



Infants recognize words spoken through opaque masks but not through clear masks

Leher Singh¹ | Agnes Tan¹ | Paul C. Quinn²

¹ Department of Psychology, National University of Singapore, Singapore

² Department of Psychological and Brain Sciences, University of Delaware, Newark, Delaware, USA

Correspondence

Leher Singh, Department of Psychology, National University of Singapore, 9 Arts Link, 117570 Singapore.
Email: psyls@nus.edu.sg

Abstract

COVID-19 has modified numerous aspects of children's social environments. Many children are now spoken to through a mask. There is little empirical evidence attesting to the effects of masked language input on language processing. In addition, not much is known about the effects of clear masks (i.e., transparent face shields) versus opaque masks on language comprehension in children. In the current study, 2-year-old infants were tested on their ability to recognize familiar spoken words in three conditions: words presented with no mask, words presented through a clear mask, and words presented through an opaque mask. Infants were able to recognize familiar words presented without a mask and when hearing words through opaque masks, but not when hearing words through clear masks. Findings suggest that the ability of infants to recover spoken language input through masks varies depending on the surface properties of the mask.

KEYWORDS

auditory-visual perception, development, infant word recognition, language

1 | INTRODUCTION

On account of COVID-19, the language learning landscape of many children has changed. In particular, many children hear at least some of their language input through a mask. Given that language comprehension is an intermodal event where listeners capitalize on both auditory and visual cues as adults (Rosenblum, 2008) and as children (Lewkowicz, 2003; Lewkowicz & Flom, 2014), a degraded visual signal, arising from hearing speech through a mask, may disrupt spoken language processing. The consequences of these disruptions for young children remain unclear. The goal of the present study was to determine whether a central component of everyday communication—spoken word recognition—is influenced by different types of masks used when speaking to infants. In this study, we compared effects of clear masks and opaque masks to unmasked speech on word recognition.¹ Each type of mask provides different types of information

about the face. Clear masks allow for greater transmission of light rays through the mask. In contrast, surgical masks are a much less transmissive medium for light than clear masks. However, some lip movements may be observable on the outer surface of the mask depending on contact between the mouth of a speaker and the inner surface of the mask.

In recent months, due to COVID-19, scientists have begun to debate the impact of masks for communication with young children (see Spitzer, 2020, for a review). In particular, this discussion has invoked research findings that speak to children's reliance on facial information for verbal communication, non-verbal communication, and other forms of social communication (e.g., emotional signaling). Although in many cultural settings, individuals habitually interact with children with face coverings and there is no reason to believe this to be harmful, the impact of COVID-19 is different: many children who previously encountered speech and language without masks have started to receive some language input through masks for the first time. The extent to which children, unattuned to masked language input, adapt to these new conditions remains unclear. It also remains undetermined how different types of masks, which provide different types of access to facial cues (e.g., clear versus opaque masks), influence social and

¹ By clear masks, we refer to transparent face shields that cover the entire face. By opaque masks, we refer to surgical masks (see Figure 1a and 1b for a photograph of both types of masks).



linguistic communication. As noted in a recent news article by Yeung et al. (2021), as the impact of COVID-19 may endure for months or years to come, the use of masks may continue to be a part of children's environments over the long term, making it important to understand children's capacity to adapt to masked language input.

With respect to language processing, there are reasons to posit that both clear and opaque masks could disrupt language processing for children unattuned to face coverings. We address each type of mask in turn. In the case of opaque masks, the nose and mouth area are largely covered, obscuring a listener's view of linguistically relevant cues originating from the mouth region. The mouth region is an important area of focus for children when listening to speech. In the few months after birth, infants are sensitive to information originating in this area of the face when listening to speech (Flom & Bahrick, 2007; Kuhl & Meltzoff, 1982, 1984; Lalonde & Werner, 2019; Lewkowicz, 1996, 2010; Lewkowicz & Hansen-Tift, 2012). This sensitivity has consequences for language processing. For example, the abilities of infants to process words are improved when verbal input is synchronous with facial cues, demonstrating an early sensitivity to visual speech cues (Hollich et al., 2005). In terms of specific visual cues to which infants attend when processing speech, they demonstrate sensitivity to both temporal and articulatory cues. In terms of temporal cues, studies have reported sensitivity to temporal synchrony between the onset and offset of speech and opening and closure of the mouth in infants, pointing to the use of temporal cues as a means of integrating auditory and visual input (Lalonde & Werner, 2019; Lewkowicz, 2010). This sensitivity has been argued to be adaptive for young and inexperienced learners, helping them to process and recognize both familiar and unfamiliar linguistic information (Lewkowicz, 2010; Pons & Lewkowicz, 2014). Beyond temporal synchrony, which is posited to be a low-level and domain-general sensitivity (Lewkowicz, 2010), there is additional evidence that infants may use articulatory cues (i.e., lip movements) to access more specific phonetic information about language input (Teinonen et al., 2008) and information about words in their language (Weatherhead & White, 2017). Therefore, for a range of reasons, covering the mouth area may tax speech and language processing by providing reduced access to informative temporal and articulatory signals.

Like opaque coverings, transparent coverings also pose challenges to visual processing, which may impact linguistic processing. Viewing objects behind transparent surfaces poses unique challenges to our visual system (Anderson, 2011). When viewing objects through transparent surfaces, individuals experience information from different surfaces (or layers) within their line of sight. They experience both the transparent medium and the surface behind the medium, both of which need to be simultaneously recovered by perceptual systems. Perception through transparent surfaces can be computationally complex: although information from both sources (the medium and the background) is collapsed into one retinal image, to compensate for optical distortions and interpret the scene, individuals have to "decompose" the image, correctly assigning surface properties to the transparent medium in order to visually define objects behind the medium (Dövençioğlu et al., 2018; Singh & Anderson, 2002). In addition, with transparent surfaces, information that lies at the boundaries

RESEARCH HIGHLIGHTS

- Recently, more children have begun to receive linguistic input through masks, the consequences of which remain unknown.
- We investigated spoken word recognition through clear masks (i.e., transparent face shields), opaque masks, and with no masks in 2-year-old infants.
- Results demonstrated that infants recognized words with no masks and through opaque masks, but not through clear masks.

of transparent surfaces ("X-junctions") introduces discontinuities in perception (e.g., changes in the geometric properties of a pen sitting in a glass of water, above and below the surface of the water).

Viewing objects through transparent media differs from viewing the same object with no barrier. Without a barrier, visual perception of an object depends on its intrinsic properties and lighting conditions. For the same objects viewed through transparent surfaces, the visual percept is optically distorted due to refraction and reflection. Transparent materials, inclusive of plastic film and glass, are refractive such that light rays change direction when transmitted through these media. The consequence of refraction is a change in the direction of the transmission of light by a specific quotient (the index of refraction). The index of refraction of transparent surfaces is not uniform within regions of a transparent surface and is generally difficult for human observers to predict (Singh & Anderson, 2002; see also Fleming, 2014; Fleming et al., 2011). Furthermore, the index of refraction is more complex for a curved transparent medium, as is the case for clear masks, making prediction of the index of refraction even more challenging. The added complexity arises because the direction of curvature (convex/concave) as well as the extent of curvature lead to diffusion (concave) where light rays diverge, or focused refraction (convex) where light rays converge, whereas flat transparent surfaces typically refract light without diffusing or focusing light (Dickinson, 1895).

Second, in addition to refraction, the reflection of light differs for transparent and opaque surfaces. Transparent surfaces transmit light, but also reflect light (Metelli, 1970). Reflections from transparent objects are multifarious as light can hit the surface of a transparent object from the outside of the transparent surface (first-order reflections, which reach the observer directly) and from the inside of a transparent surface (second-order reflections). Images projected from reflecting surfaces are somewhat unstable in that all orders of reflectance change when lighting conditions change (e.g., when sunlight casts a shadow on an object) or when the object or the observer move (Murry et al., 2013). Therefore, reconstructing an underlying image from a partially reflected projection of the image requires complex perceptual inference (Fleming et al., 2004). The type of reflection that occurs with transparent surfaces (specular reflection) differs from the type of reflection occurring with opaque surfaces (diffuse reflection) in a way that influences the visual percept of a transparent surface. High



specular reflection can cause a mirror-like image on the surface of the transparent medium, which must be reconciled with the visual percept of the object behind the medium. High reflectance can also be distracting and cause a glare that obscures an observer's view of the stimulus behind the reflecting surface.

In addition to refraction and reflection, viewing objects through transparent media can have consequences for the perception of visual contrast, color, and luminance due to the attenuation of light through transparent media (Szeliski et al., 2000). When viewing a stimulus through a transparent surface, reflectance of the transparent surface reduces the transmission of light through the surface, which can reduce the visual contrast of an object behind the surface, making it appear more dull (Anderson, 1997; Kingdom, 2011; Metelli, 1970). Transparent surfaces can also reduce perceived luminance differences of objects behind the surface (Anderson, 2003). Transparent media can further alter the perception of color: Depending on the extent of transparency of a medium, the way in which a surface transmits different wavelengths of light can influence color perception of the object behind the medium. Surfaces that are not completely transparent (erring towards translucency) absorb certain wavelengths of light, which can distort the color percept. Clear luminance boundaries and color contrast are both important to auditory-visual speech perception: the availability of these cues facilitates the accurate identification of visual cues to speech (Daubias, 2005; Jordan et al., 2000; McCotter & Jordan, 2003). Overall, then, transparent surfaces introduce both geometric (contour and shape) and photometric (luminance, contrast, and color) distortions. These factors may make it challenging to perceive visual cues to language through a transparent medium.

It remains unclear whether the intrinsic properties of clear and opaque masks, discussed above, influence the perception of speech and language. A series of studies have investigated this question in adult listeners. In a study that examined speech perception in adults spoken to without a mask or through surgical masks, there was no cost to speech perception when a surgical mask was used (Mendel et al., 2008). Under noisy conditions, however, there was a marginal cost to speech perception, which applied equally to unmasked and masked conditions. In a similar study, Atcherson et al. (2017) compared opaque masks with clear masks, both of which covered the mouth region only, on speech perception under noisy conditions. For adult listeners, there was no significant decrement in performance for either type of mask relative to no mask. In another study, Cohn et al. (2021) investigated effects of speech style on speech intelligibility through masks. In normal speech, there was no difference in intelligibility of speech produced with cloth masks or without a mask. In emotional speech, there was a significant cost associated with cloth masks. Additionally, when speakers were explicitly asked to produce speech clearly, there was a mask advantage, suggesting that when asked to enunciate clearly, speakers produce more clarity adjustments with a mask versus without a mask. The preceding studies measured intelligibility (whether speech can be accurately repeated), leaving the question open as to whether comprehension (whether speech is also understood) is similarly resilient to the effects of masks. In a study comparing effects of three types of opaque mouth coverings (N95 masks, surgical masks, and cloth cover-

ings) on speech perception and word comprehension in adults, Magee et al. (2020) reported that speech perception as measured by intelligibility was not adversely affected by any of the mask types, but comprehension was equally negatively affected by all three types of masks. This outcome suggests that linguistic meaning may be particularly challenging to extract through masked language input.

For the most part, studies on language processing through masks have focused on adults. There is currently little indication as to how young children negotiate speech through masks. Perceptual recovery through masked language input may be very different for adults, who are equipped with much larger vocabularies and heightened top-down knowledge of speech and language. In contrast, for infants, who are in the process of building up a native vocabulary, it is important to know whether everyday language processing is affected by masked input. In the present study, we tested 2-year-olds on their abilities to recognize spoken words when words were presented with no mask, through a clear mask, or through an opaque mask. We employed a standard preferential looking paradigm, which has consistently demonstrated that by 2 years of age, when presented with two objects on-screen (a target and a distractor), there is preferential fixation of the target over the distractor upon hearing the target labeled (e.g., Ballem & Plunkett, 2005; Mani & Plunkett, 2007; Mani et al., 2008; Singh et al., 2015; Swingley & Aslin, 2000, 2002; Wewalaarachchi et al., 2017; White & Morgan, 2008). In line with the studies cited above, we hypothesized that infants would preferentially fixate the target object when its label was presented without a mask. As both opaque and clear masks introduce different types of challenges to word recovery, we sought to investigate how these types of coverings would influence language comprehension in relation to each other and in relation to no mask.

2 | METHOD

2.1 | Participants

Twenty-four infants participated in this study (12 males and 12 females). All infants were monolingual speakers of English. The mean age was 22.6 months (range = 22 months, 1 day – 23 months, 27 days). Four additional infants were tested. One was excluded from the study due to technical error, one was excluded for having insufficient data for inclusion due to maintaining an exclusive focus on the target or distractor, and two were excluded due to being statistical outliers (fixation times exceeded 2 SDs of the group mean). Effect sizes are typically large in this type of task (see von Holzen & Bergmann, 2021, for a meta-analysis of studies using this paradigm). The effect size used for this computation was derived from a meta-analysis on the current paradigm (correct pronunciation trials) by von Holzen and Bergmann (2021), which yielded an effect size of 1.04 (Hedges' *g*) at 22 to 24 months of age. These estimates were established prior to testing and used as a guide to sample size. Assuming this effect size applies to 24-month-old participants (von Holzen & Bergmann, 2021) using a power criterion of .8, a minimum of 10 participants would be required to detect recognition of correctly pronounced words presented without

TABLE 1 Acoustic analyses of target words

	No Mask		Opaque Mask		Clear Mask	
	M	SD	M	SD	M	SD
Duration (s)	2.16	0.074	2.12	0.053	2.13	0.074
Average Pitch (Hz)	226.29	6.92	223.78	7.10	226.06	10.68
Average Pitch Range	173.31	20.73	161.73	23.38	163.89	18.52
Loudness (dB)	73.28	1.50	74.14	1.42	74.29	1.70

a mask (measured by a significant increase in fixation to a visual target after hearing it labeled relative to before hearing it labeled).

2.2 | Stimuli

Eighteen monosyllabic and imageable test words served as targets (*bear, bird, boat, book, cake, car, chair, cheese, door, fork, keys, milk, shoe, sock, soup, spoon, star, train*). All target words were early-acquired words in English monolingual infants (Fenson et al., 2007). Labels for target objects were recorded within the carrier phrase "Can you see the ___?". All stimuli were recorded by a female speaker originating from the same city as the participants. Distractor stimuli consisted of 18 images of common objects (e.g., a ball, pen, hand, tree) and were not labeled.

Stimuli were audio- and video-recorded under natural lighting conditions by a female speaker. The female speaker wore the same clothes and accessories, and maintained a consistent facial expression across the three conditions. The speaker posed a neutral facial expression throughout the videos. While we experimented with greater facial expressiveness during piloting, a smiling expression was far more salient in the clear mask and no mask conditions than in the opaque mask condition, where the actor's smile was largely obscured. A neutral expression was chosen in order to minimize the affective contrast between conditions and to avoid the clear mask and no mask conditions being more appealing.

The actor produced five tokens of each test word within the carrier phrase for each of the three conditions. From these, a selection of one token per carrier phrase was chosen based on clarity and quality of the recording. Each carrier phrase and word was recorded with a clear mask, without a mask, and with an opaque mask. We measured target word duration, mean pitch, pitch range, and loudness of words across the three presentation conditions (no mask, clear mask, opaque mask). Across the three conditions, stimuli were matched on target word duration, ($F[2, 51] = 1.53, p = .23$), mean pitch ($F[2, 51] = 0.49, p = .62$), and pitch range ($F[2, 51] = 1.55, p = .22$). Mean loudness of target words during presentation did not vary across the three conditions, ($F[2, 51] = 2.22, p = .12$). Acoustic analyses are presented in Table 1.

2.3 | Procedure

The study took place in a child-friendly room. Infants were seated next to their parents, who wore masking music headphones throughout the

task. Participants were seated 60 cm away from the screen, in alignment with the center of a computer monitor. Auditory stimuli were presented via speakers at a conversation level commensurate with infant-directed speech (65 to 70 dB). A video camera recorded the eye movements of the participants throughout each trial. Video records were coded frame-by-frame offline at a frame rate of 30 frames per second (33 ms/frame) using the ELAN coding system (Lausberg & Sloetjes, 2009). The coder was blind to the objectives and goals of the experiment and had no access to the speech stimuli, nor to the condition of the study. The coder was only told that each trial had a central object and a left and right object. All coding was done with a silent video track, so it was not possible to know what stimuli were being played.

The experiment began with two practice trials where participants viewed two common objects on the left and right side of the screen, a target and distractor. The target was labeled in the carrier phrase "Can you see the ___?". There was no face presented on screen in the practice trials as the objective of these trials was to familiarize infants with the presence of objects on the left and right of the screen. After the practice trials, 18 test trials were presented, each consisting of a target and distractor appearing on the left and right sides of the screen. The visual angle subtended by the left and right stimuli was 14.5 degrees. During each test trial, a video of a woman's face appeared in the center of the screen. The woman labeled one of the objects using the carrier phrase "Can you see the ___?" while fixating the center of her field of view. Both objects and the woman's face appeared on screen for the entire trial. In six trials, the woman had no mask; in six trials, she wore an opaque mask; and in six trials, she wore a clear mask (see Figures 1a to 1c). The opaque mask was a normal surgical mask. The clear mask used was a Starise re-usable transparent face shield. Trial order was randomized within and across participants. Left-right positioning of the target was counterbalanced across participants on the first trial, and also counterbalanced within participants across trials. Although all infants received the same targets and distractors, and targets were rotated between conditions (no mask, opaque mask, clear mask) across infants. Target-distractor pairings remained constant across infants as target and distractor images were roughly matched in visual salience on each trial.

As in past studies using preferential looking to measure infant word recognition, trials were divided into pre-naming (0-2500 msec from trial onset) and post-naming (2501-5000 msec from trial onset) phases (Ballem & Plunkett, 2005; Mani & Plunkett, 2007; Singh et al., 2015; Zangl et al., 2005). On each trial, the target word appeared at the mid-way mark (2500 msec). Target fixation during the pre-naming phase provides a measure of baseline attention to the target object. If participants associate verbal labels with the target object, they typically demonstrate an increase in fixation to the target during the post-naming phase. For the post-naming window, PTL was calculated from 367 msec after the onset of the target word based on prior evidence that eye movements prior to this point are unlikely to be responses to the auditory label (Canfield et al., 1997). In addition to the word recognition task, parents of all participants completed the MacArthur-Bates Communicative Development Inventories (Words and Sentences) (Fenson et al., 2007) to derive an estimate of vocabulary size.

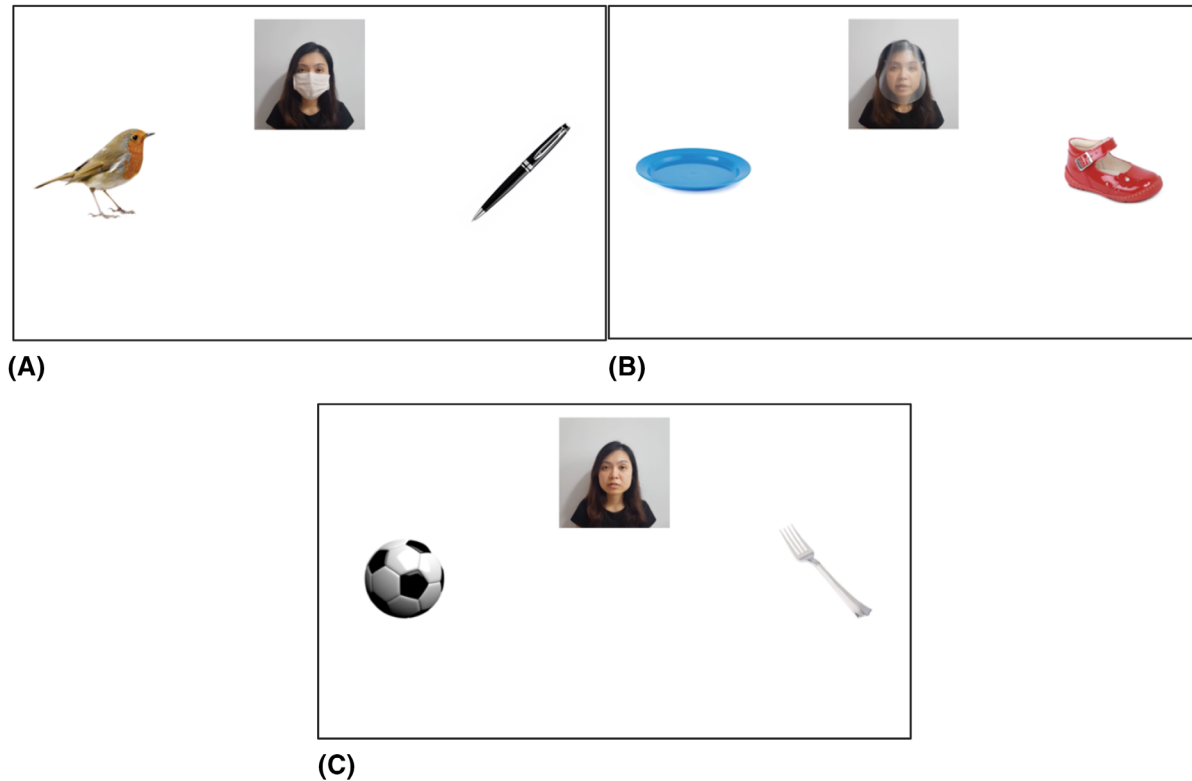


FIGURE 1 a) An example of an opaque mask trial; b) An example of a clear mask trial; c) An example of a no mask trial

TABLE 2 Descriptive statistics for preferential looking paradigm

	No Mask		Opaque Mask		Clear Mask	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	(95% CI)		(95% CI)		(95% CI)	
Pre-naming	.47	.20	.49	.18	.54	.21
	(.33 - .55)		(.41 - .57)		(.45 - .63)	
Post-naming	.64	.16	.65	.18	.59	.19
	(.57-.71)		(.57-.73)		(.51 - .67)	

3 | RESULTS

The dependent measure consisted of proportion of fixation to labeled targets during pre- and post-naming phases. Descriptive statistics for proportion of fixation to labeled targets in each phase are reported in Table 2. Data analyses were conducted using SPSS (version 27) (IBM Corporation). As in past research using this paradigm, a significant increase in fixation to the target during the post-naming phase relative to the pre-naming phase suggests that participants have associated the verbal label with the image of the target. As a first step, we analyzed attention to videos of the speaker for the no mask, opaque mask, and clear mask conditions to ensure that participants engaged equally with faces in all conditions. A repeated-measures ANOVA revealed no difference in fixation to videos of the speaker without a mask, with an opaque mask, or with a clear mask, $F(2, 46) = 1.21, p = .31$.

We then sought to investigate effects of two relevant background variables on word recognition: prior experience with masks and vocabulary size. In the first analysis, within the participant sample, we asked parents about their children's prior exposure to clear masks and opaque masks. Twelve participants were fully cared for at home and received no significant language input through a mask, 8 participants attended daycare and their primary caregiver at daycare wore an opaque mask at all times, and finally, 4 participants attended daycare and their primary caregivers wore an opaque mask at all times, but switched to a clear mask (i.e., transparent face shield) when engaging in language-related activities (e.g., vocabulary instruction, singing). Via a one-way ANOVA, we examined whether the extent of increase in proportional fixation to the target between pre- and post-naming phases (i.e., naming effects) differed based on prior mask experience (opaque mask, clear mask, no mask). There was no effect of type of mask experience on naming effects, $F(2, 23) = .37, p = .69$. In terms of vocabulary factors, parents were asked whether their children understood each of the 18 words in the experiment. The mean number of words reported to be understood was 17.23 (range: 12 to 18). Eighteen infants understood all of the words. Of the remaining words, three infants did not know the word 'key', three infants did not know the word "fork," two infants did not know the word "soup," two infants did not know the word "cheese," two infants did not know the word "boat," one infant did not know the word "cake," one infant did not know the word "door," one infant did not know the word "star," one infant did not know the word "sock," and one infant did not know the word "train." There were no words reportedly unknown across the sample.

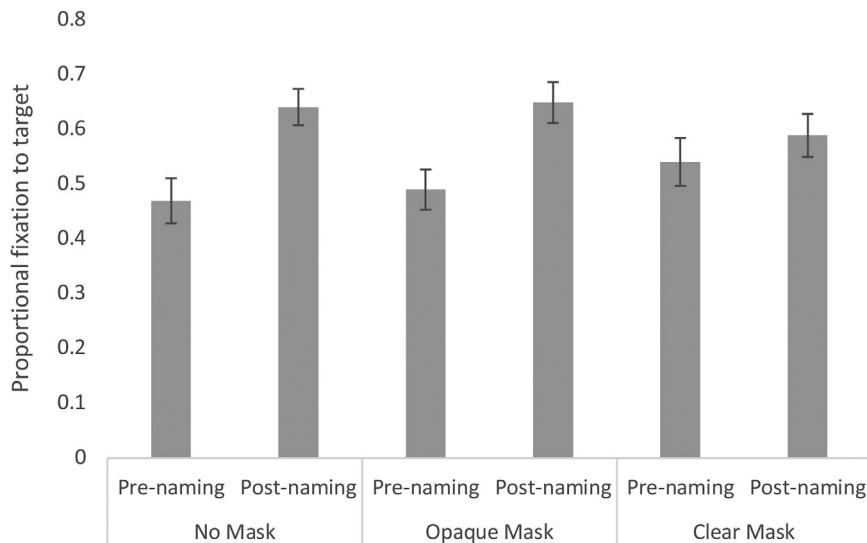


FIGURE 2 Proportion of target fixation by phase and condition

Note: Error bars indicate SEM

Vocabulary size estimates were collected on all infants in light of prior evidence that vocabulary size has been associated with the capacities of infants to restore a degraded signal. In particular, it has been suggested that infants with larger vocabularies may have stronger lexical representations that allow them to recover target words under sub-optimal listening conditions (Newman, 2004). Using a similar paradigm where infants view paired images displayed side-by-side accompanied by familiar labels, Zangl et al. (2005) demonstrated that toddlers with larger overall vocabulary size estimates were better able to recover the underlying target word and preferentially fixate the labeled object when the auditory signal was degraded or incomplete. In our study, we correlated naming effects (post-naming versus pre-naming fixation times) with vocabulary size as measured by the MCDI. Vocabulary size referred to words that were understood and said by the infants. Mean vocabulary size across participants was 158 words (range: 0 to 509 words). Vocabulary size was positively correlated with mean naming effects, $r(24) = .43, p = .03$. On account of this association and prior evidence that high vocabulary size may protect infants against the disruptive effects of degraded input (Zangl et al., 2005), we included vocabulary size as a covariate in our analyses.

Our primary analysis sought to determine whether spoken word recognition varied on account of whether participants heard words without masks, through clear masks, or through opaque masks. Participants had to have at least 1 valid trial per condition to be included. Trials were excluded if participants did not look at both the target or distractor. As a result of exclusion based on not looking at both objects, for the no mask condition, there were a total of 102 trials included in the analysis out of a total of 144 trials. For the opaque mask condition, a total of 120 trials were included. For the clear mask condition, there were a total of 108 trials included. To determine whether baseline interest in the objects varied by condition, we conducted a repeated-measures ANOVA with proportion of fixation to objects before they were labeled as the dependent variable and condition (no mask, clear mask, opaque mask) as the independent variable. Pre-naming target fixation did not differ by condition ($F[2,$

48] = .66, $p = .52$, $BF_{10} = .22$). Vocabulary size was not entered as a covariate in this analysis, because vocabulary knowledge is not believed to influence basic visual attention to objects, but rather, vocabulary size factors into word recognition after objects are labeled.

We then conducted a 2×3 (phase: pre-naming/post-naming \times condition: no mask, clear mask, opaque mask) repeated-measures ANOVA with the proportion of target looking (PTL) as the dependent variable and vocabulary size as a covariate. Bayes Factors were computed via JASP (2020) using default priors for scale and location. Using a significance criterion of .05, there was no main effect of phase, $F(1, 46) = 3.12, p = .10$ ($BF_{10} = 276.04$) and no main effect of condition, $F(2, 46) = .17, p = .85$ ($BF_{10} = .07$). However, there was a significant interaction of phase and condition, $F(2, 46) = 3.68, p = .02$ partial $\eta^2 = .16$ ($BF_{10} = 19.85$).

To investigate the interaction further, pre- and post-naming PTL were compared for each condition (see Figure 2). There was a significant increase in fixation between pre- and post-naming phases for words presented with no mask, $t(23) = 3.01, p = .006$, Cohen's $d = .93$ ($BF_{10} = 7.29$), and for words presented through an opaque mask $t(23) = 3.51, p = .002$, Cohen's $d = .86$ ($BF_{10} = 20.01$). However, there was no significant difference in fixation to target between pre- and post-naming phases for words presented through a clear mask, $t(23) = .71, p = .49$, ($BF_{10} = .27$), providing strong support for an effect in the opaque mask condition and in the no-mask condition, and moderate support for a null effect in the clear mask condition (Wagenmakers et al., 2018). Pairwise comparisons remained significant following Bonferroni correction for multiple comparisons.

We repeated the analyses above with all trials removed that contained words that infants reportedly did not understand. This procedure led to the exclusion of 15 trials. Across the sample, a total of 17 words were reported to not be understood by the infants, but 2 of these trials had already been excluded because participants did not fixate both the target and distractor. The pattern of results

with unknown words excluded was highly similar with a significant interaction of phase and trial type, $F(2, 44) = 5.59, p = .007$, partial $\eta^2 = .20$ ($BF_{10} = 5.36$). As before, a significant increase in fixation between pre- and post-naming phases was evident in trials with no mask, $t(23) = 3.27, p = .003$, Cohen's $d = .97$ ($BF_{10} = 12.09$) and trials with an opaque mask, $t(23) = 3.48, p = .002$, Cohen's $d = .89$ ($BF_{10} = 18.82$), but not in trials with a clear mask, $t(23) = .14, p = .89$, ($BF_{10} = .22$).

Finally, we compared the extent of increase in PTL (i.e., naming effects) across conditions via a repeated-measures ANOVA, again with vocabulary size as a covariate. The difference between pre- and post-naming fixation times served as the dependent variable. Trials where participants did not fixate target and distractor objects were excluded, as before. There was a main effect of condition on the extent of increase in PTL, $F(2, 44) = 3.62, p = .03$, partial $\eta^2 = .14$, $BF_{10} = .37$. Within-subject contrasts comparing each type of face covering (clear masks, opaque masks) to no-mask trials revealed no difference in naming effects for opaque masks versus no mask trials $F(1, 22) = .57, p = .46, BF_{10} = .22$, but a significant increase in naming effects for no masks versus clear masks, $F(1, 22) = 7.63, p = .01$, partial $\eta^2 = .26$, $BF_{10} = .45$.

4 | DISCUSSION

The purpose of the current study was to investigate the abilities of infants to recognize spoken words through different types of masks. In particular, we examined the abilities of infants to identify visual targets corresponding to words produced with no masks, opaque masks, and clear masks (i.e., transparent face shields). Results demonstrated preferential fixation of visual targets upon hearing them labeled with no mask and through an opaque mask, but not through a clear mask. Our findings suggest that infants are able to recover linguistic information through opaque masks. In contrast, clear masks appeared to be more challenging. The difficulties in extracting linguistic information through a clear mask were evidenced by the lack of target preference only in the clear mask condition. Against a substantial backdrop of evidence suggesting that infants reliably fixate labeled targets by 2 years of age under clear listening conditions (e.g., Ballew & Plunkett, 2005; Mani & Plunkett, 2007; Mani et al., 2008; Singh et al., 2015; Swingley & Aslin, 2000, 2002; Swingley et al., 1999; Wewalaarachchi et al., 2017; White & Morgan, 2008) as well as with degraded auditory input (Zangl et al., 2005), our findings suggest that these abilities are preserved with opaque masks and degraded with clear masks. To our knowledge, this study provides the first published data that compares language processing in children with different types of masks, contributing much-needed evidence on how linguistic communication can be optimized in the current environment.

It may be natural to ascribe linguistic advantages to clear masks over opaque masks as they provide a more extensive view of the face. This reasoning was reflected in our study where a subset of our participants attended daycares where caregivers intentionally switched from opaque masks to clear masks only when engaged in language-related

activities. It is also reflected in guidance provided by the Centers for Disease Control and Prevention (US, 2021), which noted that clear masks provide an alternative for young children learning to read and for those learning a new language. However, as suggested in the Introduction, clear masks may result in a less accurate visual percept of the face based on structural properties of transparent media, specifically, reflectant and refractive properties. In addition, transparent surfaces interact with environmental conditions in more complex ways which are relevant to conversational interactions. For example, movement on the part of the person in front of the mask or behind the mask, as well as changing lighting conditions, can alter visual information transmitted through a clear mask, creating moment-to-moment disruptions in the visual signal. In natural interactions, perceptual restoration of a rapidly changing visual signal in response to spontaneous object movement or lighting conditions or both may be challenging to accomplish in real time. Opaque masks are more resistant to these disruptions and although they occlude significant parts of the face, in some ways, they may provide a more stable visual signal. It is possible that some cues may remain perceptible (e.g., lip movements). These cues, if perceptible, may be less warped by both the structural properties of the mask as well as less susceptible to changing environmental conditions. The extent to which articulatory cues are available to listeners through opaque masks likely depends on the properties and fit of the mask.

Our study investigated one aspect of language processing, spoken word recognition. In typical instantiations of preferential looking paradigms used to measure spoken word recognition, it is possible to arrive at the visual target using auditory cues alone and the task does not necessitate accessing any facial information. However, in natural interactions, children encounter a range of additional cues to word meanings. For example, adults often visually fixate an intended referent while naming it, providing gaze cues to word meaning (Brooks & Meltzoff, 2005, 2008). Having an unobscured view of the eye region—as is the case with opaque masks—may therefore facilitate referential communication in natural interactions. In preferential looking experiments, leading gaze cues are typically absent, as they were in this study. However, when gaze cues are available, infants utilize these cues along with auditory information to guide word recognition (e.g., Graham et al., 2010; Paulus & Fikkert, 2014). Future studies could explore whether social cues to reference (e.g., eye gaze) are less accessible with clear masks, which often cover the eye region and may therefore distort visual perception of eye gaze, versus opaque masks, which leave the eye region uncovered.

In addition to understanding spoken language, other important aspects of social communication may be influenced by masks, such as the perception of emotion. In natural interactions, children use a range of cues to identify facial expressions of emotion (Gross & Ballif, 1991; Nelson & Russell, 2011). Both children and adults make use of the eye region and the mouth region, although in differing ways, to identify emotions in the face (Leitzke & Pollak, 2016). In a recent study, Ruba and Pollak (2020) contrasted effects of lower-face opaque coverings (opaque masks), upper-face opaque coverings (sunglasses), and no face coverings on emotional identification in the face. School-aged children ranging from 7 to 13 years of age were more accurate in



identifying emotions in unmasked faces relative to those covered by sunglasses or by opaque masks. However, children were still above chance when identifying emotions obscured by sunglasses and opaque masks. Performance did not differ across the two masked conditions (sunglasses versus face masks). The authors concluded that children readily adapt to the varying ways in which emotion is conveyed in natural discourse and when specific cues are inaccessible, children harness other available cues to identify emotions. This conclusion is consistent with broader evidence that linguistic cues are not necessarily localized to one area of the face. For example, whole-head movements provide predictive cues to vocal pitch and amplitude, which originate from a talker's mouth (Munhall et al., 2004) and convey vocal emotion. Similarly, movement of the articulators—that specify phonetic information—can be predicted by more global facial movement (Yehia et al., 1998). In this sense, it is possible that listeners recover linguistic information associated with the mouth by accessing cues elsewhere in the face when the mouth is obscured. Further empirical work could investigate whether facial cues that predict mouth movements and voice quality (e.g., vocal pitch) are more easily accessed with opaque masks or clear masks.

While the present study suggests that infants are able to recognize words presented through opaque masks, future research could compare caregiver communication through clear masks versus opaque masks. For example, it is possible that caregivers provide compensatory information, such as increased gaze cues or greater vocal effort or both, when communicating through opaque masks, given that the mouth area is occluded. Although this possibility has not been studied in adult-child interactions, amongst adults, speakers report committing greater vocal effort when speaking with an opaque mask (Ribeiro et al. 2020). It could be that speaking through an opaque mask leads speakers to make articulatory adjustments to compensate for the medium, as demonstrated by Cohn et al. (2021). Whether these adjustments are comparable when speaking through clear masks remains unknown. Furthermore, understanding the extent to which adults compensate for either type of mask when speaking with infants would better inform our understanding on the co-regulation of communication between child and caregiver when interacting with masks.

Currently, little is known about how rapidly or effectively language learners adapt to masked visual information as they gain more experience with face coverings, clear or opaque. Many children around the world receive language input from caregivers who wear face coverings. We do not suggest that this is in any way negative for language development. Instead, our study investigates perceptual adaptation to masks, that is, the ease with which infants attune to a change in the medium through which speech is produced. Perceptual adaptation to novel listening conditions has been widely studied in adults (see Kleinschmidt & Jaeger, 2015). In contrast, it is less clear how effectively young children adapt to novel listening conditions. Future research could chart within-participant development in the abilities of infants to negotiate masked language input over time to determine whether perceptual restoration of masked language input improves with increased exposure to masks and whether any observed improvement would differ for opaque versus clear masks. Along similar lines, investigating effects of individ-

ual differences in the age of infants, their vocabulary size, and their working memory capacity (Nagaraj & Magimairaj, 2020), could provide insight into the conditions under which infants can best recover accurate linguistic information from masked input.

Our findings have implications for learning language through clear and opaque face masks, but they may also be relevant to language and communication through other transparent media. For example, many schools use Plexiglass barriers between students (Hyde, 2020). Similar to clear masks, speech perception through Plexiglass barriers results in optical distortion of the visual signal which significantly degrades speech perception. In an empirical study on perceiving speech through a Plexiglass barrier, effects of the barrier were perceptually similar to viewing a speaker through significantly blurred vision, reducing the accuracy of auditory perception of speech to almost half the level with no barrier (Erber, 1979). Distortion is particularly high when speakers and listeners are situated at relatively large distances from each other (> 60 cm), as is the case in a socially distanced classroom. Overall, the consequences of transparent barriers for linguistic communication remain largely unknown and merit further testing.

Our study provides a first step towards understanding the impact of different types of masks on language comprehension. However, there were limitations to our study. First, it was conducted in a laboratory setting, which is important for obtaining speech-responsive eye movements without background noise. In addition, stimuli were recorded and presented under clear conditions. In natural interactions, however, background noise is ubiquitous, including in educational settings. Past studies investigating language input through clear and opaque masks in adults suggest that there is little to no decrement in speech intelligibility when listening to speech through both types of masks in noisy environments, with signal to noise ratios of +5 or +10 (Atcherson et al., 2017; Mendel et al., 2008). However, these studies may underestimate the perceptual challenges introduced by background noise. As is typical in laboratory studies investigating effects of background noise, the methods included continuous (and therefore, predictable) background streams of conversational babble (Atcherson et al., 2017) or the noise of a specific dental procedure (Mendel et al., 2008) overlaid on target words and sentences. In natural environments, background noise can be continuous or intermittent and more varied in spectral quality than the sources of noise used in prior studies. It is not clear how infants would fare with masked input in the context of natural sources of background noise. Future research could examine how infants contend with masked input under more typical listening conditions (e.g., in a classroom setting or on a playground). It is possible that any sort of masked input may impact word recognition under these conditions. In addition, our task presented infants with dissimilar objects with distinct labels. Tasks that require infants to use more fine-grained phonological knowledge may yield different findings. Future studies could compare fixation to target and distractors consisting of minimal pairs (e.g., “cat” and “bat”) to determine whether masked input compromises performance on tasks that require more granular phonological sensitivities.

The goal of the present study was to investigate the effects of masked speech on language comprehension in infants. Our findings

suggest that early learners can recover linguistic input from opaque masks, but that clear masks are more challenging for infants, even when recognizing familiar words. While both opaque and clear masks degrade access to visual information, optical distortions from transparent media may limit the transmission of visual information through clear masks. The present findings are relevant to the current climate where little is known about how best to optimize language input to children while prioritizing health and safety in children's environments.

ACKNOWLEDGMENTS

This research was supported by an ODPRT research excellence grant to Leher Singh. We are grateful to Annabel Tan, Alexandra Paquette, Glinys Lee, and Stella Png for assistance with study preparation, participant testing, and data coding.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, 26(4), 419–453. <https://doi.org/10.1068/p260419>
- Anderson, B. L. (2003). The role of occlusion in the perception of depth, lightness, and opacity. *Psychological Review*, 110(4), 785–801. <https://doi.org/10.1037/0033-295X.110.4.785>
- Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology*, 21(24), R978–R983. <https://doi.org/10.1016/j.cub.2011.11.022>
- Atcherson, S. R., Mendel, L. L., Baltimore, W. J., Patro, C., Lee, S., Pousson, M., & Spann, M. J. (2017). The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss. *Journal of the American Academy of Audiology*, 28(1), 58–67. <https://doi.org/10.3766/jaaa.15151>
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32(1), 159–173. <https://doi.org/10.1017/S0305000904006567>
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543. <https://doi.org/10.1111/j.1467-7687.2005.00445.x>
- Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, 35(1), 207–220. <https://doi.org/10.1017/S030500090700829x>
- Canfield, R. L., Smith, E. G., Breznsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development*, 62, 1–160. <https://doi.org/10.2307/1166196>
- Centers for Disease Control and Prevention (April 6). (2021) Guidance for wearing masks: Help to stop the spread of COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html>
- Cohn, M., Pycha, A., & Zellou, G. (2021). Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. *Cognition*, 210, 104570. <https://doi.org/10.1016/j.cognition.2020.104570>
- Daubias, P. (2005). Is color information really useful for lip-reading? (Or what is lost when color is not used). In Proceedings of the Annual Conference of the International Speech Communication Association (INTER-SPEECH), 1193–1196.
- Dickinson, F. (1895). Errors of refraction. *JAMA: The Journal of the American Medical Association*, XXIV(18), 665. <https://doi.org/10.1001/jama.1895.02430180011002d>
- Dövényöglu, D. N., van Doorn, A., Koenderink, J., & Doerschner, K. (2018). Seeing through transparent layers. *Journal of Vision*, 18(9):25, 1–19. <https://doi.org/10.1167/18.9.25>
- Erber, N. P. (1979). Auditory-visual perception of speech with reduced optical clarity. *Journal of Speech and Hearing Research*, 22(2), 212–223. <https://doi.org/10.1044/jshr.2202.212>
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd edn.). Brookes.
- Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, 94, 62–75. <https://doi.org/10.1016/j.visres.2013.11.004>
- Fleming, R. W., Jäkel, F., & Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological Science*, 22(6), 812–820. <https://doi.org/10.1177/0956797611408734>
- Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, 4(9), 798–820. <https://doi.org/10.1167/4.9.10>
- Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, 43(1), 238–252. <https://doi.org/10.1037/0012-1649.43.1.238>
- Graham, S. A., Nilsen, E. S., Collins, S., & Olineck, K. (2010). The role of gaze direction and mutual exclusivity in guiding 24-month-olds' word mappings. *The British Journal of Developmental Psychology*, 28(2), 449–465. <https://doi.org/10.1348/026151009X424565>
- Gross, A. L., & Ballif, B. (1991). Children's understanding of emotion from facial expressions and situations: A review. *Developmental Review*, 11(4), 368–398. [https://doi.org/10.1016/0273-2297\(91\)90019-K](https://doi.org/10.1016/0273-2297(91)90019-K)
- 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. <https://doi.org/10.1111/j.1467-8624.2005.00866.x>
- Hyde, Z. (2020). COVID-19, children and schools: Overlooked and at risk. *The Medical Journal of Australia*, 213(10), 444–446. <https://doi.org/10.5694/mja2.50823>
- IBM Corporation (2020). IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY. [Computer software].
- JASP Team (2020). JASP (Version 0.14.1)[Computer software].
- Jordan, T. R., McCotter, M. V., & Thomas, S. M. (2000). Visual and audiovisual speech perception with color and grayscale facial images. *Perception & Psychophysics*, 62(7), 1394–1404. <https://doi.org/10.3758/BF03212141>
- Kingdom, F. A. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, 51(7), 652–673. <https://doi.org/10.1016/j.visres.2010.09.012>
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141. <https://doi.org/10.1126/science.7146899>
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, 7(3), 361–381. [https://doi.org/10.1016/S0163-6383\(84\)80050-8](https://doi.org/10.1016/S0163-6383(84)80050-8)
- Lalonde, K., & Werner, L. A. (2019). Infants and adults use visual cues to improve detection and discrimination of speech in noise. *Journal of Speech, Language, and Hearing Research*, 62(10), 3860–3875. https://doi.org/10.1044/2019_JSLHR-H-19-0106
- Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41(3), 841–849. <https://doi.org/10.3758/BRM.41.3.841>

- Leitzke, B. T., & Pollak, S. D. (2016). Developmental changes in the primacy of facial cues for emotion recognition. *Developmental Psychology*, 52(4), 572–581. <https://doi.org/10.1037/a0040067>
- Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1094–1106. <https://doi.org/10.1037/0096-1523.22.5.1094>
- Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, 39(5), 795–804. <https://doi.org/10.1037/0012-1649.39.5.795>
- Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, 46(1), 66–77. <https://doi.org/10.1037/a0015579>
- Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, 109(5), 1431–1436. <https://doi.org/10.1073/pnas.1114783109>
- Lewkowicz, D. J., & Flom, R. (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development*, 85(2), 685–694. <https://doi.org/10.1111/cdev.12142>
- Magee, M., Lewis, C., Noffs, G., Reece, H., Chan, J. C. S., Zaga, C. J., Paynter, C., Birchall, O., Azocar, S. R., Ediriweera, A., Caverlé, M. W., Schultz, B. G., & Vogel, A. P. (2020). Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. *Journal of Acoustical Society of America*, 148(6), 3562–3568. <https://doi.org/10.1101/2020.10.06.327452>
- Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57(2), 252–272. <https://doi.org/10.1016/j.jml.2007.03.005>
- Mani, N., Coleman, J., & Plunkett, K. (2008). Phonological specificity of vowel contrasts at 18 months. *Language and Speech*, 51(1–2), 3–21. <https://doi.org/10.1177/00238309080510010201>
- Mendel, L. L., Gardino, J. A., & Atcherson, S. R. (2008). Speech understanding using surgical masks: A problem in health care? *Journal of the American Academy of Audiology*, 19(9), 686–695. <https://doi.org/10.3766/jaaa.19.9.4>
- Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, 13(1), 59–66. <https://doi.org/10.1080/00140137008931118>
- McCotter, M. V., & Jordan, T. R. (2003). The role of facial colour and luminance in visual and audiovisual speech perception. *Perception*, 32(8), 921–936. <https://doi.org/10.1068/p3316>
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, 15(2), 133–137. <https://doi.org/10.1111/j.0963-7214.2004.01502010.x>
- Murphy, A. A., Welchman, A. E., Blake, A., & Fleming, R. W. (2013). Specular reflections and the estimation of shape from binocular disparity. *Proceedings of the National Academy of Sciences*, 110(6), 2413–2418. <https://doi.org/10.1073/pnas.1212417110>
- Nagaraj, N. K., & Magimairaj, B. M. (2020). Auditory processing in children: Role of working memory and lexical ability in auditory closure. *Plos One*, 15(11), e0240534. <https://doi.org/10.1371/journal.pone.0240534>
- Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, 110(1), 52–61. <https://doi.org/10.1016/j.jecp.2011.03.014>
- Newman, R. (2004). Perceptual restoration in children versus adults. *Applied Psycholinguistics*, 25(4), 481–493. <https://doi.org/10.1017/S0142716404001237>
- Paulus, M., & Fikkert, P. (2014). Conflicting social cues: Fourteen- and 24-month-old infants' reliance on gaze and pointing cues in word learning. *Journal of Cognition and Development*, 15(1), 43–59. <https://doi.org/10.1080/15248372.2012.698435>
- Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, 149, 142–147. <https://doi.org/10.1016/j.actpsy.2013.12.013>
- Ribeiro, V. V., Dassie-Leite, A. P., Pereira, E. C., Santos, A. D. N., Martins, P., & Irineu, R. d. A. (2020). Effect of wearing a face mask on vocal self-perception during a pandemic. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2020.09.006>
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17(6), 405–409. <https://doi.org/10.1111/j.1467-8721.2008.00615.x>
- Ruba, A. L., & Pollak, S. D. (2020). Children's emotion inferences from masked faces: Implications for social interactions during COVID-19. *Plos One*, 15(12), e0243708. <https://doi.org/10.1371/journal.pone.0243708>
- Singh, L., Goh, H. H., & Wewalaarachchi, T. D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition*, 142, 1–11. <https://doi.org/10.1016/j.cognition.2015.05.010>
- Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, 109(3), 492–519. <https://doi.org/10.1037/0033-295x.109.3.492>
- Spitzer, M. (2020). Masked education? The benefits and burdens of wearing face masks in schools during the current Corona pandemic. *Trends in Neuroscience and Education*, 20, 100138. <https://doi.org/10.1016/j.tine.2020.100138>
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147–166. [https://doi.org/10.1016/s0010-0277\(00\)00081-0](https://doi.org/10.1016/s0010-0277(00)00081-0)
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484. <https://doi.org/10.1111/1467-9280.00485>
- Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71(2), 73–108. [https://doi.org/10.1016/s0010-0277\(99\)00021-9](https://doi.org/10.1016/s0010-0277(99)00021-9)
- Szeliski, R., Avidan, S., & Anandan, P. (2000). Layer extraction from multiple images containing reflections and transparency. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1, 246–253. <https://doi.org/10.1109/CVPR.2000.855826>
- 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. <https://doi.org/10.1016/j.cognition.2008.05.009>
- von Holzen, K., & Bergmann, C. (2021). The development of infants' responses to mispronunciations: A meta-analysis. *Developmental Psychology*, 57(1), 1–18. <https://doi.org/10.1037/dev0001141>
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. <https://doi.org/10.3758/s13423-017-1343-3>
- Wewalaarachchi, T. D., Wong, L. H., & Singh, L. (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English-Mandarin toddlers. *Journal of Experimental Child Psychology*, 159, 16–33. <https://doi.org/10.1016/j.jecp.2017.01.009>
- Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, 160, 103–109. <https://doi.org/10.1016/j.cognition.2017.01.002>
- White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59(1), 114–132. <https://doi.org/10.1016/j.jml.2008.03.001>
- Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26(1–2), 23–43. [https://doi.org/10.1016/S0167-6393\(98\)00048-X](https://doi.org/10.1016/S0167-6393(98)00048-X)
- Yeung, H., Curtin, S., & Werker, J. (2021). *Face-mask use and language development: Reasons to worry?* Retrieved 26 February



2021, from <https://www.theglobeandmail.com/canada/article-face-mask-use-and-language-development-reasons-to-worry/>
Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*, 6(2), 179–208. https://doi.org/10.1207/s15327647jcd0602_2

How to cite this article: Singh L, Tan A, Quinn PC. Infants recognize words spoken through opaque masks but not through clear masks. *Developmental Science*, 2021;24:e13117. <https://doi.org/10.1111/desc.13117>