

Supplemental Material

Supplementary Analyses of Experiment 1

Musical experience

In a further analysis, we tested whether differences in musical experience between Dutch and Farsi speakers affected performance. In addition to linguistic height-pitch metaphors in Dutch, there are also associations between height and pitch in musical notation, in both Dutch and Farsi culture: Higher on the staff is higher in pitch. Therefore it is possible that differences in experience reading musical notation could have contributed to the observed effect. During a debriefing, participants rated how well they read music on a scale of 1-7. Musical experience did not interact with Language and Task, $F(1,79)=1.52$, $MSE=.41$, *ns*, and the interaction between Language and Task remained highly significant when the effect of musical experience was controlled, $F(1,79)=10.83$, $MSE=.29$, $p=.002$.

Figure S1

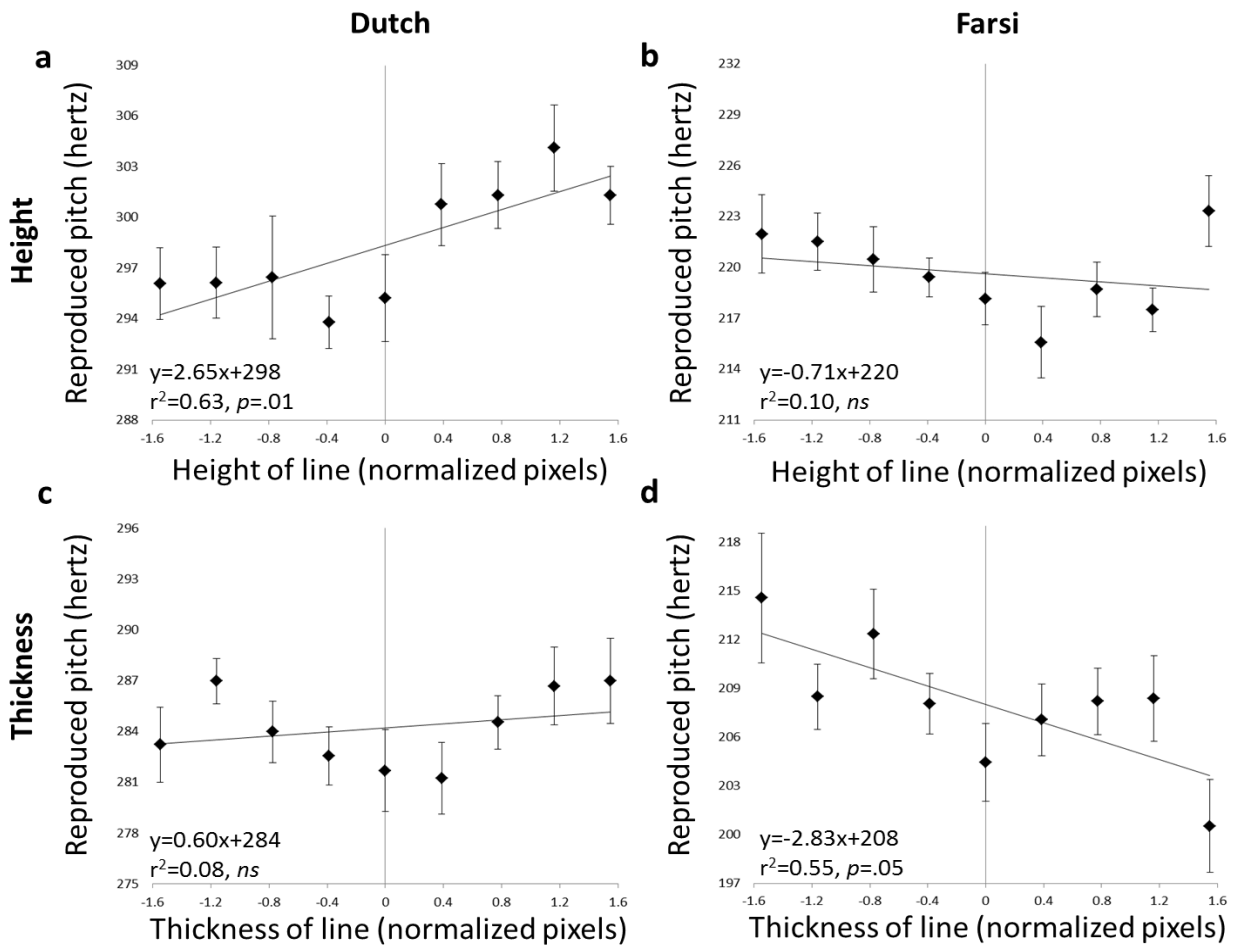


Figure S1. Results of Experiment 1. Effects of height interference (top) and thickness interference (bottom) on pitch estimates in speakers of Dutch (left) and Farsi (right). Error bars represent standard error of the mean.

Figure S2

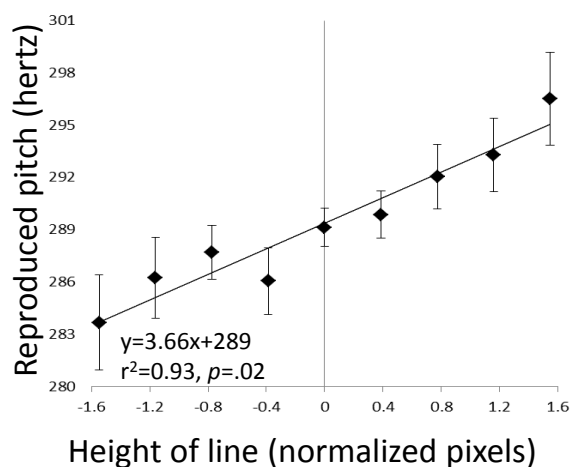


Figure S2. Results of Experiment 2. Spatial height still influenced pitch estimates, even under verbal interference. Error bars represent standard error of the mean.

Figure S3

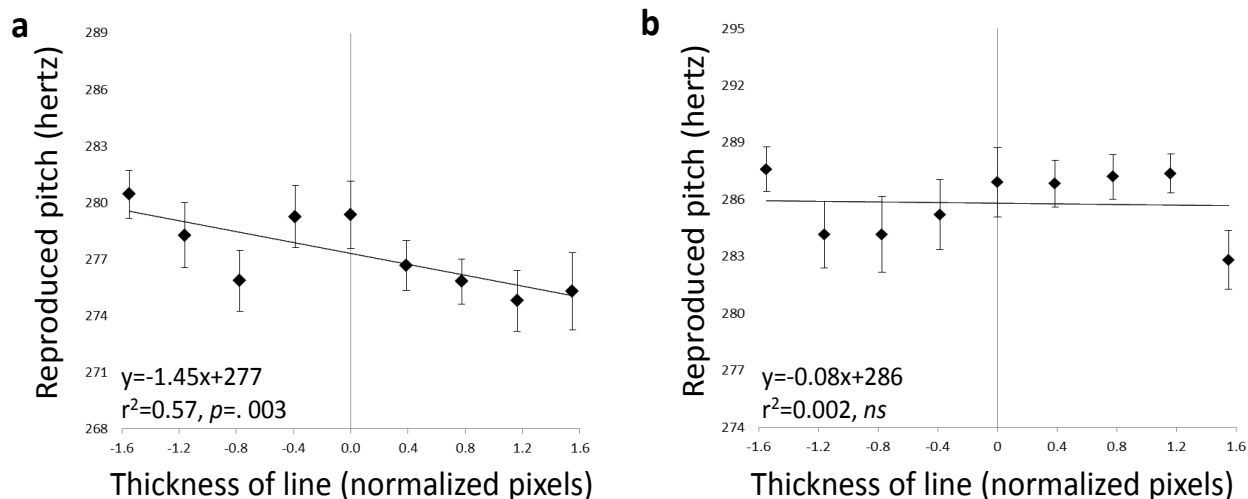


Figure S3. Results of Experiment 3. Effects of thickness interference on pitch estimates in speakers of Dutch after Thickness Training (left) and after Height Training (right). Thickness influenced pitch reproduction following Thickness Training, Slope=1.45, $p=.003$, but not following Height Training, Slope=0.08, *ns*. Error bars represent standard error of the mean.

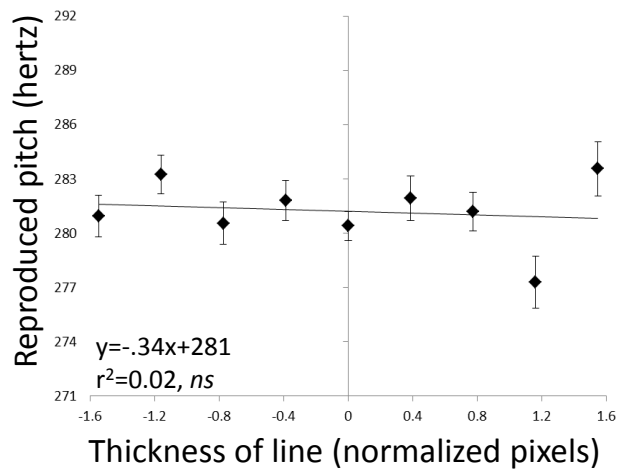
Figure S4

Figure S4. Results of Experiment 4. Effects of thickness interference on pitch estimates in speakers of Dutch after Reversed Thickness Training. Thickness did not influence pitch reproduction following Reversed Thickness Training, (Slope= $-.34$, *ns*). Error bars represent standard error of the mean.

Table S1

Experiment	Language	Condition	Slope	Range of averaged pitch estimates in Hertz (Hz) and cents (C)
1	Dutch	Height Interference	2.65 *	303-294 Hz; Difference = 10 Hz, 58 C
	Dutch	Thickness Interference	0.60	287-283 Hz; Difference = 4 Hz, 24 C
	Farsi	Height Interference	-0.71	216-222 Hz; Difference = -6 Hz, -47 C
	Farsi	Thickness Interference	-2.83 *	201-215 Hz; Difference = -14 Hz, -117 C
2	Dutch	Height Interference (with verbal suppression)	3.66 *	297-284 Hz; Difference = 13 Hz, 78 C
3	Dutch	Thickness Interference (after Thickness Training)	-1.45 *	275-280 Hz; Difference = -5 Hz, -31 C
		Thickness Interference (after Height Training)	-0.08	283-287 Hz, Difference = -4 Hz, -24 C
4	Dutch	Thickness Interference (after Reversed Thickness Training)	-0.34	277-283 Hz; Difference = -6 Hz, -37 C

Table S1. Range of averaged pitch estimates in Hertz (Hz) and cents (C) for all four experiments. The direction of the difference (positive vs. negative) corresponds to the slope of the effect of space on pitch estimation.