Supplemental Digital Content 1:

Sample

The sample size for the research project was estimated a priori with a power analysis. Previous studies indicated that a difference in emotion understanding between DHH and TH children could be observed with small-to-medium effect sizes (Torres et al., 2016; Wiefferink et al., 2013). Thus, a minimum sample size of 82 was required to detect a group difference (d = .4; α = .05; power = .90). Note that we planned to use mixed model ANOVAs when estimating sample size a priori and later changed to multilevel models considering the two-level structure in the data.

Supplemental Digital Content 2:

Stimuli

Stimuli

Table S2.1.

Overview of videos. There were two sets of videos, and children were randomly assigned to one of the sets. The first sentence describes the contextual scene. The second sentence (bolded) describes the key-action scene (i.e., the scene included in the analyses).

Trial	Set A	Set B
1§	A woman is crying and a man approaches. **The man gives her a flower.***	A man is enjoying himself (slightly shaking body with a tempo) and a woman approaches. **The woman gives him a well wrapped gift.**
2§	A woman is hurt and a man approaches. **The man does not help her.**	A man is enjoying himself (smoking) and a woman approaches. **The woman shows that he is forbidden to be here.***
3	A woman is happily checking smartphone and another woman approaches. **The second woman gives her a high five.**	A woman is happily checking smartphone and another woman approaches. **The second woman pushes her away with elbow while walking by.***
4	A man is sitting in a cafeteria, looking hungry, and a woman approaches with a pizza. **The woman refuses to share the pizza with him.**	A man is sitting in a cafeteria, looking hungry, and a woman approaches with a pizza. **The woman shares the pizza with him.***
5	A woman is waving with a smile and crossing a road. **A man on the other side of the road pushes her down to the ground.***	A woman is waving with a smile and crossing a road. **A man on the other side of the road gives her a well wrapped gift.**

6	A man is hurt, walking with sticks, and a woman approaches. **The woman shows him a cake.***	A man is hurt, walking with sticks, and a woman approaches. **The woman laughs.**
7	A woman is happily climbing across monkey bars. **She successfully makes it to the end and a man gives her a cold drink.**	A woman is happily climbing across monkey bars. **She falls from it and a man points at her.***
8	A man is riding a bike, almost falling down, and a woman approaches. **The woman holds the bike.***	A man is riding a bike, almost falling down, and a woman approaches. **The woman throws a rock at him.**

§ While in trials 3 to 8 the videos in set A and set B were parallel, the first two trials had a different structure. We designed two videos with conceptually similar (but not the same) contextual scenes (e.g., crying vs. feeling pain) and ending with an emotion in the opposite valence. These two videos were placed in the same set.

*The video had a twist in the plot, i.e., from positive to negative emotion, or vice versa. We assumed that videos with a twist would be more difficult for children as they needed to answer based on their prediction of the emotion triggered in the key-action scene, rather than on the initial emotion expressed during the contextual scene. Thus, we ensured that in each set of the videos, half of the videos were with a twist, and half without.

Video validation

Before the study started, the emotion triggered in the videos were rated by 17 typically developing adults. The inter-rater reliability was moderate (Light's kappa = .58; Landis & Koch, 1977; Light, 1971). Further inspection showed that the percentage of agreement was above 82% for 15 of the 16 videos (M = 90.86%, SD = 6.27). Yet, one video had a low consent rate (58.82%) because some raters mistook the surprised face among the response options as fearful. This face image was replaced by a retaken photo and was agreed by the research team.

Supplemental Digital Content 3:

Fixation Duration

Table S3.1.

Fixed and random effects in the generalized linear mixed model of fixation ratios within video frame (binomial distribution, link function = logit).

Fixed and random effect	Coefficient	95% *CI*	*z*-value (*p*-value)
Intercept	**5.30**	[4.74, 5.86]	18.51 (< .001)
Age	**.04**	[.01, .06]	3.38 (.001)
Group	*ns*		
Valence	**-.46**	[-.76, -.16]	-2.98 (.003)
Valence x Group	*ns*		
Variance - Intercept	1.90	[1.53, 2.37]	

Note: Group was coded as -1 = DHH, 1 = TH. Valence was coded as -1 = negative, 1 = positive. The last category was used as the reference. An "*ns*" indicates that the variable was removed from the final model due to insignificance. CI = confidence interval.

Table S3.2.

Mean (SD) of fixation duration (ms) and ratio within each area as a function of hearing status, device use, and chronological age.

		Hearing status		Device use		Chronological age	
Area		DHH (*n* = 57)	TH (*n* = 68)	CI (*n* = 52)	HA (*n* = 5)	< 6 years (*n* = 77)	≥ 6 years (*n* = 48)
Target Head	*ms*	647.75 (384.83)	845.92 (501.63)	665.43 (382.31)	463.95 (404.28)	634.33 (433.44)	950.02 (441.12)
	ratio[a]	0.15 (0.07)[#]	0.18 (0.08)[#]	0.16 (0.07)	0.11 (0.08)	0.14 (0.07)***	0.20 (0.07)***
Target Body	*ms*	452.53 (323.29)	494.1 (359.59)	473.34 (323.63)	236.04 (252.29)	503.63 (392.94)	429.43 (238.46)
	ratio[a]	0.13 (0.07)	0.11 (0.08)	0.13 (0.07)	0.10 (0.07)	0.12 (0.08)	0.10 (0.06)
Partner Head	*ms*	963.97 (598.61)	1077.50 (654.02)	974.10 (598.70)	858.65 (656.87)	872.12 (602.3)	1272.14 (598.29)
	ratio[a]	0.23 (0.11)	0.23 (0.11)	0.23 (0.11)	0.22 (0.13)	0.20 (0.11)***	0.27 (0.10)***
Partner Action	*ms*	466.10 (292.23)	531.74 (332.46)	482.37 (293.20)	296.91 (245.53)	470.87 (334.45)	551.44 (277.85)
	ratio[a]	0.11 (0.05)	0.11 (0.05)	0.11 (0.05)	0.07 (0.03)	0.10 (0.06)	0.12 (0.05)
Within video	*ms*	3845.59 (1495.81)	4395.06 (1587.74)	3925.78 (1440.33)	3011.55 (1979.97)	3956.52 (1709.09)	4446.05 (1258.83)
	ratio[a]	0.97 (0.13)	0.96 (0.08)	0.97 (0.14)	0.95 (0.07)	0.95 (0.13)*	0.99 (0.03)*
Within screen	*ms*	3889.36 (1498.80)	4527.17 (1532.56)	3969.32 (1444.30)	3057.86 (1976.24)	4089.27 (1696.54)	4472.24 (1243.66)
	ratio[b]	0.57 (0.22)*	0.66 (0.22)*	0.58 (0.21)	0.44 (0.28)	0.60 (0.24)	0.66 (0.18)
Off-screen	*ms*	2987.01 (1591.86)	2311.53 (1530.15)	2910.40 (1549.23)	3783.71 (2001.34)	2779.31 (1756.23)	2363.26 (1249.65)
	ratio[b]	0.43 (0.22)*	0.34 (0.22)*	0.42 (0.21)	0.56 (0.28)	0.40 (0.24)	0.34 (0.18)

Note: DHH = deaf and hard of hearing. TH = typically hearing. CI = cochlear implant. HA = hearing aid.

[a] Ratio against fixation duration within the entire screen. [b] Ratio against the duration of the video (key-action scene).

# $p < .08$; * $p < .05$; *** $p < .001$ for the differences in fixation ratios between the comparison groups according to t-tests (two-tailed).

Supplemental Digital Content 4:

Analyses on Children with Cochlear Implants

When analyses were conducted excluding the children with only a hearing aid (HA), i.e., including only children with a cochlear implant (CI), the directions of results generally remained the same. Below we discuss the differences observed between the analyses on all DHH children and the analyses on children with a CI. See Table S4.1 for the complete final models.

Encoding

Regarding the fixation ratios within the video frame, all results were in line with the previous analyses where the entire DHH group was included.

In the analysis on fixation ratios within the AOIs, all the effects remained the same as previous analyses, except for the interaction of Group x Partner Head. When only children with a CI were included, this effect became marginal, b = .03, 95% CI [-.00, .06], δ = .17. The interaction of Group x Target Body remained significant, b = .04, 95% CI [.01, .07], δ = .23.

Interpretation

All the results were congruent with previous analyses, except that an effect for Valence was observed, b = .06, p = .038, 95% CI [.003, .11], δ = .09. In children with CIs and

with TH, negative emotions were interpreted more accurately than positive emotions. No interaction effects were observed, in line with previous analyses.

Effect of Encoding on Interpretation

For nonverbal interpretation, we observed two additional interactions: Group x Target Body, b = -.39, 95% CI [-.68, -.10], δ = .60, and Group x Partner Head, b = -.30, 95% CI [-.56, -.03], δ = .46. These results suggest that, while looking longer at Target Body and Partner Head decreased the nonverbal interpretation scores in the two groups alike, these effects were even stronger in children with a CI.

For verbal interpretation, we observed an additional main effect of Partner Head, b = -.17, 95% CI [-.29, -.05], δ = .28. This indicates that longer looking times at Partner Head were associated with lower verbal scores in the two groups. We also observed two additional interactions, Group x Target Head, b = .30, 95% CI [.02, .59], δ = .49, and Group x Partner Action, b = .48, 95% CI [.18, .78], δ = 78. Although looking longer at Target Head increased verbal interpretation scores in the two groups alike, this effect was stronger in the children with a CI. Also, the association between longer looking times at Partner Action and lower verbal scores was observed only in the TH children, but not in the children with a CI.

Discussion

Despite these differences between the analyses on all DHH children and the analyses on only children with a CI, the overall picture derived from the results was similar. DHH

children decreased their attention to the target person's head and increased their attention to the target person's body. This finding further supports our claim that DHH children tend to divert their attention away from ambiguous cues to explicit, visually observable information, especially the body cues.

Also, the cues we examined in this study appear to work differently on interpretation in the two groups. DHH children were more easily misled by explicit cues, such as target person's body and interaction partner's head. This is most likely because they did not have adequate social-emotional knowledge to support their use of these explicit cues, as we discussed in the main text. The extra interactions of Group x Target Body and Group x Partner Head we observed in the analyses on only the children with a CI suggest that children with a CI might need even more support for gaining social-emotional knowledge in order to make proper interpretation when encountering social situations.

Considering that we only had five children with a HA, it is hard to draw a conclusion whether different types of amplification or degrees of hearing loss might have an effect. Future research is suggested to look further in this direction.

Table S4.1

Fixed and random effects in the generalized linear mixed models excluding children with only a hearing aid (n = 5).

	Fixation ratio				Interpretation		Effect of encoding on interpretation			
	Within AOIs		Video frame[a]				Nonverbal		Verbal	
Parameters	*b*	95% CI	*b*	95% CI	*b*	95% CI	*b*	95% CI	*b*	95% CI
Intercept	.18	[.17, .20]	**5.30**	[4.74, 5.86]	1.10	[1.03, 1.18]	1.19	[1.12, 1.26]	1.12	[1.05, 1.18]
Age	**.00**	[.00, .00]	**.04**	[.01, .06]	**.01**	[.01, .01]	**.01**	[.01, .01]	**.01**	[.01, .01]
Group	**-.02**	[-.05, -.00]	*ns*		**-.16**	[-.25, -.06]	-.10	[-.21, .01]	**-.14**	[-.24, -.04]
Valence	*ns*		**-.46**	[-.76, -.16]	**.06**	[.00, .11]	**.10**	[.06, .14]	*ns*	
Task	--		**--**		**.07**	[.02, .13]	**--**		--	
Group x Valence	*ns*		*ns*		*ns*		*ns*		*ns*	
Group x Task	--		--		*ns*		--		--	
Group x Valence x Task	--		--		*ns*		--		--	
TarHead	*ref*		--		--		.07	[-.10, .24]	**.18**	[.02, .35]
TarBody	**-.07**	[-.09, -.05]	--		--		**-.22**	[-.43, -.01]	-.11	[-.31, .09]
ParHead	**.05**	[.03, .07]	--		--		**-.31**	[-.48, -.14]	**-.17**	[-.29, -.05]
ParAction	**-.07**	[-.09, -.05]	--		--		*ns*		**-.59**	[-.80, -.39]
Group x TarHead	*ref*		--		--		**.68**	[.40, .96]	**.30**	[.02, .59]
Group x TarBody	**.04**	[.01, .07]	--		--		**-.39**	[-.68, -.10]	**-.42**	[-.69, -.15]
Group x ParHead	.03	[-.00, .06]	--		--		**-.30**	[-.56 -.03]	*ns*	
Group x ParAction	.03	[-.00, .06]	--		--		*ns*		**.48**	[.18, .78]
Var(Intercept)	.00	[.00, .00]	1.90	[1.53, 2.37]	.05	[.03, .07]	.08	[.06, .10]	.07	[.05, .09]
Residual	.03	[.02, .03]	--	--	.38	[.35, .40]	.34	[.32, .36]	.31	[.29, .32]

Note: Group was coded as -1 = DHH, 1 = TH. Valence was coded as -1 = negative, 1 = positive. Task was coded as -1 = nonverbal, 1 = verbal. AOI was coded as -2 = interaction partner's head (ParHead), -1 = interaction partner's action (ParAction), 1 = target person's body (TarBody), 2 = target person's head (TarHead). The last category was used as the reference ("*ref*"). An "*ns*" indicates that the variable was removed from the

final model due to insignificance. A "--" indicates that the effect was not included in the full model. Significant fixed effects ($p < 0.05$) are bolded. CI = confidence interval.

[a]Binomial distribution (link function = logit) was selected.

Supplemental Digital Content 5:

Correlations with Age at Amplification and Listening Experience

Table S5.1

Correlations of study variables with age at amplification and listening experience within the DHH group.

	Age at amplification	Listening experience
Fixation ratio		
Target Head	-.28	.35**
Target Body	.17	-.19
Partner Head	-.21	.20
Partner Action	-.02	.06
Video frame	-.04	.18
Within screen	-.09	.03
Off-screen	.09	-.03
Interpretation accuracy		
Nonverbal	-.22	.06
Verbal	-.15	.26

**$p < .01$ (Bonferroni-corrected significance level was set at $p < \alpha/2 = .025$)