

Psychophysical study		Loudness		Centroid	
Algorithm	Labeled <i>real</i>	Err.	<i>r</i>	Err.	<i>r</i>
Full system	40.01% \pm 1.66	0.21	0.44	3.85	0.47
- Trained from scratch	36.46% \pm 1.68	0.24	0.36	4.73	0.33
- No spacetime	37.88% \pm 1.67	0.22	0.37	4.30	0.37
- Parametric synthesis	34.66% \pm 1.62	0.21	0.44	3.85	0.47
- No RNN	29.96% \pm 1.55	1.24	0.04	7.92	0.28
Image match	32.98% \pm 1.59	0.37	0.16	8.39	0.18
Spacetime match	31.92% \pm 1.56	0.41	0.14	7.19	0.21
Image + spacetime	33.77% \pm 1.58	0.37	0.18	7.74	0.20
Random impact sound	19.77% \pm 1.34	0.44	0.00	9.32	0.02

(a) Model evaluation