Journal of Experimental Child Psychology 126 (2014) 295–312

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Children use visual speech to compensate for non-intact auditory speech

Susan Jerger [a,*], Markus F. Damian [b], Nancy Tye-Murray [c], Hervé Abdi [a]

[a] School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX 75080, USA
[b] School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK
[c] Department of Otolaryngology, Washington University School of Medicine, St. Louis, MO 63110, USA

ARTICLE INFO

Article history:
Received 5 September 2013
Revised 14 May 2014

Keywords:
McGurk effect
Speechreading
Children
Audiovisual speech perception
Lipreading
Development

ABSTRACT

We investigated whether visual speech fills in non-intact auditory speech (excised consonant onsets) in typically developing children from 4 to 14 years of age. Stimuli with the excised auditory onsets were presented in the audiovisual (AV) and auditory-only (AO) modes. A visual speech fill-in effect occurs when listeners experience hearing the same non-intact auditory stimulus (e.g., /-b/ag) as different depending on the presence/absence of visual speech such as hearing /bag/ in the AV mode but hearing /ag/ in the AO mode. We quantified the visual speech fill-in effect by the difference in the number of correct consonant onset responses between the modes. We found that easy visual speech cues /b/ provided greater filling in than difficult cues /g/. Only older children benefited from difficult visual speech cues, whereas all children benefited from easy visual speech cues, although 4- and 5-year-olds did not benefit as much as older children. To explore task demands, we compared results on our new task with those on the McGurk task. The influence of visual speech was uniquely associated with age and vocabulary abilities for the visual speech fill-in effect but was uniquely associated with speechreading skills for the McGurk effect. This dissociation implies that visual speech—as processed by children—is a complicated and multifaceted phenomenon underpinned by heterogeneous abilities. These results emphasize that children perceive a speaker's utterance rather than the auditory stimulus per se. In children, as in adults, there is more to speech perception than meets the ear.

* Corresponding author. Fax: +1 972 883 2491.
E-mail address: sjerger@utdallas.edu (S. Jerger).

http://dx.doi.org/10.1016/j.jecp.2014.05.003
0022-0965/

Introduction

When adults engage in casual conversations in noisy environments, they typically understand each other without effort. Such skilled understanding in degraded soundscapes seems related to the inherently multimodal nature of speech perception, as dramatically illustrated by the classic McGurk effect (McGurk & MacDonald, 1976). In this task, an audiovisual speech stimulus with mismatched auditory and visual onsets (e.g., hearing /ba/ while seeing /ga/) is presented to participants. Adults typically perceive a mixture of the auditory and visual inputs (i.e., /da/ or /ða/). Our ability to integrate auditory and visual speech helps us to understand speech in noisy environments as well as unfamiliar content in clear environments (Arnold & Hill, 2001; MacLeod & Summerfield, 1987). Because visual speech is so useful to communication when the message is complex/degraded or the environment is noisy (e.g., classrooms), it is paramount to investigate the development of multimodal speech perception during the preschool and elementary school years.

Development of multimodal speech perception

Extant studies with multiple types of tasks report that, compared with adults, children from around 5 years of age to the preteen/teenage years show reduced sensitivity to visual speech (e.g., Desjardins, Rogers, & Werker, 1997; Erdener & Burnham, 2013; Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009; McGurk & MacDonald, 1976; Ross et al., 2011; Sekiyama & Burnham, 2008; Tremblay et al., 2007; see Fort, Spinelli, Savariaux, & Kandel, 2010, for an exception regarding vowel monitoring). For example, McGurk and MacDonald (1976) noted that children were influenced by visual speech only roughly half as often as adults. Poorer sensitivity to visual speech in children has been attributed to developmental differences in articulatory proficiency or speechreading skills (Desjardins et al., 1997; Erdener & Burnham, 2013), linguistic experiences and perceptual tuning into language-specific phonemes (Erdener & Burnham, 2013; Sekiyama & Burnham, 2008), or differences in the perceptual weighting of visual speech cues (Green, 1998).

Supplementing these more specific theories, Jerger et al. (2009), who observed a U-shaped developmental function with a significant influence of congruent visual speech on phonological priming in 4- and 12-year-olds but not in 5- to 9-year-olds, adopted a dynamic systems viewpoint (Smith & Thelen, 2003). Jerger and colleagues hypothesized that children's poorer sensitivity to visual speech from 5 to 9 years of age was reflecting a period of dynamic growth as relevant perceptual, linguistic, and cognitive skills were reorganizing in response to external and internal factors. Externally, reorganization may be due to literacy instruction at around 5 or 6 years of age, during which time knowledge transmutes from phonemes as coarticulated nondistinct parts of speech into phonemes as separable distinct written and heard elements (Bryant, 1995; Burnham, 2003; Horlyck, Reid, & Burnham, 2012). Internally, reorganization may be due to phonological processes becoming sufficiently proficient (at around the same ages) to support the use of inner speech for learning, remembering, and problem solving (Conrad, 1971). The actuality of reorganization is confirmed by evoked potential studies indicating—during this age period—developmental restructuring of the lexical phonological system (Bonte & Blomert, 2004).

A dynamic systems viewpoint (Smith & Thelen, 2003) stresses two points motivating this research. First, periods of poorer sensitivity to visual speech may reflect a transition period—in contrast to a loss of ability—during which time relevant perceptual, linguistic, and cognitive resources are harder to access. Second, dynamic periods of reorganization and growth are characterized by less robust processing systems and decreases in processing efficiencies. So according to a dynamic system viewpoint, the influence of visual speech may vary as a function of task/stimulus demands for these softly reassembled resources. This suggests that tasks with less complex stimuli and/or lower task demands might not tax a child's limited processing resources as readily. This would make the harder to access resources more accessible; therefore, the child's performance might reveal greater sensitivity to visual speech. Below we introduce our new task to set up reviewing the related literature.