Corpus Analysis Tools for Computational Hook Discovery

Jan Van Balen
John Ashley Burgoyne
Dimitrios Bountouridis
Daniel Müllensiefen
Remco C. Veltkamp

HOOKED ON MUSIC

Select a game

RECOGNISE THAT TUNE?

WHAT'S THE HOOK?

TIME TRIAL

IN A ROW

#HookedOnMusic

3 million data points
100,000+ players
1500 song sections
300 songs

#HookedOnMusic

3 million data points
100,000+ players
1500 song sections
300 songs

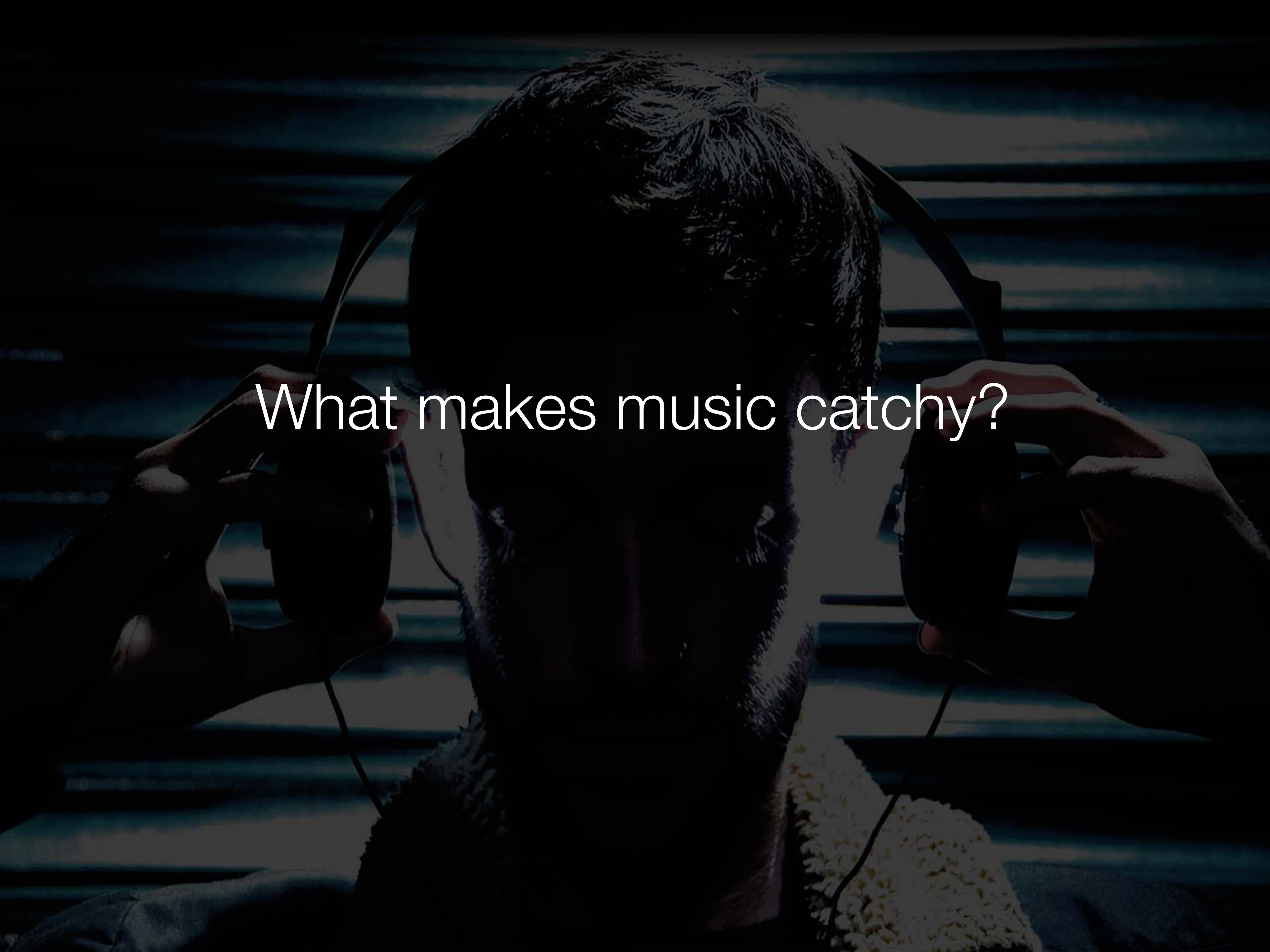
→ reliable measure of *recognizability* of each section

#HookedOnMusic

3 million data points
100,000+ players
1500 song sections
300 songs

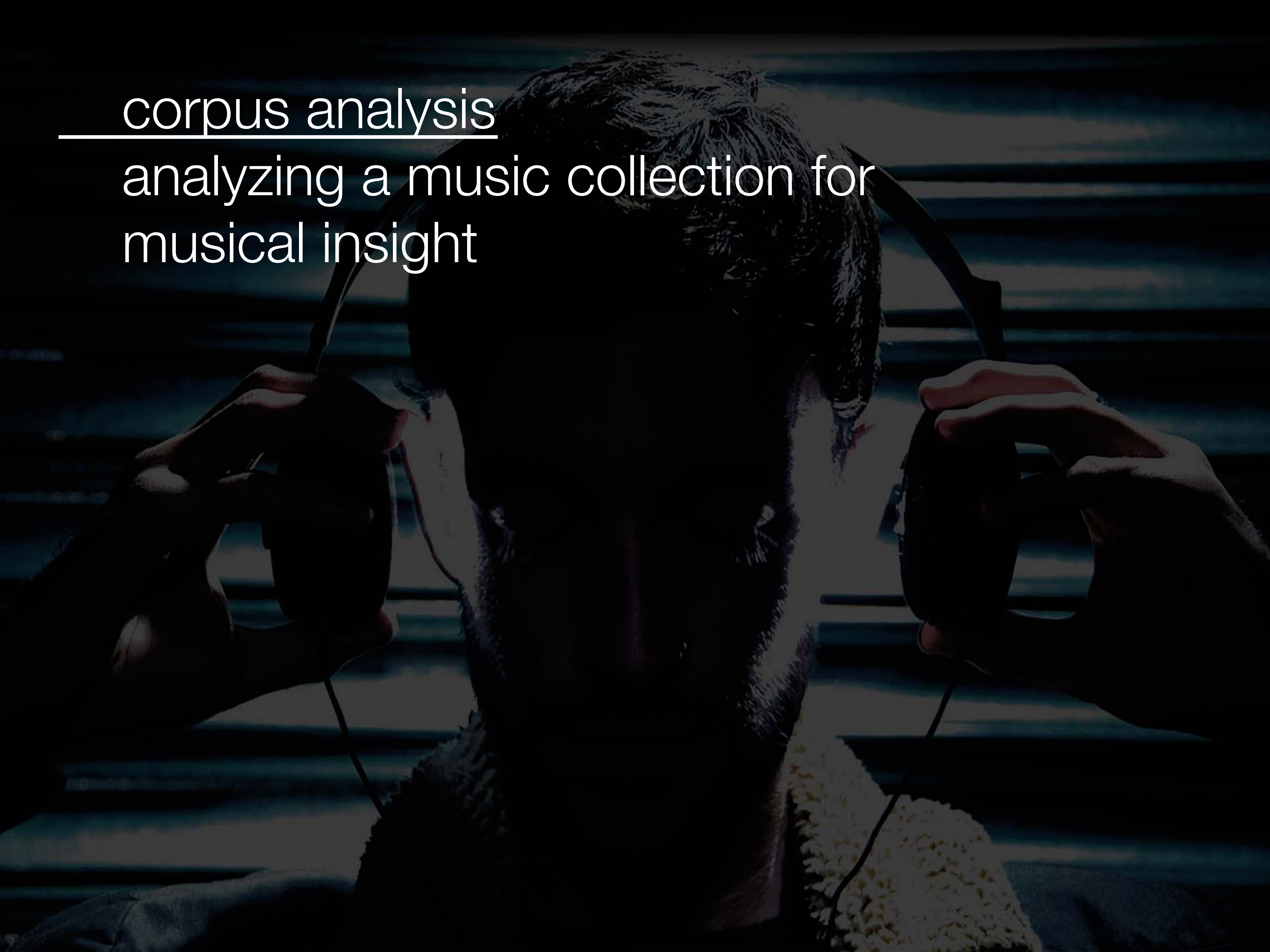
→ reliable measure of *recognizability* of each section

audio for each of these songs sections


What makes music catchy?

corpus analysis

analyzing a music collection for musical insight

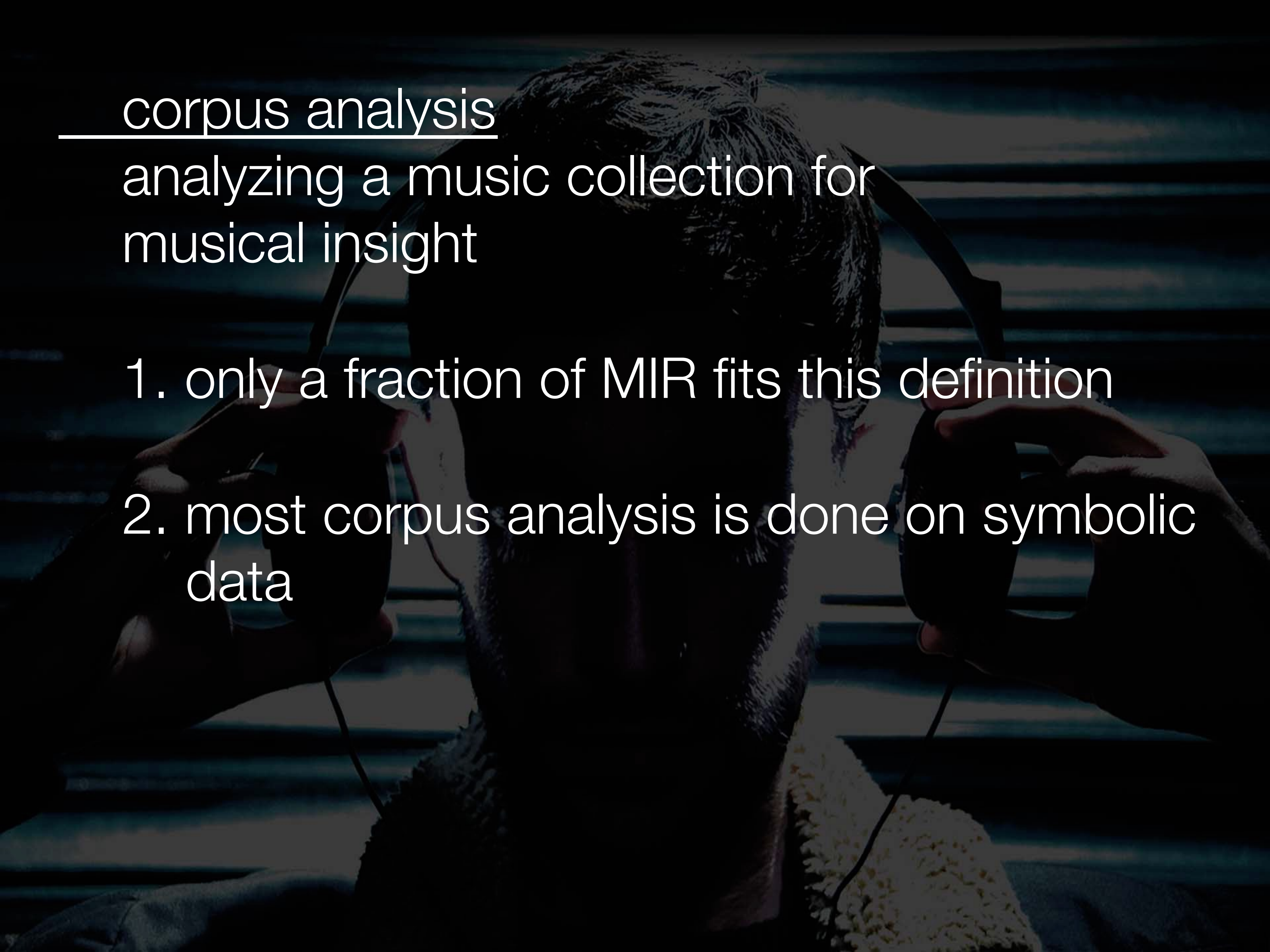


corpus analysis

analyzing a music collection for musical insight

- is popular music getting louder?
- is popular music getting less diverse?
- how has popular music harmony evolved?
- how does music affect walking speed?
- what makes melodies easy to remember?
- …

Deruty et al., Serra et al., Mauch et al., De Clercq et al.. Burgoyne et al., Leman et al., Müllensiefen & Halpern.



corpus analysis

analyzing a music collection for musical insight

1. only a fraction of MIR fits this definition
2. most corpus analysis is done on symbolic data

corpus analysis

analyzing a music collection for musical insight

1. only a fraction of MIR fits this definition

2. most corpus analysis is done on symbolic data

   lack of audio tools for corpus analysis

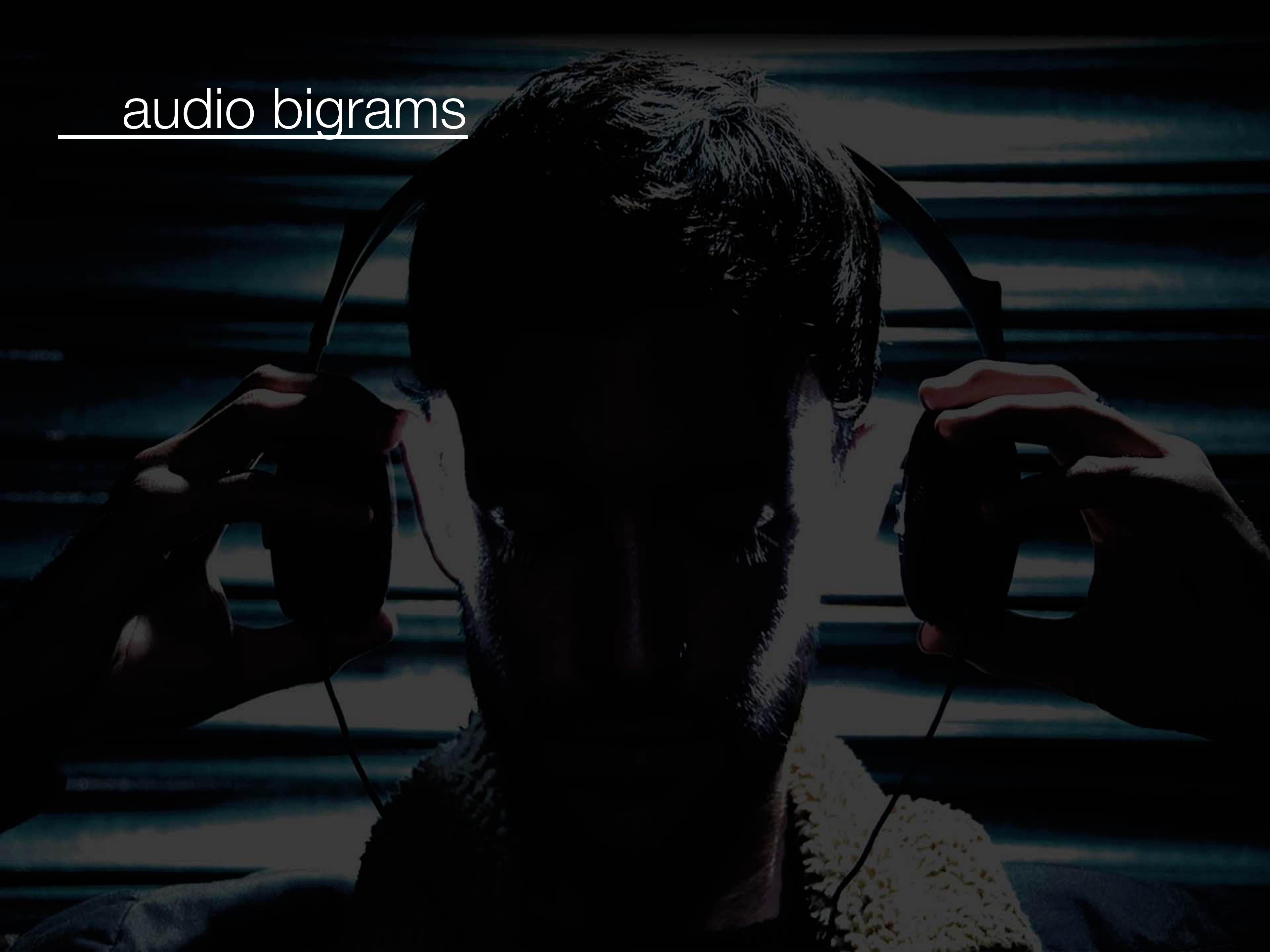
outline

three useful feature classes:

1. audio bigrams
2. second-order features
3. corpus- vs. song-based second-order features

test experiment

audio bigrams



audio bigrams

- bigrams = ordered pairs of things
  - (words, pitches, intervals)
  - common in
    - text retrieval, NLP
    - symbolic music processing

audio bigrams

bigrams = ordered pairs of things
- (words, pitches, intervals)
- common in
  - text retrieval, NLP
  - symbolic music processing

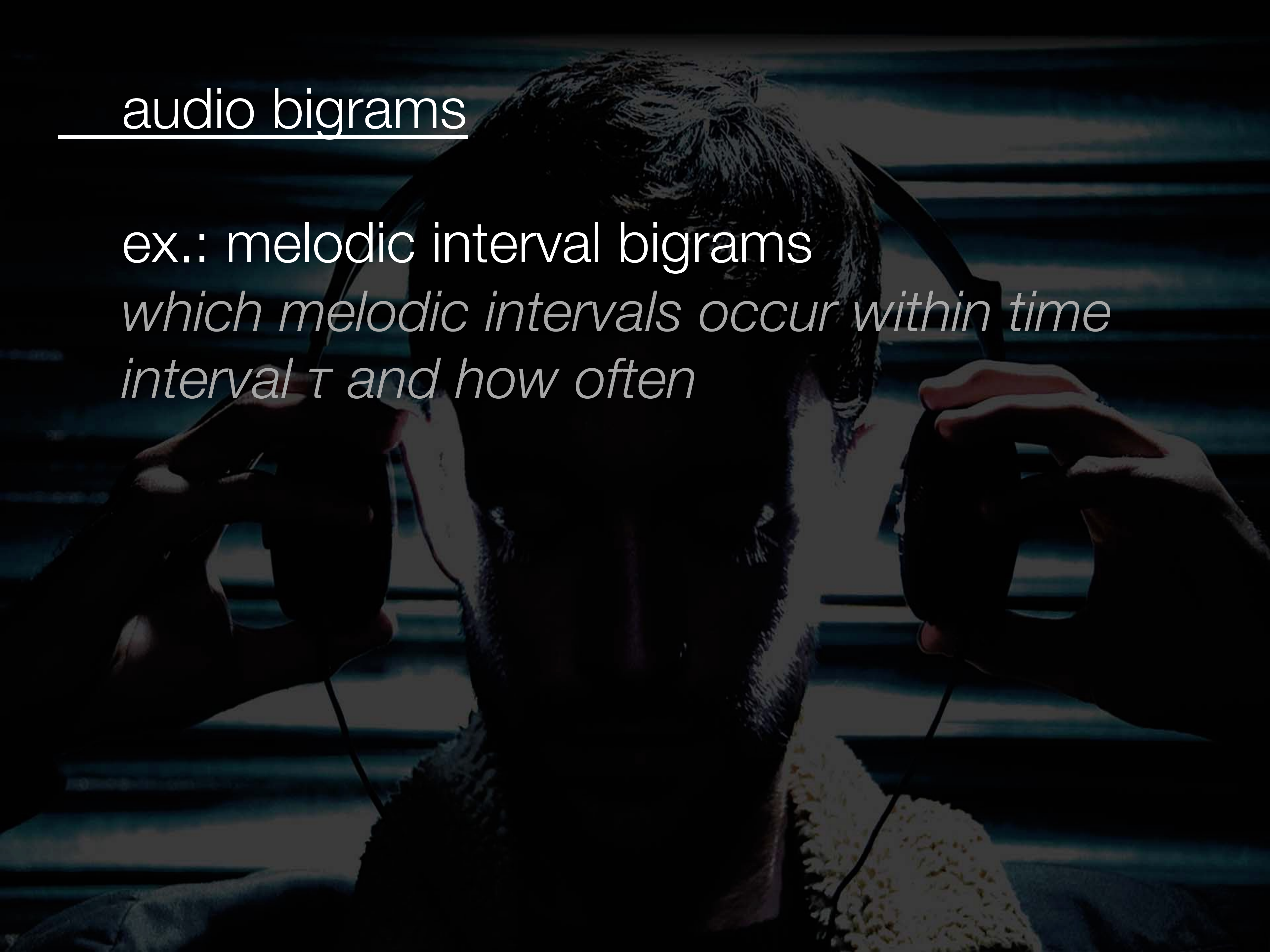
two nice properties many audio features don’t have:

1. incorporates a notion of order in time
2. can be counted

audio bigrams

ex.: melodic interval bigrams

which melodic intervals occur within time interval τ and how often



audio bigrams

ex.: melodic interval bigrams

which melodic intervals occur within time interval τ and how often

melody estimate $M_{t,\,l}$ (chroma-like 12-d array)

trigrams $T(i,j,k) = \Sigma_t \, max_\tau \, M_{t-\tau,\,i} * M_{t,\,j} * max_\tau \, M_{t+\tau,\,k}$

interval bigrams $T(i, j, k) \rightarrow T(k\text{-}i, k\text{-}j)$

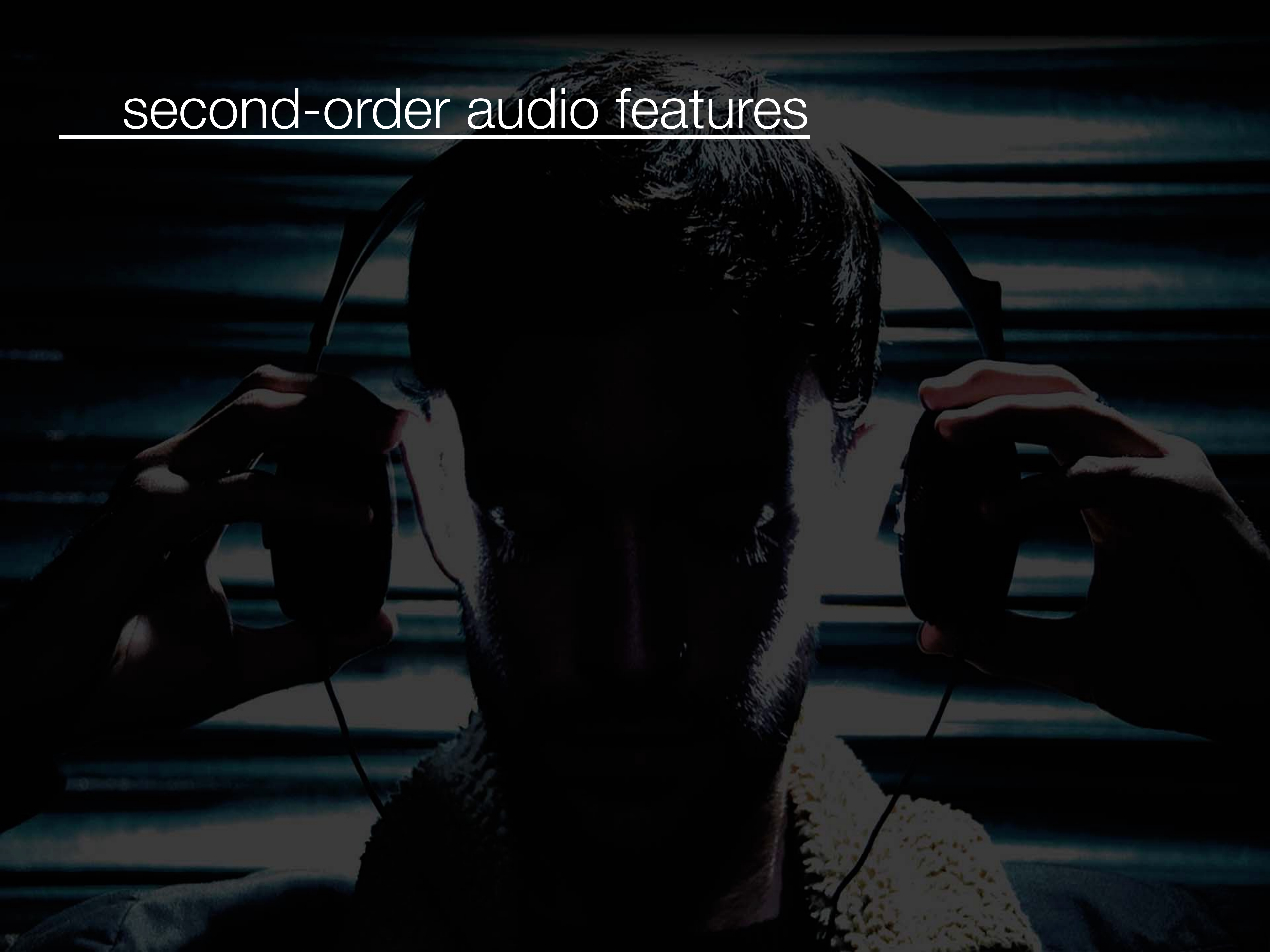
audio bigrams

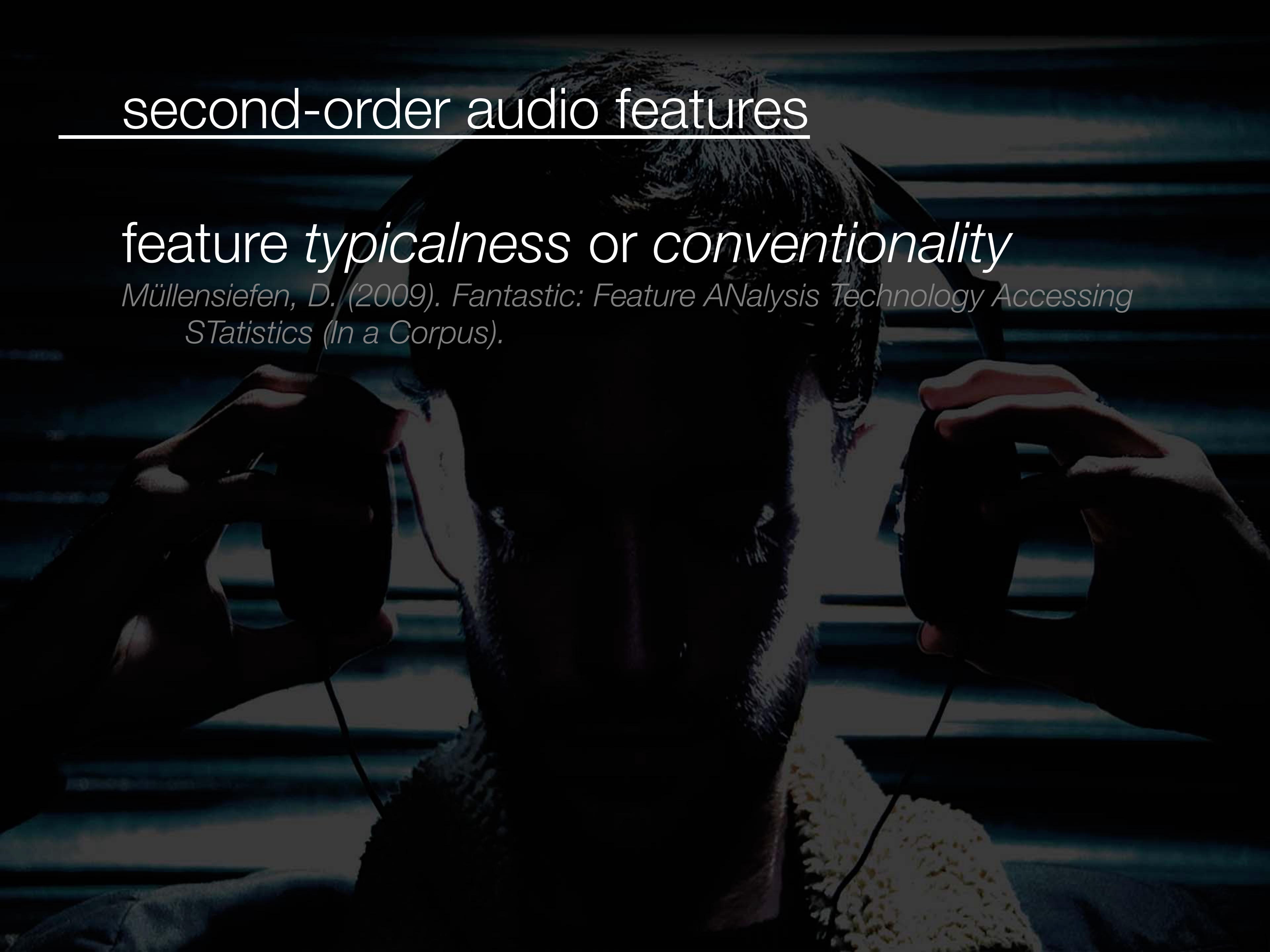
captures hamony and melodic structure
- can be used as a very simple fingerprint

essentially a distribution
- all kinds of statistics can be computed (entropy, KL divergence…)

doesn’t rely on chords

second-order audio features





second-order audio features

feature *typicalness* or *conventionality*

Müllensiefen, D. (2009). Fantastic: Feature ANalysis Technology Accessing STatistics (In a Corpus).

second-order audio features

feature *typicalness* or *conventionality*

Müllensiefen, D. (2009). Fantastic: Feature ANalysis Technology Accessing STatistics (In a Corpus).

more informative to humans than absolute feature values (‘a sharpness of 10.5’)

second-order audio features

feature *typicalness* or *conventionality*

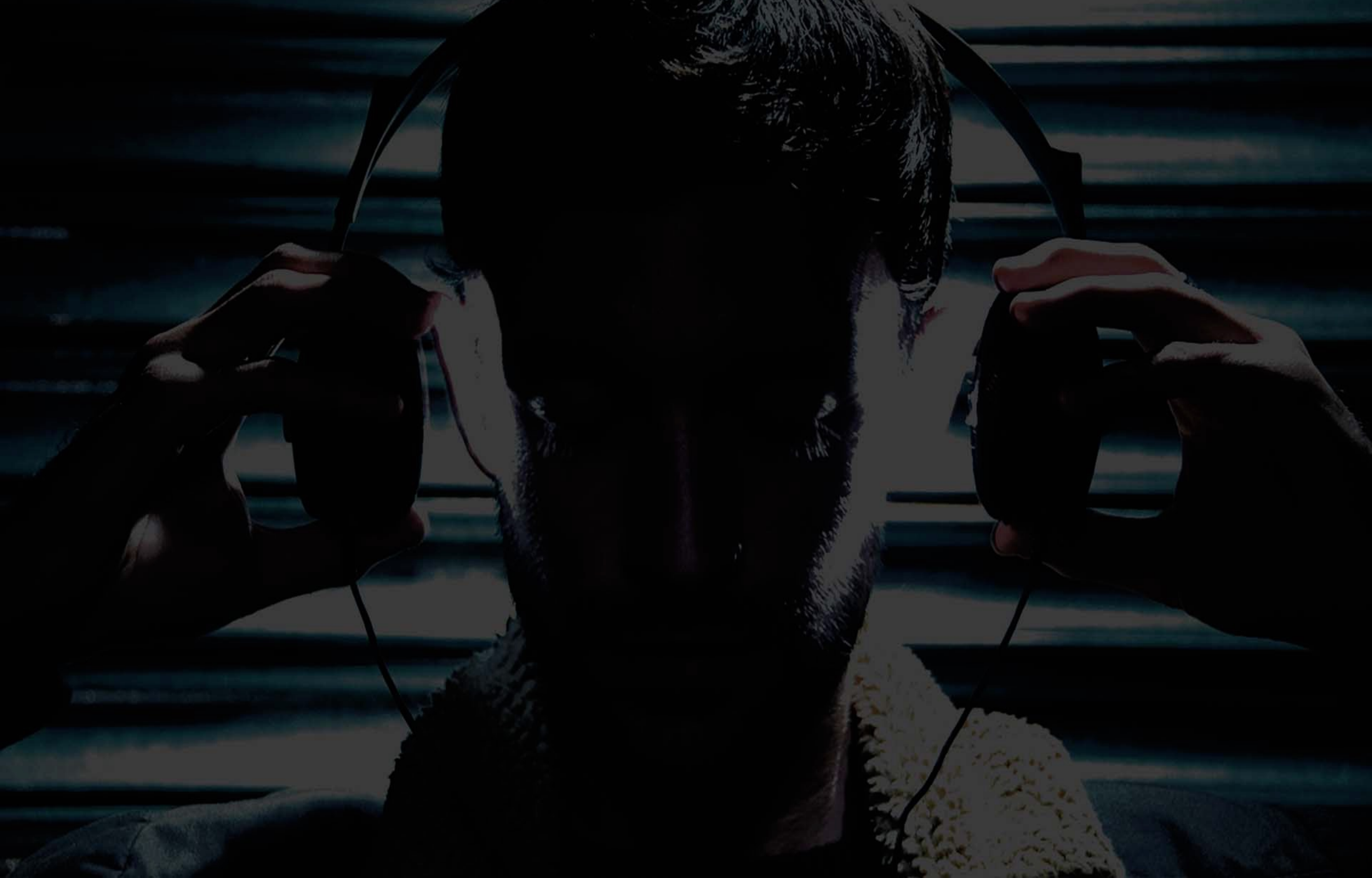
Müllensiefen, D. (2009). Fantastic: Feature ANalysis Technology Accessing STatistics (In a Corpus).

more informative to humans than absolute feature values (‘a sharpness of 10.5’)

defined for audio as:

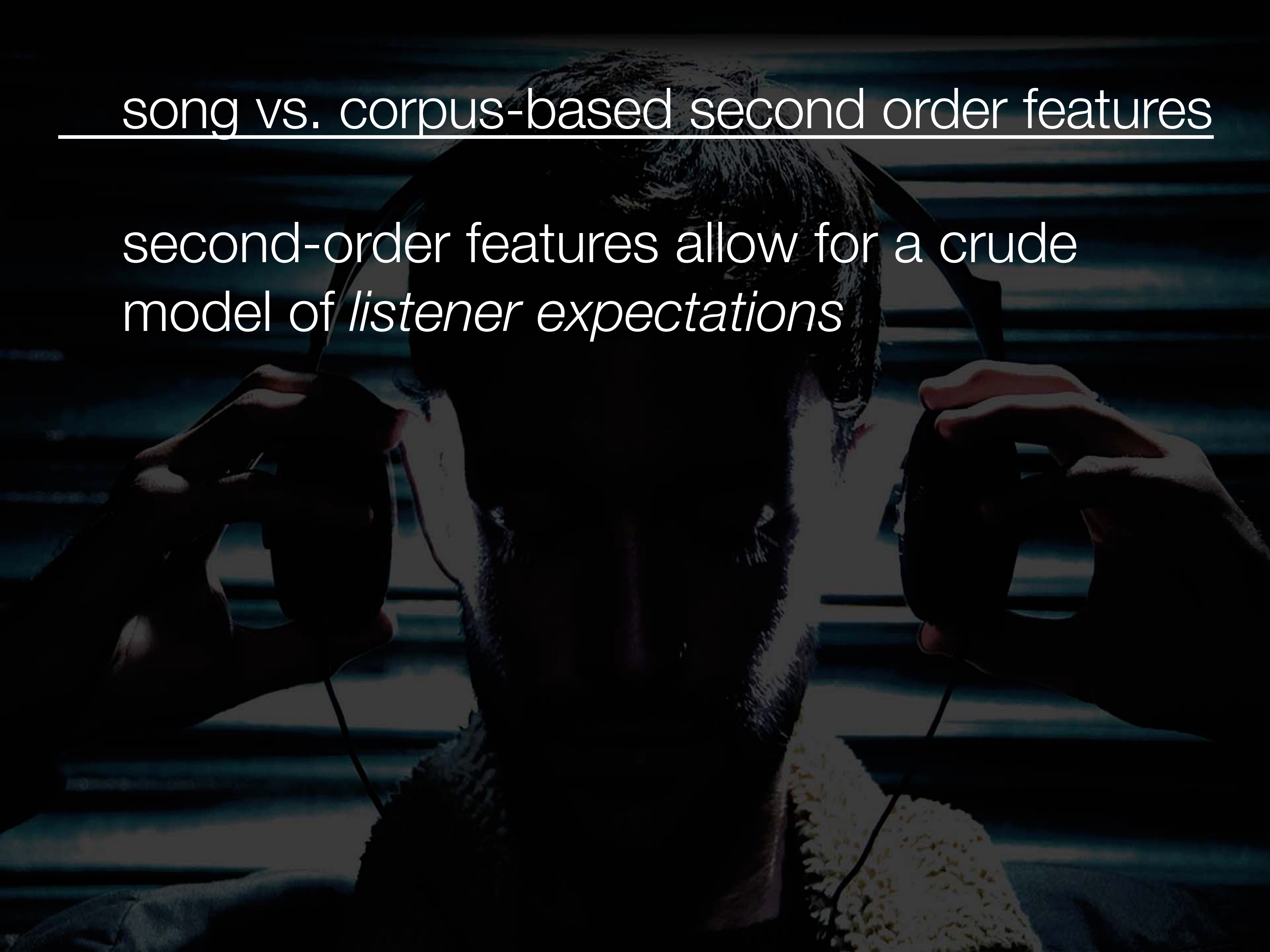
given a feature value, the probability of observing a more extreme *value in a larger reference corpus*

song vs. corpus-based second order features



song vs. corpus-based second order features

second-order features allow for a crude model of *listener expectations*



song vs. corpus-based second order features

second-order features allow for a crude model of *listener expectations*

veridical expectations

- expectations related to a particular work

schematic expectations

- broad generalizations from years of music listening

David Huron: Sweet Anticipation (2006), MIT Press

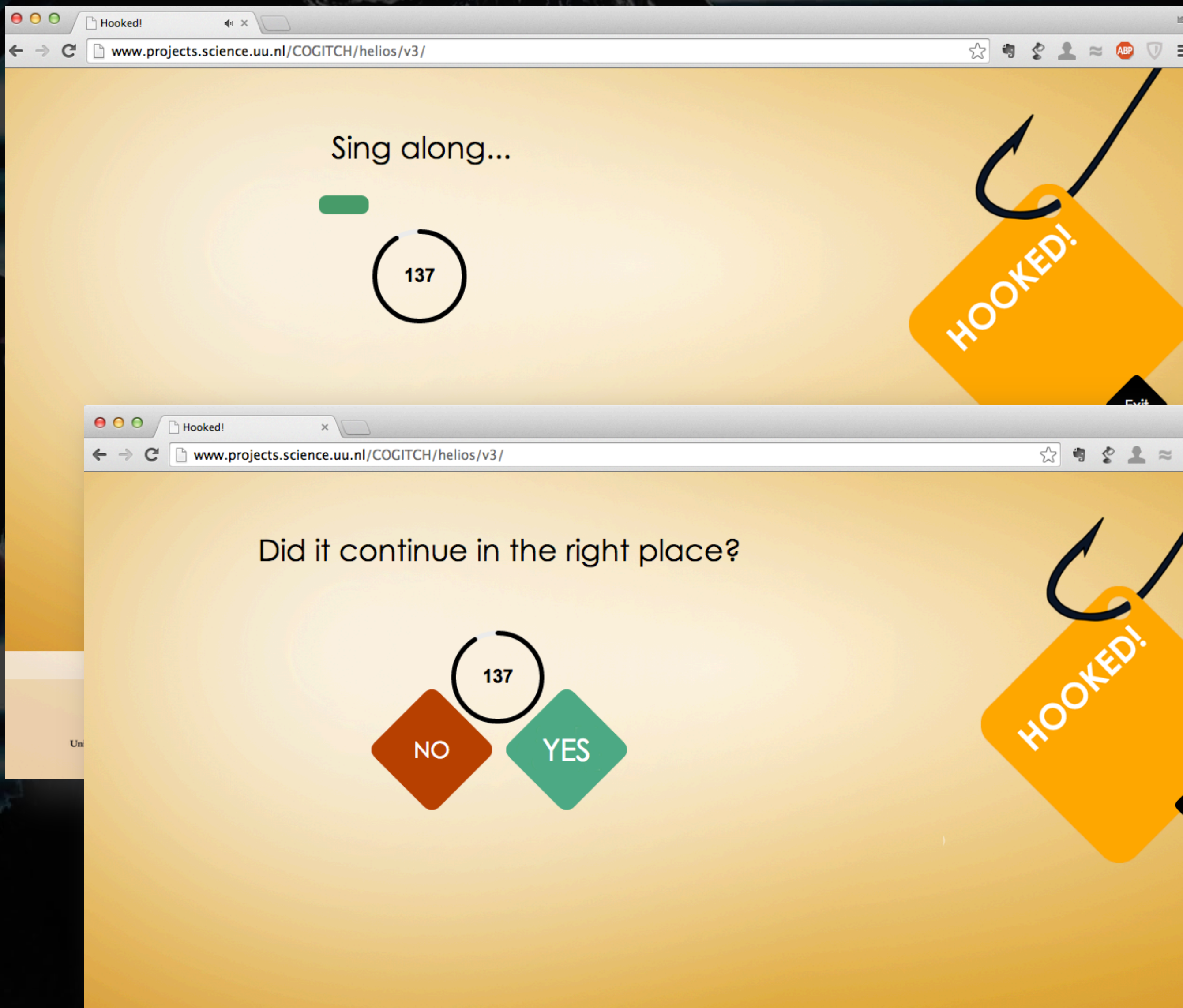
song vs. corpus-based second order features

for a song fragment,

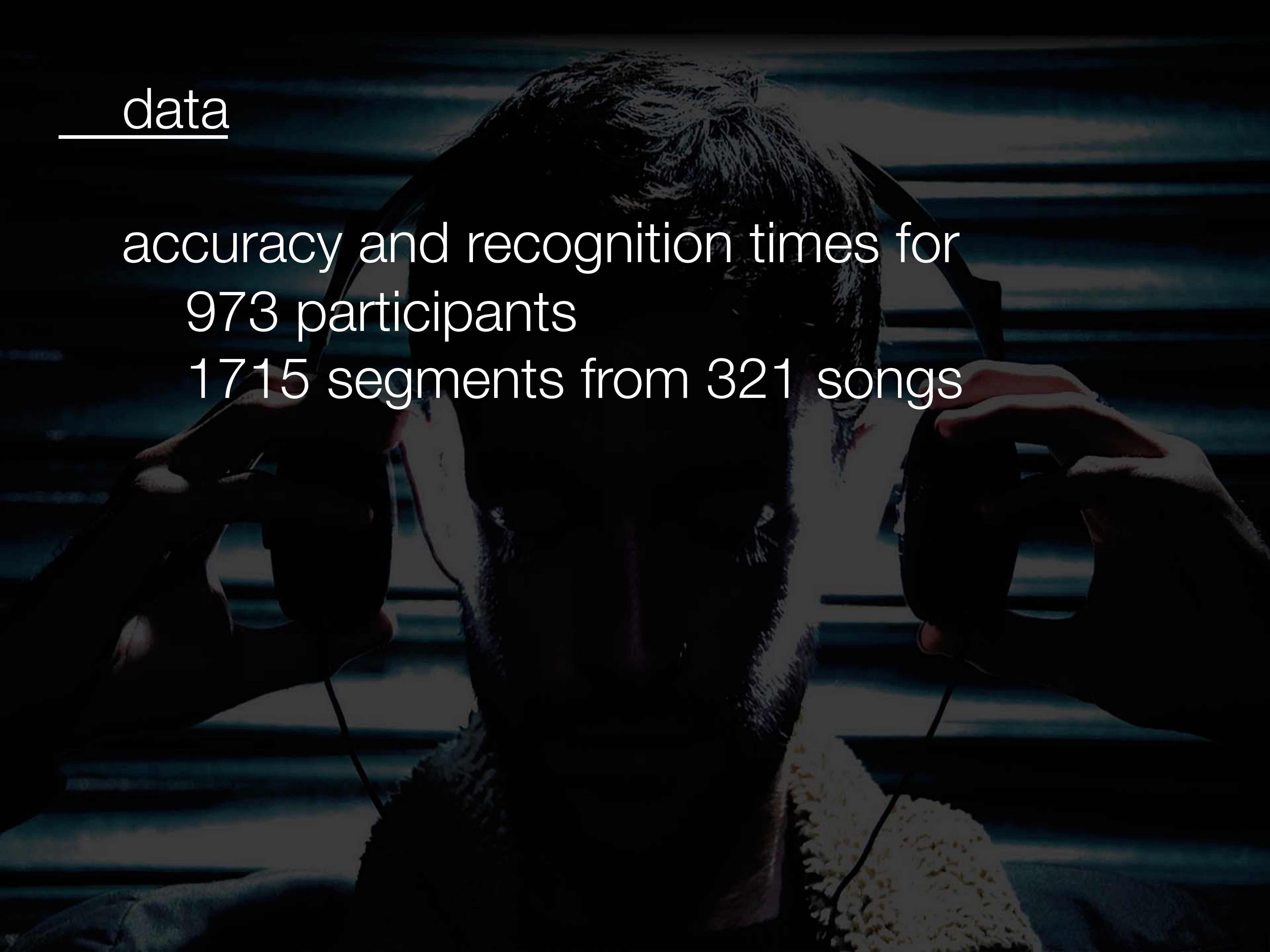
song-based second-order features:
corpus = fragments from the same song
~ veridical expectations

corpus-based second-order features:
corpus = fragments from other songs
~ schematic expectations

test experiment



www.hookedgame.nl



data

accuracy and recognition times for

- 973 participants
- 1715 segments from 321 songs

features

	first-order	second -order
pitch height, pitch range	mean	conventionality, repetition
3 x audio bigrams (melody, harmony)	entropy	conventionality, repetition
4 x psycho-acoustic features	mean, variance	conventionality, repetition
MFCC	total variance	conventionality, repetition

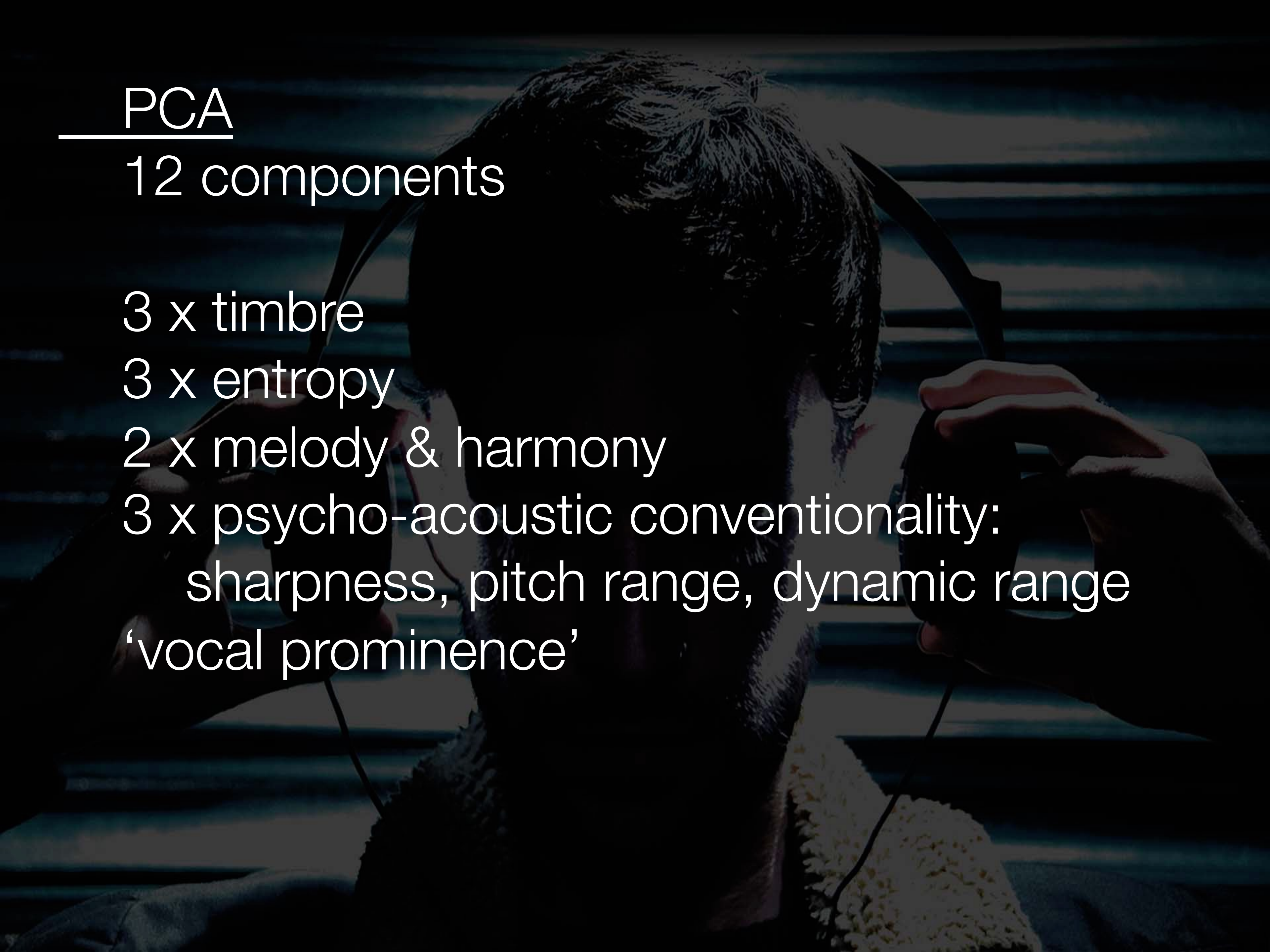
analysis method

- PCA

- mixed-effects regression on 1715 sections

  random intercepts per song
  coefficients based on *within-song* effects

  controls for several confounding factors
  (genre, tempo, recency, airplay…)



PCA

12 components

3 x timbre
3 x entropy
2 x melody & harmony
3 x psycho-acoustic conventionality:
sharpness, pitch range, dynamic range
‘vocal prominence’

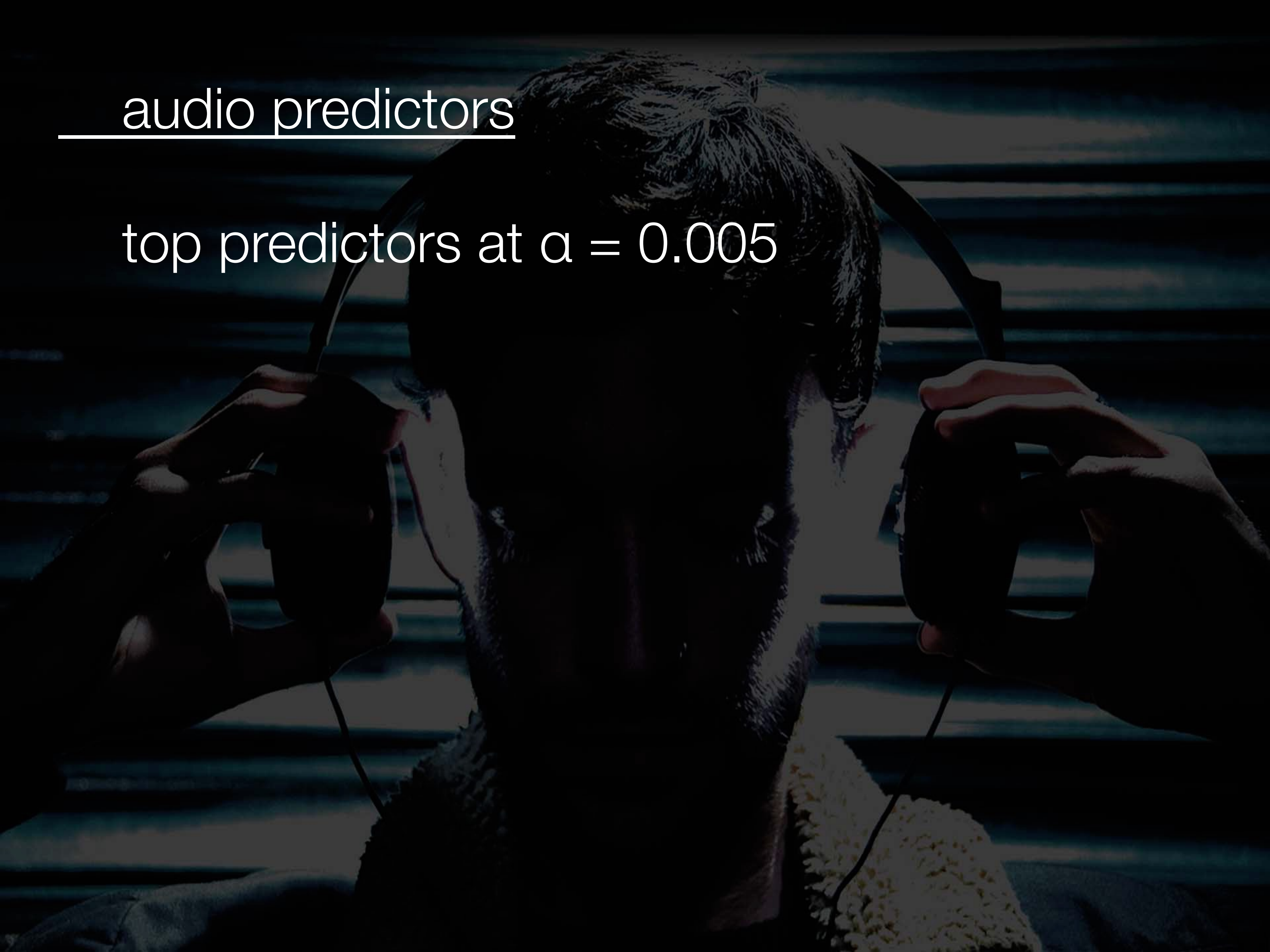
prediction

audio features compare favorably to symbolic features (Fantastic Toolbox) on the same task

	audio	symbolic	audio + symbolic
$R_{marginal}^2$	.06	.07	.10
$R_{conditional}^2$	.46	.47	.47

audio predictors

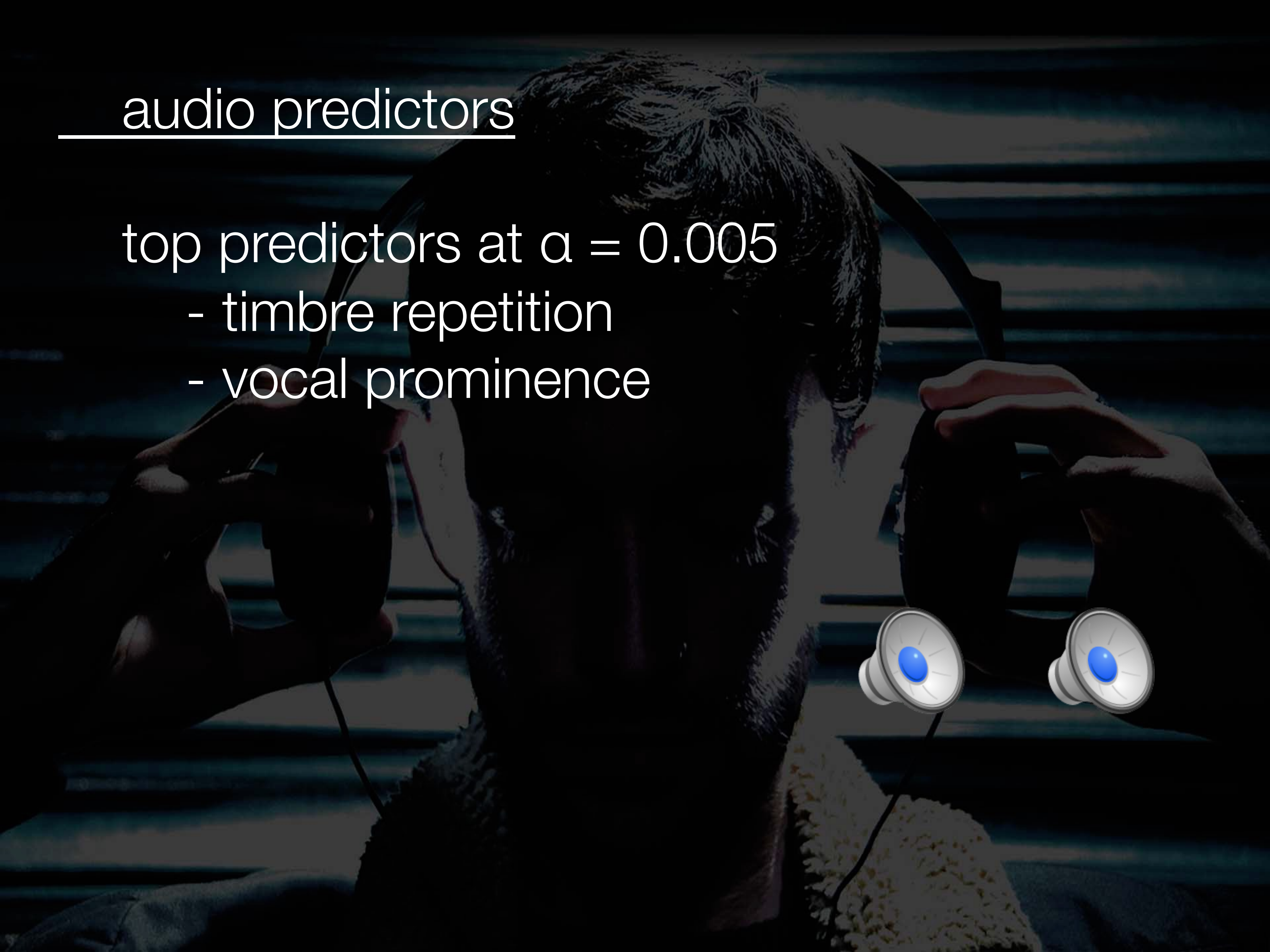
top predictors at $\alpha = 0.005$



audio predictors

top predictors at α = 0.005

- timbre repetition


audio predictors
top predictors at α = 0.005
- timbre repetition
- vocal prominence

audio predictors

top predictors at α = 0.005

- timbre repetition
- vocal prominence
- 6 kinds of conventionality
  ex.: melodic range

summary

three useful feature classes:

1. audio bigrams
2. second-order features
2. corpus- vs. song-based second-order features

conventionality, a prominent melody, and repetition make music catchy.

CATCHY toolbox

second-order audio features

inspired by a toolbox for symbolic music analysis (FANTASTIC) and LSA

e.g.: non-parametric, one dimension

$z = \text{rank}(f(x) / N)$

f estimated from size N corpus