Auditory–visual speech perception in three- and four-year-olds and its relationship to perceptual attunement and receptive vocabulary*

DOĞU ERDENER

Psychology Program, Middle East Technical University, Northern Cyprus Campus, Güzelyurt/Morphou, Northern Cyprus

AND

DENIS BURNHAM

The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

ABSTRACT

Despite the body of research on auditory–visual speech perception in infants and schoolchildren, development in the early childhood period remains relatively uncharted. In this study, English-speaking children between three and four years of age were investigated for: (i) the development of visual speech perception – lip-reading and visual influence in auditory–visual integration; (ii) the development of auditory speech perception and native language perceptual attunement; and (iii) the relationship between these and a language skill relevant at this age, receptive vocabulary. Visual speech perception skills improved even over this relatively short time period. However, regression analyses revealed that vocabulary was predicted by auditory-only speech perception, and native language attunement, but not by visual speech perception ability. The results suggest that, in contrast to infants and schoolchildren, in three- to four-year-olds the relationship between

speech perception and language ability is based on auditory and not visual or auditory–visual speech perception ability. Adding these results to existing findings allows elaboration of a more complete account of the developmental course of auditory–visual speech perception.

INTRODUCTION

Early speech perception ability predicts later language development (Elliot, Hammer & Scholl, 1989; Tsao, Liu & Kuhl, 2004; Vance, Rosen & Coleman, 2009). For example, infants' vowel discrimination at six months predicts their word comprehension and production at two years (Tsao *et al.*, 2004). In addition, the degree of perceptual attunement, the attentional bias towards native and away from non-native speech sounds (Werker & Tees, 1984) at 7 months predicts language development at 14 to 30 months (Kuhl, Conboy, Padden, Nelson & Pruitt, 2005).

These studies concern AUDITORY speech perception and language development; this study explores AUDITORY–VISUAL speech perception and language development. Speech perception is auditory–visual both in adults (e.g. Campbell, Dodd & Burnham, 1998; Sumby & Pollock, 1954) and infants. For instance, infants use visual speech information to FACILITATE auditory input by one month, to MATCH auditory and visual input by three to four months, and to INTEGRATE auditory and visual information by four months (see Burnham, 1998, and Burnham & Sekiyama, 2012, for reviews). The latter, auditory–visual integration, is often measured using the McGurk effect, in which the auditory, /ba/, dubbed onto /ga/ lip movements, is perceived as neither the auditory nor the visual component but as the emergent percept "da" or "tha" (McGurk & MacDonald, 1976). By five months, infants perceive both this emergent percept version of the McGurk effect (Burnham & Dodd, 2004), and other versions, such as the Auditory [ba] + Visual [va] (A[ba]V[va]) as "va" (Rosenblum, Schmuckler & Johnson, 1997) and A[bi]V[vi] as "vi" (Desjardins & Werker, 2004).

So, like its auditory counterpart, auditory–visual speech perception is evident very early. Moreover, like auditory speech perception, auditory–visual speech perception becomes attuned to the native language very early. Dodd and Burnham (1988) showed infants two faces, one reciting a text in their native language (English) and the other in a non-native language (Greek), with only one of the corresponding voices being played to infants over a central loudspeaker. Ten-week-old infants were able to match both the native and the non-native voice to its face, but by 20 weeks matching was language-specific – infants only matched face and voice for the native language.

While auditory–visual speech perception is well established in infancy, it continues to develop into early and later childhood. For example, McGurk

274

and MacDonald (1976) found that visual speech influence increased from preschool (three- to five-year-olds) to school (seven- and eight-year-olds) to adulthood (eighteen- to forty-year-olds), a developmental trend consistently supported by subsequent studies (Hockley & Polka, 1994; Massaro, 1984; Sekiyama & Burnham, 2008). Thus, despite relatively mature auditory–visual speech perception in infancy, the use of visual speech information continues to improve into early childhood and beyond.

The development of the use of visual speech information in childhood is related to two factors: (i) to the structure of particular languages, their phonology, orthography, and phoneme–grapheme relationship; and (ii) more generally, to a second period of perceptual attunement around the onset of reading instruction. With respect to language structure, building on reports showing more limited McGurk effects in Japanese- than English-language adults (Sekiyama & Tokhura, 1991, 1993), Sekiyama and Burnham (2008) showed that visual influence indexed by responses to McGurk stimuli is equivalent for English- and Japanese-language six-year-olds, but then increases from six to eight years in English-language children and is maintained thereafter, whereas for Japanese children there is no six to eight years increase nor any subsequent increase. Thus differences between English and Japanese adults' McGurk effect responses have their origin at the onset of reading, and Sekiyama and Burnham suggest such cross-language differences in auditory–visual speech perception might be due to the interaction of (a) learning to read a script with relatively transparent (Japanese) vs. opaque (English) phoneme-to-grapheme mapping, and (b) the relatively minor functional utility of visual speech information in a language with five vowels and no consonant clusters (Japanese) versus one with twelve to fourteen vowels, consonant clusters, and fricative sounds that are visually but not auditorally distinctive.

Following up this language-specific effect, Erdener and Burnham (2013) investigated the relationship between auditory–visual speech perception and other language skills, particularly those associated with the onset of reading. They built their study on both the Sekiyama and Burnham (2008) Japanese/English study and on research showing that, at six years of age, there is a second period of perceptual attunement related to the onset of reading. In this period, English-language children's perceptual attunement to native over non-native sounds is heightened between four and six years (over and above the attunement established in infancy), and the degree of this heightened perceptual attunement predicts reading and reading-related phonological skills (Burnham, 2003; Burnham, Earnshaw & Clark, 1991; Horlyck, Reid & Burnham, 2012). In their study, Erdener and Burnham (2013) found that the degree of perceptual attunement and visual-only speech perception (lip-reading) reliably predicted auditory–visual speech perception in five- to eight-year-old English-language

275

children, but that, in adults, auditory–visual speech perception was predicted only by auditory speech perception. Erdener and Burnham speculated that there may be a common determinant for the augmented perceptual attunement (Burnham, 2003; Burnham et al., 1991; Horlyck et al., 2012) and the increased visual influence in speech perception (Sekiyama & Burnham, 2008) around the time of reading onset.

In similarly oriented studies, Jerger, Damian, Spence, Tye-Murray, and Abdi (2009) tested four- to fourteen-year-old children on a picture–word naming task with auditory–visual, and auditory-only phonological distractors and found a temporary loss of sensitivity to visual speech around five years of age, which they suggest reflects a reorganization of phonological processing related to literacy instruction, auditory and auditory–visual speech perception, and linguistic skills. Further work by this team (Jerger, Damian, Tye-Murray & Abdi, 2014), again with four- to fourteen-year-olds, showed that the use of visual cues in a task involving visual fill-in for missing auditory phonemes improved with age and was uniquely accounted for by age and vocabulary skills, and that McGurk effect performance was uniquely accounted for by visual-only speech perception (lip-reading).

In summary, auditory–visual speech perception is clearly evident in infancy, and continues to improve over age into early and later childhood. However, around reading onset auditory–visual speech perception development (i) is affected by the particular language background, (ii) is associated with the degree of perceptual attunement (native minus non-native speech perception), and (iii) is associated with vocabulary skills.

This study was conducted to investigate auditory–visual speech perception development in the relatively uncharted age range of around three to four years. Children were presented with auditory-only, visual-only, and auditory–visual speech discrimination tests, a native minus non-native speech perception attunement test, and a measure of receptive vocabulary size (Peabody Picture Vocabulary Test-III; Dunn & Dunn, 1997). This last was included as an age-appropriate measure of language development: given the relationship of auditory–visual speech perception with language skills in studies with older children, receptive vocabulary was chosen as it has been found to be related to phonological skills in the second year of life (Bundgaard-Nielsen, Best, Kroos & Tyler, 2012; Bundgaard-Nielsen, Best & Tyler, 2011a, 2011b), and to speech discrimination in four- to five-year-olds (Vance et al., 2009).

While auditory–visual speech perception research in this age range is limited, there is sufficient evidence from older children that auditory–visual speech perception and language development may be related. So, in this age range, a tentative prediction that receptive vocabulary will be associated with the level of auditory–visual speech perception performance is warranted. There were three hypotheses, each with an auditory and a

276

visual component: (1) four-year-olds would perform better than three-year-olds on (a) auditory (auditory-only speech perception, and native vs. non-native perceptual attunement), and (b) visual (lip-reading and visual speech influence in auditory–visual integration) measures; (2) the two auditory measures would predict vocabulary size; and (3) the two visual speech perception measures would make an ADDITIONAL unique contribution to the prediction of vocabulary size.

METHOD

Participants

Forty-eight children, 24 three-year-olds ($M = 3 \cdot 1$ years, $SD = 0 \cdot 2$ years, 12 female) and 24 four-year-olds ($M = 4 \cdot 2$ years, $SD = 0 \cdot 2$ years, 12 female) were recruited from the BabyLab register at the MARCS Institute for Brain, Behaviour and Development at Western Sydney University. All children were from monolingual Australian English-speaking families and reported normal hearing (including no middle ear infection history) and vision and no history of speech impediments. Parents received $20 (AUD) to reimburse travel costs, and children received a soft toy of their choice and a Young Scientist certificate.

Stimuli, apparatus, and procedure

All children were tested individually. Due to the young age of the children, one parent was also present during testing. Parents were asked not to provide any feedback to their child during the experiment. The order for the three tasks – speech perception, native/non-native speech perception, and vocabulary – was counterbalanced across participants.

Auditory-only, visual-only, and auditory–visual speech perception test. The speech stimuli were videotaped audio–visual utterances of [ba], [da], and [ga] spoken by one adult male and one adult female native English speaker (see Erdener & Burnham, 2013, Sekiyama & Burnham, 2008, for details). These base auditory–visual stimuli were edited to create a total of sixty trials, comprising twelve auditory-only (AO), twelve visual-only (VO), and thirty-six auditory–visual (AV) trials. Each trial consisted of two stimuli presented sequentially. In each of the three conditions, half were same trials (the two stimuli were the same) and half were different trials (two different sounds were presented).

The structure of the AO and VO trials was identical: both consisted of pairs of speech stimuli (from the three base sounds) that were either the same (two repetitions each of [ba]-[ba], [da]-[da], [ga]-[ga], $n = 6$) or different (two repetitions each of [ba]-[da], [ba]-[ga], [da]-[ga], $n = 6$, $N = 12$), with stimulus order within pairs counterbalanced across the two repetitions). The only difference was mode of delivery: in AO trials, pairs

277

of sounds were each accompanied by a static image of the speaker for the duration of the speech sounds; in VO trials, each member of the pair was presented as the dynamic face of the speaker without any sound.

The thirty-six AV trials were presented in the same way as the AO and VO trials, but with both audio and dynamic visual representations. Of the thirty-six AV stimulus pairs, eighteen were pairs in which both members were congruent (e.g. A [ba] + V [ba]), and eighteen were pairs in which at least one member was incongruent (e.g. A [ba] + V [ga]). And in the eighteen trials in each of the two types (congruent/incongruent), nine were same pairs and nine were different pairs. The structure of these thirty-six AV trials is set out in Table 1. The trials were blocked (AO, VO, AV) and the order of these blocks was counterbalanced between participants.

With the sound level set at 65 dB, the stimuli were presented via a 17-inch monitor in the participants' sagittal plane 60 cm from the seating position. An AX discrimination task was used: in each trial, participants heard two sounds and were required to indicate whether the two items were the same or different. To ensure that children understood the task, they were first presented with practice items. In these, the experimenter demonstrated the concepts 'different' and 'same' by showing the child a picture of a circle and a star, and a picture of two circles. Children were then presented with a man and a woman saying different words and were asked to say if the words, not the voices or the faces, were the same or different. After as many practice trials as required, test trials were presented. In these, the first item in each trial was always produced by the male speaker and the second by the female speaker. Children were asked whether the lady was saying the same thing as the man, and the experimenter took down responses on a printed sheet, later transcribed to file. The task took around 20–30 minutes, including breaks as required. The experimenter in the room verbally alerted children to the onset of each stimulus and ensured that they maintained full attention throughout. No trial began until the child was fully attending to the screen.

The dependent variables for the AO and VO speech perception were discrimination index scores. These discrimination indices were calculated by taking the difference between the number of correct 'different' responses on trials in which the stimuli were in fact different (AB trials) (hits) and the number of incorrect 'different' responses on same (AA) trials (false positives) divided by the total number of AB trials ($n = 9$), as set out below.

$$\text{Discrimination Index (DI)} = \big[\big(\Sigma \text{ 'different' responses on AB trials}\big) \\ -\big(\Sigma \text{ 'different' responses on AA trials}\big)\big] \\ /\text{Number of AB trials}$$

278

TABLE 1. *The structure of the AV stimuli*

Total: 36 AV stimuli			
Congruent pairs (*n* = 18) Same auditory and visual components *Description*		**Incongruent pairs (*n* = 18)** Different auditory and visual components *Description*	
9 AA	9 same pair trials: auditory and visual components identically congruent in each stimulus, e.g. Auditory [ba] + Visual [ba] vs. Auditory [ba] + Visual [ba]	**9 AA**	9 same pair trials: auditory and visual components identically incongruent in each stimulus, e.g. Auditory [ba] + Visual [ga] vs. Auditory [ba] + Visual [ga]
9 AB	9 different pair trials: auditory and visual components congruent within but not between each stimulus e.g. Auditory [ba] + Visual [ba] vs. Auditory [ga] + Visual [ga]	**9 AB**	9 different pair trials: auditory & visual components congruent in 1 stimulus, incongruent in other, e.g. Auditory [ba] + Visual [ba] vs. Auditory [ba] + Visual [ga] OR Auditory [ba] + Visual [ba] vs. Auditory [da] + Visual [ga]

This yields a score between −1 and +1, where +1 indicates the highest possible level of performance, 0 indicates chance level responding, and −1 indicates below chance performance. This formula was used to derive an AO DI score and a VO DI score for each child.

For the AV speech perception trials, the construct of interest was visual speech influence and this was measured by a Visual Speech Index (VSI). The nine incongruent AB trials (see lower right quadrant of Table 1) all involved pairs of items in which the auditory and visual components were congruent in one member of the pair and incongruent in the other, with the difference being in both the auditory and the visual component ($n = 5$) or differing in only the visual component ($n = 4$). Only the four AB incongruent trials on which the visual component differed between the members of the pair were of importance (the other trials types were included to counterbalance the number of trials on which 'same' and 'different' were the correct response). Whether a 'same' or a 'different' response was made to these four AB incongruent trials determined whether children's responses were visually influenced or not, and this is set out in Table 2.

The VSI is given by a proportion: the number of 'different' (visually based) responses on the AB incongruent trials on which the visual component differed, divided by the total number of AB incongruent trials on which the visual component differed ($n = 4$):

$$\text{Visual Speech Index (VSI)} = \big(\Sigma \text{ 'different' responses on AB Incongruent trials with different visual components}\big)$$
$$/\big(\text{Number of AB Incongruent trials with different visual components}\big)$$

The maximum possible VSI of 1 indicates very strong visual speech influence, 0·5 indicates chance level, and below 0·5 indicates below chance level.

Native vs. non-native auditory language specific speech perception test. The native/non-native speech perception stimuli consisted of three Thai syllables spoken by a female native Thai speaker: the voiceless aspirated bilabial stop [pʰa], the voiceless unaspirated bilabial stop [pa], and the pre-voiced bilabial stop [ba]. Three different exemplars of each of the syllables were used. The three sounds were arranged in this task into two speech contrasts, one that is native in English, [pa] vs [pʰa] as in b̲in vs. p̲in, and a non-native contrast, [ba] vs. [pa], both perceived as the phoneme /b/ in English.

The stimuli were presented using PsyScript software (Bates & D'Oliveiro, 2003) via a high-quality loudspeaker placed on top of a computer monitor

TABLE 2. *Scoring for the Visual Speech Index. The total number of visually based responses to the Visual Component only different Incongruent AB pairs was divided by the total possible score (4) to provide the Visual Speech Index (VSI) score.*

Incongruent AB pairs differing on:	Example	'same' response indicates:	'different' response indicates:
Both auditory & visual components ($n = 5$)	Aud[ba]Vis[ba] vs. Aud[da]Vis[ga]	Error	Auditory OR Visually based
Visual component only ($n = 4$)	Aud[ba]Vis[ba] vs. Aud[ba]Vis[ga]	Auditory based	Visually based

(used to present cartoon clips for correct responses). Children were seated with their parents 60 cm away from the monitor in their sagittal plane. In front of them was a response box, a modified game controller with a large red button (7 cm diameter) which children pressed to record their responses.

A category change paradigm was used that included two types of trial, change and no-change. Each trial always consisted of ten speech sounds. In change trials, a background sound was repeated two to six times (randomly varying across trials), after which a different (change) sound was played for eight to four times such that there were always ten sounds per trial, no matter when the change was introduced. In no-change trials the same background sound was played ten times. This procedure ensured that participants could not know at the start of any given trial whether it would be a change or a no-change trial, nor when a different stimulus would be played if at all in any given series of ten. The training phase consisted of three components: demonstration, pre-task, and task competence. In the Demonstration phase, children were presented with the sound of a rooster 'crow' as a background sound, and were instructed to press the response button as soon as they heard a cow's 'moo'. There were four trials (two change and two no-change), and these were repeated if necessary. This was followed by the Pre-Task Competence Phase, in which the rooster and cow sounds were replaced by the minimal pair *rag* and *rug*, spoken by a female native Australian English speaker. There were eight trials in total in this phase (four change and four no-change), and children were required to respond correctly to six of the eight items in order to proceed to the next phases. In the Task Competence phase, there were eight trials in which children were required to discriminate randomly chosen speech contrasts used in the testing phase. In the Test phase, there were eighteen native (English native [pa] vs. [pʰa]) and eighteen non-native (Thai native [pa] vs. [ba]) trials with equal numbers of change and no-change trials. The native and non-native speech contrasts were presented in separate blocks, the order of which was counterbalanced

between subjects with randomized trial order within each block. The particular exemplar of each sound played in the contrast was randomized in order to ensure that children responded to differences between phonetic categories rather than idiosyncratic acoustic differences. Children were instructed to press the response button as soon as the sound changed, but to do nothing if the sound did not change. In all phases, each correct response was followed by a 5-second cartoon clip as a reward.

The dependent variables for the native/non-native speech perception measure was a discrimination index for each of the native (N) and the non-native (NN) speech contrasts. In the analyses of variance, each score was used, and in the regression analyses, the difference score between native and non-native DIs (N–NN) was used. The discrimination index scores were calculated as follows.

$$\text{Discrimination Index (DI)} = \big[\big(\Sigma\,\textit{different}\text{ responses on AB trials}\big) \\ -\big(\Sigma\,\textit{different}\text{ responses on AA trials}\big)\big] \\ /\text{Number of AB trials}$$

Receptive vocabulary knowledge test. Receptive vocabulary size was measured using the Peabody Picture Vocabulary Test (PPVT-III) (Dunn & Dunn, 1997), a commonly used standard receptive vocabulary test that requires minimal verbal instructions, as responses consist of a picture pointing task. Blocks of twelve words are presented in order of difficulty. Each child began with a block of twelve items appropriate for their age group. In each trial, children were shown four pictures and asked to point to the target word named by the experimenter. The test progressed until a child made eight errors in a block. A score for each child was computed as per the PPVT scoring manual.

RESULTS

Tests for age differences

Descriptive statistics for all variables over age are presented in Table 3, along with *t*- and *F*-values for tests of three- vs. four-year-olds. There were significant increases from three to four years in vocabulary, AO speech perception, and visual speech influence, and a marginal ($p = \cdot07$) difference for VO speech. For native/non-native speech perception, a two-way native × non-native discrimination index score × age analysis of variance (ANOVA) showed that overall four-year-olds performed significantly better than the three-year-olds [$F(1,46) = 7\cdot709$, $p < \cdot01$], but there was neither any significant main effect of native vs. non-native (N–NN) speech perception ($p > \cdot05$), nor any interaction of N–NN with age ($p > \cdot05$) (see Table 3).

282

TABLE 3. *Means (and standard deviations) of vocabulary (PPVT), auditory-only speech perception (AO), visual-only speech perception (VO), visual speech influence (VSI) scores, and language-specific speech perception (N, NN, N–NN) scores, by age group, and t-test results across age group.*

	3 yr olds **M** (SD)	4 yr olds **M** (SD)	Total **M** (SD)	3 vs. 4 yr old *t*- and *F*-values
PPVT	**43·3** (14·3)	**67·6** (13·0)	**55·44** (18·28)	$t(46) = -6·15$***
AO	**0·59** (0·21)	**0·79** (0·15)	**0·68** (0·21)	$t(46) = 4·17$***
VO	**0·49** (0·16)	**0·57** (0·13)	**0·53** (0·15)	$t(46) = 1·86, p = ·07$
VSI	**0·53** (0·18)	**0·78** (0·23)	**0·65** (0·24)	$t(46) = 3·92$***
N	**0·01** (0·19)	**0·16** (0·17)	**0·09** (0·20)	}
NN	**0·05** (0·17)	**0·11** (0·21)	**0·08** (0·19)	} $F(1,46) = 7·71$**
N–NN	**−0·03** (0·27)	**0·05** (0·31)	**0·01** (0·29)	}

NOTES: ** sig at $\alpha = ·01$; *** sig at $\alpha = ·001$.

One-sample *t*-tests against chance were conducted separately for the three- and four-year-old groups for AO, VO, N and NN, and VSI scores. AO and VO scores were significantly above chance for both age groups [all *t*s (23) > 14·08, all *p* values < ·001]. For the four-year-old group, both NN and N were significantly greater than chance [$t(23) = 2·52$, $p = ·019$ and $t(23) = 4·53$, $p < ·001$, respectively], but for the three-year-old group neither N nor NN differed significantly from chance [$t(23) = 1·31$, $p = ·205$, and $t(23) = 0·35$, $p = ·728$, respectively]. For visual speech influence, the four-year-olds scored significantly above chance on the VSI measure [$t(23) = 5·88$, $p < ·001$], but the three-year-olds did not [$t(23) = 0·69$, $p = ·50$].

Prediction of vocabulary size

A sequential multiple regression analysis was performed with vocabulary scores as the criterion and five predictors – age, AO, N–NN, VO, and visual speech influence. Evaluation of assumptions was satisfactory after two outliers were omitted based on Mahalanobis distance. The five predictors were entered into the model in three steps. In Step 1, exact age in months and days was entered in order to partial out any effects on vocabulary due simply to experience. In Step 2, the auditory perception variables, AO and N–NN, were entered, then in Step 3, the visual perception variables, VO and VSI scores, were entered to investigate whether performance in either or both of the auditory and visual measures contributed to vocabulary. As the focus is the role of visual information in linguistic (vocabulary) development, the visual variables were entered after the auditory variables to investigate whether visual information contributed to vocabulary after auditory contributions were considered.

TABLE 4. *Correlation coefficients, with age partialled out, for vocabulary (PPVT), auditory-only speech perception (AO), language-specific speech perception (N–NN), visual-only speech perception (VO), and visual speech influence (VSI);* N = 46.

	Vocabulary	AO	N–NN	VO	VSI
Vocabulary	1				
AO	.32*	1			
N-NN	.46**	.05	1		
VO	0	.04	.02	1	
VSI	.25	.26	.04	.08	1

NOTES: * sig. at α = ·05; ** sig. at α = ·01; *** sig. at α = ·001; two-tailed test.

Correlation coefficients were calculated with age partialled out, and are presented in Table 4. There were significant correlations between vocabulary and AO speech perception ($r = ·32$, $p < ·05$) and between vocabulary and N–NN scores ($r = ·46$, $p < ·001$). The only other correlations of note were between the VSI score and vocabulary ($r = ·25$) and the VSI score and AO speech perception ($r = ·26$), but neither of these were significant ($p = ·096$ and $.085$, respectively).

Table 5 presents the unstandardized regression coefficients, (B), standardized regression coefficients (β), and R^2 change for the predictors at their entry step, and final B and β. At Step 1, Age significantly predicted vocabulary size and significantly increased R^2. At Step 2, Age, AO, and N–NN all significantly predicted vocabulary size, and AO and N–NN significantly increased R^2 over and above the contribution of Age. At Step 3, Age and N–NN still significantly predicted vocabulary size, and AO was a marginally significant predictor ($p = ·058$), but visual speech influence and VO were not significant predictors – the addition of these two visual speech perception variables did not increase R^2. Thus, there was no unique prediction of vocabulary size by visual speech measures once age and auditory perception measures were taken into account. These results suggest that, while there was the expected predictive link between auditory speech perception (AO and N–NN) and receptive vocabulary, there was no unique predictive relationship between the visual measures (VO speech perception and visual speech influence) and vocabulary.

DISCUSSION

This study had two foci, each with two parts: (1) the development of (a) auditory (AO and N–NN), and (b) visual (VO and Visual Speech Influence) speech perception in three- and four-year-olds, and (2) how (a)

284

TABLE 5. *Sequential multiple regression of Age and AO, N–NN, VO, and VSI scores as predictors of PPVT vocabulary scores; N = 46.*

Step	Variables	B	β	R² change	F-value
1	Age	20·12	0·66***	·44	34·53***
2	Age	14·19	0·47***	·17	9·26***
	AO	26·08	0·27*		
	N–NN	21·83	0·34**		
3	Age	11·56	0·38**	·02	0·78
	AO	22·47	0·23 (p = ·058)		
	N–NN	21·61	0·34**		
	VO	−3·54	−0·02		
	VSI	14·72	0·17		

NOTES: * sig. at $\alpha = \cdot 05$; ** sig. at $\alpha = \cdot 01$; *** sig. at $\alpha = \cdot 001$.

auditory and (b) visual speech perception measures might be related to language development, specifically vocabulary.

With respect to development over age, both visual only speech perception and visual speech influence improved over age from three to four years (though VO was at $p = \cdot 07$). VO scores were significantly above chance at each age, but VSI scores were above chance in the four- but not the three-year-olds. It is possible that the auditory–visual speech perception matching task and/or the dependent variable, the VSI, was difficult or complex for the three-year-olds. In future studies a more robust (e.g. based on more trials) VSI measure would be desirable. Nevertheless, that there is improved use of visual speech over such a small age range is not only statistically, but also theoretically, significant, and is in accord with previous studies (Sekiyama & Burnham, 2008; Erdener & Burnham, 2013; Hockley & Polka, 1994; Massaro, 1984; Sekiyama & Burnham, 2008).

There were also significant improvements over age for AO speech perception, N and NN speech perception, and receptive vocabulary. However, like visual speech influence, N and NN speech perception were only significantly above chance for the four-year-olds, and again, this could be due to task difficulty for the three-year-olds.

Given the age-related improvements in all measures, correlations between these measures might be expected. Correlation coefficients, with age partialled out, showed that neither VO nor VSI correlated with any other measure or with each other, but that both auditory measures, AO and N–NN, correlated positively with vocabulary size.

So, our first hypothesis, that four-year-olds should perform better than three-year-olds on both auditory and visual measures, was confirmed (with some caveats). In addition, the auditory, but not the visual, measures were positively correlated with vocabulary size, partialled for age. This brings

285

us to the second hypothesis, that auditory speech perception measures should predict vocabulary, and that visual speech perception measures should also make a unique contribution to the prediction of vocabulary size. Only the first part of this hypothesis was supported. Both auditory speech perception measures, AO and N–NN, predicted vocabulary size even after age was accounted for, but only N–NN remained a significant predictor after VO and visual speech influence were entered (with marginal involvement of AO; $p = \cdot058$). Thus, N–NN, native language attunement, is a robust predictor of vocabulary size independent of age. This is of interest with respect to studies of the second period of perceptual attunement (Burnham, 2003; Burnham et al., 1991; Horlyck et al., 2012), which have shown that N–NN predicts reading and reading related abilities at reading onset of around six years of age. Now it can additionally be concluded that native language attunement is integrally involved in age-appropriate language development both at reading onset (with respect to reading-related measures), and BEFORE reading onset (with respect to vocabulary size).

The second half of the second hypothesis was not supported; neither VO nor VSI correlated with any other measures, nor with each other, and neither were significant predictors of vocabulary size. The most straightforward interpretation of these results is that, for this age group, the relationship between speech perception and language development is based on auditory speech perception only. This does not mean that visual speech perception is unimportant, as it appears to contribute to linguistic development either at this or later ages in other aspects, e.g. articulation ability (Desjardins, Rogers & Werker, 1997) and phonological processing (Dodd, McIntosh, Erdener & Burnham, 2008). This is similar to what was found by Erdener and Burnham (2013), in that, while reading-age children's visual speech influence was predicted by VO and N–NN, for adults, visual speech influence was only predicted by AO. So, both before and after reading onset, perceptual attunement and visual speech influence are unrelated, but during reading acquisition they are related (Erdener & Burnham, 2013, school children results).

Coupling the results of this study and those of other studies, a tentative account of auditory–visual speech perception development can be proposed. In infancy, visual information is integral in speech perception, and there are parallels between auditory-only and auditory–visual speech perception phenomena (Burnham, 1998; Burnham & Dodd, 2004; Burnham & Sekiyama, 2012; Desjardins & Werker, 2004; Rosenblum et al., 1997), including auditory and auditory–visual perceptual attunement (Burnham & Dodd, 1998; Dodd & Burnham, 1988; Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastian-Galles & Werker, 2007). Beyond infancy and throughout the early childhood period, visual speech perception skills

continue to develop (this study; Hockley & Polka, 1994; Massaro, 1984). While visual speech perception skills are not related to vocabulary in pre-readers, these visual speech perception skills are put to good use once reading instruction begins; those children who are good at using visual speech information also have sharper perceptual attunement – greater perceptual superiority for native over non-native speech sounds (Erdener & Burnham, 2013). Thus, during this reading acquisition period, attention to visual information, just like selective attention to native over non-native sounds, may well aid phoneme-to-grapheme mapping, a skill vital for reading, though investigation of any direct links here requires further research. Later, at around eight years of age, the temporarily augmented bias for native over non-native sounds is reduced as children become proficient readers (Burnham, 2003). And later, adults' visual speech influence is predicted by auditory-only speech perception but not by N–NN perceptual attunement (Erdener & Burnham, 2013).

Thus it seems that visual speech information plays a role in both perceptual attunement in infancy (Burnham & Dodd, 2004; Weikum et al., 2007; Werker & Tees, 1984) and in the second period of perceptual attunement around reading age (Burnham, 2003; Horlyck et al., 2012), but does not play a role in the interim – in three- to four-year-old pre-readers. However, in this three- to four-year-old period, native language perceptual attunement is related to vocabulary size. Thus the relationship between perceptual attunement and language development appears to be important throughout infancy, preschool, and childhood, whereas visual speech perception is possibly only related to language development in infants and at reading onset. In order to test this tentative conclusion rigorously, future research could examine whether other auditory–visual speech perception skills, (i) auditory-visual perceptual attunement (native vs. non-native auditory–visual perceptual discrimination) (Burnham, 2003; Horlyck et al., 2012) or (ii) visual facilitation of auditory speech perception in tasks other than the visual fill-in effect (Jerger et al., 2014), predict vocabulary development at this age.

REFERENCES

Bates, T. C. & D'Oliveiro, L. (2003). PsyScript: a Macintosh application for scripting experiments. *Behavior Research Methods, Instruments & Computers* **35**(4), 565–76.

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C. & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics* **33**(3), 643–64.

Bundgaard-Nielsen, R. L., Best, C. T. & Tyler, M. D. (2011a). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* **33**(3), 433–61.

287

Bundgaard-Nielsen, R. L., Best, C. T. & Tyler, M. D. (2011b). Vocabulary size matters: the assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* **32**(1), 51–67.

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

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

Burnham, D. & Dodd, B. (1998) Familiarity and novelty in infant cross-language studies: factors, problems, and a possible solution. In C. Rovee-Collier (ed.), *Advances in infancy research*, 12 (pp. 170–87). Norwood, NJ: Ablex.

Burnham, D. & Dodd, B. (2004). Auditory-visual speech integration by pre-linguistic infants: perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology* **44**, 209–20.

Burnham, D. K., Earnshaw, L. J. & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language* **18**, 231–60.

Burnham, D. & Sekiyama, K. (2012). Investigating auditory-visual speech perception development. In G. Bailly, P. Perrier & E. Vatikiotis-Bateson (eds), (in preparation). *Advances in audio-visual speech processing* (pp. 62–75). Cambridge: Cambridge University Press.

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

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.

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

Dodd, B. & Burnham, D. K. (1988). Processing speechread information. *The Volta Review: New Reflections on Speechreading* **90**, 45–60.

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

Elliott, L. L., Hammer, M. A. & Scholl, M. E. (1989). Fine-grained auditory discrimination in normal children and children with language-learning problems. *Journal of Speech, Language, and Hearing Research* **32**(1), 112–9.

Erdener, D. & Burnham, D. (2013). The relationship between auditory-visual speech perception and language-specific speech perception at the onset of reading instruction in English-speaking children. *Journal of Experimental Child Psychology* **116**, 120–38.

Hockley, N. & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *Journal of the Acoustical Society of America* **96**, 3309.

Horlyck, S., Reid, A. & Burnham, D. (2012). The relationship between learning to read and language-specific speech perception: maturation versus experience. *Scientific Studies of Reading* **16**(3), 218–39.

Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N. & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: a new multimodal picture–word task. *Journal of Experimental Child Psychology* **102**(1), 40–59.

Jerger, S., Damian, M., Tye-Murray, N. & Abdi, H. (2014). Children use visual speech to compensate for non-intact auditory speech. *Journal of Experimental Child Psychology* **126**, 295–312.

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), 237–64.

288

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development* **55**, 1777–88.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* **264**, 746–8.

Rosenblum, L. D., Schmuckler, M. A. & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics* **59**, 347–57.

Sekiyama, K. & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science* **11**(2), 306–20.

Sekiyama, K. & Tohkura, Y. (1991). McGurk effect in non-English listeners: few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America* **90**(4), 1797–805.

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics* **21**, 427–44.

Sumby, W. H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* **26**, 212–5.

Tsao, F. M., Liu, H. M. & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development* **75**(4), 1067–84.

Vance, M., Rosen, S. & Coleman, M. (2009). Assessing speech perception in young children and relationships with language skills. *International Journal of Audiology* **48**(10), 708–17.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N. & Werker, J. F. (2007). Visual language discrimination in infancy. *Science* **316**(5828), 1159.

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