A label-free task reveals semantic and acoustic features underlying speech-music categories

MAX PLANCK INSTITUTE
FOR EMPIRICAL AESTHETICS

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

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

Background

Listeners show remarkable abilities when asked whether a sound should be classified as music or speech but the mechanisms underlying this ability are speculative.

Our previous work [1]:

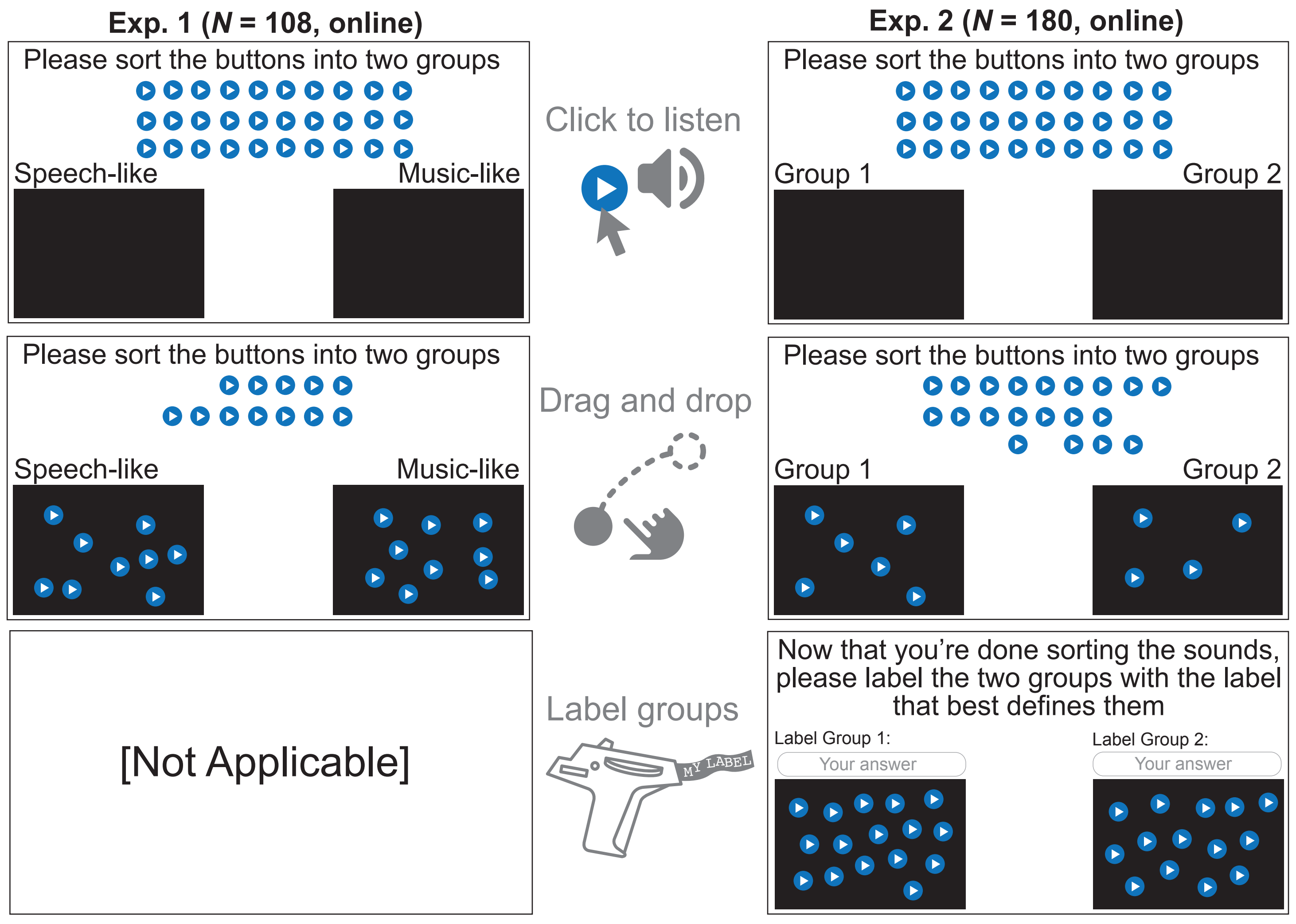
- used 6-10 sec recordings of Nigerian dùndún talking drum performances that were intended to be speech or music
- a categorization task: is the sequence music- or speech-like?
- a cross-cultural approach: Nigerian and familiar with dùndún vs. not

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 depart from the usual experimental procedure and explore the role of task demands and acoustic features in predicting naive participants' categorization.

Methods



Exp. 1 (*N* = 108, online)

Exp. 2 (*N* = 180, online)

Please sort the buttons into two groups

Speech-like | Music-like

Group 1 | Group 2

Click to listen

Drag and drop

Label groups

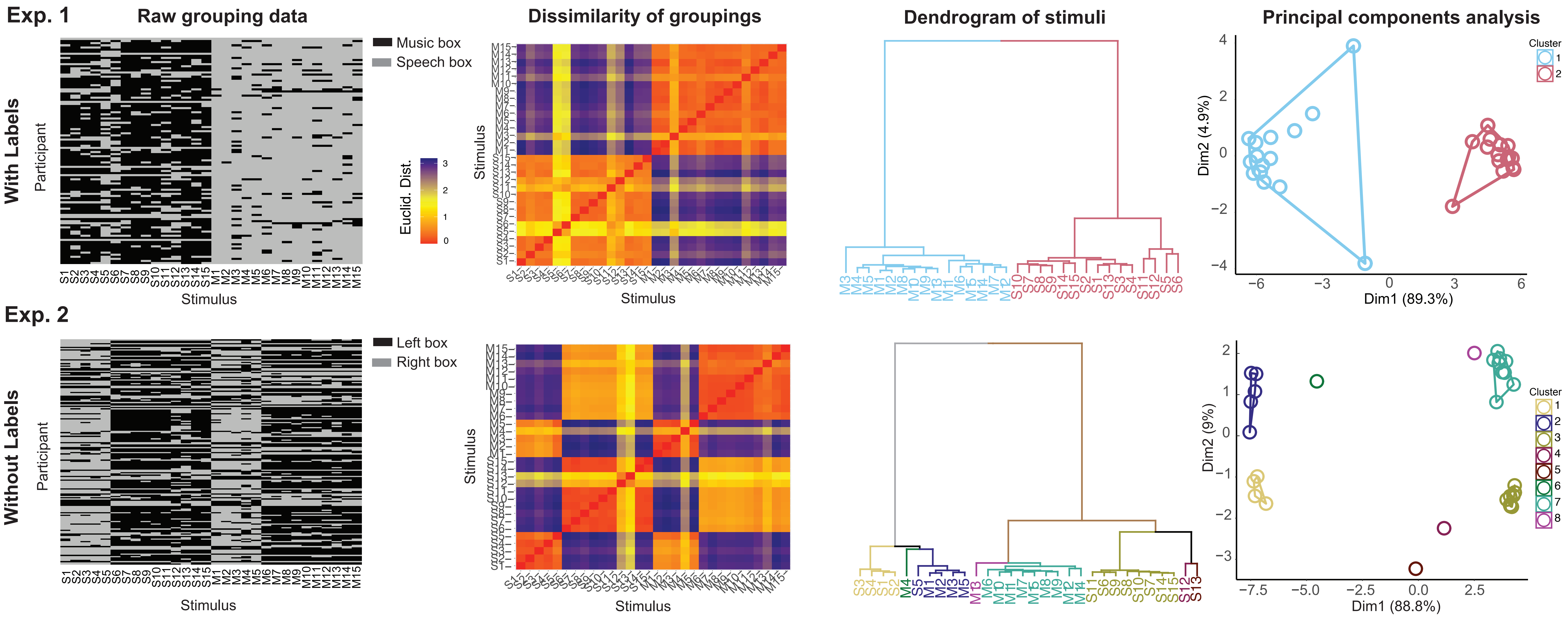
[Not Applicable]

Now that you're done sorting the sounds, please label the two groups with the label that best defines them

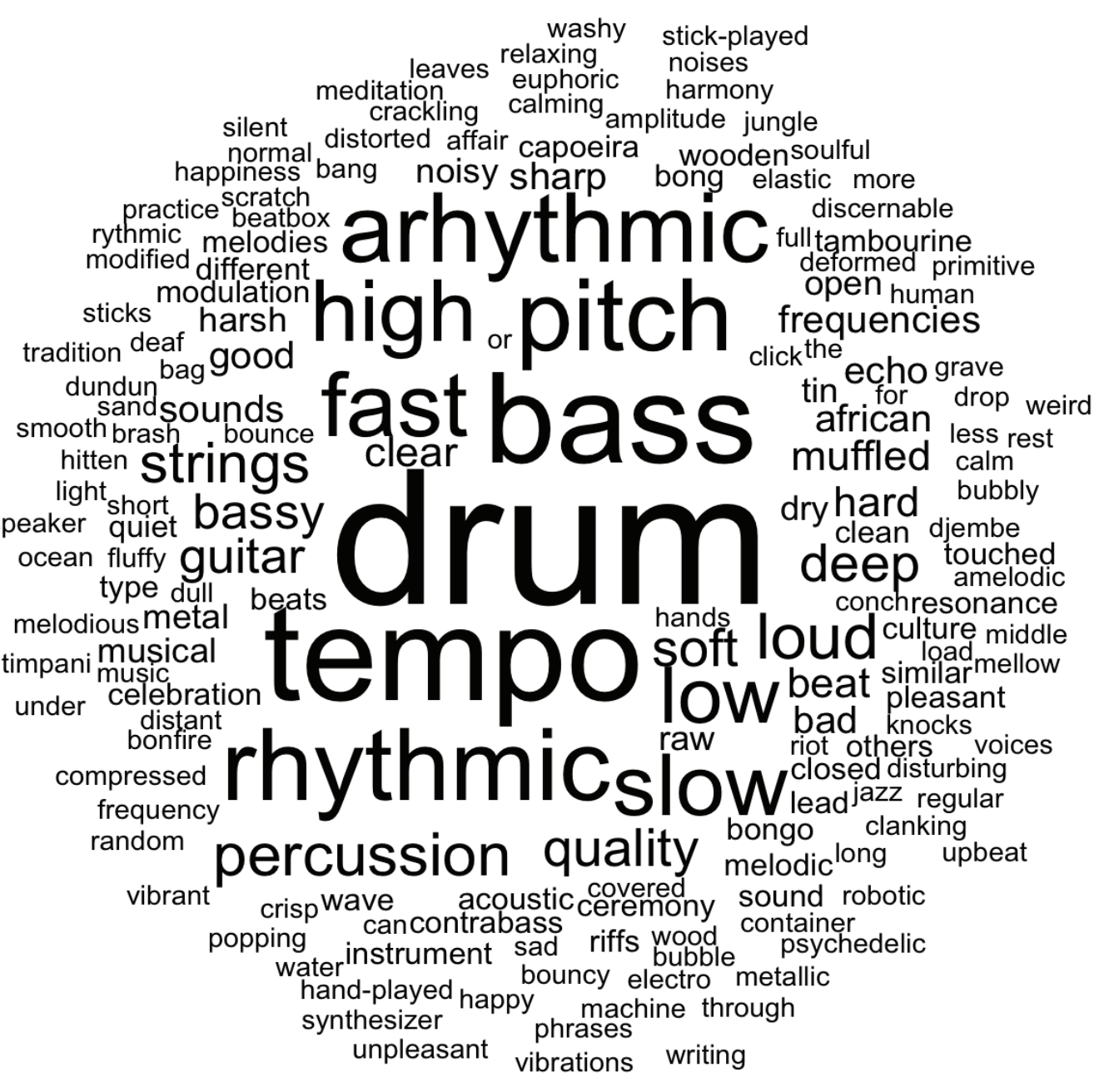
Label Group 1: Your answer

Label Group 2: Your answer

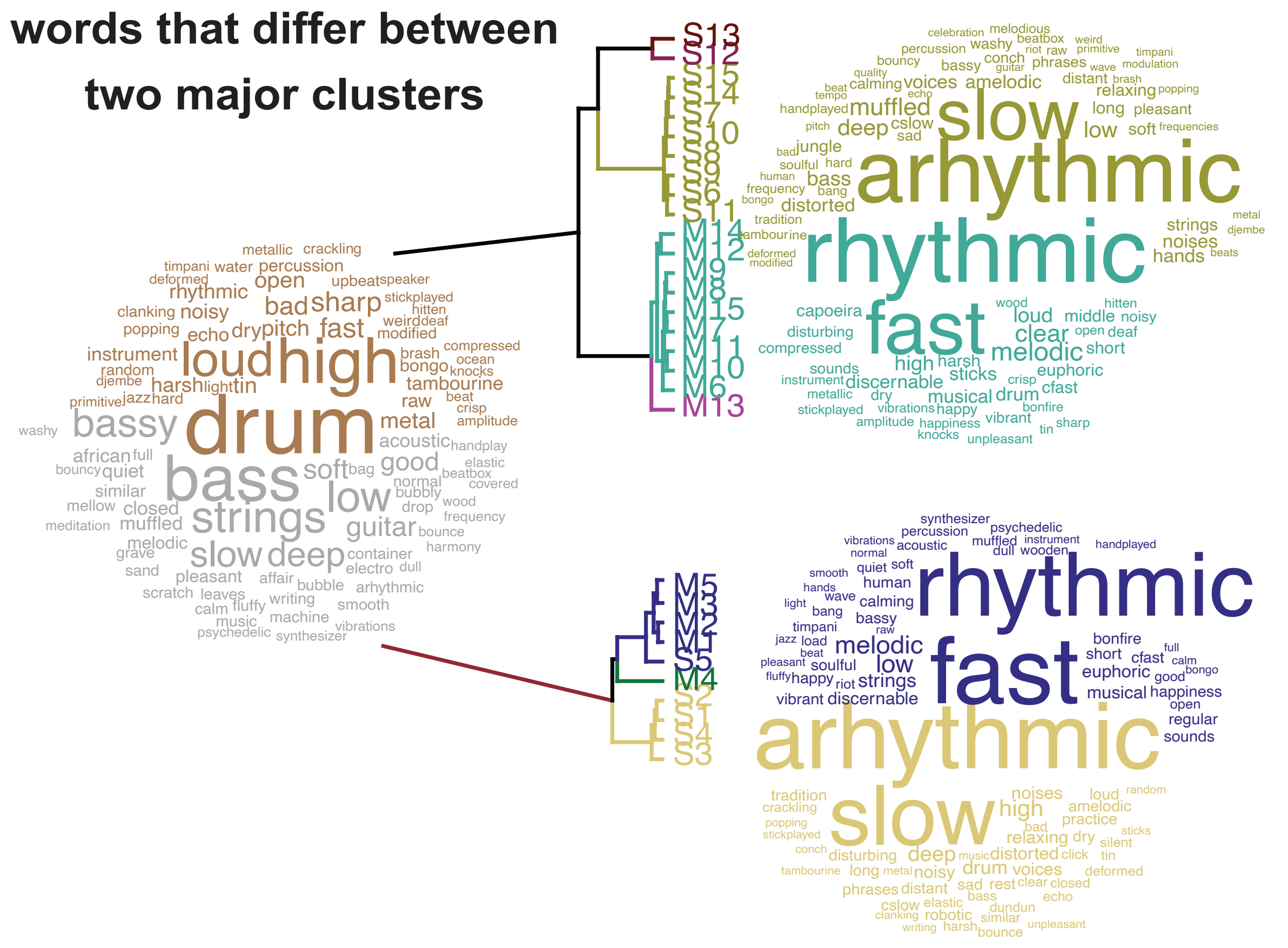
Results



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

- Results of Exp. 1 replicate Durojaye et al. (2021). Participants categorize well above chance which stimuli fall into speech or music categories.
- However, Exp. 2 shows that this speech/music distinction is not the most salient one. Thus, the type of task influences acoustic categorization.
- 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.

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

References

[1] 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.

MAX PLANCK NYU
CENTER FOR LANGUAGE, MUSIC, AND EMOTION