CURRENT RESEARCH IN THE SMART LAB ON MUSIC AND EMOTIONAL SPEECH IN HEARING IMPAIRED & HEARING AIDED LISTENERS

Frank A. Russo, Ph.D.

Ryerson University

Smart Lab

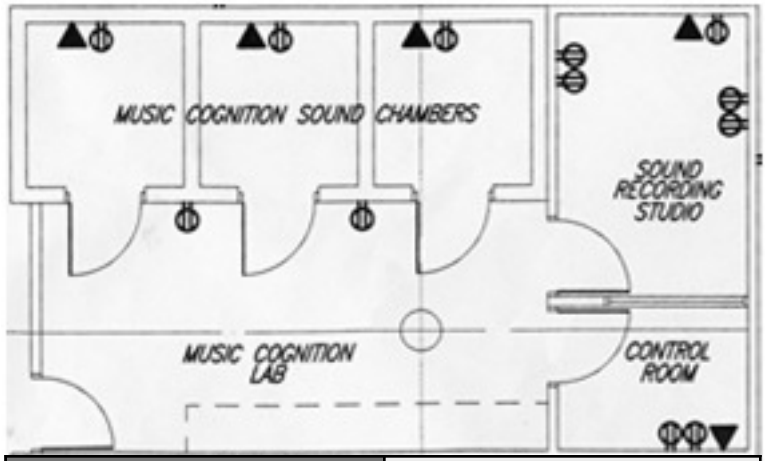
What will I cover today?

1. INFRASTRUCTURE AND METHODS
2. RAVDESS
    1. Behavior
    2. Physiological Response
    3. Modeling Emotion
3. EMO-CHEQ
4. AMP
    1. NH vs. HI
    2. HI vs. HA
5. Effects of HA on pitch tracking at the brainstem
6. Effects of Compression in music
7. Tracking Emotion in Music in HA

SMART Lab

- INFRASTRUCTURE
  - 4 IAC Chambers
  - Recording Studio
  - A, V, VT

- PRINCIPAL METHODS
  - AUDIOMETRY
  - BRAINSTEM RESPONSE
  - ELECTROPHYSIOLOGICAL
    - EEG/EMG/SCL/HR/RR
  - PSYCHOACOUSTICS

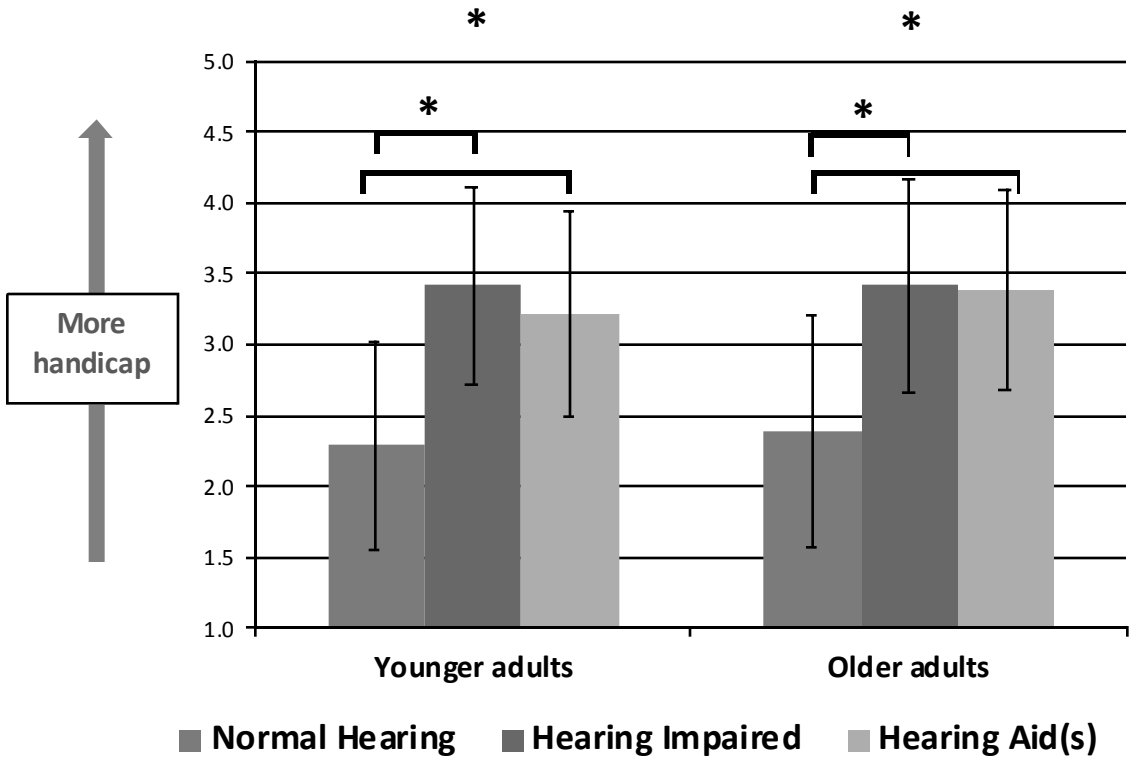


EEG Suite 1 | EEG Suite 2

Question: Do HI listeners & users of HAs report handicap when listening to emotion in speech?

Total N = 586	Normal Hearing (NH)	Hearing-impaired (HI) (do not own HAs)	Hearing aid(s) (HA)
Younger adults (18-59 years old)	n = 142 Mean age = 33.9 yrs	n = 115 Mean age = 32.9 yrs	n = 67 Mean age = 29.8 yrs
Older adults (60+ years old)	n = 101 Mean age = 63.7 yrs	n = 78 Mean age = 64.4 yrs	n = 83 Mean age = 64.8 yrs

Question: Do HI listeners & users of HAs report handicap when listening to (auditory) affective speech?

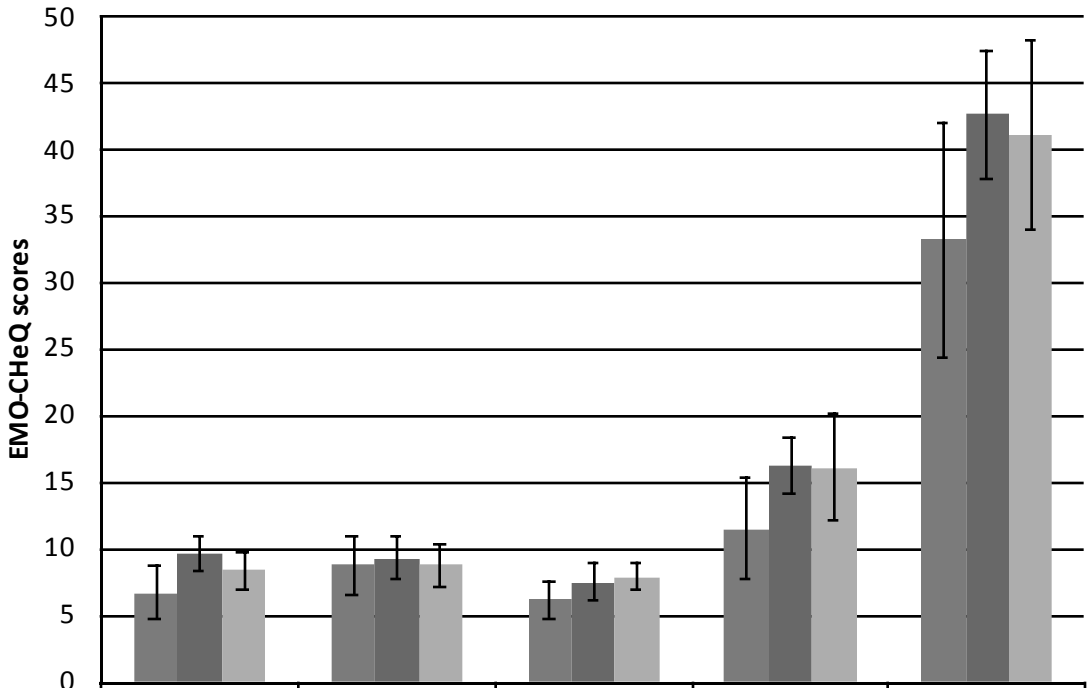


Preliminary Lab Results: Emo-Cheq

More

Impairment

Less



	Speaker Characteristics	Production	Situational Factors	Wellbeing	Total
NH	6.727272727	8.818181818	6.181818182	11.54545455	33.27272727
HA	9.6	9.3	7.5	16.3	42.7
HI	8.363636364	8.818181818	7.909090909	16.09090909	41.18181818

Error bars represent one Standard Deviation

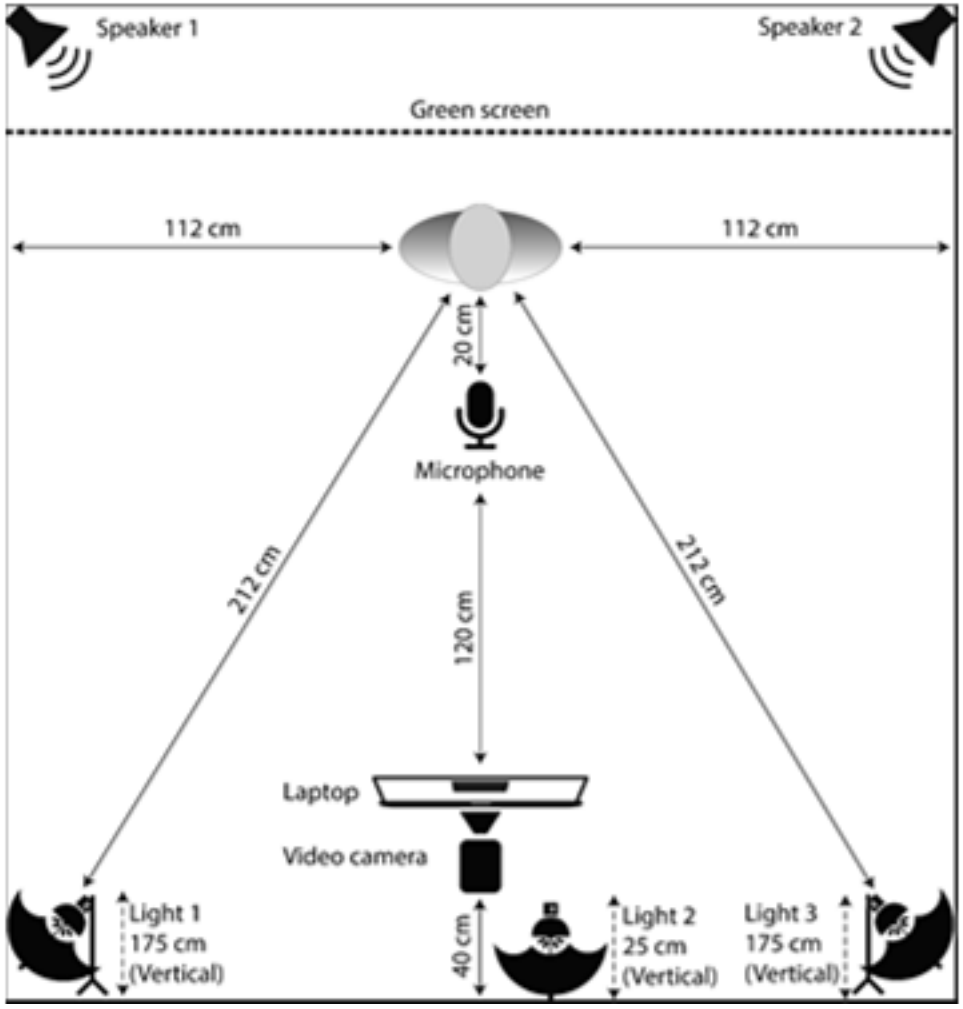
RAVDESS: Ryerson Audio-Visual Database of Emotional Speech and Song

- Collection of 7356 recordings made from 24 actors (12 male, 12 female) speaking and singing with various emotions:
  - Neutral, calm, happy, sad, angry, fearful, surprise, disgust
  - 2 levels of intensity: Strong, weak

- Audio, visual and audio-visual exemplars have been validated in a perceptual experiment involving 297 younger-adult and 48 older-adult participants

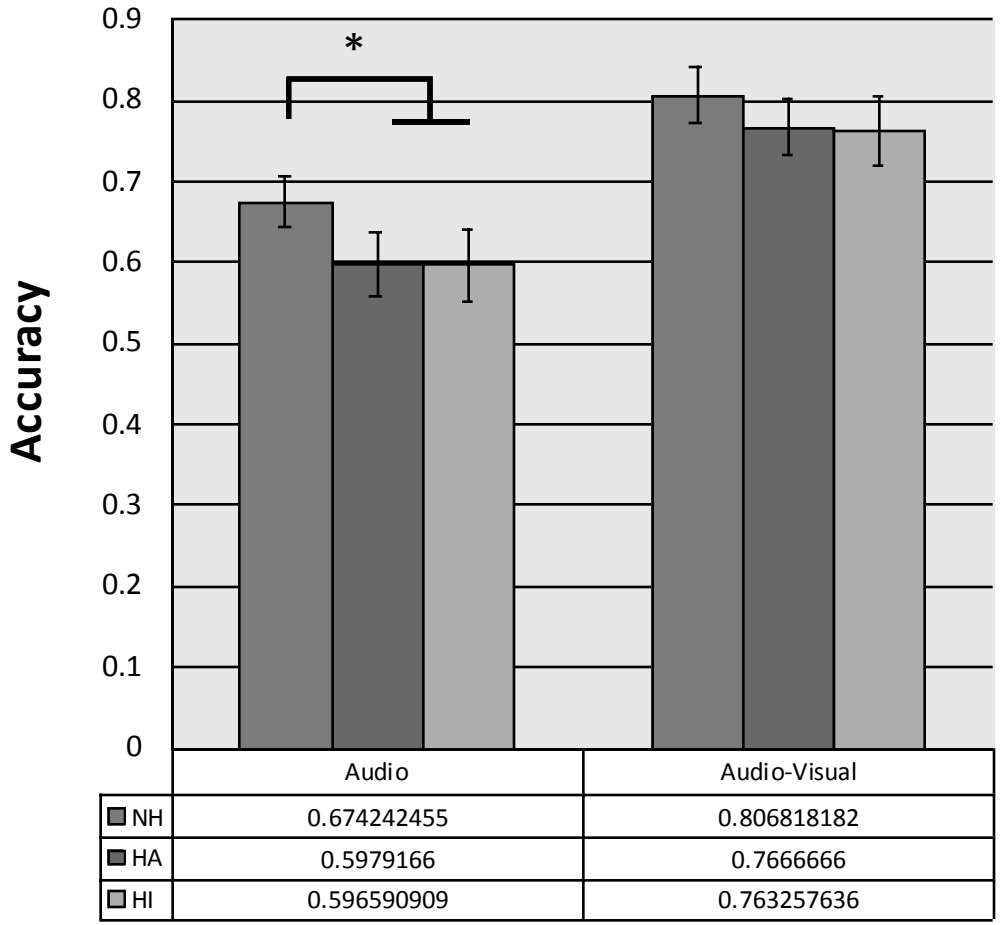
- Freely available for scientific-research and general use under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

Recording Apparatus



RAVDESS Exemplars

Emotional Speech Identification



	Audio	Audio-Visual
NH	0.674242455	0.806818182
HA	0.5979166	0.7666666
HI	0.596590909	0.763257636

- One-Way ANOVA revealed a significant main effect of group on accuracy ratings, F=3.585, p=.041.
- Planned comparisons showed a significant difference between the normal hearing group and the hearing impaired unaided and hearing aided group in the audio only condition, t(29)=-2.675, p=.012.
- No significant effect was observed in the Audio-visual condition

Physiological Responses to Emotion in Speech

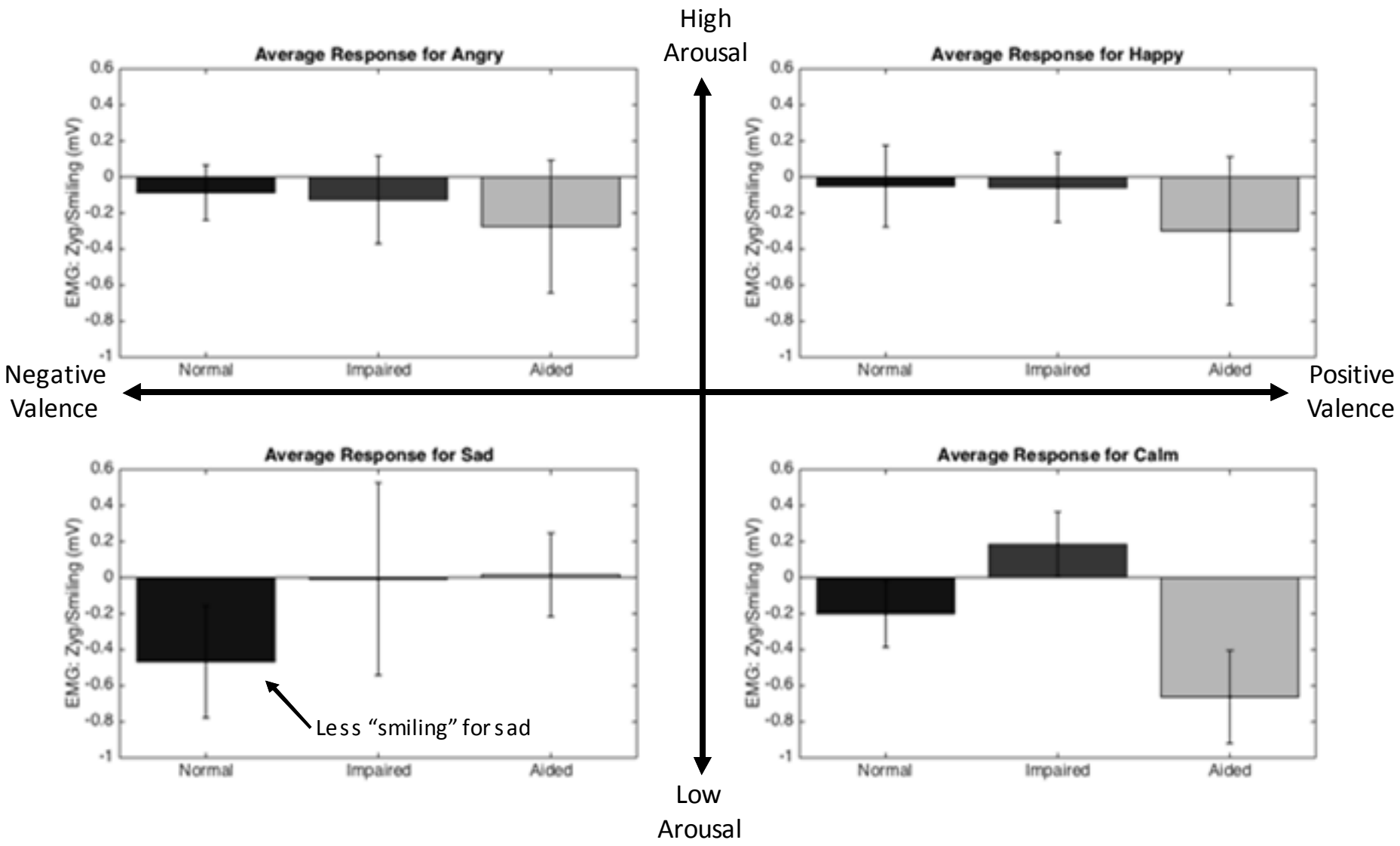
- There is a characteristic physiological response to emotion such as emotional speech
  - e.g. skin conductance, facial muscle activity (explained later)

- How is this physiological response affected by hearing loss and aids?

Electromyography (EMG)

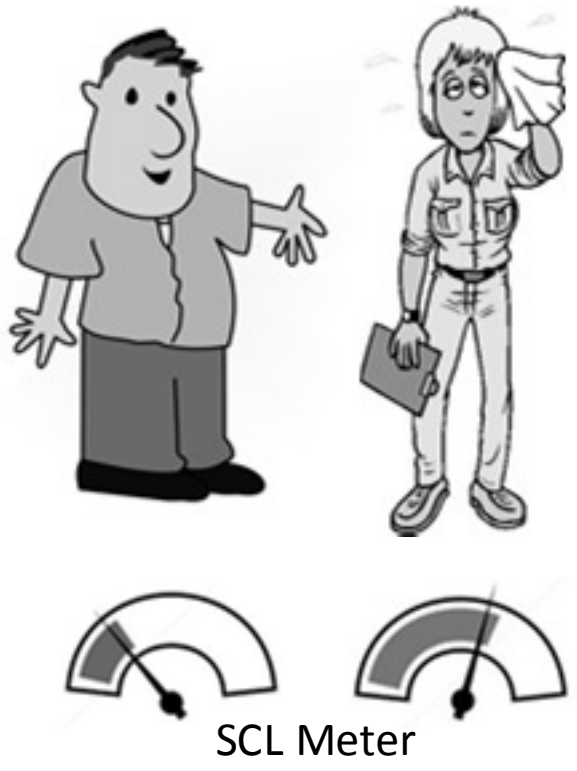
- Corrugator Supercilli muscle ("frowning")
  - Measured with electrodes above the eyebrow
- Zygomaticus Major muscle ("smiling")
  - Measured with electrodes on the cheek
- Activity in these muscles tends to mimic the emotional valence of a stimulus
- Not measuring overt movement; but rat baseline activity (presumably in preparation for movement)

“Smiling” Response



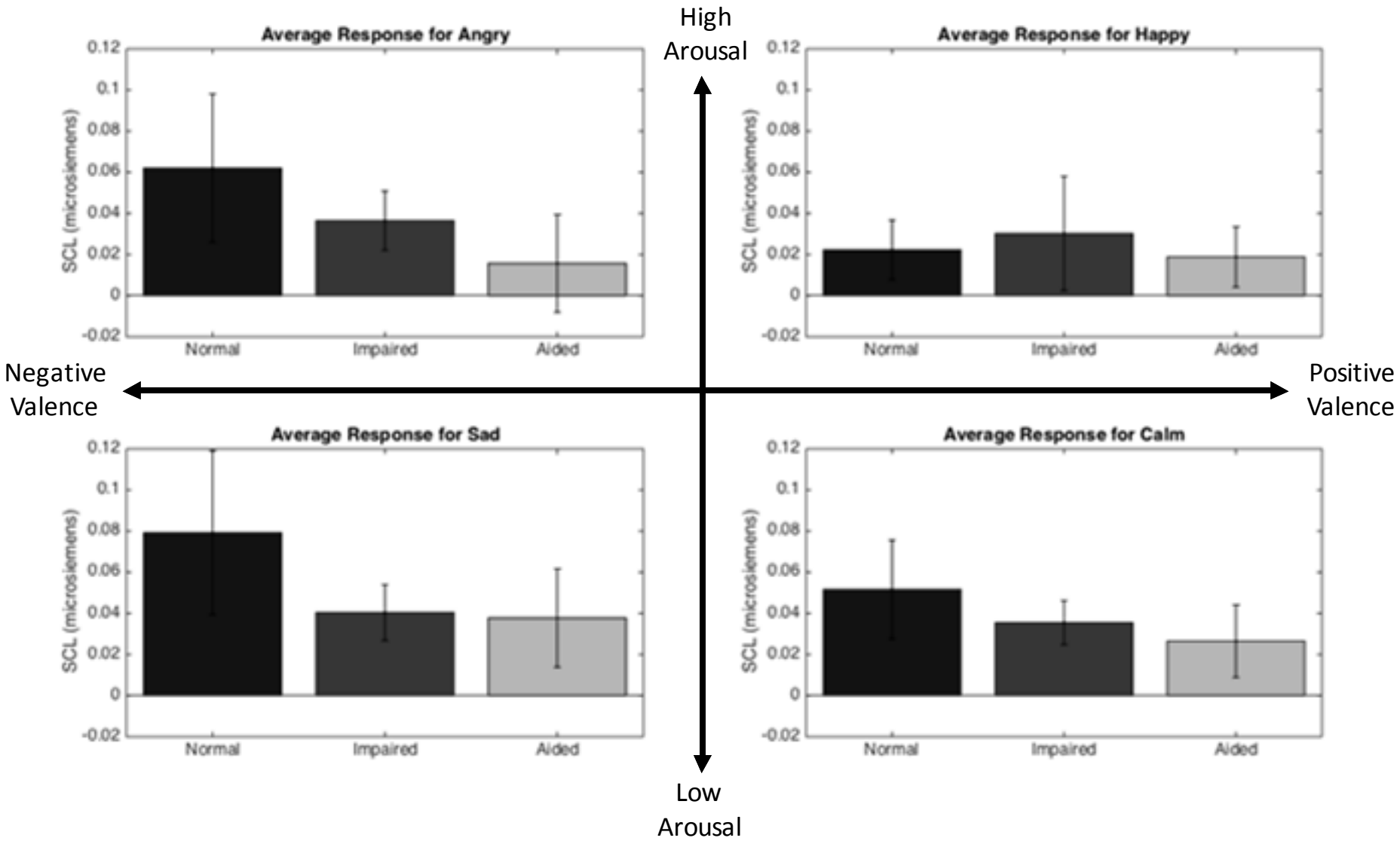
Skin Conductance Level (SCL)

- Sympathetic nervous system activity increases sweat gland secretions
- SCL measures changes in the electrical conductance of the skin due to sweat
- SCL increases are seen in response to emotional stimuli

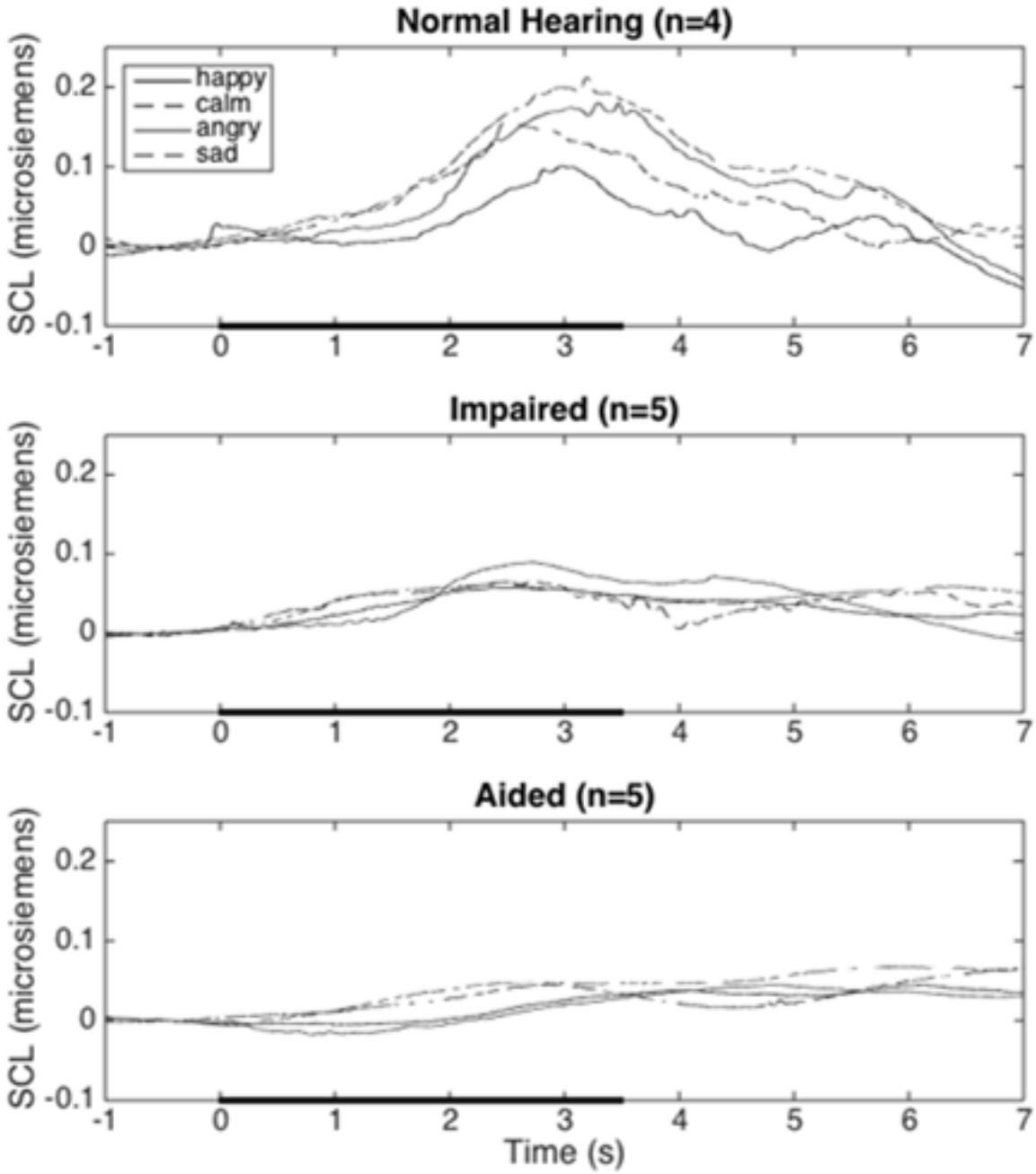


SCL Meter

SCL Average Response



SCL Time Course



→ NH show a characteristic SCL increase to emotional speech

→ HI show a dampened response to emotional speech

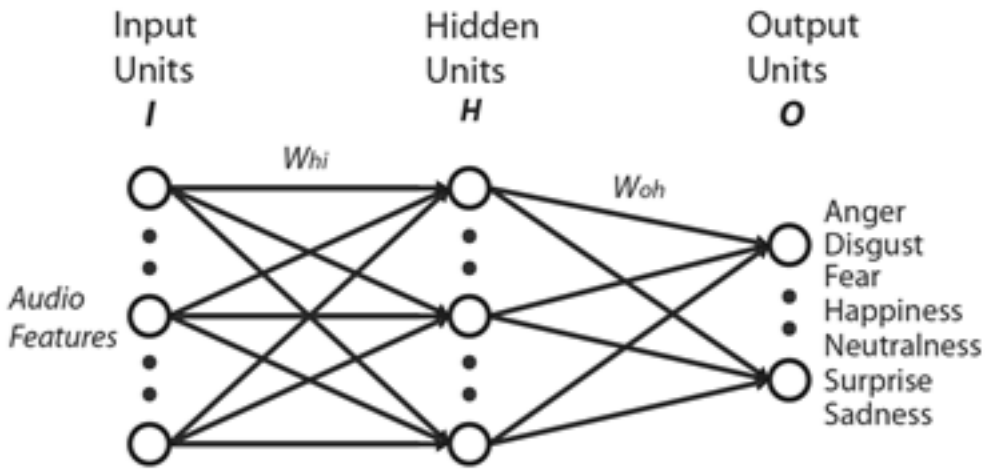
→ Using hearing aids does not recover the NH response to emotional speech

Acoustic Descriptions of Emotional Speech

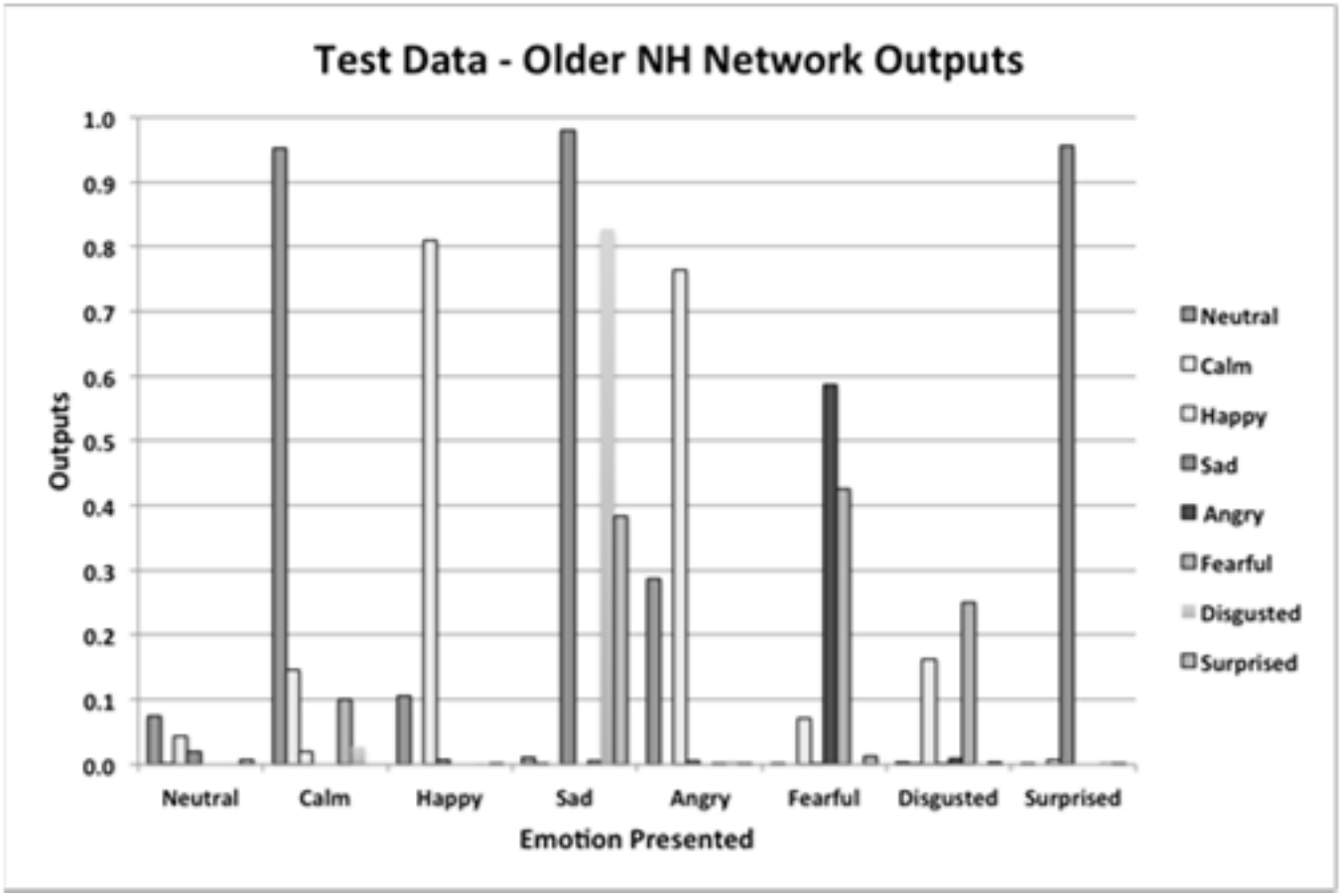
Feature		Description
Fundamental frequency (f0)		Strongly determines the perceived pitch of the voice. Periodicity with which the vocal folds open and close across the glottis.
Vocal intensity (voInt)		Heard as loudness of the voice. Due to subglottal pressure and laryngeal tension (Daniloff, Schuckers, & Feth, 1980; Zemlin, 1968).
Duration		The amount of time an utterance lasts for (a permutation of speech rate)
Formant frequency		Spectral peaks in the sound spectrum that reflect the resonance characteristics of the vocal tract and articulatory organs (Fant, 1960).
Formant bandwidth		Width of the spectral band containing significant formant energy (Scherer, 1989).
High spectral energy (HF500)		Distribution of energy in the higher-frequency range (Bezooijen, 1984; Scherer, 1989)
Harmonics to noise ratio (HNR)		Ratio of energy contained in harmonics (integer multiples of the fundamental frequency) to energy which is not (noise) (Yumoto, Gould, & Baer, 1982)
Jitter		Small perturbations in fundamental frequency caused by variations in the glottal pressure cycle (Lieberman, 1961; Scherer, 1989)
Shimmer		Perception of roughness or breathiness in the voice (Dejonckere et al., 1996) caused by rapid perturbations in amplitude during the glottal pressure cycle (Horii, 1980)

Developing a Computational Model

- Mapping acoustic features on to emotion judgments for older listeners



Preliminary Results

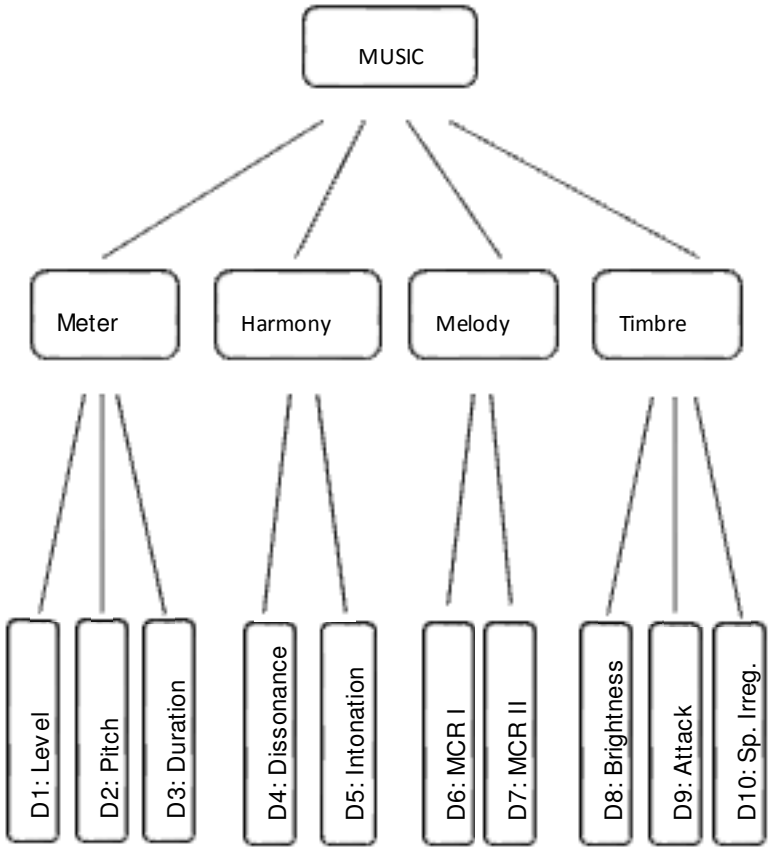


Long-term Modeling Goals

- Can the model be manipulated to resemble emotion recognition in older adults with hearing impairments?
  - applying low pass filtering and jittering

- Finally, can the features be adjusted (simulation of hearing aids) to optimize performance

The Adaptive Music Perception (AMP) Test



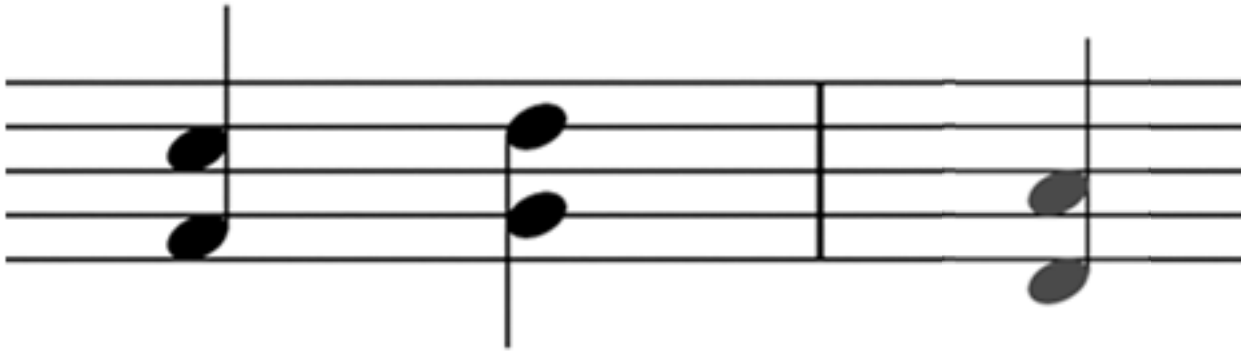
Meter

Level

Pitch

Length

Harmony

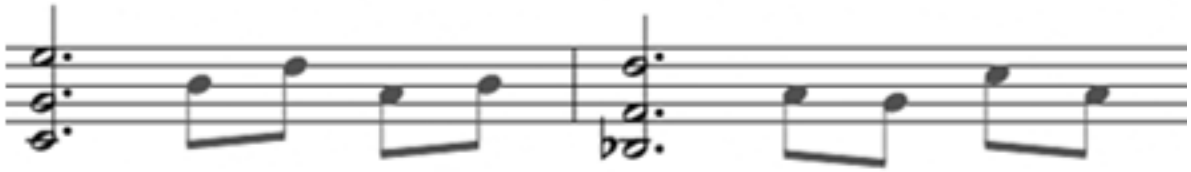


SUB DOM TON

Tuning

Pitch

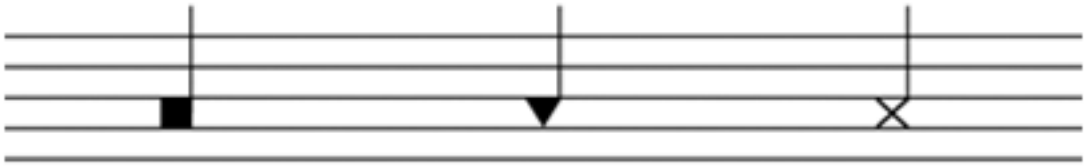
Melody



SNR I

SNR II

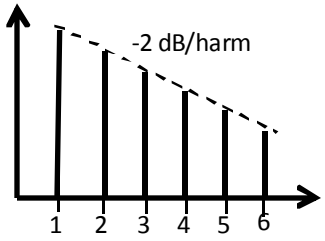
Timbre

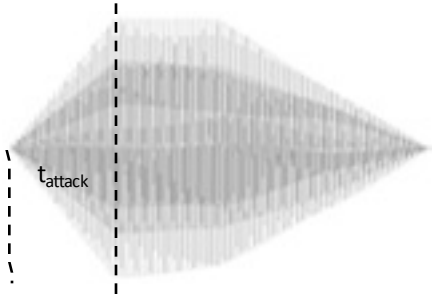


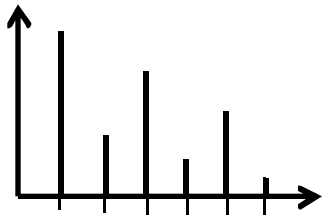
Brightness

Attack

Spec. Irreg.



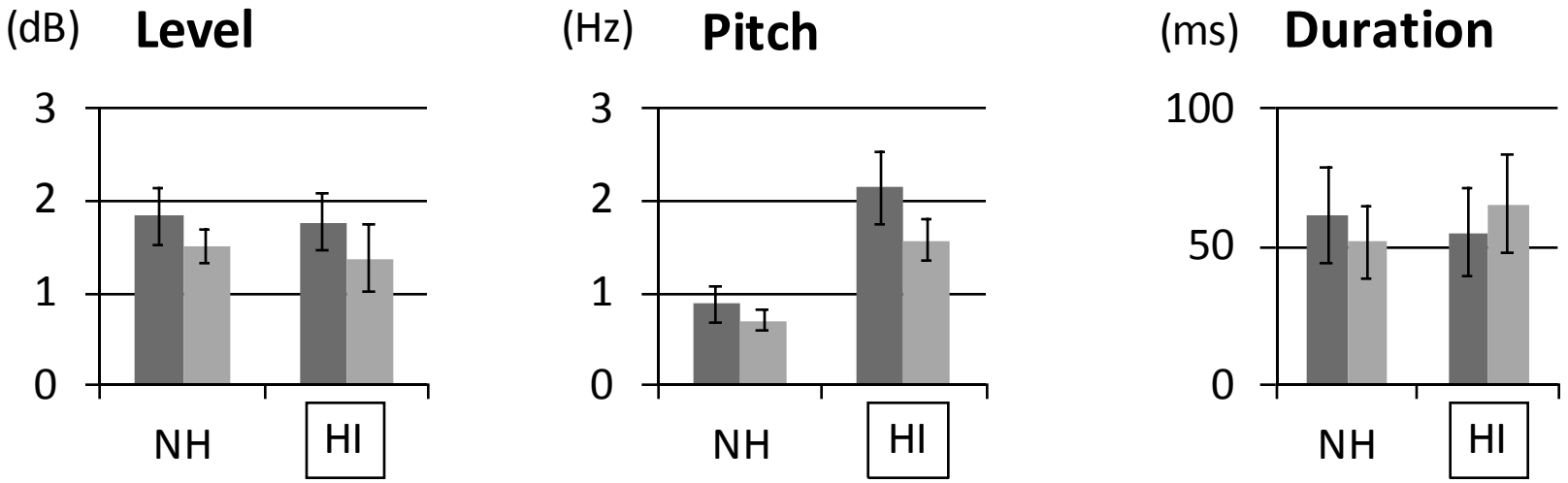




07/12/2015

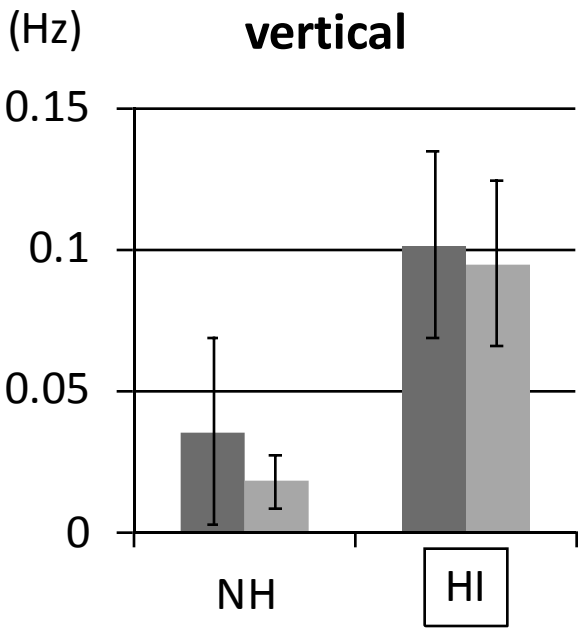
Using AMP to assess music perception in normal and hearing-impaired listeners

Meter

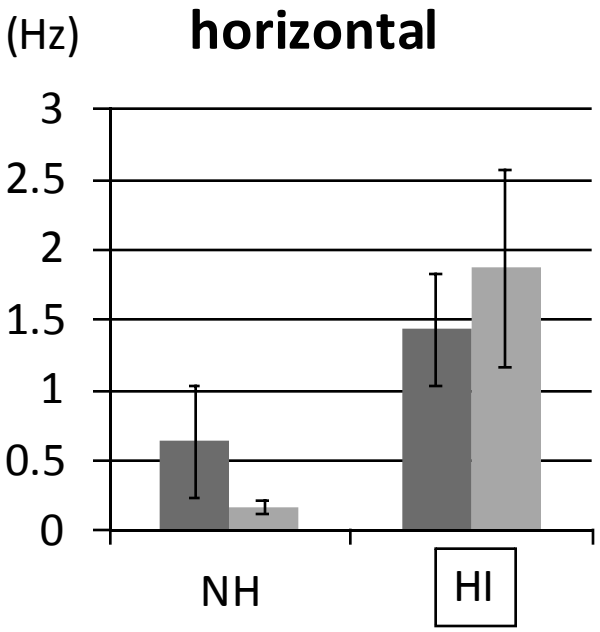


F(22) = 9.867, p<0.01

Harmony

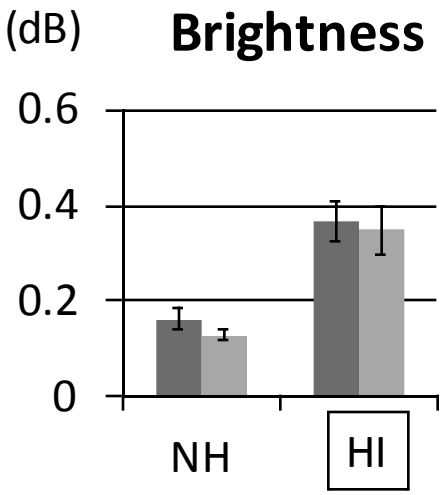


F(16) = 6.80, p<0.05

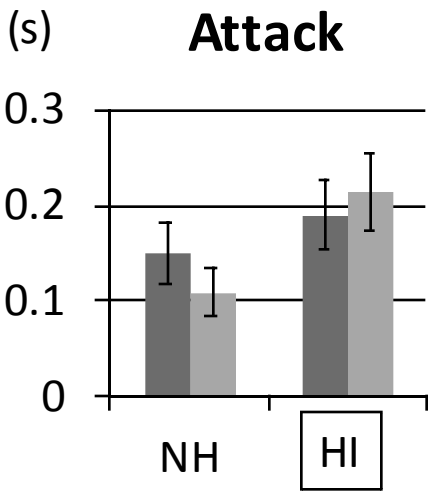


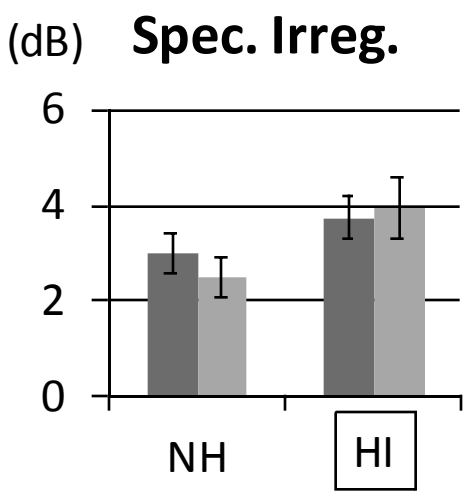
F(16) = 6.78, p<0.05

Timbre



F(22) = 14.44, p<0.01





Using AMP to assess music perception in aided and unaided hearing-impaired listeners

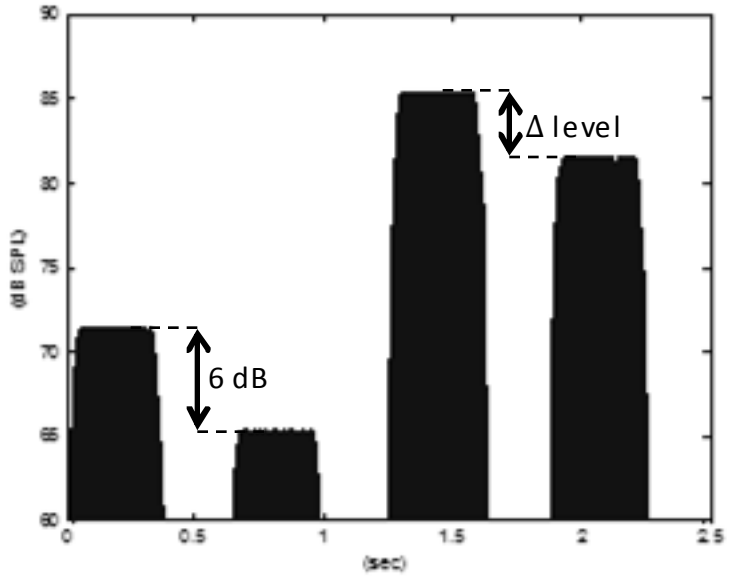
Results

		mean discrimination threshold		ANOVA	
subtest	**dimension**	**unaided**	**aided**	**F**	**P**
meter	**D1 (dB)**	1.95	2.44	3.94	0.059
	D2 (Hz)	2.89	2.94	0.05	0.826
	D3 (msec)	55.3	55.5	0.01	0.973
harmony	**D4 (Hz)**	0.213	0.245	0.55	0.468
	D5 (Hz)	4.64	3.65	1.95	0.178
timbre	**D8 (dB)**	0.417	0.415	0.01	0.972
	D9 (msec)	213	275	3.79	0.063
	D10 (dB)	5.33	6.01	2	0.17

Follow-up analysis

- Is dynamic compression responsible for the elevated thresholds?

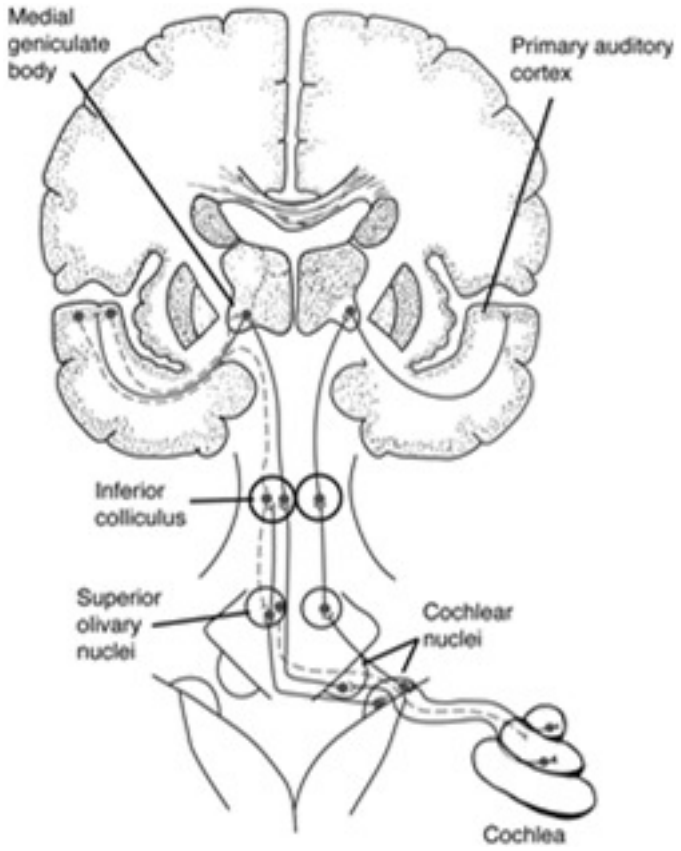
$$\boldsymbol{CF = 6dB - \Delta level_{in-situ}}$$

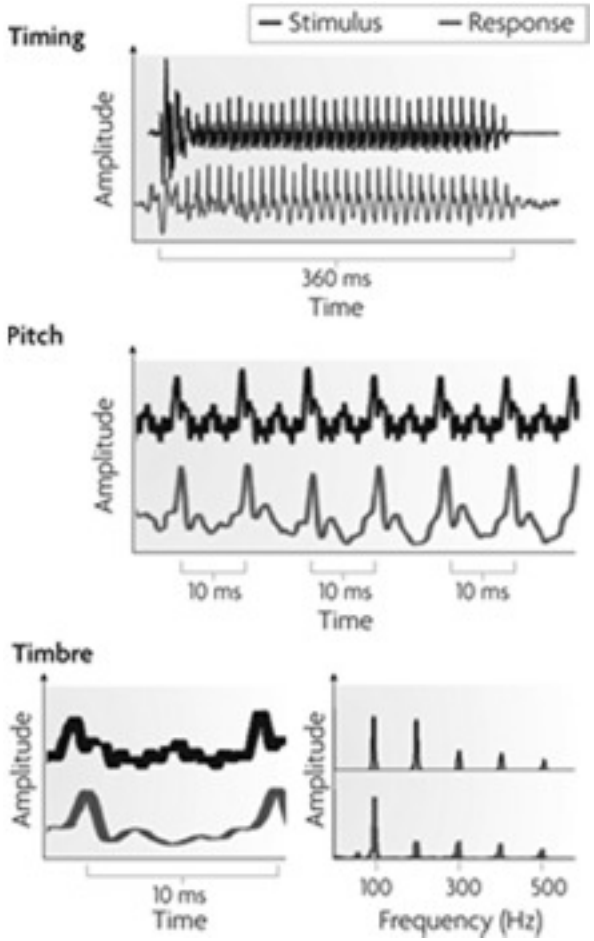


Results

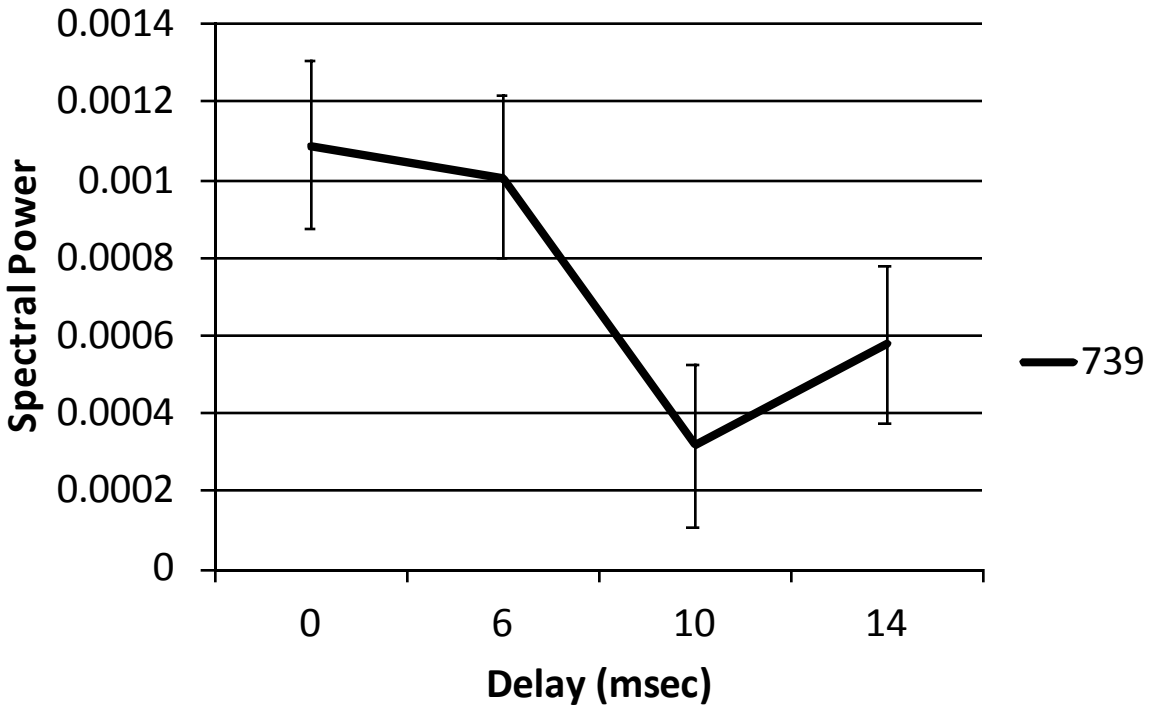
- D1: level r=0.190, p=0.363
- D9: attack r=0.453, p=0.023

Frequency Following Response

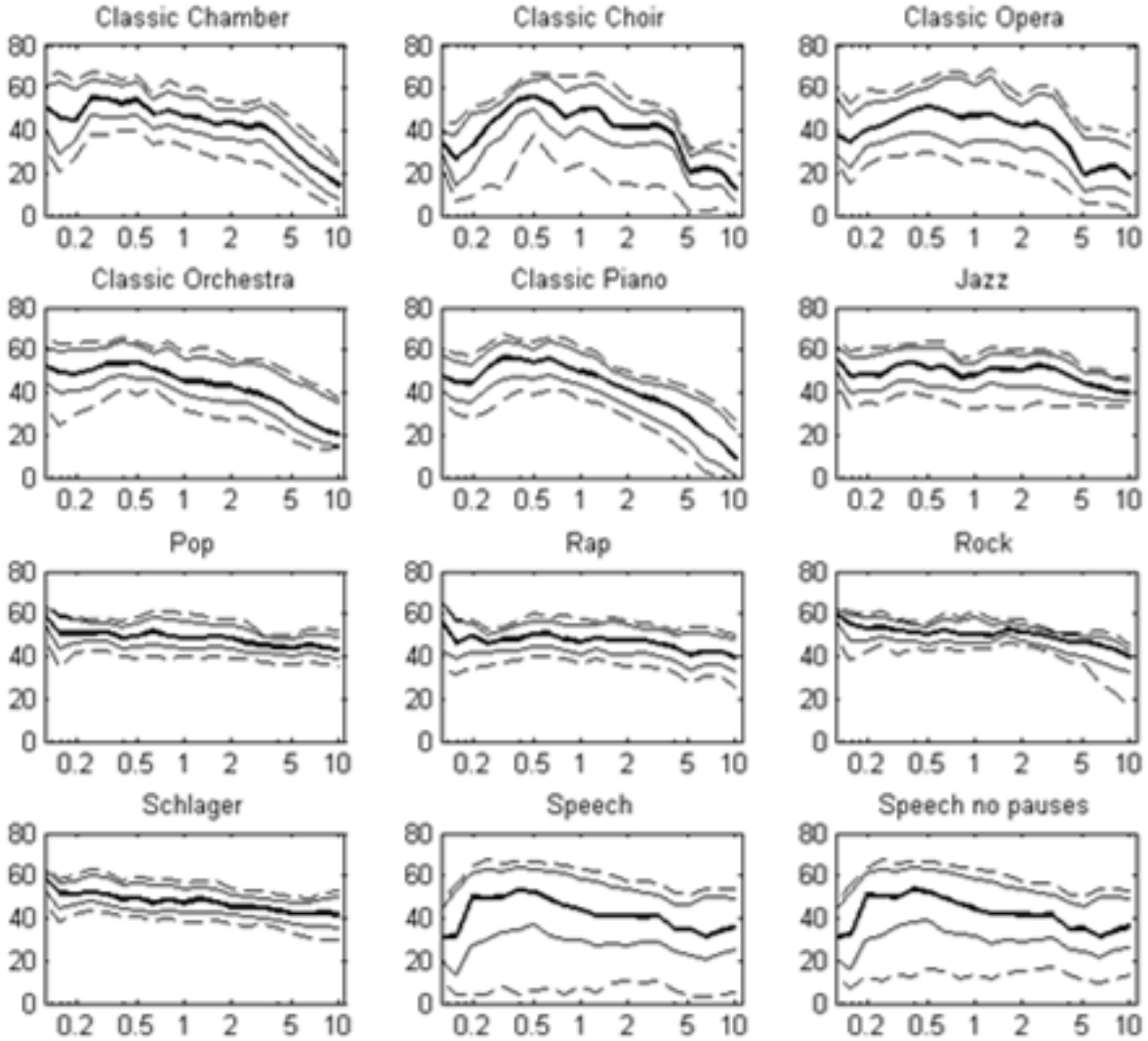




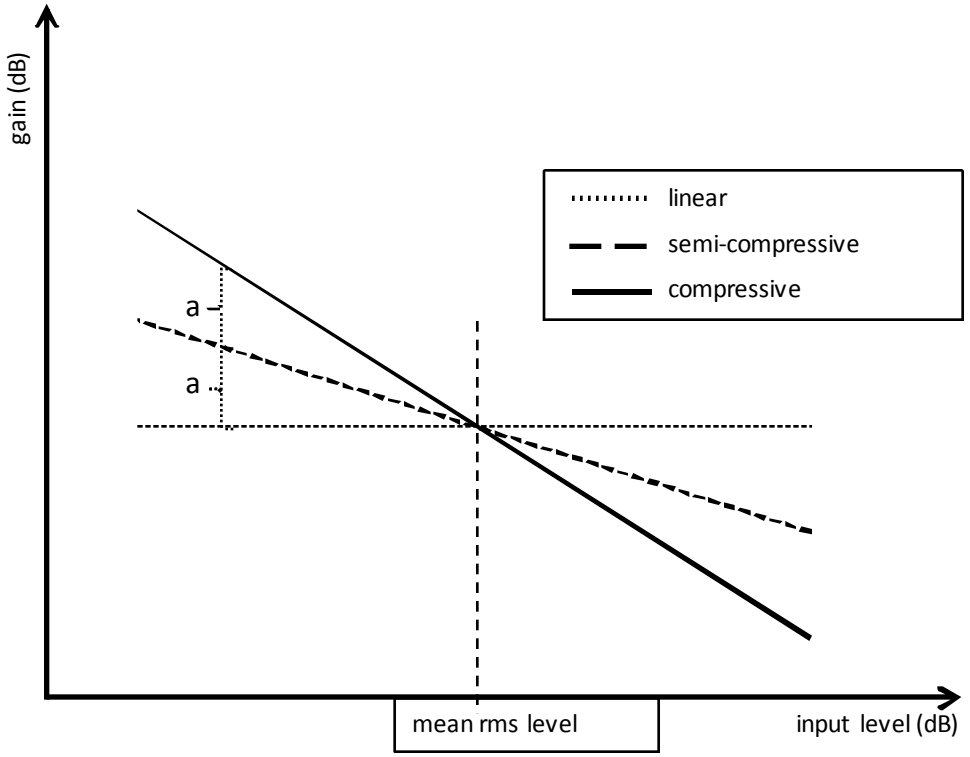
FFR as a function of HA Delay



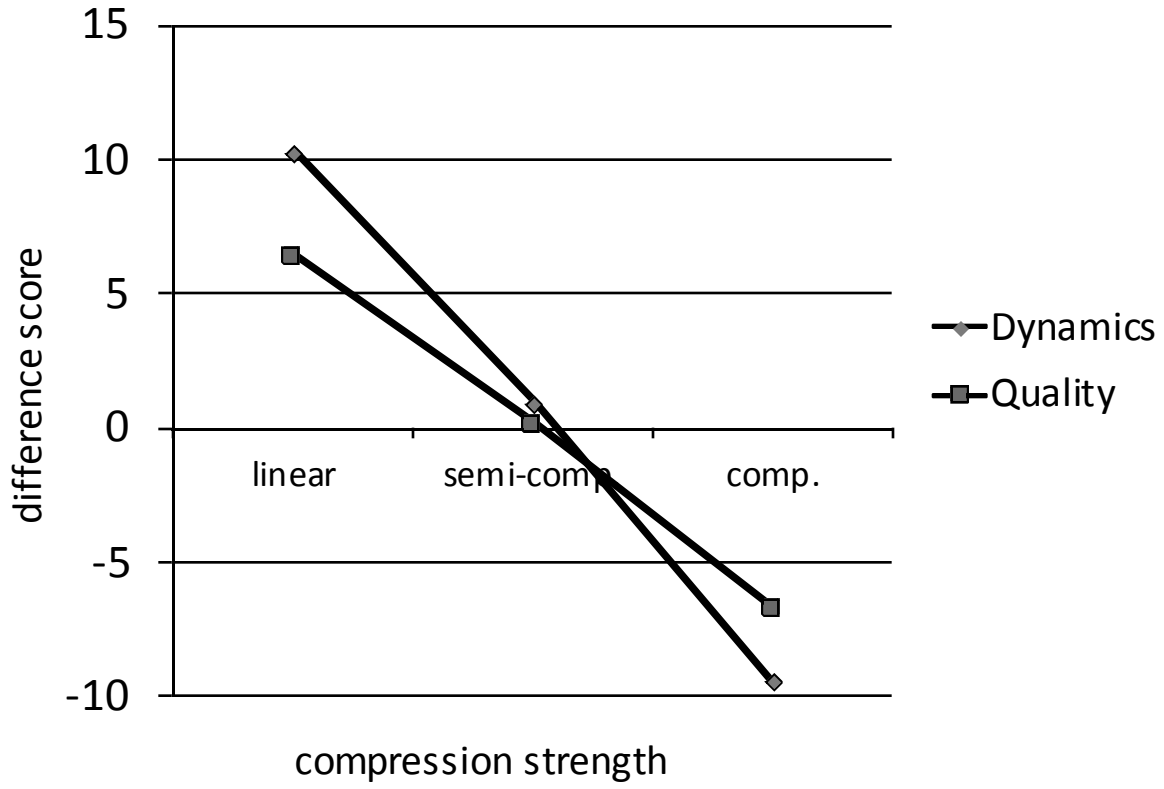
Compression in Music and HA



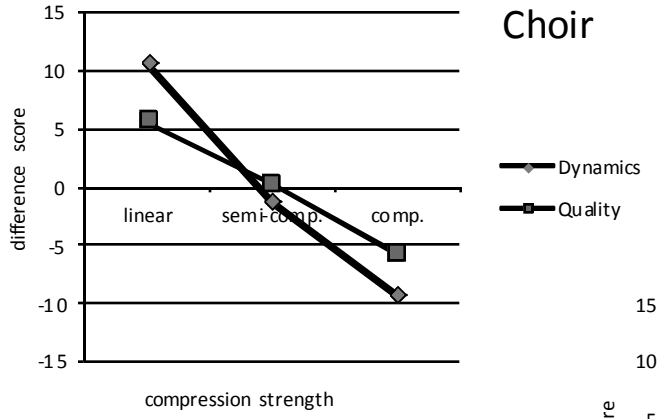
Levels of Compression

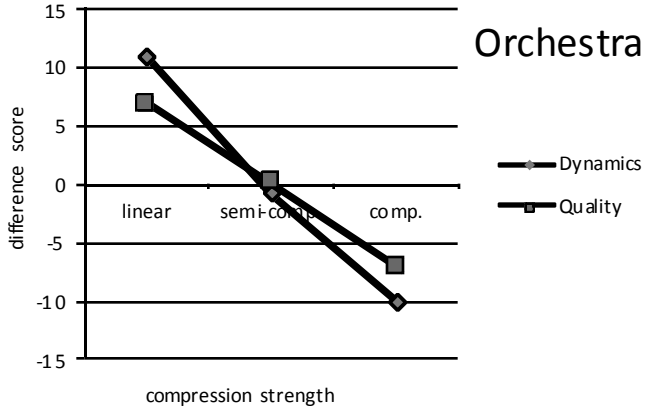


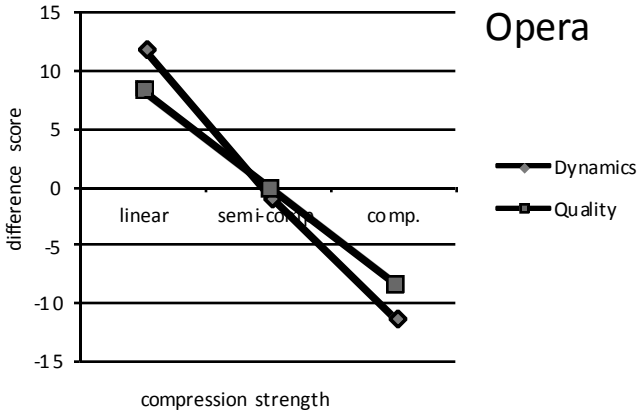
Overall Results

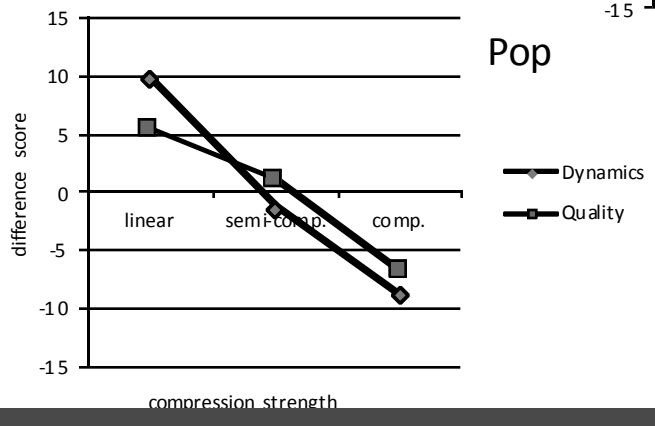


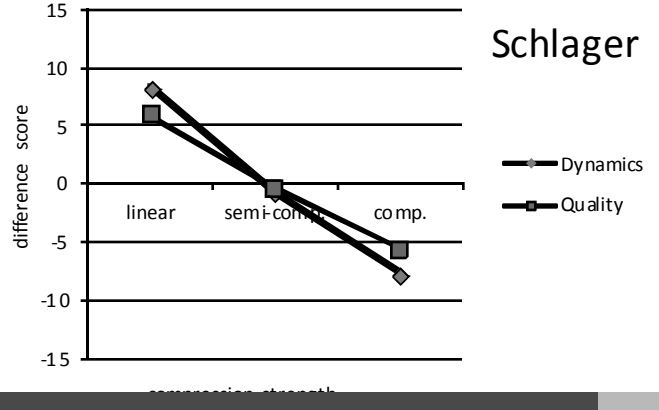
By Genre


Choir
15
10
5
0
-5
-10
-15
difference score
linear
semi-comp.
comp.
compression strength
Dynamics
Quality


Orchestra
15
10
5
0
-5
-10
-15
difference score
linear
semi-comp.
comp.
compression strength
Dynamics
Quality

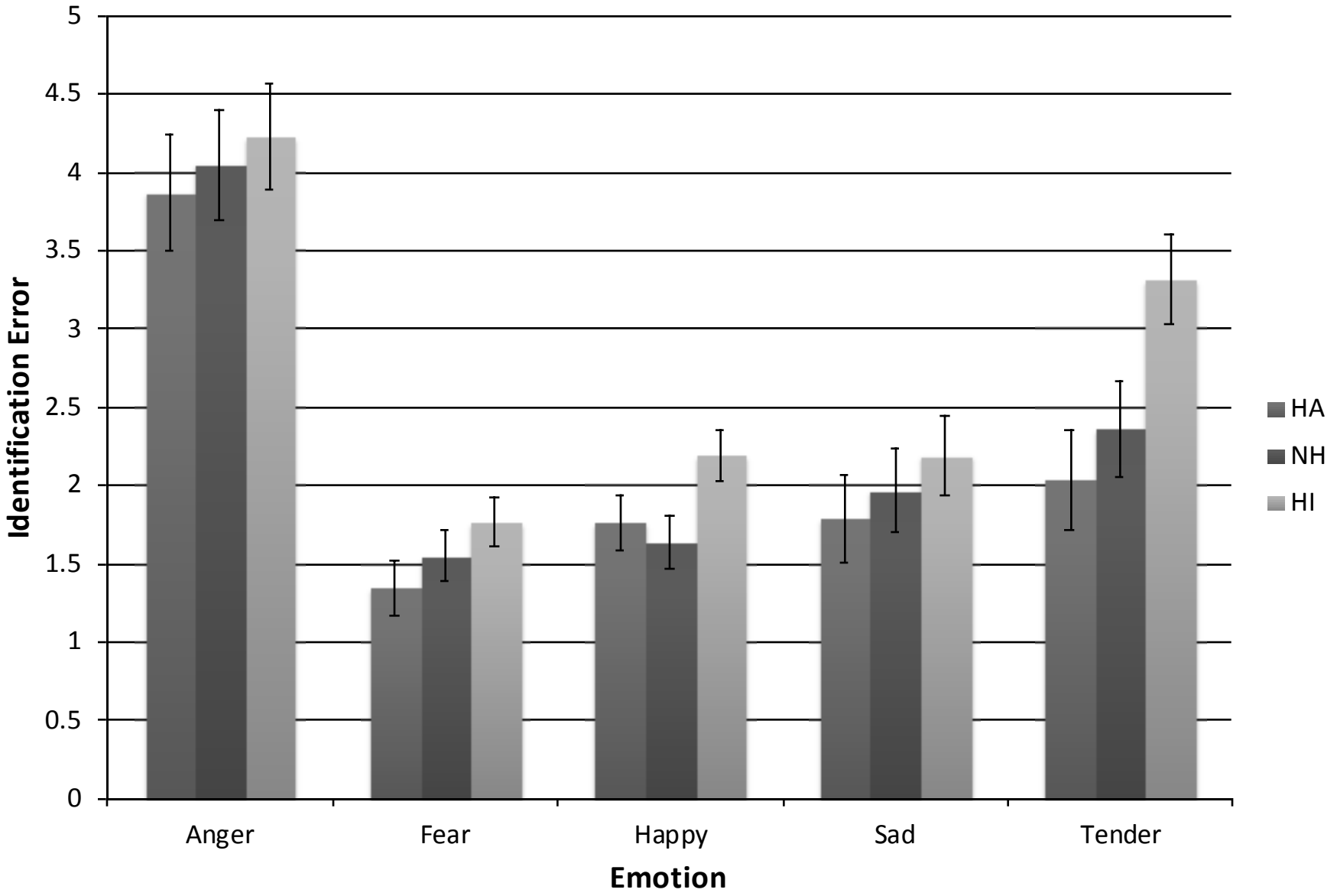

Opera
15
10
5
0
-5
-10
-15
difference score
linear
semi-comp.
comp.
compression strength
Dynamics
Quality


Pop
15
10
5
0
-5
-10
-15
difference score
linear
semi-comp.
comp.
compression strength
Dynamics
Quality


Schlager
15
10
5
0
-5
-10
-15
difference score
linear
semi-comp.
comp.
Dynamics
Quality

Perception of Music Emotion in HA

- Assessed ability to identify emotion in film music excerpts
- 3 Groups:
  - 10 HA (dB HL = 37.26)
  - 11 HI (avg dB HL = 35.51)
  - 10 NH (avg dB HL = 15.56)


5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Identification Error
Anger
Fear
Happy
Sad
Tender
Emotion
HA
NH
HI

Thank-you

Collaborators

Gurjit Singh
Martin Kirchberger
Steven Livingstone
Gabe Nespoli
Naresh Vempala
Lisa Liskovoi
Domenica Fanelli

Funding sources

Contact: russo@ryerson.ca
RAVDESS: smartlaboratory.org/RAVDESS

Follow-up tests

- We expect that pop music would require less compression; but we didn’t find this in the quality assessments
  - Consider doing AMP test as follow-up with different music genres