Monophonic vs Polyphony

Timbre in music????

Polyphonic Timbre

Global sound

Polyphonic Timbre

Global sound

musical genres typically contain prototypical sound mixtures which then describes the global sound

composers also tend to have their own “sound” because of their choice of instruments

Global Sound/Polyphonic Timbre Significance

Perceptual

- genre identification and categorization (250ms-3s) (Gjerdingen & Perrott in 1999)
- affect identification (500 ms) (Peretz et al. 1998)
- agree on basic emotions(1-2sec) (Bigand et al. 2005, Watt & Ash, 1998)

Computational (MFCC)

- genre(Tzanetakis et al.,2001, Aucouturier, 2006) ,
- mood, emotion (Liu et al., Yang & Lee, Leman et al.)
- semantics (Turnbull et al,2008, Slaney, 2002)

Global Sound/Polyhpnic Timbre Significance

"spectrum-of-a-spectrum"

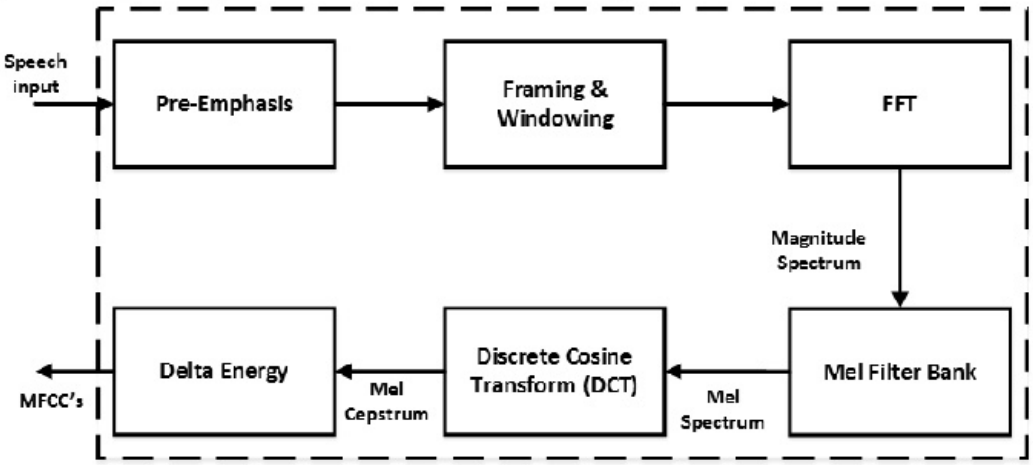


Fig. 2. Block Diagram for MFCC Computation

Computational (MFCC)

- genre(Tzanetakis et al.,2001, Aucouturier, 2006) ,
- mood, emotion (Liu et al., Yang & Lee, Leman et al.)
- semantics (Turnbull et al,2008, Slaney, 2002)

- descriptor of spectral shape based on perception
- widely used in speech research (e.g. speech recognition)
- perceptual correlates unclear

Perceptual dimensions

What do we perceive/hear when we listen to short bursts of music

What are the latent dimensions

Alluri, V. and Toiviainen, P. (2010). Exploring **perceptual** and **acoustic** correlates of polyphonic timbre. *Music Perception*, 27(3):223–241

Research Questions

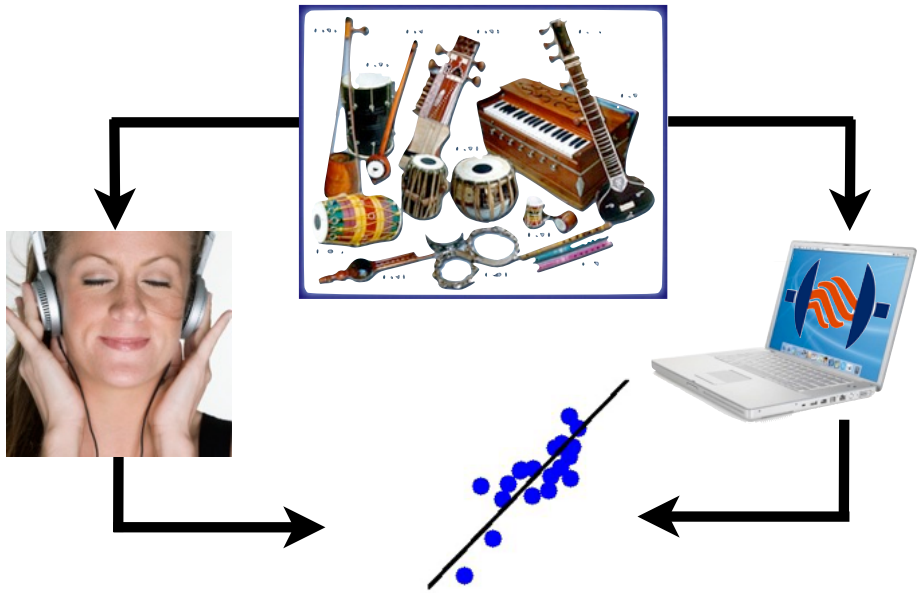
- Which terms are used to describe polyphonic timbre?
- Are these terms used consistently?
- Can they be predicted from acoustical features?

Alluri, V. and Toiviainen, P. (2010). Exploring **perceptual** and **acoustic** correlates of polyphonic timbre. *Music Perception*, 27(3):223–241

Perceptual and Acoustic Correlates of Polyphonic Timbre

- **Experiment 1**
  - Devise framework for quantitative assessments

- **Experiment 2**
  - Evaluate excerpts and uncover underlying perceptual dimensions
  - Acoustic correlates

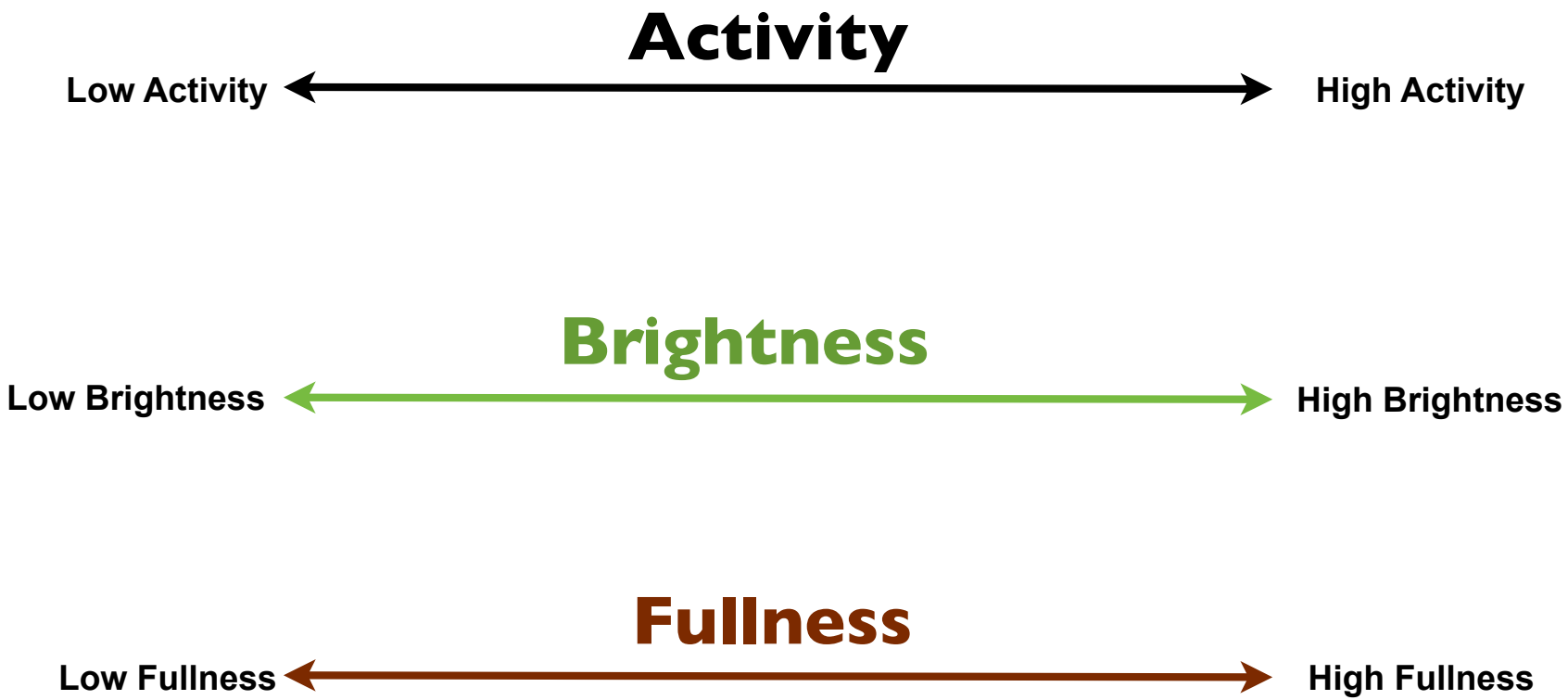


Perceptual and Acoustic Correlates of Polyphonic Timbre

- semantic associations comparable to those of monophonic timbre

Colorful-Colorless
Warm-Cold
Dark-Bright
Acoustic-Synthetic
Soft-Hard
Strong-Weak
Empty-Full
High Energy-Low Energy

Perceptual dimensions



Acoustic features

Temporal

- zero-crossing rate

Spectrotemporal

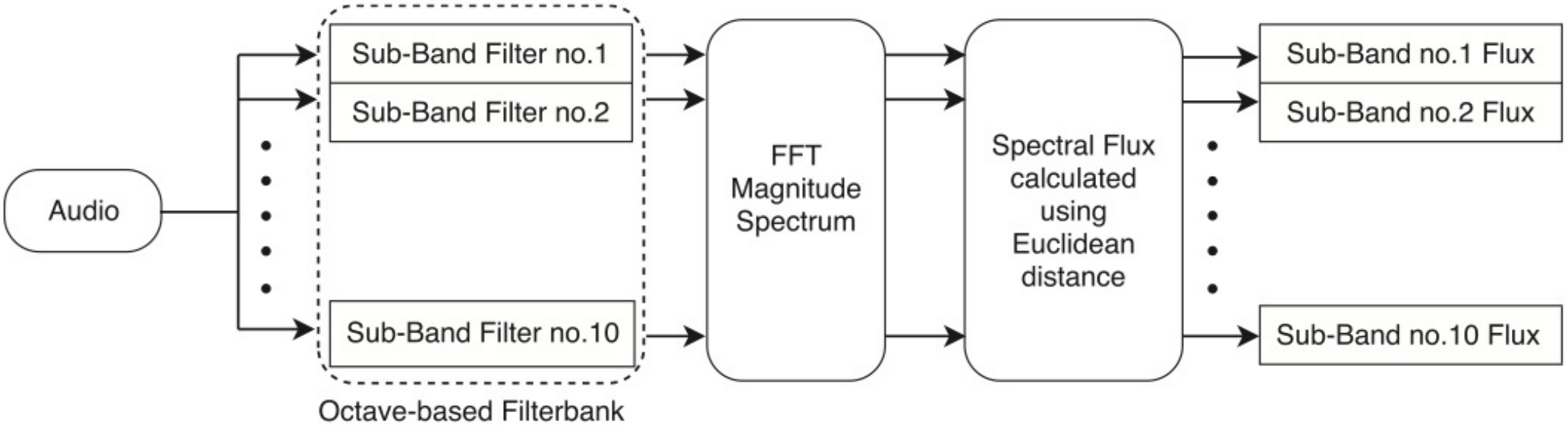
- roughness
- ***sub-band flux***

Spectral

- centroid
- high energy-low energy ratio
- entropy
- roll-off 85
- MFCC

(features extracted using MIRToolbox (Lartillot & Toiviainen)

Sub-band Flux



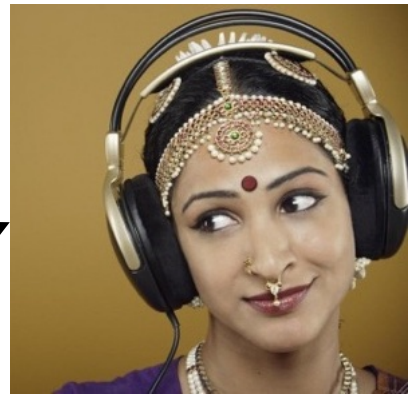
Acoustic Correlates

- acoustic correlates : subband fluxes (spectrotemporal modulations) are best predictors
- **Activity** dimension can be predicted most accurately with acoustical features
- MFCCs fail to predict

- so the Finns rated a bunch of Indian audio samples
- does that mean that our perception of global sound is the same?

Study 2 : Effect of enculturation on semantic and acoustic correlates of polyphonic timbre

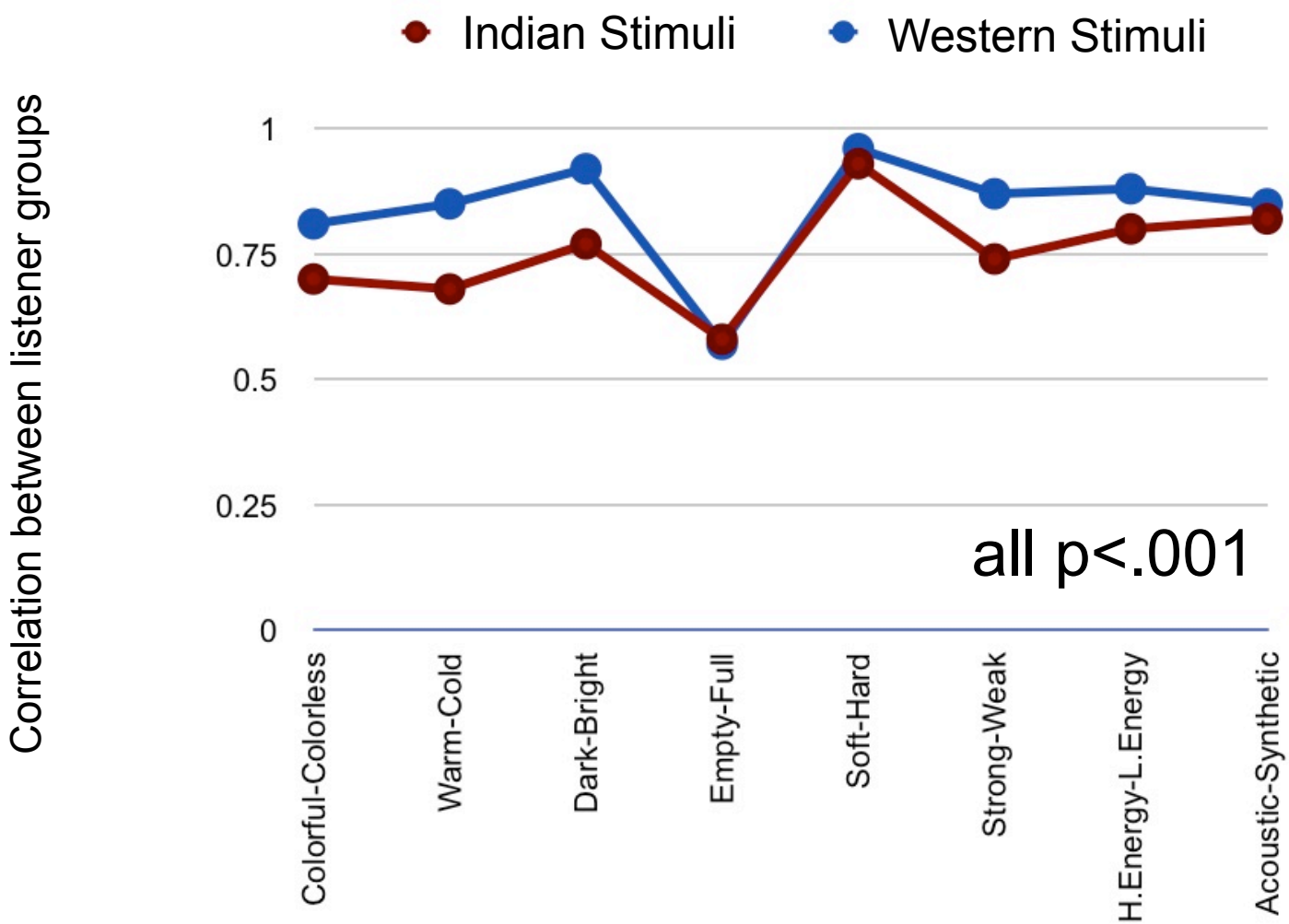
IW non-musicians WI



W W

I I

Correlation Between Listener Groups



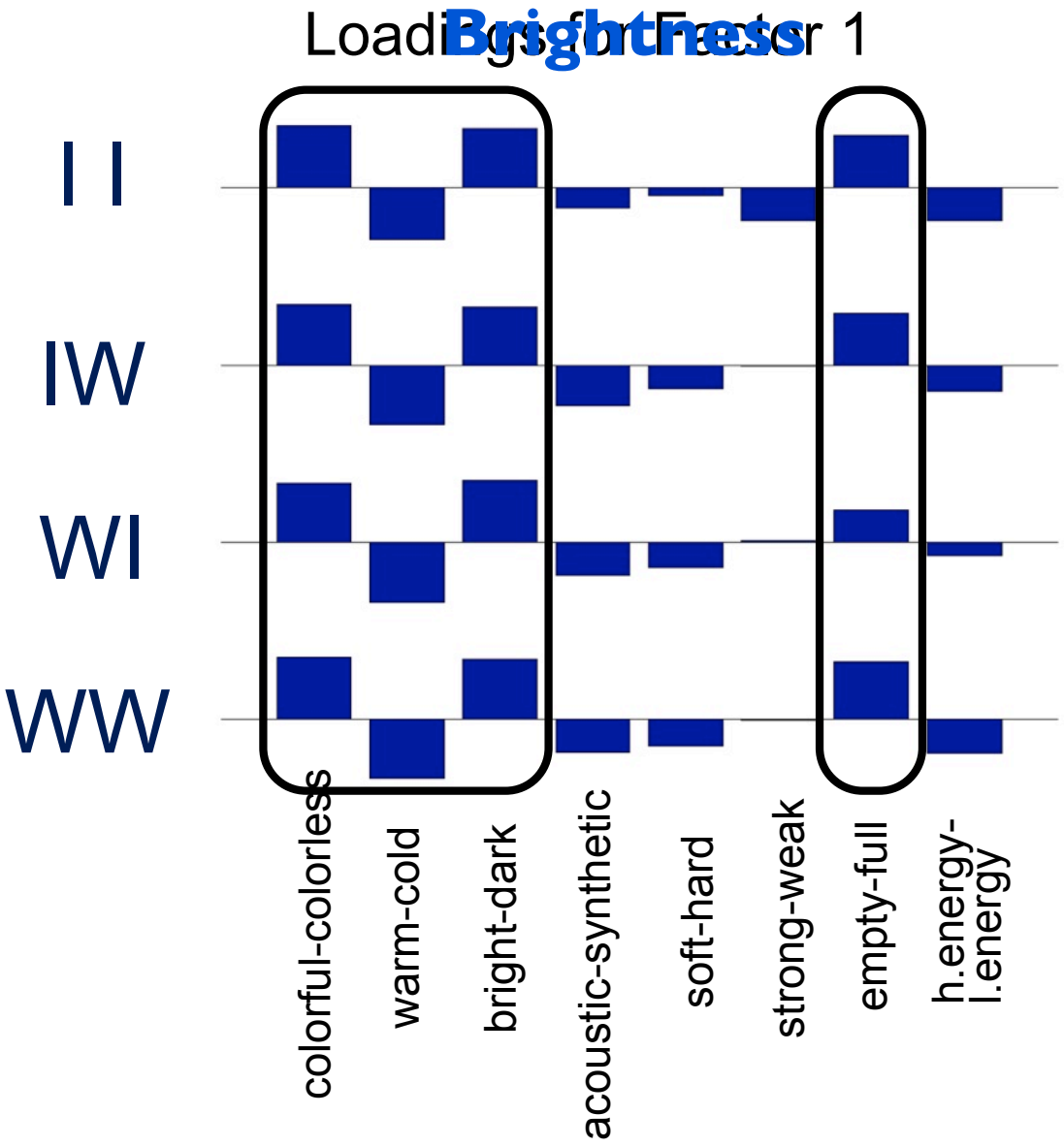
Intrinsic Dimensionality

Factor Analysis (Kaiser criterion)

	II	IW	WI	WW
No. of factors	3	2	2	3

familiarity
- increases dimensionality of perceptual timbre space
- enhances perceptual acuity

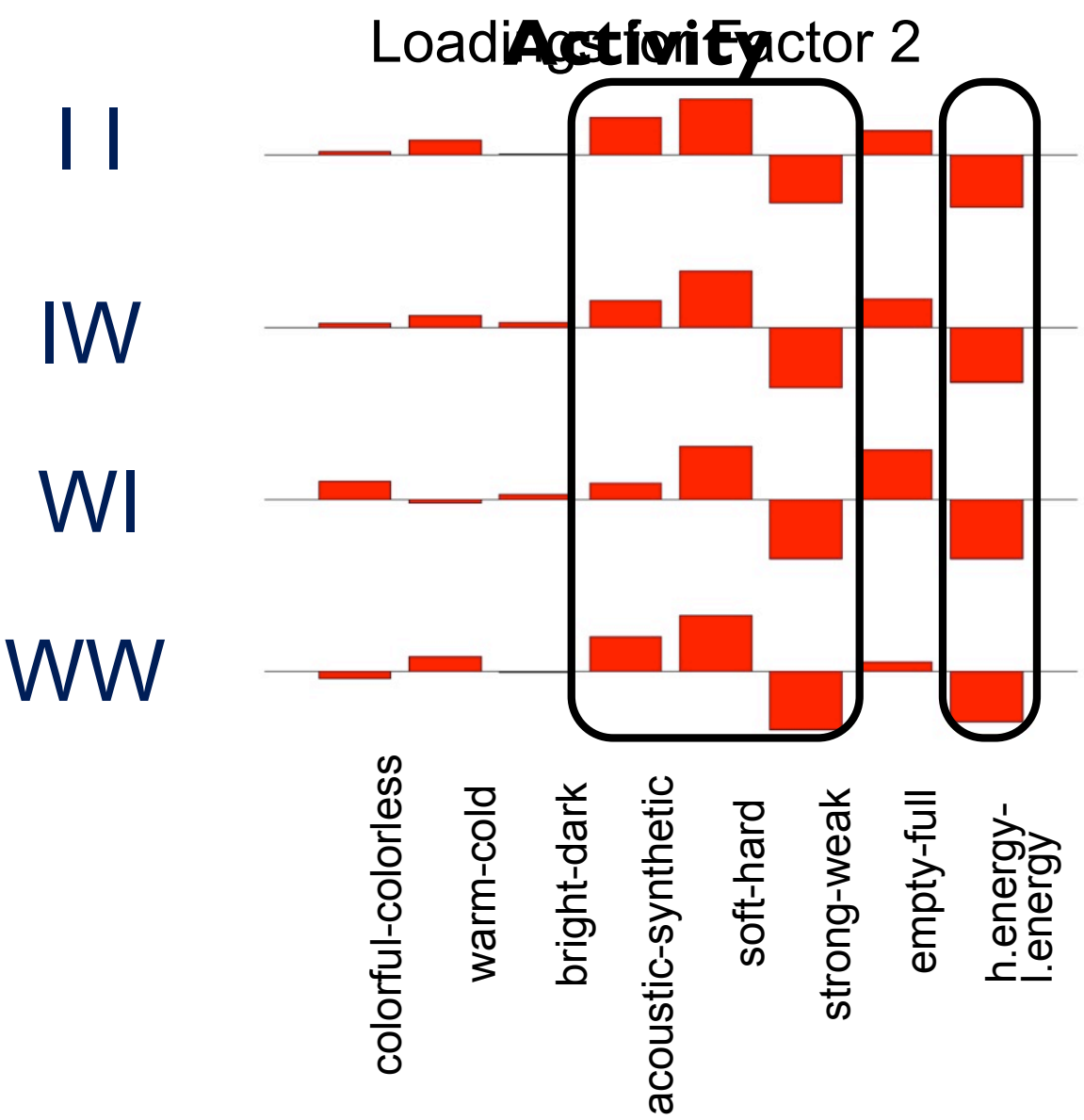
Perceptual Dimensions



High Brightness Scores

Low Brightness Scores

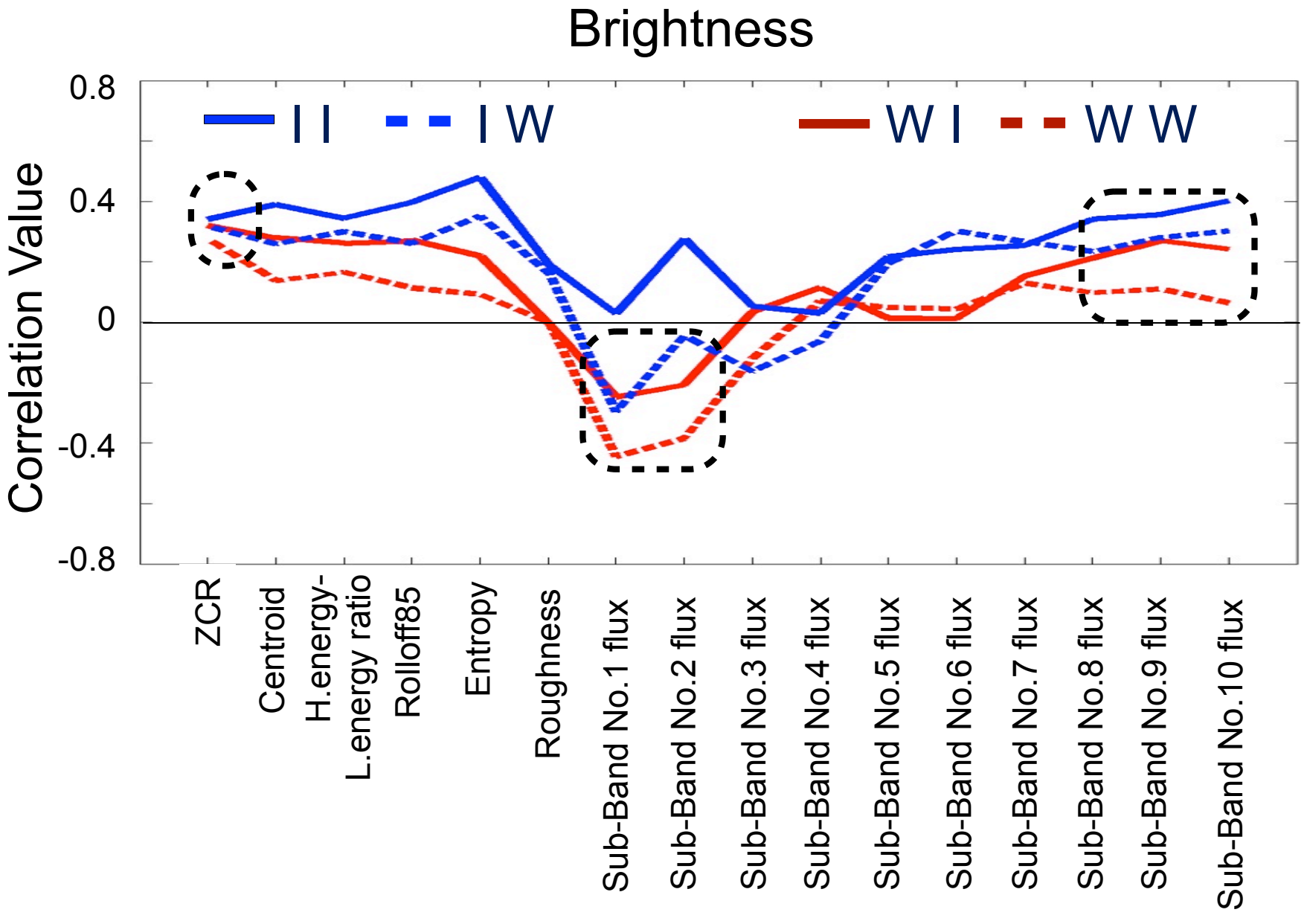
Perceptual Dimensions



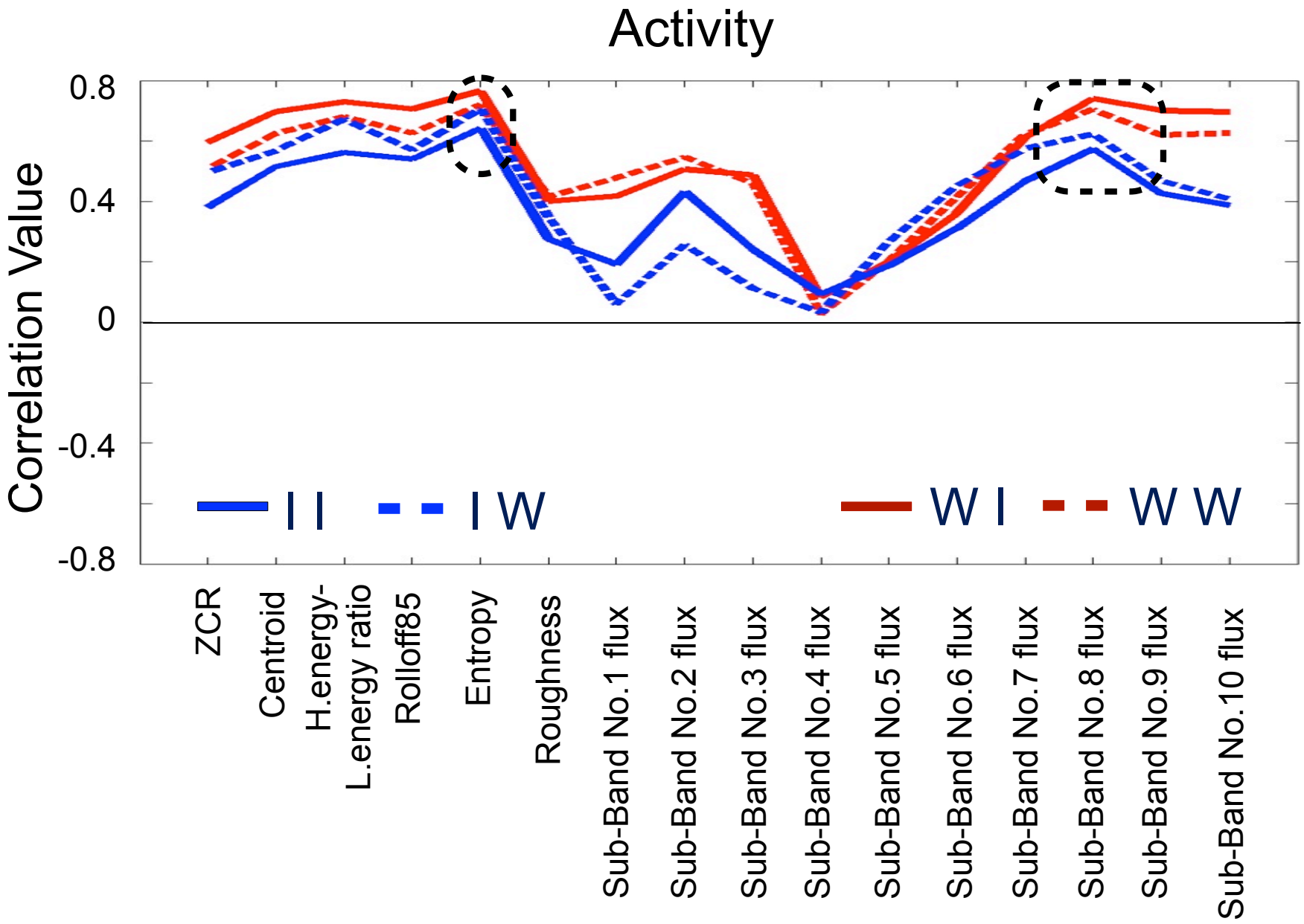
High Activity Scores

Low Activity Scores

Acoustic Feature Correlation



Acoustic Feature Correlation



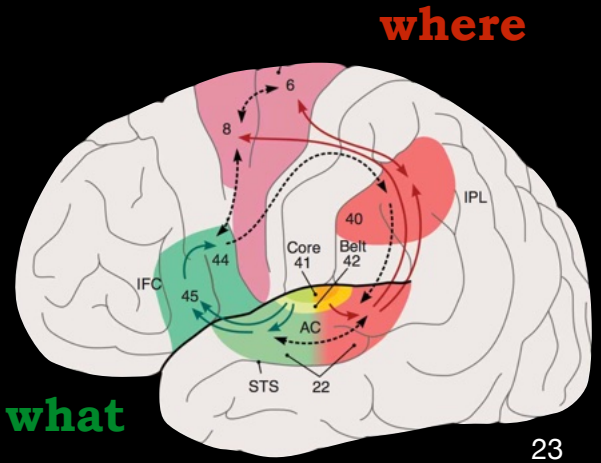
Conclusion

- Overall results hint at the existence of universal patterns in polyphonic timbre perception
  - **Activity, Brightness** - common perceptual dimensions
- increase in intrinsic dimensionality
  - enculturation
  - musical expertise

can genre, emotion be identified based on these two dimensions?

Dual Pathways in the brain

- represent the "**where**" and "**what**" of the sound
- postero-dorsal stream encompasses both **spatial** and **language** functions
- antero-ventral auditory pathway important in **auditory pattern recognition** and **object identification** and is **hierarchically** organized



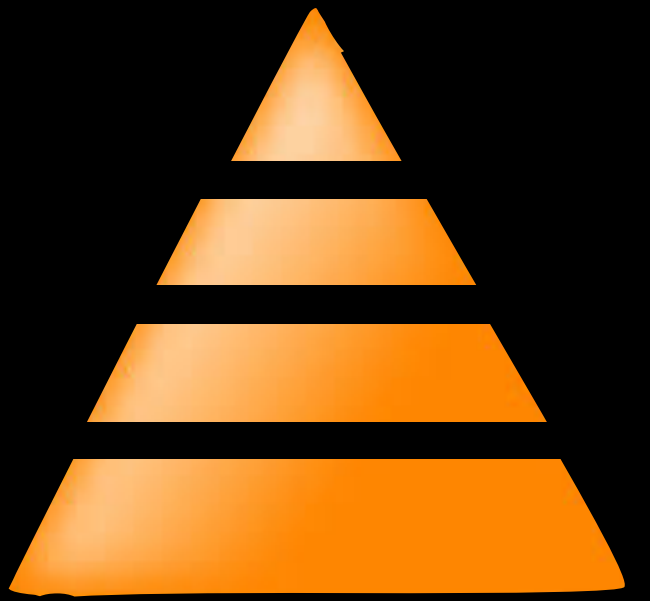
Rauschecker and Scott, 2009

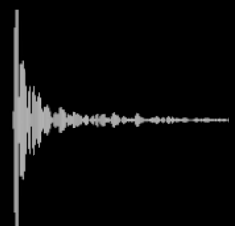
“hierarchical”

- levels of abstraction
- basis to hypothesise where timbre is encoded in the brain

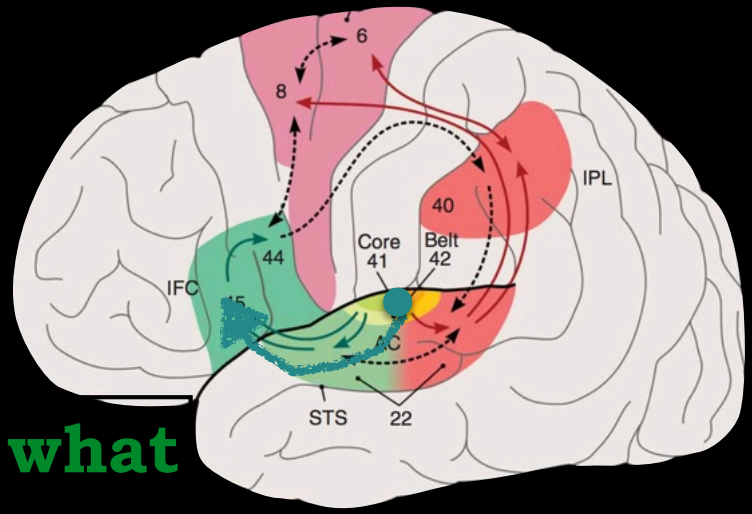
levels of abstraction

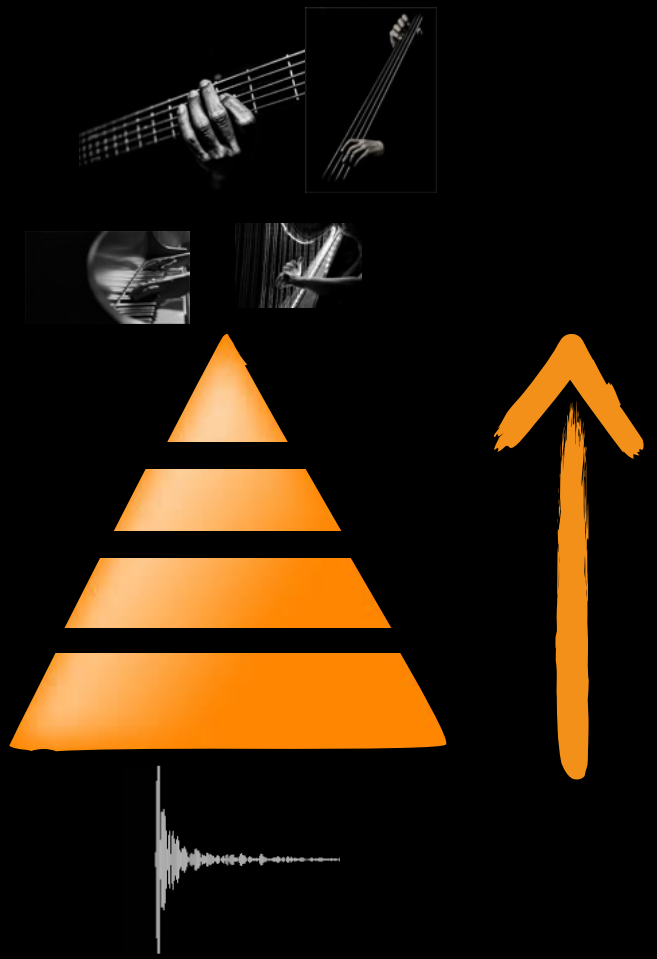
- **low** levels associated with representation of acoustic structure
- **high** levels associated with representation of sound source properties
- **higher** levels associated with identification, description, and categorization of prototypical mixtures of sources (**polyphonic timbre**/global sound)





“What” Pathway in the brain

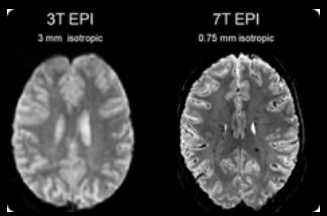


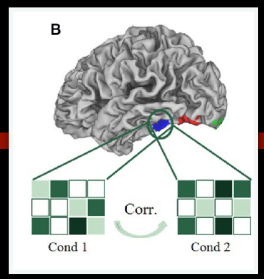


modelling: **decoding genre**

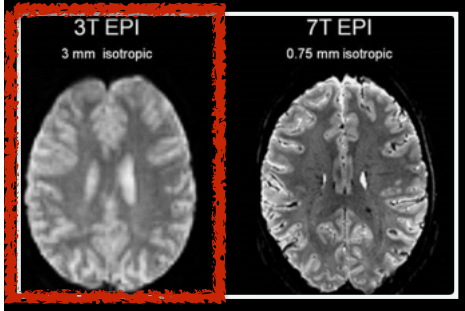
6 sec long excerpts

fMRI responses

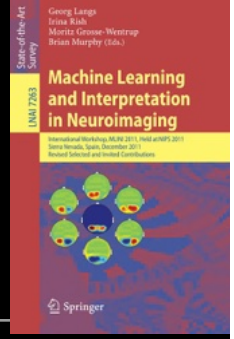




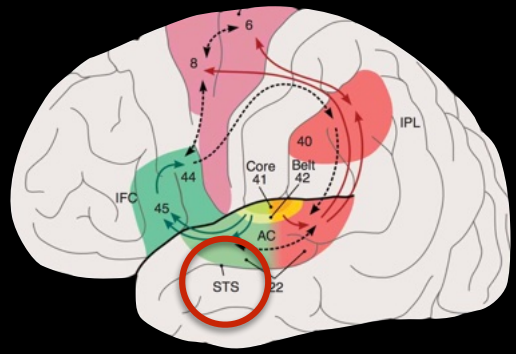
musical genre

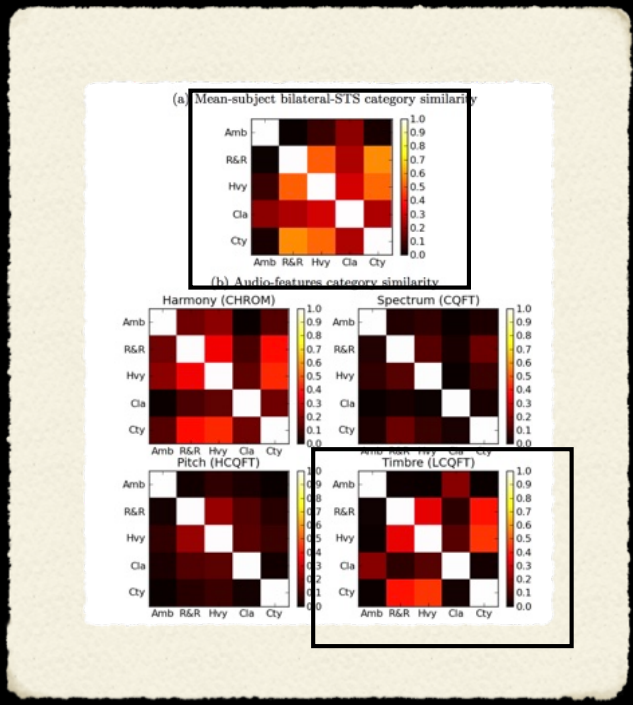


modelling: **decoding genre**



- MVPA restricted to bilateral **STS**
- highest correlation representation similarity analysis (RSA) matrix of **timbral features** (cepstral features - "the way the stimulus sounds") with that of **STS** than tonality and pitch
- "distributed population code in bilateral **STS** responsible for genre discrimintation"
- timbral features were most accurate in predicting the image response





Casey, M., Thompson, J., Kang, O., Raizada, R., and Wheatley, T. (2012). "Population codes representing musical timbre for high-level fmri categorization of music genres," in *Machine Learning and Interpretation in Neuroimaging* (Sierra Nevada: Springer), 34–41.

frontiers in Psychology

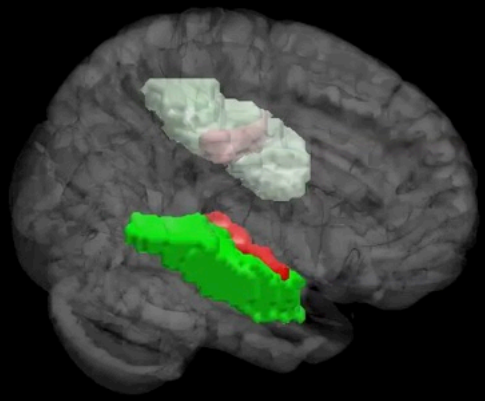
ORIGINAL RESEARCH
published: 14 July 2017
doi: 10.3389/fpsyg.2017.01179

Music of the 7Ts: Predicting and Decoding Multivoxel fMRI Responses with Acoustic, Schematic, and Categorical Music Features

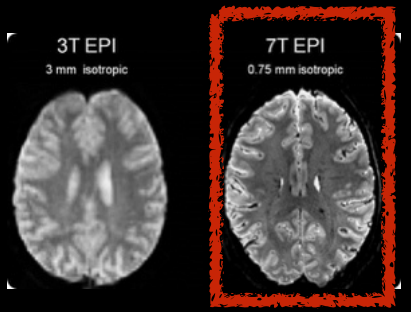
Michael A. Casey *

Bregman Music and Audio Lab, Computer Science and Music Departments, Dartmouth College, Hanover, NH, United States

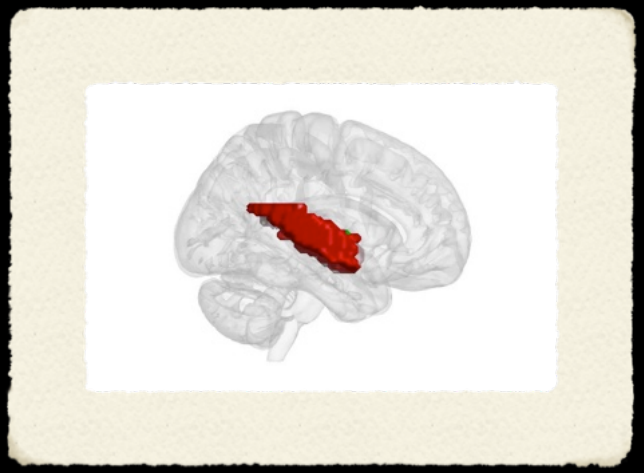
- song & genre decoding from fMRI
- Forrest database
  - 7 tesla scanner
  - 20 participants
  - 25 excerpts (6-sec), 5 genres
- ROIs: Heschl's gyrus, anterior STG, posterior STG
- Support Vector Machine classifier



modelling: **decoding genre**



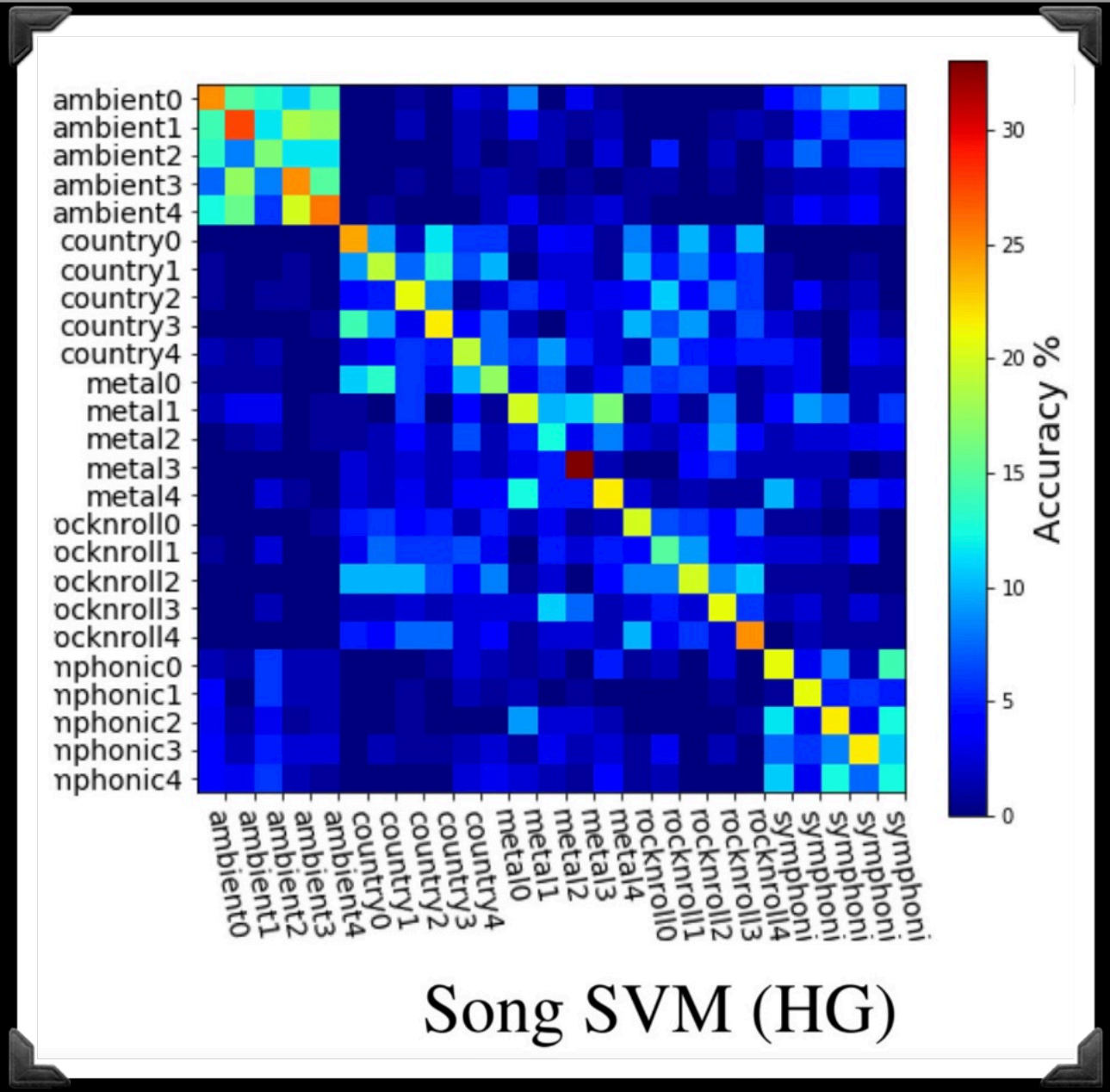
- restricted to bilateral **HG** and **STG**
- **STG** predicted better than **HG** perceptual genre classifications

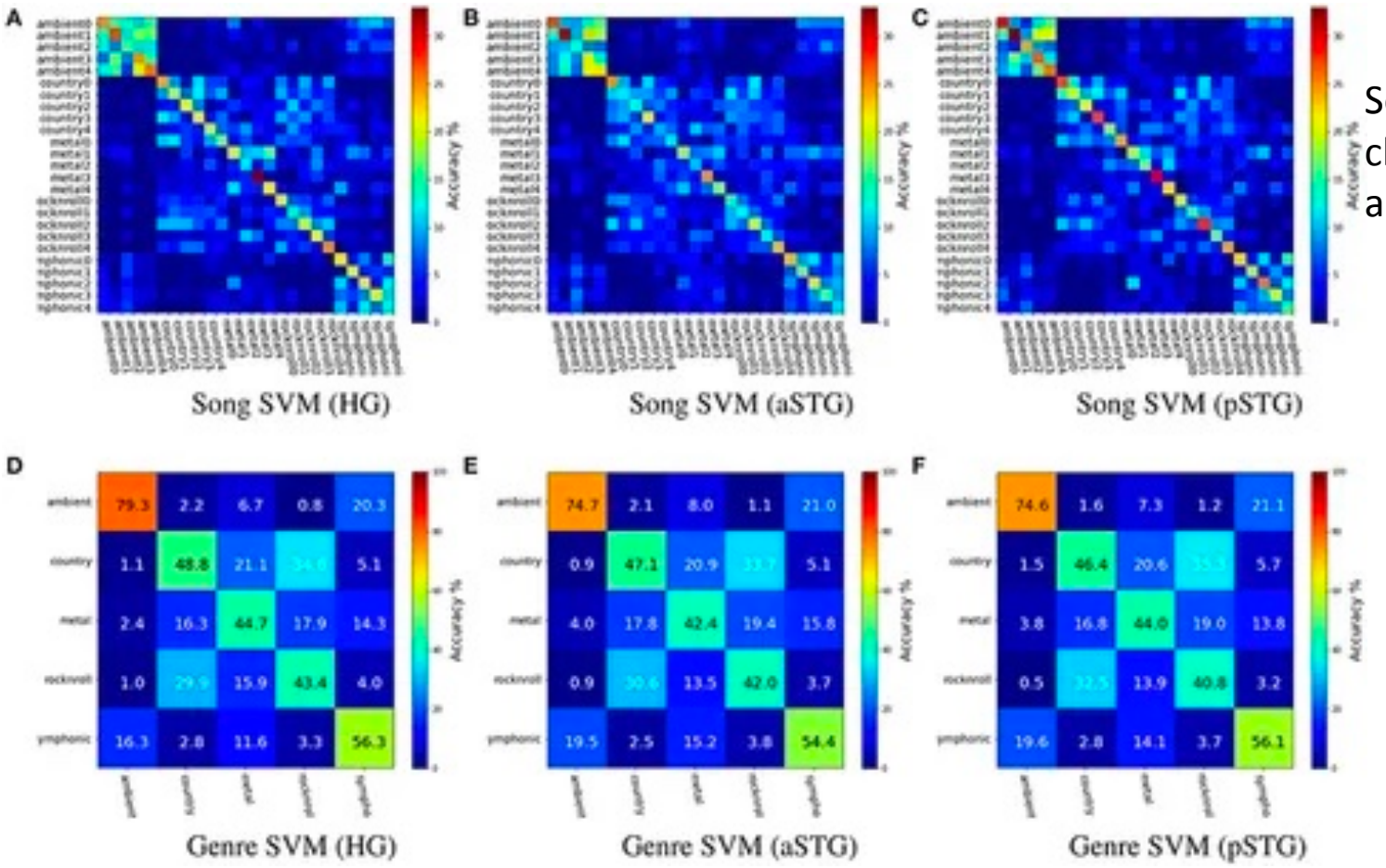


Primary and Secondary auditory cortex

Casey, M. A. (2017). Music of the 7Ts: Predicting and decoding multivoxel fMRI responses with acoustic, schematic, and categorical music features. Frontiers in Psychology, 8, 1–11. http://doi.org/10.3389/fpsyg.2017.01179

results / song classification





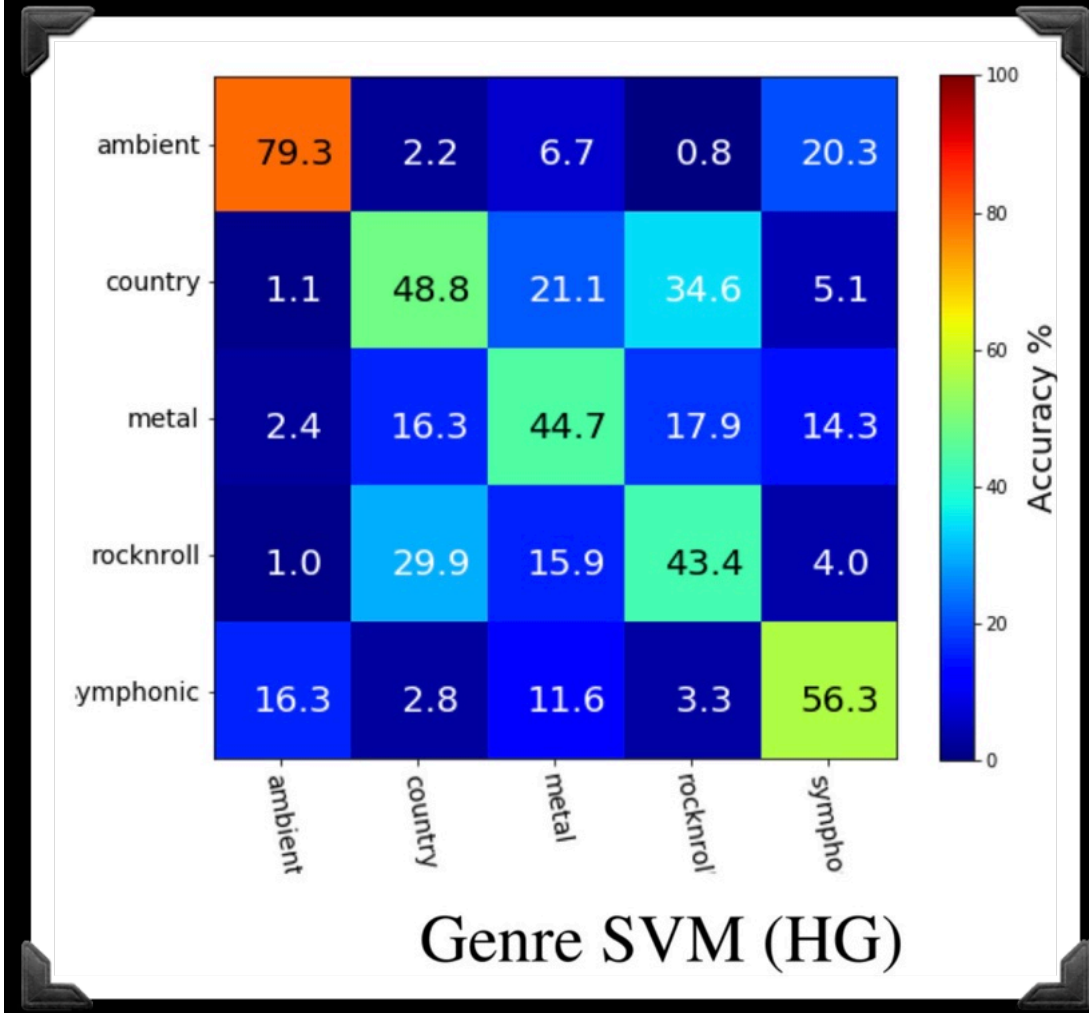
Songs mean classification accuracy = 15.95%

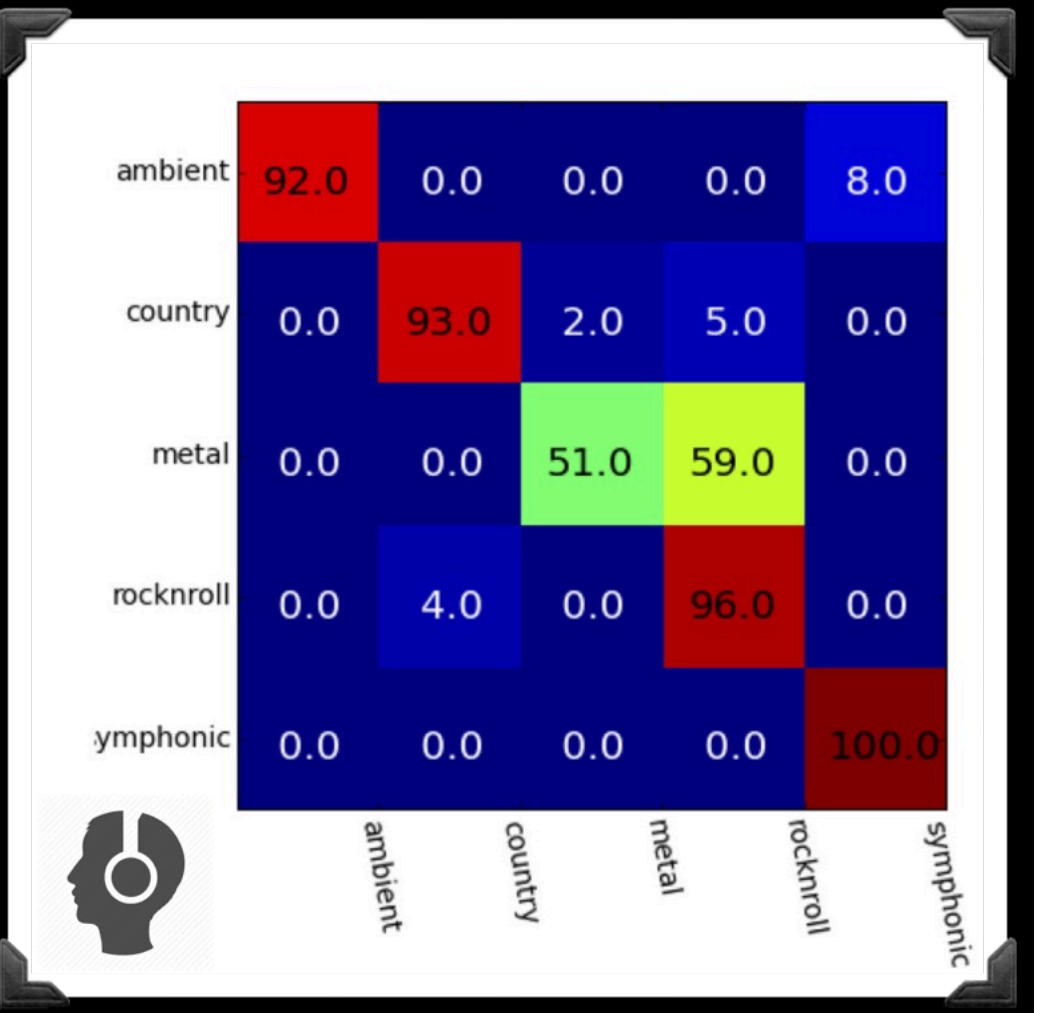
Genre mean classificatio n accuracy = 60%

The figures show that distinct anatomical ROIs yield similar relational information about music. The patterns of misclassifications show that when songs are misclassified, they are more likely to be confused with items from the same genre, or a similar sounding genre: e.g., Ambient and Symphonic; and Rock-n-Roll and Country.

Casey, M. A. (2017). Music of the 7Ts: Predicting and decoding multivoxel fMRI responses with acoustic, schematic, and categorical music features. Frontiers in Psychology, 8, 1–11. http://doi.org/10.3389/fpsyg.2017.01179

results / genre classification





—> the pattern of confusions exhibited by the classifiers was significantly correlated with confusions in a behavioral categorization task