*J. Child Lang.* **35** (2008), 809–822. © 2008 Cambridge University Press
doi:10.1017/S0305000908008817 Printed in the United Kingdom

# Audiovisual speech recalibration in children*

SABINE VAN LINDEN AND JEAN VROOMEN
*Department of Psychology, Tilburg University, the Netherlands*

(*Received* 26 *June* 2006. *Revised* 7 *March* 2007)

ABSTRACT

In order to examine whether children adjust their phonetic speech categories, children of two age groups, five-year-olds and eight-year-olds, were exposed to a video of a face saying /aba/ or /ada/ accompanied by an auditory ambiguous speech sound halfway between /b/ and /d/. The effect of exposure to these audiovisual stimuli was measured on subsequently delivered auditory-only speech identification trials. Results were compared to a control condition in which the audiovisual exposure stimuli contained non-ambiguous and congruent sounds /aba/ or /ada/. The older children learned to categorize the initially ambiguous speech sound in accord with the previously seen lip-read information (i.e. recalibration), but this was not the case for the younger age group. Moreover, all children displayed a tendency to report the stimulus that they were exposed to during the exposure phase. Methodological improvements for adjusting such a response bias are discussed.

Speech is perceived by both ear and eye. When the auditory signal is degraded or embedded in noise, seeing the speaker significantly enhances speech understanding (Sumby & Pollack, 1954). The contribution of visual information to speech is not restricted to conditions of degraded auditory input, but it also occurs when the sound is clear. As originally demonstrated by McGurk & MacDonald (1976), the perception of non-ambiguous speech tokens can be modified by the simultaneous presentation of visually incongruent articulatory gestures. Typically, when presented with an auditory syllable /ba/ dubbed onto a face articulating /ga/, participants report hearing /da/. The occurrence of this so-called McGurk effect has been taken as a particularly powerful demonstration of the use of visual information in speech perception.

[*] Address for correspondence: Jean Vroomen, Tilburg University, Dept of Psychology, Warandelaan 2, 5000 LE Tilburg, The Netherlands. tel: +31-(0)13-4662394; fax: +31-(0)13-4662067. e-mail: j.vroomen@uvt.nl



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

The use of lip-read information for speech perception has not only been demonstrated in adults, but also in children and infants. Several developmental studies suggest that integration of visual and auditory speech is present early in life. For example, infants at 4·5 months of age prefer to look at a face in which the articulatory gestures are in congruence with the heard vowel rather than to a face with incongruent speech gestures (Kuhl & Meltzoff, 1982). Five-month-olds are also susceptible to the McGurk effect, as for example demonstrated by Rosenblum, Schmuckler & Johnson (1997; see also Burnham & Dodd, 2004; Desjardins & Werker, 2004).

Despite demonstrations that lip-reading contributes to speech perception in early infancy, it has also been demonstrated that the impact of lip-reading on speech perception increases with age. In the original study by McGurk & MacDonald (1976), adults (18–40-year-olds) were compared with children. Results showed that adults were more influenced by incongruent visual information than two younger groups (3–4- and 7–8-year-olds) that did not differ consistently. This developmental trend has been confirmed in later studies (Bruce *et al*., 2000; Burnham & Sekiyama, 2004; Desjardins, Rogers & Werker, 1997; Sekiyama & Burnham, 2004; Wightman, Kistler & Brungart, 2006). For example, Hockley & Polka (1994) presented audiovisual conflicting stimuli to five different age-groups: five-, seven-, nine- and eleven-year-olds and adults. Results again showed that the impact of lip-reading on speech identification increased with age, even beyond the age of twelve, as only half of the eleven-year-old group responded in an adult-like manner. Likewise, Massaro (1984) compared six-year-olds with adults on an audiovisual speech identification task and found that children had only half the visual influence shown by adults. In this study, children and adults also performed a secondary task in which they indicated whether or not the mouth had moved. Despite this attentional focus on the region of the mouth, the difference in visual influence between the two age groups remained, thus suggesting that it was unlikely that children's low susceptibility to visual information was due to paying less attention to the face.

Such a developmental trend in the impact of visual speech may either suggest that lip-reading is an ability that simply needs to mature, or alternatively that linguistic experience is necessary, possibly because visible articulation is initially not well-specified. Exposure to audiovisual speech may then be necessary to develop phonetic representations more completely. This linguistic argument is in line with the finding that Japanese-speaking children are less skilled than English-speaking children at lip-reading at six years old (Sekiyama & Burnham, 2004). This interpretation is also in line with the finding that dyslexics with poor phonological representations are also poor lip-readers (de Gelder & Vroomen, 1998). Alternatively, rather than experience with listening to



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

speech, it may also be that experience producing speech contributes to lip-reading ability. A finding that is in agreement with the latter is that, among preschoolers (four- to five-year olds), those who substitute phonemes in speech production are also poorer lip-readers and tend to be less influenced by the visual component of an incongruent audiovisual syllable than children who do not substitute phonemes (e.g. Desjardins *et al.*, 1997).

In this present study, we further explored the developmental trend in the use of lip-read information. So far, all previous studies have looked at the biasing effect that a lip-read stimulus has on a simultaneously presented auditory speech token. Recently, though, it has been shown that besides such a direct bias effect, lip-reading may also serve a very different role in speech perception, namely in the RECALIBRATION of ambiguous auditory speech (Bertelson, Vroomen & de Gelder, 2003; van Linden & Vroomen, 2007; Vroomen, van Linden, de Gelder & Bertelson, 2007; Vroomen, van Linden, Keetels, de Gelder & Bertelson, 2004). In this view, lip-read information serves as a teacher of the phonetic system, showing how to interpret an initially ambiguous speech sound. Recalibration by lip-read speech has been demonstrated by exposing participants to an auditory ambiguous speech token (henceforth A?) intermediate between /aba/ and /ada/ dubbed onto a visible face articulating either /aba/ or /ada/ (A?Vb and A?Vd, respectively; A = Auditory, V = Visual). Following a short exposure phase (usually in the order of eight trials) to either one of these audiovisual stimuli, participants were given immediately thereafter auditory-only test trials in which they categorized the auditory ambiguous token (A?) and its two closest neighbours on the phonetic continuum (A? − 1 and A? + 1) as either /aba/ or /ada/. Results showed that following exposure to A?Vb, participants were MORE likely to categorize the ambiguous test tokens as /aba/ if compared with exposure to A?Vd. The occurrence of this after-effect (i.e. the difference in the percentage of /aba/ responses on test trials following exposure to A?Vb versus A?Vd) represents the basic recalibration phenomenon. Presumably, listeners resolved during the exposure phase the conflict between the heard and the lip-read speech information by shifting the phoneme boundary such that the ambiguous sound was categorized in accordance with the lip-read information it was combined with. This learning effect could then in turn be observed as an after-effect in the test trials.

To ensure that this phenomenon was not due to some kind of priming or response bias (e.g. in the test phase participants might simply be biased to repeat what they had heard during the exposure phase), participants in the study by Bertelson *et al.* (2003) were also exposed to auditory non-ambiguous and audiovisual congruent adapters AbVb and AdVd. These adapters were not expected to evoke recalibration – because there was no conflict between the heard and seen information – but selective speech



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

adaptation (Eimas & Corbit, 1973). Unlike recalibration, selective speech adaptation does not depend upon exposure to an audiovisual conflict, but upon prolonged exposure to an auditory non-ambiguous speech sound as it probably reflects 'fatigue' of some of the relevant acoustic processes (Eimas & Corbit, 1973; Samuel, 1986). In line with this prediction, following exposure to AbVb, participants were in subsequently delivered test trials LESS likely to categorize the auditory ambiguous test tokens as /aba/ if compared with exposure to AdVd, an effect that was in the opposite direction of recalibration. Further tests also showed that participants were unable to distinguish the exposure stimuli with ambiguous speech sounds (causing recalibration) from non-ambiguous speech sounds (causing selective adaptation). Participants could thus not reliably discriminate between A?Vb and AbVb, and between A?Vd and AdVd when explicitly asked to do so in a discrimination task. This therefore excluded the possibility that results were caused by any deliberate response strategy (Bertelson *et al.*, 2003).

Here, we wanted to test whether there is, as in the case of the McGurk effect, a developmental trend in the use of visual speech for the purpose of recalibration. Recalibration of auditory speech by lip-read information has never been tested in children, as it has been demonstrated only very recently in adults (for lexically induced recalibration, see also Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2006, 2007; Norris, McQueen & Cutler, 2003). In studies with adults, it has been observed that those who are more susceptible to the McGurk effect will also tend to show larger recalibration effects (van Linden & Vroomen, 2007). Assuming that the contribution of lip-reading to speech increases with age, one might expect a corresponding trend in lip-read-induced recalibration. Younger children might then show smaller recalibration effects than older ones, possibly because they lack a sufficiently detailed representation of the lip-read speech.

To test for cross-age differences, a group of five-year-olds and eight-year-olds participated, as in the original studies by McGurk & MacDonald (1976). Following exposure to the auditory ambiguous and incongruent audiovisual speech tokens A?Vb and A?Vd, we expected to find after-effects indicative of recalibration (i.e. more /aba/ responses following exposure to A?Vb than following exposure to A?Vd) that might be smaller or even absent in the younger age group.

As a control condition, we included the auditory non-ambiguous and audiovisual congruent adapters AbVb and AdVd. We did not expect these adapters to produce substantial selective adaptation effects, but they were nevertheless important from a methodological point of view. The effect of these adapters was expected to be small because the optimal number of exposure trials for obtaining selective adaptation is usually much bigger



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

than for recalibration. Whereas recalibration occurs fast and is at ceiling after only very few exposure trials, selective speech adaptation takes more time to develop and increases with number of exposures (Vroomen *et al.*, 2007; Vroomen *et al.*, 2004). Given that the focus of the current study was on recalibration, we used only a limited amount of exposure trials (eight) that was optimal for obtaining recalibration, but not selective speech adaptation. Another reason why selective speech adaptation was expected to be small is that a developmental trend has been reported here as well (Sussman, 1993; Sussman & Carney, 1989). Sussman & Carney (1989) explored selective speech adaptation in five–six-year-olds, seven–eight-year-olds, nine–ten-year-olds and adults. They observed that when adapted to /ba/, only adults showed a boundary shift on a /ba/–/da/ continuum indicative of selective speech adaptation (i.e. more /da/ responses following exposure to /ba/). For the seven–eight-year-olds and nine–ten-year-olds, no boundary shift was observed, while the youngest age-group actually showed a shift in the opposite direction (i.e. MORE responses consistent with the exposure stimuli), a finding for which no explanation was offered. The differences between children and adults were explained by the children's inability to process all acoustic details, as the youngest children were also less proficient in discriminating the speech tokens. (In a later study, though, it was found that the /ba/ token was not a particularly powerful adapter, and when adapted to /da/, both five-year-olds and adults showed a boundary shift (Sussman, 1993).)

Despite the fact that the adapters with non-ambiguous speech tokens were not expected to produce a substantial selective speech adaptation effect, we included them in the design of our study because they allowed us to address a methodological concern. One possible strategy that children may adopt when tested in an 'exposure-test' design is that, whenever unsure, they repeat the phoneme heard during the exposure phase. Such a deliberate response strategy would effectively boost the after-effect thought to reflect recalibration, which is a problem that might become particularly pertinent when comparing across age-groups, because different-aged children might adopt different strategies. This response strategy might also explain why children sometimes display no selective speech adaptation effects, and why five–six-year-olds may actually show a reversed trend (Sussman & Carney, 1989).

In the present study, we sought to alleviate this problem of response bias by comparing the after-effects of audiovisual adapters containing ambiguous sounds with those containing non-ambiguous sounds (i.e. we subtracted (A?Vb − A?Vd) − (AbVb − AdVd)). This comparison between after-effects of ambiguous versus non-ambiguous sounds annuls any contribution of deliberate response strategies to the extent that its contribution is equally big in both cases. This can be assumed to be the



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

case, because even adults have great difficulty discriminating adapter stimuli containing ambiguous from non-ambiguous sounds when explicitly asked to do so (Bertelson *et al.*, 2003). Hence, one can safely predict that if recalibration is at stake, after-effects of audiovisual adapters containing ambiguous sounds will be bigger than those containing non-ambiguous sounds independent of participants' response biases.

## METHOD

### *Participants*

Twenty five-year-olds and twenty eight-year-olds from a public elementary school in a Dutch village, Hapert, were tested. All children were native Dutch speakers, and only children with normal age-related speaking and hearing proficiency, as judged by their teacher, participated in the experiment. Eight boys and twelve girls with a mean age of 5;5, and ten boys and ten girls with a mean age of 8;6 were tested. Data from two of the five-year-olds were discarded from analyses because one child completed only two experimental sessions, while the other frequently failed to respond during the test. This left eighteen children in the younger age group.

### *Stimuli*

Details of the stimuli have been described elsewhere (Bertelson *et al.*, 2003). In short, a 9-point /aba/–/ada/ speech continuum was created by varying the frequency of the second (F2) formant in equal steps. The midpoint token of this continuum (A?), the most ambiguous one as determined by pilot tests, was used to create the audiovisual ambiguous exposure stimuli. It was dubbed onto the video of a face articulating /aba/ or /ada/, resulting in the two exposure stimuli used to induce recalibration: A?Vb and A?Vd. The two endpoint tokens, clear /aba/ and clear /ada/ were dubbed onto the congruent lip gestures to create the exposure stimuli AbVb and AdVd for the control condition. The three tokens closest to the phoneme boundary, A?, and its two neighbours on the continuum (A?−1 and A?+1), were used for the test trials.

### *Design and procedure*

Design and procedures were adopted from our previous study (Bertelson *et al.*, 2003). The children were tested individually in a quiet room in their school. They were seated in front of a table at a distance of about 50 cm from a 15-inch laptop (Dell Inspiron 2650) on which the visual stimuli were presented. Sounds were delivered over two loudspeakers that were placed besides the screen. The experimenter sat at the side of the table in front of a keyboard. Testing began with a series of practice trials followed by four



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

experimental sessions. Total testing took about half an hour with small pauses in between sessions.

There were four experimental sessions, each consisting of four exposure-test blocks. During an exposure-test block, one of the four bimodal exposure stimuli (AbVb, AdVd, A?Vb or A?Vd) was successively presented eight times (ISI = 1 s.), followed by six auditory-only test tokens. The test tokens consisted of the three midpoint tokens of the continuum (A? − 1, A? and A? + 1), presented twice in counterbalanced order. All four bimodal exposure stimuli were presented once in an experimental session. Total testing thus consisted of 16 exposure-test blocks with 96 test trials (4 exposure stimuli × 4 sessions × 6 test trials). Presentation order of the exposure stimuli was counterbalanced over the experimental sessions. Participants were asked to silently watch the video during exposure. In the test phase, the video turned grey. After participants heard a test stimulus, they were required to report orally whether they had heard /aba/ or /ada/. The experimenter then pressed a corresponding key on the keyboard. Trials in which the child was inattentive or failed to respond were discarded from analyses. The next test trial started after a key press by the experimenter. A practice session containing each of the four different exposure-test blocks was included at the start of the experiment. Practice continued until the child completely understood the task and successfully completed at least two exposure-test blocks. When necessary, the child was encouraged to keep attention on the face during audiovisual exposure.

## RESULTS

Figures 1a–2b present performance in the test phase for each exposure condition and age group. Figures 1a and 1b show the results for the five-year-olds, and Figures 2a and 2b for the eight-year-olds; Figures 1a and 2a show performance following exposure to the ambiguous adapters A?Vb and A?Vd, and Figures 1b and 2b for the non-ambiguous adapters AbVb and AdVd. Each graph shows the proportion of /aba/ responses for the ambiguous token (A?), and its two neighbours on the phonetic continuum (A? − 1 or A? + 1). The first thing to note is that for both the non-ambiguous and ambiguous exposure conditions, children responded more often with /aba/ following exposure to 'aba'-adapters (AbVb and A?Vb) than following exposure to 'ada'-adapters (AdVd and A?Vd). This is visible in the figures in that the solid lines are always above the dotted lines. As mentioned before, our design allowed us to subtract out such an overall tendency to repeat the phoneme heard during the exposure phase, as recalibration devoid of response bias should manifest itself in that the after-effect (the difference between the dotted and solid lines) of ambiguous sounds should be bigger than that of non-ambiguous sounds. As is clearly



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

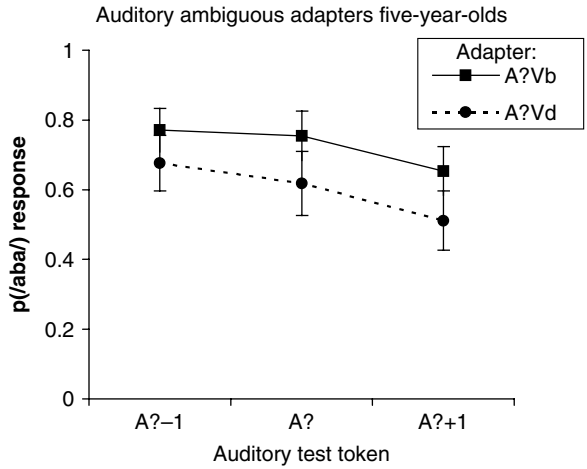



Fig. 1a. Performance on test trials for the five-year-olds after exposure to adapters containing auditory ambiguous sounds (A?) combined with visual /aba/ (A?Vb) or visual /ada/ (A?Vd). The graph shows the proportion of /aba/ responses for the test tokens A?−1 (/aba/-like), A? (ambiguous) and A?+1 (/ada/-like).

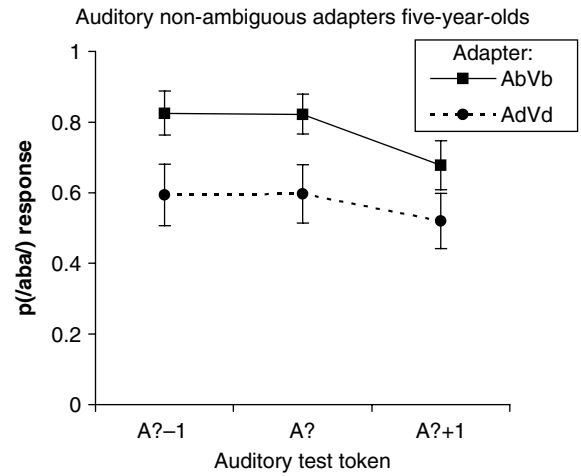



Fig. 1b. Performance on test trials for the five-year-olds after exposure to adapters containing auditory non-ambiguous sounds combined with visual /aba/ (AbVb) or visual /ada/ (AdVd).

visible, for the five-year-olds there was no difference in the size of the after-effects of ambiguous and non-ambiguous sounds (the difference between dotted and solid lines in Figures 1a vs. 1b is the same or even appears



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

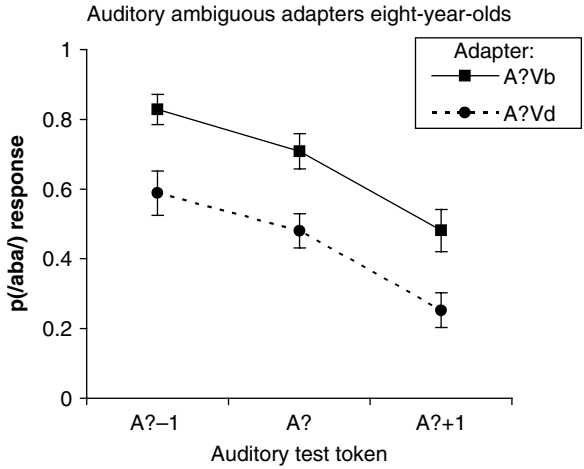



Fig. 2a. Performance on test trials for the eight-year-olds after exposure to adapters containing auditory ambiguous sounds (A?) combined with visual /aba/ (A?Vb) or visual /ada/ (A?Vd).

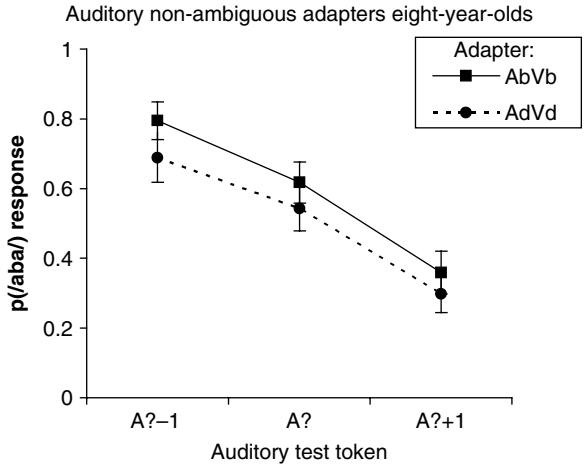



Fig. 2b. Performance on test trials for the eight-year-olds after exposure to adapters containing auditory non-ambiguous sounds combined with visual /aba/ (AbVb) or visual /ada/ (AdVd).

slightly bigger in the case of non-ambiguous sounds). For the eight-year-olds, though, after-effects of ambiguous sounds were bigger than those of non-ambiguous sounds (Figure 2a vs. 2b).

In the 2 (Age-group five- or eight-year-olds) × 2 (Adapter Sound Ambiguous or Non-ambiguous) × 2 (Identity of the visual Adapter /aba/ or



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

TABLE 1. *Mean proportion of /aba/ responses and mean after-effects induced by audiovisual ambiguous and audiovisual non-ambiguous exposure sounds on auditory-only test trials*

| | | Age-group | | | |
|---|---|---|---|---|---|
| | | Five-year-olds | | Eight-year-olds | |
| | Exposure stimulus | /aba/ responses | After-effect | /aba/ responses | After-effect |
| Non-ambiguous sounds | AbVb | 0·77 | 0·20 | 0·59 | 0·08 |
| | AdVd | 0·57 | | 0·51 | |
| Ambiguous sounds | A?Vb | 0·73 | 0·13 | 0·67 | 0·23 |
| | A?Vd | 0·60 | | 0·44 | |

/ada/) × 3 (Test Token A? − 1, A?, A? + 1) overall ANOVA on the proportion of /aba/ responses, there was a main effect of the identity of the adapting stimulus ($F(1, 36) = 16{\cdot}50$, $p < 0{\cdot}001$), because there were more /aba/ responses following exposure to /aba/-adapters than /ada/-adapters, and a main effect of test token ($F(2, 72) = 49{\cdot}37$, $p < 0{\cdot}001$), because there were more /aba/ responses for the /b/-like test token A? − 1 than for the /d/-like test token A? + 1. There was, furthermore, an interaction between test token and age-group ($F(2, 72) = 11{\cdot}75$, $p < 0{\cdot}001$). Inspection of the figures shows that the eight-year-olds had steeper identification functions than the five-year-olds, indicating that the eight-year-olds responded in a more 'categorical' way. Most importantly, there was a significant interaction between age group, ambiguity of the adapter sound and identity of the adapter stimulus ($F(1, 36) = 6{\cdot}48$, $p < 0{\cdot}015$). In the figures this interaction is clearly visible as for the eight-year-olds, the difference between the solid and dotted lines (the after-effect) is bigger for ambiguous than non-ambiguous speech sounds, while there is no such difference for the five-year-olds. No other interactions were significant.

For ease of interpretation, we computed after-effects by subtracting the proportion of /aba/ responses following exposure to A?Vd from A?Vb and AdVd from AbVb (see Table 1), thereby pooling over the three test tokens. Separate paired *t*-tests confirmed that the after-effect of the eight-year-olds for ambiguous sounds was bigger than of the non-ambiguous sounds, 0·23 versus 0·08, respectively ($t(1,19) = 2{\cdot}64$, $p = 0{\cdot}016$), while for the five-year-olds there was no such difference, 0·13 versus 0·20, respectively ($t(1,17) = 1{\cdot}12$, $p = 0{\cdot}28$). The eight-year-olds thus showed the predicted recalibration effect in that their after-effect of ambiguous sounds was bigger than that of non-ambiguous sounds (a 0·15 difference in the predicted



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

direction), while the five-year-olds showed no sign of recalibration (a non-significant 0·07 difference in the opposite direction).

## DISCUSSION

There is by now accumulating evidence that when listeners hear an ambiguous phoneme, they flexibly adjust the phonetic categories of their language in accordance with disambiguating formation that tells what the phoneme should be (Kraljic & Samuel, 2006, 2007; Norris *et al.*, 2003; van Linden, Stekelenburg, Tuomainen & Vroomen, 2007; van Linden & Vroomen, 2007; Vroomen *et al.*, 2007). Here we demonstrate, for the first time, that there is also a developmental trend in this phenomenon. Eight-year-olds, but not five-year-olds, learned to categorize an initially ambiguous speech sound in accord with previously seen lip-read information. Presumably, eight-year-olds adjusted their phoneme boundary to reduce the phonetic conflict in the audiovisual stimuli A?Vb or A?Vd. This shift may occur in the older group but not the younger one because lip-reading is not yet very effective at the age of five. One possibility why the five-year-olds are less proficient is that they lack a sufficiently detailed representation of the visual information. This interpretation is in line with findings showing that young children are less proficient in decoding lip-read speech (Hockley & Polka, 1994; Massaro, 1984; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2004). Experience producing speech (e.g. Desjardins *et al.*, 1997) may be another critical step to develop a visual or articulatory based representation with sufficient detail so that it can be used to recalibrate the existing auditory representation. Young children may in this view thus show no lip-read-induced recalibration because they do not yet decode the lip-read message in sufficient detail.

To examine this further, one might explore whether recalibration in young children can be induced by lexical information. Lexical information can, like lip-read speech, inform a listener what an initially ambiguous sound should be. For example, an ambiguous sound halfway between /b/ and /d/ in the context of *?utter* is likely to be a /b/ rather than a /d/ because *butter* but not *dutter* is a word in English. Several studies have shown that recalibration can indeed be evoked by lexical knowledge (Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2006, 2007; Norris *et al.*, 2003; van Linden *et al.*, 2007), and that lip-read and lexically induced recalibration are similar to each other (van Linden & Vroomen, 2007). If the young children's lack of recalibration is caused by their poor lip-reading abilities, it might nevertheless be possible to observe lexically induced recalibration to the extent that the words are part of the children's lexicon.

Our findings also have methodological consequences for studies measuring after-effects that use a typical exposure-test design. Here, we



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

were concerned that children might adopt a response strategy of simply repeating the phoneme perceived during the foregoing exposure phase. Such a deliberate response strategy would artificially boost the after-effect considered to reflect recalibration. The use of adapters with non-ambiguous and congruent speech sounds, though, allowed us to subtract out the contribution of this response strategy. By so doing, we clearly observed that for the eight-year-olds after-effects of ambiguous and non-ambiguous sounds differed in the expected direction, thus indicating that there was indeed recalibration, while there was no sign of this for the five-year-olds.

Our results may also shed new light on previous studies showing that adults have larger selective speech adaptation effects than children (Sussman & Carney, 1989). This has been attributed to the fact that adults are more categorical in perceiving the test stimuli, a phenomenon we also observed with the older children (see also Hazan & Barrett, 2000; Morrongiello, Robson, Best & Clifton, 1984; Nittrouer & Studdert-Kennedy, 1987). The present results, though, make it clear that in order to measure after-effects in an exposure-test design, a baseline is needed that takes into account the various response strategies that participants might use. One strategy that we observed here is that children may have a tendency to report the exposure stimuli presented during the foregoing exposure phase. For future research on selective speech adaptation or recalibration, it seems important to include a baseline that takes such strategic effects into account, especially when comparing different age-groups because children might differ on this strategy.

In conclusion, here we demonstrated that eight-year-olds, but not five-year-olds learned to categorize ambiguous speech sounds by the use of lip-read information. The five-year-olds may not yet be able to adjust their phonetic boundary because they are not adept decoders of lip-read information. One may thus expect proficiency in lip-reading not only to correlate with susceptibility to the McGurk effect (e.g. Massaro, 1987), it may also appear to be a critical skill for lip-read-induced recalibration (van Linden & Vroomen, 2007). Further testing is needed, though, to understand what critical experience is needed that helps to promote lip-reading, and how it relates to lip-read-induced recalibration of phonetic categories in detail.

## REFERENCES

Bertelson, P., Vroomen, J. & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science* **14**, 592–97.

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S. & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology* **18**, 319–33.



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

Burnham, D. & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology* **45**, 204–220.
Burnham, D. & Sekiyama, K. (2004). When auditory-visual speech perception develops: The locus of the Japanese McGurk effect. *Australian Journal of Psychology* **56**, 108.
de Gelder, B. & Vroomen, J. (1998). Impaired speech perception in poor readers: Evidence from hearing and speech reading. *Brain and Language* **64**, 269–81.
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.
Eimas, P. D. & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology* **4**, 99–109.
Eisner, F. & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics* **67**, 224–38.
Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics* **28**, 377–96.
Hockley, N. S. & Polka, L. A. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *Journal of the Acoustical Society of America* **96**, 3309.
Kraljic, T. & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* **51**, 141–78.
Kraljic, T. & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* **13**, 262–68.
Kraljic, T. & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language* **56**, 1–15.
Kuhl, P. K. & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science* **218**, 1138–41.
Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development* **55**, 1777–88.
Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* **264**, 746–48.
Morrongiello, B. A., Robson, R. C., Best, V. C. T. & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology* **37**, 231–50.
Nittrouer, S. & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research* **30**, 319–29.
Norris, D., McQueen, J. M. & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology* **47**, 204–238.
Rosenblum, L. D., Schmuckler, M. A. & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics* **59**, 347–57.
Samuel, A. G. (1986). Red herring detectors and speech perception: In defence of selective adaptation. *Cognitive Psychology* **18**, 452–99.
Sekiyama, K. & Burnham, D. (2004). Issues in the development of auditory-visual speech perception: Adults, infants, and children. Paper presented at the 8th International Conference on Spoken Language Processing, Jeiu Island, Korea.
Sumby, W. H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* **26**, 212–15.
Sussman, J. E. (1993). Auditory processing in children's speech perception: Results of selective adaptation and discrimination tasks. *Journal of Speech and Hearing Research* **36**, 380–95.
Sussman, J. E. & Carney, A. E. (1989). Effects of transition length on the perception of stop consonants by children and adults. *Journal of Speech and Hearing Research* **32**, 151–60.



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press

van Linden, S., Stekelenburg, J. J., Tuomainen, J. & Vroomen, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience Letters* **420**, 49–52.

van Linden, S. & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance* **33**, 1483–94.

Vroomen, J., van Linden, S., de Gelder, B. & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia* **45**, 572–77.

Vroomen, J., van Linden, S., Keetels, M., de Gelder, W. & Bertelson, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication* **44**, 55–61.

Wightman, F., Kistler, D. & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *Journal of the Acoustical Society of America* **119**, 3940–49.



https://doi.org/10.1017/S0305000908008817 Published online by Cambridge University Press