

**The Application of Cochlear
Audio Analysis Techniques to
Finding Percussion in Electroacoustic Music
(Human Judgement Collection)**

Anderson Mills

The University of Texas at Austin

153rd meeting of the Acoustical Society of America

Salt Lake City, Utah

presented 2007-06-05

Introduction

- Objective
- Background
 - Electroacoustic Music
 - Single, Damped, Percussive Event (SDPE)
- Human Judgement Collection
 - Stimuli
 - Judgement Collection
- Results
- Conclusions and Future Work

Objective

The Greater Goal

provide detailed, informative images corresponding to pieces of electroacoustic music (lack a standard visual representation)

My Goal

create algorithms which use models of human hearing to extract information about percussion in recorded electroacoustic music

Today's Goal

show results from a collection of human judgements about Single, Damped, Percussive Events (SDPE's)

Background

Electroacoustic Music

- definition — involves electronic technology for the compositional manipulation of sound
- requires that music is treated as pure sound

Background

Single, Damped, Percussive Event (SDPE)

- single sound event created by the impact of one object with another without either object breaking
 - a strike of a xylophone, a hand clap, or a ball bounced against a wall
- single sound event created by the direct introduction of a extremely sudden pressure change in the air
 - a balloon pop, a pistol shot, or a vocal plosive
- or any synthetic or electronically manipulated sound which is reminiscent of these
 - an electronic drum

Human Judgement Collection

Methodology

Stimuli

- the base stimulus sound is a synthesized snare drum (cSound)
- three sets of 16 stimulus sounds
- each stimulus set is organized in a 4x4 grid

dimension 2	13	14	15	16
	9	10	11	12
	5	6	7	8
	1	2	3	4
	dimension 1			

dimension 2	4	8	12	16
	3	7	11	15
	2	6	10	14
	1	5	9	13
	dimension 1			

- dimension 1 of the stimulus sounds is varied along one axis
- dimension 2 of the stimulus sounds is varied along the other axis

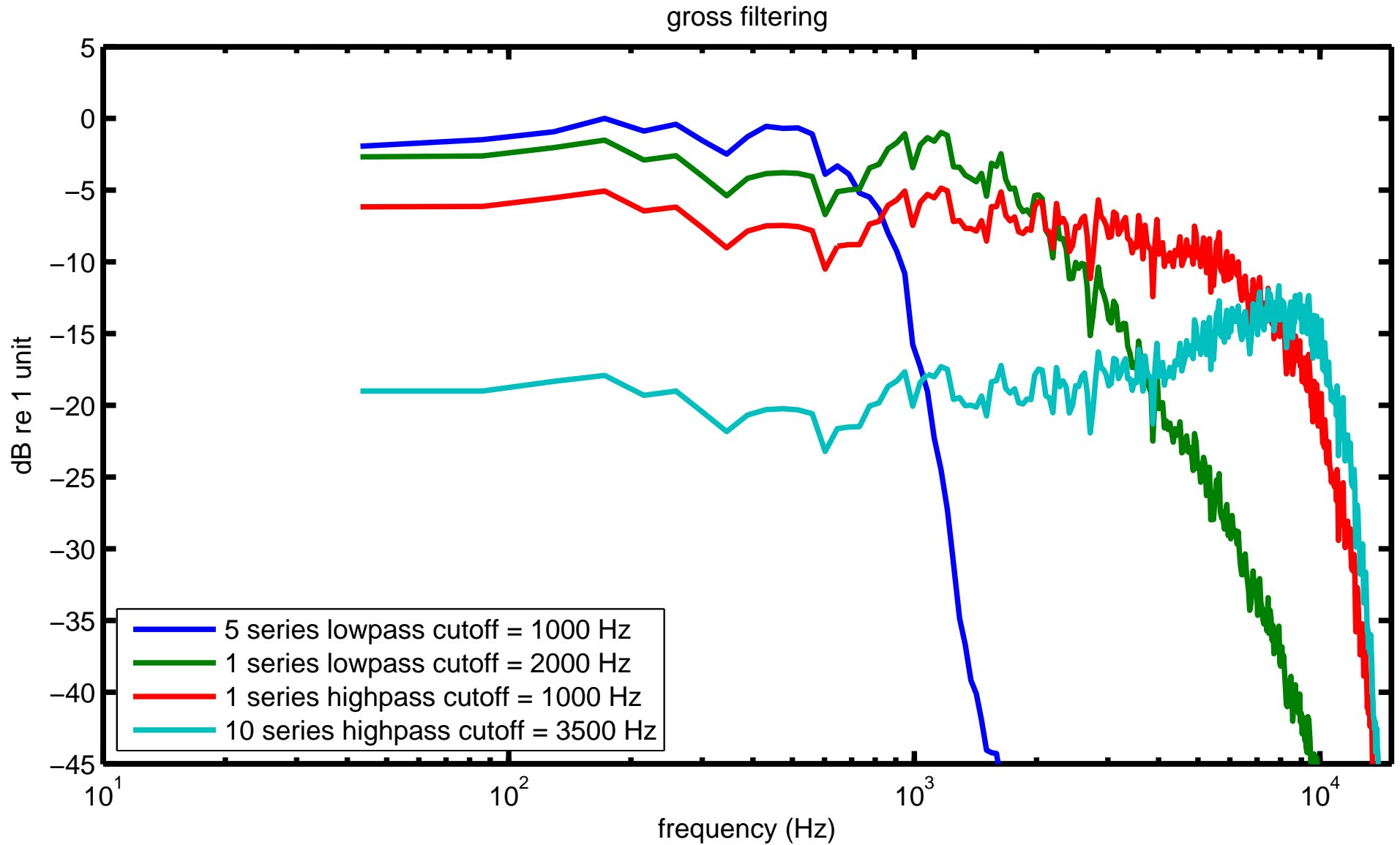
Human Judgement Collection

Stimuli

stimulus set	dimension 1	dimension 2
A	rise time (10–70 msec)	string resonance (20–80%)
B	gross filtering (low–high)	noisiness (20–80%)
C	rise time (10–70 msec)	gross filtering (low–high)

Human Judgement Collection

Gross Filtering



Human Judgement Collection

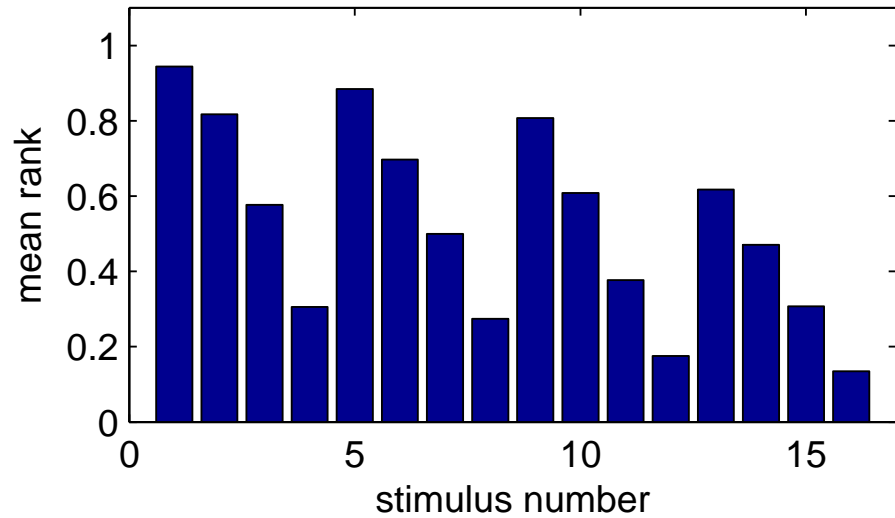
Judgement Collection

- each participant is presented the stimulus sets in random order
- within each set a participant is presented 120 stimulus pairs (+ 30 training pairs) in random order
- for each pair the participant chooses the stimulus which is “most like an SDPE” and specifies the difficulty of the judgement
- a round robin tournament algorithm is used to rank the 16 stimulus sounds
- the participant chooses an SDPE cutoff point in the ranked list

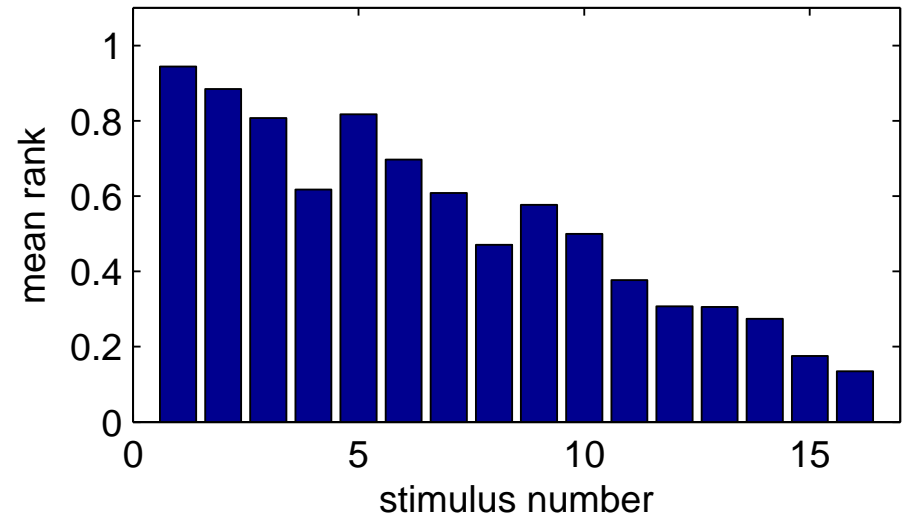
Results

Stimulus Set A, average

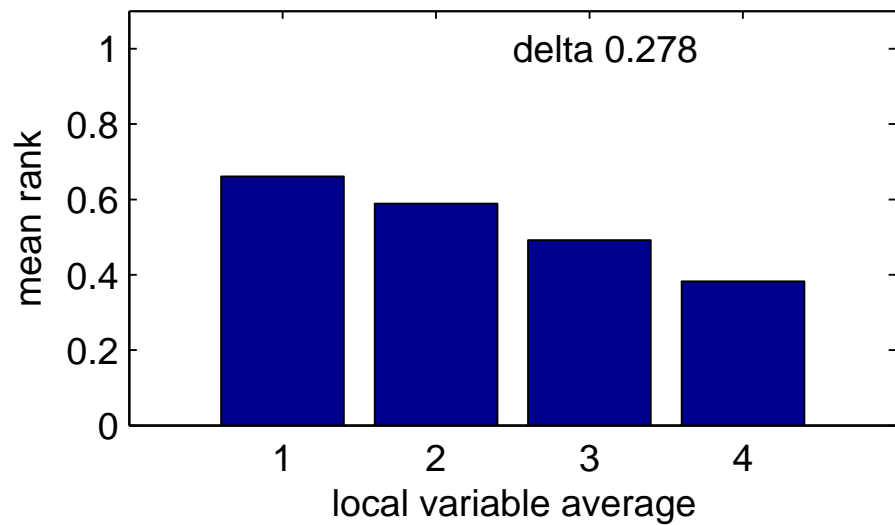
set A, global = string resonance



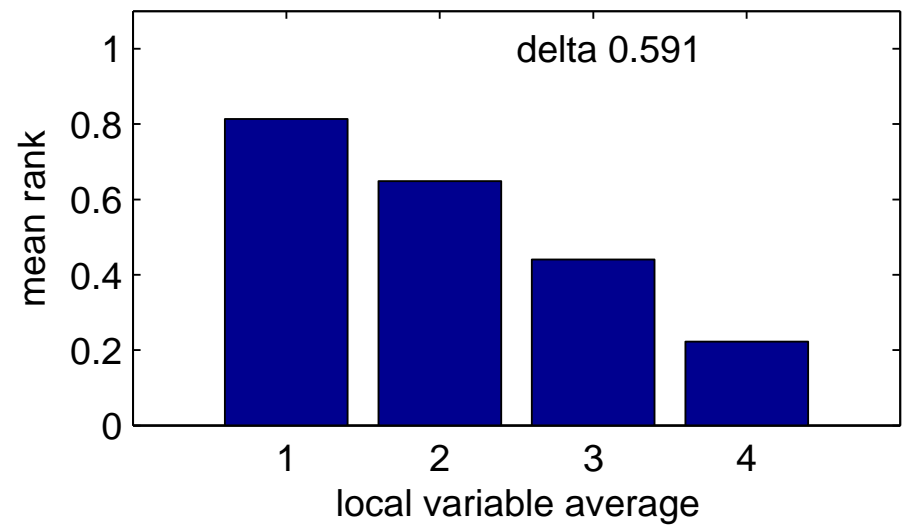
set A, global = rise time



delta 0.278



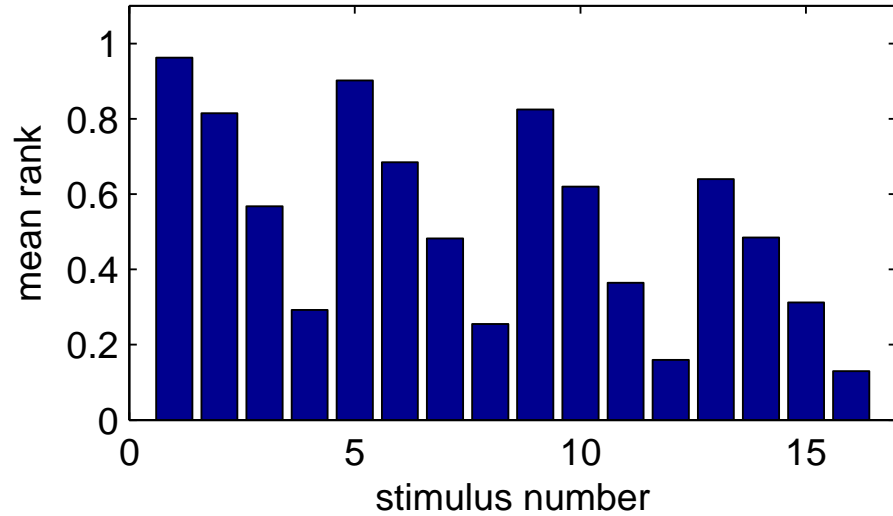
delta 0.591



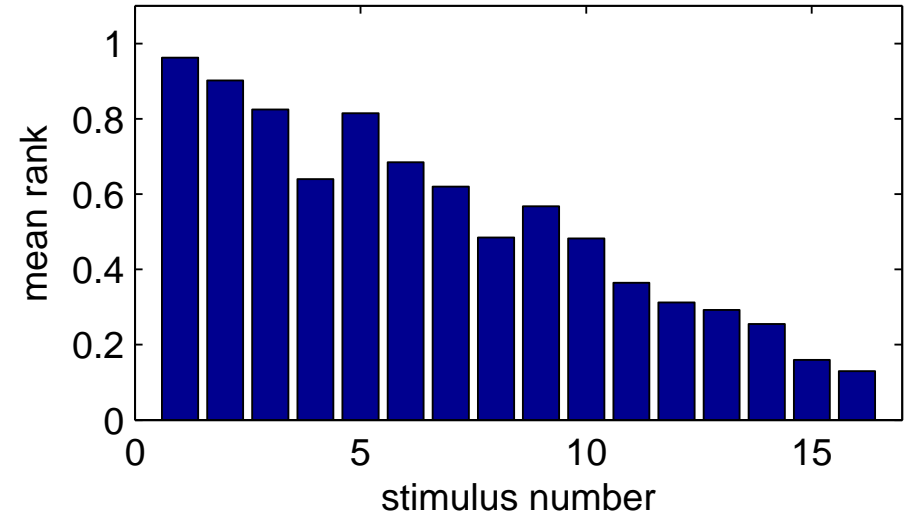
Results

Stimulus Set A, average with no outliers

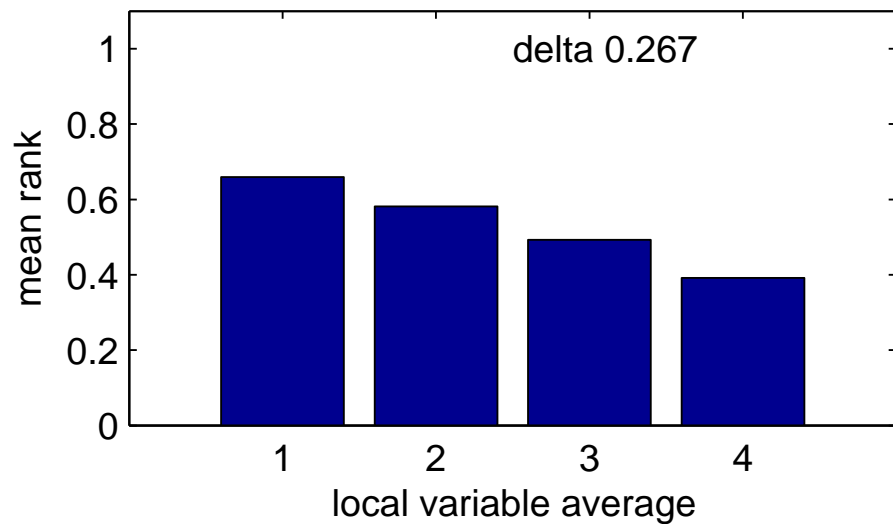
set A, global = string resonance



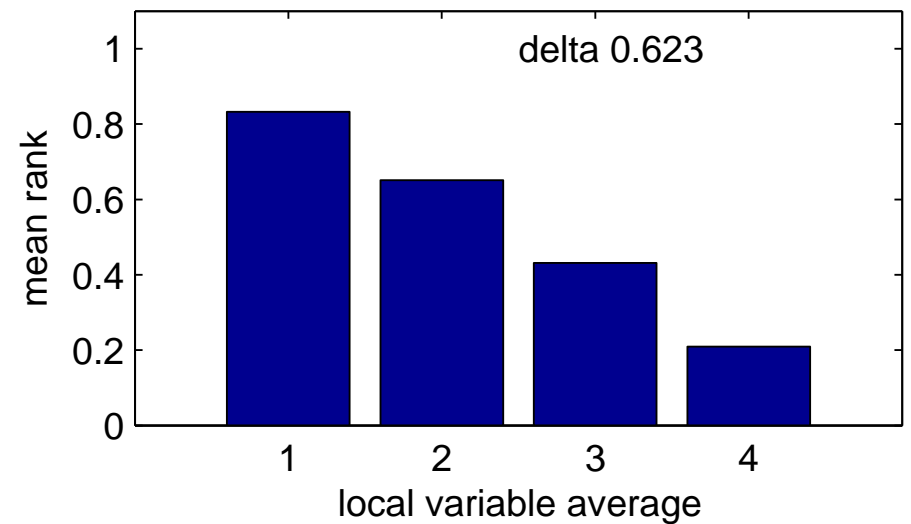
set A, global = rise time



delta 0.267



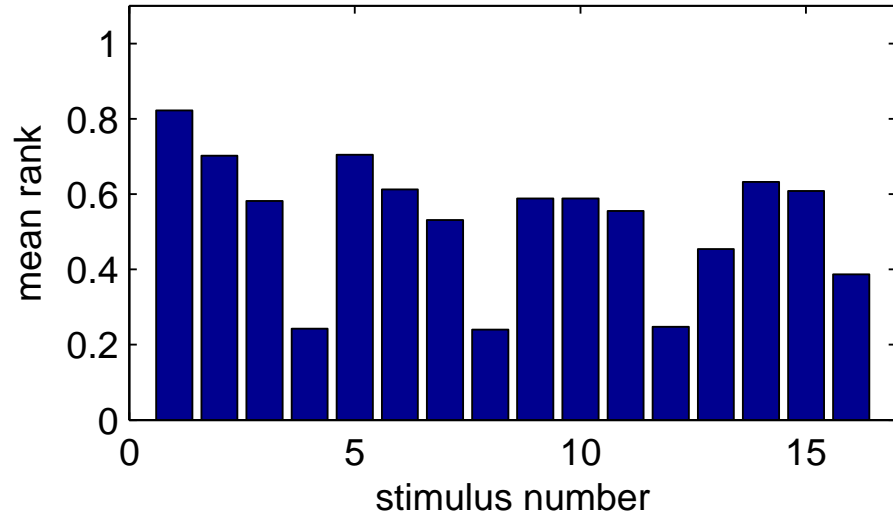
delta 0.623



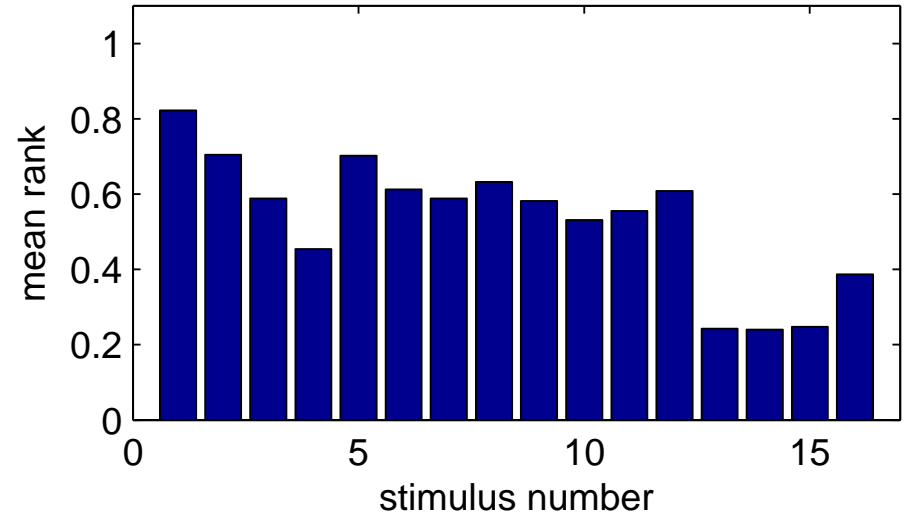
Results

Stimulus Set B, no outliers

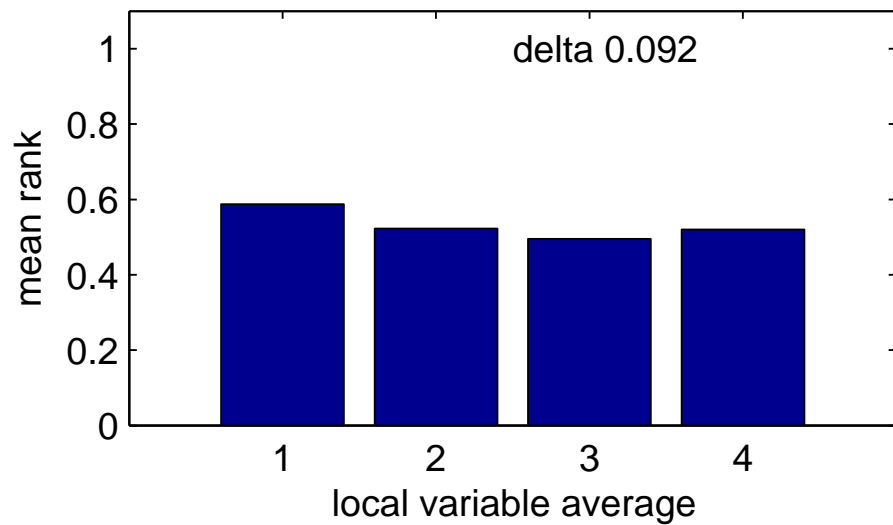
set B, global = noisiness



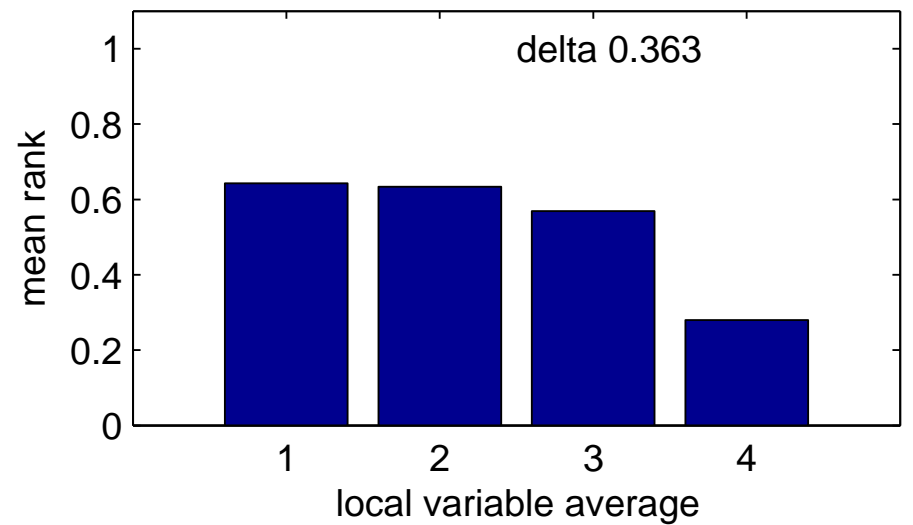
set B, global = gross filtering



delta 0.092



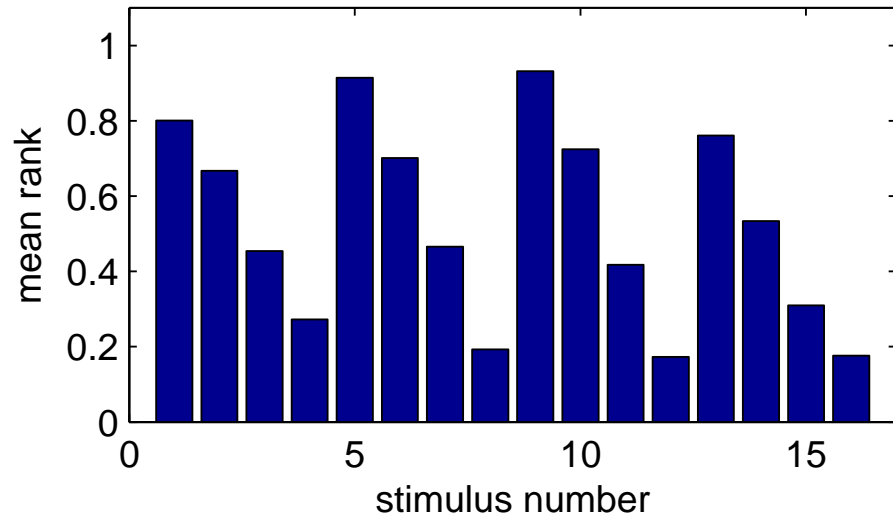
delta 0.363



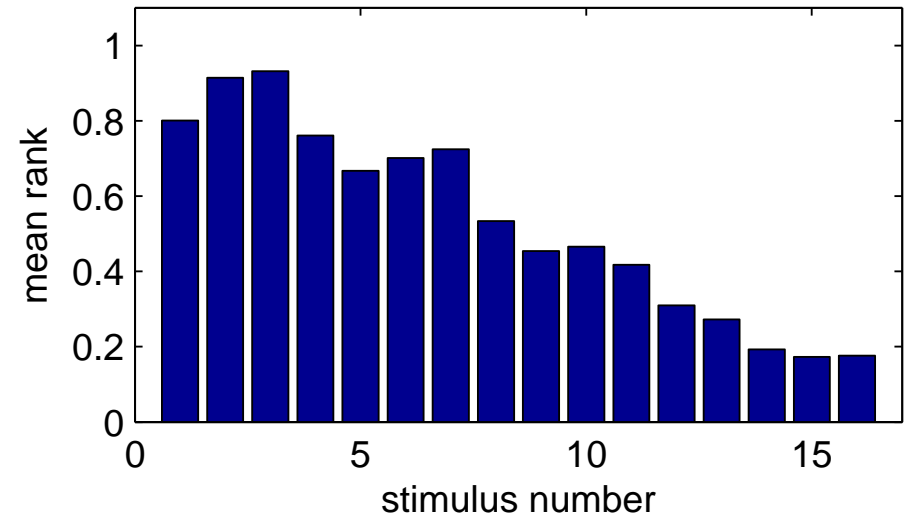
Results

Stimulus Set C, no outliers

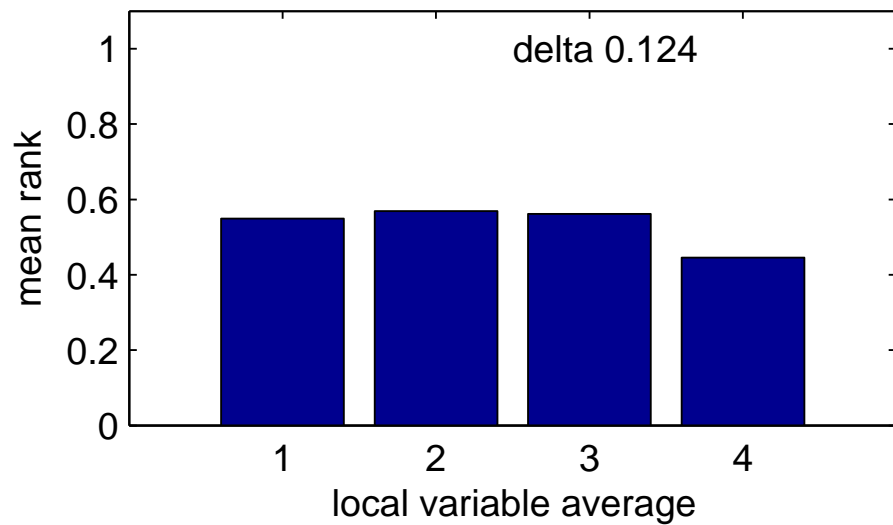
set C, global = gross filtering



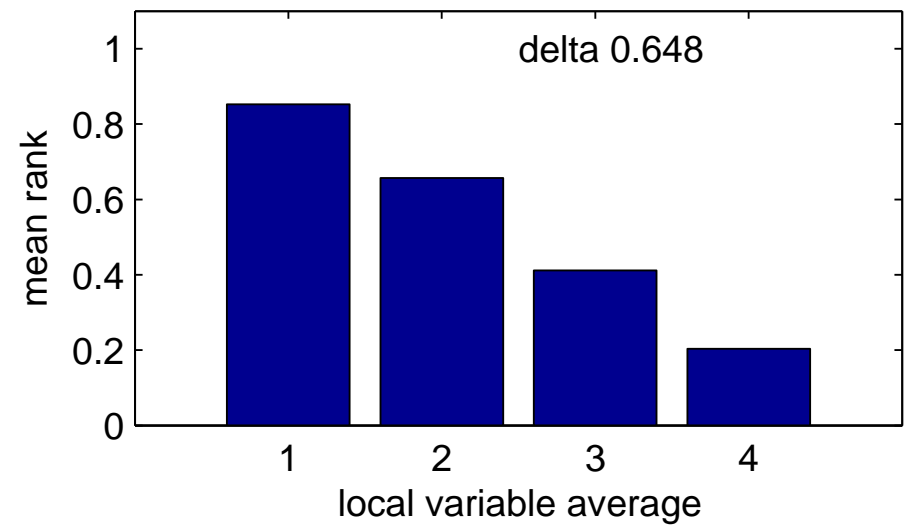
set C, global = rise time



delta 0.124



delta 0.648



Conclusions and Future Work

Conclusions

- rise time appears to be the most significant cue to a sound being percussive of the dimensions tested
- string resonance has a strong effect on the perception of a sound's percussiveness, but is not a primary dimension of the sound
- gross spectral filtering may be a secondary cue to a sound's percussiveness, but the effect appears to lessen in the presence of a strong cue

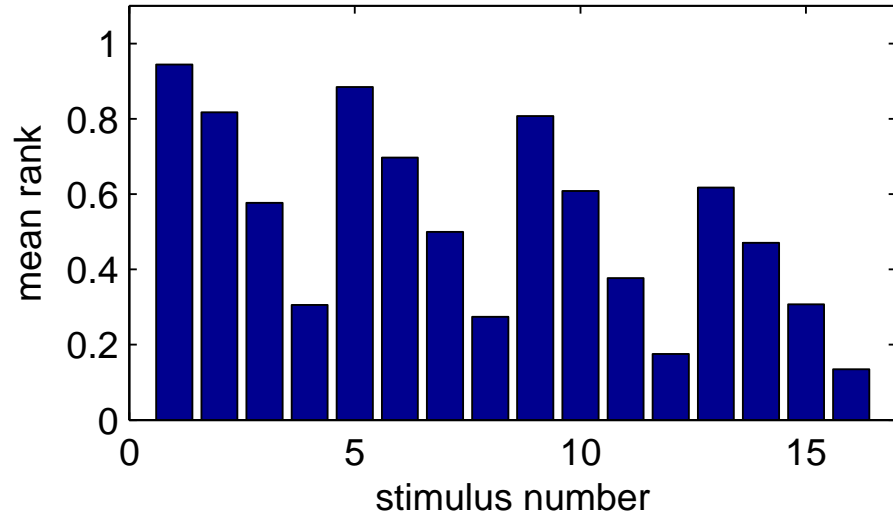
Future Work

- statistics (RM ANOVA)
- adjust the machine listening algorithm according to these results

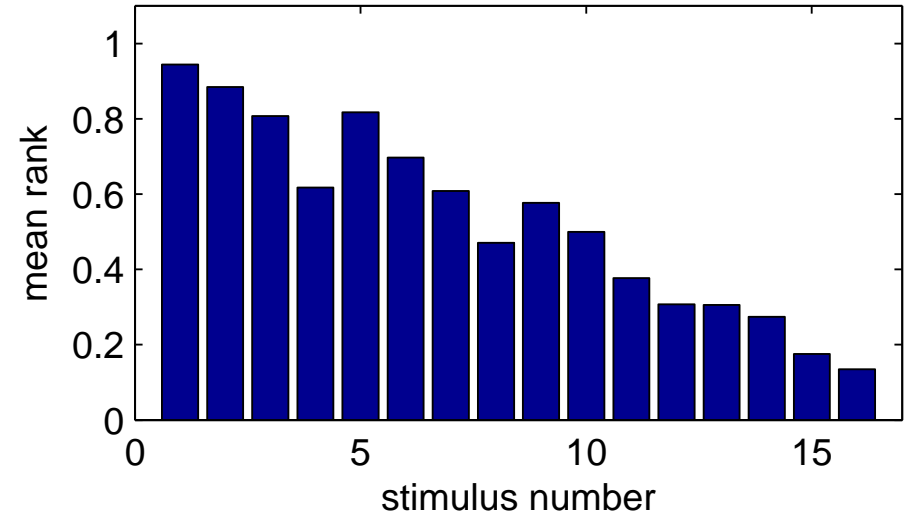
Results

Stimulus Set A, average

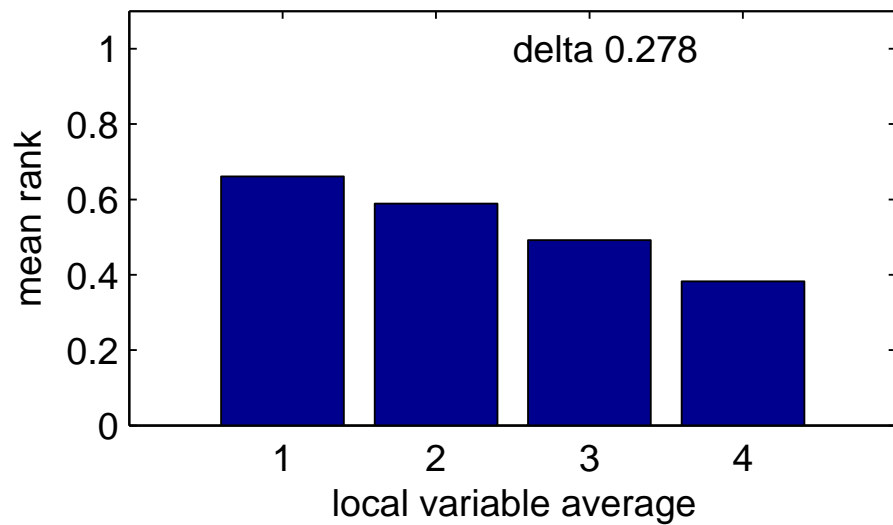
set A, global = string resonance



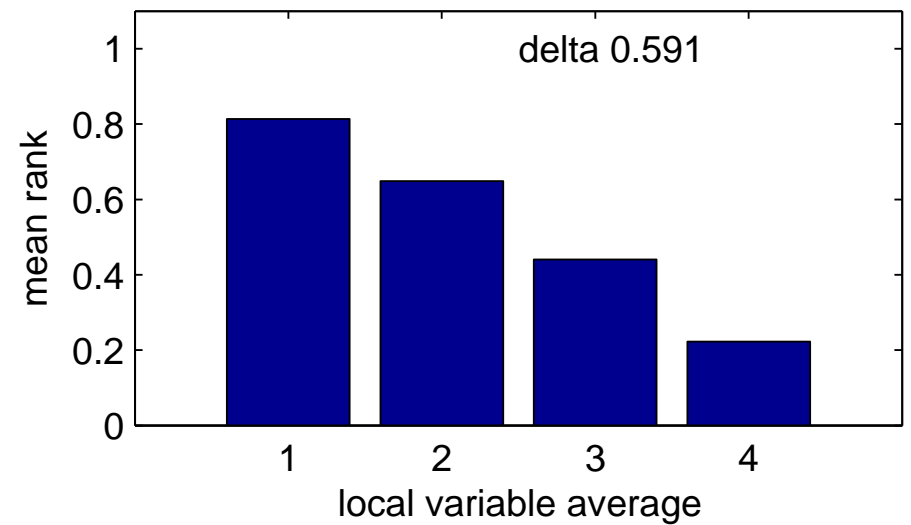
set A, global = rise time



delta 0.278



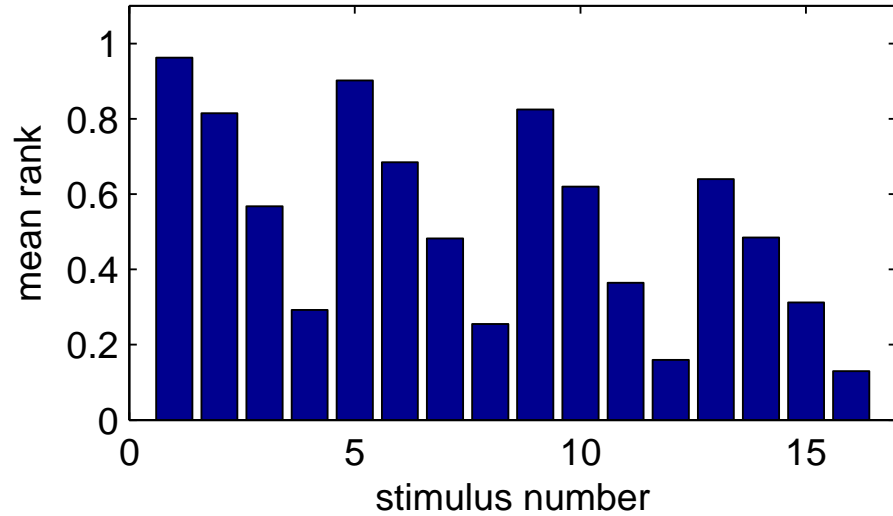
delta 0.591



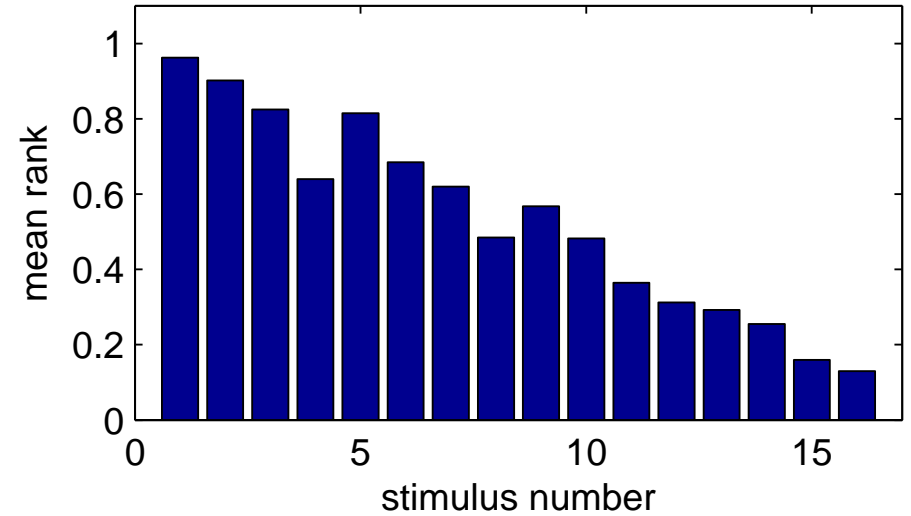
Results

Stimulus Set A, average with no outliers

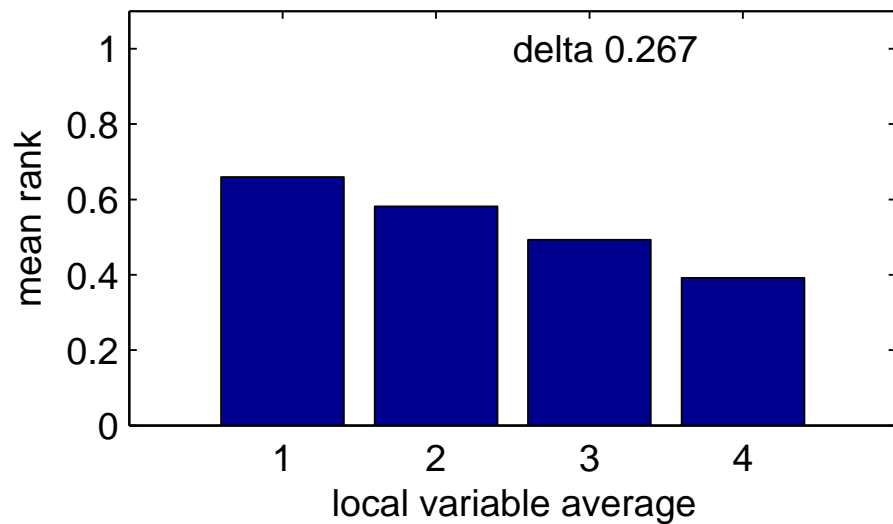
set A, global = string resonance



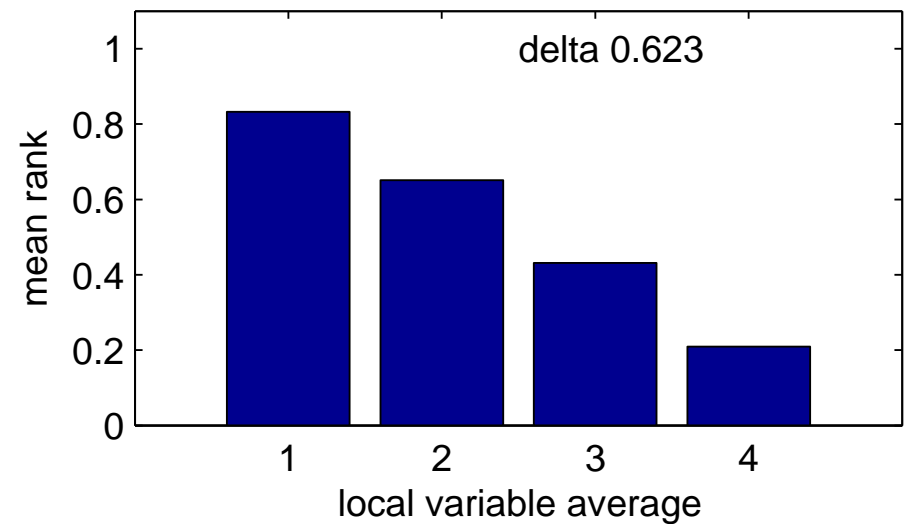
set A, global = rise time



delta 0.267



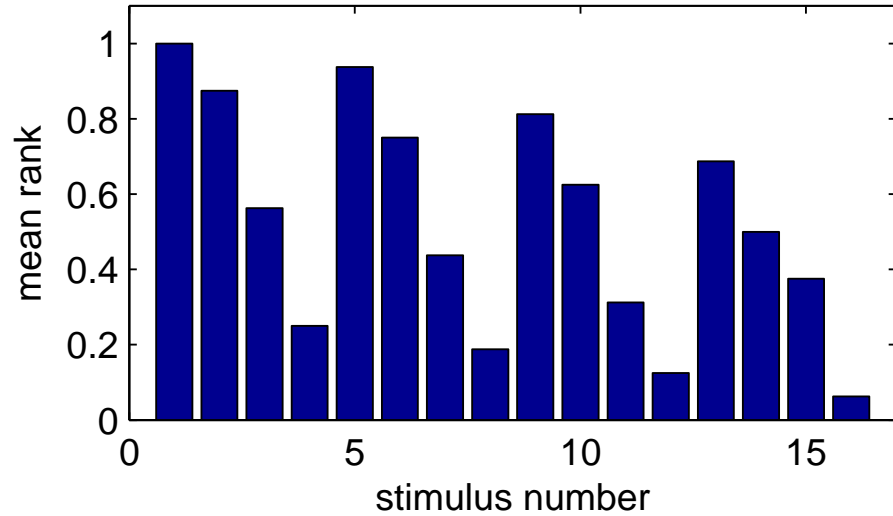
delta 0.623



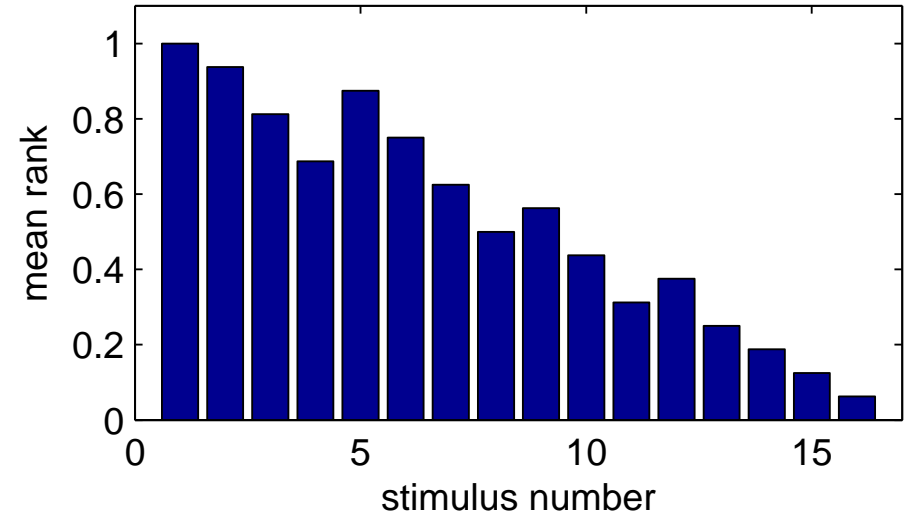
Results

Stimulus Set A, mode with no outliers

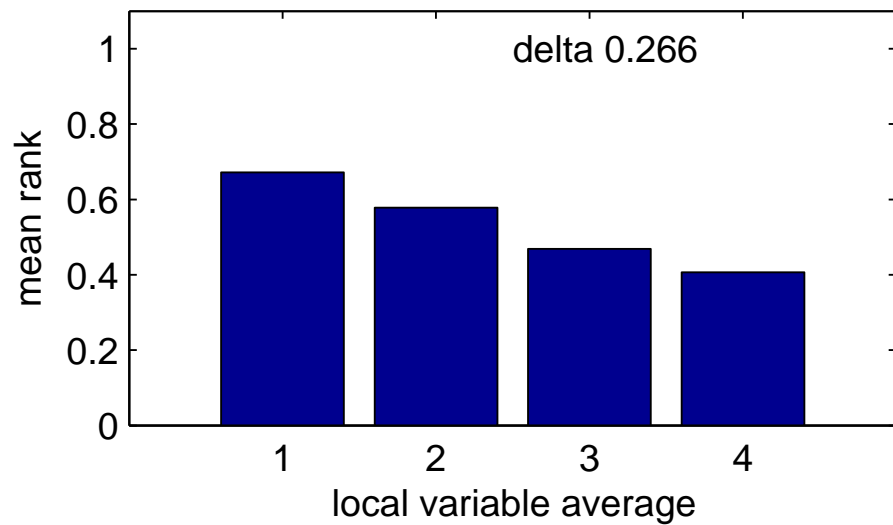
set A, global = string resonance



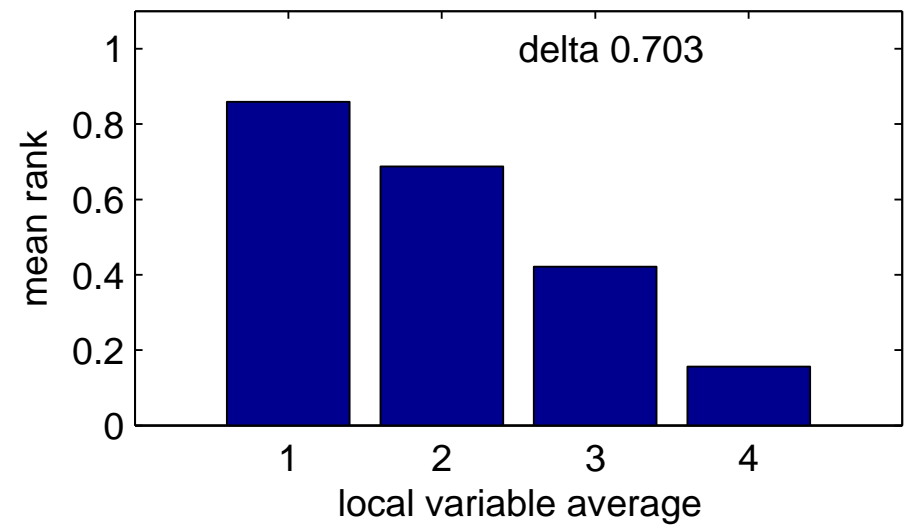
set A, global = rise time



delta 0.266



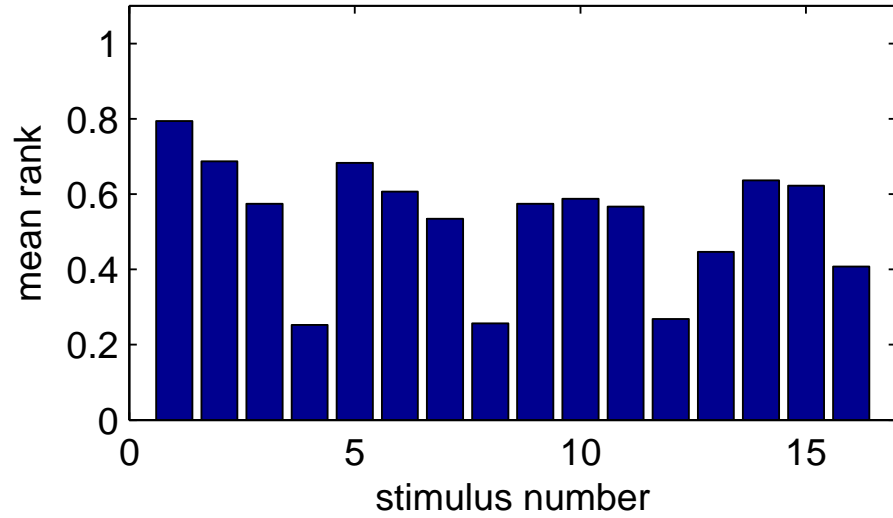
delta 0.703



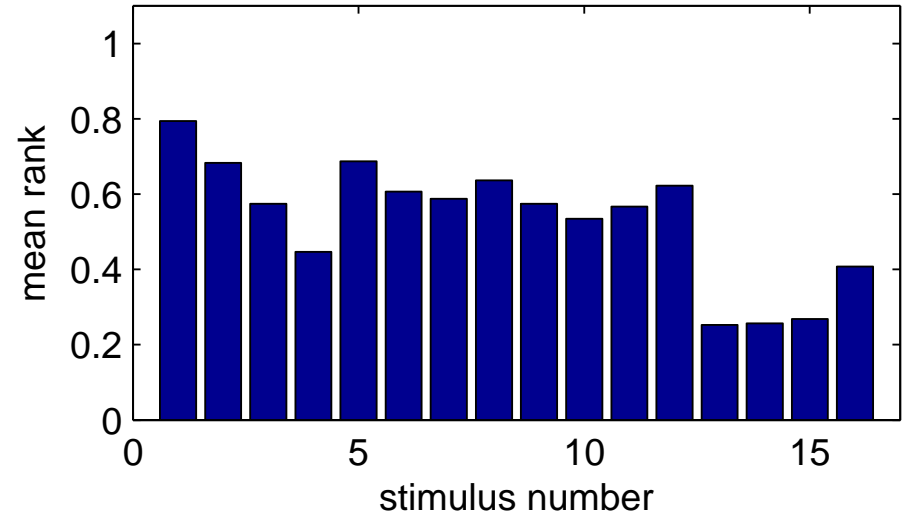
Results

Stimulus Set B

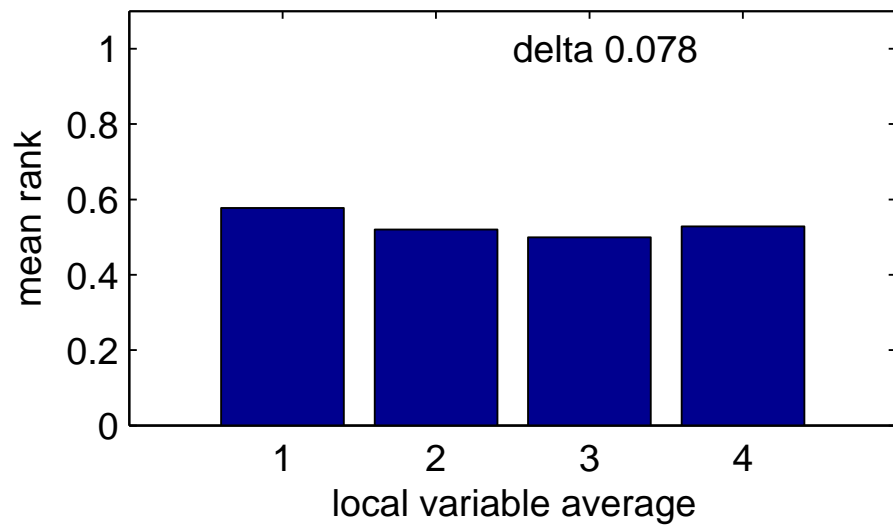
set B, global = noisiness



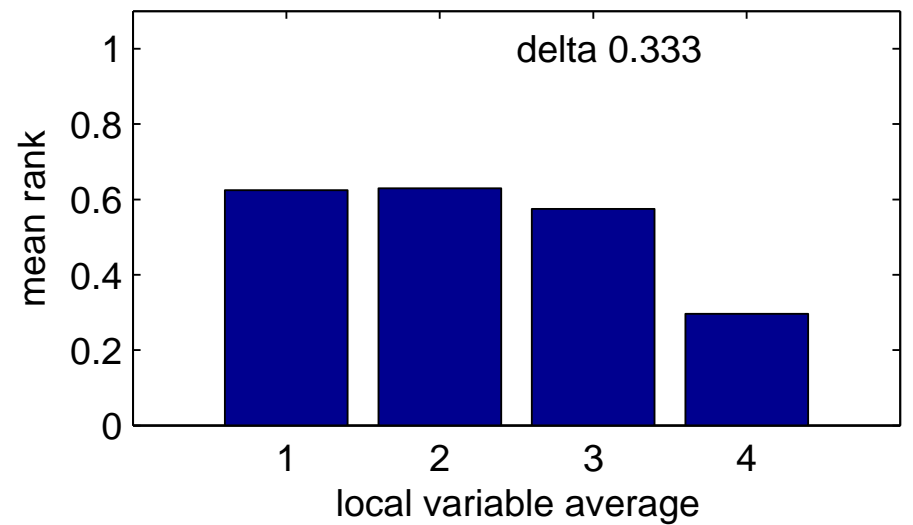
set B, global = gross filtering



delta 0.078



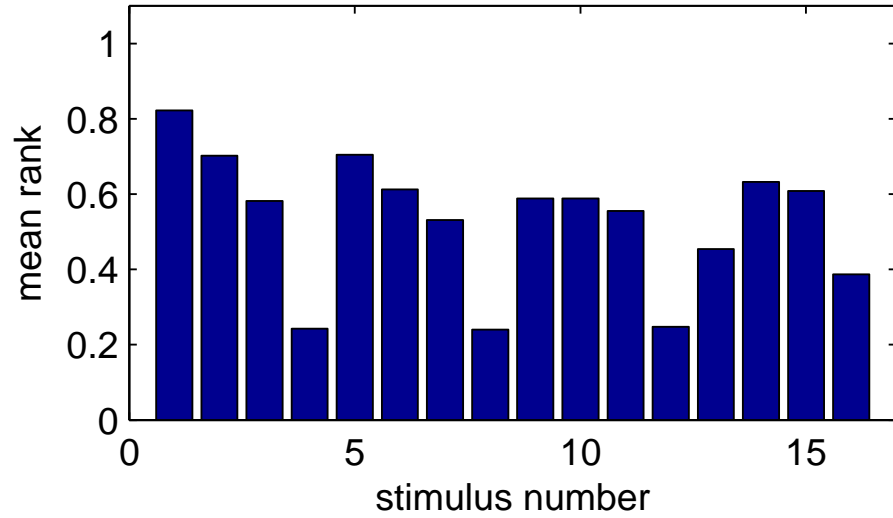
delta 0.333



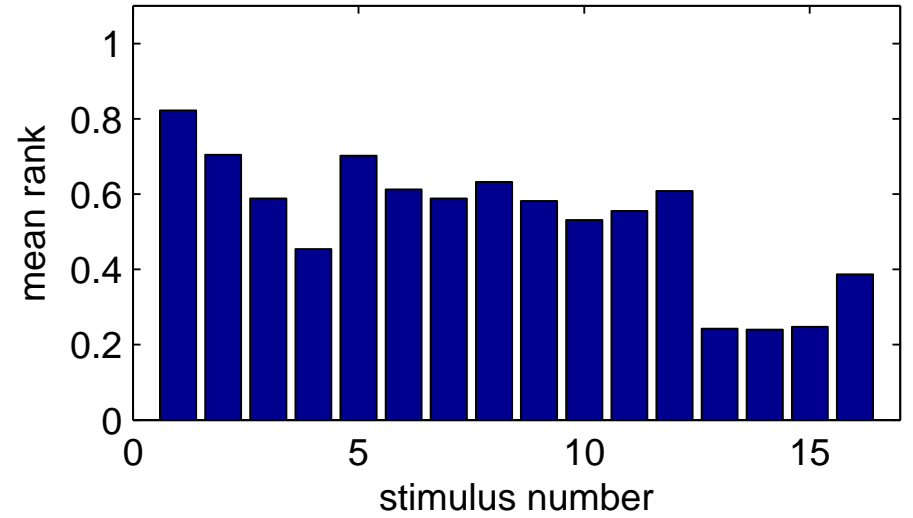
Results

Stimulus Set B, no outliers

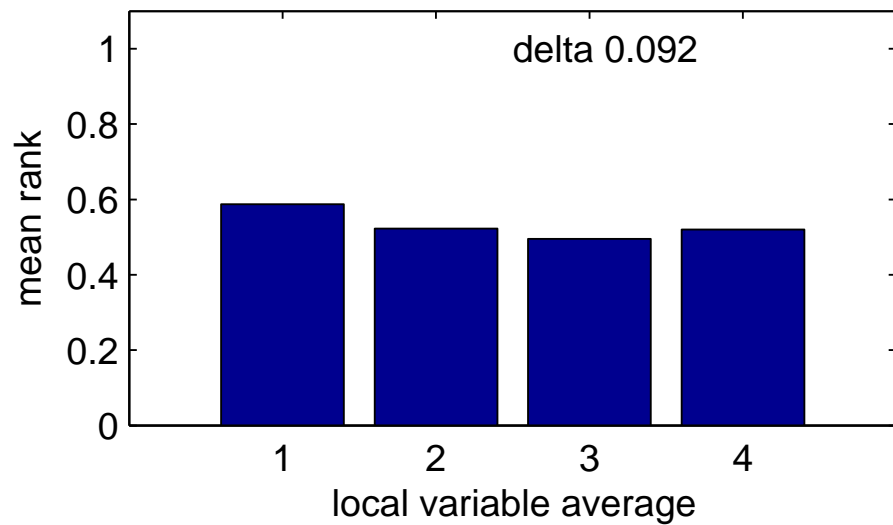
set B, global = noisiness



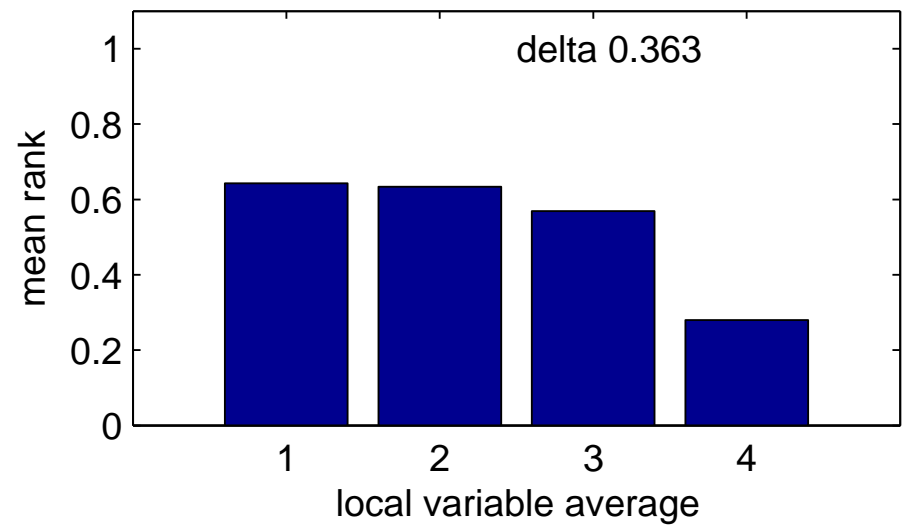
set B, global = gross filtering



delta 0.092



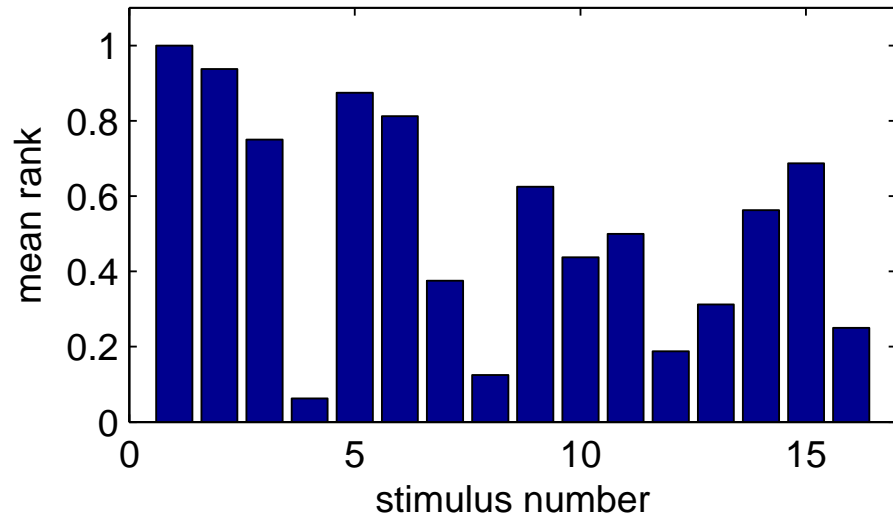
delta 0.363



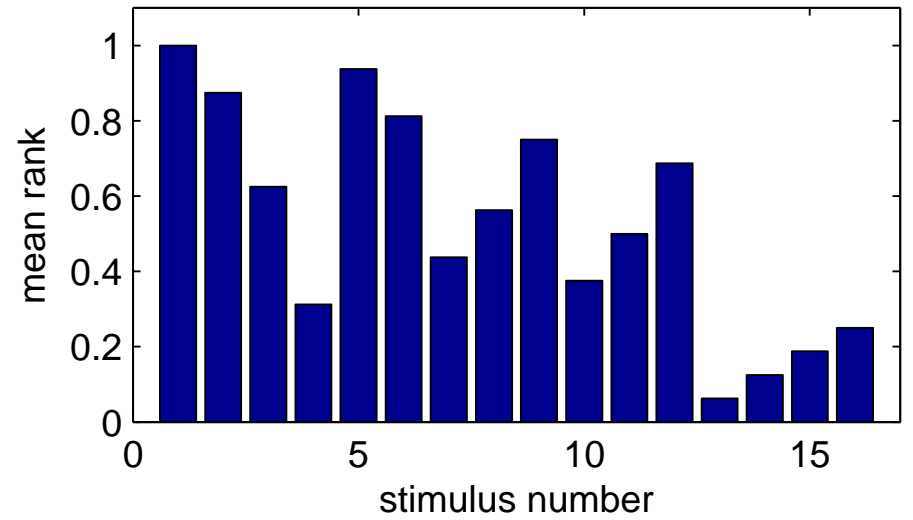
Results

Stimulus Set B, mode with no outliers

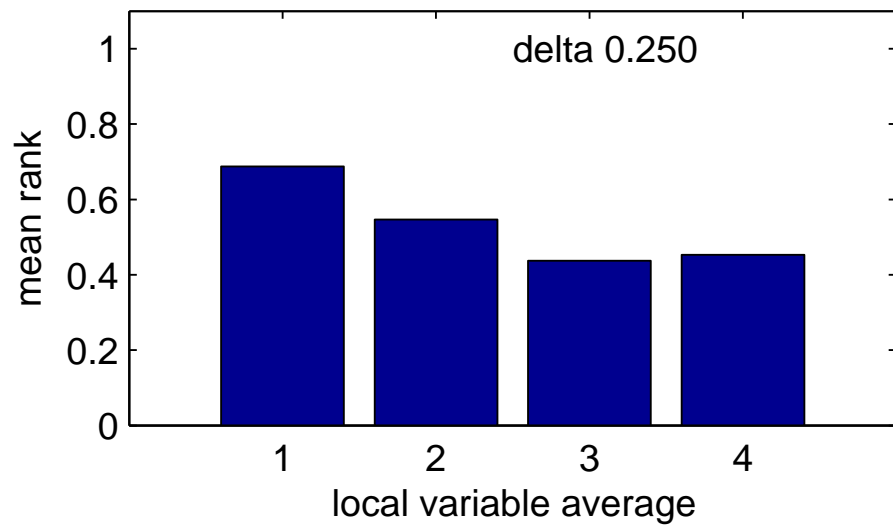
set B, global = noisiness



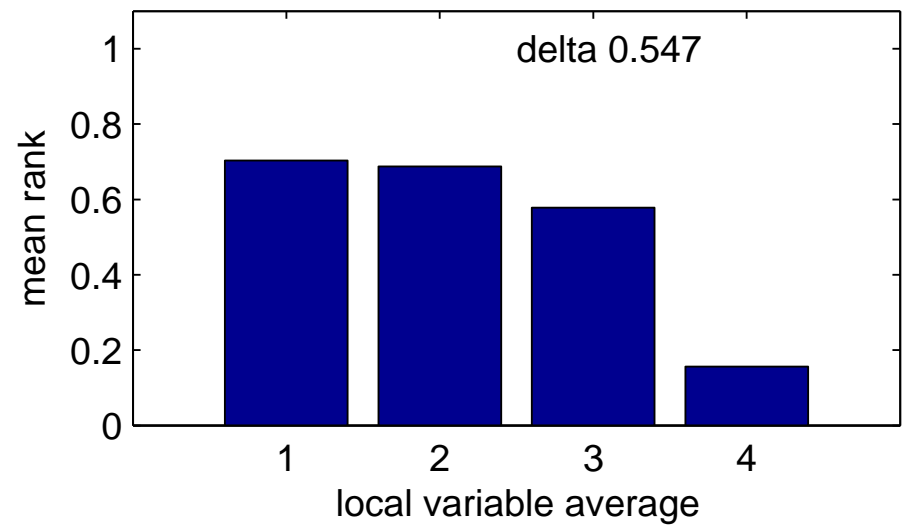
set B, global = gross filtering



delta 0.250



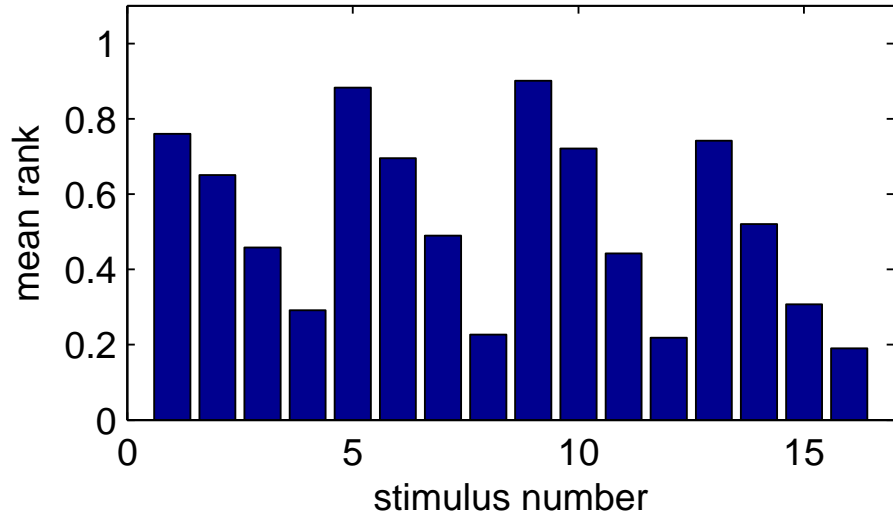
delta 0.547



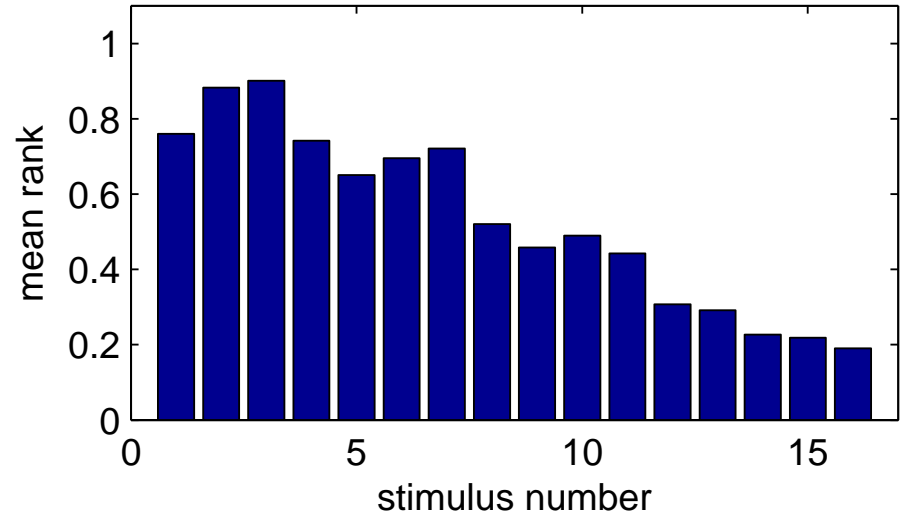
Results

Stimulus Set C

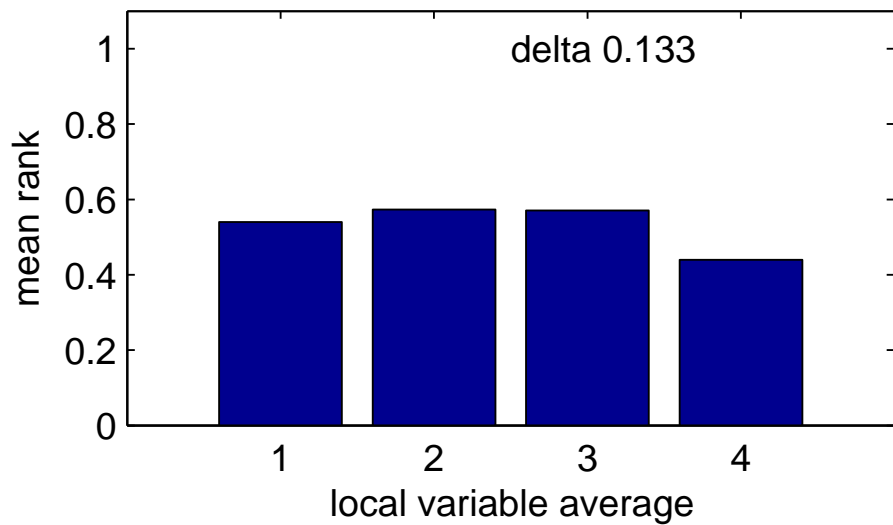
set C, global = gross filtering



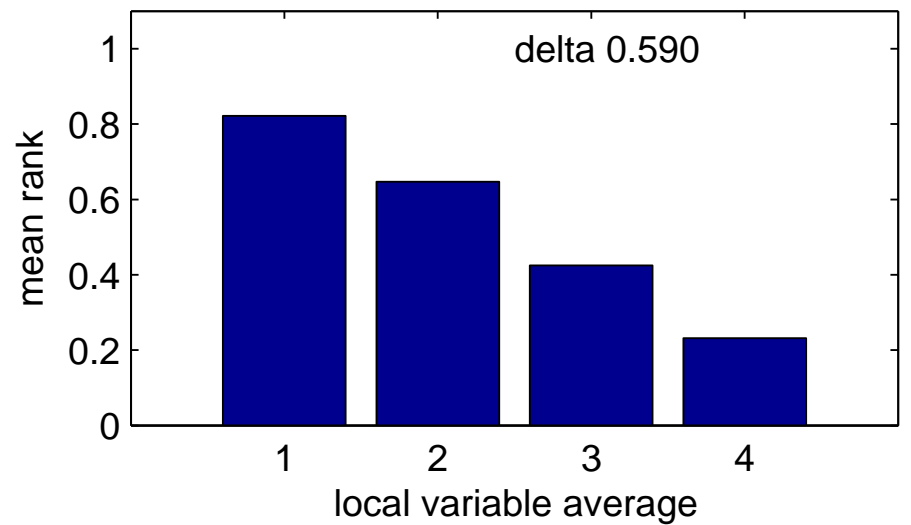
set C, global = rise time



delta 0.133



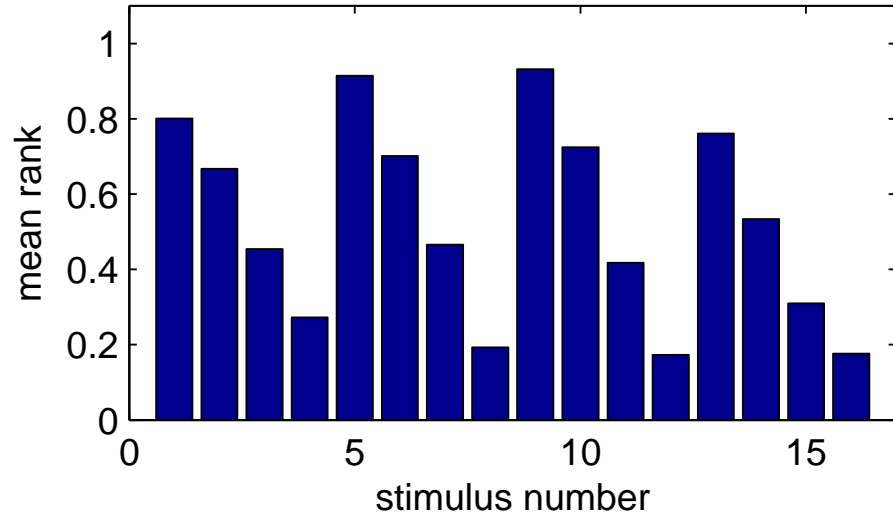
delta 0.590



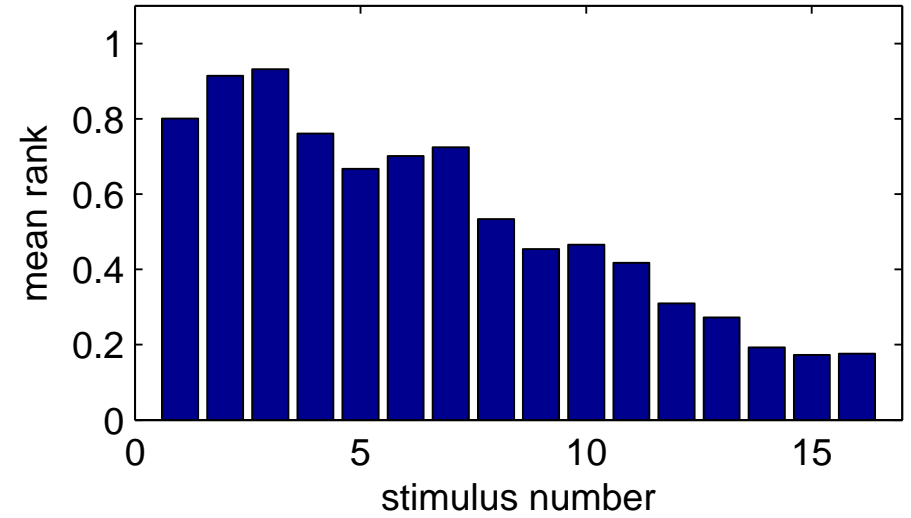
Results

Stimulus Set C, no outliers

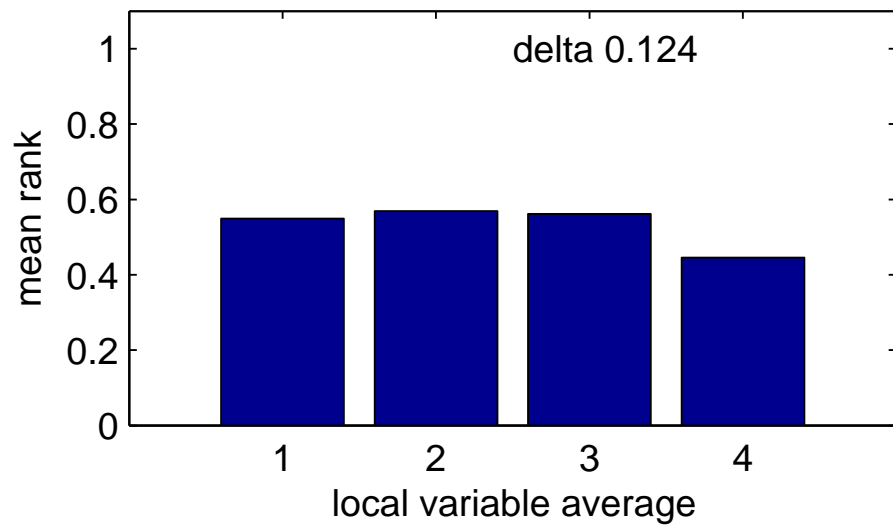
set C, global = gross filtering



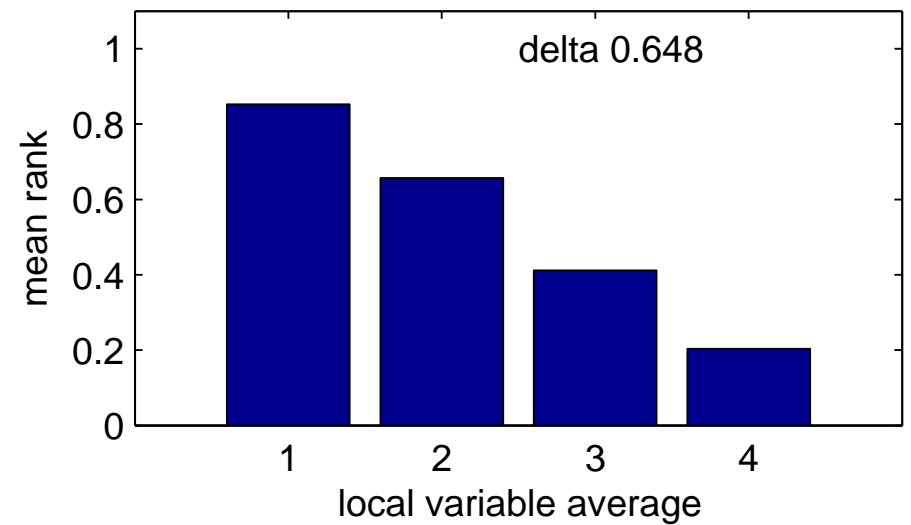
set C, global = rise time



delta 0.124



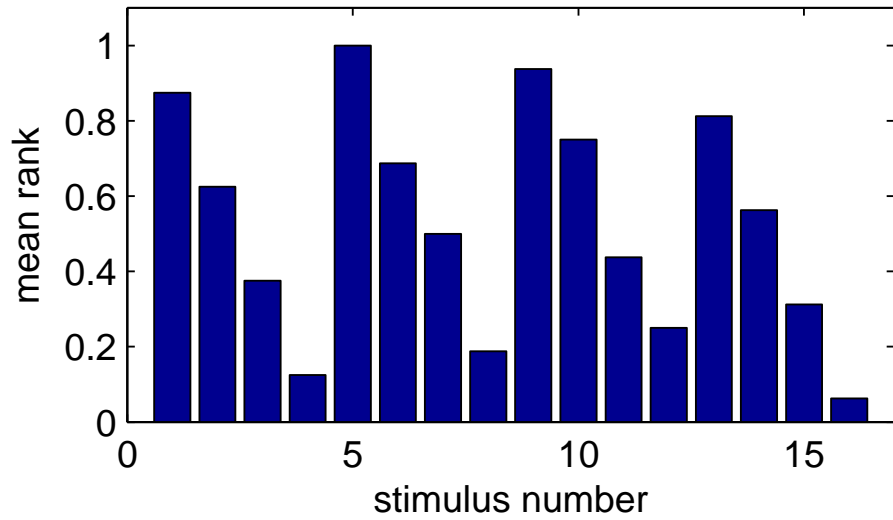
delta 0.648



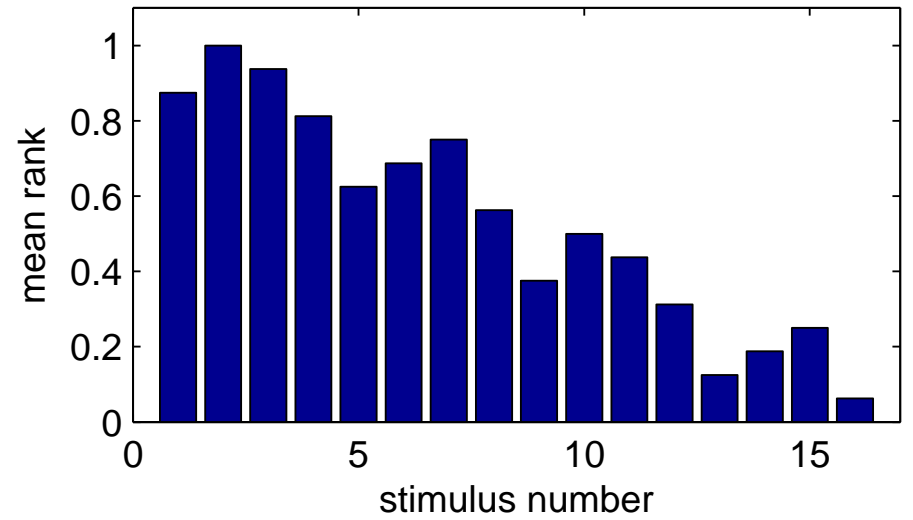
Results

Stimulus Set C, mode with no outliers

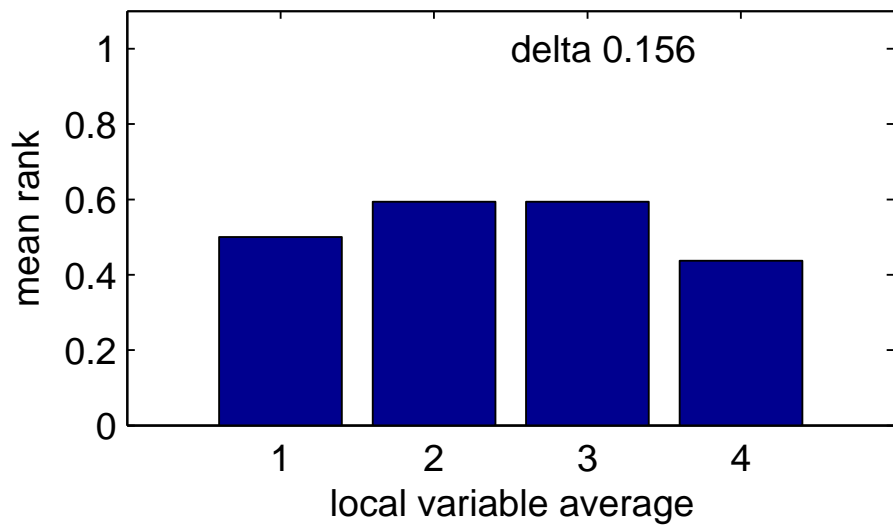
set C, global = gross filtering



set C, global = rise time



delta 0.156



delta 0.750

