

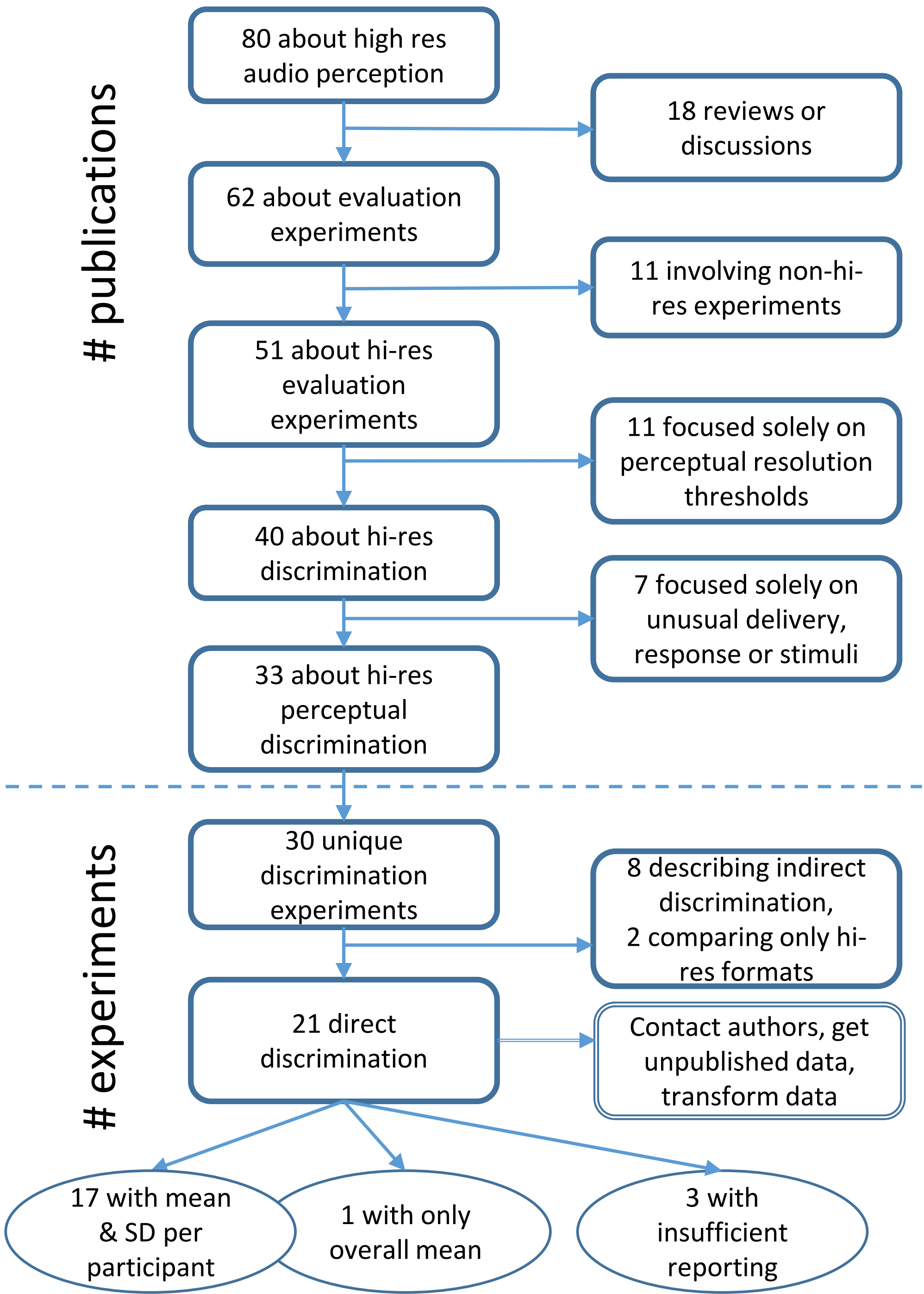
Meta-analysis of high resolution audio perception

centre for digital music

Josh Reiss, joshua.reiss@qmul.ac.uk



We assessed the ability to perceive a difference between high resolution (beyond 16 bit, 44.1 kHz) and standard CD quality audio. Over 50 publications were reviewed, and all experiments for which sufficient data could be obtained were subjected to meta-analysis. Results showed a small but significant ability to discriminate high resolution content, especially when test subjects received training. Potential biases, effect of methodology, experimental design and choice of stimuli were also investigated.

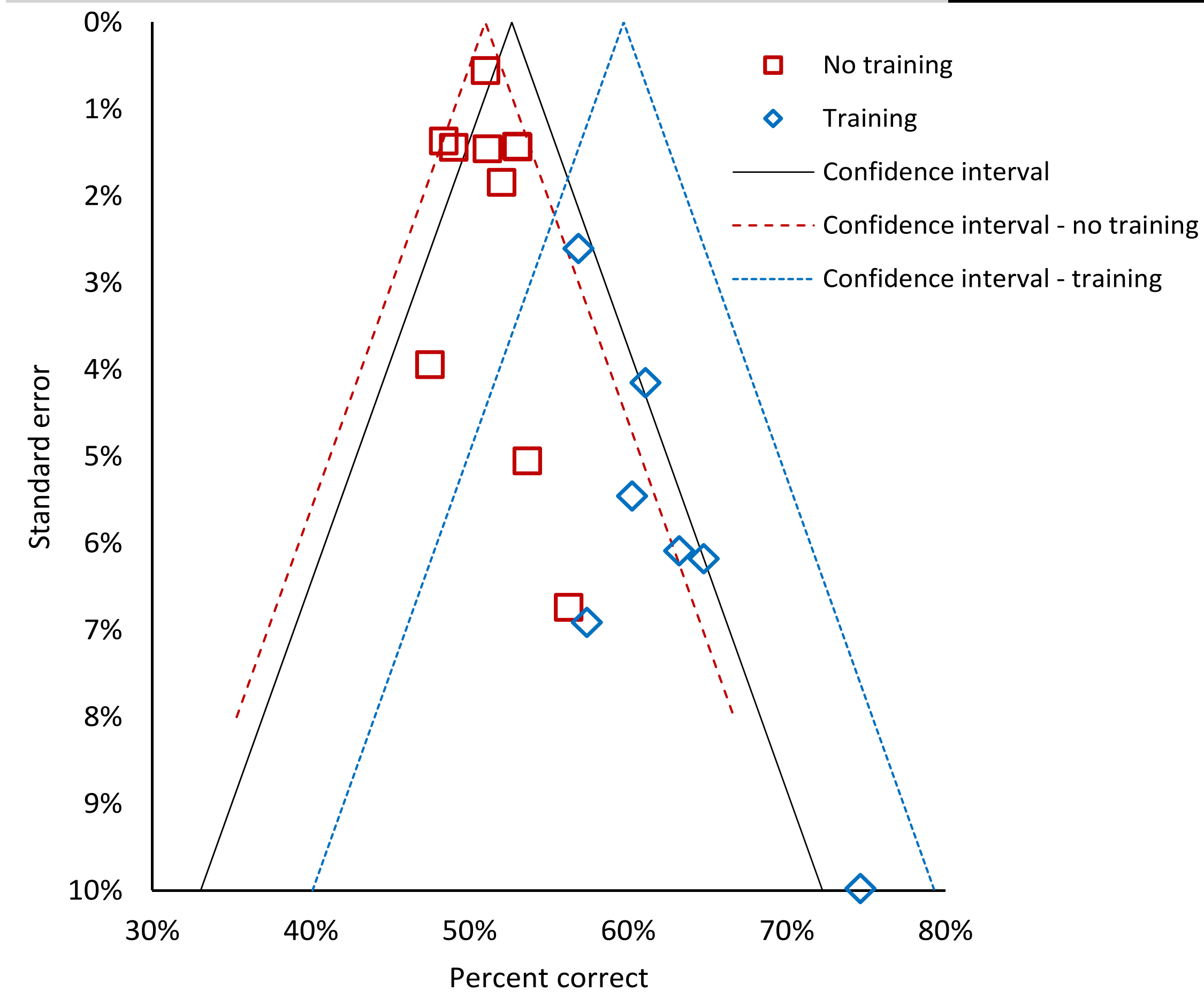
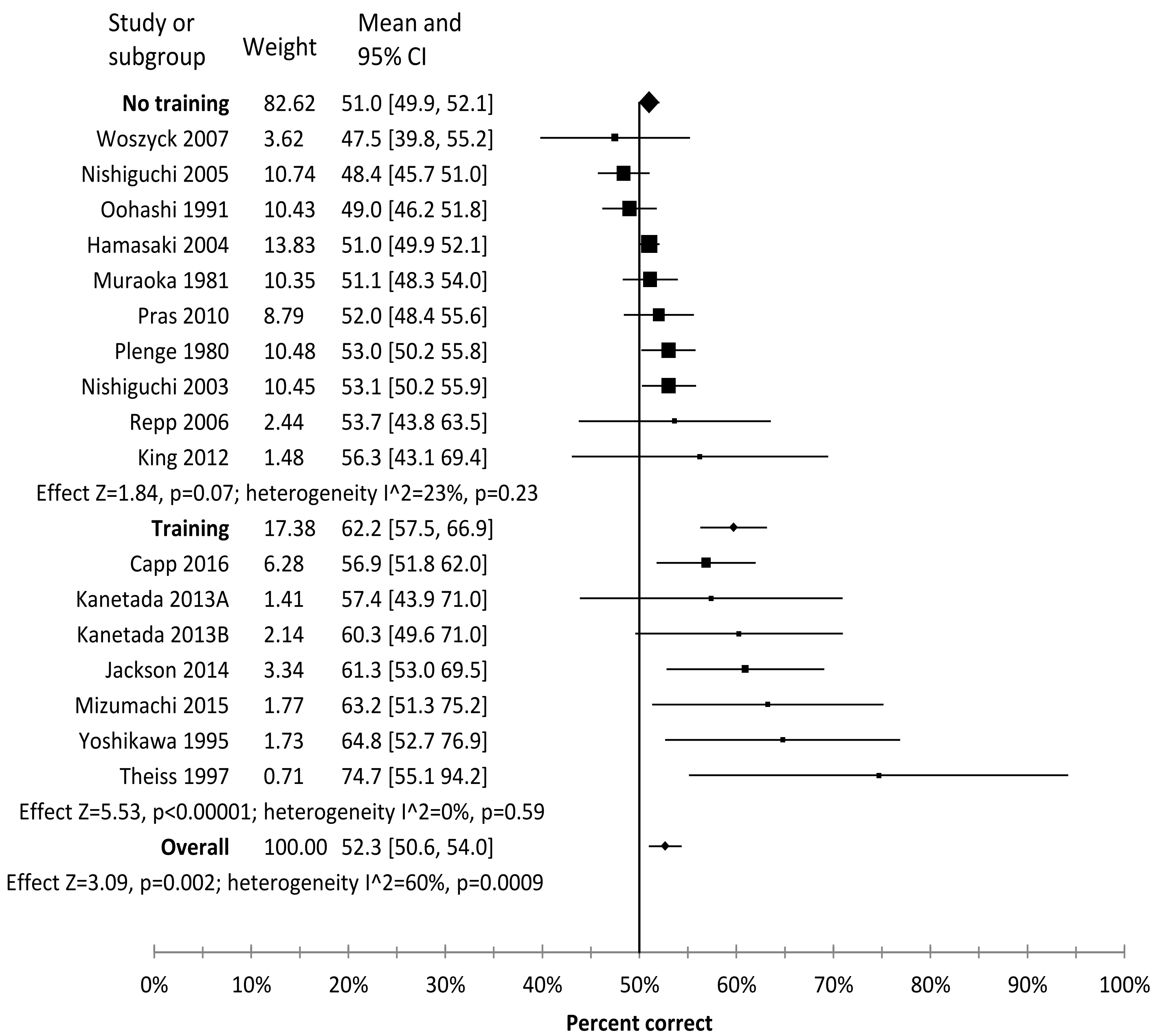


Test Type			Methodology
Auditory perception resolution	Bone conduction perception		pattern recognition, frequency JND
	Temporal resolution		2IFC, ABX
	Frequency resolution		Method of limits, 2IFC
	Joint time-frequency resolution		2IFC
Format discrimination	Indirect discrimination	Brain response	N/A
		Semantic description	DoD, Attribute rating
		Other (level, spatialisation, temporal resolution)	Method of adjustment, Method of limits
	Sufficient formats discrimination	Alternative Hi-Res Formats	ABX, AB
		Low resolution content	AXY, Same different
	High vs standard discrimination	Test signals	Same different, 2IFC
		Real world content	Same different, ABX, AXY, XY, Multistimulus rating

18 direct discrimination studies used in meta-analysis

- All studies with available data included
- Hypotheses gathered a priori
- All results subject to sensitivity analysis

A. Study		B. Risk of Bias							C. Binomial test			
Study	Year	Allocation bias	Methodology	Experimental design	Stimuli bias	Attrition bias	Reporting bias	Leading to	# correct	total	percent correct	probability
Plenge	1980	—	?	△	△	—	—	Type II errors	1367	2580	52.98%	1.294E-03
Muraoka	1981	—	?	△	?	—	—	Neutral	542	1060	51.13%	0.2400
Oohashi	1991	—	—	—	—	—	?	Neutral	392	800	49.00%	0.7261
Yoshikawa	1995	—	—	?	?	?	—	Neutral	85	132	64.39%	5.976E-04
Theiss	1997	—	?	?	?	△	—	Neutral	38	51	74.51%	3.105E-04
Nishiguchi	2003	—	—	—	△	—	—	Type II errors	489	920	53.15%	0.0301
Hamasaki	2004	—	—	—	?	—	?	Neutral	944	1848	51.08%	0.1821
Nishiguchi	2005	—	—	—	△	—	—	Type II errors	418	864	48.38%	0.8381
Repp	2006	—	△	△	△	△	—	Type II errors	42	86	48.84%	0.6267
Meyer	2007	△	—	?	△	△	△	Type II errors	276	554	49.82%	0.5507
Woszyck	2007	—	?	—	?	—	?	Type II errors	54	114	47.37%	0.7439
Pras	2010	?	—	?	?	—	?	Neutral	368	707	52.05%	0.1462
King	2012	—	△	—	?	△	?	Type II errors	34	61	55.74%	0.2213
KanetadaA	2013	?	—	?	—	—	—	Type I errors	62	108	57.41%	0.0743
KanetadaB	2013	?	—	?	—	—	—	Type I errors	135	224	60.27%	1.281E-03
Jackson	2014	—	?	—	—	—	—	Neutral	585	960	60.94%	6.352E-12
Mizumachi	2015	?	—	?	—	—	—	Type I errors	86	136	63.24%	1.279E-03
Capp	2016	—	?	—	—	—	—	Neutral	819	1440	56.88%	1.000E-07
Total									6736	12645	53.27%	1.006E-13



Small but statistically significant discrimination of high resolution audio

- Greatly improves with training
- Need for careful selection of stimuli
- ABX/AXY/XY tests give slightly stronger results than 1IFC Same/different
- Slightly stronger results with long duration stimuli and intervals
- Insufficient data to test the effect of quantization
- Evidence of Simpson's paradox in several studies
- Beware significance tests from multiple dichotomous trials