Acta Psychologica 149 (2014) 142–147

Contents lists available at ScienceDirect

Acta Psychologica

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

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 audiovisual (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.

0001-6918/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.12.013

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

asynchrony produced by presenting the auditory component earlier in time than the visual component. In contrast to the Lewkowicz (2010) study, however, here we presented fluent audiovisual speech rather than isolated audiovisual syllables. To determine whether early linguistic experience affects responsiveness to A-V asynchrony, in Experiment 1 we presented fluent speech utterances spoken in Spanish to Spanish-learning infants and to Catalan-learning infants. This enabled us to determine whether responsiveness depends on whether the fluent speech is in the infants' familiar language or in a non-native but familiar language. Because Spanish and Catalan are similar languages, in Experiment 2 we tested Spanish- and Catalan-learning infants' response to utterances spoken in English, a non-native and completely unfamiliar language to these infants.

2. Experiment 1

2.1. Method

2.1.1. Participants

Forty-eight 8-month-old infants were tested. Twenty-four of these infants were from Spanish-learning homes (M = 242 days, range = 230–256 days; 11 girls) and 24 were from Catalan-learning homes (M = 239 days, range = 228–252 days; 12 girls). All infants were full-term, healthy, and had no history of ear infections according to the parents' report. Fifteen additional infants participated in the study but were excluded from the analysis due to crying or fussiness (n = 5; 4 Spanish-learning infants, 1 Catalan-learning infant), experimental error (n = 1; 1 Spanish-learning infant), or failure to habituate (n = 9; 5 Spanish-learning infants, 4 Catalan-learning infants). Infants were recruited by visiting new mothers at the Hospital Sant Joan de Déu (Barcelona) and soliciting their participation in the study. Although both Catalan and Spanish are spoken in Barcelona (the vast majority of the population is bilingual, both languages are co-official and are taught in school), the infants in this study were from monolingual families. A detailed language questionnaire (Bosch & Sebastián-Gallés, 2001) was administered and only infants with less than 15% of direct exposure to the other language were included in the sample.

2.1.2. Apparatus and stimuli

The stimuli consisted of multimedia movies which were constructed with Premiere 6.0 (Adobe Corporation). The movies were video clips of a female speaker looking directly at the camera uttering a script in Spanish and speaking in an infant-directed manner (see Appendix A). Infant-directed speech was characterized by using a prosodically exaggerated manner, slow tempo, high pitch excursions, and continuous smiling. The original video used to make the movies was compressed with the Cinepak Codec and the movie was presented at 30 frames/s at a resolution of 1024 × 480 pixels (see Fig. 1). The audio part of the movie was sampled at 1024 kbps. Four movies were created. One of these presented the audio and video streams in synchrony with one another while the other three movies presented these two streams at different degrees of asynchrony. In these latter movies, the auditory stream was moved ahead of the visual one by 11 video frames creating a novel test trial with an asynchrony of 366 ms (NOV 366), by 15 frames resulting in a novel test trial with an asynchrony of 500 ms (NOV 500), and by 20 frames resulting in a novel test trial with an asynchrony of 666 ms (NOV 666), respectively.

Infants were seated in an infant seat during the experiment. The parent was present in the room, sitting silently behind the infant. Testing took place in a dimly lit, sound-attenuated laboratory room, with a television monitor situated 130 cm from the infant. The movies were presented on an LG 50″ television monitor. The head of the female seen and heard speaking was approximately the same size as that of a live person speaking to the infant. The experiment was controlled by the experimenter from an adjacent room using the Habit 2002 software (Cohen, Atkinson, & Chaput, 2000) running on a Power Mac G5. The experimenter, who was unaware of trial status, controlled movie presentation and recorded infant visual fixation of the movie by pressing a key on the computer keyboard while watching an image of the infant's face on a video monitor (Panasonic BT-S1460Y TV). This monitor transmitted the image of the infant's face from a Canon MV750i video camera mounted under the television monitor located in the booth.

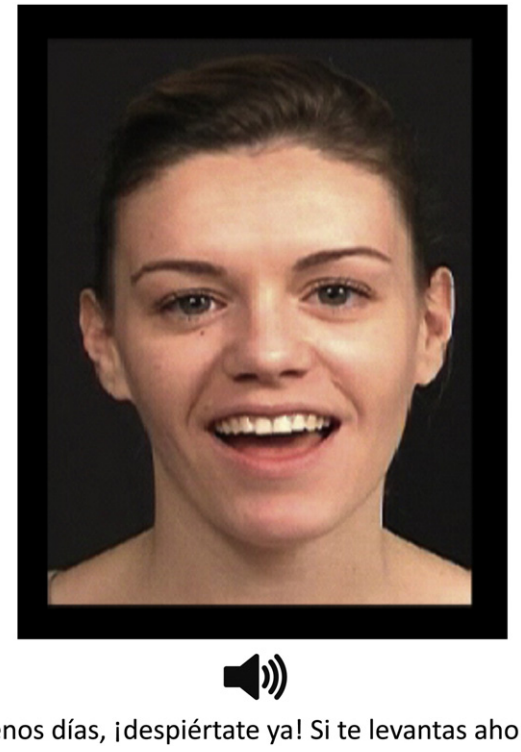

Fig. 1. Example of a single video frame of the clip of the speaker uttering the monologue.

2.1.3. Procedure

We used the habituation/test procedure to test for detection of A-V asynchrony. The experiment began with a pretest test trial during which we presented a spinning waterwheel together with a sine wave tone in order to assess infants' initial engagement in the task. Once this trial ended, the habituation phase began. The experimenter began each trial whenever infants looked at an attention-getter (a blue expanding flower presented in the middle of the monitor). As soon as the infant looked at the attention-getter, it was turned off and the movie of the woman speaking was presented. Here, the audible and visible speech streams were in perfect synchrony. Each habituation trial lasted a

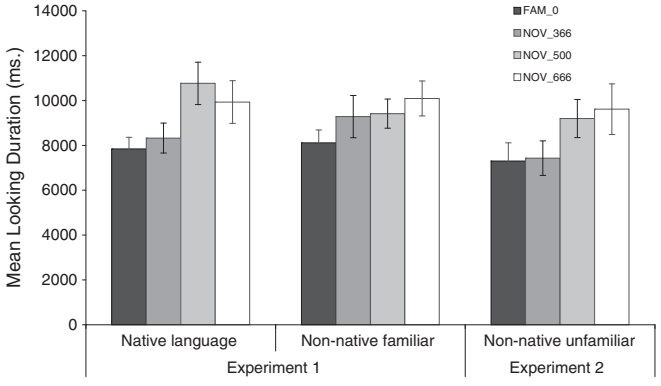

Fig. 2. Mean looking times to test trials by 8-month-old Catalan- and Spanish-learning infants tested in native, non-native but familiar, and non-native and unfamiliar language. Error bars represent standard error.

maximum of 15 s. The habituation criterion was set such that infant looking had to decline during a three-trial block to 60% of the total looking time during the longest block of three trials. When infants reached this criterion, the habituation phase ended and the test phase began. If an infant did not reach the 60% criterion within 24 habituation trials, the habituation phase was ended and that particular infant's data were excluded. There were four test trials: a synchronous test trial, dubbed the Familiar trial (i.e., FAM 0 where the 0 refers to the degree of temporal asynchrony) and the three asynchronous trials (i.e., NOV 366, NOV 500, and NOV 666). These trials were presented in counterbalanced order across infants according to a Latin Square design. This resulted in four different test-trial-order groups. Finally, after the last test trial was presented, we administered a posttest trial during which infants once again saw the spinning waterwheel and heard the tone. The duration of looking during this trial helped us to determine whether infants became fatigued during the course of the experiment and whether failure to respond in any of the asynchrony test trials was due to fatigue or to a failure to detect a specific degree of asynchrony. In the latter case, if infants exhibited response recovery in the posttest trial then failure to respond in an asynchrony trial could not be ascribed to general fatigue.

2.2. Results

2.2.1. Preliminary analyses

A preliminary 2 (language: Spanish- vs. Catalan-learning infants) by 2 (test: pretest vs. posttest) mixed analysis of variance (ANOVA), with language as the between-subjects factor and test as the within-subjects factor, was carried out on the data from the pretest and posttest trials to determine whether infants became fatigued across the experiment and whether fatigue might have been related to the language presented. There were no significant effects of language or trial, indicating that both groups of infants maintained their interest during the experimental session and that the particular language that they saw and heard did not affect their responsiveness. A second preliminary analysis of the data from the habituation phase was conducted. The habituation criterion required that infant looking during a block consisting of the last three habituation trials declines to 60% of the total looking time during a previous block of three trials with the longest amount of looking. To determine whether the two groups of infants differed during the habituation phase, we compared their habituation data by way of a 2 (block of trials: first vs. last) × 2 (infants' language: Catalan vs. Spanish) mixed analysis of variance (ANOVA). This analysis yielded a significant main effect for block, $F(1, 46) = 1966.64$, $p < .001$ partial $\eta^2 = .977$, but no other effects. Thus, the two groups of infants did not differ in terms of their rate of habituation.

2.2.2. Perception of asynchrony

As indicated earlier, prior studies have found that infants can detect an A-V speech asynchrony of 666 ms when tested with isolated audiovisual syllables (Lewkowicz, 2000, 2003, 2010). To determine whether the infants in the current experiment detected any of the three different asynchronies, we adopted the analytic approach used by Lewkowicz (2010) and, thus, conducted a set of planned contrast analyses. These analyses compared the duration of looking in the NOV 366, NOV 500, and the NOV 666 test trials, respectively, with the duration of looking in the FAM 0 test trial. To determine whether the infants' native-language background or test-trial order affected responsiveness, first we submitted the data from the four test trials to a 2 × 4 × 4 (language background × test-trial order × test-trial type) mixed, repeated-measures ANOVA, with language background and test-trial order as the between-subjects factors and test-trial type as a within-subjects factor. This analysis yielded no significant main effects or interactions indicating that responsiveness was not affected by the infants' early experience with a specific language nor by the order of the test trials. The planned contrast analyses indicated that infants did not exhibit significant response recovery in the NOV 366 test trial but that they did exhibit significant response recovery in the NOV 500 test trial, $F(1, 46) = 9.48$, $p = .003$, partial $\eta^2 = .171$, and in the NOV 666 test trial, $F(1, 46) = 9.54$, $p = .003$, partial $\eta^2 = .172$. These findings show that infants detected the 500 and 666 ms asynchrony and that they did so regardless of whether the audiovisual speech was familiar or non-familiar to them (See Fig. 2).

2.3. Discussion

The results of the current experiment indicated that 8-month-old infants detected an asynchrony of as low as 500 ms in fluent audiovisual speech and that their ability to do so was not affected by whether the speech was in their native or non-native language. At first blush, it might seem surprising that language background did not affect responsiveness given that language-specific experience influences adults' perception of temporal A-V speech relations (Navarra et al., 2010) and that infants can distinguish auditory-only Spanish from Catalan speech at four months of age (Bosch & Sebastián-Gallés, 1997). Indeed, there are, actually, several reasons why the infants' response was not affected by language background. First, the task in the current study was different from a task requiring infants to discriminate auditory-only Spanish versus Catalan speech. Here, the infants' task was to detect the temporal alignment of audible and visible speech streams, not their identity. Prior studies have shown that newborns can detect the temporal synchrony of faces and vocalizations without relying on identity cues (Lewkowicz, Leo, & Simion, 2010). If newborns can succeed at such a task then 8-month-old infants can surely do so as well, especially if they have not yet achieved sufficient multisensory expertise in their native language (Lewkowicz & Hansen-Tift, 2012; Lewkowicz & Pons, 2013; Pons et al., 2009). Second, the infants in the current study had approximately 15% exposure to the non-native language, making it unlikely that the Catalan infants had sufficient experience with Spanish to distinguish its audiovisual attributes from those of Catalan. Finally, Spanish and Catalan are highly similar languages, making their discrimination even more difficult. In sum, given that infants had to detect the temporal alignment of the audible and visible attributes of two highly similar languages – where one of them was relatively unfamiliar – it is not surprising that responsiveness was not affected by the infants' language background.

3. Experiment 2

The similarity of Spanish and Catalan makes it difficult to test the effects of early linguistic experience on the detection of A-V asynchrony directly. Therefore, in this experiment we tested 8-month-old Spanish- and Catalan-learning infants' response to A-V asynchrony inherent in a fluent-speech utterance spoken in a non-native and unfamiliar language, namely English. We expected the results to be similar to those obtained in Experiment 1. This prediction was based on the fact that perceptual narrowing of responsiveness to auditory and audiovisual speech is not complete until later in infancy (Lewkowicz & Ghazanfar, 2009; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2009; Werker & Tees, 1984). Because perceptual narrowing leads to the emergence of native-language specialization, it is only when narrowing is complete that it would be reasonable to expect evidence of the effects of early linguistic experience on the detection of A-V speech asynchrony. In other words, prior to narrowing, it is reasonable to expect that detection of A-V speech asynchrony will be the same regardless of whether audiovisual speech is native, non-native but similar, or non-native and completely unfamiliar.

3.1. Method

3.1.1. Participants

Twenty-four 8-month-old Spanish- and Catalan-learning infants were tested (M = 246 days, range = 236–260 days; 13 girls; 14 Spanish-learning infants). All infants were full-term, healthy, and had no history of ear infections according to the parents' report. Eight additional infants participated in the study but were excluded from

the analysis due to crying or fussiness (n = 5), or failure to habituate (n = 3).

3.1.2. Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were the same as those in Experiment 1. The only difference was that infants were exposed to a video clip showing a female actor speaking the same utterance that was presented in Experiment 1 except that here she spoke in English.

3.2. Results

3.2.1. Preliminary analyses

We conducted a preliminary one-way ANOVA on the data from the pretest and posttest trials to determine whether infants became fatigued during the experiment and whether fatigue might have been related to the infants' language background. This analysis yielded no significant effects of language background nor trial, indicating that infants maintained their interest during the experimental session and that the specific language that they were exposed to did not affect their responsiveness. Then, we conducted a second preliminary analysis of the data from the habituation phase to determine whether responsiveness differed in the two language groups. This consisted of a 2 (block of trials: first vs. last) × 2 (infants' language background: Catalan vs. Spanish) mixed analysis of variance (ANOVA). This analysis yielded only a significant main effect for block $F(1, 23) = 850.45$, $p < .001$, partial $\eta^2 = .974$, indicating that infants' responsiveness declined during the habituation phase and that this decline was not affected by their language background.

3.2.2. Perception of asynchrony

The data from the four test trials were submitted to a 2 × 4 × 4 (language background × test-trial order × test-trial type) mixed, repeated-measures ANOVA, with language background and test-trial order as the between-subjects factors and test-trial type as a within-subjects factor. This analysis yielded no significant main effects or interactions, indicating neither of the two between-subjects factors affected responsiveness in the test trials. As a result, like in Experiment 1, we conducted planned contrast analyses comparing the duration of looking in each of the asynchrony novel test trials, respectively, with the duration of looking in the FAM 0 test trial. These analyses revealed the same response pattern as previously. That is, infants did not exhibit significant response recovery in the NOV 366 test trial but did in the NOV 500 test trial, $F(1, 23) = 5.91$, $p = .023$, partial $\eta^2 = .204$, and in the NOV 666 test trial, $F(1, 23) = 9.54$, $p = .043$, partial $\eta^2 = .166$.

As predicted, and similar to the results from Experiment 1, the findings from this experiment indicated that infants detected the 500 and 666 ms asynchronies. This time, however, the asynchrony that they detected was inherent in a completely unfamiliar language. This finding shows that perception of A-V asynchrony in fluent audiovisual speech at eight months of age is not affected by earlier experience with a specific language.

4. Discussion

The current study investigated 8-month-old infants' perception of A-V asynchrony inherent in fluent speech and the effects of specific linguistic experience on its perception. Using a habituation/test procedure, we habituated infants to synchronous audiovisual speech first and then administered separate test trials during which the audible speech stream was moved ahead of the visible stream by 366, 500, or 666 ms. Using amount of response recovery as an index of asynchrony detection, we found that infants did not detect an A-V asynchrony of 366 ms but that they did detect an asynchrony of 500 and 666 ms. Moreover, we found that specific early experience with a particular language did not affect infants' responsiveness. This was the case in Experiment 1 where Spanish- and Catalan-learning infants responded similarly when tested with desynchronized Spanish audiovisual speech and in Experiment 2 where infants from both of these groups responded similarly when tested with desynchronized English audiovisual speech.

The findings from the current study are interesting in the context of findings from previous studies of infant response to A-V asynchrony. The previous studies investigated infant response to temporal A-V asynchrony inherent in isolated syllables and found that following habituation to a synchronous audiovisual syllable, infants detected an asynchrony of 666 ms but not asynchronies of 500 and 366 ms (Lewkowicz, 2000, 2003, 2010). Like in those studies, here we also obtained evidence that infants detected an asynchrony of 666 ms. In addition, however, we also obtained evidence that infants detected an asynchrony of 500 ms. This suggests that detection of A-V asynchrony may depend on the nature of the information in the two modalities. The principal difference between an isolated audiovisual syllable and a continuous speech utterance is that in the former case there are only two opportunities for the detection of a desynchronization: at the start and at the end of the syllable. In contrast, in the latter case, there are multiple opportunities for the detection of desynchronization because there are many points along the continuously varying auditory and visual speech streams where the signals in the two modalities are no longer aligned. Moreover, at eight months of age infants have not yet acquired any lexical nor semantic knowledge and, thus, cannot rely on these cues for determining whether the auditory and visual speech streams are coherent. Therefore, they can only rely on the temporal alignment of the dynamic cues inherent in the concurrent but misaligned auditory and visual speech streams. Nonetheless, if they attend to the dynamic cues they have the opportunity to experience multiple instances of desynchronization as the audible and visible streams of speech go by. This interpretation and our conclusion that detection of the A-V asynchrony of isolated syllables is likely to be more difficult are consistent with findings from studies of 4-, 5-, and 6-year-old children's response to A-V asynchrony inherent in isolated syllables. These findings demonstrate that it is not until five years of age that children first exhibit the ability to detect a 500 ms A-V asynchrony inherent in isolated syllables (Lewkowicz & Flom, 2013).

It should be noted that differential stimulus salience cannot explain the difference between the current results and those from Lewkowicz (2010). Although previous studies have found that stimulus salience (e.g., loudness or size of visual display) can affect infant audiovisual perception (i.e. Sekiyama, Kanno, Miura, & Sugita, 2003), this was unlikely to be the case here because we used the same auditory and visual parameters as in Lewkowicz (2010).

The finding that language experience did not affect responsiveness to A-V speech asynchrony at eight months of age is consistent with the effects of perceptual narrowing during infancy. As indicated earlier, infants only exhibit tuning to their native audiovisual speech by 12 months of age (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2009). In other words, the effects of early linguistic experience have not had their full impact prior to this age. As a result, it is not surprising that 8-month-old infants do not respond differently to the A-V synchrony relations of native, familiar non-native, and unfamiliar non-native audiovisual speech.

It is interesting to note that our infant findings are consistent with the results from studies with adults and with our experience-based interpretation. For example, 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, and crucially, that this difference tended to diminish as the amount of experience with the non-native language increased. In other words, in contrast to infants, adults respond differently to A-V temporal synchrony relations as a function of linguistic experience. We interpreted our finding that infants did not exhibit response differences as a function of language background to be a reflection of the fact that they had not yet accumulated sufficient linguistic experience. If this is correct then the adult findings most likely reflect the effects of perceptual narrowing

in infancy as well as the additional experience that leads to further specialization and expertise after infancy. In other words, the additional experience that adults acquire after infancy presumably provides them with a greater sensitivity to temporal A-V synchrony relations in their native language than in a non-native or unfamiliar language.

In conclusion, when our infant findings, the adult findings, and on our experience-based interpretation are considered together, they suggest some interesting avenues for future studies. Specifically, they raise the obvious question of how additional language exposure during infancy might affect the detection of temporal A-V speech synchrony perception. For example, how might 12 month-old infants respond to fluent A-V speech asynchrony given that the process of perceptual narrowing has completed by then and given that by this age infants possess an initial specialization in their native language? An a-priori prediction may be difficult because infants at this age do not attend more to the mouth of a talker speaking in their native language but do when they are exposed to a talker speaking in a non-native speech (Lewkowicz & Hansen-Tift, 2012). Therefore, on the one hand, 12-month-olds' ability to perceive A-V asynchrony based on the temporal relation between vocalizations and lip motions might be counteracted by the fact that they pay less attention to the mouth when they are exposed to native speech. On the other hand, it is possible that 12-month-olds' increased expertise makes it possible for them to extract invariant audiovisual speech information even without having to attend directly to the talker's mouth. Whatever the ultimate answer to the question of older infants' response to audiovisual fluent speech asynchrony might be, there is little doubt that the complexity of fluent speech (e.g., its overall tempo, prosody, and lexical and semantic complexity) is likely to interact with older infants' increased efficiency and expertise for processing audiovisual fluent speech.

Acknowledgments

This work was supported by the Spanish Ministerio de Ciencia e Innovación (PSI2010-20294) to FP, by a grant from the National Science Foundation (BCS-0751888) and grant no. R01HD057116 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to DJL, and by the European COST Action ISCH TD0904 "TMELY: Time in MEntal activity". We thank Maria Teixidó and Helena Moliné for her assistance with data collection.

Appendix A

Experiment 1: Spanish script: *¡Buenos días, despiértate ya! ¡Si te levantas ahora tendremos una hora entera para jugar! Me encantan estas mañanas largas, ¿y a ti? Ojala no se acabaran nunca. Bueno, por lo menos es viernes y tenemos todo el sábado para descansar, excepto por lo de la fiesta. Me vas a ayudar a arreglar la casa, ¿sí? Tenemos que comprar flores, preparar la comida, sacar el polvo, aspirar la casa y limpiar los discos.*

Experiment 2: English script: *Good morning! Get up. Come on now. If you get up right away, we have a whole hour to putter around the house. I love these long mornings, don't you? I wish that they could last all day. Well, at least it's Friday and we can loaf around all day Saturday, except of course, for the party. Are you going to help me fix up the house? We have to buy flowers, prepare the food, vacuum the house, dust everything and clean the records.*

References

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*(3), 99–102.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, *65*(1), 33–69.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29–49.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). *Habit 2000: A new program for testing infant perception and cognition (Version 2.2.5c) [Computer software]*. Austin, TX: University of Texas.

Dixon, N. F., & Spitz, L. T. (1980). The detection of auditory visual desynchrony. *Perception*, *9*, 719–721.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11*(4), 478–484.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778.

Grant, K. W., & Greenberg, S. (2001). Speech intelligibility derived from asynchronous processing of auditory-visual speech information. *Proceedings of the Workshop on Audio Visual Speech Processing, Scheelsminde, Denmark, September 7–9* (pp. 132–137).

Grant, K. W., van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (crossmodal) synchrony. *Speech Communication*, *44*(1–4), 43–53 (Special Issue: Audio Visual Speech Processing).

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 12639–12643.

Lewkowicz, D. J. (2000). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development*, *71*(5), 1241–1257.

Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, *39*(5), 795–804.

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46*(1), 66–77.

Lewkowicz, D. J., & Flom, R. (2013). The audio-visual temporal binding window narrows in early childhood. *Child Development*, http://dx.doi.org/10.1111/cdev.12142.

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences*, *103*(17), 6771–6774 (109 5).

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*(11), 470–478.

Lewkowicz, D. J., & Hansen-Tift, A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436.

Lewkowicz, D. J., & Kraebel, K. (2004). The value of multimodal redundancy in the development of intersensory perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processing*. Cambridge: MIT Press.

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match non-human primate faces & voices. *Infancy*, *15*(1), 46–60.

Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development*, *37*(2), 90–94.

Miner, N., & Caudell, T. (1998). Computational requirements and synchronization issues of virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, *7*, 396–409.

Navarra, J., Alsius, A., Velasco, I., Soto-Faraco, S., & Spence, C. (2010). Perception of audiovisual speech synchrony for native and non-native speech. *Brain Research*, *1323*, 84–93.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*(2), 499–507.

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*(5406), 1272–1273.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., et al. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 5297–5300.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10598–10602.

Pons, F., Teixidó, M., Garcia-Morera, J., & Navarra, J. (2012). Short-term experience increases infants' sensitivity to audiovisual asynchrony. *Infant Behavior & Development*, *34*(5), 815–818.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory–visual speech perception examined by fMRI and PET. *Neuroscience Research*, *47*(3), 277–287.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia*, *45*, 598–607.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, *1111*, 134–142.

Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In M. J. Naumer, & J. Kaiser (Eds.), *Multisensory object perception in the primate brain*. New York: Springer.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, *316*(5828), 1159.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63.