Time-course of attention to a talker’s mouth in monolingual and close-language bilingual children

Abstract

We presented 28 Spanish monolingual and 28 Catalan-Spanish close-language bilingual 5-year-old children with a video of a talker speaking in the children’s native and a non-native language and examined the temporal dynamics of their selective attention to the talker’s eyes and mouth. When the talker spoke in the children’s native language, monolinguals attended equally to the eyes and mouth throughout the trial whereas close-language bilinguals first attended more to the mouth and then distributed attention equally between the eyes and mouth. In contrast, when the talker spoke in a non-native language (English), both monolinguals and bilinguals initially attended more to the mouth and then gradually shifted to a pattern of equal attention to the eyes and mouth. These results indicate that specific early linguistic experience has differential effects on young children’s deployment of selective attention to areas of a talker’s face during the initial part of an audiovisual utterance.

Keywords: audiovisual speech, bilingualism, language proximity, selective attention, lipreading

Public Significance Statement.

This study shows that selective attention to a talker’s face is a temporally-dynamic process that depends on prior linguistic experience. Close-language 5-year-old bilingual children exhibited greater initial attention to a talker’s mouth than did monolingual children. This suggests that regular and continuous experience with two close languages modulates how audiovisual speech cues are exploited, primarily at the start of communicative bouts.

We usually hear and see our interlocutors talking during typical social interactions. Their eyes provide social and deictic cues (Birmingham & Kingstone, 2009) while their mouth provides spatiotemporally congruent auditory (A) and visual (V) speech cues. Normally, we integrate the A and V speech cues inherent in an interlocutor’s mouth and, as a result, gain access to perceptually more salient communicative information than that specified by A-only speech cues. The greater perceptual salience of audiovisual (AV) speech is evident in findings that adults exhibit increased comprehension of AV speech, relative to A-only speech, when it is presented in noise or in a second language (Arnold & Hill, 2001; Cotton, 1935; Mitchel et al., 2022; Reisberg, 1978; Reisberg et al., 1987; Sanders & Goodrich, 1971; Sumby & Pollack, 1954). It is also evident in findings that adults deploy greater attention to a talker’s mouth - the source of congruent A and V speech cues - when the speech signal is more challenging to process (Barenholtz et al., 2016; Birulés et al., 2020; Lansing & McConkie, 2003; Lusk & Mitchel, 2016; Vatikiotis-Bateson et al., 1998).

Greater attention to a talker’s mouth emerges in infancy and creates opportunities for infants to profit from the greater perceptual salience of AV speech cues both in terms of speech processing and the acquisition of speech and language. In the first study to provide evidence of a link between the emergence of selective attention to a talker’s mouth in infancy and speech processing, Lewkowicz and Hansen-Tift (2012) tracked selective attention to a talker’s eyes and mouth while 4-, 6-, 8-, 10-, and 12-month-old monolingual English-learning infants and English-speaking adults watched and listened to a talker speaking in English or in Spanish. When presented with English utterances, infants attended more to the talker’s eyes at 4 months, equally to the eyes and mouth at 6 months, more to the mouth at 8 and 10 months, and equally to the eyes and mouth at 12 months. The adults attended more to the eyes. When presented with Spanish (i.e., non-native) utterances, infants exhibited the same developmental pattern of shifting attention to the mouth by 8 months except that this time they continued to attend more to the mouth up to 12 months of age. Adults once again attended more to the eyes. Similar findings of a shift in attention to the mouth between 6 and 8 months of age in response to native AV speech have been reported in other studies (Tenenbaum et al., 2013).

Lewkowicz & Hansen-Tift (2012) interpreted the attentional shift to the talker’s mouth between 6 and 8 months of age as evidence of an emerging interest in AV speech *per se* based on the fact that the shift coincides with the emergence of canonical babbling (Oller, 2000). Moreover, Lewkowicz & Hansen-Tift (2012) interpreted the subsequent shift of attention away from a talker’s mouth at 12 months of age in response to native speech but not to non-native speech as a reflection of the differential effects of newly acquired native-language expertise. They proposed that when 12-month-old infants are confronted with native speech, they no longer need to rely as much on AV speech cues because they are familiar with them, but that they need to still rely on such cues when confronted with non-native speech because such cues are unfamiliar and, thus, harder to process. In essence, Lewkowicz & Hansen-Tift (2012) assumed that infants’ greater reliance on lipreading of a talker producing non-native speech facilitates their ability to process unfamiliar communicative information.

Subsequent studies have provided additional evidence of the emergence of selective attention to a talker’s mouth in infancy, of greater attention to the mouth when a talker speaks in a non-native language, and of a link between lipreading in infancy and language acquisition. Thus, studies have found that infants attend more to the mouth when presented with non-native than native speech (Birulés, Bosch, Brieke, Pons, & Lewkowicz, 2018; Kubicek et al., 2013; Pons, Bosch, & Lewkowicz, 2015; although see Morin-Lessard, Poulin-Dubois, Segalowitz, & Byers-Heinlein, 2019) and that greater attention to the talker’s mouth in infancy is associated with vocal imitation (Imafuku et al., 2019), rule-learning (Birulés et al., 2022), concurrent and later expressive language skills (Tenenbaum et al., 2015; Tsang et al., 2018; Young et al., 2009), and indirectly with greater vocal complexity and expressive communication (Santapuram et al., 2022). Finally, it has been reported that infants learning two rhythmically and phonologically close languages (e.g., Spanish and Catalan) rely more on AV speech cues than do monolingual infants (Fort et al., 2018; Pons et al., 2015) or than do bilingual infants learning phonologically distant languages such as Spanish and English (Birulés et al., 2018). This last set of findings is especially interesting because Spanish and Catalan are known to be more difficult for infants to discriminate than pairs of distant languages (Bosch & Sebastián-Gallés, 1997; Nazzi et al., 1998). In sum, the evidence to date indicates that infants rely on the greater perceptual salience of AV speech by increasing their attention to a talker’s mouth and that this enhances their speech processing, native-language acquisition, and the processing and learning of two closely related languages.

Interestingly, the deployment of selective attention to a talker’s mouth continues into early childhood where it now varies as a function of speech-processing demands, task difficulty, and prior linguistic experience. For example, unlike 12-month-old infants, 18-month-old toddlers deploy more attention to a talker’s mouth than eyes regardless of whether they are exposed to native or non-native speech (Hillairet de Boisferon et al., 2018). This age difference has been interpreted as reflecting differential speech-processing/task demands. Whereas 8- and 10-month-old infants’ greater attention to a talker’s mouth has been interpreted as reflecting the acquisition of phonology (Lewkowicz & Hansen-Tift, 2012), the greater attention to the mouth in 18-month-olds has been interpreted as reflecting the acquisition of vocabulary (Hillairet de Boisferon et al., 2018). Thus, selective attention to the mouth is presumed to reflect different underlying processes at different points in development. Similarly, and consistent with the previously-cited evidence that attention to the mouth is related to specific early experience with a particular type of speech input and task difficulty, studies of 5-6 year-old monolingual children - who possess substantial knowledge of their native language phonology (Bosch Galceran, 2004; Dodd et al., 2003) – have found that they attend equally to a talker’s eyes and mouth when exposed to native AV speech (Król, 2018; Morin-Lessard et al., 2019; Nakano et al., 2010). In contrast, studies of 5-6 year-old close-language bilingual children have found that they attend more to the mouth when they have to process AV speech in either of their native languages (Birulés et al., 2018; Pons et al., 2018).

The findings of greater attention to a talker's mouth in close-language bilingual children relative to monolingual or distant-language bilingual children raise interesting questions. Is it possible that, despite their substantial expertise in both of their languages, close-language bilingual children attend more to the mouth because they find it more difficult to process/parse the linguistic input in either of their relatively similar native languages? Does their reliance on the more perceptually salient AV speech cues make it easier for them to disambiguate the communicative content of their two input languages more accurately? Currently available empirical evidence does not provide answers to these questions. Indeed, the specific nature of close-language bilingual children's reliance on AV speech cues relative to distant-language bilinguals is only characterized at a global level, namely in terms of a difference in the overall magnitude of attention. Although such a difference is interesting, it is theoretically possible that the difference actually reflects distinct patterns of selective attention over the course of naturalistic interactions between interlocutors. That is, given that speech processing demands are likely to be greatest at the start of a communicative bout, it is theoretically possible that it is during the initial part of a communicative bout that the dynamics of selective attention are different in close-language than in distant-language bilinguals. The purpose of the current study was to test this possibility. To do so, we conducted a study in which we used eye tracking to investigate children's selective attention to AV speech and, in addition to measuring the overall amount of attention directed at the eyes and mouth of a talker, we examined the time-course of selective attention to these two areas of a talker's face over the course of a test trial.

Characterization of the temporal pattern of selective attention to a talker's face may provide novel insights into the process underlying the facilitation of speech processing that is usually observed when redundant and highly salient AV speech cues are available and attended. Although there is no doubt that measures of selective attention based on total looking time have clear face and *a priori* theoretical validity, they do not provide any insights into the temporal dynamics underlying the deployment of selective attention to a talker’s eyes and mouth during a communicative bout. In contrast, measures of temporal dynamics provide a finer-grained analysis of changes in selective attention to AV speech during a communicative bout and provide additional insights into underlying mechanisms.

The dynamics of selective attention to a talker’s face might manifest themselves in one of two ways. Given that a perceiver is first confronted with novel and often difficult-to-process communicative information at the beginning of a communicative bout, the principal changes in selective attention to a talker’s eyes and mouth may occur at the beginning to disambiguate the speech information. This should be manifest in an initial peak in selective attention directed to the mouth followed by a gradual decrease of mouth-directed attention as the trial proceeds in monolingual as well as distant- and close-language bilingual learners. This response pattern would be consistent with a perceptual adaptation process to speech and talker characteristics that has been observed in adults’ response to accented or non-native speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004) and would be similar to the gradual decrease in attention to the mouth observed in adults when they are becoming familiarized with novel artificial words (Lusk & Mitchel, 2016). Alternatively, given that bilinguals generally exhibit greater attention to a talker’s mouth starting in infancy (e.g., Birulés et al., 2018; Pons et al., 2015), they are likely to exhibit greater initial attention to a talker’s mouth than monolinguals and they may continue to attend more to the mouth throughout the communicative bout. In contrast, monolinguals may only attend more to the mouth at the start of a communicative bout.

Method

Participants

We recruited and tested sixty-six 4- to 6-year-old children from two schools located in Barcelona (i.e., a Catalan-Spanish bilingual environment) and one school located in Madrid, Spain (a Spanish-monolingual environment). None of the children had a history of hearing problems according to parental report. Parents completed an online language questionnaire (adapted from Bosch & Sebastián-Gallés, 2001; as used in Birulés et al., 2018[1]) to establish the language background of the participants. The questionnaires also ensured typical language development in all children. In the case of the bilingual children, language dominance was estimated from information relative to language exposure and use with parents, siblings, and in school. The bilingual children came from Catalan-dominant (n=21) or Spanish-dominant (n=7) homes and were exposed to their other native language early in life (i.e., at home or day-care, nursery centres and before entering preschool/kindergarten at age 3). In contrast, the monolingual children were only exposed to Spanish in their environment. After entering school, all participants had some limited oral exposure to English (approximately 2 hours/school week).

Ten of the 66 children tested were not included in the final data analysis because they were exposed to another language at home that was not Catalan or Spanish (n=2) or due to eye-tracking calibration failure (n=8). No participants were excluded based on the minimum looking time criterion of at least 20% looking during any trial (Birulés et al., 2018; Frank et al., 2012). The final sample of 56 children (Mean age = 5 years, 8 months, Range = 4 years, 2 months - 6 years, 9 months) consisted of 28 Spanish-Catalan bilinguals (Mean age = 5 years, 11 months, Range = 5 years, 5 months - 6 years, 6 months, 11 boys) and 28 Spanish monolinguals (Mean age = 5 years, 8 months, Range = 5 years, 5 months - 6 years, 6 months, 15 boys).

[1] Link to children's language questionnaire (in English): https://forms.gle/yfeUa54hAEQdgFBp7

Stimuli

The stimuli for this study were the same nine 60s videos presented by Birulés et al. (2020). In each of these videos, the same female actor could be seen (from the shoulders up) and heard reciting a children's story in a child-directed manner. The videos consisted of 3 sets of the same 3 different stories recited in Catalan, Spanish, and English, respectively. The actor was a 21-year-old a trilingual native speaker of Catalan, Spanish, and English who was exposed from birth to these three languages from her parents in her family environment (native English from her father and native Catalan/Spanish from her mother). The videos were recorded in a soundproof booth.

Apparatus and Procedure

Children were seated on an adjustable chair, 60 cm in front of a 17-inch monitor, in a small and dimly lit room of their school. Stimuli were presented with Tobii Studio software (Tobii Technology AB, Danderyd, Sweden) and eye gaze was recorded using a Tobii T120 eye-tracker at a sampling rate of 60 Hz. After Tobii's nine-point calibration routine, each child watched three video clips, one in each respective language. The order of the clips and the specific story presented in each language were counterbalanced across children. We designated English as the non-native language for all children, Spanish as the native language for monolingual children and Spanish or Catalan as the native language for bilinguals, depending on their dominant language at home[2]. The children were instructed to pay close attention to the videos, because they would be asked questions about the stories. These instructions were only given to ensure that they were fully engaged in the experiment.

[2] We found no differences in the pattern of results when we examined responsiveness to only Catalan or Spanish as the bilinguals' native language.

To measure selective attention, we created three areas of interest (AOIs) corresponding to the actor's eyes, mouth, and whole face (same AOIs as in Birulés et al., 2020). Figure 1 is a screenshot of one of the video frames showing the three respective AOIs.

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Bioethical committee of the University of XX. This study was not preregistered. The data, code and stimuli are publicly available at: https://osf.io/njxam/?view_only=9c4bd220c2c64129879759a6227b6c68.

Results

As in Birulés et al. (2018), we calculated a proportion of total looking time (PTLT) score for each participant for the eye and mouth AOIs, respectively, over the designated trial duration (i.e., either the first 10 s or the full 60 s). To compute the PTLT score, we divided the amount of time participants looked at each AOI, respectively, by the total amount of time they looked at the face. As can be seen in the Supplementary Materials section (SM 2), the PTLT scores were derived from relatively long looking times (i.e., 40 sec minimum in the full trial). We used the PTLT scores to conduct three separate analyses. First, to compare the current findings to those from our previous work with close-language bilinguals (Birulés et al., 2018), we used a mixed analysis of variance (ANOVA) to examine the proportion of selective attention deployed to the eyes and mouth during the first 10 s of the native and non-native trials. Second, we conducted an analysis of overall selective attention during the full 60 s test trial. Finally, we examined the time-course of selective attention to the talker's eyes and mouth over the course of a trial with a growth curve analysis (Mirman, 2014).

Overall selective attention during the first 10 s test-trial period.

We submitted the PTLT scores for the first 10 s to a mixed, repeated-measures ANOVA, with AOI (eyes, mouth) and Test Language (native, non-native) as within-subjects factors and Language Background Group (monolingual, bilingual) as the between-subjects factor. This ANOVA yielded an AOI main effect [*F*(1,54) = 13.12, *p* < .001, η_p^2 = .20], reflecting an overall preference for the mouth. The ANOVA also yielded a Language Background Group x AOI [*F*(1,54) = 4.78, *p* = .033, η_p^2 = .08], and a Language Background Group x AOI x Test Language [*F*(1,54) = 4.03, *p* = .050, η_p^2 = .07] interaction. The 3-way interaction is depicted in Figure 2.

Follow-up analyses of the Language Background Group x AOI x Test Language triple interaction consisted of separate repeated-measures ANOVAs for each respective language background group. The monolingual group ANOVA yielded a significant Test Language x AOI interaction [*F*(1,27) = 6.31, *p* = .018, η_p^2= .19]; this was due to greater attention to the mouth in the non-native but not in the native language condition. The bilingual group ANOVA only yielded a main effect of AOI [*F*(1,27) = 27.15, *p* < .001, η_p^2 = .50]; this was due to greater attention to the mouth in both language conditions.

Overall selective attention during the entire 60 s test-trial period.

Figure 3 shows the overall amount of selective attention deployed to the eyes and mouth during the entire 60 s trial period. A comparison of Figures 2 and 3 reveals that the overall pattern of responsiveness during the total 60 s trial period was similar to that found during the first 10 s of the trial. Using PTLT scores for the entire duration of each trial, we performed the same repeated-measures ANOVAs as we did for the 10 s duration data. Results yielded a near significant AOI main effect [*F*(1,54) = 3.67, *p* = .06, η_p^2 = .06], a significant Test Language main effect [*F*(1,54) = 6.31, *p* = .015, η_p^2= .1], a significant Test Language x AOI interaction [*F*(1,54) = 5.71, *p* = .02, η_p^2= .10], and a near significant Language Background Group x AOI x Test Language interaction [*F*(1,54) = 3.50, *p* = .067, η_p^2= .06].

Despite the fact that the 3-way interaction only approached significance, we conducted follow-up analyses of the Language Background Group x AOI x Test Language interaction to be consistent with the 10 s trial-period analyses. Separate repeated-measures ANOVAs for each respective language group yielded a significant Test Language x AOI interaction [$F(1,27) = 7.7$, $p = .01$, $\eta_p^2 = .22$] for the monolingual group and a main effect of AOI [$F(1,27) = 9.27$, $p < .01$, $\eta_p^2 = .26$] for the bilingual group. These results are similar to those from the 10 s test-trial period and show that monolingual children attended more to the mouth only in the non-native condition but that the bilingual children attended more to the mouth in both conditions (see table 1)

Analysis of the temporal dynamics of selective attention.

To examine the temporal dynamics of selective attention, we used a growth curve analysis (Mirman, 2014) and Post-hoc likelihood-ratio (χ2) forward model comparisons to analyze the time-course of selective attention. For this analysis, we down-sampled to a data-point each 60 ms, computed a difference score value (PTLT Eyes – PTLT Mouth) and rounded the data of each time point to 1 (Eyes-look) and 0 (Mouth-look). This method simplifies the model, avoids auto-correlation (when Eyes-look increases Mouth-look decreases and vice-versa), and allows a better fit of a binomial distribution. The overall time course (60 s) to the two AOIs was modeled with up to a third-order (cubic) orthogonal polynomial with the same fixed effects as in the previous analysis: Test Language and Language Background Group on all time terms (intercept, linear, quadratic, and cubic). The model also included participant random effects on all time terms except the cubic[3]. Native language and Monolingual children were used as the baseline in the model and relative parameters were estimated for the Bilingual group and the Non-native conditions. We decided to include time as not only a linear term but also as a quadratic and cubic term to allow the model flexibility to predict a potentially non-linear response pattern (i.e., an abrupt change in attention to the mouth over the trial). Data processing and statistical analyses were done in R 4.0.2 (R Core Team, 2020) and the dplyr (v1.0.5; Wickham, François,

[3] Estimating random effects is "expensive" in terms of the number of observations required, so this cubic term was excluded because it tends to capture less-relevant effects in the tails.

Henry, & Müller, 2021), lme4 and lmerTest packages (v.1.1-27 and v.3.1-3 respectively; Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). Full details on data processing, analysis, and models' results are available in Supplementary Materials.

Table 2 shows the Post-hoc likelihood-ratio (χ2) forward model comparisons; a comparison of each model fit with the same model plus one other variable. In this way, each variable is added one by one and only if it significantly improves the model. The results showed that the full model – containing linear, quadratic and cubic time polynomials and the two fixed effects plus all interactions – provided the best fit without compromising its convergence[4]. The model was coded in R as: [full model <- PTLT ~ (time + time^2 + time^3) * Language Background Group * Test Language + (time + time^2 | Participant), family="binomial" (link = "logit")].

Figure 4 summarizes the results of the full statistical model (full results in Supplementary Materials). The significant main effect of Test Language reflects greater attention to the mouth in the non-native condition [β= - 0.47 (0.03), z = - 18.01, *p* < .001]. Importantly, however, the absence of a Language Background Group main effect indicates that the greater attention to the mouth found in the previous analysis of the entire 60 s duration of the trial in bilingual children is no longer significant [β= - 0.96 (0.52), z = - 1.83, *p* = .07].

Examination of the time polynomials indicated that there was a significant triple interaction between the three time-terms (linear, quadratic and cubic), Language Background Group, and Test Language [β= - 0.92 (0.12), z = - 7.58, *p* < .001; β= - 0.56 (0.12), z = - 4.53, *p* < .001; β= - 1.58 (0.12), z = - 13, *p* < .001, respectively]. This means that the dynamic response profiles of the two linguistic groups across the two language conditions were significantly different.

[4] By the principle of marginality, a factor must be kept if the interaction is significant, regardless of the main effect. Moreover, changing the order of the comparisons – maximal and drop1 or forward (Barr et al., 2014) – yielded the same results (i.e. keeping the maximal model).

Visual inspection of the native-language condition data depicted in Figure 4 suggests that bilingual children initially preferred the talker’s mouth and that this mouth-preference diminished by approximately 20 s but that monolingual children exhibited a relatively flat response pattern that consisted of slightly greater attention to the talker’s eyes. In contrast, inspection of the non-native language condition data indicates that both monolingual and bilingual children exhibited a relatively similar pattern of initially greater attention to the mouth and that this diminished over time. Follow-up growth curve models built on the monolingual and bilingual group separately confirmed these results (full description of follow-up models in Supplementary Materials).

Discussion

We investigated selective attention to a talker’s face in 5-year-old monolingual and bilingual children while she spoke either in a native or a non-native language. Prior studies of infants and children (Birulés et al., 2018; Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019) and adults (e.g., Barenholtz et al., 2016; Birulés et al., 2020; Lansing & McConkie, 2003) have found that selective attention to a talker’s eyes and mouth is modulated by developmental experience and processing task. Individuals learning more than one language generally attend more to a talker’s mouth than those learning a single language. Similarly, attention to the mouth increases when speech must be processed in the context of competing noise (Buchan et al., 2007; Vatikiotis-Bateson et al., 1998), when speech is presented at a low volume (Lansing & McConkie, 2003), or when non-native speech is presented (Barenholtz et al., 2016; Birulés et al., 2020). The generally accepted interpretation of increased attention to a talker’s mouth is that this provides direct access to combined A and V speech cues. When processed together, combined AV speech cues are known to be perceptually more salient than A-only cues and, as a result, attention to such cues enhances speech comprehension (Savariaux et al., 2004).

The vast majority of the studies to date that have investigated selective attention to a talker’s face have relied on aggregate measures of attention to the eyes and mouth over the duration of the trial. It is theoretically reasonable, however, that the most important changes in selective attention to a talker's eyes and mouth occur at the start of a communicative bout because it is then that a perceiver is first confronted with novel and often difficult-to-process communicative information. If so, it would be highly informative to examine the temporal dynamics of selective attention during the course of a communicative bout. We did that in the current study by investigating the temporal dynamics of selective attention to a talker's eyes and mouth in bilingual and monolingual children's response to native and non-native speech and compared traditional aggregate measures of selective attention to our new temporal measures.

Our first two analyses examined the total amount of selective attention during the first 10 s of the test trial as well as during the entire 60 s of the test trial. Consistent with findings that learning two close languages increases attention to a talker's mouth both in infancy (Pons et al., 2015) and in early childhood (Birulés et al., 2018) , the 10 s analysis revealed that the close-language bilingual children exhibited greater mouth-looking than did their monolingual counterparts. The 60 s analysis yielded similar results to those obtained in the 10 s analysis in that bilingual children attended more to the mouth in both conditions but that monolingual children attended more to the mouth only in the non-native condition. Finally, the temporal analysis indicated that the dynamic response profiles differed as a function of language background and language presented. Specifically, in response to native speech, bilingual children deployed more attention to the mouth during the initial part of the trial whereas monolingual children deployed equal attention to the eyes and mouth throughout the trial. In response to non-native speech, both bilingual and monolingual children initially deployed more attention to the mouth and then gradually stopped doing so as the trial progressed.

The results from the temporally based analyses provide new insights into the dynamics of selective attention and pinpoint where the difference between bilinguals and monolinguals first appears. Specifically, the results show that both bilingual and monolingual children only deployed greater attention to the talker’s mouth during the initial phase of speech processing, with bilinguals exhibiting this in response to both native and non-native speech but monolinguals only to non-native speech. This shows that this early focus on a talker’s mouth - modulated by early linguistic experience - reflects a general attentional strategy and suggests that the highly salient AV speech cues initially enhance speech processing. It also shows that once children have used the highly salient speech cues to help them disambiguate the communicative signal, they then resort more to listening while also looking at the eyes where social and deictic cues are located. This interpretation is consistent with previous evidence showing a similar decrease of attention to the mouth when adults become familiarized with an AV task (Lusk & Mitchel, 2016) or when adults perceptually adapt to foreign-accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004).

The temporal dynamics data indicate that bilingual children rely on a general perceptual processing strategy that maximizes access to the highly salient AV speech cues available in a talker’s mouth regardless of language, a strategy that reduces initial uncertainty and favors a quick adaptation to the specific acoustic-phonetic characteristics of the incoming language. In contrast, monolingual children only rely on this strategy when they detect unfamiliar, non-native speech. Interestingly, the temporal change in attention from the mouth to the eyes peaked earlier in response to native than non-native speech (at around 20 s in the native and 40 s in the non-native speech condition, see Figure 4). This suggests that the rate of decrease may be proportional to the difficulty of the speech-processing task. In other words, the more difficult the processing task, the longer perceivers attend to a talker’s mouth. Crucially, here the native vs. non-native speech difference cannot be explained by differential perceptual salience features because the same trilingual speaker produced all the utterances.

It should be noted that even though the current study investigated children’s foveal attention to the eyes and mouth of a talker’s face, prior studies have shown that audiovisual speech cues also can be processed outside of foveal fixation (i.e., McGurk effect in the visual periphery, Paré et al., 2003) or under low spatial frequency (Munhall et al., 2004). Therefore, we cannot assume that children only benefit from audiovisual redundancy when they attend to a talker's mouth. Rather, the foveated attention that we measured most likely reflects children's attempt to decipher the speech utterance based on the maximally informative speech cues, namely the directly accessed redundant AV speech cues in the talker's mouth. Based on this interpretation, the reduction of attention to the mouth over the course of the utterance may reflect a shift to more automatic processing of the utterance through extra-foveal forms of attention.

It should also be noted that the current study did not include any measures of active learning or processing. In order to infer a causal link between selective attention to a talker's mouth and language acquisition, future studies will need to combine measures of selective attention with measures of language learning in tasks in which such processes as lexical segmentation, word learning, or rule-learning are manipulated. Similarly, future studies using psycho-physiological measures such as EEG, pupillometry or heart rate might help determine whether the initial attention to the talker's mouth that we observed here is correlated with greater cognitive effort.

In conclusion, the current study is the first to characterize the dynamics of selective attention in bilingual and monolingual children's processing of native and non-native speech. When children's prior linguistic experience involves more than one language, and when the languages are prosodically, phonologically, lexically, and even syntactically close, such bilingual children initially attend more to the audiovisual speech cues located in a talker's mouth than do monolingual children. Presumably, this helps the bilingual children identify the acoustic-phonetic characteristics of the input language and disambiguate the initial communicative information. Once they have done so, they then decrease their attention to the talker's mouth. In contrast, when children's prior linguistic experience is with a single language, they only exhibit the initially greater focus on the talker's mouth when they need to process speech in a non-native language. Overall, the current results demonstrate that selective attention to a talker’s face in early childhood is a dynamically fluid process that depends both on prior linguistic experience with either one or more than one language and the specific language to be processed. Furthermore, our results suggest that speech perception in 5-year-old children is facilitated by AV speech cues inherent in a talker’s mouth at the start of communicative bouts. Overall, when our findings are considered in the context of findings from prior studies, they demonstrate that deployment of selective attention to the redundant and highly salient AV speech cues inherent in a talker’s mouth continues to play an important role in speech processing into early childhood. In addition, they indicate that regular and continuous experience with two close languages modulates the manner in which AV speech cues are exploited and that close bilingual children approach the speech processing task differently than do monolinguals.

References

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339–355. https://doi.org/10.1348/000712601162220

Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, *147*, 100–105. https://doi.org/10.1016/j.cognition.2015.11.013

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2014). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 1–43. https://doi.org/10.1016/j.jml.2012.11.001.Random

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Birmingham, E., & Kingstone, A. (2009). Human social attention: A new look at past, present, and future investigations. *Annals of the New York Academy of Sciences*, *1156*, 118–140. https://doi.org/10.1111/j.1749-6632.2009.04468.x

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2018). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science*, *22*(3), 1–11. https://doi.org/10.1111/desc.12755

Birulés, J., Bosch, L., Pons, F., & Lewkowicz, D. J. (2020). Highly proficient L2 speakers still need to attend to a talker's mouth when processing L2 speech. *Language, Cognition and Neuroscience*, *35*(10), 1314–1325. https://doi.org/https://doi.org/10.1080/23273798.2020.1762905

Birulés, J., Martinez-Alvarez, A., Lewkowicz, D. J., de Diego-Balaguer, R., & Pons, F. (2022). Violation of non-adjacent rule dependencies elicits greater attention to a talker's mouth in 15-month-old infants. *Infancy*, *27*(5), 963–971. https://doi.org/10.1111/infa.12489

Bosch Galceran, L. (2004). *Evaluación fonológica del habla infantil*. Masson.

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. https://doi.org/10.1016/S0010-0277(97)00040-1

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy*, *2*(1), 29–49. https://doi.org/10.1207/S15327078IN0201_3

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 1–22. http://www.sciencedirect.com/science/article/pii/S0010027707001126

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, *2*(1), 1–13. https://doi.org/10.1080/17470910601043644

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Cotton, J. C. (1935). Normal "Visual Hearing." *Science*, *82*(2138), 592–593.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics*, *17*(8), 617–643. https://doi.org/10.1080/0269920031000111348

Fort, M., Ayneto-Gimeno, A., Escrichs, A., & Sebastián-Gallés, N. (2018). Impact of Bilingualism on Infants' Ability to Learn From Talking and Nontalking Faces. *Language Learning*, *68*, 31–57. https://doi.org/10.1111/lang.12273

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social Attention Using Free-Viewing. *Infancy*, *17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology*, *172*(April), 189–200. https://doi.org/10.1016/j.jecp.2018.03.009

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental Science*, *March*, e12825. https://doi.org/10.1111/desc.12825

Król, M. E. (2018). Auditory noise increases the allocation of attention to the mouth, and the eyes pay the price: An eye-tracking study. *PLoS ONE*, *13*(3), 1–14. https://doi.org/10.1371/journal.pone.0194491

Kubicek, C., Boisferon, A. H. de, Dupierrix, E., Loevenbruck, H., Gervain, J., & Schwarzer, G. (2013). Face-scanning behavior to silently talking faces in 12-month-old infants : The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, *37*(2), 106–110.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, *65*(4), 536–552. https://doi.org/10.3758/BF03194581

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 of the United States of America*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lusk, L. G., & Mitchel, A. D. (2016). Differential Gaze Patterns on Eyes and Mouth During Audiovisual Speech Segmentation. *Frontiers in Psychology*, *7*(February), 52. https://doi.org/10.3389/fpsyg.2016.00052

Mirman, D. (2014). Growth curve analysis and visualization using R. In J. M. Chambers, T. Hothorn, D. Temple Lang, & H. Wickham (Eds.), *The R Series* (Vol. 26, Issue 3). CRC Press/Taylor & Francis. https://doi.org/10.1177/0962280215570173

Mitchel, A. D., Lusk, L. G., Wellington, I., & Mook, A. T. (2022). Segmenting Speech by Mouth: The Role of

Oral Prosodic Cues for Visual Speech Segmentation. *Language and Speech*. https://doi.org/10.1177/00238309221137607

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Heinlein, K. B.-. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*, 1–60.

Munhall, K. G., Kroos, C., Jozan, G., & Vatikiotis-Bateson, E. (2004). Spatial frequency requirements for audiovisual speech perception. *Perception & Psychophysics*, *66*(4), 574–583.

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., Ohta, H., Kato, N., & Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2935–2943. https://doi.org/10.1098/rspb.2010.0587

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Oller, D. K. (2000). The emergence of the speech capacity. *Journal of Child Language*, *30*(3), 731–734. https://doi.org/10.1121/1.1388001.

Paré, M., Richler, C., Ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics*, *65*(4), 553–567.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, *26*(4), 490–498. https://doi.org/10.1177/0956797614568320

Pons, F., Sanz-Torrent, M., Ferinu, L., Birulés, J., & Andreu, L. (2018). Children With SLI Can Exhibit Reduced Attention to a Talker's Mouth. *Language Learning*, *June*, 180–192.

https://doi.org/10.1111/lang.12276

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Reisberg, D. (1978). Looking where you listen: visual cues and auditory attention. *Acta Psychologica*, *42*(4), 331–341. https://doi.org/10.1016/0001-6918(78)90007-0

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-reading* (pp. 97–113). Lawrence Erlbaum Associates, Inc.

Sanders, D. A., & Goodrich, S. J. (1971). The Relative Contribution of Visual and Auditory Components of Speech to Speech Intelligibility under Varying Conditions of Frequency Distortion. *Journal of Speech Language and Hearing Research*, *14*(1), 154–159. https://doi.org/10.1121/1.2143572

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., Kim, S. Y., Keceli-Kaysili, B., Bottema-Beutel, K., Lewkowicz, D. J., Wallace, M. T., & Woynaroski, T. G. (2022). Mechanisms by Which Early Eye Gaze to the Mouth During Multisensory Speech Influences Expressive Communication Development in Infant Siblings of Children with and Without Autism. *Mind, Brain, and Education*, 1–13. https://doi.org/10.1111/mbe.12310

Savariaux, C., Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), 69–78. https://doi.org/10.1016/j.cognition.2004.01.006

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased Focus on the Mouth Among Infants in the First Year of Life: A Longitudinal Eye-Tracking Study. *Infancy*, *18*(4), 534–553. https://doi.org/10.1111/j.1532-7078.2012.00135.x

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

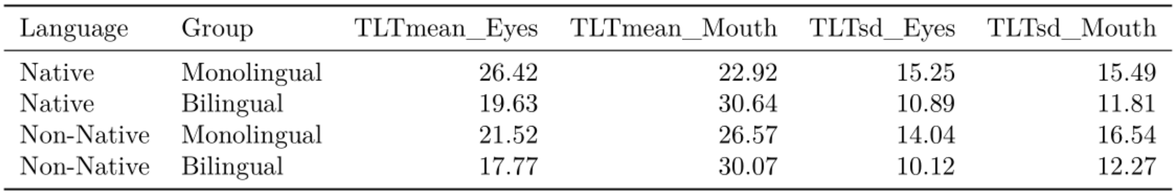
Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, *60*(6), 926–940. https://doi.org/10.3758/BF03211929

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. https://cran.r-project.org/package=dplyr

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Table 1

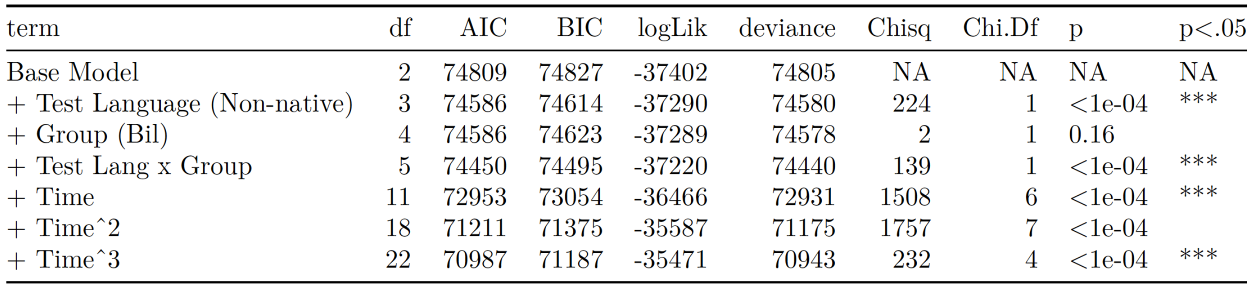
Mean and Standard Deviation (sd) of children Looking Times (TLT, in seconds) to the eyes and mouth of the talker, during the full trial (60 sec).



Language	Group	TLTmean_Eyes	TLTmean_Mouth	TLTsd_Eyes	TLTsd_Mouth
Native	Monolingual	26.42	22.92	15.25	15.49
Native	Bilingual	19.63	30.64	10.89	11.81
Non-Native	Monolingual	21.52	26.57	14.04	16.54
Non-Native	Bilingual	17.77	30.07	10.12	12.27

Table 2

Post-hoc likelihood-ratio (χ2) forward model comparisons’ statistics. Time^2 and Time^3 refer to quadratic and cubic time terms, Group refers to Language Background Group.



term	df	AIC	BIC	logLik	deviance	Chisq	Chi.Df	p	p<.05
Base Model	2	74809	74827	-37402	74805	NA	NA	NA	NA
+ Test Language (Non-native)	3	74586	74614	-37290	74580	224	1	<1e-04	***
+ Group (Bil)	4	74586	74623	-37289	74578	2	1	0.16	
+ Test Lang x Group	5	74450	74495	-37220	74440	139	1	<1e-04	***
+ Time	11	72953	73054	-36466	72931	1508	6	<1e-04	***
+ Time^2	18	71211	71375	-35587	71175	1757	7	<1e-04	
+ Time^3	22	70987	71187	-35471	70943	232	4	<1e-04	***

Note. Sig.: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 1

Still photo of the talker's face showing the eyes (blue), mouth (red), and face (white) AOIs.

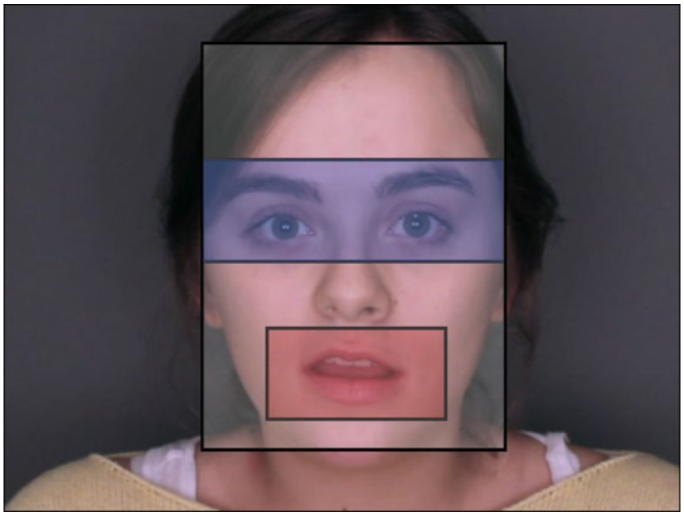


Figure 2

Selective attention directed at the eyes and mouth during the first 10 seconds of the video of native and non-native speech in monolingual and bilingual children. Dots represent each child's mean PTLT score and bars and crosses with error bars represent mean PTLT scores and standard errors of the mean (SE) for each group.

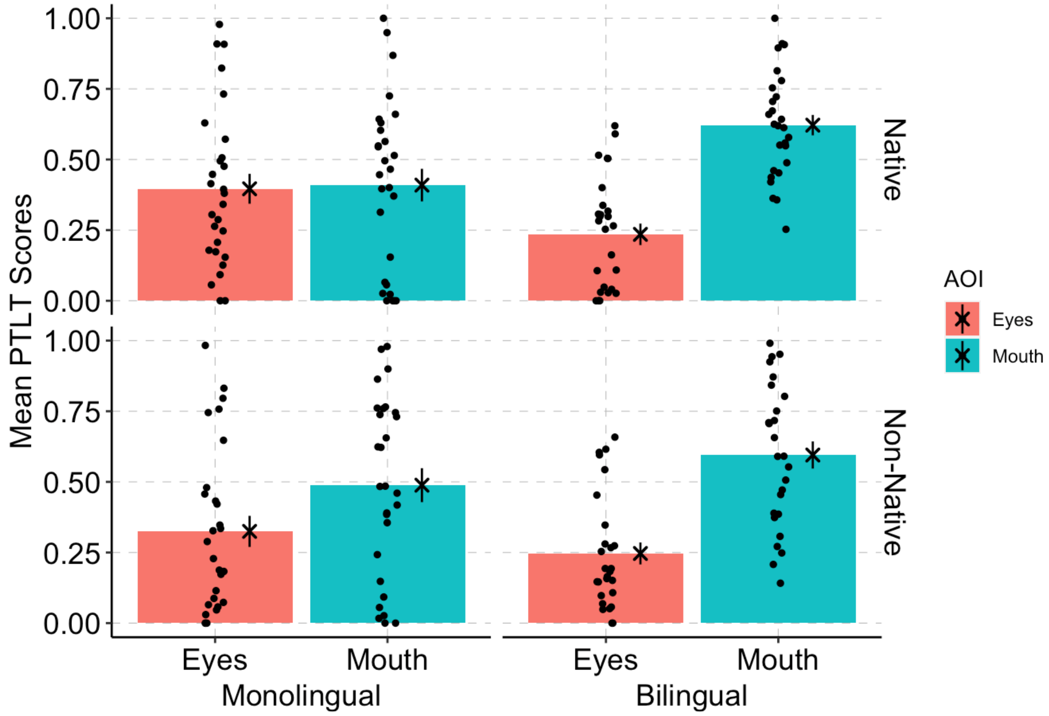


Figure 3

Selective attention directed at the eyes and mouth during the entire 60 s video of native and non-native speech in monolingual and bilingual children. Dots represent each child's mean PTLT score and bars and crosses with error bars represent mean PTLT scores and standard errors of the mean (SE) for each group.

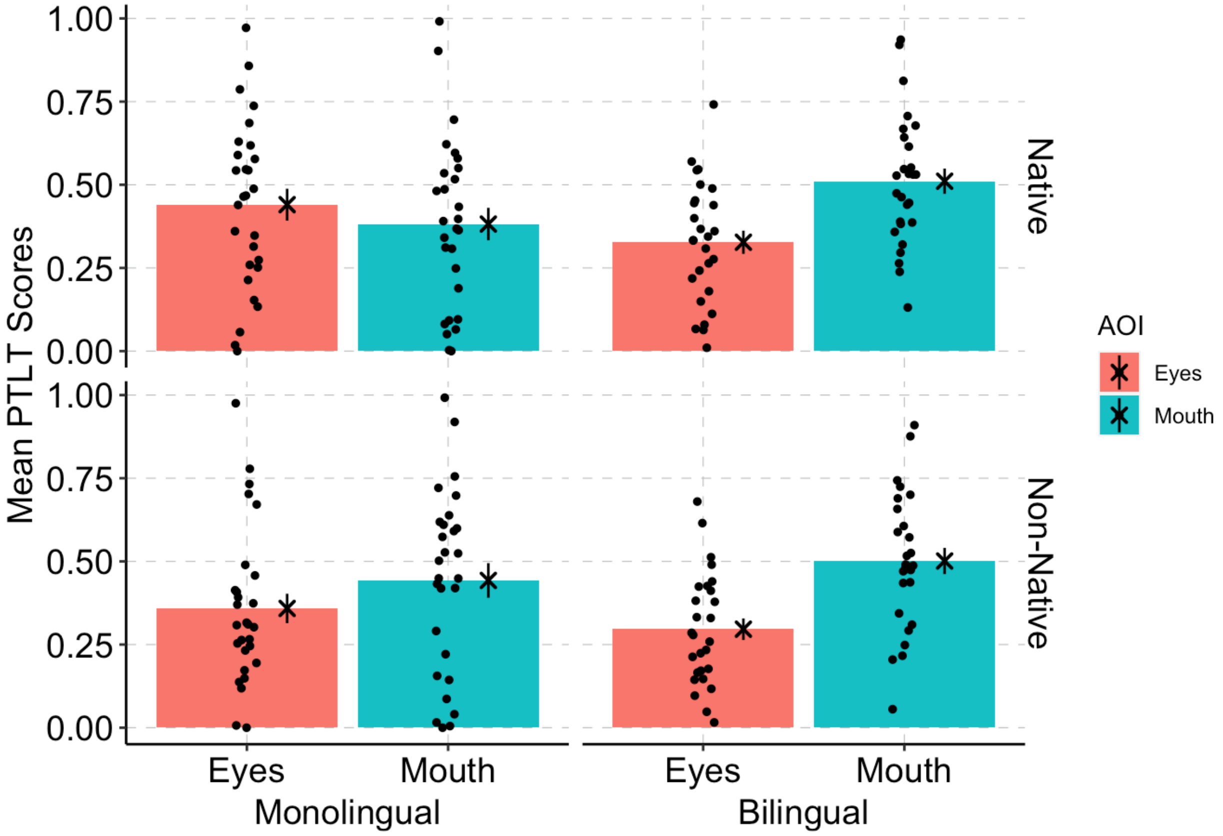
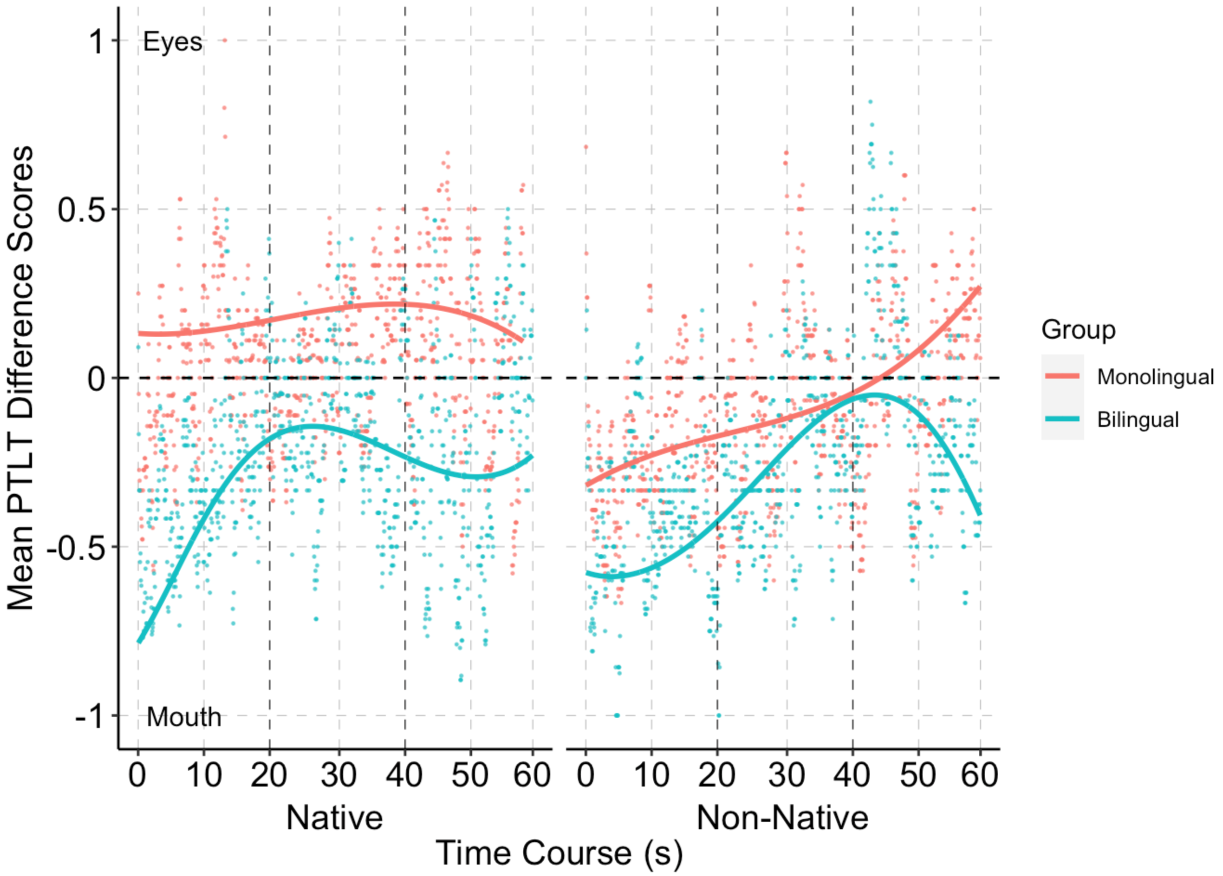


Figure 4

Time-course graphs for the monolingual (red) and close-language bilingual (blue) children's Mean Difference Score (i.e., PTLTeyes - PTLTmouth) across the 60 s trial plotted separately for the Native and Non-native conditions. The lines represent the fitted GLMM including time up to the cubic term for each condition, recoded here as 1 for 100% eyes and -1 for 100% mouth for consistency with prior studies. Dots represent each group Mean PTLT Difference Score for each time point.


Mean PTLT Difference Scores
Eyes
Mouth
Native
Non-Native
Time Course (s)
Group
Monolingual
Bilingual