

1 **Supplemental Material**

2 **Gestural enhancement effects on children at different noise levels**

3 Before conducting the study reported in the main text, we conducted a pilot study by
4 comparing two groups of 6- and 7-year-olds in order to decide which noise level to
5 use in the current study and to investigate at which noise level children could benefit
6 from gestures the most. The two groups were called Group A and Group B,
7 respectively. Group A consisted of children who eventually participated in the main
8 study. Group B included children who were newly recruited for comparison to the
9 children in Group A. Groups A and B each had 15 children with eight females in each
10 group. The mean age in Group B was 7;02 (Ten 7-year-olds and Five 6-year-olds).
11 Both groups differed in terms of the noise-vocoding levels they were presented with.
12 Group A received the 4- noise-vocoding condition, 8-band noise-vocoding condition
13 and a clear speech condition. Group B received the 8-band noise-vocoding
14 condition, 10-band noise-vocoding condition and a clear speech condition. The
15 creation of the noise-vocoded speech and video stimuli is described in the main text.
16 Cut-off frequencies for the 10-band noise-vocoding were 50 Hz, 83.1 Hz, 138 Hz,
17 229.2 Hz, 380.7 Hz, 632.5 Hz, 1050.6 Hz, 1745.2 Hz, 2899.1 Hz, 4815.9 Hz, and
18 8000 Hz.

19

20 **Correct responses**

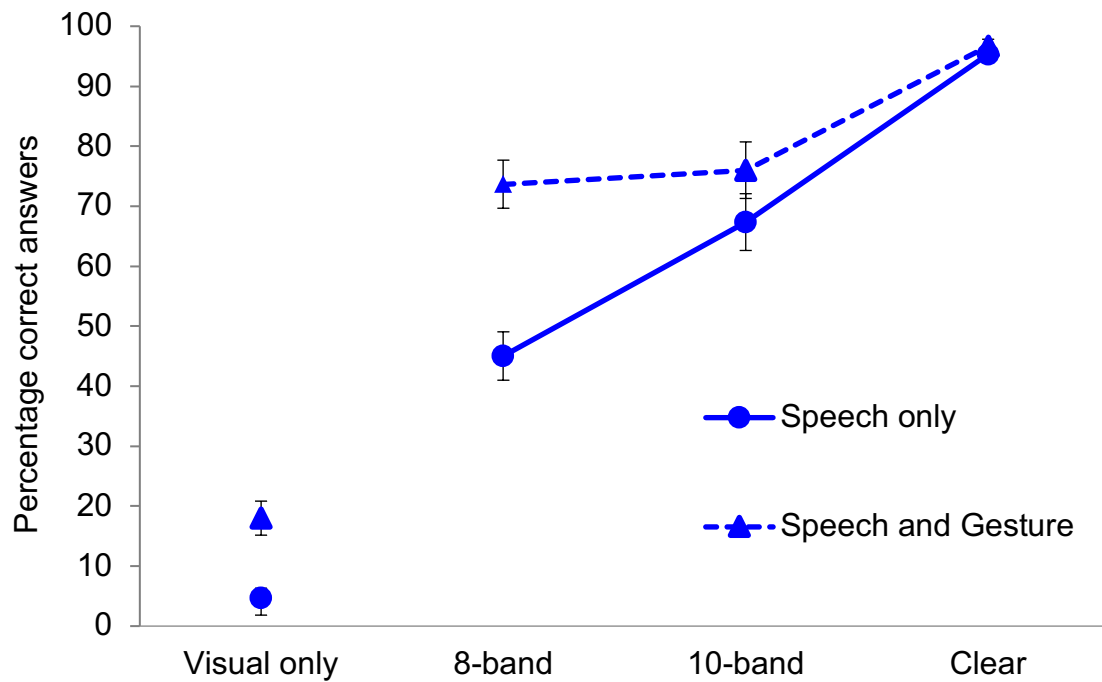
21 First, we conducted paired t-tests to compare the mean percentages of correct
22 responses between the speech-only condition and the speech and gesture
23 conditions at each noise-vocoding level for Group B, in the same way we reported
24 our statistical analyses for Group A in the main text. T-values and effect sizes are
25 shown in Table 1 and Figure 1. Group A was also included in Table 1 to note the

26 differences between the groups. Results revealed that for all groups, the mean
 27 percentages of correct responses in the speech and gesture condition were
 28 significantly higher than those in the speech-only condition at all noise-vocoding
 29 levels (degraded speech or visual-only) except for when the speech was clear and
 30 without noise.

31 Table 1. Mean percentage of correct responses with the standard deviation in
 32 parentheses, provided across the eight conditions for each child group. Results of
 33 the paired t-tests (t-values and effect sizes) are also reported below.

Speech quality	Modality			
Group A	Speech only (SO)	Speech and Gesture (SG)	T-value	effect size (r)
Visual-only	5.67 (8.00)	22.33 (11.78)	7.34***	0.89
4ch	20.33 (13.95)	38.33 (20.06)	-4.23***	0.75
8ch	48.33 (13.45)	73.33 (12.49)	-9.83***	0.94
clear	93.67 (5.50)	96 (5.73)	-0.91	0.24
Group B	Speech only (SO)	Speech and Gesture (SG)	T-value	effect size (r)
Visual-only	4.67 (6.40)	18 (10.99)	4.23***	0.75
8-band	45 (15.70)	73.67 (15.52)	-2.76*	0.59
10-band	67.33 (18.41)	76 (18.24)	-7.67***	0.9
Clear	95.33 (5.82)	96.67 (4.50)	-0.65	0.17

34



35

36 Figure 1. The percentage of correct trials per condition for Group B. The dashed
 37 lines indicate results from the speech and gesture (SG) condition, and the solid lines
 38 indicate results from the speech-only (SO) condition.

39

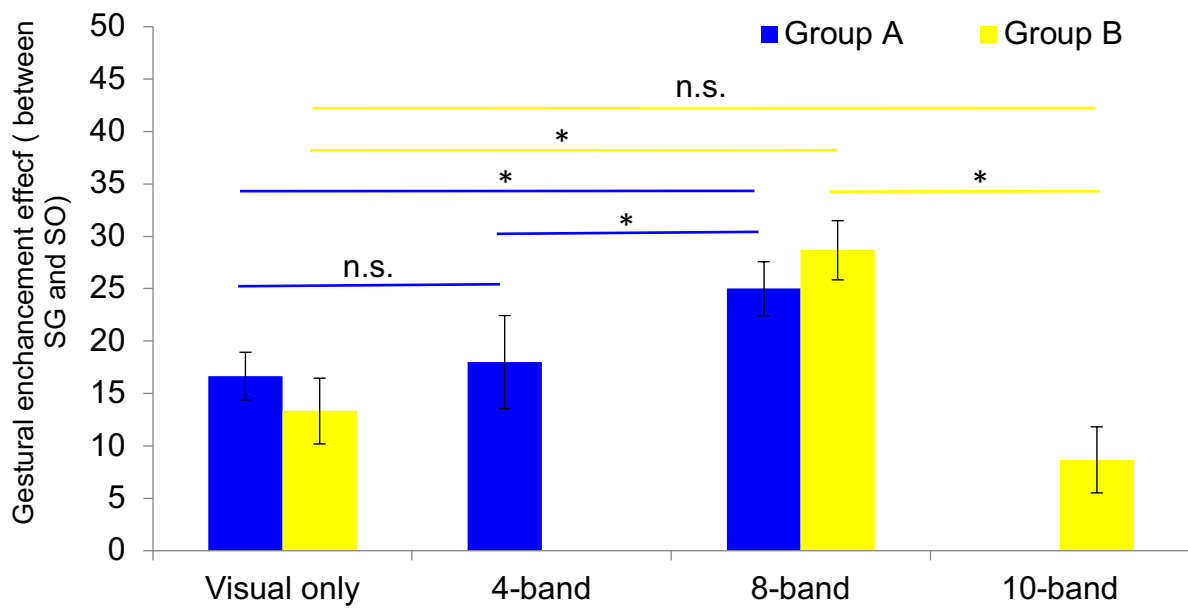
40 **Gestural enhancement effect**

41 In order to investigate the extent to which degraded speech affects children’s speech
 42 perception, we first examined the differences in gestural enhancement effects at
 43 each noise-vocoded level for each child group (Figure 2). We did not directly
 44 compare the two groups, as each group was exposed to different noise-vocoded
 45 levels. After arcsine transformation for proportional data, we conducted a repeated
 46 measure analysis of variance for each child group. The analysis revealed a main
 47 effect of *noise-vocoding* for Group A (visual-only vs. 4-band vs. 8-band), $F(2, 28) =$
 48 $3.81, p < .05, \text{partial } \eta^2 = .17,$ and for Group B (visual-only vs. 8-band vs. 10-band), F
 49 $(2, 28) = 10.23, p < .001, \text{partial } \eta^2 = .42.$ Post hoc tests using the Bonferroni
 50 correction ($p < .05$) showed that in Group A, the gestural enhancement effect was

51 significantly higher in the 8-band condition than in the 4-band and visual-only
52 conditions. The gestural enhancement effect was higher in the 8-band condition than
53 the 10-band and visual-only conditions for Group B. Additionally, statistical analysis
54 revealed that accuracy with the visual-only condition was higher than accuracy for
55 the 10-band noise-vocoding condition. These results indicate that children do not
56 gain much benefit from gestures at 4-band (Group A) and 10-band (Group B) noise-
57 vocoded speech compared to the benefits they gain from gestures with the 8-band
58 noise-vocoded speech. Thus, the 8-band noise-vocoding level appears to be the
59 level in which children benefit most from gestures.

60 Next, we compared the two child groups to see whether there were differences in
61 the gestural enhancement effect between the 4- (Group A) and 10-band (Group B)
62 noise-vocoding conditions and between the 8-band noise-vocoding conditions across
63 the two child groups. After arcsine transformation, we conducted an independent-
64 samples t-test. Results showed no significant differences ($t(28) = 1.49, p = .17$)
65 between the 4- and 10-band noise-vocoding conditions, as well as no significant
66 differences ($t(28) = 0.91, p = .37$) between the 8-band noise-vocoding conditions.

67 These results indicate that the 8-band noise-vocoded level appears to be the level in
68 which children benefit from gestures the most. The gestural enhancement effect was
69 almost the same for the 4- and 10-band noise-vocoded levels. It can be said that the
70 8-band noise-vocoded level is the moderate noise level for 6- and 7-year-olds and
71 where children benefit from gestures the most, just as the 6-band noise-vocoded
72 speech is the level in which adults benefit from gestures the most, as shown by
73 Drijven and Özyürek (2017). Thus, we decided to use Group A in the main study.



74
75 * Bonferroni correction ($p < .05$)

76 Figure 2. Gestural enhancement effect at each noise-vocoding level for Groups A
77 and B in the bottom panel. Error bars indicate standard errors.

78