO [$t(1,34) = .642$, $p>.05$] scores, whereas the no-SD group performed better in the VO condition than the SD group [$t(1,34) = 2.254$, $p<.05$] (see Figure 1).

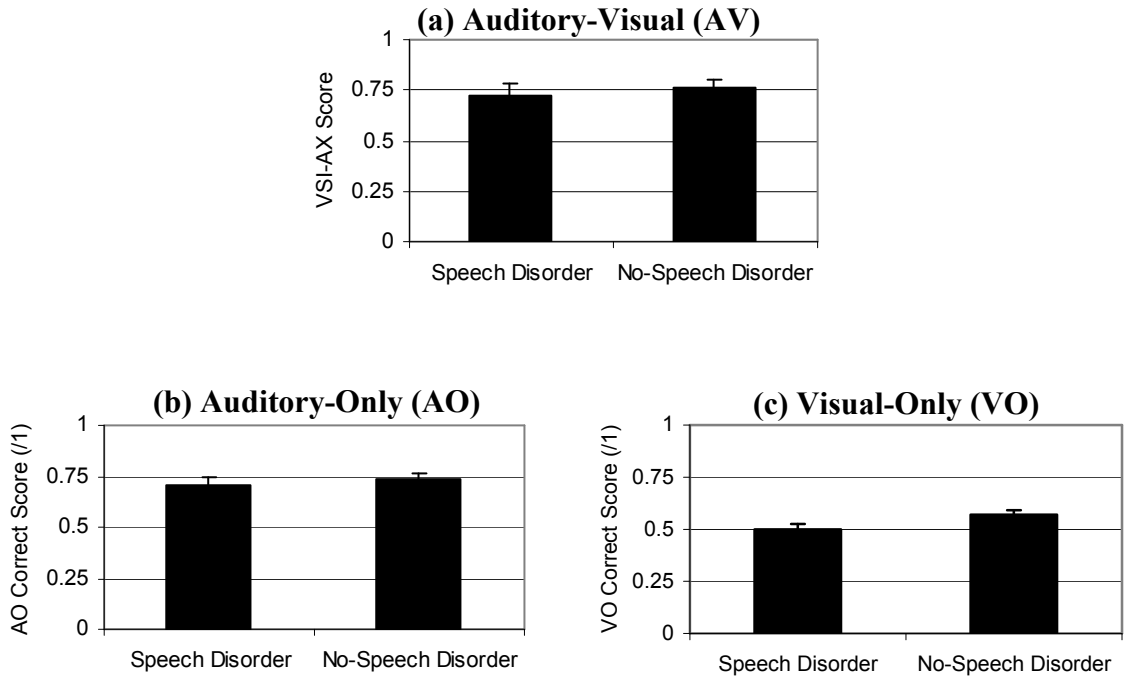

Figure 1. Mean (a)VSI-AX, (b) AO, and (c)VO scores for SD and no-SD children.

LSSP Test. N-DI and NN-DI scores were subjected to a 2 x 2 (Speech Disorder Status x Native/Non-native) ANOVA. Results showed generally better performance for native contrasts than for non-native contrasts [F (1, 34) = 5.391, p<.05], and relatively better native versus non-native speech perception for the no-SD group than with the SD group [F (1, 34) = 13.108, p<.001] (see Figure 2).

Executive Function Test. The FIST scores were subjected to a 2 x 2 (Speech Disorder Status x Rule Abstraction/Cognitive Flexibility) ANOVA. Results showed a significant difference between general rule abstraction and cognitive flexibility scores [F (1, 34) = 73.908, p<.001], but no between-group difference [F (1, 34) = 0.512, p>.05] (see Figure 2).

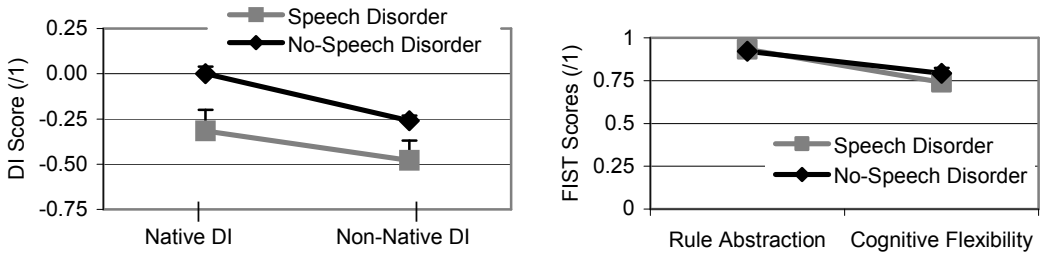

Figure 2. Mean N-DI and NN-DI (left) and FIST subtest scores (right) for SD and no-SD children.

Regression Analyses. A sequential multiple regression was run with VSI-AX scores as dependent variable and six predictors in the following order: age, speech disorder status, AO, VO, N-NN DI, and FIST scores. None of these variables reliably predicted VSI-AX scores. A second regression analysis was performed in which the dependent variable was VO scores and age, speech disorder status, AO, VSI-AX, N-NN DI and FIST scores were entered as independent variables. Results showed that only speech disorder status in step 2 predicted speechreading scores reliably (R=.35, R^2=.11, F (1,34) =4.95, p<.05)(see Table 1).

Table 1. Multiple regression of age, speech disorder (SD) status, AO, VSI-AX, N-NN DI and FIST scores as predictors of VO scores.

Step	Variables	B at Step	β at Step	R^2 at Step	Final B	Final β
1	**Age**	.001	.033	.000	.007	.029
2	**SD Status**	-.074	-.361	.130	-.077*	-.375*
3	**AO**	-.015	-.020	.000	-.016	-.020
4	**VSI-AX**	-.012	-.024	.001	-.039	-.076
5	**N-NN DI**	-.033	-.182	.039	-.032	-.178
6	**FIST**	.001	.018	.000	.001	.018

* Sig. at α=.05, **Sig. at α= .01

3.2. Speech Disordered Children: Age Effects

AVSP Test. The AO, VO and VSI-AX scores of three groups of SD children (see 2.1) were subjected to three sets of 3 x 1 (Age x AO, Age x VO, and Age x VSI-AX) ANOVAs. Results showed that there were no group differences on AO [F (1, 36) =.54, p>.05], VO [F (1, 36) = 1.56, p>.05], or VSI-AX [F (1, 36) = 1.43, p>.05] scores (see Figure 3).

LSSP Test. The DI scores were subjected to a 3 x 2 (Age x Contrast type) ANOVA. Results revealed no age-related differences with respect to DI scores overall [F (1, 36) = 0.01, p>.05], although N- DI scores were higher than NN-DI scores [F (1, 36) = 54.472, p<.05] (see Figure 4).

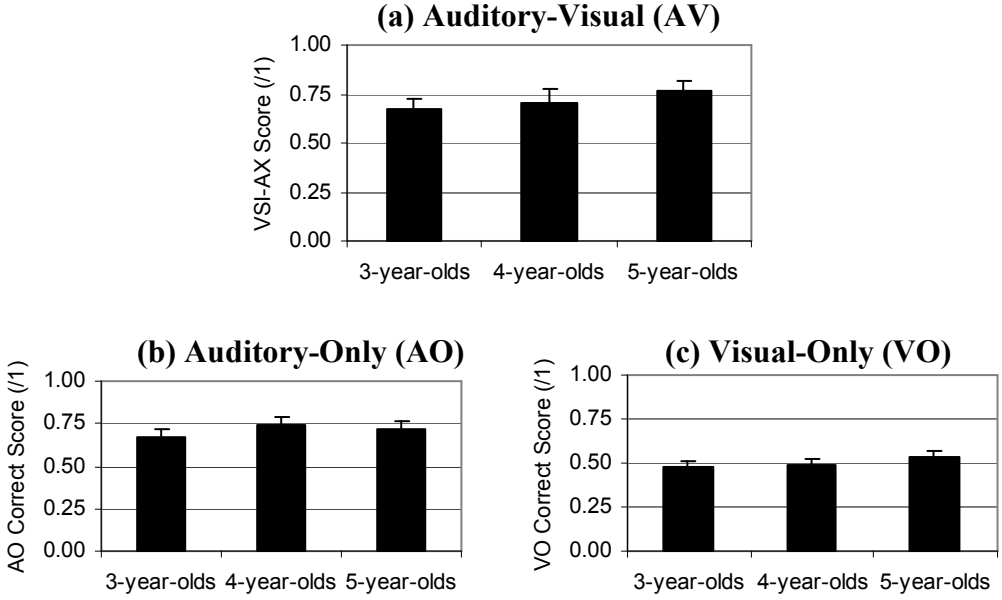

Figure 3. Mean VSI-AX, AO, and VO scores for SD children.

Executive Function Test. The FIST scores were subjected to a 3 x 2 (Age x rule abstraction / cognitive flexibility) ANOVA. Results showed that FIST scores increased with age [F (1, 36) = 52.47, p<.001]. Also, rule abstraction scores were significantly higher than cognitive flexibility scores [F (1, 36) = 78.305, p<.05], interacting with age [F (1, 36) = 14.531, p<.005] (see Figure 4).

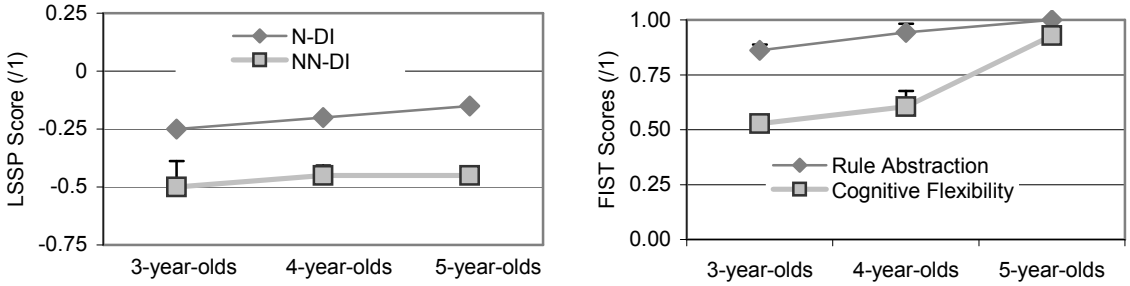

Figure 4. Mean N-DI and NN-DI (left) and FIST subtest scores (right) for SD children.

Regression Analyses. A sequential multiple regression was performed with VSI-AX scores as the dependent variable and six predictors in the order of age, age group, AO, VO, N-NN DI and FIST scores. Results showed that only N-NN DI scores (entered at step 5) significantly predicted VSI-AX scores (R=.34, R^2=.11, F (1,33) =5.10, p<.05) (see Table 3).

Table 3. Multiple regression of age, speech disorder (SD) status, AO, VSI-AX, N-NN DI and FIST scores as predictors of VSI-AX scores.

Step	Variables	B at Step	β at Step	R^2 at Step	Final B	Final β
1	**Age**	.063	.208	.043	.131	.436
2	**Age Groups**	.004	.014	.000	-.070	-.275
3	**AO**	.186	.151	.023	.215	.175
4	**VO**	.211	.101	.009	.281	.135
5	**N-NN DI**	-.362	-.355	.124*	-.360	-.353*
6	**FIST**	.002	.038	.001	.002	.038

* Sig. at α=.05, **Sig. at α= .01

An additional regression analysis, in which the VO scores were entered as the dependent and age, age groups, AO, VSI-AX, N-NN DI and FIST scores as independent variables showed that age groups at step 2 (R=.372, R^2=.13, F (1,37) =5.558, p<.01) and FIST scores at step 6 (R=.578, R^2=.33, F (1,32) =7.665, p<.01) predicted VO scores (Table 4). A further analysis in which FIST elements were entered as separate predictors revealed that cognitive flexibility but not rule abstraction reliably predicted VO scores (R=.578, R^2=.33, F (1,31) =5.445, p<.05).

Table 4. Multiple regression of age, speech disorder (SD) status, AO, VSI-AX, N-NN DI and FIST scores as predictors of VO scores.

Step	Variables	B at Step	β at Step	R^2 at Step	Final B	Final β
1	**Age**	.010	.072	.005*	-.123	-.853*
2	**Age Groups**	.119	.969	.133**	.149	1.218**
3	**AO**	-.054	-.091	.008	-131	-.222
4	**VSI-AX**	.044	.092	.008	.054	.112
5	**N-NN DI**	.042	.086	.006	.023	.048
6	**FIST**	-.013	-.485	.173**	-.013	-.485**

* Sig. at α=.05, **Sig. at α= .01

4. Discussion

Results show that SD and no-SD children are comparable on AO and AV speech measures, but that no-SD children are better speechreaders than SD children. These results provide support for two conclusions. First, it appears veridical experience in speech production (in the no-SD group) results in better speechreading than experience of incorrectly producing speech. Second, they suggest that experience of incorrect articulatory experience (SD group) had no effect on the *integration* of auditory and visual speech components. The latter suggests that AVSP is a separate construct to auditory and visual speech components. This finding also provides an indirect support to previous studies that auditory and visual speech components are processed at an earlier e.g. phonetic (Rosenblum & Saldaña, 1996) than a later stage, e.g. phonological (Massaro et al., 1995), as the SD children in this study are speech disordered precisely because they have a problem with the execution of phonological rules in their speech production.

The SD children showed a greater LSSP effect than no-SD children. This is presumably a reflection of phonological processing differences between the two groups. One model that might explain this difference is the motor theory of speech perception (Liberman & Mattingly, 1985). In light of the motor theory, what may be lacking in SD children is the link between the perception of phonological speech categories and the perception of articulatory movements, which have visual instantiations. When a speaker's face is viewed during speech, the visible articulatory movements also carry significant phonological information from the vocal tract (Yehia, Rubin, & Vatikiotis-Bateson, 1998). In addition to the speechreading (VO) superiority of no-SD over SD children, regression analyses also show that the SD status is a significant predictor of speechreading ability. SD children's data show there were no age-based differences in AVSP, unlike previous findings with no-SD children (Erdener & Burnham, 2005; Sekiyama & Burnham, 2004). The only variable that predicts speechreading in the SD group was cognitive flexibility, the ability to abstract and apply rules to new events. This is also in agreement with recent results we obtained with no-SD children aged three and four (Erdener & Burnham, in preparation). This finding makes good intuitive sense: when speechreading is difficult (as it is for SD children), those children with greater cognitive flexibility will be those who better perceive visual speech just as Erdener and Burnham (2005) found with no-SD children.

In summary this study showed that the ability to *produce* speech sounds correctly is related to the amount of visual speech information available, partly supporting earlier evidence with young children (Desjardins et al., 1997) and cerebral palsy adults (Siva et al., 1995). Furthermore, LSSP predicts the amount of visual speech influence, which in turn might affect the resultant speech production. One of the causes of disordered speech in children may be the lack of a link between the representation of speech sound categories in the form of prototype articulatory gestures (Liberman & Mattingly, 1985) and the perception of speech, which in turn, feeds back to speech production. Further research is required in order to understand the auditory and visual aspects of speech perception and production, from which research and therapy in speech production disorders can benefit.

Acknowledgements

We thank Dr. Sharon Crosbie for her valuable comments and help with the testing process, Mr. Brad McIntosh for writing the code for the McGurk AX discrimination task and the children and their parents without whom this study could not have been conducted.

References

Bernstein, L. E., Burnham, D., and Schwartz, J.-L. Special session: Issues in audiovisual spoken language processing (when, where, and how?). In *Proceedings of ICSLP2002,* Vol. 3, pages 1445-1448, 2002.

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

Burnham, D., & Dodd, B. Auditory-visual speech integration by pre-linguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology,* 44: 209-220, 2004.

Desjardins, R. N., Rogers, J., & Werker, J. F. An Exploration of Why Preschoolers Perform Differently Than Do Adults in Audiovisual Speech Perception Tasks. *Journal of Experimental Child Psychology,* 66: 85-110, 1997.

Dodd, B. *Differential Diagnosis and Treatment of Children with Speech Disorder.* Whurr, 1995.

Erber, N. P. Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research,* 12: 423-425, 1969.

Erdener, V. D., & Burnham, D. K. Development of auditory-visual speech perception in English-speaking children: The role of language-specific factors. In E. Vatikiotis-Bateson, D. Burnham & S. Fels (Eds.), In *Proceedings of Auditory-Visual Speech Processing International Conference,* pages 57-62, 2005.

Fox, A.V. & Dodd, B. Phonologically Disordered German-Speaking Children. *Journal of Speech-Language Pathology,* 10(3): 291-307, 2001.

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

Liberman, A. M., & Mattingly, I. G. The motor theory of speech perception revised. *Cognition,* 21: 1-36, 1985.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. Cross-linguistic comparisons in the integration of visual and auditory speech. *Memory and Cognition,* 23(1): 113-131, 1995.

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

Oates, J., & Greyson, A. *Cognitive and Language Development in Children*, Blackwell Publishing, 2004.

Rosenblum, L. D., & Saldaña, H. M. An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 22(2): 318-331, 1996.

Sekiyama, K., & Burnham, D. Issues in the Development of Auditory-Visual Speech Perception: Adults, Infants, and Children. In *Proceedings of International Conference on Spoken Language Processing 2004,* Vol. 2:1137-1140, 2004.

Siva, N., Stevens, E. B., Kuhl, P. K., & Meltzoff, A. N. A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions, *Journal of the Acoustical Society of America,* 98: 2893, 1995.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. Quantitative association of vocal-tract and facial behavior. *Speech Communication,* 26: 23-43, 1998.