Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-026-07281-1 OPEN

Language development in the shadow of COVID-19 pandemic: evidence from audiovisual speech perception

Yi Weng[1,4], Yicheng Rong[2,3,4] & Gang Peng[1]✉

During the coronavirus disease 2019 (COVID-19) pandemic, social distancing and mask-wearing were effective preventive measures implemented globally, which dramatically changed the environment where children acquire language. However, whether these long-term protective measures, designed to reduce interpersonal interaction, impact speech development among children remains unclear. Audiovisual speech perception, a crucial component of language acquisition, was investigated using the McGurk paradigm among two groups of 5–6-year-old children native to the same language (i.e., Cantonese) but from regions implementing differing public measures against the pandemic, i.e., Hong Kong (HK) vs. mainland China. Results showed that HK children experiencing stricter and prolonged mandatory measures demonstrated reduced visual reliance when processing incongruent stimuli, especially under the noisy condition, compared to their mainland China counterparts. Combining a tendency to identify congruent stimuli less accurately, slower development in audiovisual speech processing is suggested among HK children. The results raise the possibility that social changes during public health crises could influence language development, perhaps via changes in the surrounding environment. Policymakers may need to exercise heightened caution when considering the developmental needs of children.

[1] Department of Language Science and Technology, Hong Kong Polytechnic University, Hong Kong, China. [2] School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China. [3] National Research Centre for Language and Well-Being, Shanghai Jiao Tong University, Shanghai, China. [4] These authors contributed equally: Yi Weng, Yicheng Rong. ✉email: gpeng@polyu.edu.hk

Introduction

Audiovisual speech perception develops gradually across childhood, and young children rely increasingly on visual cues, particularly lip movements, to support speech understanding (Hirst et al., 2018; Tremblay et al., 2007). Because this developmental process depends considerably on consistent sensory availability, reductions in speech input, especially from the visual modality, may hinder the maturation of audiovisual integration (Chládková et al., 2021). The coronavirus disease 2019 (COVID-19) pandemic has resulted in an unprecedented public health crisis, profoundly disrupting lifestyles worldwide and exerting far-reaching effects on the environments in which children grow, learn, and develop. As a respiratory disease, COVID-19 spreads primarily through airborne droplets and aerosols, making it highly contagious, especially in crowded or poorly ventilated spaces. Thus, protective measures, such as mask-wearing and physical distancing, are strongly advocated by the World Health Organization (WHO, 2023) and governments worldwide to reduce virus transmission and protect vulnerable populations. Taking the Hong Kong Special Administrative Region (HK hereafter) as an example, which was the epicenter of the 2003 Severe Acute Respiratory Syndrome (SARS) outbreak, the implemented policies can be considered as cautious, or even strict. Many of these control measures targeted children, such as suspension of in-person classes and mandatory mask-wearing throughout all school hours on campus. The long-lasting social distancing and mask mandate have raised concerns among linguists and psychologists regarding the potential impact of global public health crises on children's social and language development (e.g., Charney et al., 2021; Wenner Moyer, 2022). In particular, under the shadow of the COVID-19 pandemic, protective measures aiming at reducing interpersonal interactions may directly impact the quality and quantity of auditory and visual speech input, potentially leading to delayed development of audiovisual speech perception mechanisms in children. Accordingly, this study seeks to explore whether children exposed to varying intensities of pandemic-related regulations exhibit different developmental progress at the same chronological age.

Social distancing and its impact on speech perception

Social distancing breaks the chain of viral transmission by limiting close contact and minimizing exposure to respiratory droplets, which is regarded as a key measure for controlling the COVID-19 pandemic. HK government had mandated regulations, including a 1.5-m separation, gathering restrictions, or even stay-at-home measures, to reduce the risk of viral transmission since the early stage of the outbreak (Zhao et al., 2020).

According to the classic theory of social distance proposed by Hall (1966), interpersonal distance can be categorized into four ranges based on proxemics: intimate distance (0–46 cm), personal distance (46–122 cm), social distance (122–210 cm), and public distance (above 210 cm). Although subsequent studies have further demonstrated that social distance is influenced by factors such as gender, age, and familiarity, it can be reasonably argued that controlling for social distancing inevitably disrupts closer interpersonal contact (Hecht et al., 2019; Little, 1965; Uzzell and Horne, 2006). Notably, existing research has indicated a tendency for compressed interpersonal space in interactions among children (Paulus, 2018; Willis et al., 1979), suggesting that children may generally prefer closer proximity compared to adults. Consequently, the imposed 1.5-m social distancing requirement is likely to have a more pronounced impact on social interactions among children than among adults. Beyond this physical dimension, pandemic-related social distancing also extended to broader restrictions on daily activities, including periods of lockdown and quarantine that further limited children's opportunities for everyday social interaction. In many regions, children experienced substantially reduced opportunities for everyday social interactions due to limitations on outdoor activities, fewer visits to parks and playgrounds, and reduced participation in group settings. Such reductions in daily social exposure restrict children's access to naturalistic conversational exchanges and diverse communicative partners. These broader social constraints may further limit the richness of auditory and visual speech input available to children, compounding the effects of physical distancing on speech perception and language development.

On the one hand, the clearness of auditory speech cues degrades as distance magnifies due to several consequences, including the attenuation of sound intensity (Talbot-Smith, 1993), reduced high-frequency components (Bass et al., 1995), and the increased interference of environmental factors such as reverberation (Kressner et al., 2018) and background noise (Meyer et al., 2013), ultimately diminishing intelligibility and comprehension. In terms of the transmission of visual speech, increased talker-to-listener distance may reduce the visibility of facial movement during speech production (Jordan and Sergeant, 2000) while increasing the effort to discern finer-grained speech cues from the perceiver's side (Zheng and Samuel, 2019), also leading to a lowered quality of speech communication. As alternatives to in-person interactions, virtual meetings became increasingly important and widely adopted during the COVID-19 pandemic, as online platforms allowed individuals and organizations to maintain communication and collaboration, despite physical restrictions. However, the quality of voice transmission and the synchronization between audio and video depend on hardware and network stability, which can compromise the clarity of speech information (Karl et al., 2022).

Overall, cautious social distancing primarily limits close interpersonal communication, diminishing the amount and fidelity of auditory and visual speech input, potentially impeding audiovisual integration and language development. Broader pandemic-related restrictions also reduced children's everyday social interactions, further limiting access to naturalistic speech input. Together, these constraints raise important questions about how children's audiovisual speech perception may have been shaped in the context of pandemic-related social distancing.

The impact of wearing masks on speech perception

Wearing masks prevents the entry of viruses into the respiratory tract by covering the nose and mouth areas, making it another effective measure for controlling the COVID-19 pandemic. However, the masked area—the face, particularly the mouth—is critical for conveying social cues and speech-related information. As early as the 2003 SARS outbreak, the impact of mask-wearing on speech communication has drawn the attention of researchers. For instance, Coniam (2005) investigated the impact of wearing masks on oral test scores during the SARS outbreak in HK. Though no differences regarding scores reached significance between mask and unmask conditions, the author found that test takers who felt disadvantaged in a mask-wearing condition had to compensate with strategies like louder speech and increased eye contact to mitigate the adverse impact of masks on exam results. From this earlier study, it can be learned that test takers needed to pay extra effort to ensure the quality of raters' perception, indicating the obstruction to speech perception caused by masks is readily and widely experienced.

Years later, the universal and prolonged enforcement of mandated mask-wearing due to the COVID-19 pandemic has drawn increasing attention to its impact on speech perception.

Empirical studies consistently highlight the detrimental effects of mask-wearing on speech perception from unimodal auditory and/or bimodal audiovisual perspectives, even though they employ diverse types of visual stimuli ranging from authentic filmed mask conditions to digitally superimposed masks (Aguillon-Hernandez et al., 2022; Bandaru et al., 2020; Bottalico et al., 2020; Brown et al., 2021; Giovanelli et al., 2021; Kleinman et al., 2022; Lalonde et al., 2022; Lipps et al., 2021; Moon et al., 2022; Rahne et al., 2021; Sönnichsen et al., 2022; Thibodeau et al., 2021; Yi et al., 2021; Zhou et al., 2022). From the unimodal auditory perspective, the impact of mask can be particularly seen in the acoustic attenuation induced by their material (tissue or plastic) on sound pressure levels, which eventually elevates the listening efforts of perceivers, especially for higher-frequency sounds (Lalonde et al., 2022) and for noisy conditions (Bandaru et al., 2020; Bottalico et al., 2020; Brown et al., 2021; Rahne et al., 2021), although the extent of such attenuation—particularly for surgical masks—still remains debated. For instance, Rahne et al. (2021) found that the speech reception threshold (SRT) of listeners with normal hearing was significantly lifted when the loudspeakers were enclosed with masks, with N95 masks posing greater attenuation relative to the surgical ones. Similar findings were also reported by Bottalico et al. (2020), who observed reduced speech intelligibility in college students with normal hearing when word recognition tasks were presented through audio masked by various types of face coverings.

In typical face-to-face communication, visual information can serve as a compensation for speech processing when unimodal auditory information is degraded. Under mask mandates, this compensatory function of visual modality becomes ineffective (Giovanelli et al., 2021), given that masks obscure the mouth region, especially the lips, which provide the most detailed visual speech cues. The impact of depriving visual speech cues can be most evidently reflected in the speech decoding of individuals with hearing impairment, who rely heavily on visual speech cues to achieve speech comprehension. Existing studies suggested that populations with hearing impairment indeed faced a particularly challenging period during the mask mandate (Lalonde et al., 2022; Lipps et al., 2021; Moon et al., 2022; Sönnichsen et al., 2022; Thibodeau et al., 2021). For instance, Sönnichsen et al. (2022) measured the SRT of both cochlear implant (CI) users and normal-hearing individuals in various conditions, finding a significant increase in SRT in both groups if the speaker's mouth was covered by a mask under the audiovisual condition, with a more significant effect in CI users relative to their normal-hearing controls, especially for those who solely relied on CI for speech processing. On the other hand, when the speaker's face was visible, either because no mask was worn or because a transparent mask was used, individuals with hearing impairment were able to make fuller use of visual speech cues. Studies have shown that transparent masks, in particular, allow hearing-impaired listeners to benefit more from visual enhancement compared with opaque masks (Moon et al., 2022; Thibodeau et al., 2021). Taken together, the experience of hearing-impaired populations, who are highly dependent on visual information to achieve language comprehension, indicates that mask-wearing leads to a substantial deprivation of visual unimodal information to support speech processing.

For individuals with normal hearing, mask-wearing also greatly impacts their audiovisual speech processing, which stems from two aspects: the absence of visual signals along with the degradation of audio signal quality, resulting in reduced efficiency of audiovisual speech perception (Lalonde et al., 2022; Lipps et al., 2021; Thibodeau et al., 2021; Yi et al., 2021; Zhou et al., 2022). For instance, Zhou et al. (2022) measured the SRT, defined as the signal-to-noise ratio (SNR) required for 50% accuracy, under auditory-only and audiovisual settings with various types of masks among participants with normal hearing. They found that SRTs were significantly higher when speakers wore surgical masks or N95 masks with a face shield compared with conditions in which visual information was available (i.e., no mask or transparent mask). Although transparent masks provide visible mouth movements and were shown by Zhou et al. (2022) to improve perceptual accuracy in low-SNR conditions relative to opaque masks, other studies indicate that they can still hinder the transmission of auditory information. For instance, Thibodeau et al. (2021) assessed sentence recognition in noise under unmask, transparent mask, and opaque mask conditions through an online study. Results showed that, albeit the transparent mask significantly improved speech recognition due to visual cues, participants performed worse with transparent masks relative to opaque masks in the auditory-only condition.

To summarize, mask-wearing reduces the transmission quality of auditory and visual speech information, and the restriction on visual cues is particularly severe. Visual cues play an important role in supporting speech perception, especially when listening conditions are difficult. Therefore, the extent to which visual information is available has a substantial influence on audiovisual speech perception during the pandemic.

The development of audiovisual speech perception in the pandemic context

Although children start to show preferences for talking faces at a very young age, it takes a long journey for them to effectively utilize these cues to assist language processing (Hirst et al., 2018; Tremblay et al., 2007; Weng et al., 2024). As a form of multisensory processing, audiovisual speech perception develops along a prolonged trajectory, likely due to its reliance on the continuous recalibration of multiple sensory systems that mature at different rates (Ernst, 2008; Ernst and Banks, 2002). Previous developmental studies have shown that children take years to establish an adult-like manner in processing audiovisual language input (Hirst et al., 2018; Tremblay et al., 2007; Weng et al., 2024). Specifically, this developmental process initiates with a natural bias for auditory information and gradually evolves into an audiovisual-integrated manner. Hence, the effective utilization of visual speech cues remains the key to this maturation process. With the protective measures implemented during the COVID-19 pandemic, the source of meaningful visual speech input was largely blocked. Unlike infants who predominantly engage in interactions within familial contexts, which are less likely influenced by protective measures, young children might face more marked challenges not only because they have been put into more complex social networks but also because they were in a critical period of experiencing the developmental shift of audiovisual speech processing (Hirst et al., 2018; Maidment et al., 2015; Tremblay et al., 2007). For example, among Mandarin-speaking children, a marked shift in sensory dominance has been observed around ages 5–6 during audiovisual speech perception (Weng et al., 2024).

Given that this developmental shift is still underway in early childhood, any reduction in access to clear speech cues may interfere with children's progression toward more integrated audiovisual processing. Existing research indicates that the protective measures implemented during the pandemic have substantially hindered adults' speech perception, affecting both auditory-only and audiovisual contexts (Aguillon-Hernandez et al., 2022; Brown et al., 2021; Rahne et al., 2021; Thibodeau et al., 2021). This raises important concerns for children, whose language skills are still developing: Are young children facing even greater challenges under such restrictions (e.g., Charney et al., 2021; Wenner Moyer, 2022)? This concern arises from two practical observations. On the one hand, measures such as social distancing and mask-wearing have significantly reduced meaningful in-person interactions among children, thereby obscuring critical social cues that support language development (Charney et al., 2021). More importantly, covering parts of the face undermines speech expression by obscuring key facial cues, such as lip movements. These visual cues play a crucial role in enhancing speech perception and comprehension. When these cues are hidden by masks, children may face difficulties in accurately interpreting spoken words, which are essential for effective communication and language development (Wenner Moyer, 2022). Considering the contradiction between the vitalness of facial information on early development and the necessity of facial masking during the pandemic, Carnevali et al. (2022) systematically reviewed papers to illustrate the impact of the visibility of full-face and various facial features on the infants' development of socio-communicative and language skills. Included studies have confirmed that infants direct their visual attention to the mouth region of speakers during early development because they are already aware of the close relationship between mouth movements and speech input. The clear access to audiovisual information serves as pillars for early speech learning, granting the intake of visual speech cues and supporting multisensory synchronization. Therefore, Carnevali et al. (2022) concluded that prolonged face masking during the pandemic would hinder spoken language processing among children, and future research is strongly advocated to delve deeper into this topic.

Recent studies have demonstrated that children are more likely to be interfered with in perceiving audiovisual speech stimuli produced under various masking conditions. Lalonde et al. (2022) examined the consonant recognition ability among normal-hearing adults, normal-hearing children, and hearing-impaired children. Results confirmed that all groups of participants were consistently influenced by face masks in noisy conditions. Furthermore, although children with normal hearing behaved comparably with their adult counterparts in the auditory-only condition, they exhibited suppressed visual benefit in the audiovisual condition, indicating greater challenges faced by children in the pandemic context likely resulted from their immature audiovisual-integrative mechanism. Furthermore, a study carried out before and after mask mandate showed that perceivers' sensory weighting would be altered by social surroundings. Chládková et al. (2021) examined audiovisual speech perception using the McGurk paradigm on both cross-sectional and longitudinal samples, finding that, with highly efficient adaptation to changes in social surroundings, younger participants showed a significant drop in visual reliance when processing the audiovisual incongruent stimuli after one-month deprivation of visual speech information due to the mask mandate. Findings indicate that the reliance on visual modality was lowered due to the sparsity of meaningful visual speech cues from the environment, and this perceptual manner would be transferred to the scenario where speakers' faces were unmasked. Combining the immaturity of using visual speech cues, the attentional shift driven by the impoverished visual cues due to the mask mandate is highly likely to place barriers to the development of audiovisual speech perception in children during the COVID-19 pandemic.

The current study

Existing evidence suggests that pandemic-related regulations may have altered children's access to audiovisual speech cues, potentially influencing the developmental course of audiovisual speech perception, yet this possibility has rarely been examined directly. To explore this issue, the current study focuses on the audiovisual speech perception in 5–6-year-old Cantonese-speaking children from HK and mainland China using the McGurk paradigm. The choice of the 5–6 age range was motivated by two reasons: First, converging evidence from previous studies indicates that ages 5–6 constitute an important developmental window for audiovisual integration, making this period well suited for examining potential impacts of pandemic-related policies (Maidment et al., 2015; Weng et al., 2024). Second, children within this age range might encounter a greater impact of preventive policies than infants since they were considerably engaged with extensive social interactions (Hay et al., 2004). The two regions, namely, HK and the mainland, implemented different levels of self-protective policies during the three years of the pandemic, which allow us to examine the potential impacts of preventive policies on their developmental progress of audiovisual speech perception. For instance, according to the Prevention and Control of Disease Regulation issued by the Chief Executive in Council (Cap.599) issued on 14 July 2020, group gathering or not wearing masks in specified public places committed offenses with a fixed penalty in HK (The Government of the Hong Kong Special Administrative Region, 2020), and the regulation on mask-wearing had not expired after 969 days on 1 March 2023 (The Government of the Hong Kong Special Administrative Region, 2023). In accordance with a Letter to School from the Education Bureau of Hong Kong on 8 October 2020, children were required to wear masks at all times (Education Bureau, 2020), whereas government officials from the mainland had earlier stated that masks were not necessary in schools located in low-risk areas (Ministry of Education & National Health Commission, 2020). These contrasting policies provide a valuable opportunity to explore how social regulations in response to public health crises may relate to differences in the developmental trajectory of audiovisual speech perception, particularly when linguistic background is held constant.

Materials and methods

Participants. Audiovisual speech perception was measured using the McGurk paradigm among forty-four Cantonese-speaking children, with 20 born and raised in HK and 24 in mainland China from March to October 2023. The 20 HK children aged from five to six years were recruited through the departmental website and the official Facebook account. The other group of Cantonese-speaking children of the same age range was recruited from the *Guangfu* Region in Guangdong Province, mainland China. All participants were recruited from mainstream educational institutions. Children in the HK group attended local Chinese-medium kindergartens, where Cantonese was the primary language of instruction, and Mandarin was also taught as part of the curriculum. Children in the Mainland group were enrolled in public kindergartens in Guangdong Province, where Mandarin served as the main instructional language, consistent with the broader linguistic environment of mainland China. To ensure comparability in spoken language experience, we confirmed with teachers and parents that all children in the mainland group were Cantonese speakers who used Cantonese at home and could communicate fluently in daily interactions. Both groups attended kindergartens that were subject to local government regulations related to COVID-19 prevention and control. Both groups experienced substantial preschool closures during the COVID-19 pandemic, with HK and mainland China implementing comparable periods of school suspension (Tan et al., 2024; Wong et al., 2023). However, the policies in place once schools reopened differed markedly. HK maintained a strict mask mandate for young children at all times on school premises and imposed tighter social-distancing regulations in public settings, whereas mainland schools in low-risk areas were permitted to operate without mandatory masking (The Government of the Hong Kong Special Administrative Region, 2020; Ministry of Education and National Health Commission, 2020). As a result, children in HK were exposed to considerably more prolonged deprivation of visual speech cues during daily communication than their mainland peers.

Table 1 summarizes the characteristics of Cantonese-speaking child participants from HK and mainland China. The sample size was determined on availability with reference to previous studies employing similar tasks. We further confirmed the required sample size using the G*power software opting for a moderate sample size ($\eta_p^2 = 0.06$), 0.80 power, an alpha of 0.05, and 0.5 as the correlation among repeated measures for pursuing the *Region* (HK vs. mainland China) × *Noise* (clean vs. noisy) interaction, which turned out to be a total sample size of 24 (i.e., 12 per group). The experiment design had been reviewed and approved by the The Hong Kong Polytechnic University Institutional Review Board (Approval Number: HSEARS20220213003; Date of approval: 21 February 2022).

According to caregiver reports, all children had no known intellectual, behavioral, auditory, or neurological difficulties. To ensure that all participants met the criteria for typical development, caregivers completed two standard screening questionnaires: the Autism Spectrum Quotient—Children's Version (traditional Chinese version; Auyeung et al., 2008) and the Vanderbilt Assessment Scale (Parent Informant; Wolraich, 2003). These questionnaires were used only to rule out neurodevelopmental or psychiatric conditions that might affect social communication or eye-movement behavior. As shown in Table 1, all child participants scored below the cut-off thresholds, specifically 76 for ASD and 6 for ADHD, indicating that no child participants were at high risk of suffering from ASD or ADHD. All participants had normal or corrected-to-normal vision. Written consent was obtained from caregivers, and monetary compensation was provided for participation upon completion.

Table 1 Information on Cantonese-speaking Child Participants from HK and mainland China.

Group		HK Child	Mainland Child
N (Female)		20 (10)	24(12)
Chronological Ages	*Mean*	6	6.01
	(Range)	(5.11-6.93)	(5.33-6.99)
	SD	0.56	0.46
Autism Spectrum Quotient—Children's Version	*Mean*	57.6	54.82
	(Range)	(37-69)	(37-72)
	SD	9.29	9.77
Inattentive Symptoms	*Mean*	1.52	1.3
	(Range)	(0-5)	(0-5)
	SD	1.77	1.55
Hyperactive Symptoms	*Mean*	1.76	1.61
	(Range)	(0-5)	(0-4)
	SD	2.22	1.3

Stimuli. A young female speaker aged 24 years native in Cantonese, from whom written consent was obtained, was invited to film the articulation process when uttering three consonant-vowel (CV) syllables: "Ba", "Da" and "Ga" (transcribed as [pa], [ta] and [ka] using International Phonetic Alphabet) with high-level tone (around 270 Hz) where the consonants were voiceless unaspirated stops in a quiet room. All videos were filmed with a high resolution of 1920 × 1080 pixels and a framerate of 30 frames per second. The speaker's face and neck were presented at the center of the frame against the background with a solid color. The duration of each audiovisually presented stimulus was scaled to 2000 milliseconds using Adobe Premiere Pro CC 2018, within which the duration of auditory utterances was normalized to 500 milliseconds using Praat (Boersma and Weenink, 2024). Each stimulus started with the speaker's still face, followed by the articulation process, and ended with the still face. In addition, an incongruent stimulus was created with the auditory component of "Ba" dubbed onto the muted video of "Ga", giving rise to the classic fused stimulus of the McGurk paradigm (i.e., "AbVg", McGurk and Macdonald, 1976). In the context of the McGurk paradigm, typical perceivers often report a fused percept, such as "da", which represents an audiovisual-integrated response that is inconsistent with both the heard "ba" and the seen "ga", thereby indicating the presence of audiovisual integration. In contrast to the clean condition, the noisy condition was generated by adding pink noise at -10dB SNR using a MATLAB script (R2018a version). With the same script, the intensity level of the auditory components of the audiovisual stimuli was scaled to 70 dB with regard to the root mean square.

Procedure. Before performing the experiment, participants were seated in front of a monitor with a resolution of 1920 × 1080 pixels at a distance of around 55 centimetres in a soundproof room. Soundtracks were presented using a pair of external speakers posited at both sides of the monitor with a medium level of volume. During the experiment, participants needed to complete all the tasks without their caregivers, but in the presence of an examiner. The experimental procedure was similar to Weng et al. (2024), which could be split into two training sessions and a formal experimental session. The first training session was set to ascertain that participants were able to differentiate the three CV syllables associated with the current design. As shown in Fig. 1a, participants would be presented with three slides, on each of which there was one rectangular pattern with a certain

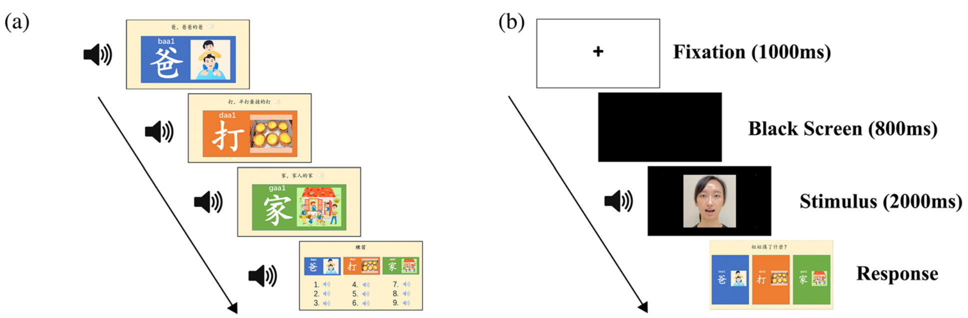

Fig. 1 Experimental procedures for the training and testing sessions. a Procedure of the first training session where participants were familiarized with the three CV syllables involved in the classic McGurk fusion. **b** Procedure shared by the second training session and formal experimental session.

background color. Each rectangular pattern contained the Chinese character, the Romanized form of the pronunciation of the character (*Jyut Ping*, which developed by Linguistics Society of Hong Kong in 1983) and one picture semantically related to the character. To strengthen the correspondence between the patterns and stimuli, the examiner would play the recording of the syllable when displaying the specific slide and instruct the participants to repeat the syllable. At the end of the first training session, there was a nine-trial tiny test where participants were expected to point to the correct rectangular pattern when hearing the syllable just played and feedback regarding the correctness of every response would be provided. During the second training session, the three congruent stimuli would be randomly presented twice following the procedure of the formal experimental session (see below for a detailed introduction). Participants who gained full accuracy were eligible to participate in the formal experiment, and all child participants successfully proceeded to the formal experiment.

The formal experiment comprised 56 audiovisually presented trials with the three congruent stimuli (i.e., "Ba", "Da" and "Ga") and the incongruent stimulus (i.e., "AbVg") randomly repeated seven times in both clean and noisy conditions, whose procedure was almost identical to that of Weng et al. (2024). Within each trial, as shown in Fig. 1b, participants would be sequentially exhibited an 800-ms fixation, a 1000-ms black screen, a 2000-ms stimulus, and an infinite response screen using E-prime. Participants were required to point to one of the three patterns according to what they had perceived. The examiner recorded the response from the participants by pressing the first, third or fifth button from the left on Chronos referring to "Ba", "Da" and "Ga", respectively.

Data analysis. Because accuracy and percentage scores are bounded and therefore may be non-normally distributed.Shapiro-Wilk tests confirmed significant deviations from normality (all $ps < 0.001$). Accordingly, non-parametric tests, namely, permutation-based repeated measures ANOVA models, were constructed using the R package "*permuco*" (Frossard and Renaud, 2021; R core team, 2023). For incongruent stimuli, the dependent variable was the percentage of trials in which each participant produced one of the three possible perceptual responses: audio-dominant ("Ba"), audiovisual-integrated ("Da"), and visual-dominant ("Ga"). These three response categories reflect the perceptual strategy adopted during multisensory conflict. The within-subject factors were *Response Type* (audio-dominant, audiovisual-integrated and visual-dominant) and *Noise* (clean vs. noisy) while the between-subject factor was *Region* (HK vs. mainland China). For the identification of congruent stimuli, likewise, a $2 \times 3 \times 2$ repeated measures permutation ANOVA was conducted to examine the impact on identification accuracy with *Stimulus Type* ("Ba", "Da" and "Ga") and *Noise* (clean vs. noisy) as within-subject factors and *Region* (HK vs. mainland China) as the between-subject factor. Wilcoxon tests with Bonferroni correction were performed to conduct post-hoc pairwise comparisons where appropriate, using the *wilcox_test* function from the "*rstatix*" package (Kassambara, 2023). Effect sizes were computed as r using the *wilcox_effsize* function from the same package, following the definition proposed by Fritz et al. (2012). Descriptive statistics for the percentage of responses to incongruent stimuli and the identification accuracy for congruent stimuli are provided in Supplementary Materials.

Results

The perception of incongruent stimuli. Figure 2 demonstrates the percentage of each type of response to the incongruent stimuli made by 5–6-year-olds residing in HK and mainland China under clean and noisy conditions, respectively. Permutation-based ANOVA revealed a significant main effect of *Response Type* ($F(2, 84) = 3.56$, permutation $p = 0.03$, $\eta_p^2 = 0.03$), a *Response Type* × *Noise* two-way interaction ($F(2, 84) = 131.27$, permutation $p < 0.001$, $\eta_p^2 = 0.51$), and a *Region* × *Response Type* × *Noise* three-way interaction ($F(2, 84) = 5.38$, permutation $p = 0.01$, $\eta_p^2 = 0.04$).

For the main effect of *Response Type*, post-hoc pairwise comparisons failed to detect any significant differences. As for the significant two-way interaction between *Response Type* and *Noise*, the percentage of audio-dominant responses made by both groups of participants sharply decreased, while the percentage of visual-dominant ones dramatically increased as auditory intelligibility declined from clean to -10dB SNR (both $ps < 0.01$). The three-way interaction reached significance, which was further analyzed under *Noise* (See Supplementary Materials for detailed results of post-hoc analysis). In the clean condition, no significant between-group differences were found for any of the three response types (all $ps > 0.05$). Within each group, children from HK were detected to make significantly more audio-dominant responses relative to audiovisual-integrated ones ($r = 0.58$, $p = 0.03$), while such differences failed to reach significance in their mainland peers ($r = 0.36$, $p = 0.23$). Besides, both HK and mainland children produced significantly more audio-dominant than visual-dominant responses in the clean condition (both $ps < 0.05$). As for the noisy condition, HK group made significantly more audiovisual-integrated responses while significantly fewer visual-dominant responses than the mainland group (both $ps < 0.05$). The two groups did not differ in the proportion of audio-dominant responses ($r = 0.11$, $p = 0.47$). Within-group comparisons further showed distinct response patterns across regions. For HK children, both audiovisual-integrated and visual-dominant responses were made significantly more often than audio-dominant responses (both $ps < 0.05$), while the percentages of audiovisual-integrated and visual-dominant responses did not significantly differ ($r = 0.44$, $p = 0.16$). In contrast, mainland children produced a comparable number of audio-dominant and audiovisual-integrated responses ($r = 0.43$, $p = 0.11$), whereas visual-dominant responses were made significantly more often than both (both $ps < 0.01$).

The identification of congruent stimuli. Figure 3 exhibits the accuracy of identifying congruent stimuli achieved in both clean and noisy conditions by the two groups of children. Permutation-based repeated measures ANOVA unraveled the significant main effects of *Region* ($F(1,42) = 6.35$, permutation $p = 0.01$, $\eta_p^2 = 0.02$), *Stimulus Type* ($F(1.76, 73.95) = 19.03$, permutation $p < 0.001$, $\eta_p^2 = 0.13$) as well as *Noise* ($F(1, 42) = 243.06$, permutation $p < 0.001$, $\eta_p^2 = 0.49$). Besides, *Region* × *Stimulus Type* ($F(1.76, 73.95) = 4.00$, permutation $p = 0.02$, $\eta_p^2 = 0.03$) and *Stimulus Type* × *Noise* ($F(1.92, 80.79) = 17.41$, permutation $p < 0.01$, $\eta_p^2 = 0.12$) interactions were also significant. Yet other interaction effects failed to attain the significance level ($ps > 0.05$).

The main effect of *Region* was probably given rise by the higher accuracy achieved by children from the mainland relative to their peers from HK ($Mdn_{\mathrm{Mainland}} = 1.00$, $Mdn_{\mathrm{HongKong}} = 0.86$, permutation $p = 0.01$). The main effect of *Stimulus Type* was driven by a significantly lower accuracy in identifying "Da" relative to "Ba" and "Ga" (both $ps < 0.05$), while the accuracy of identifying "Ba" and "Ga" did not significantly differed ($r = 0.16$, $p = 0.22$). For the *Noise* main effect, participants significantly performed better in the clean relative to the noisy condition ($r = 0.82$, $p < 0.01$). Moreover, the significant *Region* × *Stimulus Type* interaction indicated that the overall advantage of the mainland

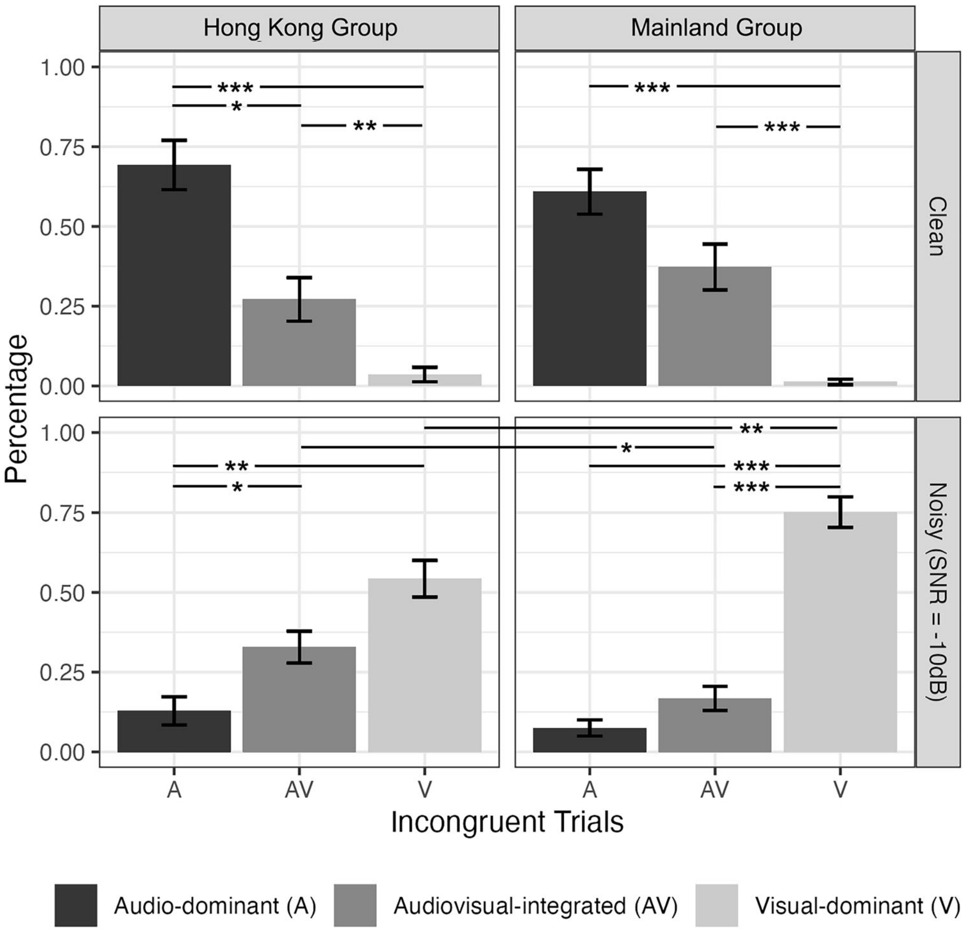

Fig. 2 The percentage of three response types to the incongruent stimuli by Cantonese-speaking children residing in HK and the mainland. Error bars indicate standard errors of the mean. Asterisks denote significance levels for post-hoc comparisons: *$p < .05$, **$p < .01$, ***$p < .001$.

group was driven by their higher accuracy in identifying "Ga" relative to the HK group ($r = 0.23$, $p = 0.03$), whereas no significant group differences were found for the "Ba" or "Da" trials (both ps > 0.05). In addition, the post-hoc analysis regarding *Stimulus Type* × *Noise* interaction revealed that participants showed higher accuracy in identifying "Ba" and "Ga" than "Da" in the noisy condition (both ps < 0.05), while accuracy for "Ba" and "Ga" did not differ from each other ($r = 0.30$, $p = 0.13$). In contrast, no accuracy differences across *Stimulus Type* were observed in the clean condition (all ps > 0.05).

Discussion

To explore how children growing up under different pandemic-related public health measures perform in audiovisual speech perception, the current study compared Cantonese-speaking 5–6-year-olds from Hong Kong and mainland China using the McGurk paradigm in both clean and noisy conditions. For the responses to the incongruent stimuli in the clean condition, an auditory reliance was witnessed by the significantly more audio-dominant compared to audiovisual-integrated responses in children residing in HK but not in their peers living in the mainland. Regarding the noisy condition, children from mainland China relied more on the visual modality than the HK group. For the congruent stimuli, the significant between-groups difference was demonstrated by the higher accuracy of identifying congruent stimuli, especially for the "Ga" trials, found in the mainland group relative to that in the HK group. With the comparison between two groups of children from different regions but native to the same language, these findings suggest that differences in social surroundings may have contributed to variation in perceptual strategies when processing audiovisual speech stimuli as those who experienced the tougher social restrictions in HK appeared less efficient in utilizing the visual information than their mainland counterparts.

Visual reliance was affected by social regulations during the pandemic. The comparison between Cantonese-speaking 5–6-year-old children growing up under differing social restrictions may provide insight into how regulation could influence the development of audiovisual speech perception. Overall, children in HK showed reduced reliance on visual information, particularly in the noisy condition, where they identified congruent stimuli less accurately and produced significantly more audiovisual-integrated responses but significantly fewer visual-dominant responses than their mainland peers when perceiving incongruent stimuli. A similar but weaker trend was also observed in the clean condition, as reflected by the reduced audiovisual integration within the HK group. Implementing preventive measures against COVID-19 may be associated with these results in two aspects.

First, impoverished meaningful visual cues in the social environment might alter the cue weighting in audiovisual speech perception by lowering the reliance on visual modality. Findings from the current study were in accordance with Chládková et al. (2021) who found perceivers lowered their reliance allocated to visual modality after one-month mandatory mask-wearing, even

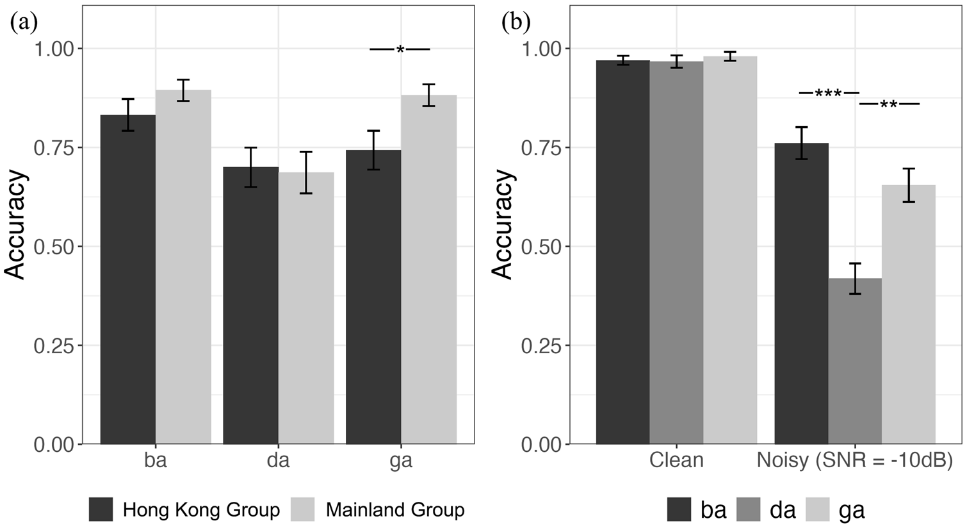

Fig. 3 Identification accuracy for congruent audiovisual speech stimuli across groups and conditions. a Identification accuracy for each stimulus type (/ba/, /da/, /ga/) in children from Hong Kong and mainland China. **b** Identification accuracy for each stimulus type in clean and noisy conditions, illustrating the Noise × Stimulus Type interaction. Error bars indicate standard errors of the mean. Asterisks denote significance levels for post-hoc comparisons: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

when the speaker's facial cues in their experiment were not masked. Similarly, in our experiment, child participants living in HK showed weaker visual utilization in the noisy condition, with a similar but weaker trend in the clean condition. Specifically, in the clean condition, responses made to the incongruent stimuli by HK Cantonese-speaking children were mainly based on the auditory component of the stimuli in the clean condition, while their peers from mainland China made a comparable number of audio-dominant and audiovisual-integrated responses. In the noisy condition where auditory information was severely disrupted, the compensative role of visual modality became more marked, which could be inferred in their performance on the congruent trials: /ba/, which involves a clear bilabial closure and thus offers highly salient visual information, was identified more accurately than /da/, whose alveolar articulation provides less visually distinct cues (van Wassenhove et al., 2005; Weng et al., 2024). Such increased visual dependence was corroborated by a shift to a visual-dominant strategy observed in both groups in the perception of incongruent stimuli, where HK children were also found to make significantly fewer visual-dominant responses to the incongruent stimuli than children from mainland China. According to the policy implemented, HK children had spent over 900 days in an environment where mask-wearing and social distancing were mandated by law, which significantly cut off access to visual speech information from children's social environment (Cheng, 2022). These policies also affected children's everyday experiences; reduced outdoor play and consistent masking during classroom activities further restricted their exposure to naturalistic visual speech input. Given the importance of input variability for language acquisition, this reduced variability in visual speech cues available to HK children may contribute to their differences in audiovisual speech development (Hoff, 2006; Samuel, 2011). A particularly critical factor concerns peer interaction. HK children were required to wear masks "at all times" at school, which heavily restricted their access to visual speech input during peer interactions (Charney et al., 2021). This is noteworthy as previous studies have highlighted the importance of peer interactions in supporting language and academic development (Burchinal et al., 2023). By comparison, children in low-risk areas in mainland China encountered far less masking on campus, as face coverings were not required in most preschool settings after reopening. Consequently, their access to visual speech cues in classroom communication was more consistent, which may help explain their greater reliance on visual information when processing audiovisual stimuli in both clean and noisy conditions.

Second, reduced and/or unclear exposure to human mouths in the pandemic era might lead to decreased sensitivity to fine-grained visual speech information, thereby affecting the development of speech recognition in noise. Combining the lower identification accuracy of congruent stimuli achieved by the HK group, the decreased visual utilization seemed to lead to weaker word identification ability, especially in auditorily adverse conditions. The ability to read speech through mouth movements was found to develop along with age, which was nurtured by experience (Tye-Murray et al., 2014). However, because children in HK had limited exposure to talking faces under COVID-19 preventive measures, the correspondence between facial movement and phonological components may be weakened. This weakened mapping may, in turn, have contributed to their lower effectiveness in using visual information for speech decoding (Latif et al., 2017) and thereby, to reduce reliance on visual modality (Bandaru et al., 2020; Brown et al., 2021; Charney et al., 2022; Giovanelli et al., 2021; Yi et al., 2021). Since the facilitative role of visual cues in audiovisual speech processing has been witnessed by a plethora of studies (e.g., Sumby and Pollack, 1954; van Wassenhove et al., 2005), it would be expected that HK children, who showed reduced efficiency and reliance on visual information, also demonstrated lower accuracy relative to the children from mainland China in identifying the real-word syllable, namely, "Ga", especially in auditorily adverse conditions (Hirst et al., 2018; Weng et al., 2024).

To sum up, the intensified preventive regulations implemented in HK may substantially reduce children's opportunities to view the facial movements of speaking partners, which may have weakened their reliance on visual cues in audiovisual speech and curtailed children's experience in acquiring the correspondence between facial movement and phonological components from the ambiance (Charney et al., 2022). This reduced efficiency in using visual information is also consistent with the HK group's lower accuracy in identifying congruent trials compared to their mainland peers, suggesting that more effective audiovisual integration may have contributed to the higher accuracy observed in the mainland group.

Implications for future practice. Findings from the current study align with concerns raised by scholars regarding the potential adverse effects of COVID-19 preventive policies on children's language development (e.g., Charney et al., 2021, 2022). One relevant manifestation is the possible disruption of the developmental course of audiovisual speech perception, a fundamental component of language acquisition. Children initially show limited sensitivity to visual speech cues and require a prolonged developmental period to acquire an adult-like ability to integrate auditory and visual information effectively (Robinson and Sloutsky, 2004, 2010). According to Weng et al. (2024), typically developing children gradually shift toward an audiovisual-integrated strategy in clean conditions and toward a visual-dominant strategy under severe noise, with corresponding improvements in speech recognition as they age. Against this developmental backdrop, the reduced visual utilization observed in the HK group may be linked to their differing pandemic-related experiences, which likely limited consistent access to talking faces and visual speech cues. Although causal claims cannot be made, this possibility carries meaningful implications for policymakers and manufacturers when considering how public health measures or device designs may influence children's multisensory learning environments.

Our findings suggest that cautious policymaking, such as implementing compulsory mask-wearing for children, should be carefully considered in the future. Children whose language ability is not yet fully mature are more susceptible to changes in their surrounding social and linguistic environment (Hoff, 2006). The substantial reduction in meaningful visual speech inputs during the COVID-19 pandemic not only increased communication difficulties among adults but also among children, who are naturally less sensitive to visual cues (Robinson and Sloutsky, 2004, 2010), possibly resulting in delays in their audiovisual speech development. Given the unprecedented scale of the pandemic, we believe that this outcome was both unexpected and challenging to address. Therefore, we suggest that in future public health crises, policies and measures will be developed with greater consideration for the developmental needs of children. For instance, the WHO recommends that children aged five years and under should not wear masks, as they may be unable to wear them correctly without guidance or supervision (WHO, 2022). Hence, it may be worth considering loosening the demand for mask-wearing on campus. Furthermore, prior studies also suggest that transparent masks can more effectively balance the dual goals of protection and communication (Brown et al., 2021; Lalonde et al., 2022; Thibodeau et al., 2021; Yi et al., 2021; Zhou et al., 2022). Thus, it also appears feasible for educators to wear transparent masks to better convey visual speech information during their teaching activities.

Given that the use of transparent masks has been revealed to considerably improve the quality and efficiency of communication, with some studies even suggesting that the perceptual accuracy achieved by adults using transparent masks can be comparable to conditions without masks (Thibodeau et al., 2021), mask manufacturers may consider further research into producing transparent masks. This would allow for clear access to visual information during face-to-face conversations, ensuring that children can receive visual speech information while being protected.

Limitations and future directions. The current study acknowledges the following limitations. First, the present study focused exclusively on 5–6-year-old children, and future work should include a broader age range to determine whether the observed patterns generalize across different developmental stages. Also, we did not assess children's internal state (e.g., discomfort or anxiety related to mask wearing) before or after the experiment. Given that many HK children arrived wearing masks and some showed reluctance to remove them, such factors may have influenced their comfort levels and should be considered in future work.

Furthermore, the present findings point to the importance of longitudinal research for clarifying how audiovisual speech perception develops over time. Our cross-sectional design provides only a snapshot of group differences, and it remains unclear whether the reduced visual reliance observed in HK children reflects a transient effect of pandemic-related experiences or a more persistent developmental difference. A longitudinal follow-up with the same cohort would help determine whether these group differences diminish, stabilize, or widen with age, thereby offering critical insight into the long-term developmental course and plasticity of audiovisual cue weighting. Additionally, future work could investigate whether different visual stimulus formats commonly used in masked-speech research (such as authentic versus digitally superimposed masks) exert distinct influences on children's perception of masked faces and audiovisual speech comprehension. Also, although all children in the current study scored within the typical range on ASD- and ADHD-related screening measures, future research with greater variability in these traits would be valuable for clarifying whether subclinical differences in social communication or attention modulate audiovisual speech perception.

Conclusion

The present study investigated the impact of social regulations implemented during the COVID-19 pandemic on audiovisual speech perception in Cantonese-speaking children. By comparing the performance of Cantonese-speaking 5–6-year-olds from HK and mainland China, we found that children in HK, who underwent stricter and more prolonged social distancing and mask mandates, tended to be more reliant on auditory cues while less efficient in utilizing visual information during audiovisual speech perception. These findings highlight that changes in the social environment resulting from public health crises may be one of the contributing factors that shape children's speech perception strategies, potentially influencing their developmental trajectories. The results demonstrate the importance of paying greater attention to the needs of children in sensitive developmental periods when formulating future public health policies.

Data availability

Summary data (means and standard deviations) are provided in the Supplementary Materials. Raw data are available from the corresponding author upon reasonable request.

Received: 15 June 2025; Accepted: 8 April 2026;
Published online: 23 April 2026

References

Aguillon-Hernandez N, Jusiak R, Latinus M, Wardak C (2022) COVID-19 masks: a barrier to facial and vocal information. Front Neurosci 16:982899. https://doi.org/10.3389/fnins.2022.982899

Auyeung B, Baron-Cohen S, Wheelwright S, Allison C (2008) The Autism Spectrum Quotient: Children's Version (AQ-Child). J Autism Dev Disord 38(7):1230–1240. https://doi.org/10.1007/s10803-007-0504-z

Bandaru SV, Augustine AM, Lepcha A, Sebastian S, Gowri M, Philip A, Mammen MD (2020) The effects of N95 mask and face shield on speech perception among healthcare workers in the coronavirus disease 2019 pandemic scenario. J Laryngol Otol 134(10):895–898. https://doi.org/10.1017/S0022215120002108

Bass HE, Sutherland LC, Zuckerwar AJ, Blackstock DT, Hester DM (1995) Atmospheric absorption of sound: further developments. J Acoust Soc Am 97(1):680–683. https://doi.org/10.1121/1.412989

Bottalico P, Murgia S, Puglisi GE, Astolfi A, Kirk KI (2020) Effect of masks on speech intelligibility in auralized classrooms. J Acoust Soc Am 148(5):2878–2884. https://doi.org/10.1121/10.0002450

Brown VA, Van Engen KJ, Peelle JE (2021) Face mask type affects audiovisual speech intelligibility and subjective listening effort in young and older adults. Cogn Res Princ Implic 6(1):49. https://doi.org/10.1186/s41235-021-00314-0

Burchinal M, Pianta R, Ansari A, Whittaker J, Vitiello V (2023) Kindergarten academic and social skills and exposure to peers with pre-kindergarten experience. Early Child Res Q 62:41–52. https://doi.org/10.1016/j.ecresq.2022.07.012

Carnevali L, Gui A, Jones EJH, Farroni T (2022) Face processing in early development: a systematic review of behavioral studies and considerations in times of COVID-19 pandemic. Front Psychol 13. https://doi.org/10.3389/fpsyg.2022.778247

Charney SA, Camarata SM, Chern A (2021) Potential impact of the COVID-19 pandemic on communication and language skills in children. Otolaryngol–Head Neck Surg 165(1):1–2. https://doi.org/10.1177/0194599820978247

Charney SA, Camarata SM, Chern A (2022) The impact of the COVID-19 pandemic on children's speech and language development. In COVID-19 and speech-language pathology. Routledge, p 20–37

Chládková K, Podlipský VJ, Nudga N, Šimáčková Š (2021) The McGurk effect in the time of pandemic: Age-dependent adaptation to an environmental loss of visual speech cues. Psychon Bull Rev 28(3):992–1002. https://doi.org/10.3758/s13423-020-01852-2

Coniam D (2005) The impact of wearing a face mask in a high-stakes oral examination: an exploratory post-SARS study in Hong Kong. Lang Assess Q 2(4):235–261. https://doi.org/10.1207/s15434311laq0204_1

Education Bureau (2020) Letter to Schools on "Stay Vigilant and Fight COVID-19 Together". https://www.edb.gov.hk/attachment/en/sch-admin/admin/about-sch/diseases-prevention/edb_20201008_eng.pdf

Ernst MO (2008) Multisensory integration: a late bloomer. Curr Biol 18(12):R519–R521. https://doi.org/10.1016/j.cub.2008.05.002

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870):429–433. https://doi.org/10.1038/415429a

Fritz CO, Morris PE, Richler JJ (2012) Effect size estimates: current use, calculations, and interpretation': Correction to Fritz et al. (2011). J Exp Psychol Gen 141(1):30–30. https://doi.org/10.1037/a0026092

Frossard J, Renaud O (2021) Permutation tests for regression, ANOVA, and comparison of signals: the permuco package. J Stat Software, 99(15). https://doi.org/10.18637/jss.v099.i15

Giovanelli E, Valzolgher C, Gessa E, Todeschini M, Pavani F (2021) Unmasking the difficulty of listening to talkers with masks: lessons from the COVID-19 pandemic. I-Percept 12(2):204166952199839. https://doi.org/10.1177/2041669521998393

Hay DF, Payne A, Chadwick A (2004) Peer relations in childhood. J Child Psychol Psychiatry 45(1):84–108. https://doi.org/10.1046/j.0021-9630.2003.00308.x

Hecht H, Welsch R, Viehoff J, Longo MR (2019) The shape of personal space. Acta Psychologica 193:113–122. https://doi.org/10.1016/j.actpsy.2018.12.009

Hirst RJ, Stacey JE, Cragg L, Stacey PC, Allen HA (2018) The threshold for the McGurk effect in audio-visual noise decreases with development. Sci Rep 8(1). https://doi.org/10.1038/s41598-018-30798-8

Hoff E (2006) How social contexts support and shape language development. Dev Rev 26(1):55–88. https://doi.org/10.1016/j.dr.2005.11.002

Jordan TR, Sergeant P (2000) Effects of distance on visual and audiovisual speech recognition. Lang Speech 43(1):107–124. https://doi.org/10.1177/00238309000430010401

Karl KA, Peluchette JV, Aghakhani N (2022) Virtual work meetings during the COVID-19 pandemic: the good, bad, and ugly. Small Group Res 53(3):343–365. https://doi.org/10.1177/10464964211015286

Kassambara A (2023) rstatix: Pipe-friendly framework for basic statistical tests (Version 0.7.2) [R package]. https://rpkgs.datanovia.com/rstatix/

Kleinman D, Morgan AM, Ostrand R, Wittenberg E (2022) Lasting effects of the COVID-19 pandemic on language processing. PLOS ONE 17(6):e0269242. https://doi.org/10.1371/journal.pone.0269242

Kressner AA, Westermann A, Buchholz JM (2018) The impact of reverberation on speech intelligibility in cochlear implant recipients. J Acoust Soc Am 144(2):1113–1122. https://doi.org/10.1121/1.5051640

Lalonde K, Buss E, Miller MK, Leibold LJ (2022) Face masks impact auditory and audiovisual consonant recognition in children with and without hearing loss. Front Psychol 13. https://doi.org/10.3389/fpsyg.2022.874345

Latif N, Alsius A, Munhall KG (2017) Seeing the Way: The Role of Vision in Conversation Turn Exchange Perception. https://doi.org/10.1163/22134808-00002582

Lipps, E, Caldwell-Kurtzman, J, Motlagh-Zadeh, L, Blankenship, CM, Moore, DR, & Hunter, LL (2021). Impact of face masks on audiovisual word recognition in young children with hearing loss during the Covid-19 pandemic. J Early Hear Detect Interv 9(1). https://doi.org/10.26077/4fda-c155

Little KB (1965) Personal space. J Exp Soc Psychol 1(3):237–247. https://doi.org/10.1016/0022-1031(65)90028-4

Maidment DW, Kang HJ, Stewart HJ, Amitay S (2015) Audiovisual integration in children listening to spectrally degraded speech. J Speech, Lang Hear Res 58(1):61–68. https://doi.org/10.1044/2014_JSLHR-S-14-0044

McGurk H, Macdonald J (1976) Hearing lips and seeing voices. Nature 264(5588):746–748. https://doi.org/10.1038/264746a0

Meyer J, Dentel L, Meunier F (2013) Speech recognition in natural background noise. PLOS ONE 8(11):e79279. https://doi.org/10.1371/journal.pone.0079279

Ministry of Education & National Health Commission (2020) Press Conference of the Joint Preventive and Control Mechanism of the State Council. https://www.gov.cn/xinwen/gwylflkjz122/index.htm

Moon I-J, Jo M, Kim G-Y, Kim N, Cho Y-S, Hong S-H, Seol H-Y (2022) How does a face mask impact speech perception?. Healthcare 10(9):1709. https://doi.org/10.3390/healthcare10091709

Paulus M (2018) Preschool children's and adults' expectations about interpersonal space. Front Psychol 9. https://doi.org/10.3389/fpsyg.2018.02479

R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rahne T, Fröhlich L, Plontke S, Wagner L (2021) Influence of surgical and N95 face masks on speech perception and listening effort in noise. PLOS ONE 16(7):e0253874. https://doi.org/10.1371/journal.pone.0253874

Robinson CW, Sloutsky VM (2004) Auditory dominance and its change in the course of development. Child Dev 75(5):1387–1401. https://doi.org/10.1111/j.1467-8624.2004.00747.x

Robinson CW, Sloutsky VM (2010) Development of cross-modal processing. Wiley Interdiscip Rev: Cognit Sci 1(1):135–141. https://doi.org/10.1002/WCS.12

Samuel AG (2011) Speech perception. Annu Rev Psychol 62(1):49–72. https://doi.org/10.1146/annurev.psych.121208.131643

Sönnichsen R, Tó GL, Hohmann V, Hochmuth S, Radeloff A (2022) Challenging times for cochlear implant users – effect of face masks on audiovisual speech understanding during the COVID-19 pandemic. Trends Hear 26:233121652211343. https://doi.org/10.1177/23312165221134378

Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26(2):212–215. https://doi.org/10.1121/1.1907309

Talbot-Smith M (1993) Sound, speech and hearing. In Telecommunications Engineer's Reference Book. Elsevier, p 8–16 https://doi.org/10.1016/B978-0-7506-1162-6.50014-9

Tan TX, Wang JH, Zhou Y (2024) COVID-19 school closures and Chinese children's school readiness: results from the natural experimental data. Br J Educ Psychol 94(3):976–994. https://doi.org/10.1111/bjep.12699

The Government of the Hong Kong Special Administrative Region. (2020) Prevention and Control of Disease Regulation. https://www.info.gov.hk/gia/general/202007/14/P2020071400037.htm

The Government of the Hong Kong Special Administrative Region. (2023) Government lifts all mandatory mask-wearing requirements. https://www.info.gov.hk/gia/general/202302/28/P2023022800431.htm

Thibodeau LM, Thibodeau-Nielsen RB, Tran CMQ, Jacob RTDS (2021) Communicating during COVID-19: the effect of transparent masks for speech recognition in noise. Ear Hear 42(4):772–781. https://doi.org/10.1097/AUD.0000000000001065

Tremblay C, Champoux F, Voss P, Bacon BA, Lepore F, Théoret, H (2007) Speech and non-speech audio-visual illusions: a developmental study. https://doi.org/10.1371/journal.pone.0000742

Tye-Murray N, Hale S, Spehar B, Myerson J, Sommers MS (2014) Lipreading in school-age children: the roles of age, hearing status, and cognitive ability. J Speech Lang, Hear Res 57(2):556–565. https://doi.org/10.1044/2013_JSLHR-H-12-0273

Uzzell D, Horne N (2006) The influence of biological sex, sexuality and gender role on interpersonal distance. Br J Soc Psychol 45(3):579–597. https://doi.org/10.1348/014466605X58384

van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci USA 102(4):1181–1186. https://doi.org/10.1073/pnas.0408949102

Weng Y, Rong Y, Peng G (2024) The development of audiovisual speech perception in Mandarin-speaking children: evidence from the McGurk paradigm. Child Dev 95(3):750–765. https://doi.org/10.1111/cdev.14022

Wenner Moyer M (2022) The COVID generation: how is the pandemic affecting kids' brains?. Nature 601(7892):180–183. https://doi.org/10.1038/d41586-022-00027-4

Willis FN, Carlson R, Reeves D (1979) The development of personal space in primary school children. Environ Psychol Nonverbal Behav 3(4):195–204. https://doi.org/10.1007/BF01127363
Wolraich ML (2003) Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. J Pediatr Psychol 28(8):559–568. https://doi.org/10.1093/jpepsy/jsg046
Wong KM, Wang Z, Fu L (2023) Lessons Hong Kong Kindergarten Teachers Learned From the Pandemic. Child Educ 99(4):56–61. https://doi.org/10.1080/00094056.2023.2232282
World Health Organization (2022) Coronavirus disease (COVID-19): Children and masks. https://www.who.int/news-room/questions-and-answers/item/q-a-children-and-masks-related-to-covid-19
World Health Organization (2023) Coronavirus disease (COVID-19): Masks. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-masks
Yi H, Pingsterhaus A, Song W (2021) Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. Front Psychol 12:682677. https://doi.org/10.3389/fpsyg.2021.682677
Zhao SZ, Wong JYH, Wu Y, Choi EPH, Wang MP, Lam TH (2020) Social distancing compliance under COVID-19 pandemic and mental health impacts: a population-based study. Int J Environ Res Public Health, 17(18), Article 18. https://doi.org/10.3390/ijerph17186692
Zheng Y, Samuel AG (2019) How much do visual cues help listeners in perceiving accented speech?. Appl Psycholinguist 40(1):93–109. https://doi.org/10.1017/S0142716418000462
Zhou P, Zong S, Xi X, Xiao H (2022) Effect of wearing personal protective equipment on acoustic characteristics and speech perception during COVID-19. Appl Acoust 197:108940. https://doi.org/10.1016/j.apacoust.2022.108940

Acknowledgements
This research was partly supported by a fellowship award from the Research Grants Council of the Hong Kong SAR, China (Project No. PolyU/RFS2122-5H01).

Author contributions
YW, YR, and GP designed research; YW and YR performed research; YW and YR analyzed data; and YW, YR, and GP wrote the paper. YW and YR contributed equally to this research.

Competing interests
The authors declare no competing interests.

Ethical approval
This study was approved by The PolyU Institutional Review Board (Approval Number: HSEARS20220213003; Date of approval: 21 February 2022). All procedures involving human participants were conducted in accordance with the guidelines and regulations of The Hong Kong Polytechnic University and the principles outlined in the Declaration of Helsinki. The approved research involved behavioral investigation of speech perception in Cantonese-speaking children with and without Autism Spectrum Disorder (ASD). Written informed consent was obtained from the legal guardians of all participating children prior to data collection.

Informed consent
Written informed consent was obtained from the parents of all child participants between March and October 2023. Prior to the onset of the study, parents and participants were thoroughly informed about the study's background, purpose, procedures, potential risks, and data usage through an information sheet. Parents were given the opportunity to ask questions to ensure they could make an informed and voluntary decision. After fully understanding the purpose of the study, they signed a consent form to indicate their agreement for their children to participate. To maintain confidentiality and anonymity, participants were assigned unique codes once they were enrolled in the study. Parents and participants were also informed of their right to withdraw from the study at any time without penalty. All informed consent documents are securely stored to protect the privacy of the child participants but can be made available upon reasonable request to the corresponding author.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-026-07281-1.

Correspondence and requests for materials should be addressed to Gang Peng.