Western listeners’ perception of music and speech is reflected in acoustic and semantic descriptors

Lauren Fink[1,2], Madita Hörster[3,4], David Poeppel[2,4,5,6], Melanie Wald-Fuhrmann[1,2], Pauline Larrouy-Maestri[1,2,4]

[1] Music Dept., Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany [2] Max-Planck-NYU Center for Language, Music, and Emotion (CLaME), New York, USA & Frankfurt am Main, Germany
[3] Department of Psychology, Ludwig-Maximilians-University, Munich, Germany [4] Neuroscience Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[5] Psychology Department, New York University, New York, USA [6] Ernst Struengmann Institute for Neuroscience, Frankfurt am Main, Germany

Max Planck Institute for Empirical Aesthetics

Background

Listeners show remarkable abilities when asked whether a sound should be classified as music or speech.

Our previous work (Durojaye et al., 2021):

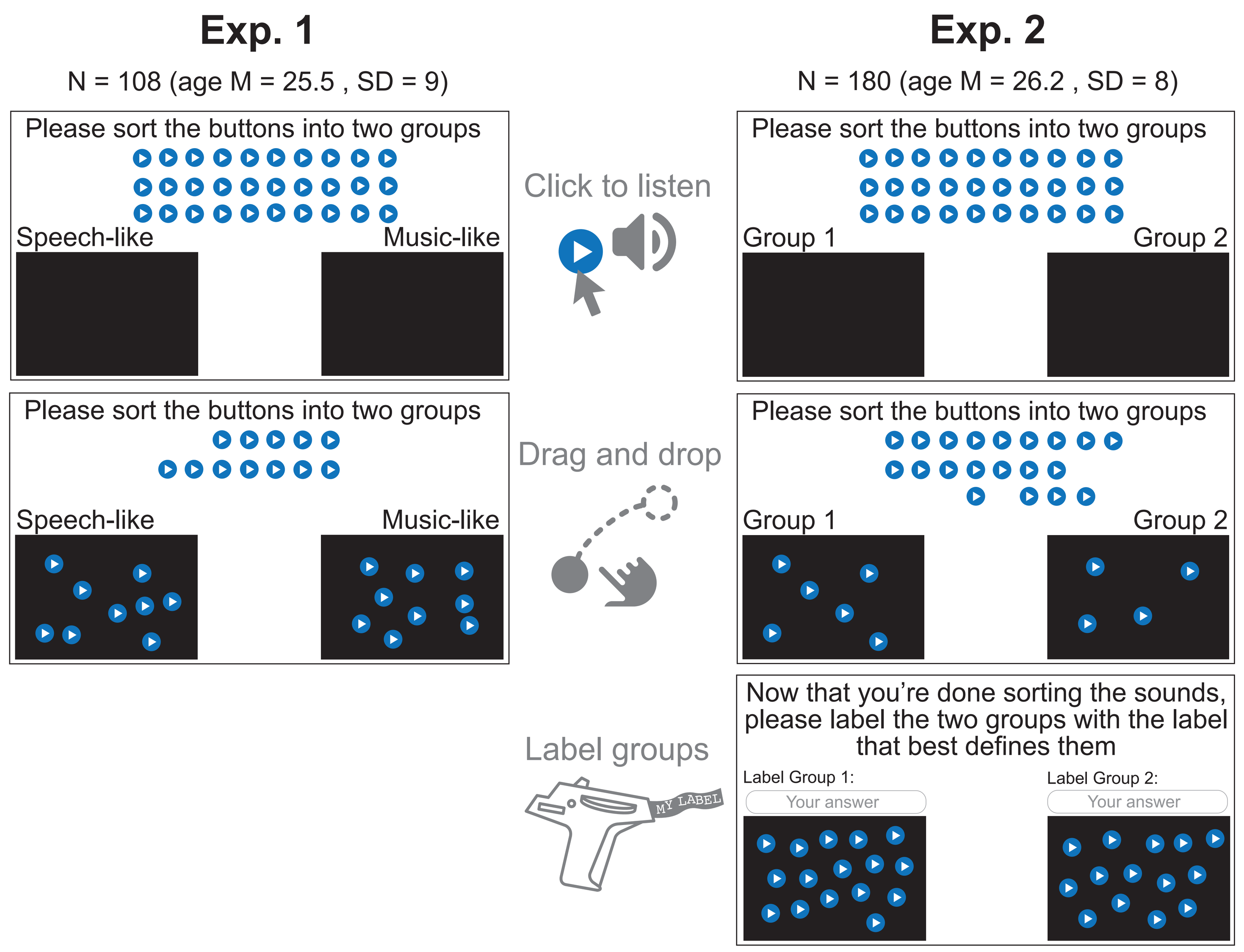
- used 6-10 sec recordings of Nigerian dùndún talking drum performances that were intended to be speech or music.
- a categorization task: is the sequence music- or speech-like?

We found: familiarity and acoustic features shape listeners’ categorizations. However, even unfamiliar participants could categorize above chance whether the drum was talking or playing music.

BUT the labels “speech” and “music” were given to participants, whereas categorization of our auditory environment is usually label-free.

HERE we explore the role of task demands and acoustic features in predicting participants’ categorization.

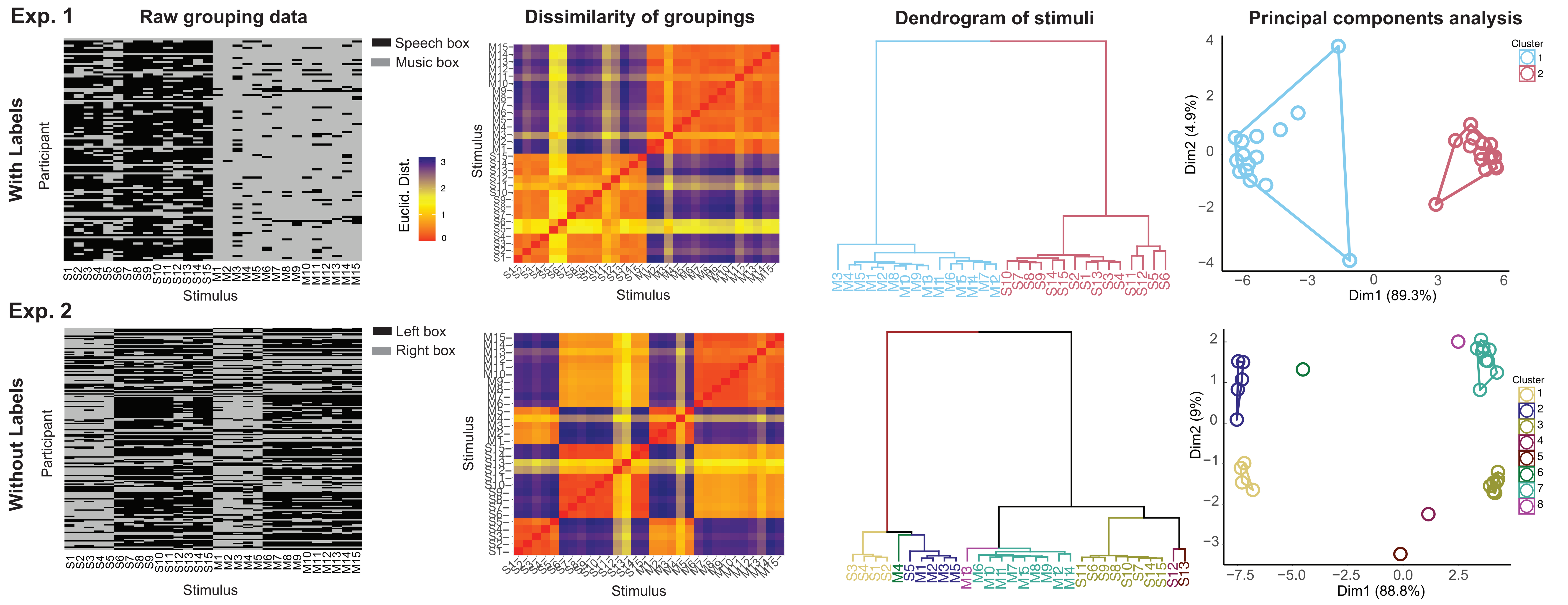
Methods



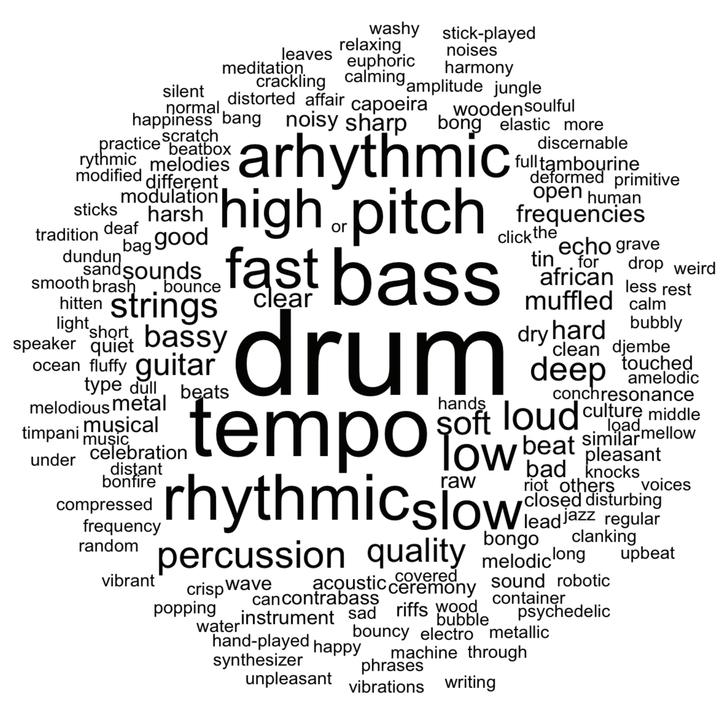
Material. Cleaned versions (removed background noise, clipping, etc.) of the recordings used in Durojaye et al. (2021).

Feature extraction. Pitch, spectral entropy (timbre), amplitude envelope (intensity), inter-onset-intervals (IOI), ratio of IOIs, amplitude modulation spectrum (AMS) peak, and pulse clarity, were calculated using custom scripts and third-party toolboxes in MATLAB.

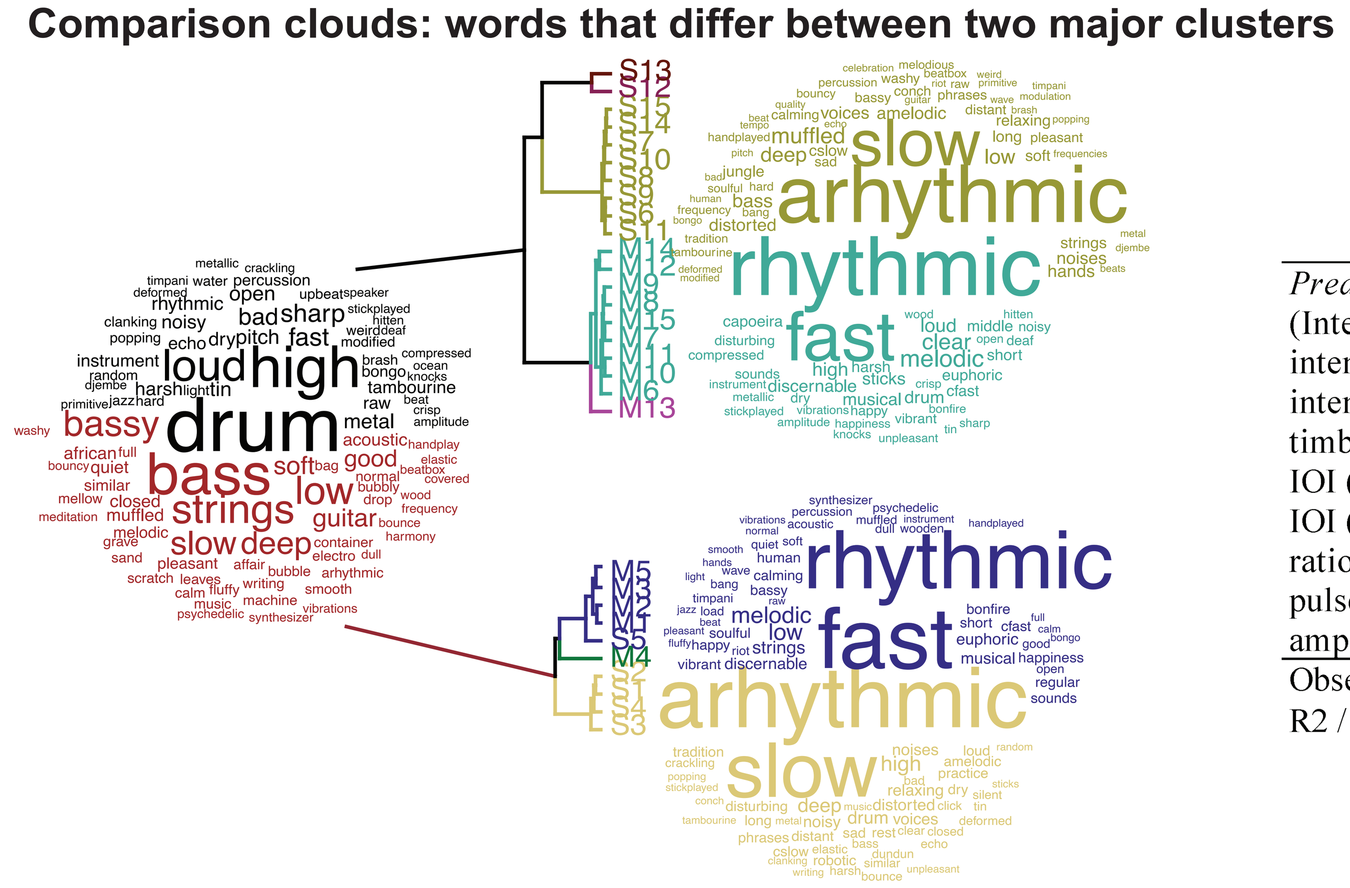
Results



Word cloud from all participant-provided labels



Comparison clouds: words that differ between two major clusters



Acoustic predictors of stimulus position in PCA space

	Exp. 2, Dim1 position			Dim2 position		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	70.82	-105.35 – 247.00	0.413	19.36	-34.58 – 73.30	0.464
intensity (mean)	**-10.18**	**-19.78 – -0.57**	**0.039**	-2.71	-5.65 – 0.23	0.069
intensity (difference)	1,11	-6.84 – 9.05	0.775	1,93	-0.50 – 4.36	0.114
timbre (mean)	**-75.53**	**-96.56 – -54.51**	**<0.001**	-2.42	-8.85 – 4.02	0.443
IOI (mean)	**0.03**	**0.01 – 0.06**	**0.016**	-0.01	-0.01 – 0.00	0.148
IOI (difference)	-0.03	-0.07 – 0.00	0.073	**-0.01**	**-0.02 – -0.00**	**0.036**
ratio (mean)	-65.01	-407.27 – 277.26	0.697	-31.34	-136.13 – 73.45	0.541
pulse clarity	-6.3	-22.44 – 9.85	0.426	0.5	-4.45 – 5.44	0.836
amp. mod. spectrum peak	-0.11	-0.60 – 0.38	0.647	0.08	-0.07 – 0.22	0.301
Observations	30			30		
R2 / R2 adjusted	0.843 / 0.783			0.854 / 0.798		

Discussion

- Participants categorize well above chance which stimuli fall into speech or music categories (replication of Durojaye et al., 2021). BUT this speech/music distinction is not the most salient one.
- When no labels are presented, participants first tend to form mixed groups of speech-like and music-like stimuli, along timbral and intensity dimensions.
- The speech/music distinction emerges on a lower hierarchical level; it is associated with labels like “arhythmic” / “rhythmic” and is predicted by timing characteristics.
- Participant labels converge with acoustic predictors.

Poster PDF

Contact: LKF
Lauren Fink
lauren.fink@ae.mpg.de
website: lkfink.github.io

Contact: PLM
Pauline Larrouy-Maestri
plm@ae.mpg.de
website: pauline-lm.github.io

Reference Durojaye*, C., Fink*, L., Roeske, T., Wald-Fuhrmann, M., & Larrouy-Maestri, P. (2021). Perception of Nigerian dùndún talking drum performances as speech-like vs. music-like: The role of familiarity and acoustic cues. *Frontiers in psychology, 12*, 1760.

CLaME NYU MAX PLANCK GESELLSCHAFT