Cognitive Development 42 (2017) 37–48

Contents lists available at ScienceDirect

Cognitive Development

The organization and reorganization of audiovisual speech perception in the first year of life

D. Kyle Danielson[a,*], Alison G. Bruderer[b,a], Padmapriya Kandhadai[a], Eric Vatikiotis-Bateson[c], Janet F. Werker[a]

[a] Department of Psychology, The University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada
[b] School of Audiology and Speech Sciences, The University of British Columbia, 2177 Wesbrook Mall, Vancouver, BC V6T 1Z3, Canada
[c] Department of Linguistics, The University of British Columbia, 2613 West Mall, Vancouver, BC V6T 1Z4, Canada

ARTICLE INFO

Article history:
Received 5 May 2016
Received in revised form 16 January 2017
Accepted 8 February 2017
Available online 6 March 2017

Keywords:
Multisensory perception
Speech perception development
Sensitive periods
Eye tracking

ABSTRACT

The period between six and 12 months is a sensitive period for language learning during which infants undergo auditory perceptual attunement, and recent results indicate that this sensitive period may exist across sensory modalities. We tested infants at three stages of perceptual attunement (six, nine, and 11 months) to determine (1) whether they were sensitive to the congruence between heard and seen speech stimuli in an unfamiliar language, and (2) whether familiarization with congruent audiovisual speech could boost subsequent non-native auditory discrimination. Infants at six- and nine-, but not 11- months, detected audiovisual congruence of non-native syllables. Familiarization to incongruent, but not congruent, audiovisual speech changed auditory discrimination at test for six-month-olds but not nine- or 11-month-olds. These results advance the proposal that speech perception is audiovisual from early in ontogeny, and that the sensitive period for audiovisual speech perception may last somewhat longer than that for auditory perception alone.

1. Introduction

1.1. Background

When we talk with one another, we not only hear each other's voice, but typically also see one another speaking. It has long been known that both heard and seen speech are important in adult speech processing, where information in either modality has the potential to modify perception in the other modality (e.g., McGurk & MacDonald, 1976; Sumby & Pollack, 1954). While an increasing number of studies have explored the extent to which speech perception is multisensory in young infants, there are still many unanswered questions. First and foremost is the question of whether information in one modality can affect processing in the other, and how this might influence and/or be influenced by the timing of the sensitive period for perceptual attunement, the process in the first year of life by which infants' discrimination of non-native speech contrasts declines and their discrimination of native speech contrasts improves (see Werker & Gervain, 2013 for a review). The current set of studies addresses these issues.

* Corresponding author.
E-mail address: kdanielson@psych.ubc.ca (D.K. Danielson).

http://dx.doi.org/10.1016/j.cogdev.2017.02.004
0885-2014/

We asked two specific questions. First, using sounds with which infants were unfamiliar, we tested whether or not young infants are sensitive to the congruence between the auditory and visual information in the speech signal and, if so, whether their sensitivity is independent of experience with specific sound-sight pairings from the native language. We explored the possibility that such sensitivity, if revealed, might decline in tandem with perceptual attunement. Second, we asked whether experimental exposure to congruent versus incongruent audiovisual speech can alter subsequent auditory-only speech perception, and possibly reveal sensitivity to non-native auditory distinctions beyond the age at which infants typically discriminate non-native sounds.

1.2. Perceptual attunement

From a young age, infants auditorily discriminate many of the similar consonant sounds used across the world's languages, regardless of whether such sounds are used to contrast meaning between two words (phonemically) in the language(s) that the child hears. For example, at six to eight months of age, both English- and Hindi-learning infants discriminate between the voiced dental and retroflex consonants of Hindi ([d̪] and [ɖ], respectively), though no such phonemic distinction exists in English, and English-speaking adults exhibit no such discrimination (Werker & Tees, 1984; Werker & Lalonde, 1988; Werker, Gilbert, Humphrey, & Tees, 1981). However, by the time they are nine months old, English-learning infants exhibit reduced discrimination of non-native consonantal phonemic distinctions. By 11 months, auditory discrimination of many non-native consonantal phonemes has declined even further, while discrimination of native phonemes has improved (Kuhl, Tsao, Liu, Zhang, & Boer, 2006; Narayan, Werker, & Beddor, 2010).

This pattern of decline in sensitivity to non-native consonant contrasts and improvement in sensitivity to native contrasts across the first year of life is called *perceptual attunement*. Similar findings have emerged for discrimination of tone distinctions (Mattock & Burnham, 2006; Yeung, Chen, & Werker, 2013), and even for the discrimination of handshape distinctions in visual-only sign language (Palmer, Fais, Golinkoff, & Werker, 2012) and for discrimination of articulatory configurations in silent visual-only speech (Sebastian-Galles, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007). The same pattern is seen for perception of vowel distinctions, but may develop earlier than for consonants (Polka & Werker, 1994). The consistency in the timing of this pattern of change, particularly for perception of consonant contrasts, suggests a critical or sensitive period in development between six and 12 months of age, during which the speech input plays an especially important role in changing perceptual sensitivities (Doupe & Kuhl, 1999; Kuhl, 2010; Friederici & Wartenburger, 2010; Maurer & Werker, 2013; Werker & Tees, 2005; Werker & Hensch, 2015).

1.3. Audiovisual speech perception

Although the bulk of research in speech perception—and in perceptual attunement—has been conducted by investigating the role of individual modalities, the audiovisual nature of speech perception has nevertheless been well attested in adults. A commonly observed piece of evidence in support of a multisensory view of speech perception is adults' robust ability to speechread: to use visual information from an interlocutor's eyes and mouth to aid in perceiving speech in noise (Grant & Seitz, 2000; Macleod & Summerfield, 1987; Sumby & Pollack, 1954; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998). Even more evidence comes from the imposition of *incongruent* visual information onto the auditory speech signal. Under certain conditions, when adult listeners are presented with simultaneous auditory and visual signals that conflict with each other (e.g., a visual/ba/and an auditory/ga/), an entirely different illusory percept arises (adults report perceiving/da/), a phenomenon known as the McGurk effect (McGurk & MacDonald, 1976; Massaro, Cohen, & Smeele, 1996; Rosenblum & Saldaña, 1992; *inter alia*).

A growing body of work suggests that speech perception is audiovisual for the infant as well. Infants exhibit the same McGurk effect that adults do (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997), although perhaps less strongly (Desjardins & Werker, 2004). Like adults, infants' auditory perception of speech in noise is improved when visual information is added (Hollich, Newman, & Jusczyk, 2005).

Much research on audiovisual processing of speech in infancy has involved cross-modal matching. When shown first a video display of two side-by-side identical faces, one articulating one syllable and the other articulating a different syllable, and then are shown the same video display accompanied by the sound for one of the syllables, infants as young as two months of age look longer to the side articulating the syllable that matches the sound that they hear (Kuhl & Meltzoff, 1982, 1984; Mackain, Studdert-Kennedy, Spieker, & Stern, 1983; Patterson & Werker, 1999, 2002, 2003). This evidence indicates that infants' perception of heard and seen speech is audiovisual from early in life. Moreover, in the first six months of life, infants match audiovisual speech combinations from languages with sounds that are unfamiliar to them (Kubicek et al., 2014; Pons, Lewkowicz, Soto-Faraco, & Sebastian-Galles, 2009; Walton & Bower, 1993), and even with pairs of non-human animal faces and their vocalizations (Lewkowicz & Ghazanfar, 2006; Lewkowicz, Leo, & Simion, 2010; Vouloumanos, Druhen, Hauser, & Huizink, 2009;).

Just as the perception of auditory speech attunes in the infant's first year to just those distinctions used in the native language, so too does the matching of the auditory and visual signal. By 11 months of age, infants no longer match heard and seen speech if the stimuli are from a non-native language. For example, six-month-old Spanish-learning infants look longer at a face articulating /ba/ (than a face articulating /va/), when hearing the sound /ba/, and longer at the face articulating /va/