Emotion recognition and cochlear implants

Christina Fuller, Deniz Baskent, Rolien Free (UMCG – ENT/Audiology)

Dicky Gilbers, Steven Gilbers (RUG – Linguistics)

16 May, 2013
Methodology and Statistics Linguistic Research

Outline

Introduction

- Cochlear implant
- Phonetics of emotional speech
- Recordings of emotional speech

Pitch analyses — Independent samples t-test

- Pitch range, mean pitch, modality — Mann-Whitney U test

Emotion recognition — Independent samples t-test

- NH vs. CI — Mann-Whitney U test

NH and CI recognition patterns — Kruskal-Wallis H Test

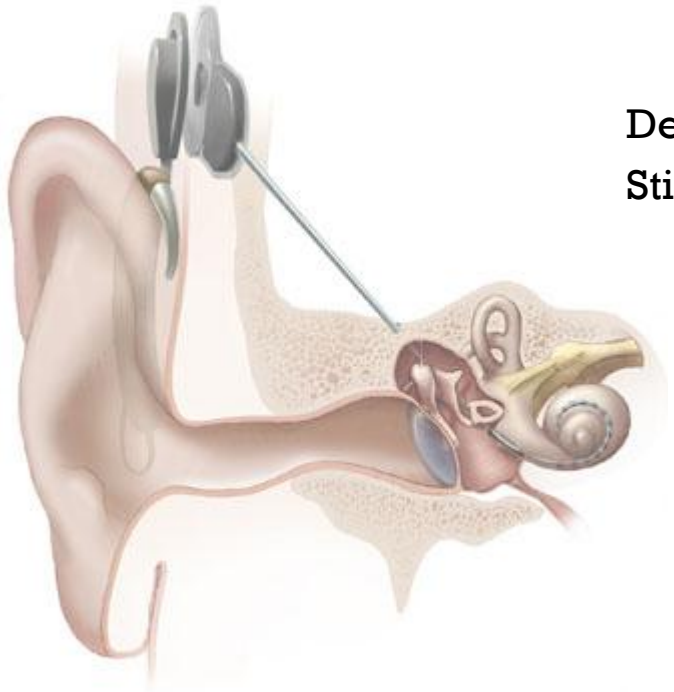
- Per speaker — Mann-Whitney U test

Optimality Theory account

Conclusion

Questions

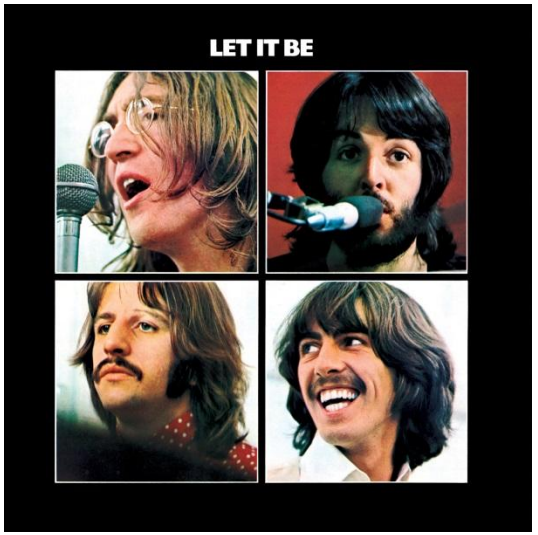
Cochlear implant



Device surgically placed into a deaf patient's cochlea
Stimulates cochlear nerves electrically

CI simulation for music

The Beatles – Let It Be



Phonetics of emotional speech

Arousal and Valence parameters

		Valence	
		Positive	Negative
Arousal	High	Joy Pride	Anger Fear
	Low	Tenderness Relief	Sadness Irritation

Goudbeek & Broersma (2010)

Recordings of emotional speech

Nonce word

[nutohɔmsɛpikɑŋ]

(satisfying both Korean & Dutch phonotactics)

Audio recordings of 8 actors

4 recordings per emotion

4 best recognized emotions selected

2 best selected for analysis

fear			**joy**		
anger			pride		
sadness			tenderness		
irritation			**relief**		

Pitch analyses

"**[A]ngry and happy** speech exhibits **higher mean pitch** [and] a **wider pitch range**, (…) whereas **sad speech** exhibits **lower mean pitch** [and] a **narrower pitch range**."

Luo, Fu & Galvin (2007)

Pitch range

Hypothesis: high arousal → wider pitch range

Mean pitch

Hypothesis: high arousal → higher mean pitch

Extra variable *(D. Gilbers, Force of Articulation Model)*

Modality

Hypothesis: high arousal → more frequency peaks

Method

Pitch range

> "Post hoc Bonferroni *t* tests showed that the female talker had a significantly larger range of F_0 variation than the male talker ($p < .03$) for all target emotions."
>
> *Luo, Fu & Galvin (2007)*

Possibly erroneous conclusion due to measuring pitch range in terms of Hz

A (110Hz) [1 octave – range = **110 Hz**] [1 octave – range = **12 semitones**]

A (220 Hz) [1 octave – range = **220 Hz**] [1 octave – range = **12 semitones**]

A (440 Hz)

Frequency range not in Hertz (Hz) but in amount of semitones

→ allows fair comparison between male and female speakers

→ allows fair comparison of pitch range for high and low arousal

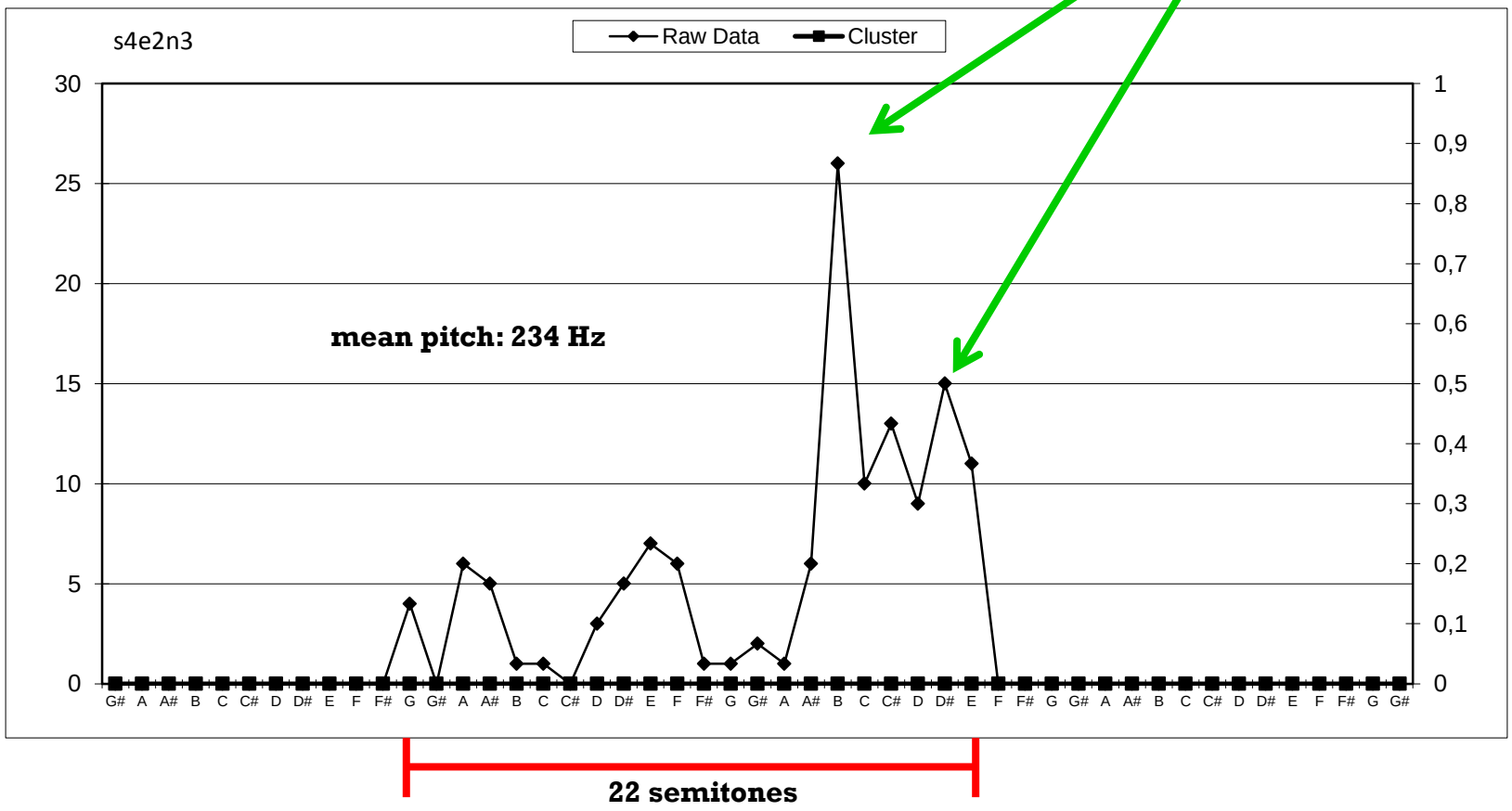
(assuming the "high arousal → higher mean pitch" hypothesis is true)

Method

Anger

High arousal, negative valence

Speaker 4, male

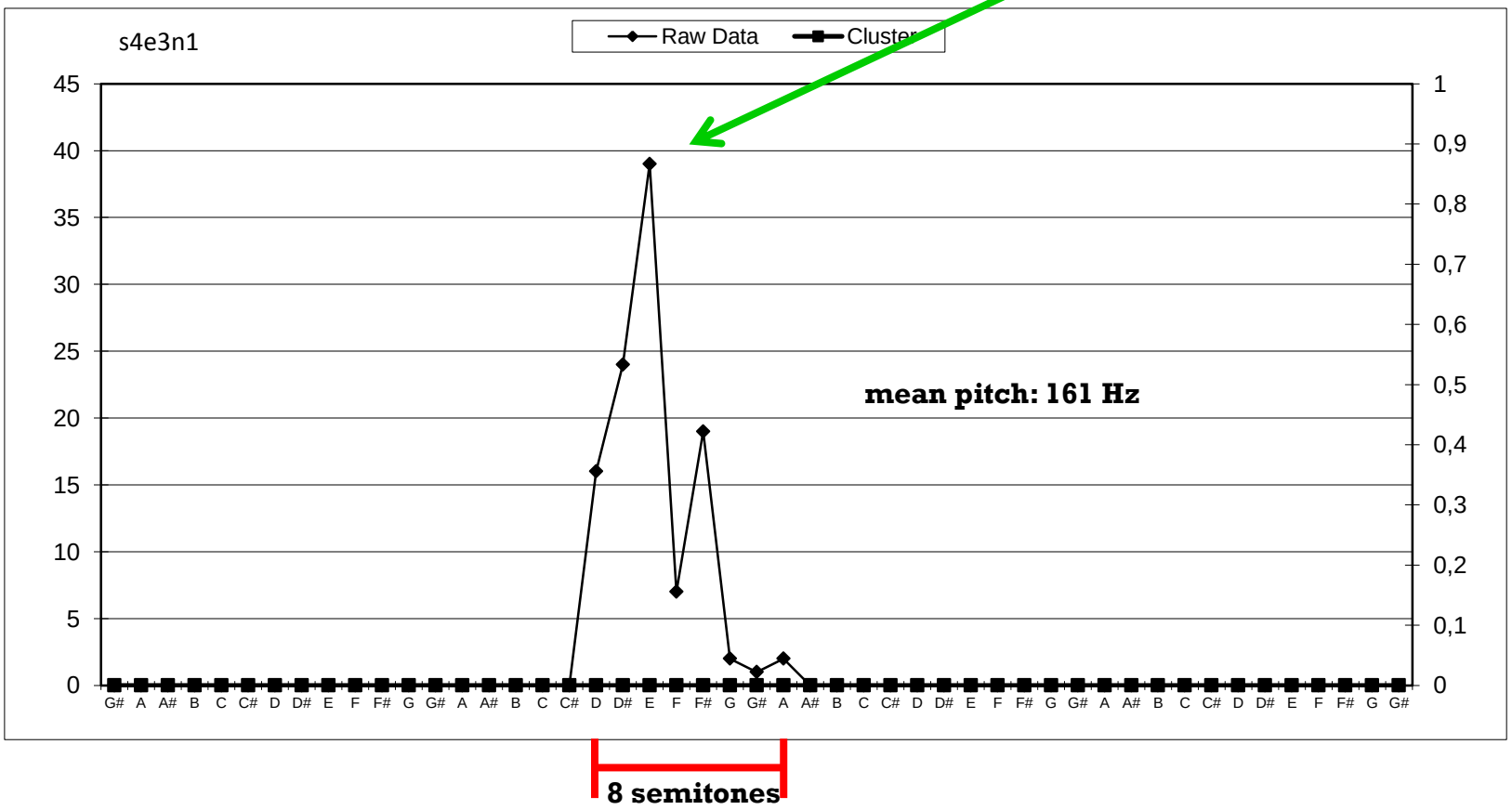


Method

Sadness

Low arousal, negative valence

Speaker 4, male

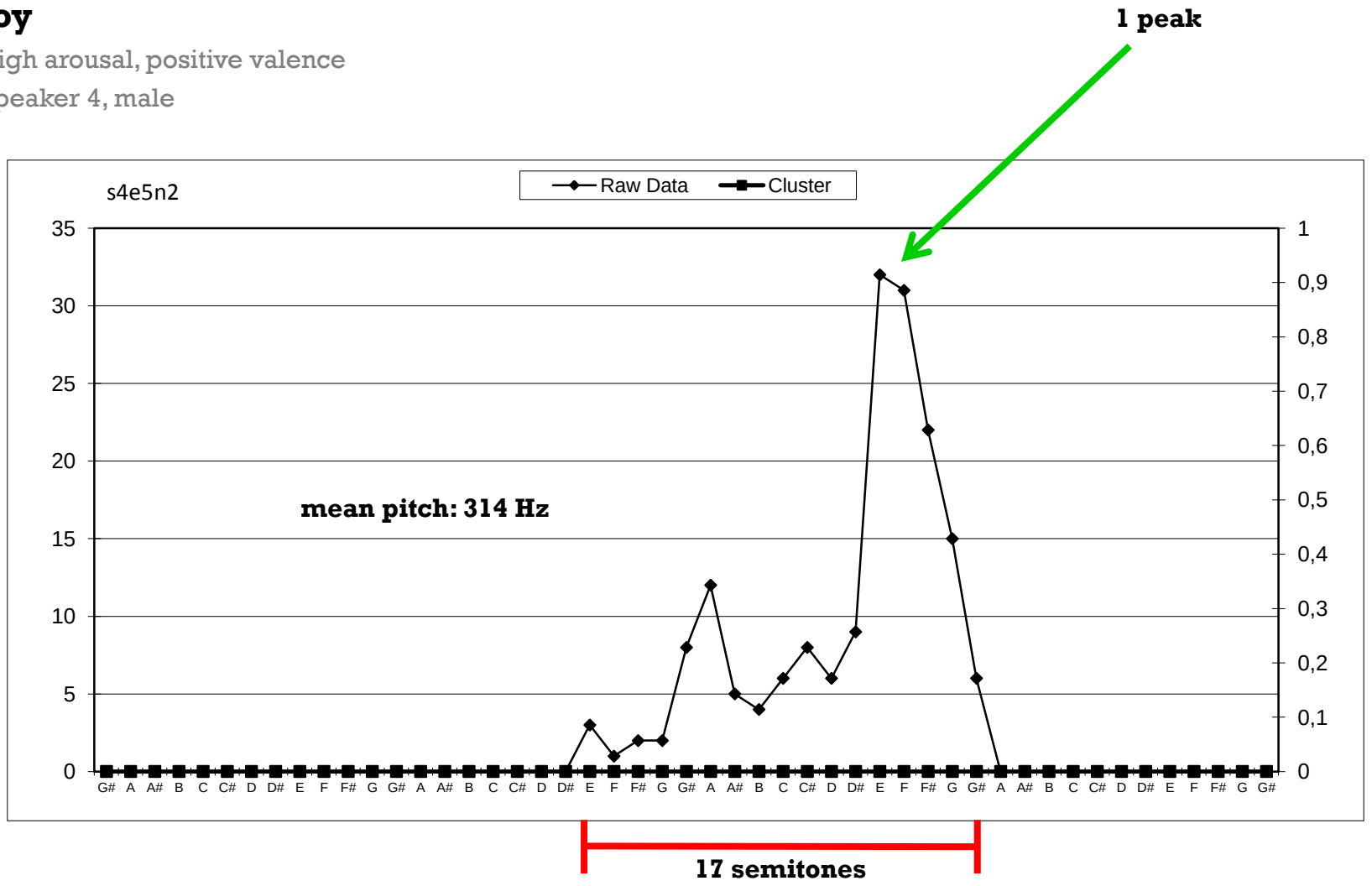


Method

Joy

High arousal, positive valence

Speaker 4, male

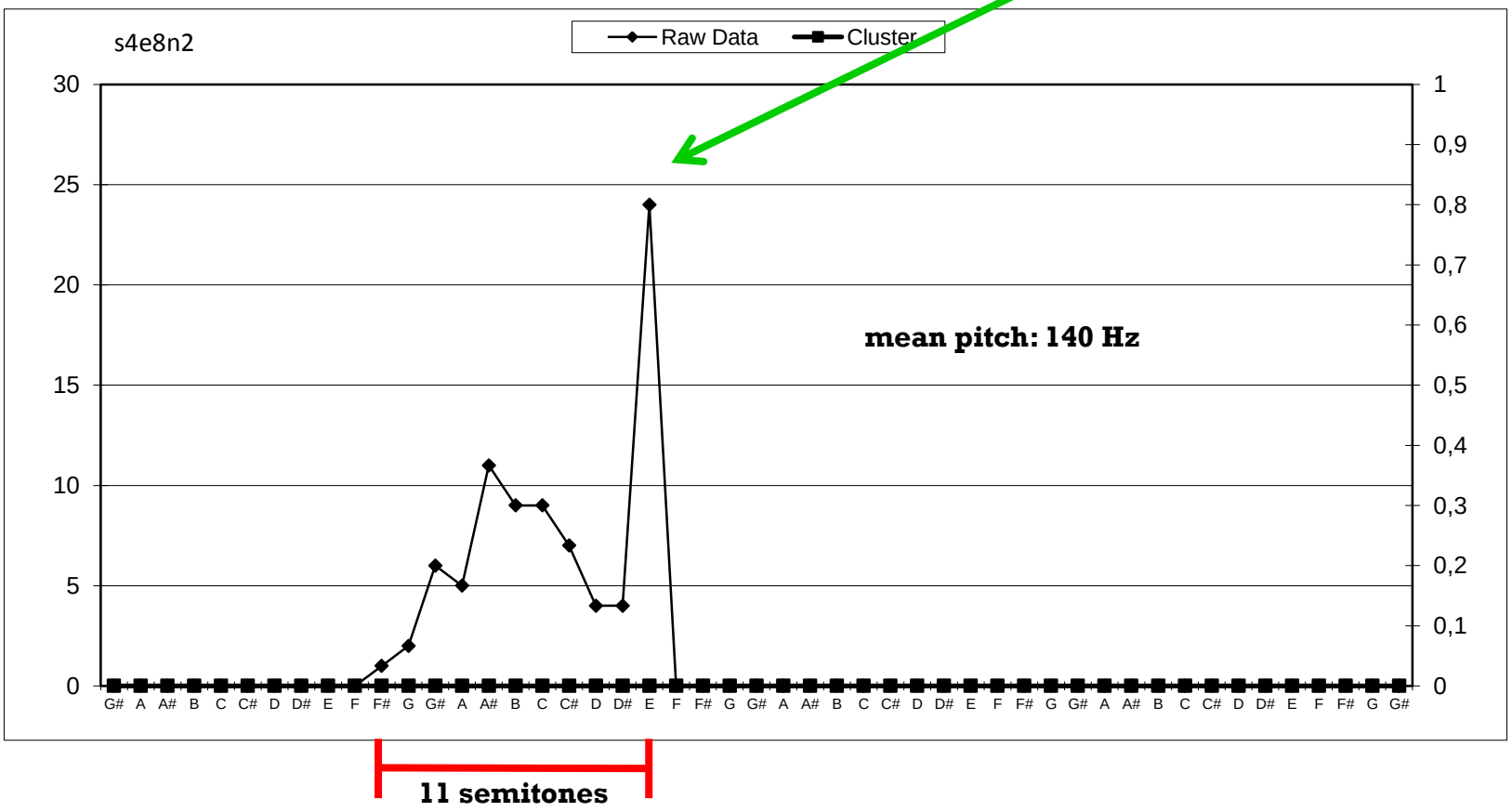


Method

Relief

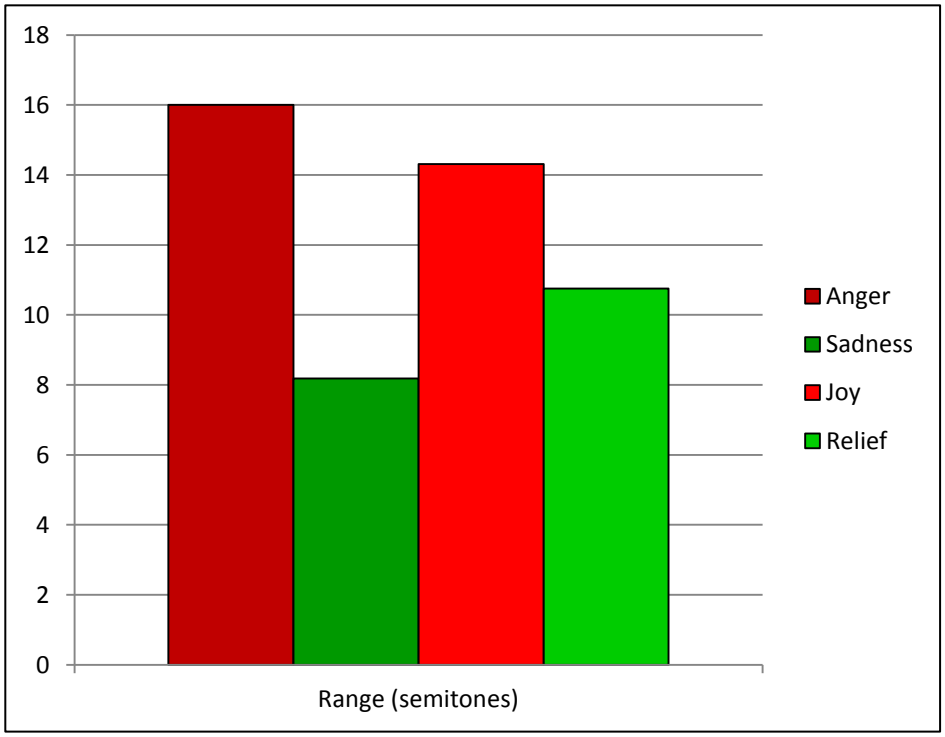
Low arousal, positive valence

Speaker 4, male



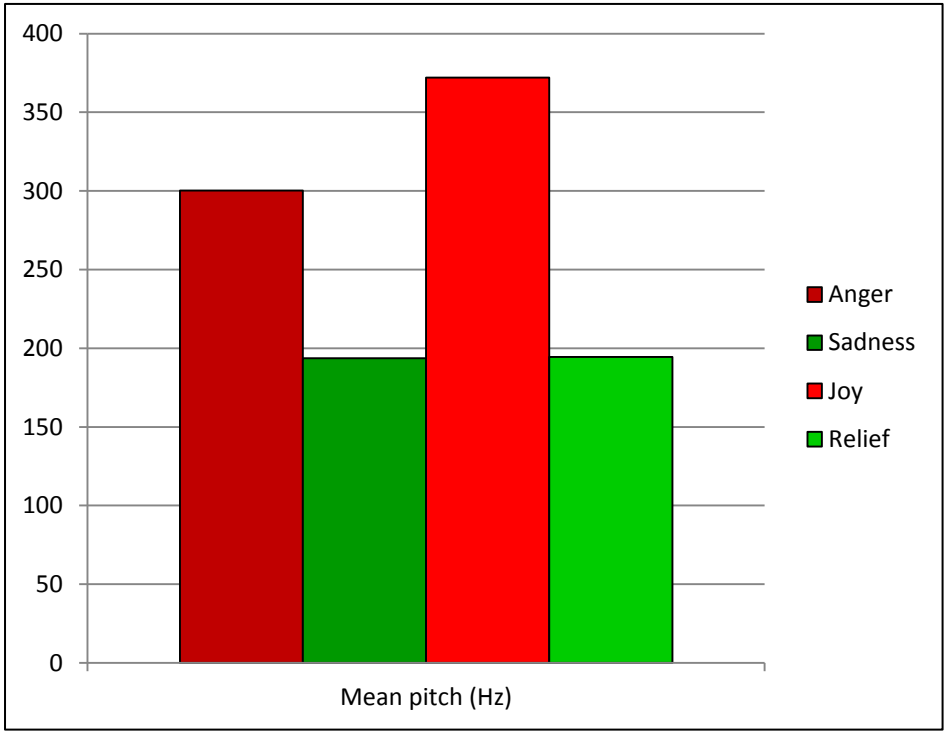
Method

	Range (st)	Mean (Hz)	Peaks
Anger	16	300	1,375
Sadness	8,1875	194	1,125
Joy	14,3125	372	1,625
Relief	10,75	194	1,3125



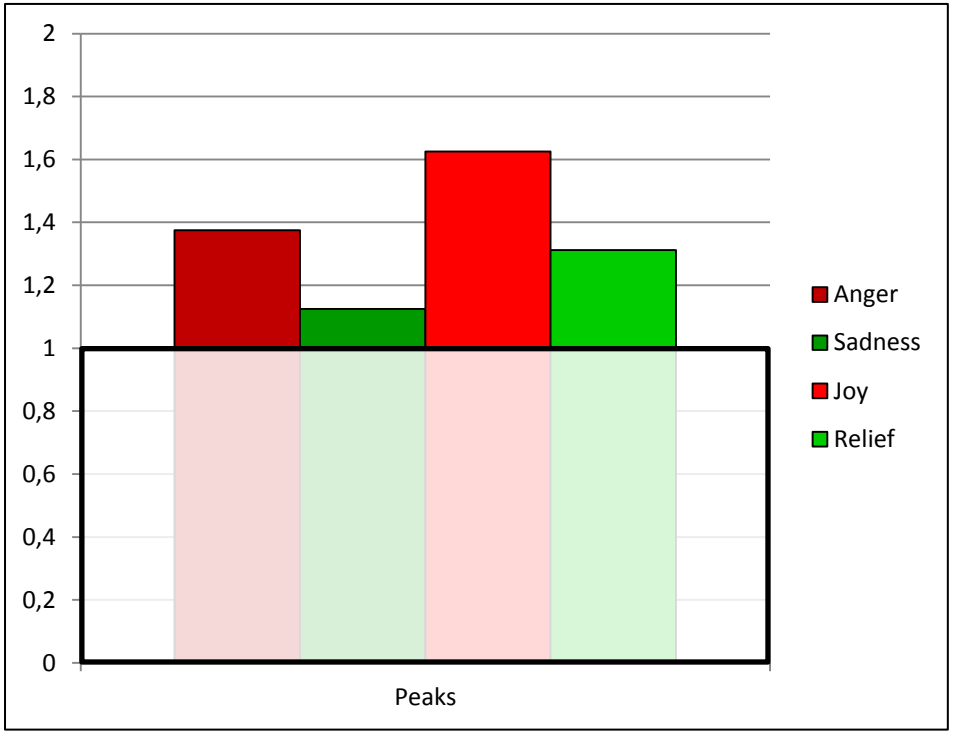
Method

	Range (st)	Mean (Hz)	Peaks
Anger	16	300	1,375
Sadness	8,1875	194	1,125
Joy	14,3125	372	1,625
Relief	10,75	194	1,3125



Method

	Range (st)	Mean (Hz)	Peaks
Anger	16	300	1,375
Sadness	8,1875	194	1,125
Joy	14,3125	372	1,625
Relief	10,75	194	1,3125



Statistical analysis

Pitch patterning along the arousal parameter?

Pitch range

H_0: No difference between high and low arousal emotions for pitch range

H_1: Difference between high and low arousal emotions for pitch range, namely wider for high arousal emotions

Mean pitch

H_0: No difference between high and low arousal emotions for mean pitch

H_1: Difference between high and low arousal emotions for mean pitch, namely higher for high arousal emotions

Modality

H_0: No difference between high and low arousal emotions for amount of peaks

H_1: Difference between high and low arousal emotions for amount of peaks, namely more for high arousal emotions

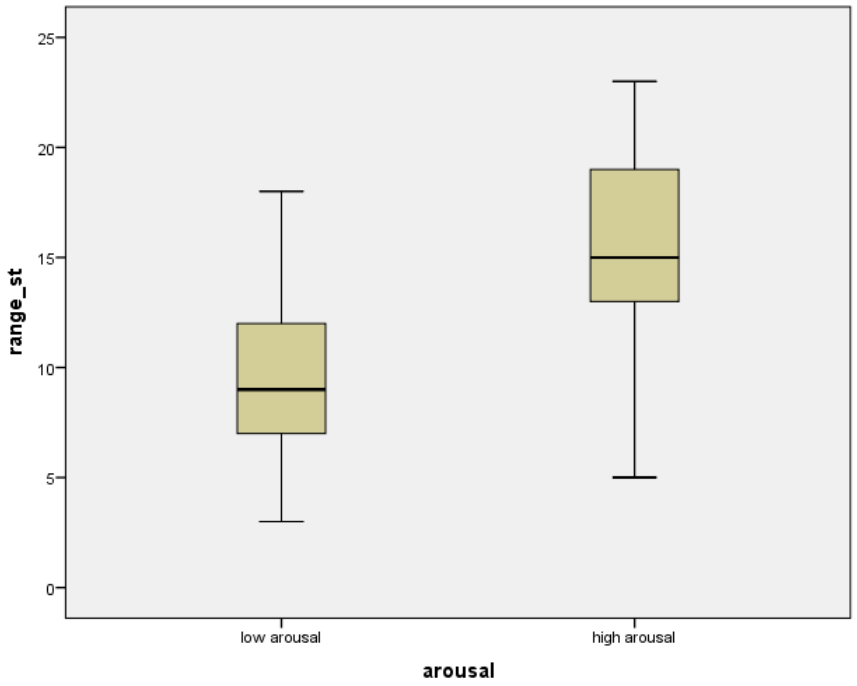
Statistical analysis

Pitch range

Independent samples t-test

The independent samples t-test found that, on average, high arousal emotions had a wider pitch range in semitones (M=15.16, SE=.699) than low arousal emotions (M=9.47, SE=.654). This difference was significant t(62)=5.944, $p < .001$. Therefore, H_0 can be safely rejected in favor of H_1.

	N	Mean	SD	SE
High arousal	32	15,16	3,952	0,699
Low arousal	32	9,47	3,698	0,654



Statistical analysis

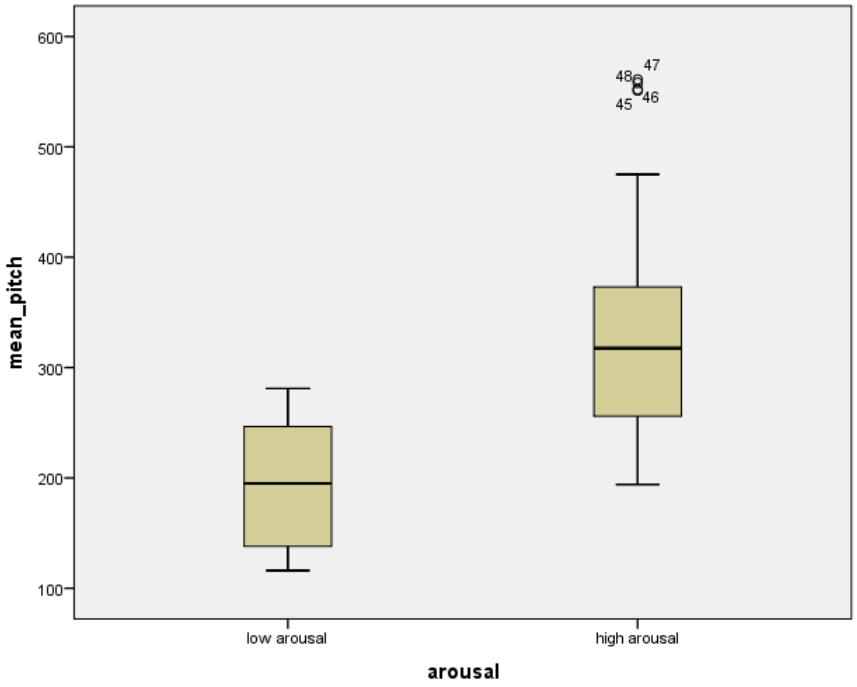
Mean pitch

Mann-Whitney U-test

The Mann-Whitney U-test found that the mean pitch for high arousal emotions was significantly higher than for low arousal emotions, $U(n_1=32, n_2=32) = 96.5, p < .001$. Therefore, H_0 can be safely rejected in favor of H_1.

	N	Mean	SD	SE
High arousal	32	209,2	79,72	14,09
Low arousal	32	321	108,2	19,12

	N	Mean rank
High arousal	32	209,19
Low arousal	32	320,97



Statistical analysis

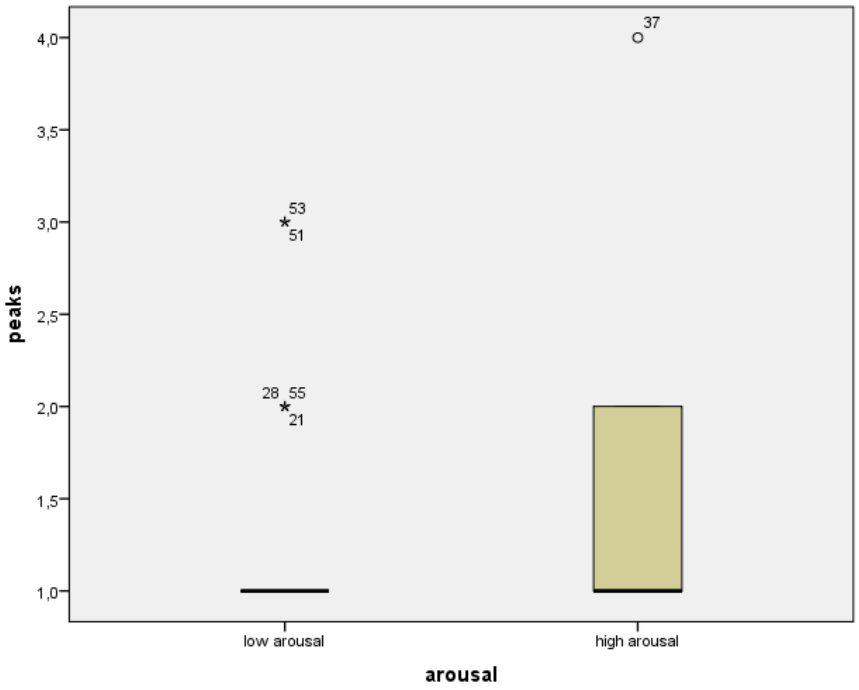
Modality

Mann-Whitney U-test

The Mann-Whitney U-test found that the modality for high arousal emotions was significantly higher than for low arousal emotions, $U(n_1=32, n_2=32) = 378.5, p < .05$. Therefore, H_0 can be safely rejected in favor of H_1.

	N	Mean	SD	SE
High arousal	32	1,5	0,672	0,119
Low arousal	32	1,22	0,553	0,098

	N	Mean rank
High arousal	32	28,33
Low arousal	32	36,67



Emotion recognition

Normal hearing participants (NH)

Training

Emotion

Emotion (normalized for intensity)

CI simulation

CI simulation (normalized for intensity)

Cochlear implant users (CI)

Training

Emotion

Emotion (normalized for intensity)

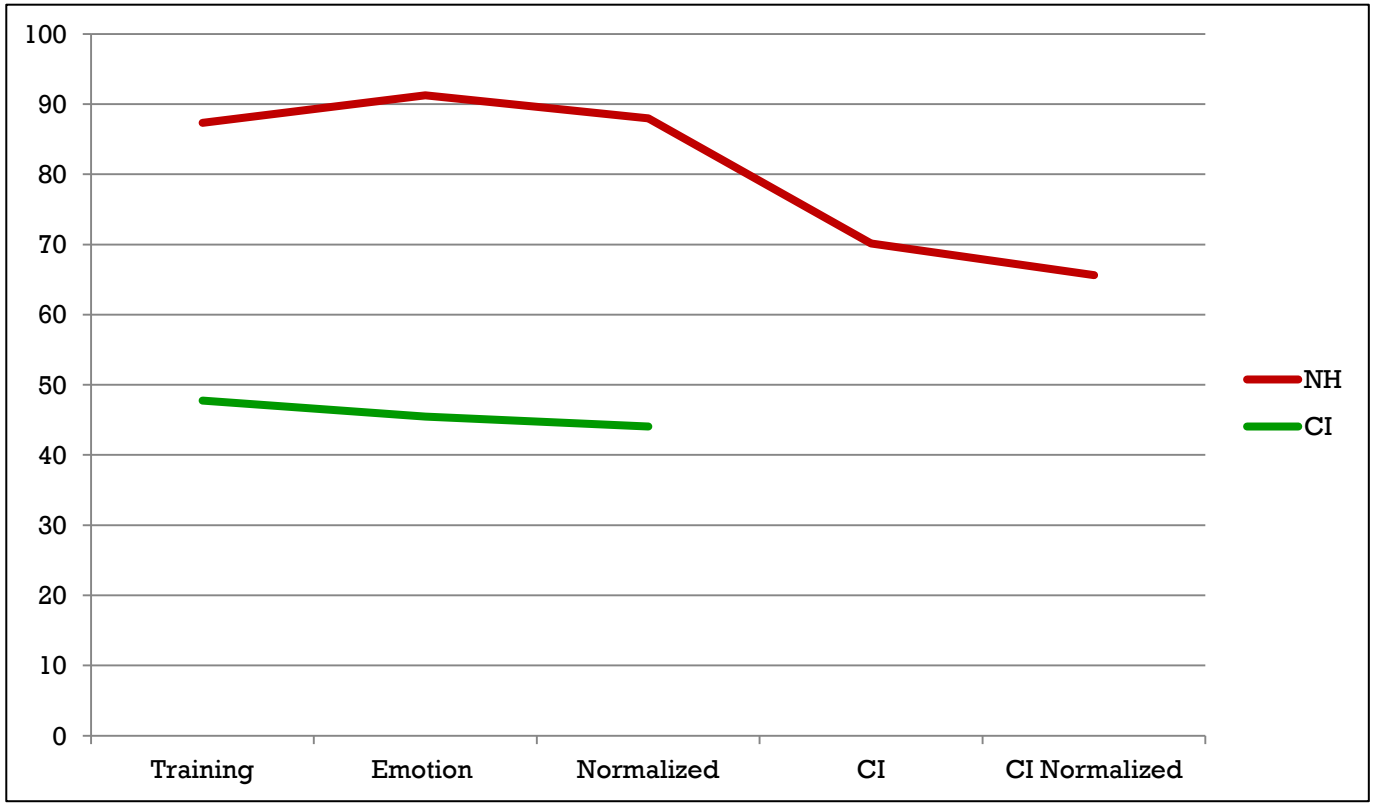
H_0: No difference between NH and CI regarding emotion recognition

H_1: Difference between NH and CI regarding emotion recognition

Method

NH vs. CI

	Training	Emotion	Emotion norm.	CI	CI norm.
NH	87,35	91,25	87,96	70,16	65,63
CI	47,75	45,48	44,07		



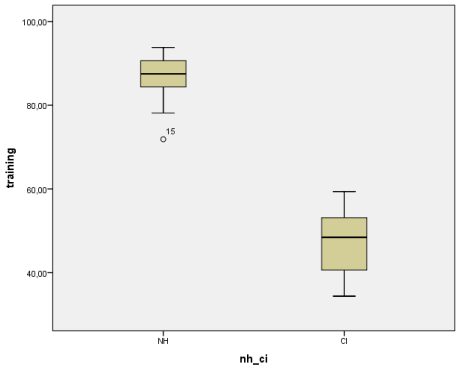
Statistical analysis

NH vs. CI

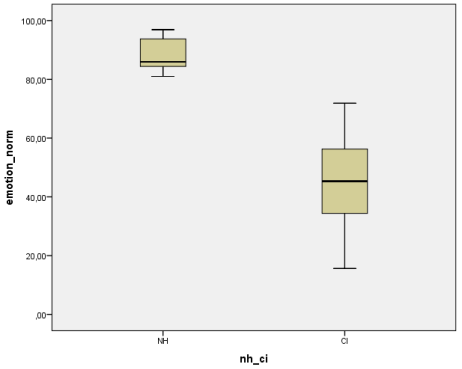
'Training', 'emotion', 'emotion normalized'

Mann-Whitney U test

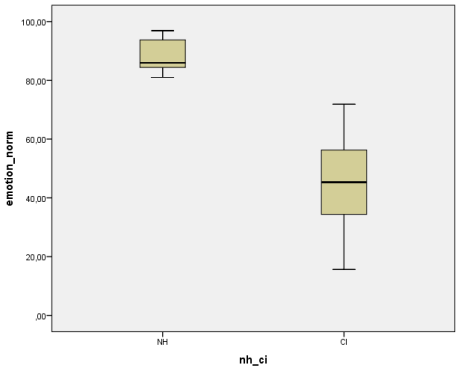
The Mann-Whitney U-test found that NH scored significantly better than CI in the 'training' condition, $U(n_1=20, n_2=18) = .000, p < .001$.



The Mann-Whitney U-test found that NH scored significantly better than CI in the 'emotion' condition, $U(n_1=20, n_2=20) = .000, p < .001$.



The Mann-Whitney U-test found that NH scored significantly better than CI in the 'emotion normalized' condition, $U(n_1=20, n_2=20) = .000, p < .001$.



Statistical analysis

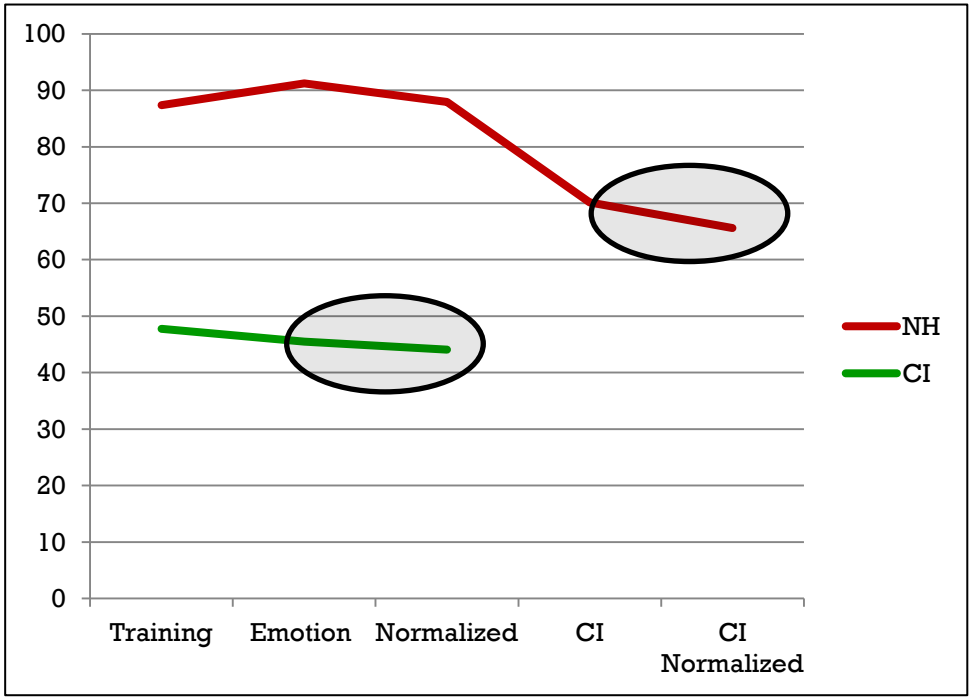
NH's CI simulation vs. CI's emotion

NH's 'CI simulation' & 'CI normalized' CI's 'emotion' & 'emotion normalized'

Independent samples t-test

H_0: No difference between NH and CI regarding emotion recognition w/ CI sound

H_1: Difference between NH and CI regarding emotion recognition



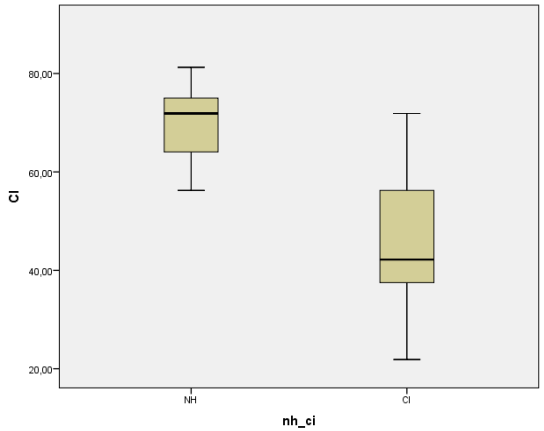
Statistical analysis

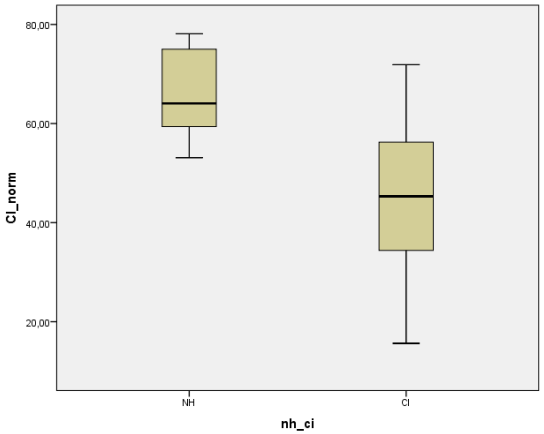
NH's CI simulation vs. CI's emotion

NH's 'CI simulation' & 'CI normalized'

CI's 'emotion' & 'emotion normalized'

Independent samples t-test





		N	Mean	SD	SE
Emotion/CI	NH	20	70,16	6,759	1,511
	CI	20	45,48	14,53	3,249
Emotion/CI normalized	NH	20	65,63	8,051	1,8
	CI	20	44,07	14,15	3,164

The independent samples t-test found that, on average, NH performed better at the recognition task with CI sound (M=70.16, SE=1.511) than CI (M=45.48, SE=3.249). This difference was significant t(38)=6.888, $p < .001$.

The independent samples t-test found that, on average, NH performed better at the recognition task with normalized CI sound (M=65.63, SE=1.8) than CI (M=44.07, SE=3.164). This difference was significant t(38)=5,921, $p < .001$.

NH and CI recognition patterns

Comparison between NH and CI recognition patterns (per speaker)

How well were the emotions portrayed by individual actors recognized?

H_0: No difference between the different speakers regarding the degree to which their emotions are correctly identified

H_1: Difference between the different speakers regarding the degree to which their emotions are correctly identified

H_0: No difference between NH's and CI's recognition patterns

(i.e. an apparently less well performing speaker's emotions will be recognized worse than the other speakers' by both NH and CI listeners and vice versa)

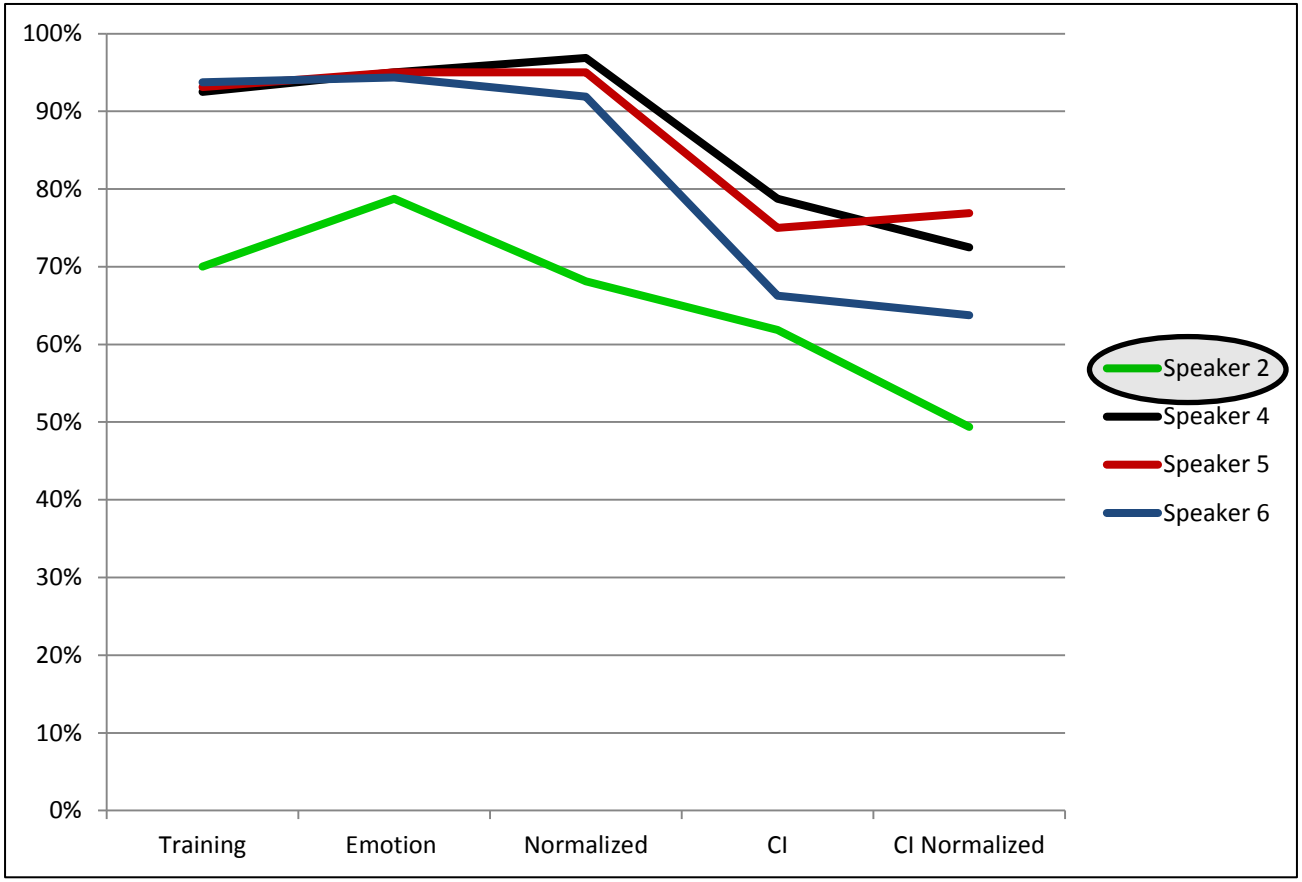
H_1: Difference between the different speakers regarding the degree to which their emotions are correctly identified

(i.e. an apparently less well performing speaker's emotions will be recognized worse than the other speakers' by both NH and CI listeners and vice versa)

NH and CI recognition patterns

Comparison between NH and CI recognition patterns (per speaker)

NH listeners



NH and CI recognition patterns

Comparison between NH and CI recognition patterns (per speaker)

NH listeners

The Kruskal-Wallis test found that the recognition scores of the four speakers in the '**emotion**' condition differed significantly, $\chi^2(3) = 18.343, p < .001$.

A Post Hoc analysis using the Mann-Whitney U test revealed that only the differences between speaker 2 and speaker 4, 5, and 6 were significant at $p < 0.01$ for said pairs.

The Kruskal-Wallis test found that the recognition scores of the four speakers in the '**emotion normalized**' condition differed significantly, $\chi^2(3) = 44.326, p < .001$.

A Post Hoc analysis using the Mann-Whitney U test revealed that only the differences between speaker 2 and speaker 4, 5, and 6 were significant at $p < 0.001$ for said pairs.

The Kruskal-Wallis test found that the recognition scores of the four speakers in the '**CI**' condition differed significantly, $\chi^2(3) = 13.527, p < .01$.

A Post Hoc analysis using the Mann-Whitney U test revealed that the differences between speaker 2 and speaker 4, 5, and 6 were significant at $p < 0.01$ (for the speaker 2-4 pair) and at $p < 0.05$ (for the speaker 2-5 pair), and the difference between speaker 6 and 4 was significant at $p < 0.05$.

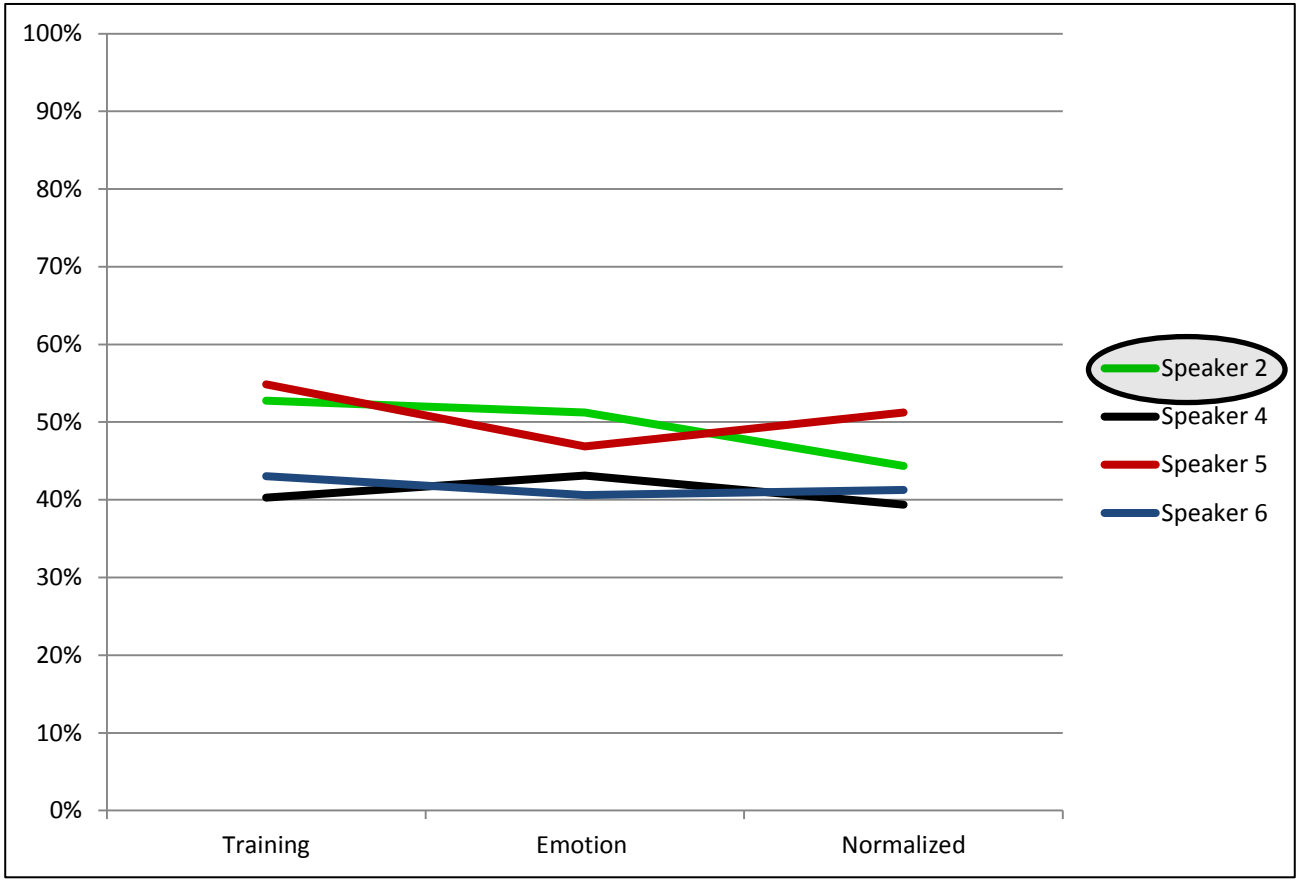
The Kruskal-Wallis test found that the recognition scores of the four speakers in the '**CI normalized**' condition differed significantly, $\chi^2(3) = 27.44, p < .001$.

A Post Hoc analysis using the Mann-Whitney U test revealed that the differences between speaker 2 and speaker 4, 5, and 6 were significant at $p < 0.001$ (for the speaker 2-4 and 2-5 pairs) and at $p < 0.01$ (for the speaker 2-6 pair), and the difference between speaker 6 and speaker 5 was significant at $p < 0.01$.

NH and CI recognition patterns

Comparison between NH and CI recognition patterns (per speaker)

CI listeners



NH and CI recognition patterns

Comparison between NH and CI recognition patterns (per speaker)

CI listeners

The Kruskal-Wallis test found that the recognition scores of the four speakers in the '**emotion**' condition did not differ significantly, $\chi^2(3) = 2.977, p = .395$.

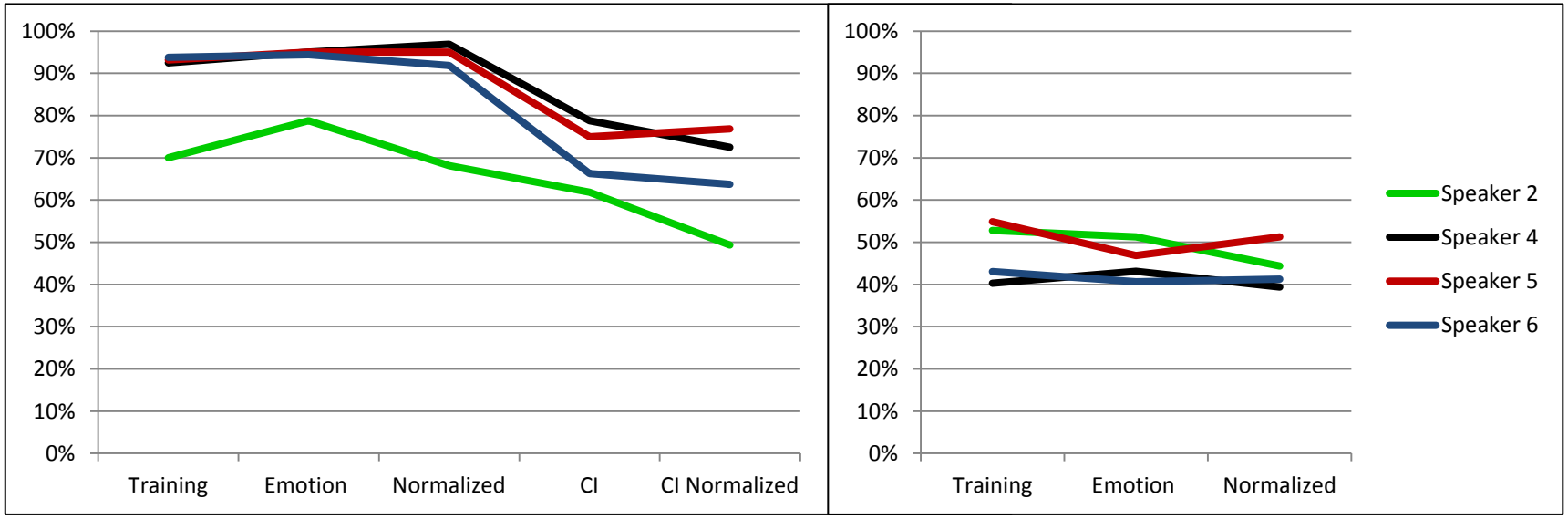
The Kruskal-Wallis test found that the recognition scores of the four speakers in the '**emotion normalized**' condition did not differ significantly, $\chi^2(3) = 2.800, p = .423$.

NH and CI recognition patterns

Comparison between NH and CI recognition patterns (per speaker)

Speaker 2's emotions (and to a lesser extent speaker 6) recognized worse for NH listeners, but not for CI listeners

Apparently, speaker 2's recordings differ from the others' in such a way that NH but not CI listeners recognize his emotions less well



Optimality Theory account

How does speaker 2 differ from the other speakers?

Many phonetic cues play a role in recognizing emotions

Force of Articulation Model (D. Gilbers)

Mean pitch	Pitch range
Hyperarticulation	Word length
Syllable isochrony	Occlusion duration
Voice Onset Time (VOT)	Plosive release duration
…	

Optimality Theory account

Hypothesis

NH and CI users use the same phonetic cues in determining which emotion they are confronted with, but the relative importance they assign to these phonetic cues differs between them

Perhaps due to for instance limitations of cochlear implants, extensive lack of exposure to sound, etc.

Optimality Theory account

Mean Pitch

Speaker 2 has a lower voice; CI cut-off 160 Hz → CI-user misses prevoicing cues (plus some F_0 information)

mean pitch less important cue for CI-users

Pitch Range

Ratio range anger-sadness Speaker 2	→	3:1
Ratio range anger-sadness Speakers 4, 5, and 6	→	~2:1

pitch range important cue for CI-users

Hyperarticulation

Speaker 2 has less hyperarticulated speech

hyperarticulation less important for CI-users

Word length

Speaker 2 speaks fastest

(slow) speech rate less important for CI-users

Isochrony

Speaker 2 scores relatively best

isochrony important for CI-users

Occlusion/VOT/plosive duration

No differences

Optimality Theory account

Cues for emotion recognition – NH and CI orderings

Preliminary

NH

Acoustic signal	MEAN PITCH	HYPER-ARTICULATION	SPEECH RATE	PITCH RANGE	ISOCHRONY	OCCLUSION VOT SEGM. LENGTH
Anger						
Sadness						
Joy						
Relief						

CI

Acoustic signal	PITCH RANGE	ISOCHRONY	MEAN PITCH	HYPER-ARTICULATION	SPEECH RATE	OCCLUSION VOT SEGM. LENGTH
Anger						
Sadness						
Joy						
Relief						

Conclusion

Pitch analyses

Pitch range: significantly wider for high arousal

Mean pitch: significantly higher for high arousal

Modality: significantly more peaks for high arousal

Emotion recognition

NH scored significantly better than CI on emotion recognition task (also when comparing NH's CI simulation scores to CI's actual implant scores)

NH and CI recognition patterns

Speaker 2's emotions are less well recognized than the others' by NH

Speaker 2's emotions are equally well recognized as the others' by CI

Optimality Theory account

NH and CI users use the same phonetic cues in determining which emotion they are confronted with, but the relative importance they assign to these phonetic cues differs between them

NH: mean pitch > pitch range **CI:** pitch range > mean pitch

Questions