Journal of Experimental Child Psychology 129 (2015) 157–164

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Brief Report

Phonetic matching of auditory and visual speech develops during childhood: Evidence from sine-wave speech

Martijn Baart [a,*], Heather Bortfeld [b,c], Jean Vroomen [d]

[a] BCBL. Basque Center on Cognition, Brain, and Language, 20009 Donostia (San Sebastián), Spain
[b] Department of Psychology, University of Connecticut, Storrs, CT 06269, USA
[c] Haskins Laboratories, New Haven, CT 06511, USA
[d] Department of Cognitive Neuropsychology, Tilburg University, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 16 May 2014
Revised 12 August 2014
Available online 23 September 2014

Keywords:
Audiovisual speech
Cross-modal correspondence
Phonetic cues
Temporal cues
Sine-wave speech
Development

ABSTRACT

The correspondence between auditory speech and lip-read information can be detected based on a combination of temporal and phonetic cross-modal cues. Here, we determined the point in developmental time at which children start to effectively use phonetic information to match a speech sound with one of two articulating faces. We presented 4- to 11-year-olds ($N = 77$) with three-syllabic sine-wave speech replicas of two pseudo-words that were perceived as non-speech and asked them to match the sounds with the corresponding lip-read video. At first, children had no phonetic knowledge about the sounds, and matching was thus based on the temporal cues that are fully retained in sine-wave speech. Next, we trained all children to perceive the phonetic identity of the sine-wave speech and repeated the audiovisual (AV) matching task. Only at around 6.5 years of age did the benefit of having phonetic knowledge about the stimuli become apparent, thereby indicating that AV matching based on phonetic cues presumably develops more slowly than AV matching based on temporal cues.

* Corresponding author.
E-mail address: m.baart@bcbl.eu (M. Baart).

http://dx.doi.org/10.1016/j.jecp.2014.08.002
0022-0965/

Introduction

Although human infants are sensitive to audiovisual (AV) phonetic congruence in speech (e.g., Burnham & Dodd, 1996; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003), the ability to extract phonetic content from visual speech improves dramatically during childhood and into puberty (e.g., Desjardins, Rogers, & Werker, 1997; Erdener & Burnham, 2013; Hockley & Polka, 1994; Kushnerenko, Teinonen, Volein, & Csibra, 2008; Massaro, 1984; McGurk & MacDonald, 1976; Ross et al., 2011; Sekiyama & Burnham, 2008). Although this may possibly be explained by a U-shaped trajectory of AV speech development (see, e.g., Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014, for a similar argument), infants' use of phonetic information is not mandatory (Desjardins & Werker, 2004).

Recently, Baart, Vroomen, Shaw, and Bortfeld (2014) argued that infants might not need phonetic information to detect correspondence in AV speech whenever salient non-phonetic cues are available. They compared adults and infants on AV matching of three-syllable strings with one of two simultaneously delivered lip-read videos. The speech sounds were either natural speech or artificial sine-wave speech (Remez, Rubin, Pisoni, & Carrell, 1981). Critically, the temporal dynamics of natural speech are retained in sine-wave speech, and this information was thus available to all listeners. AV correspondence detection was 25% higher for adults who heard natural speech than for those who heard sine-wave speech, which shows that phonetic knowledge was beneficial to them. However, adults who heard sine-wave speech did match the sound with the lip-read information significantly above chance, presumably because they detected the temporal AV correspondence. In contrast, infants did not seem to benefit from the phonetic information given that their above-chance performance was alike for natural speech and sine-wave speech, which led to the conclusion that infants had presumably relied only on the temporal AV cues. If so, it is conceivable that children would also be able to rely on temporal cues because sensitivity to AV synchrony increases during development (e.g., Grant, van Wassenhove, & Poeppel, 2004; Lewkowicz, 2010). In the same vein, van Linden and Vroomen (2008) showed that whereas 8-year-olds learn to categorize ambiguous speech based on previously seen lip-read information, 5-year-olds do not. This supports the notion that somewhere in between 5 and 8 years of age, phonetic information in the AV speech signal becomes beneficial.

Here, we directly assessed this hypothesis by testing 4- to 11-year-olds on their ability to match a sine-wave speech token with one of two simultaneously presented lip-read speech videos. The elegance of sine-wave speech is that listeners can be tested in a perceptual non-speech mode and/or a perceptual speech mode. In the first mode, listeners do not have access to the phonetic auditory content; in the second, they do. Once listeners are in speech mode, they cannot switch back to the non-speech mode. Therefore, a within-participant design requires the speech mode test to be preceded by the non-speech mode test (see, e.g., Tuomainen, Andersen, Tiippana, & Sams, 2005). Thus, we first established children's AV matching capacity while participants were in non-speech mode, assuming that they could rely only on temporal cues to detect AV correspondence. The critical manipulation consisted of subsequent training in which children were informed about the speech-like nature of the sine-wave tokens so that they perceived the phonetic identity of the sounds (children were now in speech mode, which presumably affects AV integration based on phoneme-to-viseme mapping), after which we again measured AV matching. The difference in performance on each task (the "speech mode effect") was interpreted as a perceptual benefit of phonetic information in detecting AV speech correspondence. In keeping with the literature (e.g., van Linden & Vroomen, 2008), we expected this benefit to become apparent between 5 and 8 years of age and to further increase with age.

Method

Participants

A total of 77 Dutch children between 4 and 11 years of age with normal hearing and normal or corrected-to-normal vision participated in the experiment. Children were divided into three groups according to elementary school grade. In the youngest group ($n = 23$), the ages ranged between 4.2