



Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech



Ferran Pons ^{a,b,*}, David J. Lewkowicz ^c

^a Departament de Psicologia Bàsica, Universitat de Barcelona, Pg. Vall d'Hebron 171, 08035, Barcelona, Spain

^b Institute for Brain, Cognition and Behavior (IR3C), Barcelona, Spain

^c Department of Psychology & Center for Complex Systems & Brain Science, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA

ARTICLE INFO

Article history:

Received 25 July 2013

Received in revised form 23 December 2013

Accepted 27 December 2013

Available online 25 February 2014

Keywords:

Infancy

Speech perception

Audiovisual perception

ABSTRACT

We investigated the effects of linguistic experience and language familiarity on the perception of audio-visual (A-V) synchrony in fluent speech. In Experiment 1, we tested a group of monolingual Spanish- and Catalan-learning 8-month-old infants to a video clip of a person speaking Spanish. Following habituation to the audiovisually synchronous video, infants saw and heard desynchronized clips of the same video where the audio stream now preceded the video stream by 366, 500, or 666 ms. In Experiment 2, monolingual Catalan and Spanish infants were tested with a video clip of a person speaking English. Results indicated that in both experiments, infants detected a 666 and a 500 ms asynchrony. That is, their responsiveness to A-V synchrony was the same regardless of their specific linguistic experience or familiarity with the tested language. Compared to previous results from infant studies with isolated audiovisual syllables, these results show that infants are more sensitive to A-V temporal relations inherent in fluent speech. Furthermore, the absence of a language familiarity effect on the detection of A-V speech asynchrony at eight months of age is consistent with the broad perceptual tuning usually observed in infant response to linguistic input at this age.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Our perceptual experiences are typically multisensory and, as a result, we usually have access to concurrent and often highly redundant sensory inputs in different modalities. In general, the multisensory redundancy of speech as well as all other multisensory events is known to facilitate perception, learning, and discrimination (Bahrick, Lickliter, & Flom, 2004; Lewkowicz & Kraebel, 2004; Partan & Marler, 1999; Stein & Stanford, 2008). Critically, multisensory facilitation depends on our ability to perceive the unity of audible and visible perceptual attributes and this, in turn, depends on our ability to perceive the temporal synchrony of these attributes. Fortunately, our perceptual system can tolerate a fair degree of temporal desynchronization of the audible and visible attributes of multisensory events. That is, we can perceive audiovisual events as perceptually coherent even if they are desynchronized to some degree. This is especially the case for audiovisual speech (Dixon & Spitz, 1980; Grant, van Wassenhove, & Poeppel, 2004) where there is a relatively large temporal window during which physically desynchronized auditory and visual speech attributes are perceived as part of a coherent speech event. Interestingly, this window is much larger in infancy (Lewkowicz, 2000, 2003, 2010), indicating that infants tolerate a much larger audio-visual (A-V) asynchrony than adults do.

To date, only one study has asked whether language-specific experience might affect responsiveness to A-V synchrony relations and this study investigated this question in adults and found that it does (Navarra, Alsius, Velasco, Soto-Faraco, & Spence, 2010). Here, we asked whether the effects of language-specific experience on the detection of A-V asynchrony might already be evident in infancy and prior to the emergence of language-specific expertise which typically occurs by the end of the first year of life.

1.1. Responsiveness to A-V synchrony

Adults are highly sensitive to A-V synchrony. They can detect an A-V asynchrony of as little as 180–240 ms when the visual attributes of an audiovisual event precede the auditory attributes, and as little as 60–200 ms when the auditory attributes precede the visual attributes (Dixon & Spitz, 1980; Grant & Greenberg, 2001; Grant et al., 2004; Miner & Caudell, 1998; Navarra et al., 2005; van Wassenhove, Grant, & Poeppel, 2007). Importantly, studies with adults have found that perception of A-V temporal relations is affected by the specific nature of the stimulus presented (i.e., speech vs. non-speech) and the complexity of the stimuli (Vatakis & Spence, 2010). For example, Vatakis and Spence (2006) found that adults exhibit greater sensitivity to temporal order when responding to shorter and less complex stimuli (e.g., syllables) than to longer and/or more complex stimuli (e.g., sentences). A number of studies also have found short-term adaptation effects when adults are required to detect A-V synchrony relations. Specifically, when

* Corresponding author at: Departament de Psicologia Bàsica, Universitat de Barcelona, Pg. Vall d'Hebron 171, 08035, Barcelona, Spain.

adults are first tested with audiovisually asynchronous events, they perceive them as such but after short-term exposure to such events they begin to respond to them as if they are synchronous (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Navarra et al., 2005; Vroomen, Keetels, de Gelder, & Bertelson, 2004).

From a developmental standpoint, one especially interesting finding is that adults' sensitivity to the temporal synchrony of audible and visible speech depends on their specific language experience (Navarra et al., 2010). To understand this effect it is important to note that adults are not only sensitive to A-V speech asynchrony but that they also expect visual speech to precede auditory speech by some minimum amount of time to compensate for the slower neural transmission time of visual as opposed to auditory signals. Given that, Navarra et al. (2010) found that in order for adults to perceive A-V speech simultaneity, the visual speech stream had to lead the auditory speech stream by a significantly larger interval in the participants' native language than in their non-native language. Of particular interest, Navarra et al. also found that this difference tended to diminish as the amount of experience with the non-native language increased.

1.2. Infant responsiveness to A-V synchrony

Studies with infants also have found that experience affects responsiveness to A-V synchrony but, importantly, the effects are opposite to those reported in adults. Specifically, Lewkowicz (2010) found that after infants are habituated to a syllable whose auditory and visible attributes are discriminably asynchronous, they subsequently exhibit better detection of A-V asynchrony than do infants who are habituated to a temporally synchronous syllable (Lewkowicz, 2000, 2003). Pons, Teixidó, Garcia-Morera, and Navarra (2012) have replicated this finding in a study with non-speech stimuli. Thus, the size of the A-V temporal binding window decreases following familiarization with an asynchronous event whereas in adults the window increases. Regardless of the developmental differences in the direction of adaptation effects, the adult and infant studies show that detection of A-V temporal relations is affected by prior experience. This conclusion is based on the assumption that infants pay more attention to audiovisual temporal relations simply because they have had less opportunity to acquire a perceptual bias for multisensory unity which, under normal ecological conditions, is the default state in our multisensory world (Lewkowicz, 2010).

1.3. Effects of linguistic experience in infancy

Experience plays a key role in infant response to auditory, visual, and audiovisual speech. The most direct evidence of this comes from studies of perceptual narrowing in infant response to speech and other communicative signals. This evidence shows that there is a critical difference between younger and older infants in that younger infants exhibit broad perceptual tuning for such inputs whereas older infants exhibit narrower tuning. Specifically, it has been found that younger infants can perceive native and non-native auditory (Werker & Tees, 1984) and visual (Weikum et al., 2007) speech attributes, the audiovisual coherence of native and non-native phonemes (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009), and the coherence of audible and visible vocalizations of another species (Lewkowicz & Ghazanfar, 2006). In contrast, older infants no longer respond to non-native auditory and visual speech attributes nor to the audiovisual coherence of non-native phonemes and vocalizations. Studies in which infants have been provided with "extra" experience with non-native stimuli during the narrowing period have found that infants maintain their responsiveness to such stimuli (Hannon & Trehub, 2005; Pascalis et al., 2005). This is direct evidence that specific early experience is responsible for narrowing.

The narrowing effects of early experience also can be seen in developmental changes in infant selective attention to audiovisual speech. For example, when infants begin to babble around six months of age, they gradually become interested in speech production as well as in speech

perception. As this happens, infants begin to shift their attention to the mouth of a talker starting at six months and by eight and ten months of age they spend significantly more time looking at a talker's mouth than eyes (Lewkowicz & Hansen-Tift, 2012). This attentional shift enables infants to gain direct access to the synchronous, redundant and, thus, highly salient audiovisual speech cues which, in turn, enables them to gradually acquire native-language expertise. During this time, infants look longer at the talker's mouth regardless of whether she is speaking in their native language or in a foreign language. By 12 months of age, however, when perceptual narrowing has completed and when infants have acquired their initial native-language expertise, they no longer attend more to the talker's mouth when she is speaking in their native language presumably because they no longer require access to redundant audiovisual speech cues. If, however, the talker is seen and heard speaking in a non-native language, 12-month-old infants look longer at the mouth presumably because they require the more salient redundant audiovisual cues to disambiguate what has now become an unfamiliar language. Overall, these findings indicate that infants' response to audiovisual speech at different points during the first year of life depends on specific linguistic experience at each age. Thus, infants respond the same way to different languages before narrowing has completed but differently once their perceptual tuning has narrowed to their native language.

1.4. The present research

Currently, it is not known whether early experience plays a role in infant response to the temporal coherence of fluent audiovisual speech. Dodd (1979) reported that 2.5 to 4-month-old infants attend less to fluent desynchronized speech than to fluent synchronized speech. This finding indicates that young infants can perceive audiovisual speech synchrony in fluent speech but it does not provide any information on whether linguistic experience affects such responsiveness. As already indicated, evidence from studies with adults suggests that language-specific experience influences the perception of temporal A-V speech relations (Navarra et al., 2010). When this is combined with the fact that experience plays a key role in the development of speech and language in infancy, it raises questions regarding the effects of experience on responsiveness to A-V fluent speech synchrony in infancy.

We hypothesized that the effects of experience on the detection of A-V synchrony relations are likely to depend on the degree of exposure that infants have accumulated with their native language at the time of test. That is, at an age when infants have not yet become experts in their native language (i.e., prior to the completion of perceptual narrowing), they should respond to A-V synchrony relations inherent in fluent speech regardless of their linguistic experience. To test our prediction, we tested 8-month-old infants' response to fluent native speech, fluent non-native but familiar speech, and fluent non-native, unfamiliar speech. Specifically, we tested monolingual Spanish-learning and monolingual Catalan-learning infants' response to audiovisual monologues spoken either in Spanish or in English. Because Spanish is the native language for Spanish-learning infants and a nonnative but familiar language for Catalan-learning infants, the Spanish monologue enabled us to determine whether experience with the native language versus experience with a different but familiar and similar language affects responsiveness to A-V synchrony relations. Because English is unfamiliar to both Spanish- and Catalan-learning infants, it enabled us to go one step further and determine whether responsiveness to A-V asynchrony in a completely unfamiliar language differs from that in a familiar language.

Prior studies in infants have shown that the A-V asynchrony detection threshold for audiovisual speech is 666 ms when the auditory speech leads visual speech (Lewkowicz, 2010) but those findings are based on responsiveness to isolated syllables. Thus, here we not only investigated whether A-V asynchrony detection is affected by linguistic experience but also whether it might differ for fluent speech. Like in the Lewkowicz (2010) study, we habituated infants to an audiovisually synchronous stimulus first and then tested their response to different degrees of A-V

Download English Version:

<https://daneshyari.com/en/article/7277706>

Download Persian Version:

<https://daneshyari.com/article/7277706>

[Daneshyari.com](https://daneshyari.com)