frontiers | Frontiers in Communication

TYPE Brief Research Report
PUBLISHED 13 June 2025
DOI 10.3389/fcomm.2025.1582217

OPEN ACCESS

EDITED BY
Weijun Li,
Liaoning Normal University, China

REVIEWED BY
Yu-Fu Chien,
Fudan University, China
Yudong Chen,
Communication University of China, China

*CORRESPONDENCE
William Choi
willchoi@hku.hk

RECEIVED 24 February 2025
ACCEPTED 29 May 2025
PUBLISHED 13 June 2025

CITATION
Choi W, Chu T and Zu J (2025) Can you hear me clearly? The differential effects of surgical mask on Cantonese consonant, vowel, and tone perception.
Front. Commun. 10:1582217.
doi: 10.3389/fcomm.2025.1582217

Can you hear me clearly? The differential effects of surgical mask on Cantonese consonant, vowel, and tone perception

William Choi[1,2,3]*, Tianyu Chu[1,2] and Jiaqing Zu[1,2]

[1]Academic Unit of Human Communication, Learning, and Development, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong SAR, China, [2]Speech and Music Perception Laboratory, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong SAR, China, [3]Centre for Advancement in Inclusive and Special Education, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

This study examined the differential effects of surgical mask on Cantonese consonant, vowel, and tone perception. Forty native Cantonese adults were tested with the Cantonese consonant, vowel, and tone identification tasks. Each task contained four blocks: quiet-no mask, noisy-no mask, quiet-surgical mask, and noisy-surgical mask. Bayesian analyses revealed that the Cantonese listeners identified consonants, vowels, and tones with similar accuracies across the four blocks. However, in the presence of noise, surgical mask was found to increase the response time in identifying vowels. From a theoretical perspective, this study offers a phonological account to explain why surgical mask may impede sentence comprehension. Practically, the findings suggest that surgical mask has little bearing on the ability to accurately identify Cantonese consonants, vowels, and tones, though it affects the efficiency in vowel identification.

KEYWORDS

COVID, face mask, speech, tone, vowel, consonant, perception, Cantonese

1 Introduction

The use of face mask has become increasingly prevalent in our daily lives, especially in the wake of the COVID-19 pandemic (Badillo-Goicoechea et al., 2021; Fischer et al., 2021). While the primary purpose of face mask is to reduce the transmission of respiratory droplets, their potential impact on communication has garnered attention (Badh and Knowles, 2023; Mendel et al., 2022; Zhou et al., 2022; Sinagra and Wiener, 2022). Many studies showed that face mask negatively impacted speech perception, purportedly due to degraded auditory and visual information (Moon et al., 2022; Schwarz et al., 2022; Zhou et al., 2022; c.f. Toscano and Toscano, 2021). However, these studies have only tested speech perception at the sentence level. Linguistically, sentences are composed of words, which in turn consist of consonants, vowels, and in the case of tone languages, tones. To gain a deeper understanding of how face mask hinders speech perception, it is crucial to examine not just the sentence level, but also the phonological level. To this end, we examined the differential effects of face mask on consonant, vowel, and tone perception.

During the COVID-19 pandemic, governments implemented face mask mandates in the interest of public health and safety. For example, Hong Kong citizens were legally required to wear a face mask in public areas from 2020 to 2023 (Centre for Health Protection, 2023). Although face mask mandates have been lifted, they will likely be reinstated in the unavoidable event of future respiratory disease pandemics or during influenza seasons. Therefore, face mask will always remain a relevant topic to the public and researchers. In the university and

school settings, the ability to perceive teachers' speech is essential to learning success. However, face mask has been frequently reported to impede speech perception (Badh and Knowles, 2023).

Auditory information plays a crucial role in speech perception. All spoken languages in the world contain two basic phonological units, namely consonants and vowels. Acoustically, formant frequencies play a key role in defining and distinguishing vowels and consonants in speech (Cutler, 2015). Vowels are primarily characterized by their distinct formant patterns, with each vowel having a unique combination of formant frequencies (Ladefoged and Johnson, 2011). For instance, the vowel /i/ is defined by a low first formant frequency (F1) and a relatively high second formant frequency (F2), resulting in its perception as a front, close vowel. Consonants, on the other hand, are characterized by their formant transitions and spectral properties (Ladefoged and Johnson, 2011). For example, the consonant /s/ is characterized by a high-frequency spectral noise, with a rapid decrease in formant frequencies during its production. These formant cues provide crucial information for accurate perception and discrimination of vowels and consonants in spoken language. In tone languages such as Cantonese, tones have distinctive fundamental frequency (F0) patterns (Choi et al., 2017a, 2017b; Choi and Chiu, 2023). For example, the Cantonese high level tone has a high F0 height and flat F0 contour, whereas the low rising tone has a low F0 height and a rising F0 contour.

Besides auditory information, visual information plays some role in speech perception. Researchers have suggested that visual information, primarily derived from observing mouth movements and shape, provide complementary information that enhances speech perception (Balan and Maruthy, 2018; Hazan et al., 2006; Yuan et al., 2021). For example, the vowel /a/ has an open mouth shape whereas the vowel /i/ has a closed mouth shape. Regarding mouth movement, the consonant /w/ involves a lip protrusion movement whereas the consonant /k/ does not. In the articulatory aspect, tone production involves modulating the rate of laryngeal vibration instead of supralaryngeal articulators, so visual information does not cue tones (Hong et al., 2023). Although visual information is not essential for consonant and vowel perception in normal conditions (e.g., we can still identify speech accurately on phone), their functions will become more evident when acoustic signals are degraded (e.g., noisy) or poorly detected (e.g., deafness) (Giovanelli et al., 2021; Yi et al., 2021). However, visual information will be blocked when we wear a face mask, because it physically covers our mouth. While wearing transparent face mask can be a potential alternative, they are less capable of transmitting auditory information than the traditional ones (Atcherson et al., 2017, 2021), and thus are not adopted in our study.

In addition to obscuring visual information, face mask is also found to degrade auditory information (Atcherson et al., 2021; Giuliani, 2020; Zhang et al., 2025). A recent study analyzed the sentence productions of three Cantonese talkers across various conditions: no mask, surgical mask, KF94 mask, face shield, and face shield with surgical mask (Zhang et al., 2025). It was found that surgical and KF94 masks not only attenuated sound pressure levels more intensely at higher frequency ranges, but they also shrank the vowel spaces derived from formant frequencies. Moreover, they influenced tone production, by increasing the F0 height for the high level tone, and shortening the tone duration for both the high and low level tones. Collectively, the results suggest that face mask can degrade the acoustic information of consonants, vowels, and some tones, supporting the view that face mask may act as a physical low-pass filter which attenuates acoustic information, especially in higher frequency regions.

Of interest to us is the perceptual consequence of face mask. Theoretically, face mask degrades and obscures the auditory and visual information of speech, so it should negatively affect speech perception. A recent study assessed Mandarin listeners' signal-to-noise ratio (SNR) at which they could understand 50% of the 20 sentences heard (Zhou et al., 2022). There were four conditions, namely no-mask, surgical mask, N95 mask with face shield, and transparent mask. Relative to their own performance in the no-mask and transparent mask conditions, Mandarin listeners required a higher SNR in the surgical mask and N95 mask with face shield conditions. This suggested that surgical and KN95 masks hindered sentence comprehension.

However, the findings have not always been consistent (Mendel et al., 2022; Moon et al., 2022; Schwarz et al., 2022). In a recent study, English adults and children were tasked with repeating the last word of sentences read aloud by a talker with and without surgical mask (Schwarz et al., 2022). As expected, the listeners had a lower accuracy and longer response time in the surgical mask condition. However, in the presence of semantic cues, the negative effect of surgical mask was eliminated among the adults and mitigated among the children. In another study, there was simply no significant effect of surgical mask and KN95 mask on English sentence repetition (Mendel et al., 2022). Complicating the picture, another study found a negative effect of KF94 mask on Korean sentence repetition, but the effect disappeared when listeners could see the talker (albeit wearing a mask) (Moon et al., 2022). Collectively, there is some inconsistent evidence suggesting that face mask may hinder speech perception at the sentence level.

The present study set out to address a research gap, namely the differential effects of face mask on consonant, vowel, and tone perception. As reviewed, previous studies have focused on sentence comprehension or repetition (e.g., Mendel et al., 2022; Moon et al., 2022; Zhou et al., 2022). While it is more functional to assess the impact of face mask at the sentence level, assessing *only* sentences will limit our understanding of how face mask impedes speech perception. From a phonological perspective, consonants, vowels, and tones (in tone languages) are the basic phonological units of speech (Bauer and Benedict, 1997; Ladefoged and Johnson, 2011). If face mask impedes speech perception at the sentence level, is it because it impedes consonants, vowels, tones, or all of them?

We hypothesize that face mask has more detrimental effects on consonant and vowel perception than on tone perception. Acoustically, face mask dampens sounds at higher frequencies over 1.5 kHz (Giuliani, 2020). This high frequency range corresponds to formant frequencies which cue consonants and vowels. On the other hand, tones are F0 variations and the average F0 of human voices are only 195.8 Hz for female voices and 112 Hz for male voices, far below 1.5 kHz (Oliveira et al., 2021). Visually, face mask occludes mouth shape and movement which cue consonants and vowels especially in noise (Giovanelli et al., 2021; Yi et al., 2021). However, visual cues are not relevant to tones (Hong et al., 2023). Collectively, face mask attenuates auditory information and occlude visual ones, but these may be more detrimental to the perception of consonants and vowels

compared to tones. Therefore, we posit that face mask would hinder consonant and vowel perception, but not tone perception.

In daily lives, speech is seldom heard in a quiet environment like the sound booth. To generalize our potential findings beyond the laboratory setting, we added a noisy condition in which audio stimuli were presented along with environmental noise. While most previous studies only measured accuracy, our study measured response time as well (e.g., Mendel et al., 2022; Moon et al., 2022). This enabled us to reflect the impact of face mask on both *accuracy* and *efficiency* in perceiving consonants, vowels, and tones.

In short, *what is known* is that face mask hinders speech perception at the sentence level. *What is not known* is whether face mask hinders speech perception at the phonological level. *We hypothesize that* face mask hinders the perception of consonants and vowels but not tones. As the surgical mask is the most commonly used type of face mask in Hong Kong, we focused on it for maximal practical impact. To recap, our research question is: What are the differential effects of surgical mask on consonant, vowel, and tone perception?

2 Methods

We recruited 40 native Cantonese listeners (14 male and 26 female) from The University of Hong Kong. They met the following inclusion criteria: (1) learnt Cantonese as mother tongue, (2) self-reported typical hearing, (3) typical or typical corrected vision, and (4) no history of speech and language disorder. As one participant dropped out during the experiment, the final sample contained 39 participants (mean age = 28 years, *SD* = 11.84 years).

The experiment took place in a sound booth at The University of Hong Kong. The participants completed the consonant, vowel, and tone identification tasks in a randomized order. During the experiment, the participants wore Sennheiser HD280 PRO headphones. The tasks were run on PsychoPy using a laptop. The internal consistency of the identification tasks was high (Cronbach's alpha reliability = 0.89).

2.1 Consonant identification task

2.1.1 Stimuli

The stimuli contained /pa1/ 爸, /tsʰa1/ 叉, /fa1/ 花, /ka1/ 家, /ha1/ 蝦, and /wa1/ 蛙. They are real Cantonese words that differ only in their consonants, but not vowels and lexical tones. As Supplementary Table S1 shows, the consonants /p/, /tsʰ/, /f/, /k/, /h/, and /w/ represent different places (bilabial, alveolar, labiodental, velar, glottal, and labial-velar) and manners of articulation (plosive, affricate, fricative, and glide).

Two native Cantonese talkers, one male and one female, took part in the stimuli recording. They were undergraduate students majoring in Speech-language Pathology. Audio recording was done in a sound booth with a Shure SM58 Dynamic Vocal Mic at a sampling rate of 44.1 kHz. Video recording was done with a camera of professional quality. The two talkers were asked to sit 10 cm away from the microphone and naturally produce the words in isolation. After recording the unmasked stimuli, they were given a surgical mask. They then recorded the same words while wearing the surgical mask. Earlier research indicated that individuals might adjust their speaking style under challenging conditions (Nguyen et al., 2021). Thus, we requested the talkers to read the words in their usual manner and avoid making deliberate modifications to their voices while wearing a surgical mask, such as speaking louder or articulating excessively (Zhang et al., 2025). The audio stimuli are available on OSF (https://osf.io/7snq8/?view_only=7d240ee8c715446ba8173fee2ea7de61).

2.1.2 Stimuli presentation

We adopted a forced-choice identification task with four blocks: quiet-no mask, noisy-no mask, quiet-surgical mask, and noisy-surgical mask. In the no-mask blocks, the talkers did not wear any surgical mask. In the surgical mask blocks, the talkers wore a surgical mask (see Supplementary Figure S1). On each trial, an audio and a video were simultaneously presented with a headphone and a computer screen, respectively. In the noisy but not the quiet blocks, a white noise of 52 dBSPL collected from a school environment was presented along with the audio stimuli. This level of noise was on a par with the noise levels suggested in previous studies (Atcherson et al., 2017; Hampton et al., 2020). After stimulus presentation, six written words 爸, 叉, 花, 家, 蝦, and 蛙 were simultaneously shown on the computer screen (see Supplementary Figure S2). The participants then chose as quickly as possible the written word which corresponded to the stimulus presented. The order of presentation of blocks was randomized. The task started with four practice trials, in which feedback on accuracy was provided. Then, there were 96 experimental trials without feedback (6 syllables × 2 talkers × 2 repetitions × 4 blocks). On each trial, we recorded the accuracy and response time.

2.2 Vowel identification task

2.2.1 Stimuli

The stimuli contained /si1/ 詩, /sy1/ 書, /sa1/ 沙, /sou1/ 蘇, /sɵy1/ 需, and /sɐu1/ 收. They are real Cantonese words that differ only in their vowels, but not consonants and lexical tones. Half of the vowels are monophthongs and the other half are diphthongs. The vowels differ in tongue height, blackness, and/or lip rounding (see Supplementary Table S2). They were recorded by the same talkers with the same set-up and procedure as above. Figure 1 illustrates the vowel space areas encompassed by the three extreme vowels /a/, /i/, and /ou/.

2.2.2 Stimuli presentation

We adopted the same procedure as the consonant identification task. There were 96 experimental trials in total (6 syllables × 2 talkers × 2 repetitions × 4 blocks).

2.3 Tone identification task

2.3.1 Stimuli

The stimuli contained /ji1/ 衣, /ji2/ 椅, /ji3/ 意, /ji4/ 兒, /ji5/ 耳, and /ji6/ 二. They are real Cantonese words that differ only in their tones, but not consonants and vowels. The tones differ mainly in F0 height, F0 contour, or both (see Supplementary Table S3). They were recorded by the same talkers with the same set-up and procedure as above.

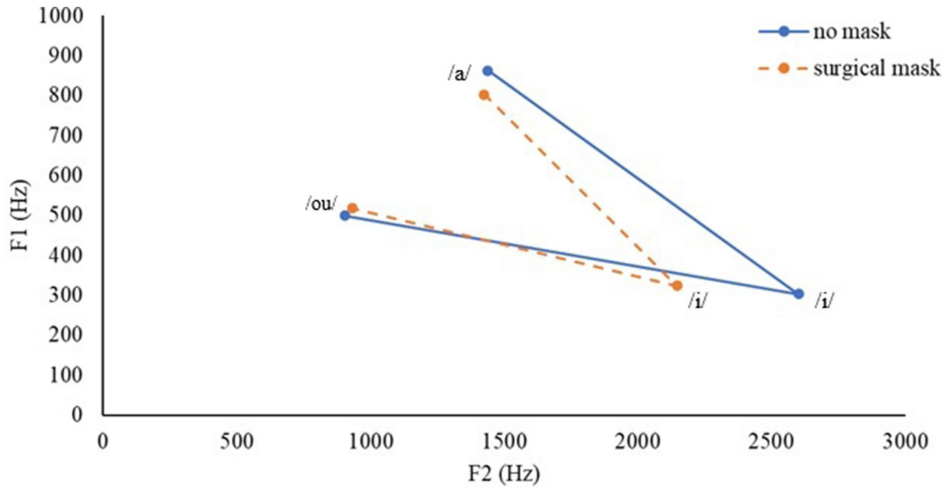

FIGURE 1
Mean vowel space areas of the vowel stimuli produced by both talkers with and without a surgical mask.

TABLE 1 Model comparisons relative to the best-fit models of mean accuracy.

Model	P(M)	P(M|data)	BF_M	BF_{01}	Error %
Consonant models					
Null	0.20	0.63	6.82	1.00	
Noise	0.20	0.17	0.81	3.73	2.89
Mask	0.20	0.15	0.71	4.16	1.93
Mask + Noise	0.20	0.04	0.17	15.63	2.26
Mask + Noise + Mask * Noise	0.20	0.01	0.04	68.82	2.55
Vowel models					
Null	0.20	0.49	3.90	1.00	
Mask	0.20	0.27	1.50	1.81	4.43
Noise	0.20	0.14	0.64	3.57	2.32
Mask + Noise	0.20	0.07	0.32	6.63	2.28
Mask + Noise + Mask * Noise	0.20	0.02	0.08	23.86	3.37
Tone models					
Null	0.20	0.66	7.83	1.00	
Mask	0.20	0.15	0.71	4.39	1.08
Noise	0.20	0.15	0.68	4.54	0.94
Mask + Noise	0.20	0.03	0.14	19.75	1.39
Mask + Noise + Mask * Noise	0.20	0.01	0.03	81.90	2.73

2.3.2 Stimuli presentation

We adopted the same procedure as above and there were 96 experimental trials (6 syllables × 2 talkers × 2 repetitions × 4 blocks).

3 Results

3.1 Mean accuracy analysis

We conducted Bayesian two-way ANOVAs on the mean accuracy in the consonant, vowel, and tone identification tasks. The within-subject factors were mask (no-mask and surgical mask) and noise (quiet and noisy). We used a uniform prior that is relatively objective (Jaynes, 2003). When interpreting the Bayes factor, we took reference of the strength of evidence which was heuristically defined by Lee and Wagenmakers (2013).

In the consonant, vowel, and tone identification tasks, the best fit models were the null model (see Table 1 and Supplementary Figure S3). Their BF_{01} ranged from 3.73 to 68.82, 1.81 to 23.86, and 4.39 to 81.90, respectively. In other words, there was moderate to very strong, anecdotal to strong, and moderate to very strong evidence favoring the null model in the consonant, vowel, and tone perception tasks, respectively.

TABLE 2 Model comparisons relative to the best-fit models of mean response time.

Model	P(M)	P(M|data)	BF_M	BF_{01}	Error %
Consonant models					
Null	0.20	0.55	4.90	1.00	
Noise	0.20	0.24	1.24	2.33	1.85
Mask	0.20	0.14	0.64	4.02	2.85
Mask + Noise	0.20	0.06	0.26	8.90	2.80
Mask + Noise + Mask * Noise	0.20	0.01	0.06	39.95	2.33
Vowel models					
Mask + Noise + Mask * Noise	0.20	0.77	13.59	1.00	
Noise	0.20	0.10	0.42	8.11	7.84
Null	0.20	0.09	0.40	8.58	8.06
Mask + Noise	0.20	0.02	0.09	36.06	7.96
Mask	0.20	0.02	0.08	37.45	9.21
Tone models					
Noise	0.20	0.67	8.16	1.00	
Mask + Noise	0.20	0.17	0.82	3.96	7.17
Mask + Noise + Mask * Noise	0.20	0.14	0.67	4.68	3.96
Null	0.20	0.01	0.05	50.59	3.08
Mask	0.20	0.00	0.01	223.69	3.75

3.2 Mean response time analysis

We conducted similar Bayesian two-way ANOVAs as above, but with mean response time in correct trials as the dependent variable. For consonant perception, the best-fit model was the null model, with the BF_{01} ranging from 2.33 to 39.95, indicating anecdotal to very strong evidence (see Table 2 and Figure 2).

For vowel perception, the best-fit model was the full model with the main effects of mask and noise, and their interaction (see Table 2). The BF_{01} ranged from 8.11 to 37.45, indicating moderate to very strong evidence. Next, we unpacked the interaction between mask and noise at the level of noise (see Supplementary Table S4). In the quiet condition, the null model was favored (BF_{01} = 1.44). However, in the noisy condition, the best fit model was the model with the main effect of mask (BF_{01} = 2.63). Post-hoc comparison showed that the mean response time was longer in the masked condition than in the unmasked condition (BF_{10} = 2.45).

For tone perception, the best-fit model was the model with the main effect of noise (see Table 2). The BF_{01} ranged from 3.96 to 223.69, indicating moderate to extreme evidence. Post-hoc comparison showed that the response time was longer in the noisy condition than in the quiet condition (BF_{10} = 459.69).

4 Discussion

This study set out to investigate the differential effects of surgical mask on consonant, vowel, and tone perception. We hypothesized that surgical mask would hinder the identification of consonants and vowels but not tones. Consistent with our hypothesis, surgical mask did not hinder tone identification. Unlike previous studies which used frequentist analysis (e.g., Toscano and Toscano, 2021), the present study used Bayesian analysis. Therefore, the similar performance of the Cantonese listeners across the no-mask and surgical mask conditions cannot be attributed to the lack of statistical power (e.g., sample size) in detecting the effect of surgical mask. Instead, we have yielded moderate to extreme evidence supporting the absence of the effect of surgical mask on tone perception.

Why does surgical mask not hinder tone perception? One likely explanation is that surgical mask does not acoustically influence Cantonese tones. As mentioned, the F0 range of tones is far below the high frequency range that face mask typically dampens (Giuliani, 2020; Oliveira et al., 2021). Indeed, our Cantonese tone stimuli produced with and without surgical mask did not differ significantly in F0 height, onset F0, offset F0, F0 slope, and even duration (see Supplementary material). As surgical mask does not acoustically alter Cantonese tones, and visual cues are irrelevant to tones (Hong et al., 2023), surgical mask does not hinder tone perception.

Hypothetically, surgical mask might have acoustically altered our Cantonese tone stimuli, but our number of stimuli was too small to register the acoustic difference. In particular, a production study found surgical mask to decrease the durations of the Cantonese high level and low level tones (Zhang et al., 2025). Furthermore, the study reported an increase in the F0 height in the Cantonese high level tone. Even if we assume that surgical mask acoustically alters Cantonese tones, these acoustic alterations should have little bearing on

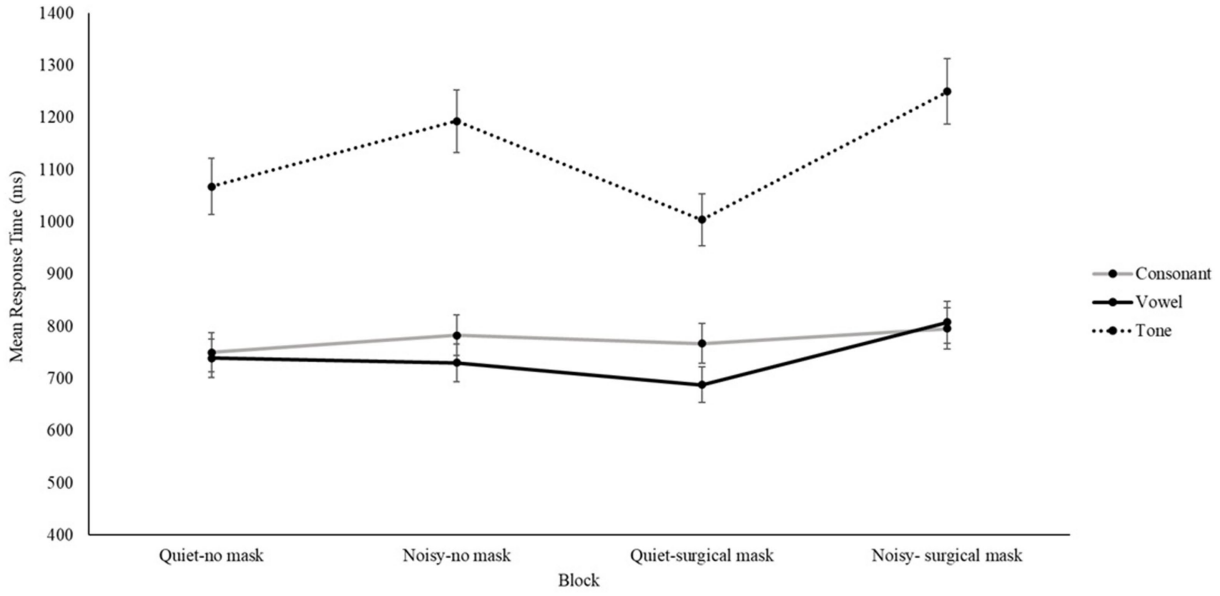

FIGURE 2
Mean response time in correct trials across the four blocks in the consonant, vowel, and tone identification tasks.

Cantonese tone identification, for two possible reasons. First, all six Cantonese tones are not contrastive in duration, so duration is not a reliable cue. Although the three entering tones (i.e., allophones of high-, mid-, and low level tones) have shorter durations compared to the other tones, they always end with /p/, /t/, or /k/, unlike the latter. According to our consonant identification results, surgical mask does not affect Cantonese consonant identification, indicating the final consonant cues of the Cantonese entering tones are still available. Thus, duration is a redundant cue even for the entering tones in the surgical mask condition. In other words, even if we hold the assumption that face mask decreases the durations of the Cantonese high level and low level tones (Zhang et al., 2025), it bears little impact on tone identification. Second, Cantonese high level tone has the highest F0 height compared with the other Cantonese tones. Further increasing the F0 height would not affect the phonemic status of the high level heard. In other words, listeners may at worst perceive the altered high level tone as higher in F0 than usual, but are still able to identify it as the high level tone.

Inconsistent with our hypothesis, surgical mask did not affect listeners' accuracy in identifying consonants and vowels. As we requested our talkers to avoid excessive articulation and other deliberate vocal modifications while wearing a surgical mask, surgical mask indeed shrank the vowel space of our vowel stimuli (see Figure 1; see also Joshi et al., 2023; Zhang et al., 2025). Importantly, the Cantonese listeners identified the Cantonese vowels with similar accuracy across no-mask and surgical mask conditions, suggesting that the acoustic alterations did not affect vowel identification. The same was true for consonant identification, with moderate to very strong evidence. Even in the noisy condition, the removal of visual information did not affect how accurately the Cantonese listeners identified consonants and vowels. For typical hearing listeners, visual information serves as a complementary cue for speech perception in challenging conditions. It appeared that, even after some auditory attenuation, the face-masked speech has preserved adequate auditory information characterizing the consonants and vowels. Even in noise, our Cantonese listeners could still accurately utilize the auditory information for consonant and vowel identification.

However, surgical mask increased the response time in identifying vowels in noise. Although surgical mask does not affect how accurately the Cantonese listeners identify consonants, vowels, and tones, it affects how *efficiently* they perceive vowels. When the vowels are masked with noise, visual information can help resolve ambiguities (Giovanelli et al., 2021; Yi et al., 2021). When visual information is occluded, listeners need to solely rely on the auditory process and devote extra cognitive resources to discriminate the vocalic signal from the noise. Regarding consonants, visual information can cue the place of articulation (e.g., bilabial versus labiodental). Nevertheless, surgical mask did not affect how efficiently the Cantonese listeners identified consonants in noise. A possible reason is that consonantal contrasts may be more acoustically salient than vowel contrasts, and therefore less susceptible to the influence of noise.

4.1 Implications and future directions

In the broader literature, the present findings can potentially explain why face mask affects sentence comprehension (Mendel et al., 2022; Moon et al., 2022; Zhou et al., 2022). In the noisy condition, listeners may still accurately identify consonants, vowels, and tones in face-masked speech. However, listeners need to devote more cognitive resources (e.g., working memory) to identify vowels with impoverished auditory and visual information. As a result, less cognitive resources become available for other cognitive processes involved in sentence comprehension, such as the retrieval of words, syntax, and prior knowledge (Gillam et al., 2019).

Our findings have some practical implications for the implementation of face mask in universities and schools during respiratory disease outbreaks or influenza seasons (World Health Organization, 2021). Despite the public health value of face mask, parents in Hong Kong and potentially elsewhere express concerns

about their children's ability to hear teachers clearly, which may impact their learning at school or university. To address these concerns, our study indicates that face mask does not bear a detrimental effect on how accurately listeners perceive consonants, vowels, and tones. However, since face mask affects how efficiently listeners perceive vowels, it is still possible for it to affect Cantonese sentence comprehension. Future studies are needed to investigate this possibility.

There are several avenues to extend our study. Given the paucity of related research, we examined consonant, vowel, and tone perception. To maximize the practical implication for the Cantonese population, future studies should examine the effect of face mask on Cantonese sentence comprehension. Moreover, we tested adults in our university rather than school-age children. To better inform parents and teachers in primary schools and pre-schools, testing children in a classroom-like environment is necessary. Lastly, we only focused on surgical mask since it is most commonly used in classrooms in Hong Kong. Including medical grade protections (e.g., KN95 and face shields) in future studies can have implications for medical workers.

5 Conclusion

Despite its documented negative impact on the transmission of auditory signals, surgical mask was found to have little bearing on the ability to accurately identify Cantonese consonants, vowels, and tones in our study. This holds true across quiet and noisy conditions. However, surgical mask affects how efficiently Cantonese listeners identify vowels in noise. In the theoretical aspect, this study offers a phonological account to explain why surgical mask may impede sentence comprehension (Mendel et al., 2022; Moon et al., 2022; Zhou et al., 2022). Concerning the practical impact of surgical mask, our findings suggest that Cantonese students may not show marked difficulties in identifying consonants, vowels, and tones, but they may indeed be less efficient in identifying vowels in noise. Ultimately, we believe that the decision to wear a face mask or not is a value judgment.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Human Research Ethics Committee of the University of Hong Kong. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

WC: Conceptualization, Formal analysis, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition. TC: Data curation, Investigation, Writing – review & editing. JZ: Data curation, Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the University of Hong Kong: Faculty Research Fund and Project-Based Research Funding (supported by Start-up Fund).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. Generative AI was utilized for proofreading and improving the grammar of the article.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2025.1582217/full#supplementary-material

References

Atcherson, S. R., McDowell, B. R., and Howard, M. P. (2021). Acoustic effects of non-transparent and transparent face coverings. *J. Acoust. Soc. Am.* 149, 2249–2254. doi: 10.1121/10.0003962

Atcherson, S. R., Mendel, L. L., Baltimore, W. J., Patro, C., Lee, S., Pousson, M., et al. (2017). The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss. *J. Am. Acad. Audiol.* 28, 58–067. doi: 10.3766/jaaa.15151

Badh, G., and Knowles, T. (2023). Acoustic and perceptual impact of face masks on speech: a scoping review. *PLoS One* 18:e0285009. doi: 10.1371/journal.pone.0285009

Badillo-Goicoechea, E., Chang, T.-H., Kim, E., LaRocca, S., Morris, K., Deng, X., et al. (2021). Global trends and predictors of face mask usage during the COVID-19 pandemic. *BMC Public Health* 21:Article 2099. doi: 10.1186/s12889-021-12175-9

Balan, J. R., and Maruthy, S. (2018). Dynamics of speech perception in the auditory-visual mode: an empirical evidence for the management of auditory neuropathy spectrum disorders. *J. Audiol. Otol.* 22, 197–203. doi: 10.7874/jao.2018.00059

Bauer, R. S., and Benedict, P. K. (1997). *Trends in linguistics: Modern cantonese phonology*. Berlin: Walter de Gruyter.

Centre for Health Protection. (2023). *Infection control advice in community vaccination centres for COVID-19 vaccination*. Hong Kong Department of Health, Centre for Health Protection. Available online at: https://www.chp.gov.hk/files/pdf/infection_control_advice_in_cvc.pdf (Accessed May 27, 2024).

Choi, W., and Chiu, M. M. (2023). Why aren't all Cantonese tones equally confusing to English listeners? *Lang. Speech* 66, 870–895. doi: 10.1177/00238309221139789

Choi, W., Tong, X., Gu, F., Tong, X., and Wong, L. (2017a). On the early neural perceptual integrality of tones and vowels. *J. Neurolinguist.* 41, 11–23. doi: 10.1016/j.jneuroling.2016.09.003

Choi, W., Tong, X., and Singh, L. (2017b). From lexical tone to lexical stress: a cross-language mediation model for Cantonese children learning English as a second language. *Front. Psychol.* 8:492. doi: 10.3389/fpsyg.2017.00492

Cutler, A. (2015). "Lexical stress in English pronunciation" in Handbook of English pronunciation. eds. M. Reed and J. Levis (Hoboken, NJ: Wiley-Blackwell), 106–124.

Fischer, C. B., Adrien, N., Silguero, J. J., Hopper, J. J., Chowdhury, A. I., and Werler, M. M. (2021). Mask adherence and rate of COVID-19 across the United States. *PLoS One* 16:e0249891. doi: 10.1371/journal.pone.0249891

Gillam, R. B., Montgomery, J. W., Evans, J. L., and Gillam, S. L. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: implications for assessment and treatment. *Int. J. Speech Lang. Pathol.* 21, 240–251. doi: 10.1080/17549507.2018.1559883

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

Giuliani, N. (2020) For speech sounds, 6 feet with a mask is like 12 feet without: social distancing and masks protect against COVID-19–but add another layer of difficulty for audiologists treating patients with hearing loss. ASHA LeaderLive. Available online at: https://leader.pubs.asha.org/do/10.1044/leader.AEA.25112020.26/full/

Hampton, T., Crunkhorn, R., Lowe, N., Bhat, J., Hogg, E., Afifi, W., et al. (2020). The negative impact of wearing personal protective equipment on communication during coronavirus disease 2019. *J. Laryngol. Otol.* 134, 577–581. doi: 10.1017/S0022215120001437

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., and Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *J. Acoust. Soc. Am.* 119, 1740–1751. doi: 10.1121/1.2166611

Hong, S., Wang, R., and Zeng, B. (2023). Incongruent visual cues affect the perception of mandarin vowel but not tone. *Front. Psychol.* 13:971979. doi: 10.3389/fpsyg.2022.971979

Jaynes, E. T. (2003). Probability theory: The logic of science. Cambridge: Cambridge University Press.

Joshi, A., Procter, T., and Kulesz, P. A. (2023). COVID-19: acoustic measures of voice in individuals wearing different facemasks. *J. Voice* 37, 971.e1–971.e8. doi: 10.1016/j.jvoice.2021.06.015

Ladefoged, P., and Johnson, K. (2011). A course in phonetics. *6th* Edn. Belmont, CA: Wadsworth.

Lee, M. D., and Wagenmakers, E.-J. (2013). *Bayesian* cognitive modelling: A practical course. Cambridge: Cambridge University Press.

Mendel, L. L., Pousson, M. A., Shukla, B., Sander, K., and Larson, B. (2022). Listening effort and speech perception performance using different facemasks. *J. Speech Lang. Hear. Res.* 65, 4354–4368. doi: 10.1044/2022_JSLHR-22-00081

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

Nguyen, D. D., McCabe, P., Thomas, D., Purcell, A., Doble, M., Novakovic, D., et al. (2021). Acoustic voice characteristics with and without wearing a facemask. *Sci. Rep.* 11:5651. doi: 10.1038/s41598-021-85130-8

Oliveira, R. C., Gama, A. C. C., and Magalhães, M. D. C. (2021). Fundamental voice frequency: acoustic, electroglottographic, and accelerometer measurement in individuals with and without vocal alteration. *J. Voice* 35, 174–180. doi: 10.1016/j.jvoice.2019.08.004

Schwarz, J., Li, K. K., Sim, J. H., Zhang, Y., Buchanan-Worster, E., Post, B., et al. (2022). Semantic cues modulate children's and adults' processing of audio-visual face mask speech. *Front. Psychol.* 13:879153. doi: 10.3389/fpsyg.2022.879156

Sinagra, C., and Wiener, S. (2022). The perception of intonational and emotional speech prosody produced with and without a face mask: an exploratory individual differences study. *Cogn. Res. Princ. Implic.* 7:89. doi: 10.1186/s41235-022-00439-w

Toscano, J. C., and Toscano, C. M. (2021). Effects of face masks on speech recognition in multi-talker babble noise. *PLoS One* 16:e0246842. doi: 10.1371/journal.pone.0246842

World Health Organization. (2021) Coronavirus disease (COVID-19) advice for the public: when and how to use masks. Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks

Yi, H., Pingsterhaus, A., and 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. doi: 10.3389/fpsyg.2021.682677

Yuan, Y., Lleo, Y., Daniel, R., White, A., and Oh, Y. (2021). The impact of temporally coherent visual cues on speech perception in complex auditory environments. *Front. Neurosci.* 15:678029. doi: 10.3389/fnins.2021.678029

Zhang, T., He, M., Li, B., Zhang, C., and Hu, J. (2025). Acoustic characteristics of Cantonese speech through protective facial coverings. *J. Voice* 39, 560.e11–560.e19. doi: 10.1016/j.jvoice.2022.08.029

Zhou, P., Zong, S., Xi, X., and Xiao, H. (2022). Effect of wearing personal protective equipment on acoustic characteristics and speech perception during COVID-19. *Appl. Acoust.* 197:108940. doi: 10.1016/j.apacoust.2022.108940