Beat Tracking and Reaction Time

Nick Collins and Ian Cross
{nc272, ic108}@cam.ac.uk

Centre for Music and Science
Faculty of Music
University of Cambridge
UK

To investigate the weaknesses of current generation (real-time, causal) computational beat trackers:

Reaction time at phase/period jumps due to changing stimuli

Signal representation and phase alignment

Reference	Stimuli	Task	Reaction times (seconds)	Notes
Moelants and McKinney [2004]	polyphonic audio	tapping the beat	2-3	data set from the referenced paper. First tap taken as indicator, (preliminary results)
Toiviainen and Synder [2003]	Bach MIDI organ	"tap the beat of the music"	1.6-2.4	"do not begin tapping until you have found the beat mentally"
Dixon and Goebl [2002]	Mozart piano sonatas	"tap the beat in time"	1.3 to 1.87	synchronisation time calculated from average responses in beats and average IBIs of stimuli
Repp [2001]	isochronous tones	tapping to a step tempo change to slower rate	up to 4 beats, around 2.1s	time to adaptation
Repp [2001]	isochronous tones	tapping to a step tempo change to faster rate	up to 7 beats, around 3.325s	time to adaptation
Pouliot and Grondin [2005]	Chopin piano prelude	detect abrupt 1-5% tempo change	1.45- 4.76	

Table 1: Reaction time measurements from the rhythm perception and production literature

Exploring ecologically valid stimuli, ie pop/dance music with a mixture of transient rich drum heavy material and smoother, more pitch cued instrumentation.

The sort of polyphonic music I need computational beat trackers to follow in concert situations.

Subject tapping was assessed with respect to a given ground truth prepared with an Annotation GUI: 5 possible tapping modes.

Find the tapping mode with minimal error:

$$\text{error score} = \frac{\text{numfalsepositives}}{\text{numtaps}} + \frac{\text{numfalsenegatives}}{\text{numground}} \quad (1)$$

With a match tolerance:

$$\text{tolerance} = \frac{0.125}{\text{extract tempo in bps}} \quad (2)$$

Reaction time is taken as first of three consecutive subject taps matched to ground truth in that mode.

Experiment 1: Phase Determination from Degraded Signals

12 musicians/11 non-musicians

Between factor: subject type

musician/non-musician

Within factor: stimulus type

three signal qualities: 1-band vocoded white noise, 6-band vocoded white-noise and CD (Scheirer 1998).

15 source extracts of around 10 seconds length (15.8 beats, starting phase of 0.2), tempi from 100-130 bpm. From Blur's *Girls and Boys* to John William's *Indiana Jones*.

Each presented twice in each signal quality condition. Thus 90 trials, 20 minute experiment.

Dependent variable: minimum phase error, averaged over the two repeats and fifteen tracks, for each condition.

Experiment run using the SuperCollider software (quick demo)

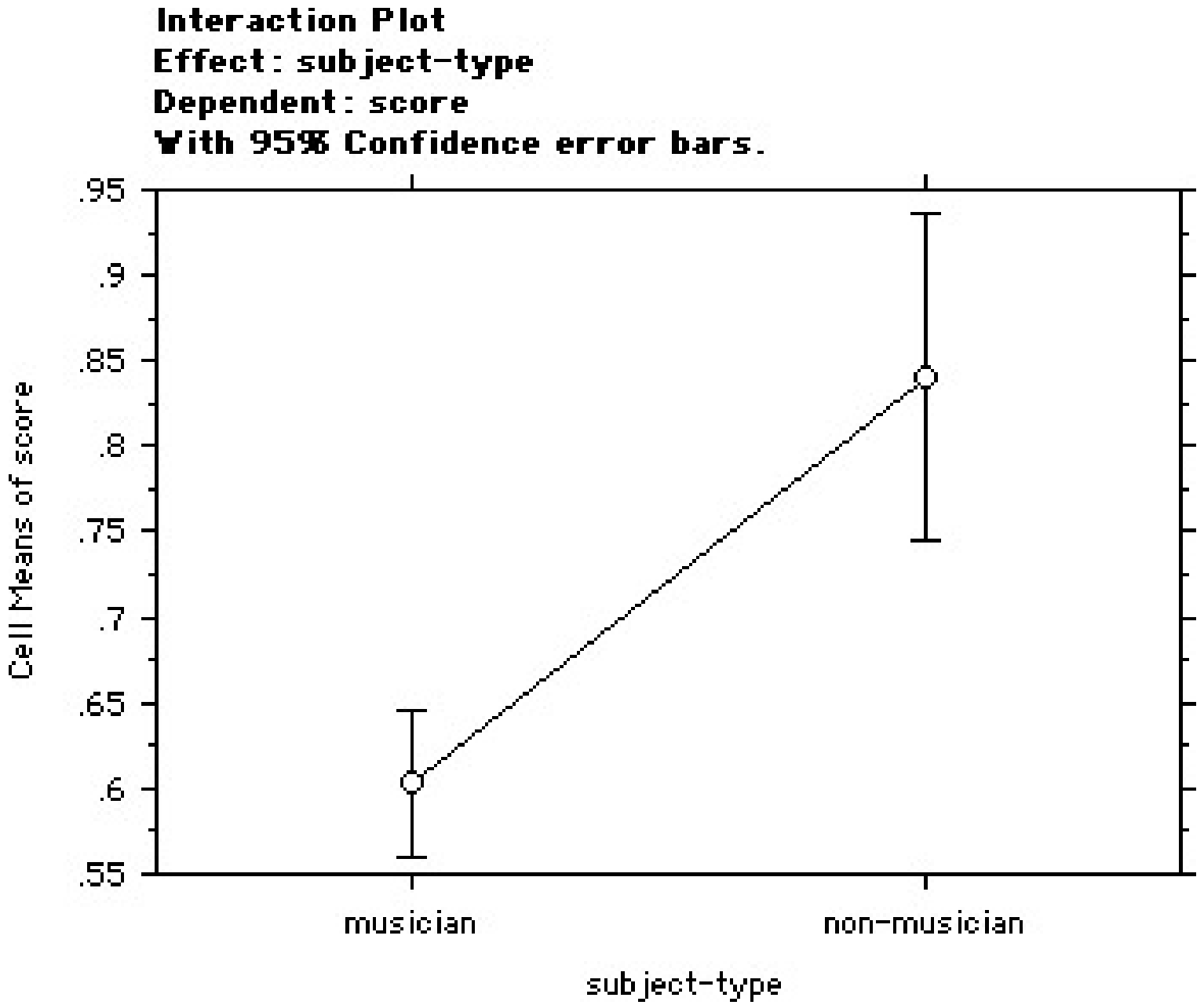
Analysed with a 1-within, 1-between ANOVA using SuperANOVA

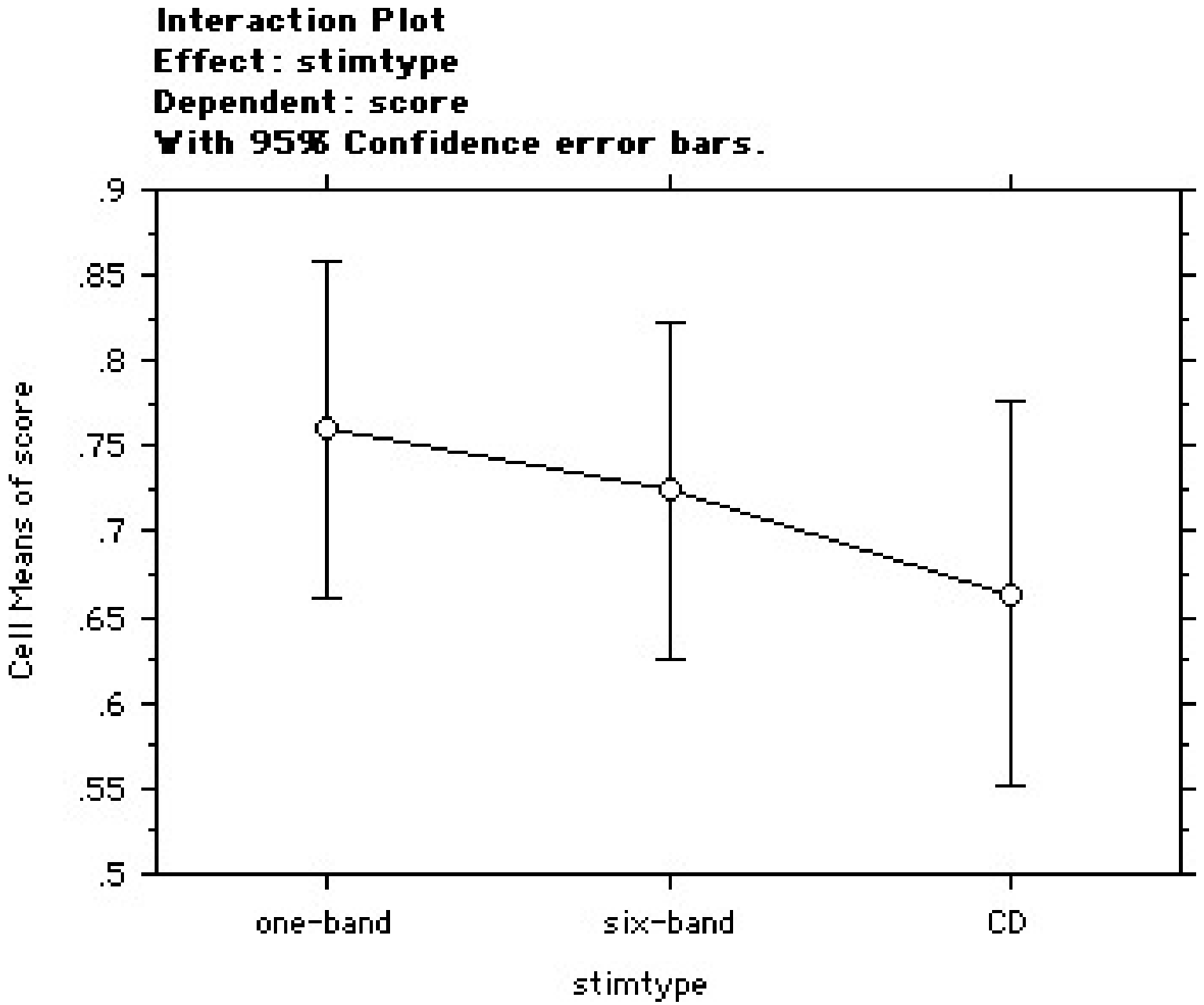
Results

Significant effect of subject type ($F(1,21)=7.949$, $p=0.0103$)

Significant effect of stimulus type ($F(2,42)=9.863$, $p=0.0004$ (G-G correction))

No significant interaction.


Interaction Plot
Effect: subject-type
Dependent: score
With 95% Confidence error bars.
Cell Means of score
.95
.9
.85
.8
.75
.7
.65
.6
.55
musician
non-musician
subject-type


Interaction Plot
Effect: stimtype
Dependent: score
With 95% Confidence error bars.
Cell Means of score
.9
.85
.8
.75
.7
.65
.6
.55
.5
one-band
six-band
CD
stimtype

Least Squares Means Table
Effect: stimtype
Dependent: score

	Vs.	Diff.	Std. Error	t-Test	P-Value
one-band	six-band	.035	.021	1.661	.1041
	CD	.096	.022	4.398	.0001
six-band	CD	.061	.022	2.737	.0091

Means Table
Effect: stim type
Dependent: reaction time

	Count	Mean	Std. Dev.	Std. Error
one band	23	2.097	.499	.104
six band	23	2.104	.443	.092
CD	23	1.888	.426	.089

Experiment 2: Reaction Time After Abrupt Transitions

13 mus/9 non-mus

Between factor: subject type

musician/non-musician

Within factors:

transition type

T→T, T→S, S→S, S→T where T is a transient rich signal and S is smoother

repetition

first and second presentation.

20 source extracts of around 6 seconds length (11.25 beats, starting phase of 0.0), tempi from 100-130 bpm. All sources were different to experiment 1, and in a mixture of styles.

Each subject took the test twice to also consider repetition as a factor.

Dependent variable: reaction time after transition averaged over the transitions in each category.

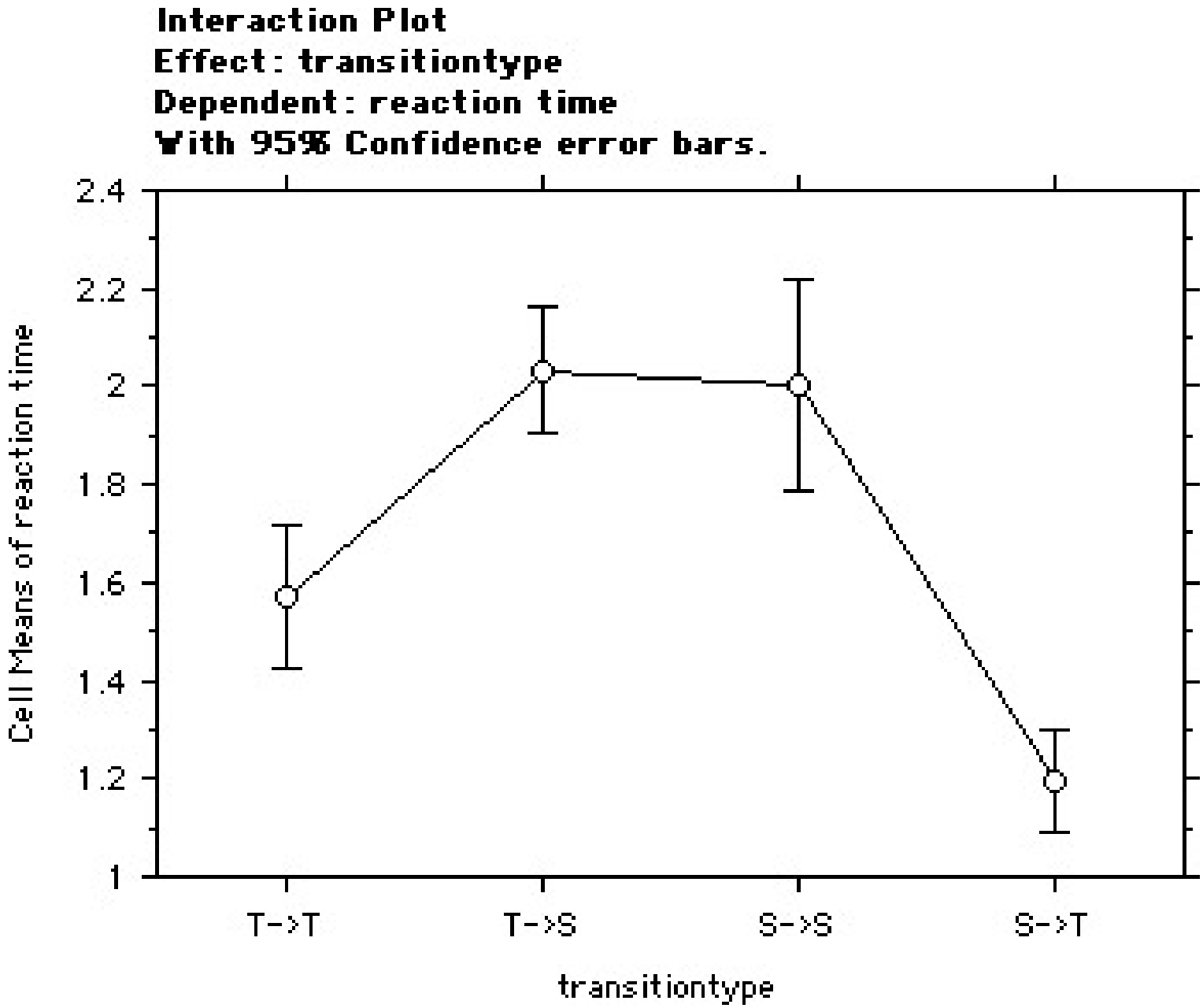
Experiment run using the SuperCollider software

Analysed with a 2-within, 1-between ANOVA using SuperANOVA

Results

Significant effect of transition type ($F(3,60)=25.987$, $p=0.001$ (G-G correction))

No significant main effect of subj type or repeat. There was a subject type/repeat interaction ($F(1,20)= 6.397$, $p=0.02$ (G-G)).


Interaction Plot
Effect: transitiontype
Dependent: reaction time
With 95% Confidence error bars.
Cell Means of reaction time
2.4
2.2
2
1.8
1.6
1.4
1.2
1
T->T
T->S
S->S
S->T
transitiontype

Least Squares Means Table
Effect: transitiontype
Dependent: reaction time

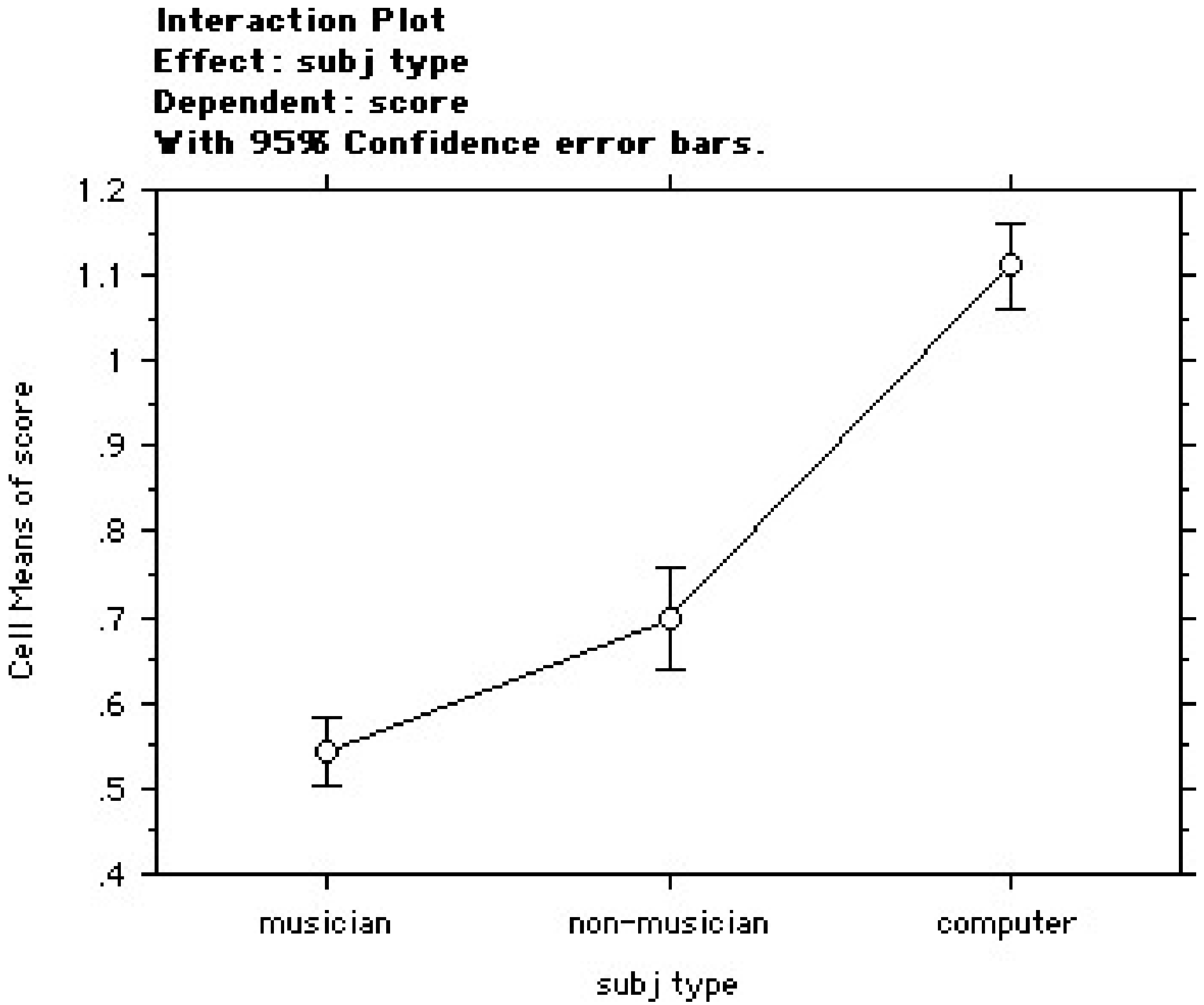
	Vs.	Diff.	Std. Error	t-Test	P-Value
T->T	T->S	-.463	.106	-4.386	.0001
	S->S	-.431	.110	-3.924	.0002
	S->T	.378	.116	3.271	.0018
T->S	S->S	.032	.070	.462	.6458
	S->T	.841	.110	7.656	.0001
S->S	S->T	.809	.112	7.194	.0001

Means Table
Effect: transitiontype * subj type
Dependent: reaction time

	Count	Mean	Std. Dev.	Std. Error
T->T, musician	26	1.587	.462	.091
T->T, non-musician	18	1.548	.514	.121
T->S, musician	26	1.969	.325	.064
T->S, non-musician	18	2.128	.542	.128
S->S, musician	26	2.020	.596	.117
S->S, non-musician	18	1.976	.857	.202
S->T, musician	26	1.123	.298	.058
S->T, non-musician	18	1.294	.395	.093

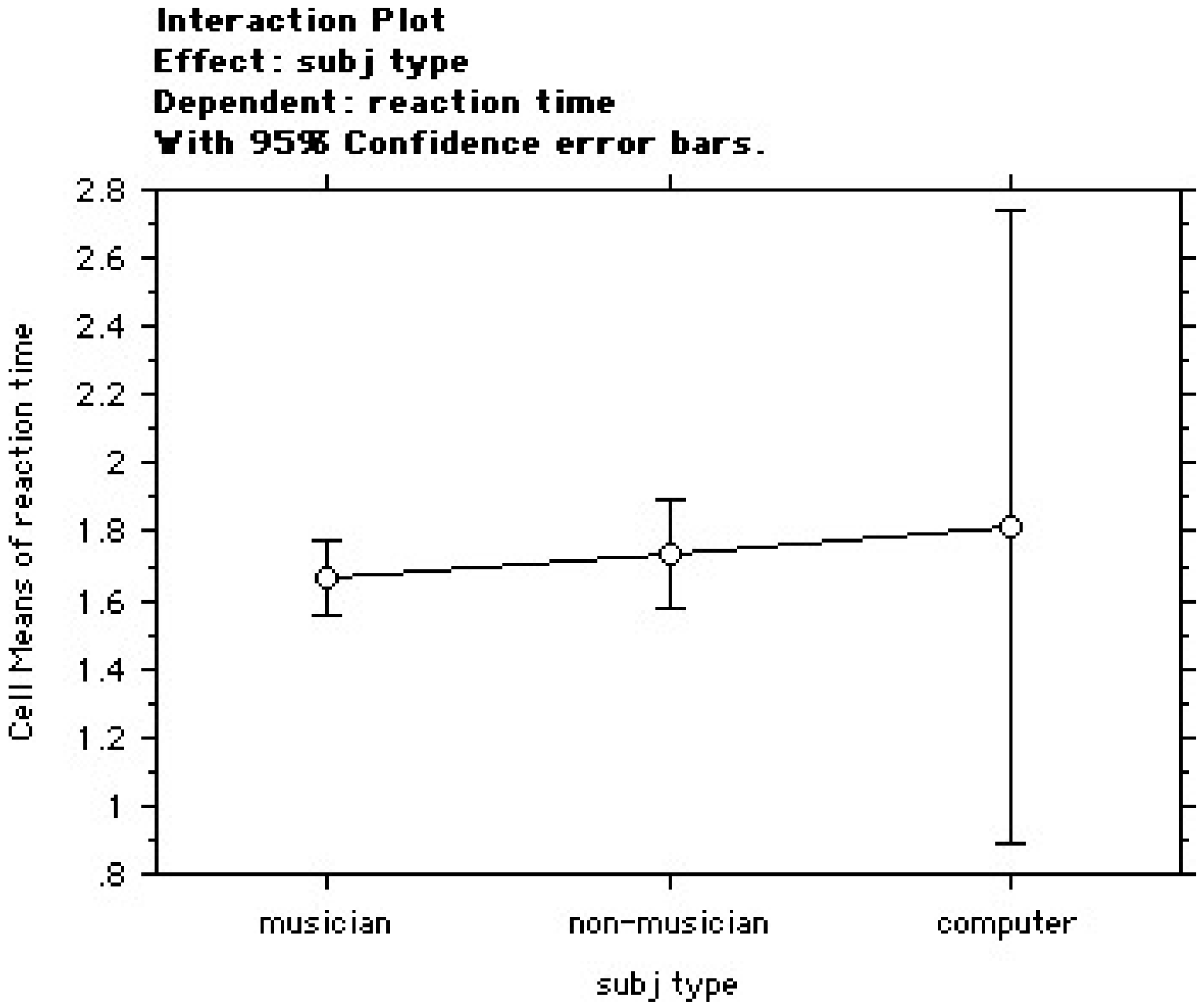
As a side analysis: same set-up, but using dependent variable of phase error score, and a three way between test on musician/non-musician/computer where computational beat trackers (Auto-Track (adapted from Davies and Plumbley 2005) and DrumTrack (Collins 2005)) are assessed as one group.

Significant effect of subject type ($F(2,21)=13.751$, $p=0.0002$)


Interaction Plot
Effect: subj type
Dependent: score
With 95% Confidence error bars.
1.2
1.1
1
.9
.8
.7
.6
.5
.4
Cell Means of score
musician
non-musician
computer
subj type

Least Squares Means Table
Effect: subj type
Dependent: score

	Vs.	Diff.	Std. Error	t-Test	P-Value
musician	non-musician	-.155	.064	-2.440	.0236
	computer	-.567	.111	-5.083	.0001
non-musician	computer	-.411	.115	-3.586	.0017


Interaction Plot
Effect: subj type
Dependent: reaction time
With 95% Confidence error bars.
Cell Means of reaction time
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
.8
musician
non-musician
computer
subj type

Computer reaction times:

- Sometimes lucky priors from a previous extract
- Mostly no adequate reaction within the short extract after a transition

Demo of computational beat tracker vs best human musician, rendering taps live.

Conclusions

Can't say that reaction time of humans faster than computational beat trackers, but certainly more reliable, even for non-musicians

Humans perform significantly less well on white noise vocoded signals; so why should we expect Scheirer's representation to be the best one for computer trackers?

Reaction times average around 1-2s; some individual musicians are faster than this.

More speculatively:

Event cues based on sound object recognition and pitch segmentation are an important mechanism; a lack of computational auditory scene analysis is holding back beat induction techniques. Event cues are degraded in energy envelope representations, particularly for classical smooth signals; the same problems are seen in computational onset detection.

Long correlation windows are not the answer for effective human-like beat tracking!

Need to spot overt piece transitions to force fast re-evaluation based on new information only (without tainting from the previous material), from knowledge of dominant instruments etc

Some support:

D. Perrot and R. O. Gjerdingen, "Scanning the dial: An exploration of factors in the identification of musical style," abstract only, presented at Society for Music Perception and Cognition, 1999.

computational transcription studies: Hainsworth 2004, Klapuri 2005

Thankyou for listening