Seeing a Talking Face Matters: Infants' Segmentation of Continuous Auditory-Visual Speech

S. H. Jessica Tan[1], Marina Kalashnikova[2,3], Denis Burnham[1]
[1]The MARCS Institute of Brain, Behaviour and Development, Western Sydney University,
[2]The Basque Center on Cognition, Brain and Language
[3]IKERBASQUE, Basque Foundation for Science

This research was funded by a doctoral scholarship to the first author funded by the MARCS Institute at Western Sydney University and the HEARing Cooperative Research Centre (CRC), and by HEARingCRC funding to the last author. The second author's work is supported by the Basque Government through the BERC 2018-2021 program, and PIBA PI-2019-0054, and by the Spanish Ministry of Science and Innovation through the Ramon y Cajal Research Fellowship, PID2019-105528GA-I00.

Correspondence concerning this article should be addressed to S. H. Jessica Tan at j.tan@westernsydney.edu.au.

Abstract

Visual speech cues from a speaker's talking face aid speech segmentation in adults, but despite the importance of speech segmentation in language acquisition, little is known about the possible influence of visual speech on infants' speech segmentation. Here, to investigate whether there is facilitation of speech segmentation by visual information, two groups of English-learning 7-month-old infants were presented with continuous speech passages, one group with auditory-only (AO) speech and the other with auditory-visual (AV) speech. Additionally, the possible relation between infants' relative attention to the speaker's mouth versus eye regions and their segmentation performance was examined. Both the AO and the AV groups of infants successfully segmented words from the continuous speech stream, but segmentation performance persisted for longer for infants in the AV group. Interestingly, while AV group infants showed no significant relation between the relative amount of time spent fixating the speaker's mouth vs. eyes and word segmentation, their attention to the mouth was greater than that of AO group infants, especially early in test trials. The results are discussed in relation to the possible pathways through which visual speech cues aid speech perception.

Keywords: auditory-visual speech perception, speech segmentation, looking behaviour, infants, language development

1. Introduction

One of the earliest tasks that language learners must master is accurate identification of word boundaries and segmentation of utterances into words. This is by no means an easy feat—word boundaries are not systematically marked by acoustic cues (Aslin, Woodward, LaMendola, & Bever, 1996; Cole, Jakimik, & Cooper, 1980), no two utterances are acoustically identical, and speaker characteristics are highly variable. Despite such complexities, there is substantial evidence that, regardless of linguistic background, infants segment continuous speech within the first year of life (e.g., Butler & Frota, 2018; Jusczyk & Aslin, 1995; Marquis & Shi, 2008), and this ability is further shaped by the surrounding language, becoming increasingly robust with age (e.g., Schmale, Cristia, Seidl, & Johnson, 2010; Thiessen & Saffran, 2004).

English-learning infants are able to parse words from fluent speech streams by 7.5 months (Jusczyk & Aslin, 1995). This segmentation ability develops over time: it is preceded by infants' use of familiar words such as their own name and 'mommy' by 6 months (Bortfeld, Morgan, & Golinkoff, 2005), and then enhanced as they begin to draw on different sources of information such as the stress patterns of their native language (Curtin, Mintz, & Christiansen, 2005; Jusczyk, Houston, & Newsome, 1999b), statistical probabilities between syllables by 7.5 months (Johnson & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003) and then prosodic boundaries by 8 months (Seidl & Johnson, 2006), phonotactic constraints by 9 months (Mattys, Jusczyk, Luce, & Morgan, 1999), and allophonic cues by 10.5 months (Jusczyk, Hohne, & Bauman, 1999a).

Several factors can influence infants' ability to identify word boundaries in continuous speech. One such factor is the gender of the speaker's voice: when familiarised with repetitions of target words spoken by a female talker, 7.5-month-olds segment target words from test passages only when the test passages are spoken by a same gender (female)

talker, and not by a different gender (male) talker (Houston & Jusczyk, 2000). Another factor is reduplication: 9-month-olds perform better on a word segmentation task when familiarised with passages containing reduplicated words than with passages containing non-reduplicated words (Ota & Skarabela, 2018). Accent and lexical stress can also influence speech segmentation: English-learning 9-month-olds are unable to segment across different English accents (Schmale et al., 2010), and they mis-segment the less frequent iambic (weak-strong) disyllabic words. Presumably this occurs due to infants' sensitivity to the predominance of trochaic (strong-weak) disyllabic words in English, such that they treat stressed syllables as word onsets (Thiessen & Saffran, 2003; 2004). These studies on lexical stress and accent additionally demonstrate infants' developmental improvement of word segmentation over time: by 12 months, English-learning infants can segment iambic words (Thiessen & Saffran, 2004) and words from passages produced in a non-local dialectal accent (Schmale et al., 2010) successfully, and by 13 months, infants can segment words from passages in an unfamiliar foreign accent (Schmale & Seidl, 2009).

Together, these studies illustrate that infant segmentation ability is evident by 7.5 months, but it continues to be consolidated into the second year of life. Individual differences in segmentation performance at 7 months can be observed and are correlated with later vocabulary mastery. For instance, a retrospective analysis revealed that children who successfully segmented words at 7.5 months had greater expressive vocabulary at 24 months than children who were unsuccessful at the word segmentation task as infants (Newman et al., 2006). Similar findings were reported by Singh et al. (2012) and Newman et al. (2015) who found that segmentation performance—indexed by greater listening preference for target words over non-target words—was correlated with expressive vocabulary at 24 months. Furthermore, the relationship between segmentation performance and later language abilities is also evident at the neurophysiological level: in a series of studies, Junge, Kooijman and

colleagues found that infants who showed a left frontal negativity to target words at 7 months had significantly higher language comprehension and word production skills compared to infants who showed a left frontal positivity (Junge et al., 2012; Kooijman et al., 2013). Together, these findings illustrate that, at the young age of 7 months, there is meaningful variability in segmentation abilities even though there is successful segmentation performance at the group level. These findings also suggest that the infants with weaker segmentation capacities may benefit from additional information from the speech stream and in the context in which they receive language input.

Even though infant speech perception is now known to be a multimodal process (e.g., Burnham & Dodd, 2004; Desjardins & Werker, 2004), studies of infant speech segmentation have largely focused on auditory-only speech. With adults, segmentation studies using artificial languages have shown better word identification within an artificial language stream when the stream is paired with a congruent dynamic video of a speaker's talking face than when it is presented only in the auditory modality (Lusk & Mitchel, 2016; Mitchel & Weiss, 2010; 2014). This raises the possibility that visual speech information may aid *infants'* speech segmentation, especially since behavioural studies have found that infants benefit from visual speech information in other speech perception tasks (Teinonen, Aslin, Alku, & Csibra, 2008). In the only study that has examined infants' auditory-visual speech segmentation, Hollich, Newman, and Jusczyk (2005) presented 7.5-month-olds with target auditory recordings paired with a background distractor passage to simulate speech-in-noise with (1) a still photo of the speaker's face, (2) a congruent video of the speaker's talking face, (3) an unsynchronised video of the speaker's talking face, or (4) an oscilloscope pattern synchronised with the auditory recordings. Infants were able to segment words when the visual display consisted of the congruent video of the speaker's talking face and the synchronous oscilloscope pattern, but not when the visual display consisted of the still photo

of the speaker's face or the unsynchronised video of the speaker's talking face. As the visible amplitude deviations of the oscilloscope pattern were temporally aligned with the syllabic lip movements, these results suggest that the facilitation that visual speech cues bring to speech perception is associated with the temporal concordance of the auditory and visual information. No subsequent study has investigated whether the same augmentation can be found even without the distractor background speech.

To derive benefits from an interlocutor's talking face, infants must first *attend* to the talking face. Research on gaze behaviour to the eye and mouth regions of a speaker's talking face has shown that infants treat the mouth as an important source of linguistic information: infants already seek linguistic information from the mouth within their first six months (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013), attend more to the speaker's mouth than the eye region at 8 months whether the speaker talks in a native or a non-native language (Lewkowicz & Hansen-Tift, 2012), and continue to do so at 12 months but only when the speaker talks in a non-native language (Pons, Bosch, & Lewkowicz, 2019). Together, these findings demonstrate that infants can efficiently deploy their attention to relevant visual speech cues.

What is not clear from the infant speech segmentation research is whether there is a direct link between infant gaze behaviour, specifically fixation to the mouth, and their speech segmentation performance. It is possible that individual differences in mouth-looking behaviour may modulate infants' speech segmentation performance. As the temporal pattern of mouth movements is highly correlated with the acoustic timescale of syllables, researchers have proposed that mouth movements provide information regarding the onset and offset of syllables (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009), and that infants can make use of the alignment between auditory and visual components of speech, such as mouth movements, to segment speech (Kitamura, Guellaï, & Kim, 2014). Notably,

previous studies have found that infants' looking time to a speaker's mouth correlates with their concurrent (Tsang et al., 2018) and later expressive (Young et al., 2009) and later receptive (Imafuku & Myowa, 2016) language abilities, demonstrating the link between infant looking behaviour and their language skills. Even so, the relationship between infant gaze behaviour and speech segmentation has not yet been studied directly.

This study addresses the possible modulating effects of visual speech information on infant speech segmentation and the mechanisms that might underly any such visual speech benefit. There are two aims: (1) to examine whether visual speech information enhances infants' segmentation performance even in the absence of background distractor speech as used in the Hollich et al. (2005) study; and (2) to assess whether individuals' differential gaze to the speaker's mouth versus eye regions modulates segmentation performance.

As in previous segmentation studies employing the passage-to-word paradigm (e.g., Hollich et al., 2005; Jusczyk & Aslin, 1995), the English-learning 7.5-month-old infants in this study were first familiarised with passages containing target words and then tested with isolated tokens of target and non-target words. One group of infants was presented with familiarisation and test stimuli in the auditory-visual (AV) modality while a second group of infants was presented with stimuli in the auditory-only (AO) modality. Unlike previous segmentation studies that have traditionally used the head-turn preference paradigm, this study uses a familiarisation-test procedure with a single central screen (as in Thiessen, 2010) to accommodate the use of an eye-tracker to record infants' gaze patterns. This allowed us to measure the time that infants in the AV condition spent fixating the speaker's eyes and mouth during familiarisation and test. Another novel aspect of this design is that our task included two test blocks separated by a re-familiarisation phase. Thus, we assessed infants' immediate and delayed responses to the target words.

The first aim of this study was to assess whether visual information enhances infants' speech segmentation, and the second aim was to investigate individual differences. To those ends, we tested infants at 7.5 months. This is an age at which, with respect to Aim 1, at the *group* level, infants exhibit successful word segmentation, both in the various auditory-only studies and in the single extant auditory-visual word segmentation study (which included background distractor speech); and with respect to Aim 2, at the *individual* level, individual differences that are significant and meaningful (i.e., differences that predict individual lexical development, Newman et al., 2015; Singh et al., 2012) continue to be observed (Newman et al., 2006). We constructed two hypotheses. First, it was hypothesised that if visual speech information augments speech segmentation, then infants in the AV condition would be expected to outperform infants in the AO condition. To this end, our design included an index of speech segmentation performance, the *d* score, given by the difference between looking times to target versus non-target words, with larger *d* scores indicating better segmentation performance (e.g., Singh et al., 2012). This index was taken on two occasions, after the first block of familiarisation trials (Block 1) and after the second block of familiarisation trials (Block 2). The second block of trials was included to allow sufficient exposure for infants to become familiarised to the passages, and for us to capture additional data by which our second hypothesis could be interrogated (see 1.4 for further details). Moreover, the inclusion of this Block variable (Block 1, Block 2) in the experimental design enabled us to measure speech segmentation *per se* (Block 1 *d* scores), and in a more exploratory vein, examine the persistence of speech segmentation over time (Block 2 *d* scores versus Block 1 *d* scores). The second hypothesis concerned individual gaze patterns. We hypothesised that if the speaker's mouth is the main source of visual speech benefit, then (a) AV group infants should attend (i) more to the mouth than the eyes, and (ii) more to the mouth than infants in the AO condition,

and (b) AV group infants' differential attention to the mouth region should be positively correlated with segmentation performance, as measured by *d* scores.

Method

1.1. Participants

Thirty-seven 7.5-month-old monolingual Australian-English learners (20 females, mean age = 7.21 months, range = 7.03-7.90 months) took part in this study. Eighteen infants participated in the auditory-only (AO) condition (9 females, mean age = 7.38 months, range = 7.03-7.87 months), and 19 in the auditory-visual (AV) condition (11 females, mean age = 7.19 months, range = 7.07-7.90 months). All infants were born full-term, with no vision or hearing deficits, and were not at risk for any language or cognitive delay and had no history of ear infections. Data from eight additional infants were excluded because they were either fussy and failed to complete the experiment (n = 6) or had less than 40% gaze samples recorded by the eye-tracker (n = 2). The present study was conducted according to guidelines laid down in Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. This study was approved by the Human Research Ethics Committee at Western Sydney University (approval number H11517). The approved protocol regarding participant recruitment, data collection and data management was adhered to. All parents provided informed written consent prior to their infants' participation in this study.

1.2. Stimuli

A female native Australian English speaker was recorded producing, in infant-directed speech (IDS), the four different 6-sentence passages used by Jusczyk & Aslin (1995) (Appendix A). Passages centred around the target words 'cup', 'dog', 'bike', and 'feet'. The recordings were auditory-visual encompassing the speaker's head, face and neck. The

average duration of the passages was 24s (range = 23 to 27s, SD = 2s). Additionally, for the single word test stimuli the speaker spoke each of the four target words 8 times in succession incorporating some degree of variation in intonation. As these were video recordings, to maintain the naturalness of the dynamics of the speaker's talking face, the speaker was instructed to say these words with a 1-second pause between each word. These video recordings and the corresponding audio were used in the auditory-visual (AV) condition. In the auditory-only (AO) condition, auditory recordings were extracted from the video recordings and paired with a still image of the speaker's smiling face.

1.3. Apparatus

Visual stimuli were presented via a 17-inch DELL LCD monitor and auditory recordings were played via two loudspeakers (Edirol MA-15 Digital Stereo Micro Monitors) placed at the left and right side of the monitor. A Tobii X120 eye tracker was placed below the screen to record infants' gaze patterns throughout the session. The eye movements of each infant were calibrated using a 5-point calibration routine before the session began.

1.4. Procedure

A familiarisation-then-test design was employed using a single central screen (Thiessen, 2010). Infants sat on their parent's lap approximately 60cm from the screen. An experimenter was stationed in the adjacent control room throughout the experiment. The experiment consisted of two familiarisation and two test phases. Each infant completed two experimental phases, each consisting of familiarisation and test presented in the following order: Familiarisation Phase 1, Test Phase 1, Familiarisation Phase 2, and Test Phase 2. The duration of each phase was identical across participants; since trial duration was not infant-controlled, each trial was presented until completion. An attention-getter animation was played between trials and phases to recapture infants' attention to the screen. A schematic representation of the procedure is shown in Figure 1.

In Familiarisation Phase 1, infants were presented with two repetitions of two passages, a total of four trials. Half of the infants were familiarised with two repetitions of the passages containing 'cup' and 'dog' while the other half were familiarised with two repetitions of the passages containing 'bike' and 'feet'. The two familiarisation passages were presented on alternate trials, e.g., 'cup' repetition 1, 'dog' repetition 1, 'cup' repetition 2, 'dog' repetition 2, with order of passages counterbalanced between infants. The mean duration of this familiarisation phase was 98.90s.

In Test Phase 1, all four words (*cup*, *dog*, *bike*, and *feet*) were presented to the infants either as targets or non-targets. Target words are the words that appeared in the familiarisation passages, i.e., for infants who heard passages that contained 'cup' or 'dog', 'cup' and 'dog' were target words while 'bike' and 'feet' were non-targets. In a given test trial, infants heard eight different tokens of a single word (e.g., eight repetitions of 'cup'). The test trials alternated between target and non-target words. For example, infants who were familiarised with passages containing 'cup' and 'dog' could hear any one of the following sequences: (1) 'cup', 'bike', 'dog', 'feet', (2) 'dog', 'feet', 'cup', 'bike', (3) 'bike', 'dog', 'feet', 'cup', or (4) 'feet', 'dog', 'bike', 'cup'. The test phase consisted of two blocks of four trials—each block contained each of the two target and the two non-target words, resulting in a total of 8 trials. Each test block of target and non-target trials was on average 64.22s (therefore, the duration of Test Phase 1 was on average 128.44s).

In Familiarisation Phase 2, the two familiarisation passages were presented once more (mean duration of this phase was 49.45s). Following this, in Test Phase 2, a single block of test trials was presented. Test Phase 2, consisting of 4 x 8 = 32 individual words, was on average 64.22s in duration (see Figure 1a).

Phase 2 (Familiarisation and Test) was added to the paradigm for a combination of three reasons. Firstly, Phase 2 was added to ascertain the robustness (in terms of temporal

persistence) of infants' word segmentation. Secondly, Phase 2 was added to ensure that infants had sufficient exposure to the familiarisation passages. In typical word segmentation studies, a version of the head turn preference paradigm is employed in which familiarisation stimuli continue to play until the infant reaches a certain listening criterion (e.g., Jusczyk & Aslin, 1995; Hollich et al., 2005) in order to ensure that infants actively listen for a certain duration. For example, the familiarisation phase in the Jusczyk and Aslin study (1995) continued until the infant accumulated at least 45s of listening time per familiarisation passage. Thirdly, Phase 2 was added because, as infant gaze patterns are of interest, it was necessary for stimulus presentation durations to be kept constant across infant participants to ensure that the overall duration of the stimuli in each phase was sufficient to capture enough gaze data. Therefore, Familiarisation Phase 2 and Test Phase 2 were added in this study first to examine the temporal robustness of word segmentation, second to ensure that infants had a sufficient amount of exposure to the familiarisation passages, and third to provide sufficient and equivalent data between participants for gaze data analyses. A single block of test trials was used in the second phase to minimise the overall duration of the test session.

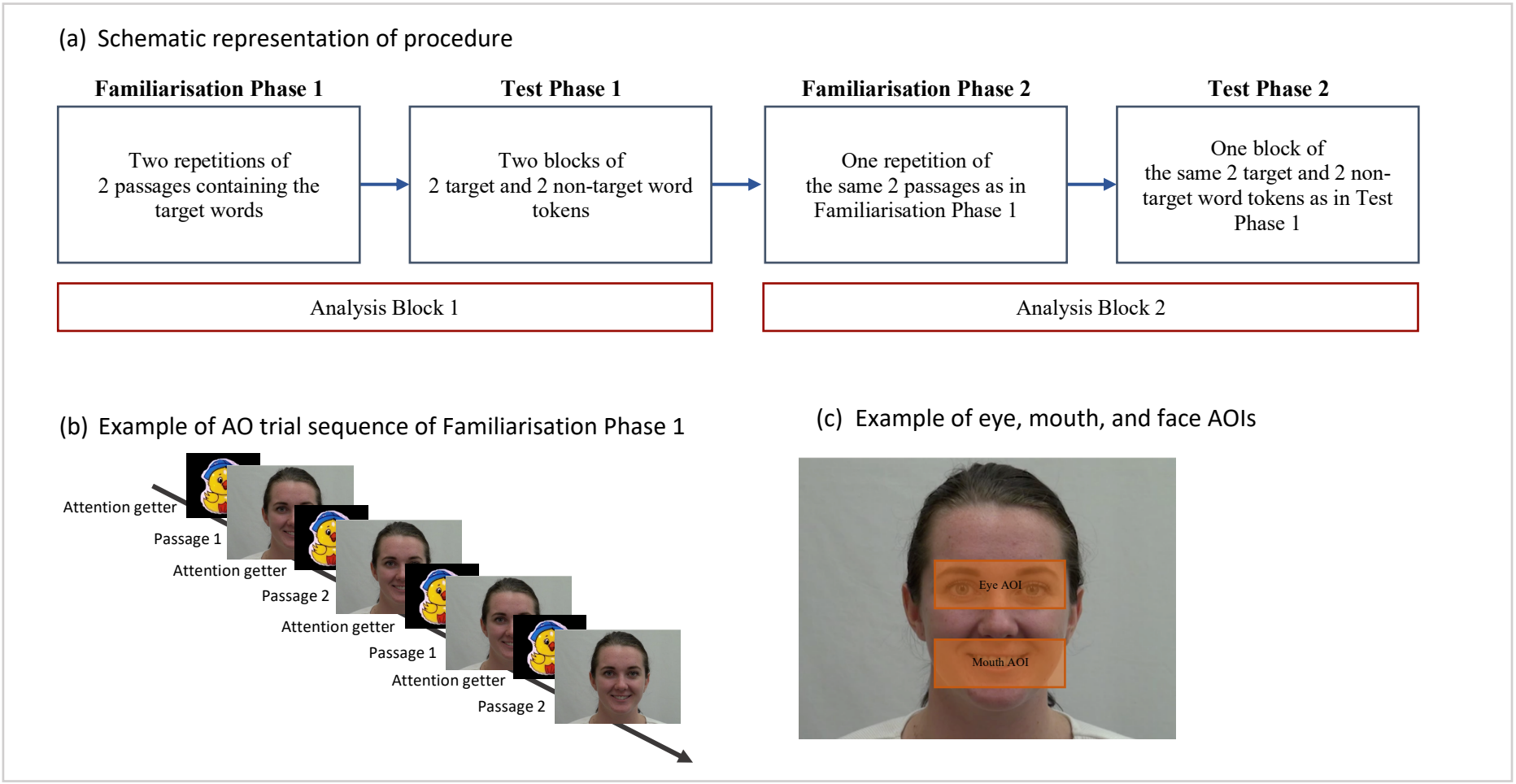

Figure 1. (a) schematic representation of the procedure, (b) an example of a trial sequence (AO Familiarisation Phase 1) illustrating an attention getter preceding each trial, and (c) an example of the dynamic eye, mouth and face AOIs that were defined for each trial (and frame for AV stimuli).

1.4.1 Auditory-only condition

Trials were initiated once infants attended to the attention getter. During familiarisation, a static image of the speaker's face appeared on the screen while the auditory recordings of the passages were played through to completion regardless of infant gaze behaviour.

Each test phase began immediately after each familiarisation phase. Once the infant attended to the attention-getter, the experimenter initiated the test trials. The auditory recording of the word tokens paired with a static image of the speaker's face played until completion. The trial ended with the static face fading from the screen. Looking times to the screen during each test trial were recorded as an index of infants' relative preference for the target and non-target words.

1.4.2 Auditory-visual condition

The procedure for the auditory-visual condition was identical to the AO condition except that familiarisation and test stimuli were presented in auditory-visual modality—dynamic videos of the speaker reciting the passages (in familiarisation) and word tokens (in test) were presented.

1.5. Eye-Tracking Analyses

Two dynamic areas of interest (AOIs) for the speaker's eye and mouth regions were defined. The AOIs for the eye and mouth regions are of equal sizes (343 by 153 pixels) (Figure 1). Looks to these two defined regions plus looks to the screen in general were collected and recorded in Tobii Studio v 3.4.5.

Raw looking times were extracted using *dplyr* (Wickham, Francois, Henry, & Muller, 2020) and *tidyr* (Wickham & Henry, 2020) packages in custom R scripts (R Core Team,

2020). Proportions of total looking times (PTLs) to these regions were then calculated from the raw looking times for each familiarisation and test trial. The PTLs of interest were:

(1) mouth vs total looks $[\frac{fixation\ duration\ to\ mouth}{total\ fixation\ duration\ to\ screen}]$,

(2) eyes vs total looks $[\frac{fixation\ duration\ to\ eyes}{total\ fixation\ duration\ to\ screen}]$, and

(3) mouth preference $[\frac{attention\ to\ mouth}{(attention\ to\ mouth + attention\ to\ eyes)}]$.

Additionally, overall attention on each trial was defined as $[\frac{total\ fixation\ duration\ to\ screen}{trial\ duration}]$.

Equations (1) and (2) were used to examine whether gaze behaviour differed between conditions for each trial type, and equation (3) was used to investigate whether individual differences in attention to the mouth (vs. the eyes) influence speech segmentation performance. A drawback of using PTL is that it does not provide us with information on how infants' looking behavior may vary across the trial time window. It may be the case that infants shift their gaze from one AOI to another, and such shifts are not reflected in PTL calculations. Therefore, additional time-course analyses were performed with the *eyetrackingR* (Dink & Ferguson, 2018) package to examine whether differences in looking behavior between the two groups of infants emerge differentially across time.

2. Results

Data were analysed in four parts. First, preliminary analyses were conducted to examine whether the two groups of infants differed in their attention to the screen during familiarisation. Second, speech segmentation was examined by comparing attention to target with non-target test trials in two ways: (i) to investigate recognition of target versus non-target *per se*, *t*-tests of infants' *d* scores in each test block against chance (zero) were used, and (ii) Analysis of Variance was used to assess whether the degree of infants' preference for the familiarised target word in each test phase varied as a function of the stimulus modality

(AO vs. AV). Third, regression analyses were conducted to investigate whether looking behaviour during familiarisation predicts segmentation performance at test. Finally, time course analyses were conducted to explore infants' word segmentation as a function of their looking behaviour to the speaker's eyes and mouth over time. Familiarisation Phase 1 and Test Phase 1 were analysed as Block 1, and Familiarisation Phase 2 and Test Phase 2 were analysed as Block 2. All statistical analyses were conducted in R (R Core Team, 2020) using *eyetrackingR* (Dink & Ferguson, 2018) and R *stats* (R Core Team, 2020) packages. All the frequentist analyses conducted to compare group performance in this study were supplemented with additional Bayesian analyses (ANOVAs and *t*-tests) using JASP (JASP Team, 2020). The full output of Bayesian ANOVAs is presented in the Supplementary Materials. This additional approach was taken to evaluate the likelihood that our evidence supports our experimental hypothesis (significant difference in performance between the AO and AV groups) or the null hypothesis (comparable performance between the AO and AV groups). According to the classification scheme by Jeffreys (1939), BF_{10} values above 1 provide evidence for the experimental hypothesis (H1), with BF_{10} values < 3 considered as weak or anecdotal evidence for H1, values between 3 to 10 as moderate evidence for H1, and values > 10 as strong evidence for H1. At the other end, values below 1 provide evidence for the null hypothesis (H0), with BF_{10} from 1/30 to 1/10 being interpreted as strong evidence for H0, values from 1/10 to 1/3 as moderate evidence for H0, and values from 1/3 to 1 as weak or anecdotal evidence for H0 (Etz & Wagenmakers, 2017; Kass & Raftery, 1995). In other words, BF_{10} values that are between 1/3 and 3 are interpreted as not offering conclusive evidence for either H0 or H1.

2.1. Attention during familiarisation

A 2 (Condition: AV vs. AO) by 2 (Block: 1 vs. 2) mixed-measures analysis of variance (ANOVA) was conducted with attention during familiarisation as the dependent

variable. The main effects of Condition and Block were both significant (Condition: $F(1, 70) = 11.71$, $p = .001$, $\eta_p^2 = .14$; Block: $F(1, 70) = 10.89$, $p = .002$, $\eta_p^2 = .13$), whereas the Condition x Block interaction was not ($F(1, 70) = 0.77$, $p = .38$, $\eta_p^2 = .01$). Infants in the AV condition attended more to the screen during familiarisation than their counterparts in the AO condition, and both groups of infants attended more in Block 1 than in Block 2 (Figure 2a).

2.2. Test Trials: Speech segmentation performance

Speech segmentation performance was quantified by the difference in attention (that is, the overall proportion of looks to the screen) to Target versus Non-Target test trials. Each participant was given a difference score *d* (proportion looking to Target minus Non-Target trials) calculated for each block: larger difference scores index stronger preferences for Target trials, hence indicating better speech segmentation performance. These difference scores were used instead of raw looking times to account for the group difference in overall attention to the screen identified in the familiarisation phase (see 3.1 – infants in the AV condition attended more to the screen than infants in the AO condition).

To examine whether there was successful speech segmentation within each group of infants, *d* scores for each group were compared against zero via one-sample *t*-tests. In the AO condition *d* scores were significantly greater than zero in Block 1 but not Block 2 (Block 1: $t(17) = 2.68$, $p = .016$, Cohen's $d = 0.63$; Block 2: $t(17) = 0.76$, $p = .46$, Cohen's $d = 0.19$), whereas in the AV condition *d* scores were significantly greater than zero in *both* blocks (Block 1: $t(18) = 2.51$, $p = .02$, Cohen's $d = 0.66$; Block 2: $t(18) = 2.49$, $p = .02$, Cohen's $d = 0.59$). Additional one-tailed Bayesian one-sample *t*-tests indicated moderate evidence in favour of the alternative hypothesis (*d* scores > 0) for the AO group in Block 1 ($BF_{10} = 7.12$), and for the AV group in Block 1 ($BF_{10} = 5.37$) and Block 2 ($BF_{10} = 5.19$), and inconclusive evidence for H0 or H1 in the AO group in Block 2 ($BF_{10} = 0.48$).

Next, a 2 (Condition: AV vs. AO) x 2 (Block: 1 vs. 2) mixed-measures ANOVA was conducted with *d* scores as the dependent variable. Neither main effect, Condition or Block, nor the Condition x Block interaction were significant (Condition: $F(1, 70) = 1.57$, $p = .21$, $\eta_p^2 = .02$; Block: $F(1, 70) = 0.18$, $p = .68$, $\eta_p^2 = .003$; Condition x Block: $F(1, 70) = 1.13$, $p = .29$, $\eta_p^2 = .02$). Figure 2b illustrates attention to Target and Non-Target trials and Figure 2c depicts *d* scores of infants in AO and AV conditions for Blocks 1 and 2. Supplementary Bayesian analyses indicated inconclusive support for H0 or H1 (AV ≈ AO; Condition: $BF_{10} = 0.48$).

2.3. Regression analyses: Does gaze behaviour differ as a function of condition?

Separate Condition (AO vs. AV) x AOI Type (Eyes vs. Mouth) x Block (Block 1 vs. Block 2) mixed-measures ANOVAs for each trial type (Familiarisation, Target, and Non-Target) were conducted to examine whether gaze behaviour differed between conditions for each trial type (Figure 3). None of the main effects or interactions were significant for either trial type (see Figure 2 for means and standard deviations; all Fs < 2.07, all ps > .15, full output of these analyses is presented in Appendix B). Bayes Factor values for Condition (Familiarisation: $BF_{10} = 0.39$; Target: $BF_{10} = 0.37$, Non-Target: $BF_{10} = 0.41$) and Condition x AOI Type (Familiarisation: $BF_{10} = 0.08$; Target: $BF_{10} = 0.07$, Non-Target: $BF_{10} = 0.36$) (see Tables S3-5 for full output) did not provide conclusive evidence for whether infant gaze behaviour differed between conditions across trial types.

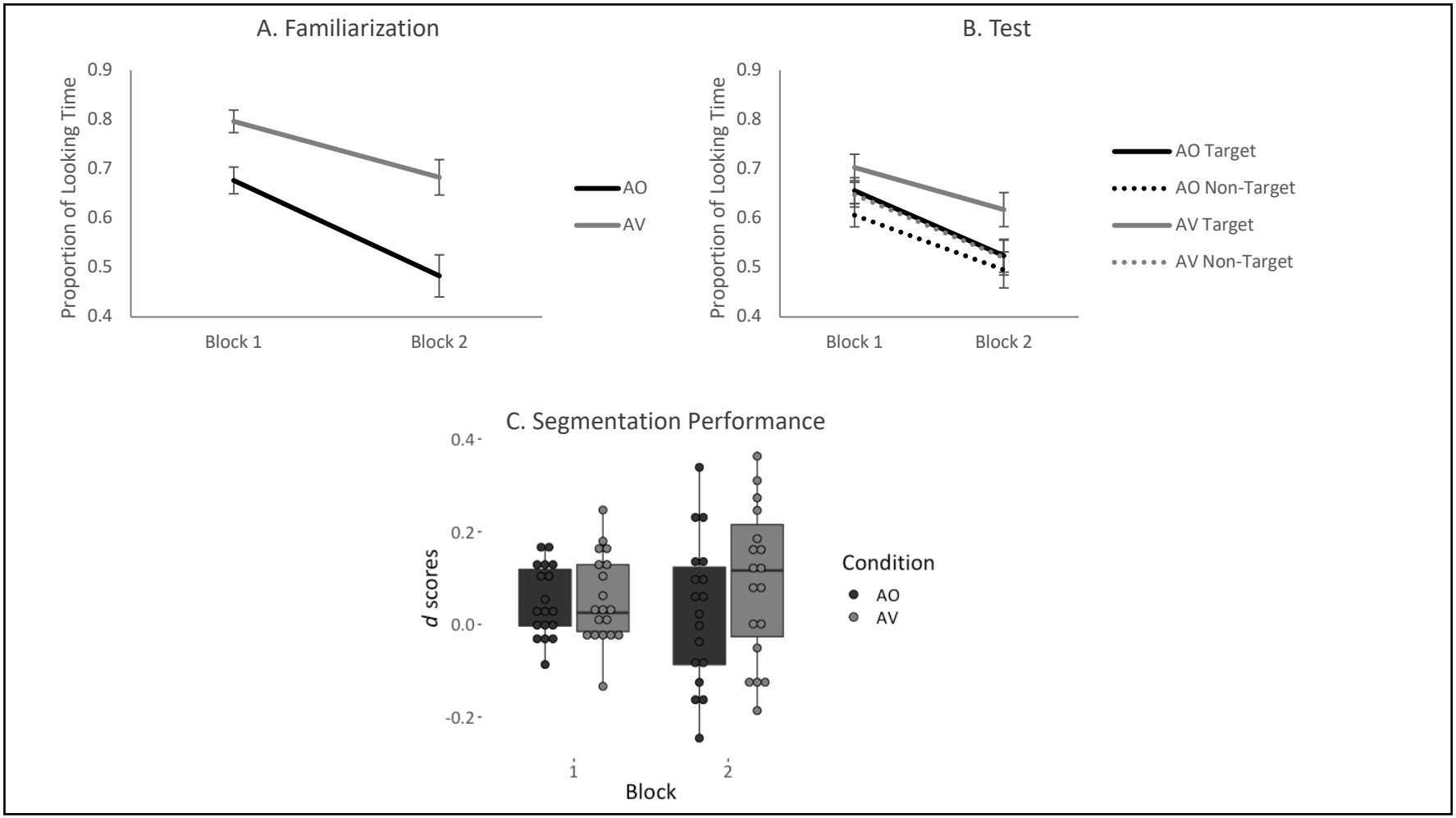

Figure 2. (a) Line graph representing attention during Familiarization Phase, (b) line graph depicting attention to Target and Non-Target trials during Test Phase, and (c) box plot illustrating individual *d* scores that quantify word segmentation performance. Error bars represent standard error means.

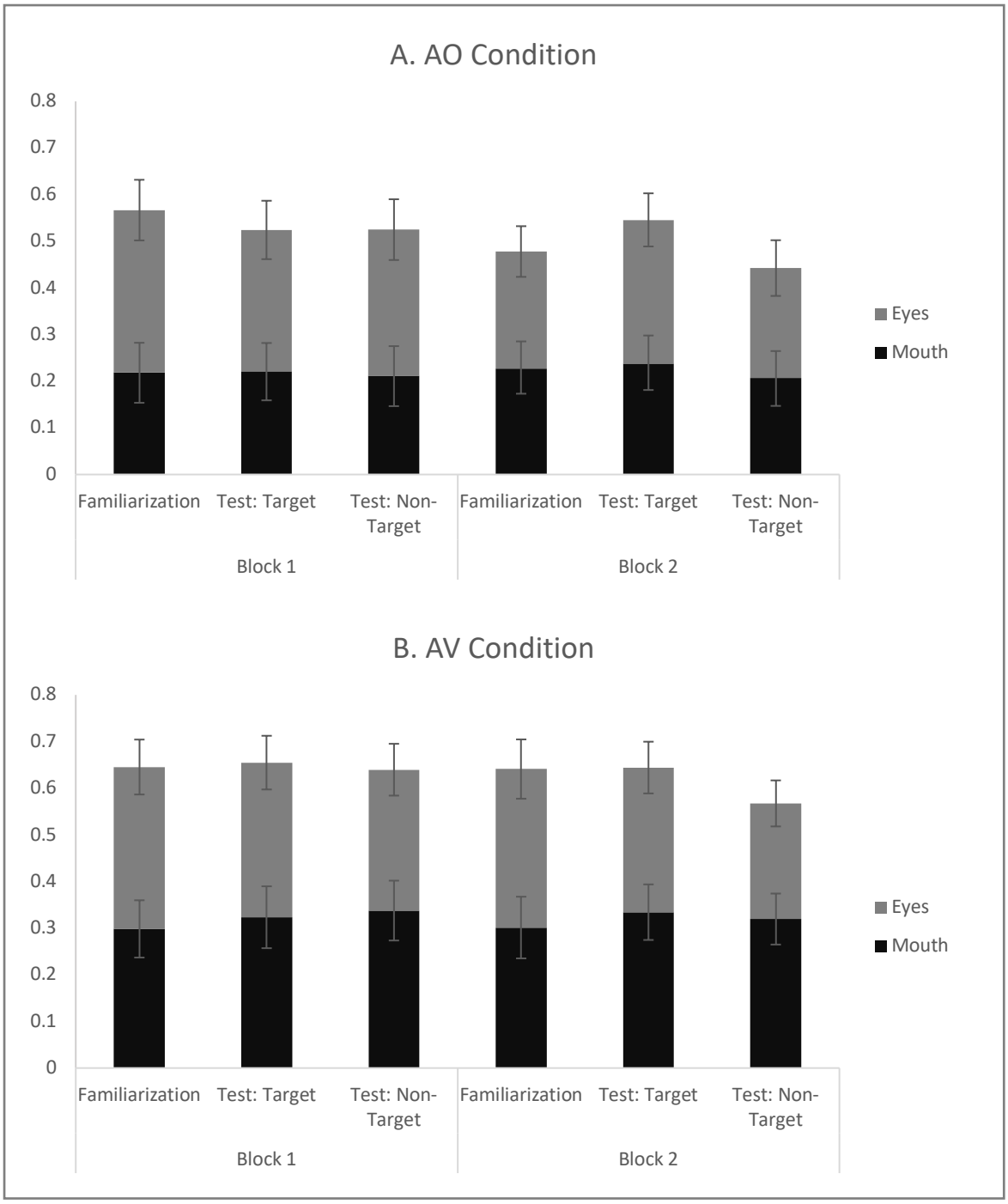

Figure 3. Bar graphs depicting means of attention to the eye and mouth regions of the speaker across trial types for (a) AO, and (b) AV conditions. Error bars represent standard errors.

2.5. Regression analyses: Do individual differences in gaze behaviour influence speech segmentation?

To investigate whether individual differences in gaze behaviour influence speech segmentation, a hierarchical linear regression analysis was conducted with Attention (proportion of looks to screen during familiarisation), Condition, and Block as predictor variables, and segmentation performance (*d* scores) as the outcome variable. Attention during familiarisation was entered in Step 1, and Condition and Block were entered in Step 2. Condition and Block were entered into the model in the second step to evaluate whether Condition and Block moderated the relationship between attention during familiarisation and segmentation performance. After entering Attention in the first step of the regression analysis, the model was not significant ($F(1, 72) = 1.22$, $p = .27$), indicating that attention to the screen during familiarisation did not explain a significant amount of variance in segmentation performance ($B = .08$ [-0.06, 0.21], $p = .27$). The addition of Condition and Block to the model was not significant ($\Delta R^2 = .02$, $\Delta F\ (2, 70) = 0.69$, $p = .51$, $\eta_p^2 = .02$), suggesting that Condition and Block were not statistically significant moderators of the relationship between attention to the screen during familiarisation and segmentation performance (see Table 1 for the results).

Table 1

Summary of Hierarchical Regression Analysis Predicting Segmentation Performance

Predictors	***B***	**95% CI [lower, upper]**	***t***	***F***	***p***	R^2	ΔR^2
Step One				1.22	.27	.02	.02
Attention during Familiarisation	.08	[-0.06, 0.21]	1.11		.27		
Step Two				0.86	.47	.04	.02
Attention during Familiarisation	.07	[-0.09, 0.23]	0.92		.36		
Condition	.03	[-0.04, 0.09]	0.81		.42		
Block	.02	[-0.04, 0.09]	0.72		.47		

To investigate whether individual differences in attention to the mouth (vs. the eyes) influence speech segmentation performance, a hierarchical regression analysis was conducted with mouth preference during familiarisation and block as predictor variables and segmentation performance (*d* scores) as the outcome variable. This analysis only included data from infants in the AV condition. Mouth preference during familiarisation was entered in Step 1, and Block was entered in Step 2 to examine whether the relationship between segmentation performance and mouth preference was moderated by the analysis block. When mouth preference during familiarisation was entered in the first step of the regression analysis, the model was not significant ($F(1, 35) = 0.28$, $p = .60$, $\eta_p^2 = .008$), indicating that the variance in speech segmentation performance of infants in the AV condition cannot be explained by their mouth preference during familiarisation. The addition of Block in Step 2 was not significant ($\Delta R^2 = .04$, $\Delta F\ (2, 34) = 1.25$, $p = .27$, $\eta_p^2 = .07$), suggesting that Block was not a statistically significant moderator of the relationship between segmentation performance and mouth preference during familiarisation (see Table 2 for the results).

Table 2

Summary of Hierarchical Regression Analysis of Mouth Preference and Segmentation Performance (Auditory-Visual Condition Only)

Predictors	***B***	**95% CI [lower, upper]**	***t***	***F***	***p***	R^2	ΔR^2
Step One				0.28	.60	.008	.008
Mouth Preference (Familiarisation)	.04	[-0.10, 0.17]	0.53		.60		
Step Two				0.77	.47	.04	.03
Mouth Preference (Familiarisation)	.04	[-0.10, 0.17]	0.56		.58		
Block	.05	[-0.04, 0.15]	1.12		.27		

2.6. Time course analyses: Do gaze patterns change over time?

The lack of statistically significant relationship between word segmentation and gaze behavior was unexpected given that it has been previously established that infants attend to the eye and mouth regions differentially depending on whether the face is producing speech or not (Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2013). One possible explanation for these unexpected findings is that speech segmentation performance here is indexed by difference scores which were derived from PTLs averaged across time, and this may mask any potential temporal variations in looking behavior (Birulés et al., 2022). To explore this possibility, the time course of infants' looking behaviour was examined. Gaze data were aggregated into 20 time-bins (~1.24s per bin for Familiarisation trials, and ~0.80s per bin for Target and Non-Target trials) and compared sequentially using ANOVAs to identify any particular time period(s) during which infants' looking patterns to the face, eye and mouth AOIs diverged. The false discovery rate (FDR) method (Benjamini & Yekutieli, 2001) was used to correct for multiple comparisons. For these analyses, PTLs to the face, eyes, and mouth AOIs were derived from equations (1-3) in which total fixation duration serves as the denominator.

First, infants' attention during Familiarisation, Target and Non-Target trials was examined. Separate time course analyses for attention to the eyes and for attention to the mouth revealed that the only significant differences between AO and AV groups were for attention to the mouth (Figure 4). As can be seen in Figure 4, attention to the mouth region was generally greater in the AV than in the AO condition early in the time course but this difference gradually declined over time. In Block 1, the greater attention to the mouth in the AV over the AO group was significant for Non-Target trials from 0.00 to 6.42s (ps < .05 for all time bins). In Block 2, the initial AV > AO difference in attention to the mouth region was significant from 0.00-5.62s (ps < .05 for all time bins) for Non-Target trials, and for Target

trials from 0.00-5.63s (ps < .02 for all time bins). Interestingly, there were no significant differences between attention to the mouth and the eyes regions in separate time-course analyses of the AO or the AV group data (all ps > .41), suggesting that the specific times at which the AV group attended more to the mouth than the AO group was a product of the comparative looking behaviour of the two groups.

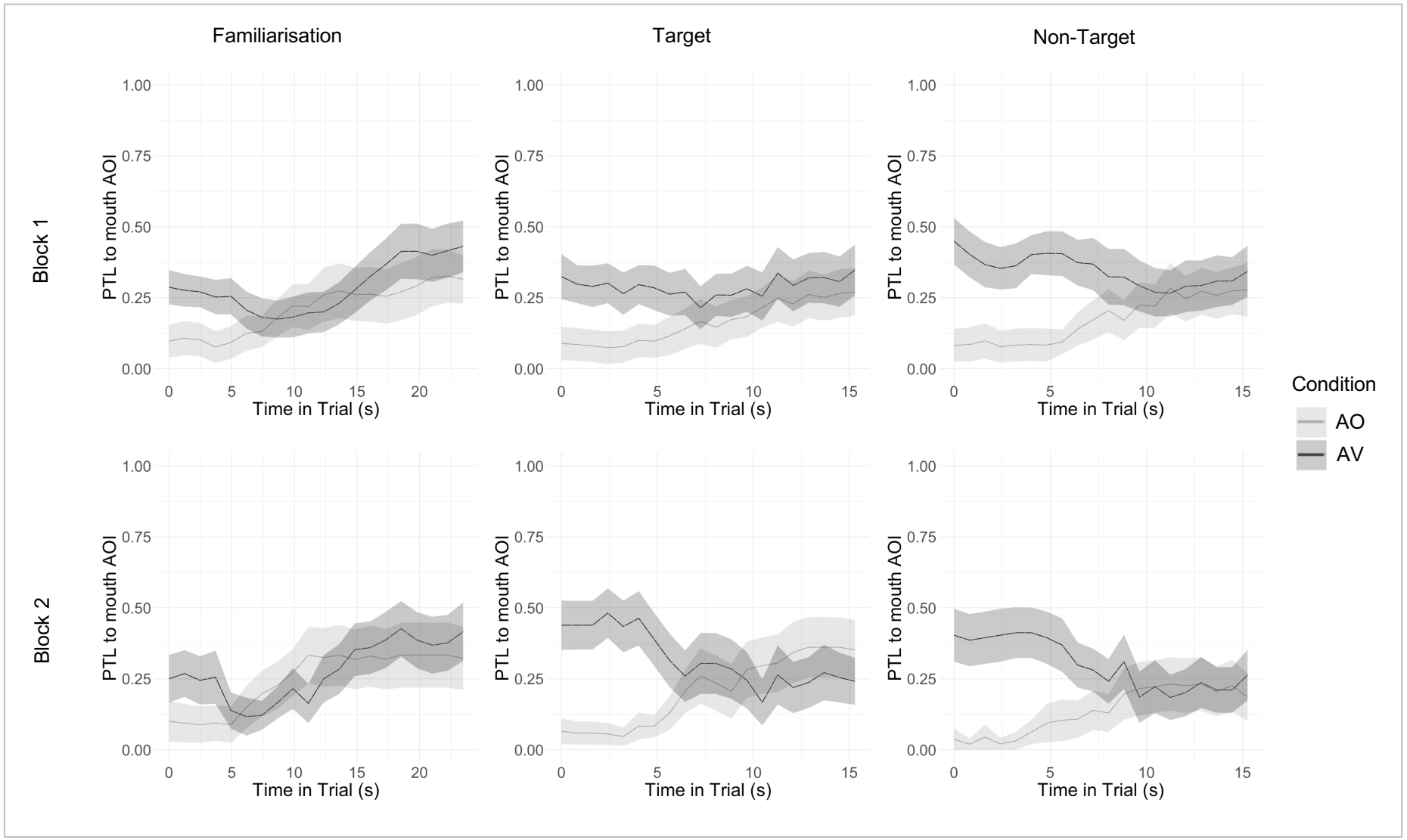

Figure 4. Time courses of attention to the mouth AOI for each trial type. Time units are in seconds.

3. Discussion

By 7.5 months of age, infants develop the ability to extract words from continuous speech. At this age, individual infants continue to vary widely in their performance on speech segmentation tasks, and these individual differences are significant predictors of later linguistic abilities. This study aimed to assess whether 7.5-month-old infants' performance in a speech segmentation task in clear speech can be augmented by access to visual speech information, as it has been shown in a previous study on auditory-visual segmentation from speech in noise (Hollich et al., 2005). Over and above that auditory-visual study and all other studies of speech segmentation, this study included two novel aspects; firstly it assessed whether individual differences in segmentation performance are related to infants' attentional patterns to visual information from a speaker's face, and secondly it included time course analyses of infants' recognition of target and non-target words. To these ends, two groups of 7.5-month-old infants were tested. One group was presented with a static image of the speaker's face paired with auditory recordings of Familiarisation passages, then Target and Non-Target word token tests, while the second was presented with dynamic videos of the speaker's talking face reciting the passages and word tokens, then Target and Non-Target word token tests.

The first set of analyses revealed that in familiarisation trials infants in the auditory-visual condition attended more to the screen than infants in the auditory-only condition. In the test phase, analyses using frequentist and Bayesian approaches yielded inconclusive evidence about the effects of test condition on infants' performance. However, follow up analyses of infants' behavior in the first and second test blocks showed evidence for the hypothesis that access to visual information led to more persistent segmentation performance: Infants in both the AO and AV groups showed successful segmentation in the first test block, but in the second test block only the AV group showed successful segmentation.

The third set of analyses investigated whether gaze behavior influences speech segmentation, and this revealed three unexpected findings. First, both AO and AV groups of infants attended similarly to the speaker's eye and mouth regions in all three types of trials (Familiarisation, Target and Non-Target). Second, attention to the screen during familiarisation did not predict speech segmentation performance, and this relationship was not moderated by condition or block. This suggests that successful speech segmentation performance in Block 2 by infants in the AV condition cannot be entirely explained by their greater attention to the screen during familiarisation. Third, mouth preference did not predict speech segmentation performance for infants in the AV condition. These results were surprising because previous studies have found that infants attend more to the mouth region than the eye region of a talking face and this shift in attention increases with age (e.g., Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2013), leading to the postulation that infants are able to make use of visual (e.g., mouth movements) and auditory modalities to segment speech (Kitamura et al., 2014).

To ensure that the null findings here were not due to the effects being masked by the use of PTLs which are averaged across time, our final set of analyses investigated the time course of infants' looking behaviour. There was no AO - AV group difference in looking behavior to the face generally over time, indicating that infants' attention to the face was similar regardless of whether a static image or a dynamic talking face was presented to them. However, analyses of gaze behaviour to the eye and mouth regions uncovered an interesting pattern of results. First, as was hypothesised, infants in the AV condition attended more to the mouth region than did infants in the AO condition; second, interestingly, this greater attention to the mouth was only at specific time periods: during the first five seconds of Non-Target trials in both Blocks 1 and 2, and during the first five seconds of Target trials in Block 2. These results provide a nice and theoretically cogent parallel with the fact that infants in the

AV but not the AO condition showed word segmentation in *both* Blocks 1 and 2. Third, there was a general decrease (across both Target and Non-Target trials) in attention to the mouth over the time course for the AV group, but the opposite, a general increase in attention to the mouth over the time course for the AO group. This suggests that infants extract the information required for word segmentation early in AV presentations, whereas in AO presentations infants may seek linguistic information from the speaker's mouth over time (even though the mouth was not moving) because they understand from prior experience that the mouth is the main source of linguistic information (Tenenbaum et al., 2015). Interestingly, attention to the mouth AOI over time did not differ between Target and Non-Target trials even for infants in the AV condition, suggesting that infants' looking behavior to the mouth region was not influenced by the familiarity of the words. Attention to visual speech cues during the production of *both* the target and non-target words, i.e., to all possible candidates, appears to provide useful information for word segmentation.

Given that infants in both the AV and AO conditions showed word segmentation in the first block, it was somewhat unexpected that only the AV infants showed word segmentation in the second block. This result is likely to be an effect of a decay in attention to the stimuli and task. Although infants' attention decreased from Block 1 to Block 2 in both conditions, infants in the AV condition attended more to the screen during re-familiarisation in Block 2 compared to infants in the AO condition (Fig. 2A). Taking this into consideration, it is not only the case that the AV group attended more to the test trials in Block 2, they also likely benefited more than did the AO group from the re-familiarisation phase in Block 2.

When the successful segmentation performance by AV infants in Block 2 is coupled with the finding that attention does not predict segmentation performance even for the AV group of infants, the scope of potential pathways via which visual speech information might augment infant speech perception is narrowed. First, visual speech cues may provide an

additional modality through which information is processed, especially since infants' memories for newly-segmented words may decay over time (Karaman & Hay, 2018). As phonological short-term memory has been shown to be fundamental for successful speech segmentation (Minagawa, Hakuno, Kobayashi, Naoi, & Kojima, 2017), and in accordance with the dual coding theory of memory (Paivio, 1991), it stands to reason that the visual speech cues available to infants in the AV condition strengthen the memory trace of the primitive phonological representation of newly-segmented words, resulting in the sustained segmentation found in the second block only for the AV group. Second, as the videos presented to the AV group involved a close-up of the speaker talking in IDS, infants in the AV group may have picked up socially meaningful cues (e.g., eye contact), which may have enhanced their segmentation performance. This is in line with neurophysiological findings that by 5 months of age infants are already sensitive to ostensive signals such as visual (gaze) and auditory (speech, IDS) information that is directed specifically to them (Parise & Csibra, 2013), and with behavioural findings that communicative social contexts such as contingent responding (e.g., Goldstein & Schwade, 2008; Goldstein, Schwade, Briesch, & Syal, 2010; Mackensen & Grossmann, 2015) and gestures (Yoon, Johnson, & Csibra, 2008) foster learning in infants. Anecdotal evidence from observing infants' behaviour during test sessions supports this possibility as infants in the AV group tended to display communicative behaviour such as reaching for the screen and babbling in response to the video presentations. Third, the visual cues from an IDS speaker involve exaggerated facial cues (Chong, Werker, & Russell, 2003; Green, Nip, Wilson, Mefferd, & Yunusova, 2010) that are temporally synchronised with the auditory exaggerations of prosody in this register (Shepard, Spence, & Sasson, 2012). Here such multimodal information from the speaker's face and voice may have facilitated the formation of higher level, amodal representations of speech that support language perception (Bahrick & Lickliter, 2000; Gogate & Bahrick, 1998; Gogate, Bolzani,

& Betancourt, 2006). Thus, over and above the increased attention that may stem from the exaggerated facial expressions in IDS, this directed attention may provide stronger grist for the formation of amodal (auditory + visual) representations that are more robust in terms of temporal longevity of word segmentation in the AV group. Examining whether the visual speech benefit found in this study is also evident when adult-directed speech is presented may further clarify the role of visual speech cues in infant language perception since IDS plays an important facilitatory role in auditory-only speech segmentation (Thiessen, Hill, & Saffran, 2005) and word learning studies (e.g., Graf Estes & Hurley, 2013).

An intriguing alternative account of our findings is that the observed group difference in the second test block reveals a change in infants' listening preferences rather than more persistent segmentation in the AV compared to the AO group. It has long been known that infants typically exhibit a familiarity preference in tasks involving spoken word recognition (Jusczyk, Friederici, Wessells, Svenkerud & Jusczyk, 1993; Jusczyk, Houston & Newsome, 1999b). However, it is noteworthy that familiarity preferences are associated with partial processing, they are observed in younger infants, and occur under challenging experimental settings or in response to complex stimuli (Burnham & Dodd, 1998; Houston-Price & Nakai, 2004; Hunter & Ames, 1988). When presented with a repeated stimulus, infants can transition from an initial familiarity preference to a novelty preference (i.e., preference for non-target stimuli) within the duration of a single experimental session indicating that they have fully processed and encoded the familiar stimulus and freed attentional and processing resources to encode a novel stimulus (Fantz, 1964; Roder, Bushnell, & Sasseville, 2000). More specifically, Roder et al. (2000) gave 4.5-month-old infants a series of trials showing two stimuli side by side, one that remained the same throughout repetitions, and the other that changed on each trial in two conditions faces and objects. They found that the vast majority of infants showed a familiarity to novelty shift. This shift occurred later for faces than objects

(11 versus 8 trials) and in fewer infants for faces than objects. Put together with the suggestion by Wagner and Sakovits (1986) (see also Burnham & Dodd, 1998) that cross-modal stimuli should more readily elicit a familiarity preference, and given that the AV condition was cross-modal and the AO condition unimodal, it could be argued that as the stimuli here were faces then a familiarity preference should persist for some time (Roder et al., 2000), and as the AV condition is of course cross-modal and the AO condition is not, infants in the AO group may have processed the information in familiar presentations more quickly than did infants in the AV condition, and by the second block, had begun to transition to a novelty preference for the non-target items. However, inspection of our data does not provide compelling evidence for this possibility: the difference between the number of infants who transitioned from a familiarity (target) to a novelty (non-target) response between the first and second blocks in the AO (6 out of the 18 infants) and the AV condition (3 out of the 19 infants) was not significant (χ^2 (1, N = 37) = 1.55, p = .21).

We also note that while the AV stimuli may be considered more complex, previous studies have found that infants benefit from visual speech information (e.g., Hollich et al., 2005; Teinonen et al., 2008), so it would be equally reasonable to expect the opposite; that the provision of visual speech information in the AV condition would allow infants to process linguistic information more efficiently resulting eventually in a novelty preference (in which case *d* scores would be negative) or at least a transition from a familiarity to a novelty preference (in which case *d* scores would not deviate from zero). More efficient encoding in the AV condition could also be predicted given that infants attended more to familiarisation stimuli in AV than in the AO condition. This is also not supported by the data—*d* scores for the AV group were significantly more positive than zero even in Block 2, indicating that the AV group of infants continued to show a familiarity preference throughout the experiment.

Despite these performance differences between infants in the AO and AV conditions in the second test block of the speech segmentation task, it is noteworthy that our analyses of variance did not reveal significant effects of condition on infants' preference for target words. This is contrary to the only other extant comparison of AO and AV word segmentation Hollich et al. (2005), in which the auditory recordings were always paired with background noise and 7.5-month-olds successfully segmented speech in the AV, but not in the AO condition. It may be that infants can segment words from clear speech quite effectively by 7 months and that the addition of visual speech information does not provide substantial incremental benefit. Our additional Bayesian analyses do not lend support to either the null or the research hypothesis and call for more research to be conducted.

As we cannot conclusively determine that our data demonstrate a lack of differences between the two conditions, we consider here some limitations of our study and directions for future research that could lead to more conclusive findings. First, our study included a conservative sample size, so it is possible that the strength of support for the null hypothesis would increase with a larger sample. Second, the inclusion of a second block of trials, Block 2, is not typical of a traditional segmentation paradigm. However, without the inclusion of Block 2, the difference in segmentation performance between AO and AV conditions would not have been uncovered. So, the inclusion of Block 2 provided greater insights into infant segmentation of auditory-visual speech, but as a second block is not usually included in segmentation studies, then no comparison is available in the literature, and conclusions from results in this block cannot be drawn with confidence. For instance, instead of reflecting stronger memory for newly segmented words, the Block 2 results may reflect segmentation based solely on the items in the brief refamiliarization phase. To directly measure memory trace, other more sensitive measures such as pupil dilation or mismatch negativity (MMN) amplitude could be included as these measures have been associated with recognition and

strength of memory trace in young infants (pupil dilation: Hellmer et al., 2018; MMN: Cheour et al., 2002). Third, it is possible that the benefit of visual speech information, which was observed in the presence of background noise (Hollich et al., 2005) can be detected in clear speech in younger infants whose speech segmentation skills are manifested less consistently than at 7 months of age. For example, English-learning (Jusczyk & Aslin, 1995) and German-learning (Höhle & Weissenborn, 2003) 6-month-olds are unable to recognise words embedded in passages after being familiarised with them (words-then-passages paradigm), but 6-month-olds do succeed in segmenting speech when the target word was placed next to their own name or a highly familiar word (Bortfeld et al., 2005). Additionally, 6-month-olds recognise the meanings of familiar/target words (Bergelson & Swingley, 2012), indicating that they have already isolated these words from the speech stream. These findings, coupled with the finding that newborns can integrate visual and auditory information (Guellaï, Streri, Chopin, Rider, & Kitamura, 2016), raise the possibility that with younger infants segmentation in AO may be significantly improved in AV, i.e., that the difference in *d* scores between infants in AO and AV conditions may be more pronounced.

In addition to including younger participants, future research would benefit from the use of neurophysiological techniques that have already been successfully used to assess speech segmentation in young infants (e.g., Kidd et al., 2018; Kooijman et al., 2009, Snijders et al., 2020). While looking times have been used widely in the literature (e.g., Altvater-Mackensen & Mani, 2013; Tsui et al., 2020; Thiessen & Erickson, 2013), they may not be ideal for comparing performance across modalities. A static photograph is less appealing than a dynamic video, so the segmentation ability in the AO group of infants in this study may have been underestimated since they may have attended auditorily to the stimulus while not looking at the screen, especially in the second test block. These effects should be explored in future studies by including conditions in which there are static familiarization phases

followed by dynamic test phases and vice versa. In this study, the same modality was used across familiarization and test trials with a central visual fixation paradigm as this enabled us to measure infants' gaze patterns to a speaking face at the same time as they were performing a speech segmentation task. It must be noted, however, that the central visual fixation paradigm differs from the traditional head-turn preference paradigm (HPP) and thus a direct comparison with traditional HPP studies is difficult.

Other future directions include introducing a 10-minute delay between familiarization and test to examine directly whether the effect of AV is to provide an additional modality that strengthens the memory trace of newly-segmented words, and isolating socially-meaningful cues by including conditions where the speaker's face is partially occluded (e.g., wearing a pair of shades to cover the eyes or a mask to cover the mouth) to investigate the relationship between these cues and segmentation performance.

Gaze behaviour to the eye and mouth regions was not associated with the degree of visual speech benefit in this task. This pattern of results differs from studies that have found a positive relationship between looking behaviour to the speaker's mouth and language development (e.g., Elsabbagh et al., 2013; Young, Merin, Rogers, & Ozonoff, 2009). Additionally, looking behaviour, when measured by PTL, was surprisingly equivalent for target and non-target words here. It was expected that infants would attend to the mouth longer in non-target test trials than in target test trials because infants were not exposed to the non-target words in the familiarisation phase and because it has been previously suggested that infants direct their attention to the mouth to gather linguistic information (Tenenbaum et al., 2013). A likely explanation for the equivalent looking behaviour during target and non-target trials (in PTL and the time-course analyses) is that the non-target words were not entirely novel— *cup*, *dog*, *bike*, and *feet* are words that infants commonly hear in their

everyday lives. Further work on looking behaviour using non-words would determine whether this is indeed the case.

Speech segmentation is arguably one of the most important skills in infants' language acquisition. Most infant segmentation studies have employed auditory-only stimuli despite the multimodality of speech. Studying the role of visual speech cues in infant language development has important implications particularly for infants who do not have access to a clear auditory signal. This study shows that 7.5-month-olds segment words from continuous speech in both AO and AV speech conditions, but that they do so in a more sustained fashion in AV. This is the first finding of its kind and so further work is required especially with younger infants, with a larger sample size, and with further manipulations of the presence/absence of visual information in familiarisation versus test. Results from this study suggest that at 7.5 months there is a visual speech benefit for word segmentation observed in terms of a sustained benefit over time. This is only the second examination of a visual speech benefit in the process of early speech segmentation, so critically, there is still much to learn about the mechanisms via which visual speech information enhances linguistic skills in infancy.

Acknowledgements

We are grateful to all infants and families who participated in this study. In addition, we are grateful for funding support as follows. The first author was supported by a scholarship from Western Sydney University and MARCS Institute for Brain, Behaviour and Development. The project itself as supported by the Hearing Cooperative Research Centre, Australia and its grant 82631, "The Seeds of Language Development" to the 2nd author. The 3rd author's work was supported by the Basque Government, Basque Country, Spain through the BERC 2018–2021 program and by the Spanish Ministry of Science and Innovation through the Ramon y Cajal Research Fellowship, PID2019–105528GA-I00.

The authors declare no conflicts of interest with regards to the funding source for this study.

Data used for analyses are available open access at https://osf.io/uwx9k/.

References

Altvater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science, 16*(6). doi:10.1111/desc.12071

Altvater-Mackensen, N., & Grossmann, T. (2015). Learning to match auditory and visual speech cues: Social influences on acquisition of phonological categories. *Child Development*, *86*(2), 362–378. http://doi.org/10.1111/cdev.12320

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). *Models of word segmentation in fluent maternal speech to infants*. In J. L. Morgan & K. Demuth (Eds.), (pp. 117–134). Lawrence Erlbaum Associates, Inc.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. http://doi.org/10.1073/pnas.1113380109

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,* 963-971. http://doi.org/10.1111/infa.12489

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. http://doi.org/10.1111/j.0956-7976.2005.01531.x

Burnham, D., & Dodd, B. (1998). Familiarity and novelty preferences in infants' auditory-visual speech perception: Problems, factors, and a solution. In C. Rovee-Collier (Ed.), *Advances in Infancy Research* (pp. 170–187).

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220. http://doi.org/10.1002/dev.20032

Butler, J., & Frota, S. (2018). Emerging word segmentation abilities in European Portuguese-learning infants: new evidence for the rhythmic unit and the edge factor. *Journal of Child Language*, *45*(6), 1294–1308. http://doi.org/10.1017/S0305000918000181

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436. http://doi.org/10.1371/journal.pcbi.1000436

Cheour, M., Čėponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., ... & Näätänen, R. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian journal of psychology*, *43*(1), 33-39. https://doi.org/10.1111/1467-9450.00266

Chong, S., Werker, J. F., & Russell, J. A. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development, 12*(3), 211–232. http://doi.org/10.1002/icd.286

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America*, *67*(4), 1323–1332. http://doi.org/10.1121/1.384185

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition*, *96*(3), 233–262. http://doi.org/10.1016/j.cognition.2004.08.005

Desjardins, R., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*(4), 187–203. http://doi.org/10.1002/dev.20033

Dink, J., & Ferguson, B. (2018). eyetrackingR [Computer software manual]. http://www.eyetracking-R.com (R package version 0.1.8.)

Elsabbagh, M., Bedford, R., Senju, A., Charman, T., Pickles, A., Johnson, M. H., The BASIS Team. (2013). What you see is what you get: contextual modulation of face scanning in typical and atypical development. *Social Cognitive and Affective Neuroscience*, *9*(4), 538–543. http://doi.org/10.1093/scan/nst012

Etz, A., & Wagenmakers, E.-J. (2017). J. B. S. Haldane's contribution to the Bayes factor hypothesis test. *Statistical Science, 32*, 313–329. http://doi.org/10.1214/16-STS599

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670. https://doi.org/10.1126/science.146.3644.668

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*, 133–149. https://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word-object relations. *Infancy*, *9*(3), 259–288. https://doi.org/10.1207/s15327078in0903_1

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. http://doi.org/10.1111/j.1467-9280.2008.02117.x

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, *15*(4), 362–391. http://doi.org/10.1111/j.1532-7078.2009.00020.x

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. http://doi.org/10.1111/infa.12006

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1529–1542. http://doi.org/10.1044/1092-4388(2010/09-0005)

Guellaï, B., Streri, A., Chopin, A., Rider, D., & Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1275–1281. http://doi.org/10.1037/xhp0000208

Hellmer, K., Söderlund, H., & Gredebäck, G. (2018). The eye of the retriever: developing episodic memory mechanisms in preverbal infants assessed through pupil dilation. *Developmental Science*, *21*(2), e12520. https://doi.org/10.1111/desc.12520

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*(3), 598–613. http://doi.org/10.1111/j.1467-8624.2005.00866.x

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. http://doi.org/10.1037//0096-1523.26.5.1570

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*(4), 341–348. http://doi.org/10.1002/icd.364

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, *6*(2), 122–127. http://doi.org/10.1111/1467-7687.00261

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research.* 5, 69–95.

Imafuku, M., & Myowa, M. (2016). Developmental change in sensitivity to audiovisual speech congruency and its relation to language in infants. *Psychologia*, *59*, 163–172. http://doi.org/ 10.2117/psysoc.2016.163

JASP Team (2020). JASP (Version 0.14.1). [Computer software]. Amsterdam, The Netherlands. Retrieved from http://jasp-stats.org.

Jeffreys, H. (1939). *Theory of probability* (1st ed.). Oxford University Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567. http://doi.org/10.1006/jmla.2000.2755

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science, 15*(4), 463–473. http://doi.org/ 10.1111/j.1467-7687.2012.1144.x

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. http://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*(3), 402-420. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999a). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465–1476. http://doi.org/10.3758/BF03213111

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999b). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3-4), 159–207. http://doi.org/10.1006/cogp.1999.0716

Karaman, F., & Hay, J. F. (2018). The longevity of statistical learning: When infant memory decays, isolated words come to the rescue. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(2), 221–232. http://doi.org/10.1037/xlm0000448

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. https://doi.org/10.2307/2291091

Kidd, E., Junge, C., Spokes, T., Morrison, L., & Cutler, A. (2018). Individual differences in infant speech segmentation: Achieving the lexical shift. I*nfancy, 23*(6), 770-794. https://doi.org/10.1111/infa.12256

Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PloS One*, *9*(10), e111467. http://doi.org/10.1371/journal.pone.0111467

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy, 14*(6), 591-612. https://doi.org/10.1080/15250000903263957

Kooijman, V. K., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, *4*, 25.

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. http://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*, 52. http://doi.org/10.3389/fpsyg.2016.00052

Marquis, A., & Shi, R. (2008). Segmentation of verb forms in preverbal infants. *The Journal of the Acoustical Society of America*, *123*(4), EL105–EL110. http://doi.org/10.1121/1.2884082

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. http://doi.org/10.1006/cogp.1999.0721

Minagawa, Y., Hakuno, Y., Kobayashi, A., Naoi, N., & Kojima, S. (2017). Infant word segmentation recruits the cerebral network of phonological short-term memory. *Brain and Language*, *170*, 39–49. http://doi.org/10.1016/j.bandl.2017.03.005

Mitchel, A. D., & Weiss, D. J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, *25*(4), 456–482. http://doi.org/10.1080/01690960903209888

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: Using facial cues to locate word boundaries in continuous speech. *Language, Cognition and Neuroscience, 29*(7), 771–780. http://doi.org/10.1080/01690965.2013.791703

Newman, R.S., Ratner, N.B., Jusczyk, A.M., Jusczyk, P.W., & Dow, K.A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology, 42* (4), 643–655. http://doi.org/10.1037/0012-1649.42.4.643

Newman, R., Rowe, M., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills

in language development. *Journal of Child Language, 43*(5), 1158–1173. http://doi.org/10.1017/S0305000915000446

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation. *Journal of Child Language*, *45*(1), 204–218. http://doi.org/10.1017/S0305000916000660

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(33), 255–287. https://doi.org/10.1037/h0084295

Parise, E., & Csibra, G. (2013). Neural Responses to Multimodal Ostensive Signals in 5-Month-Old Infants. *PloS One*, *8*(8), e72360. http://doi.org/10.1371/journal.pone.0072360

Pons, F., Bosch, L., & Lewkowicz, D. J. (2019). Twelve-month-old infants' attention to the eyes of a talking face is associated with communication and social skills. *Infant Behavior and Development*, *54*, 80–84. http://doi.org/10.1016/j.infbeh.2018.12.003

R Core Team. (2020). R: A language and environment for statistical computing. [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491–507. https://doi.org/10.1207/S15327078IN0104_9

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. http://doi.org/10.1126/science.274.5294.1926

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, *12*(4), 583–601. http://doi.org/10.1111/j.1467-7687.2009.00809.x

Schmale, R., Cristia, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, *15*(6), 650–662. http://doi.org/10.1111/j.1532-7078.2010.00032.x

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, *9*(6), 565–573. http://doi.org/10.1111/j.1467-7687.2006.00534.x

Shepard, K. G., Spence, M. J., & Sasson, N. J. (2012). Distinct Facial Characteristics Differentiate Communicative Intent of Infant-Directed Speech. *Infant and Child Development, 21*(6), 555–578. http://doi.org/10.1002/icd.1757

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental Science, 15*(4*), 482–495.* http://doi.org/10.1111/j.1467-7687.2012.01141.x

Snijders, T. M., Benders, T., & Fikkert, P. (2020). Infants segment words from songs—An EEG study. *Brain sciences, 10*(1), 39. https://doi.org/10.3390/brainsci10010039

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855. http://doi.org/10.1016/j.cognition.2008.05.009

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. http://doi.org/10.1111/j.1532-7078.2012.00135.x

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. 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. http://doi.org/10.1017/S0305000914000725

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*(6), 1093–1106. http://doi.org/10.1111/j.1551-6709.2010.01118.x

Thiessen, E. D., & Erickson, L. C. (2013). Discovering words in fluent speech: The contribution of two kinds of statistical information. *Frontiers in Psychology, 3*, 590. https://doi.org/10.3389/fpsyg.2012.00590

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. http://doi.org/10.1037/0012-1649.39.4.706

Thiessen, E. D., & Saffran, J. R. (2004). Spectral tilt as a cue to word segmentation in infancy and adulthood. *Perception & Psychophysics*, *66*(5), 779–791. http://doi.org/10.3758/BF03194972

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

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. http://doi.org/10.1016/j.jecp.2018.01.002

Tsui, A. S. M., Erickson, L. C., Mallikarjunn, A., Thiessen, E. D., & Fennell, C. T. (2020). Dual language statistical word segmentation in infancy: Simulating a language-mixing bilingual environment. *Developmental Science, 24*(3). http://doi.org/10.1111/desc.13050

Wagner, S. H., & Sakovits, L. J. (1986). A process analysis of infant visual and cross-modal recognition memory: Implications for an amodal code. *Advances in infancy research*, *4,* 195–217, 240–245.

Wickham, H., & Henry, L. (2020). tidyr: Tidy messy data [Computer software manual]. http://CRAN.R-project.org/package=tidyr (R package version 1.0.2)

Wickham, H., François, R., Henry, L. & Müller, K. (2020). dplyr: A grammar of data manipulation [Computer software manual]. https://CRAN.R- project.org/pacpkage=dplyr (R Package version 0.7.6.)

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-Induced Memory Biases in Preverbal Infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13690–13695. http://doi.org/10.2307/25464105

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. http://doi.org/10.1111/j.1467-7687.2009.00833.x

Appendix A

Taken from Jusczyk and Aslin (1995).

Target word	Six-sentence passages
Cup	The cup was bright and shiny.
	A clown drank from the red cup.
	The other one picked up the big cup.
	His cup was filled with milk.
	Meg put her cup back on the table.
	Some milk from your cup spilled on the rug.
Dog	The dog ran around the yard.
	The mailman called to the big dog.
	He patted his dog on the head.
	The happy red dog was very friendly.
	The dog barked only at squirrels.
	The neighborhood kids played with your dog.
Feet	The feet were all different sizes.
	This girl has very big feet.
	Even the toes on her feet are large.
	The shoes gave the man red feet.
	His feet get sore from standing all day.
	The doctor wants your feet to be clean.
Bike	His bike had big black wheels.
	The girl rode her big bike.
	Her bike could go very fast.
	The bell on the bike was really loud.

	The boy had a new red bike. Your bike always stays in the garage.

Appendix B

Results of Condition (AO vs. AV) x AOI (Eye vs. Mouth) x Block (Block 1 vs. Block 2) mixed-measures ANOVAs for Familiarisation, Target and Non-Target trials.

Predictor	**Sum of Squares**	***df***	**Mean Square**	***F***	***p***
Familiarisation					
(Intercept)	9.81	140	0.07		
Condition	0.13	1	0.13	1.92	.17
AOI	0.13	1	0.13	1.87	.17
Block	0.02	1	0.02	0.27	.60
Condition x AOI	0.01	1	0.01	0.14	.71
Condition x Block	0.02	1	0.02	0.24	.62
AOI x Block	0.03	1	0.03	0.43	.51
Condition x AOI x Block	0.02	1	0.02	0.31	.58
Target					
(Intercept)	9.31	140	0.07		
Condition	0.12	1	0.12	1.82	.18
AOI	0.04	1	0.04	0.59	.44
Block	0.0002	1	0.0002	0.004	.95
Condition x AOI	0.06	1	0.06	0.96	.33
Condition x Block	0.002	1	0.002	0.04	.85
AOI x Block	0.005	1	0.005	0.07	.80
Condition x AOI x Block	0.001	1	0.001	0.02	.90
Non-Target					
(Intercept)	8.97	140	0.06		
Condition	0.13	1	0.13	2.07	.15
AOI	0.0007	1	0.0007	0.01	.92
Block	0.06	1	0.06	0.86	.36
Condition x AOI	0.13	1	0.13	2.04	.16
Condition x Block	0.0002	1	0.0002	0.004	.95
AOI x Block	0.03	1	0.03	0.42	.52
Condition x AOI x Block	0.003	1	0.003	0.05	.83

Online Supplementary Materials

Table S1

Results of Condition (AO vs. AV) x Block (Block 1 vs. Block 2) Bayesian Mixed-Measures ANOVAs for Attention During Familiarisation

Models	P(M)	P (M|data)	BF_M	BF_{10}	Error %
Block + Condition	0.2	0.57	5.29	1.00	
Block + Condition + Block*Condition	0.2	0.33	1.95	0.58	2.72
Block	0.2	0.10	0.45	0.18	1.88
Condition	0.2	$5.21e^{-4}$	0.002	$9.14e^{-4}$	2.24
Null model (incl. subject)	0.2	$1.03e^{-4}$	$4.13e^{-4}$	$1.81e^{-4}$	1.60

Table S2

Results of Condition (AO vs. AV) x Block (Block 1 vs. Block 2) Bayesian Mixed-Measures ANOVAs for Segmentation Performance

Models	P(M)	P (M|data)	BF_M	BF_{10}	Error %
Null model (incl. subject)	0.2	0.51	4.14	1.00	
Condition	0.2	0.25	1.31	0.48	1.86
Block	0.2	0.14	0.63	0.27	1.50
Block + Condition	0.2	0.07	0.30	0.14	9.72
Block + Condition + Block*Condition	0.2	0.04	0.16	0.08	2.84

Table S3

Results of Condition (AO vs. AV) x AOI (Eye vs. Mouth) x Block (Block 1 vs. Block 2) Bayesian Mixed-Measures ANOVAs for Familiarisation Trials

Models	P(M)	P (M|data)	BF_M	BF_{10}	Error %
Null model (incl. subject)	0.05	0.39	11.33	1.00	
AOI	0.05	0.17	3.80	0.45	3.18
Condition	0.05	0.15	3.21	0.39	1.50
Block	0.05	0.08	1.50	0.20	1.13
AOI + Condition	0.05	0.07	1.31	0.18	3.10
Block + AOI	0.05	0.04	0.65	0.09	2.61
Block + Condition	0.05	0.03	0.53	0.08	1.18
AOI + Condition + AOI*Condition	0.05	0.02	0.33	0.05	5.86
Block + AOI + Condition + Block*AOI + Block*Condition	0.05	0.02	0.32	0.05	93.94
Block + AOI + Condition	0.05	0.01	0.24	0.03	2.40
Block + AOI + Block*AOI	0.05	0.01	0.18	0.03	2.44
Block + Condition + Block*Condition	0.05	0.01	0.14	0.02	2.41
Block + AOI + Condition + Block*Condition	0.05	0.004	0.08	0.01	9.92
Block + AOI + Condition + Block*AOI	0.05	0.004	0.07	0.01	2.29
Block + AOI + Condition + AOI*Condition	0.05	0.003	0.06	0.008	2.31
Block + AOI + Condition + Block*AOI + AOI*Condition	0.05	$9.56e^{-4}$	0.02	0.002	5.76
Block + AOI + Condition + Block*Condition + AOI*Condition	0.05	$8.42e^{-4}$	0.02	0.002	2.78
Block + AOI + Condition + Block*AOI + Block*Condition + AOI*Condition	0.05	$2.48e^{-4}$	0.004	$6.42e^{-4}$	4.26
Block + AOI + Condition + Block*AOI + Block*Condition + AOI*Condition + Block*AOI*Condition	0.05	$1.14e^{-4}$	0.002	$2.94e^{-4}$	16.28

Table S4

Results of Condition (AO vs. AV) x AOI (Eye vs. Mouth) x Block (Block 1 vs. Block 2) Bayesian Mixed-Measures ANOVAs for Target Trials

Models	P(M)	P (M|data)	BF_M	BF_{10}	Error %
Null model (incl. subject)	0.05	0.48	16.77	1.00	
Condition	0.05	0.19	3.87	0.37	0.87
AOI	0.05	0.11	2.28	0.23	1.82
Block	0.05	0.08	1.66	0.18	1.97
AOI + Condition	0.05	0.05	0.92	0.10	12.87
Block + Condition	0.05	0.03	0.58	0.07	1.66
Block + AOI	0.05	0.02	0.35	0.04	1.50
AOI + Condition + AOI*Condition	0.05	0.02	0.31	0.04	8.53
Block + Condition + Block*Condition	0.05	0.008	0.14	0.02	2.69
Block + AOI + Condition	0.05	0.007	0.12	0.01	14.26
Block + AOI + Block*AOI	0.05	0.005	0.09	0.01	3.07
Block + AOI + Condition + AOI*Condition	0.05	0.003	0.05	0.006	2.68
Block + AOI + Condition + Block*AOI	0.05	0.002	0.03	0.004	5.22
Block + AOI + Condition + Block*Condition	0.05	0.002	0.03	0.004	3.53
Block + AOI + Condition + Block*AOI + AOI*Condition	0.05	$6.95e^{-4}$	0.01	0.001	5.83
Block + AOI + Condition + Block*Condition + AOI*Condition	0.05	$6.32e^{-4}$	0.01	0.001	2.67
Block + AOI + Condition + Block*AOI + Block*Condition	0.05	$4.34e^{-4}$	0.008	$8.99e^{-4}$	4.10
Block + AOI + Condition + Block*AOI + Block *Condition + AOI*Condition	0.05	$1.57e^{-4}$	0.003	$3.25e^{-4}$	3.06
Block + AOI + Condition + Block*AOI + Block *Condition + AOI*Condition + Block*AOI *Condition	0.05	$5.03e^{-5}$	$9.05e^{-4}$	$1.04e^{-4}$	4.47

Table S5

Results of Condition (AO vs. AV) x AOI (Eye vs. Mouth) x Block (Block 1 vs. Block 2) Bayesian Mixed-Measures ANOVAs for Non-Target Trials

Models	P(M)	P (M|data)	BF_M	BF_{10}	Error %
Null model (incl. subject)	0.05	0.29	7.43	1.00	
Block	0.05	0.27	6.73	0.93	1.07
Condition	0.05	0.12	2.43	0.41	5.54
Block + Condition	0.05	0.11	2.11	0.36	1.71
AOI Type	0.05	0.05	0.96	0.17	0.82
Block + AOI Type	0.05	0.05	0.90	0.16	1.20
Block + Condition + Block*Condition	0.05	0.03	0.54	0.10	3.84
AOI Type + Condition	0.05	0.02	0.36	0.07	1.61
Block + AOI Type + Condition	0.05	0.02	0.36	0.07	4.65
Block + AOI Type + Block*AOI Type	0.05	0.01	0.22	0.04	2.43
Block + AOI Type + Condition + AOI Type*Condition	0.05	0.009	0.17	0.03	8.09
AOI Type + Condition + AOI Type*Condition	0.05	0.008	0.15	0.03	1.93
Block + AOI Type + Condition + Block*Condition	0.05	0.005	0.10	0.02	11.82
Block + AOI Type + Condition + Block*AOI Type	0.05	0.005	0.08	0.02	2.37
Block + AOI Type + Condition + Block*Condition + AOI Type*Condition	0.05	0.002	0.04	0.008	3.97
Block + AOI Type + Condition + Block*AOI Type + AOI Type*Condition	0.05	0.002	0.04	0.007	3.42
Block + AOI Type + Condition + Block*AOI Type + Block*Condition	0.05	0.001	0.02	0.004	2.15
Block + AOI Type + Condition + Block*AOI Type + Block*Condition + AOI Type*Condition	0.05	$5.09e^{-4}$	0.009	0.002	3.16
Block + AOI Type + Condition + Block*AOI Type + Block*Condition + AOI	0.05	$1.76e^{-4}$	0.003	$6.03e^{-4}$	6.17

Type*Condition + Block*AOI	
Type*Condition	