Appendix A

Biases towards objects analyses

To test if infants had any biases towards any of the novel objects and attended more to one object than the other (due to shape or other properties). A paired-samples t-test was performed to compare looking time for the first instance of each object presentation (i.e., first fixation duration for each object) in the training phase. Results indicated that first fixation durations were not significantly different between the object 1 (the star like object, *M* = 1.87, *SD* = .54) and object 2 (the taller, cone shaped object, *M* = 1.76, *SD* = .73), *t*(71) = .99, *p =* .326, 95 % CI [-.11, .33]. These results reflect that the two novel objects were equally attractive to infants.

Appendix B

Acoustic analyses of videos

In order to eliminate the prosodic interference of audios, we extracted and analyzed duration (in second), the mean fundamental frequency (mean f0, in Hz), and mean intensity (in dB) of the target word in both the training and testing phases using Praat script (Boersma, 2007). For fundamental frequency, 10 points were extracted across the duration of the target word in equal intervals. The mean fundamental frequency was then calculated by averaging the total value across the 10 points. Comparisons were then made between the No Mask and With Mask Conditions.

First, paired samples t-tests revealed that in the training phase, there was no significant difference in 1) duration of target word, *t(14) =* 1.42, *p* = 0.18 for, 2) fundamental frequency of target word, *t(14)* = -1.90, *p =* 0.07 and 3) intensity of target word *t(14) =* 1.29, *p =* 0.22 and 3) between the two conditions (see Table 1) .

Table 1. Acoustic information in the training phase

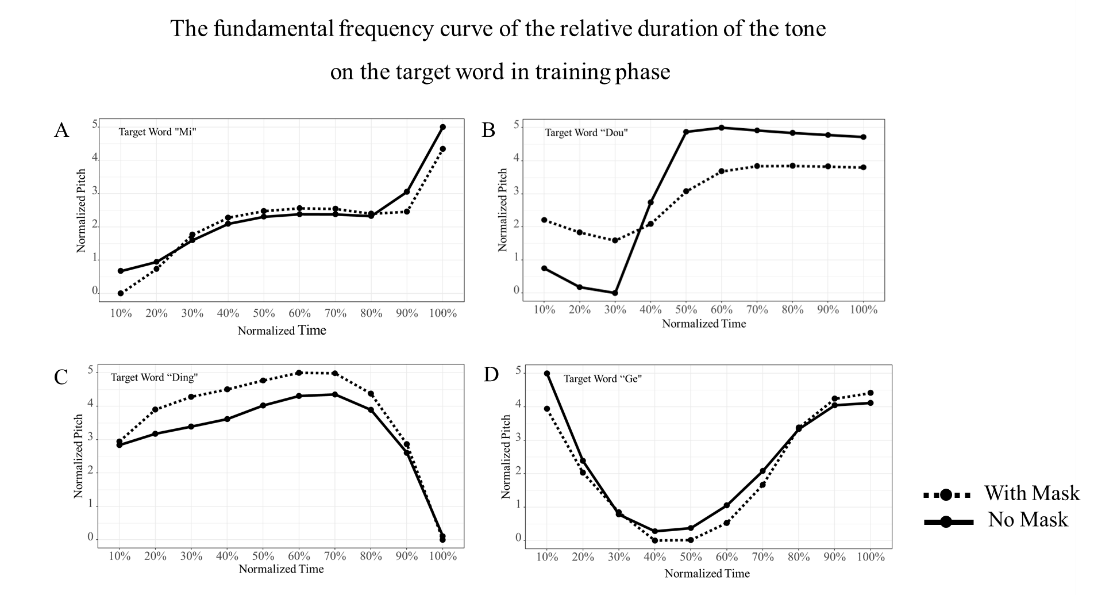
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Training Phase** | **No mask** | | **With mask** | |
| **M** | **SD** | **M** | **SD** |
| **Duration (s)** | .77 | .15 | .80 | .20 |
| **Fundamental frequency (Hz)** | 237.90 | 96.10 | 198.60 | 62.70 |
| **Intensity (dB)** | 55.90 | 2.14 | 56.60 | 1.42 |

In the testing phases, there was also no significant difference for 1) duration, *t(15) =* -0.91, *p =* 0.38 , 2) fundamental frequency, *t(15) = -1.42*, *p =* 0.18 and 3) intensity, *t(15) = -0.03*, *p =* 0.97 between conditions (see Table 2).

Table 2. Acoustic information in the test phase

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Testing Phases** | **No mask** | | **With mask** | |
| **M** | **SD** | **M** | **SD** |
| **Duration (s)** | .79 | .30 | .72 | .13 |
| **Fundamental frequency (Hz)** | 245.50 | 108.40 | 219.10 | 82.90 |
| **Intensity (dB)** | 56.70 | 1.69 | 56.70 | 2.72 |

Note that the differences of fundamental frequency (pitch) between the two conditions is close to significance in the training trials. Given that Mandarin is a tone language, more detailed analyses were conducted to explore whether any difference exists in pitch contour of the lexical token. Fundamental frequency curves of the relative duration of the tone on the target words were sketched for each syllable of the target words. The figures revealed no observable differences (see Figure 1 and 2).



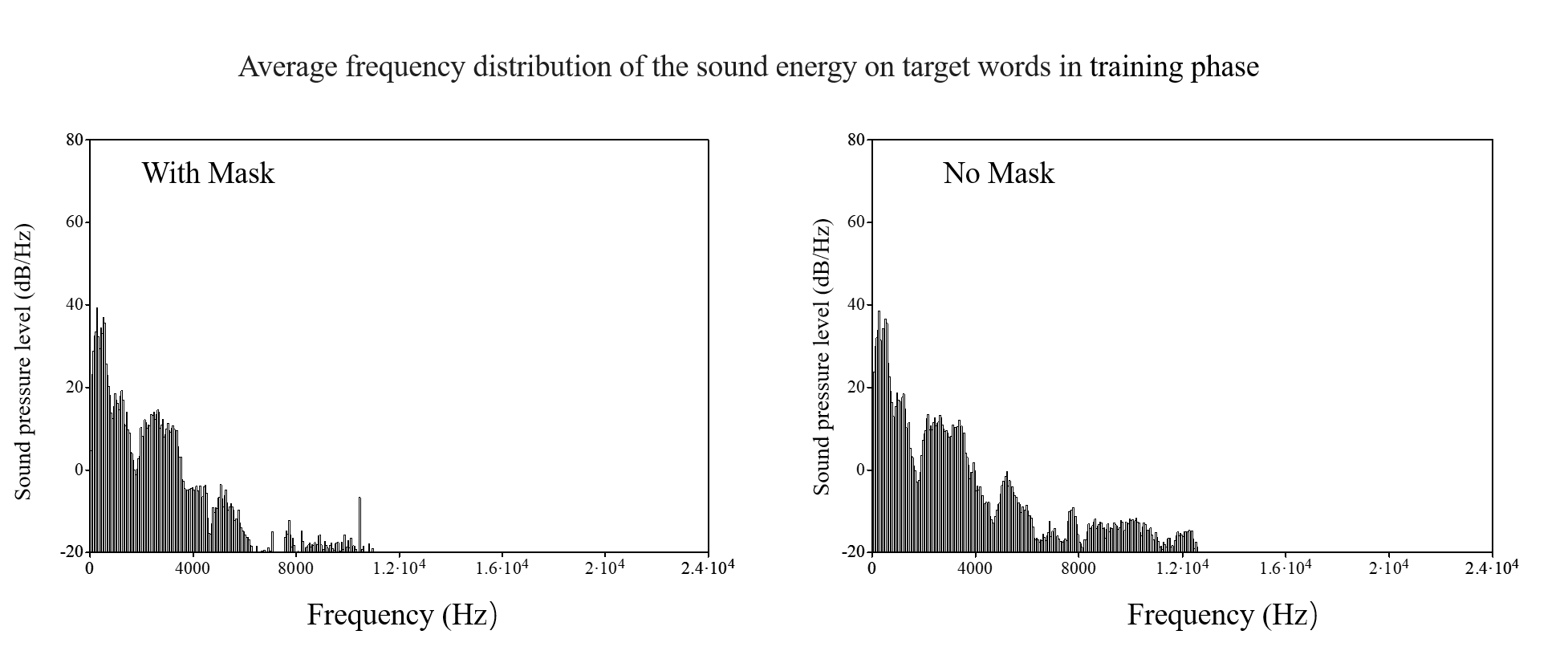
*Figure 1*. The fundamental frequency curve of the relative duration of the tone on target word in training phase that is normalized in a five-point scale. A) target word ”Mi”; B) target word ”Dou”; C) target word ”Ding”; D) target word ”Ge”.

A graph of different types of lines

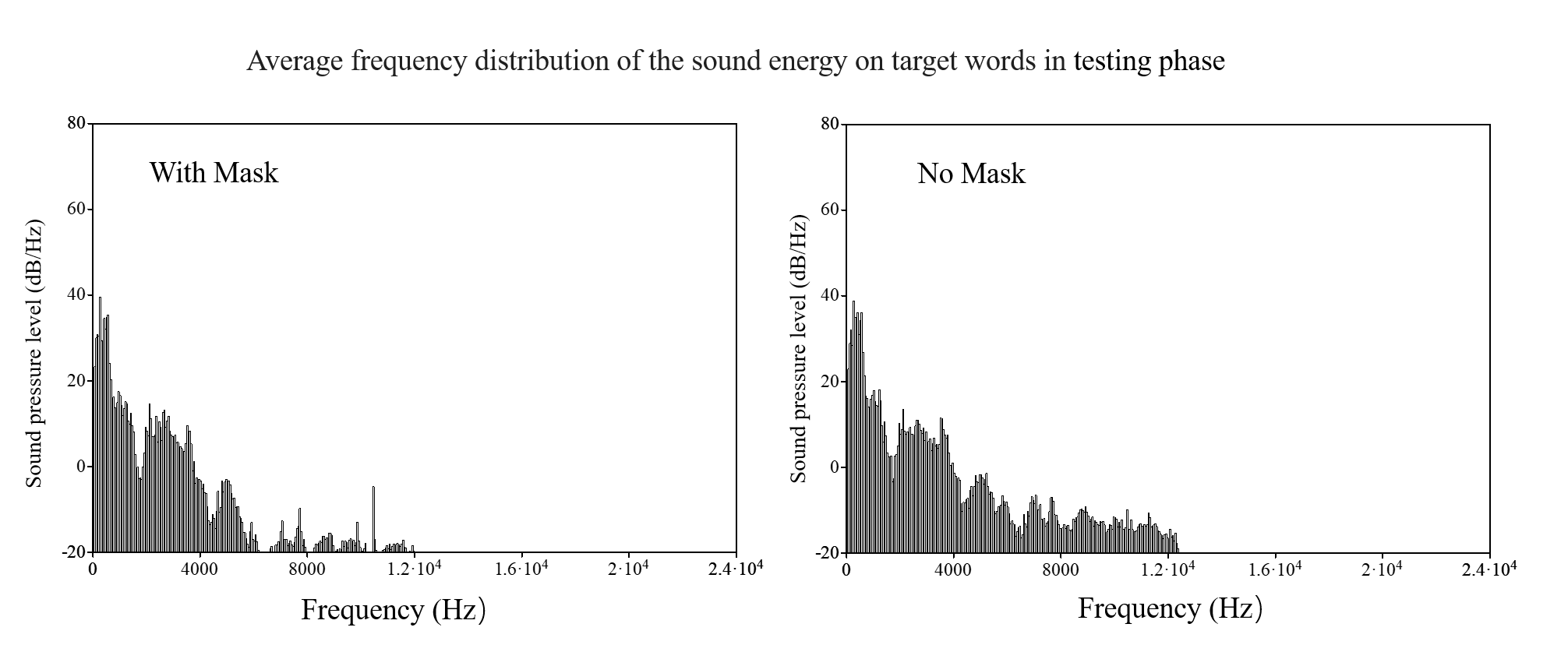
Description automatically generated with medium confidence

*Figure 2.* The fundamental frequency curve of the relative duration of the tone on target word in testing phase that is normalized in a five-point scale. A) target word ”Mi”; B) target word ”Dou”; C) target word ”Ding”; D) target word ”Ge”.

Next, previous research indicates that face masks can impact high-frequency acoustics (Knowles, 2022). To analyze the average frequency distribution of sound energy of the target word, Long-Term Average Spectrum (LTAS) analyses (Leino, 2009) were performed on the target words in the No Mask and With Mask Condition. Inspections of the LTAS results (see Figure 3 and 4) revealed no observable differences in frequency distribution of sound energy of the target word between conditions.



*Figure 3.* The average frequency distribution of the sound energy on target words under different conditions in training phase. A) With Mask Condition; B) No Mask Condition.



*Figure 4.* The average frequency distribution of the sound energy on target words under different conditions in testing phase. A) With Mask Condition; B) No Mask Condition.

**References**

Boersma, P. (2007). Praat: doing phonetics by computer. *http://www. praat. org.*

Knowles, T., & Badh, G. (2022). The impact of face masks on spectral acoustics of speech: Effect of clear and loud speech styles. *The Journal of the acoustical society of America*, *151*(5), 3359-3368.

Leino, T. (2009). Long-term average spectrum in screening of voice quality in speech: untrained male university students. *Journal of voice*, *23*(6), 671-676.