

Auditory–visual fusion in speech perception in children with cochlear implants

Efrat A. Schorr*, Nathan A. Fox*, Virginie van Wassenhove^{†‡}, and Eric I. Knudsen^{§¶}

*Department of Human Development/Institute of Child Study, University of Maryland, 3304 Benjamin Building, College Park, MD 20742; [†]Visual and Multisensory Perception Laboratory, Department of Psychology, University of California, Los Angeles, CA 90095-1563; [‡]Shimojo Psychophysics Laboratory, Division of Biology, California Institute of Technology, Pasadena, CA 91125; and [§]Department of Neurobiology, Stanford University School of Medicine, Stanford, CA 94305-5125

Contributed by Eric I. Knudsen, October 10, 2005

Speech, for most of us, is a bimodal percept whenever we both hear the voice and see the lip movements of a speaker. Children who are born deaf never have this bimodal experience. We tested children who had been deaf from birth and who subsequently received cochlear implants for their ability to fuse the auditory information provided by their implants with visual information about lip movements for speech perception. For most of the children with implants (92%), perception was dominated by vision when visual and auditory speech information conflicted. For some, bimodal fusion was strong and consistent, demonstrating a remarkable plasticity in their ability to form auditory–visual associations despite the atypical stimulation provided by implants. The likelihood of consistent auditory–visual fusion declined with age at implant beyond 2.5 years, suggesting a sensitive period for bimodal integration in speech perception.

auditory visual integration | deafness | learning | sensitive periods | speech development

Speech is traditionally thought of as an exclusively auditory percept. However, when the face of the speaker is visible, information contained primarily in the movement of the lips contributes powerfully to our perception of speech. This cooperative interaction between the auditory and visual modalities improves our ability to interpret speech accurately, particularly in low-signal or high-noise environments (1–4).

The cross-modal influence of visual information on speech perception is illustrated by a compelling illusion, referred to as the McGurk effect. This illusion is evoked when a listener is presented with an audio recording of one syllable (e.g.,/pa/) while watching a synchronized video recording of a speaker's face articulating a different syllable (e.g.,/ka/). Under these conditions, the majority of adults typically report hearing the syllable/ta/. The illusion is robust and obligatory, and has been demonstrated in adults and children and in numerous languages (5, 6).

The McGurk effect demonstrates that, in most people, the central nervous system combines visual information from the face with acoustic information in creating the speech percept. Because the stimuli that the visual and auditory systems encode is of a very different nature, and because the relationship between changes in lip shape and changes in acoustic spectrum can vary across languages, experience may play a critical role in forming the audiovisual associations that underlie bimodal speech perception.

Children who have been deaf since birth and have received cochlear implants provide a unique population to examine the effects of auditory deprivation and timing of the introduction of auditory experience on the emergence of perceptual and cognitive processes involved in speech perception (7). Cochlear implants produce patterns of auditory nerve activation that differ markedly from those produced normally by the cochlea. Nevertheless, in a dramatic example of brain plasticity, a substantial proportion of children who receive cochlear implants learn to perceive speech remarkably well using their implants

(8–10) and appear able to integrate congruent audiovisual speech stimuli (11–14). However, their ability to fuse conflicting auditory and visual information in speech perception has never been tested. In addition, because these children have received implants at various ages, they offer the opportunity to investigate the importance of age at the time of implant for the development of bimodal speech perception.

Materials and Methods

We tested children who were deaf from birth and had used their cochlear implants for at least 1 year ($n = 36$; mean = 5.85 years). Each child was capable of perceiving spoken language using the implant alone. Participants in the study met the following criteria: they were 5–14 years of age at the time of testing, were profoundly deaf from birth, had a minimum of 1 year of cochlear-implant experience, used oral language as a primary mode of communication, and could perceive spoken language. Speech perception ability was assessed by using the lexical neighborhood test (LNT) and the multisyllabic lexical neighborhood test (MLNT) (15). Performance on the lexically “easy” and lexically “hard” word list from both the LNT and MLNT were averaged. Scores obtained by children with cochlear implants ranged from 50% to 88% (mean = 71.05%, SD = 9.975%). The children could also read lips but before the implant had not experienced a correspondence between lip movement and auditory signals. The implants established such a correspondence for the first time.

Seven types of stimuli were presented: the audio-alone stimuli/pa/and/ka/, the visual-alone stimuli/pa/and/ka/, congruent audiovisual pairs/pa/pa/and/ka/ka/, and the incongruent audiovisual pair consisting of audio/pa/dubbed onto visual/ka/. In this latter condition (McGurk test), people who fuse auditory and visual speech signals report hearing “ta.” Ten trials of each stimulus type were presented for a total of 70 trials. The stimuli were presented in a random, interleaved design. Participants were asked to report what they heard, and their responses were recorded verbatim. Most of the children (60 of 69) reported hearing only “pa,” “ka,” or “ta.” Stimuli were presented by using PSYSCOPE 1.2.5 (<http://psyscope.psy.cmu.edu>). Stimulus presentation software was via Mac OS 9.2. Participants were seated 50 cm from the monitor, facing it directly at 0° azimuth. Videos were displayed centered on a 15-inch Apple G4 monitor on a black background. Sounds were presented from a loudspeaker at 75-dB sound pressure level (the McGurk task was administered by using stimuli adapted from ref. 16).

The unimodal stimuli demonstrated a child's ability to distinguish between/pa/and/ka/using audition with the cochlear implant alone or lip-reading. The congruent audiovisual stimuli/pa/pa/and/ka/ka/tested the consistency of the child's percept of these syllables under bimodal conditions. The incongruent

Conflict of interest statement: No conflicts declared.

[¶]To whom correspondence should be addressed. E-mail: eknudsen@stanford.edu.

© 2005 by The National Academy of Sciences of the USA

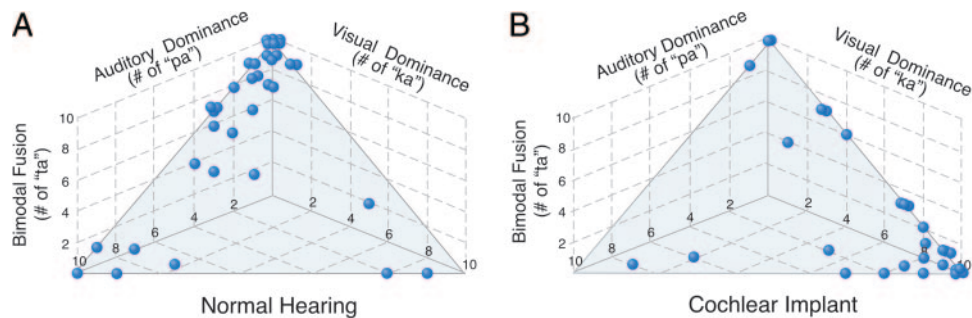


Fig. 1. Responses of individual subjects to the incongruent auditory–visual/pa/ka/stimulus (McGurk test). “ta” responses indicate audiovisual fusion, “pa” responses indicate auditory dominance, and “ka” responses indicate visual dominance. Only data from children who responded consistently with “pa” to the/pa/pa/stimulus and with “ka” to the/ka/ka/stimulus are shown. (A) Responses of children with normal hearing. (B) Responses of children with cochlear implants. Ten/pa/ka/stimuli were presented, randomly interleaved with other stimuli. For 8 children with normal hearing and 7 children with cochlear implants, 1 or 2 of the 10 responses were something other than “pa,” “ka,” or “ta.”

audiovisual stimulus/pa/ka/assessed bimodal fusion in speech perception. The responses of the children with cochlear implants were compared with those of children with normal hearing ($n = 35$). These two groups were matched for age and sex, and they met normative age levels for nonverbal IQ and language proficiency. Parents provided information about their child’s audiological history. Assessments of language functioning and nonverbal IQ were administered to all participants. Nonverbal IQ was assessed with the Kaufman Brief Intelligence Test (17). Mean nonverbal IQ for the children with cochlear implants was 110.75 (SD = 11.60, range of 90–136). Language proficiency was assessed with measures of receptive semantics and syntax. The Peabody Picture Vocabulary Test III (18) was administered to all participants. Mean language proficiency scores for the children with cochlear implants was 96.64 (SD = 21.04, range of 61–134). For participants aged 5:0–7:11, the Grammatic Understanding subtest of the Test of Language Development (Primary) was used (19). For children aged 8:0–11:11, the Grammatic Understanding subtest of the Test of Language Development (Intermediate) was administered (20). For children aged 12:0–14:11, the Listening/Grammar subtest of the Test of Adolescent and Adult Language (3rd Ed.) was used (21).

Results and Discussion

Children with normal hearing perceived both the unimodal and the congruent bimodal stimuli reliably. The congruent audiovisual stimuli/pa/pa/and/ka/ka/were reported correctly as “pa” and “ka,” respectively, on 10 of 10 trials by 20 of 35 children and on no less than 7 of 10 trials by all 35 children. In contrast, on the incongruent audiovisual stimulus/pa/ka/(McGurk test), performance was idiosyncratic (Fig. 1A): 57% (20 of 35) reported “ta” on 7 or more of the 10 trials (consistent fusion); 17% (6 of 35) reported “ta” on 4–6 trials (inconsistent fusion); and 26% (9 of 35) reported “ta” on 3 or fewer trials (poor fusion). Among the children with normal hearing who integrated poorly or inconsistently, the vast majority (80%; 12 of 15) reported “pa” (the auditory component of the stimulus) on most trials when not reporting “ta” (Fig. 1A). Thus, most of the children with normal hearing experienced consistent bimodal fusion, and for those who did not, speech perception was dominated by the auditory stimulus when confronted with visual stimuli that conflicted with auditory stimuli.

Most of the children with cochlear implants also perceived the unimodal and congruent bimodal stimuli accurately and reliably; they responded with “pa” to the/pa/pa/stimulus and with “ka”

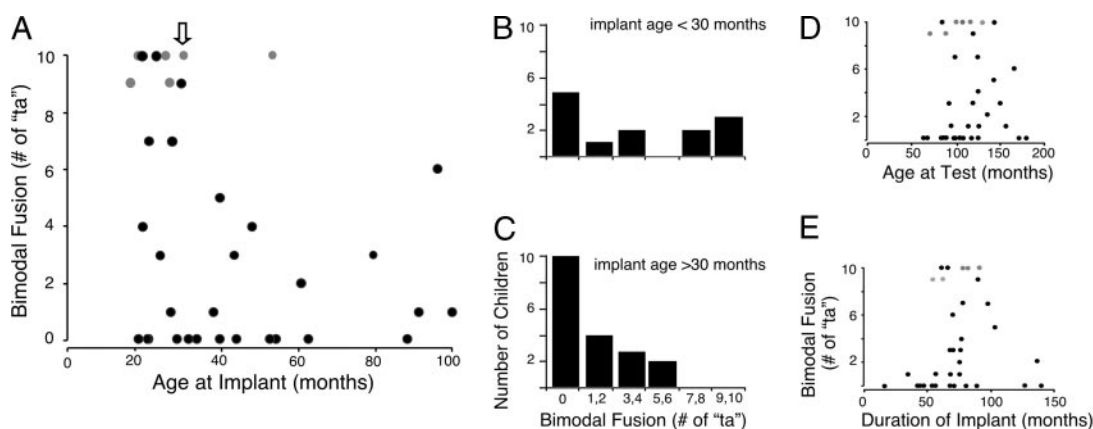


Fig. 2. Effect of age at the time of receiving a cochlear implant on auditory–visual fusion in speech perception. (A) Number of “ta” responses (of 10) to the/pa/ka/stimulus (McGurk test) plotted as a function of age at implant. Black circles represent children who responded consistently and correctly to the/pa/pa/and/ka/ka/stimuli. Gray circles represent children who responded inconsistently or incorrectly to the/ka/ka/stimulus; for example, the child who was 53 months old at implant, and who responded to the/pa/ka/stimulus with “ta” 10 times, also responded to the/ka/ka/stimulus with “ta” 7 times and with “ka” 3 times. The downward arrow indicates 30 months. (B) Number of children younger than 30 months at the time of cochlear implant, showing each level of audiovisual fusion to the/pa/ka/stimulus. Children are grouped according to the number of “ta” responses. Only children who responded consistently to both the/pa/pa/and/ka/ka/stimuli are plotted. (C) Same as in B, except that these are children who were older than 30 months at the time of implant. (D) Number of “ta” responses (of 10) to the/pa/ka/stimulus as a function of age at the time of testing. Other conventions are as in A. (E) Number of “ta” responses (of 10) to the/pa/ka/stimulus as a function of duration wearing the implant. Other conventions are as in A.

to the/ka/ka/stimulus on 7–10 trials. However, 6 (of 36) responded with other syllables on 4–10 of the/ka/ka/trials (by far the most prevalent incorrect response among these children was “ta”). The data for these children with inconsistent responses to the congruent bimodal stimuli (indicated with shaded symbols in Fig. 2*A*, *D*, and *E*) were removed from further analysis (the data for all subjects for all conditions are available as supporting information, which is published on the PNAS web site).

Like the performance of children with normal hearing, the performance of children with cochlear implants on the incongruent audiovisual stimulus was idiosyncratic (Fig. 1*B*): 20% (6 of 30) reported “ta” on 7 or more of the 10 trials (consistent fusion); 10% (3 of 30) reported “ta” on 4–6 trials (inconsistent fusion); and 70% (21 of 30) reported “ta” on 0–3 trials (poor fusion). Among the children who integrated poorly or inconsistently, 88% (22 of 24) reported “ka” (the visual component of the stimulus) on most of the trials when not reporting “ta” (Fig. 1*B*). Thus, most of the children with cochlear implants did not experience reliable bimodal fusion, and speech perception for these children was dominated by the visual stimulus under bimodal conditions. The dominance of the visual stimulus for these children indicates a higher dependence on lip reading, which these children had depended on for speech perception before the implant.

Although the proportion of children who exhibited a consistent fusion was substantially lower among children with cochlear implants (20%) than among children with normal hearing (57%), perhaps, in part, because of the increased difficulty of the task when using an implant, the performance of those children who did exhibit strong bimodal fusion was indistinguishable from that of children with normal hearing with strong bimodal fusion. These children reported “ta” on 7–10 of the/pa/ka/trials, while reporting “pa” on all/pa/pa/trials and “ka” on 8–10 of the/ka/ka/trials. Thus, children can learn to combine visual information about lip movements with the highly unnatural neural activation patterns evoked by the cochlear implants in the processing of speech.

The likelihood of consistent bimodal fusion by a child with a cochlear implant depended on the age of the child at implant (Fig. 2*A*). All of the children who exhibited consistent bimodal fusion (7–10 “ta” reports on the/pa/ka/trials) received their implants before 30 months of age; for this age group, this constituted 38% (5 of 13) of the population (Fig. 2*B*). In contrast, consistent audiovisual fusion was not exhibited by any of the children ($n = 17$) who received implants after 30 months of age (Fig. 2*C*). The difference between these groups was significant at the $P = 0.025$ level (permutation test of means) (22). Bimodal fusion depended neither on the age of the child at the time of testing ($P = 0.669$; Fig. 2*D*) nor on the amount of time wearing the implant ($P = 0.109$; Fig. 2*E*).

The results demonstrate that the consistent fusion of visual with auditory information for speech perception is shaped by experience with bimodal spoken language during early life. When auditory experience with speech is mediated by a cochlear implant, the likelihood of acquiring strong bimodal fusion is increased greatly when the experience begins before 2.5 years of age, suggesting a sensitive period (23). Children who received implants after 2.5 years of age exhibited only inconsistent or no bimodal fusion. This finding is consistent with other reports indicating improved speech perception, language skills, and auditory cortical function among children receiving implants at early ages (24–26). These studies, together with the data reported here, argue strongly for screening children for hearing capabilities and providing cochlear implants when necessary at the earliest possible ages.

We thank P. Knudsen for design of the figures, F. P. Roth for design and supervision of the language assessments, and D. Poeppel for assistance with stimuli preparation. This work was supported by National Institute on Deafness and Other Communication Disorders Grant NRSA 1-F31DC006204-02 (to E.A.S.) and the Eugene L. Derlacki, M.D. Grant for Excellence in the Field of Hearing from the American Hearing Research Foundation (to N.A.F. and E.A.S.).

- de Gelder, B. & Bertelson, P. (2003) *Trends Cognit. Sci.* **7**, 460–468.
- Grant, K. W., Ardell, L. H., Kuhl, P. K. & Sparks, D. W. (1985) *J. Acoust. Soc. Am.* **77**, 671–677.
- Summerfield, Q. (1979) *Phonetica* **36**, 314–331.
- Sumby, W. G. & Pollack, I. (1954) *J. Acoust. Soc. Am.* **26**, 212–215.
- McGurk, H. & MacDonald, J. (1976) *Nature* **264**, 746–748.
- Brancazio, L. (2004) *J. Exp. Psychol.* **30**, 445–463.
- Bergeson, T. R. & Pisoni, D. B. (2003) in *Handbook of Multisensory Integration*, eds., Calvert, G., Spence, C. & Stein, B. E. (MIT Press, Cambridge, MA), pp. 749–772.
- Svirsky, M. A., Robbins, M., Kirk, K. I., Pisoni, D. B. & Miyamoto, R. T. (2000) *Psych. Sci.* **11**, 153–158.
- Waltzman, S. B., Cohen, N. L., Gomolin, L. H., Green, J. E., Shapiro, W. H., Hoffman, R. A. & Roland, J. T. (1997) *Am. J. Otol.* **18**, 342–349.
- Balkany, T. J., Hodges, A. V., Eshraghi, A. A., Butts, S., Bricker, K., Lingvai, J., Polak, M. & King, J. (2002) *Acta Otolaryngol.* **122**, 356–362.
- Niparko, J. K. & Geers, A. E. (2004) *J. Am. Med. Assoc.* **291**, 2378–2380.
- Bergeson, T. R., Pisoni, D. B. & Davis, R. (2005) *Ear Hear.* **26**, 149–164.
- Lachs, L., Pisoni, D. B. & Kirk, K. I. (2001) *Ear Hear.* **22**, 236–251.
- Bergeson, T. R., Pisoni, D. B. & Davis, R. (2003) *Volta Rev.* **103**, 347–370.
- Kirk, K. I., Eisenberg, L. S., Martinez, A. S. & Hay-McCuthcheon, M. (1998) Progress Report No. 22 (Dept. of Otolaryngol.–Head and Neck Surgery, Indiana Univ., Bloomington).
- van Wassenhove, V., Grant, K. W. & Poeppel, D. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 1181–1186.
- Kaufman, A. S. & Kaufman, N. L. (1990) *Kaufman Brief Intelligence Test* (Am. Guidance Service, Inc., Circle Pines, MN).
- Dunn, L. M. & Dunn, L. M. (1997) *Peabody Picture Vocabulary Test III* (Am. Guidance Service, Inc., Circle Pines, MN).
- Newcomer, P. L. & Hammill, D. D. (1997) *Test of Language Development, Primary* (PRO-ED, Austin, TX), 3rd Ed.
- Hammill, D. D. & Newcomer, P. L. (1997) *Test of Language Development, Intermediate* (PRO-ED, Austin, TX), 3rd Ed.
- Hammill, D. D., Brown, V. L., Larsen, S. C. & Wiederholt, J. L. (1994) *Test of Adolescent and Adult Language* (PRO-ED, Austin, TX), 3rd Ed.
- Efron, B. & Tibshirani, R. J. (1998) *An Introduction to the Bootstrap* (Chapman & Hall, London).
- Sharma, A., Dorman, M. & Spahr, T. (2002) *Ear Hear.* **23**, 532–539.
- Ponton, C. W. & Eggermont, J. J. (2001) *Audiol. Neurootol.* **6**, 363–380.
- Tyler, R. S., Teagle, H. F., Kelsay, D. M., Gantz, B. J., Woodworth, G. G. & Parkinson, A. J. (2000) *Ann. Otol. Rhinol. Laryngol. Suppl.* **185**, 82–84.
- Kirk, K. I., Miyamoto, R. T., Lento, C. L., Ying, E., O’Neill, T. & Fears, B. (2002) *Ann. Otol. Rhinol. Laryngol. Suppl.* **189**, 69–73.