

A NEW EVALUATION METHOD

Speech features typically evaluated through a phone classification task

Requires aligned phonetic labels

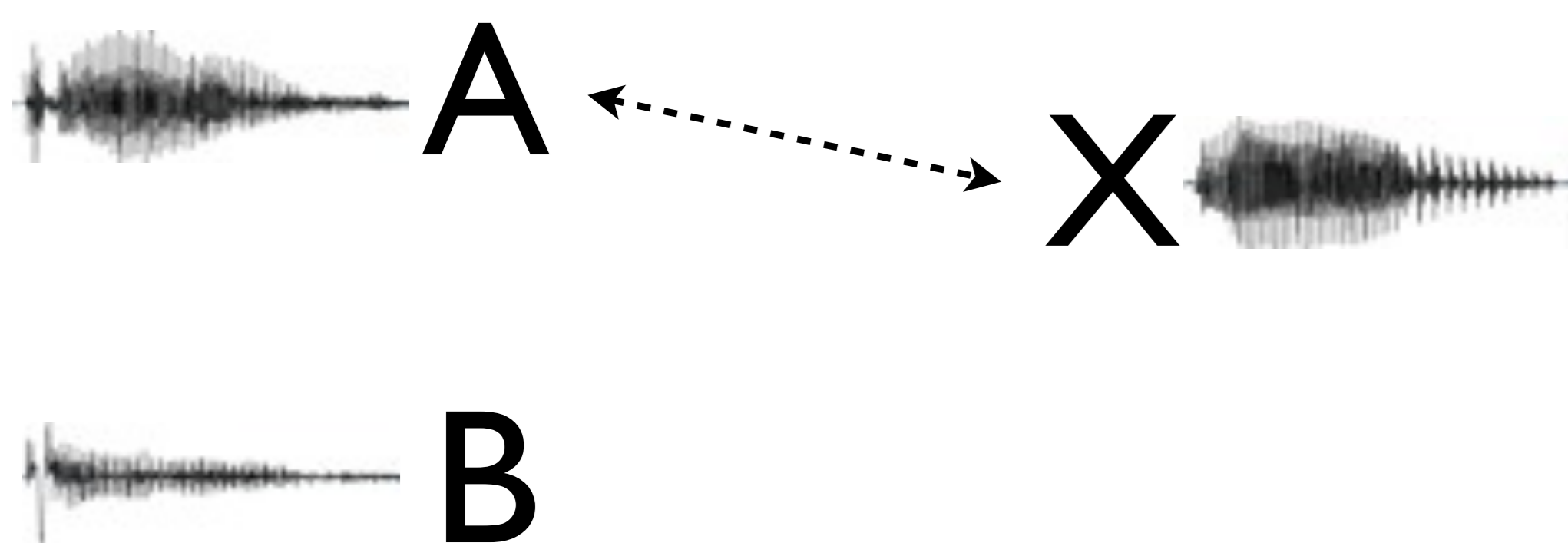
We propose an evaluation method that operates on whole words

Idea: Good speech features should discriminate between any pair of words

THE ABX TASK

A classical psychophysical task [1]

Given some sounds A, B, X,
is X closer to A or B?



Application to speech features

a, b, x features for sounds A, B, X:
 $d(a, x) > d(b, x)$?

d : DTW + cosine distance [2]

DATABASE OF SOUNDS

A subset of the Articulation Index corpus [3]

All 343 possible CV of American English (15 Vowels, 24 Consonants)

20 speakers (12 Male, 8 Female)

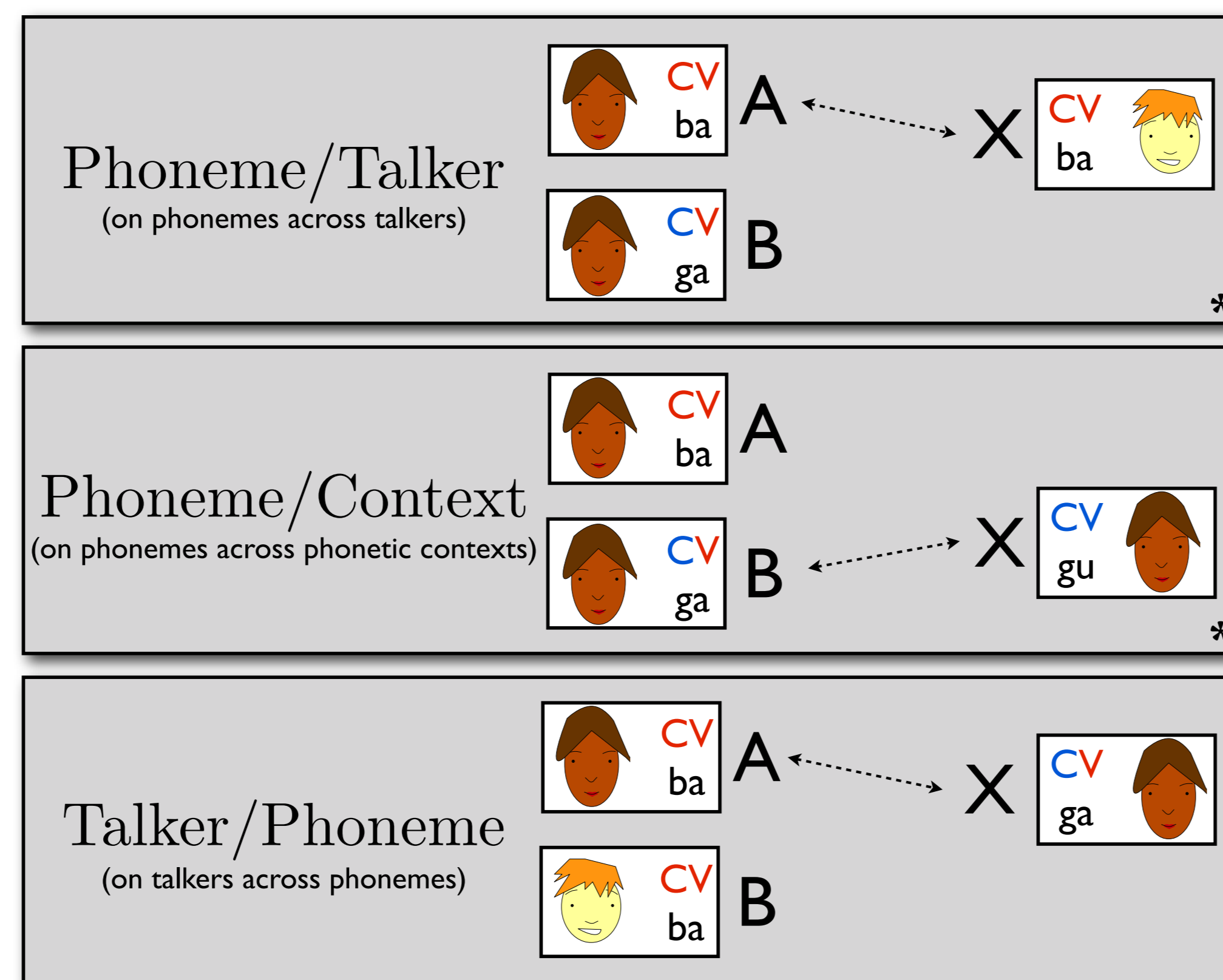
6839 syllables in total

Between 4.1 and 4.6 millions eligible ABX triplets, depending on task

SEVERAL VARIANTS OF THE TASK

Tasks based on minimal-pairs to improve interpretability of the results

Is X closer to A or B?



*Examples given for consonant discrimination. Adaptation of the design for vowel discrimination is straightforward

RESULTS ANALYSIS

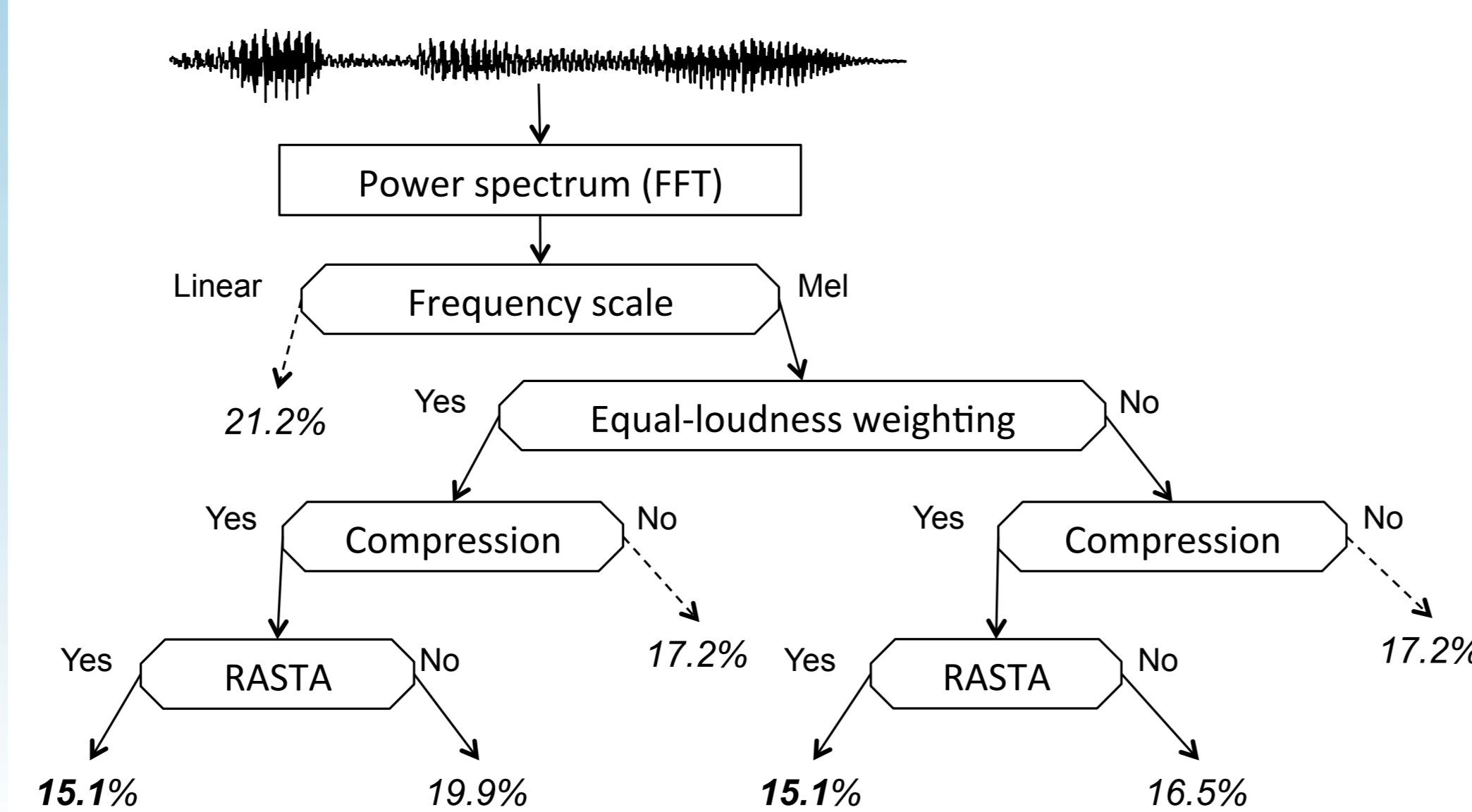
Average error rate in the ABX task for all possible combinations of the stimuli

Average error rate restricted to consonant (resp. vowel) discrimination

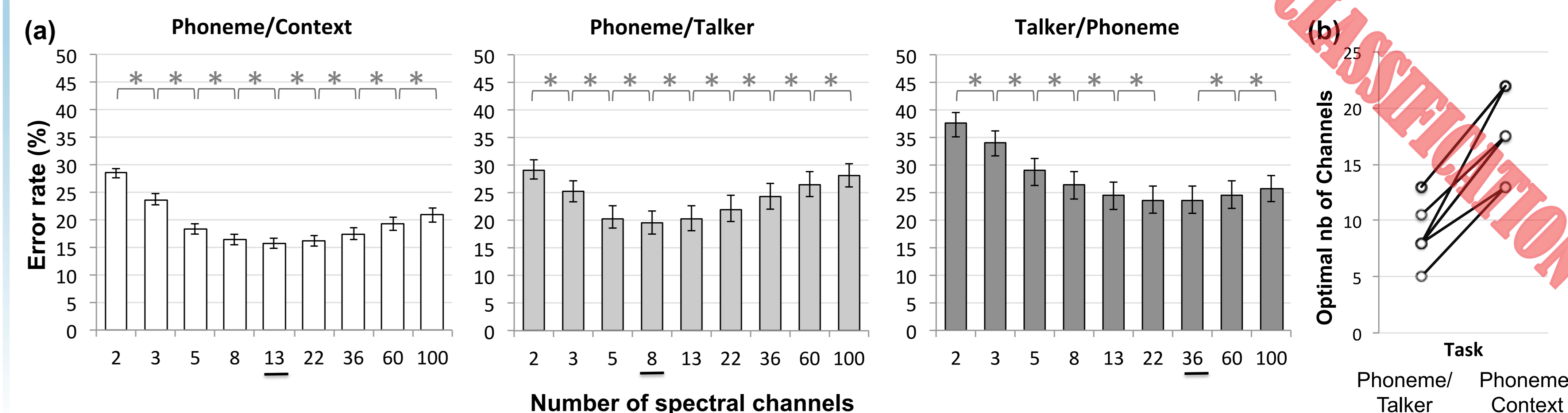
95 % bootstrap confidence intervals

SIGNAL PROCESSING PIPELINE

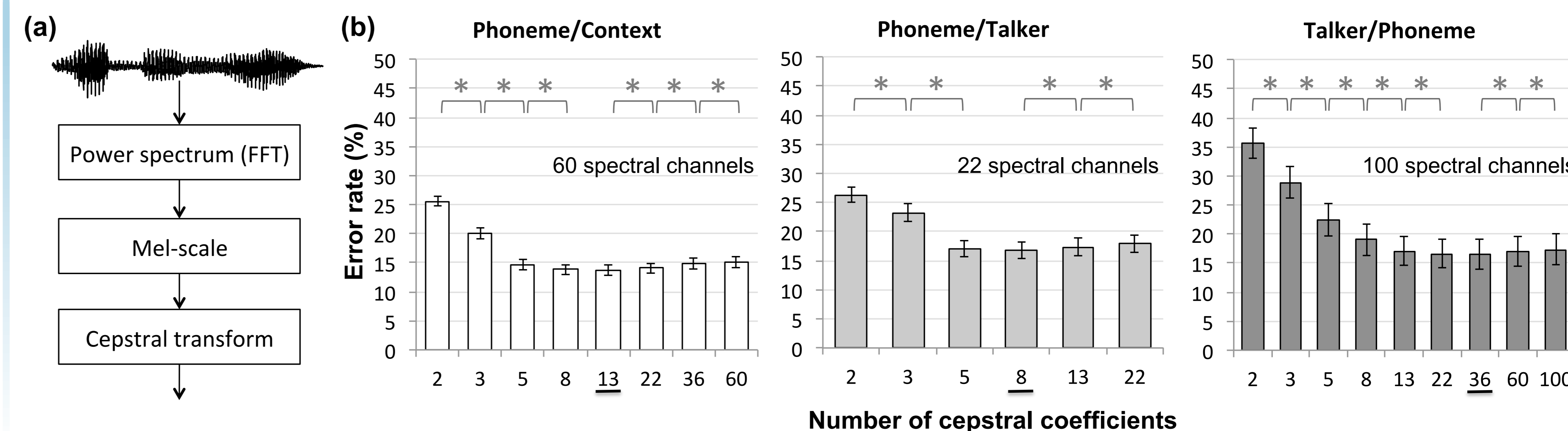
(Results on the Phoneme/Talker task)



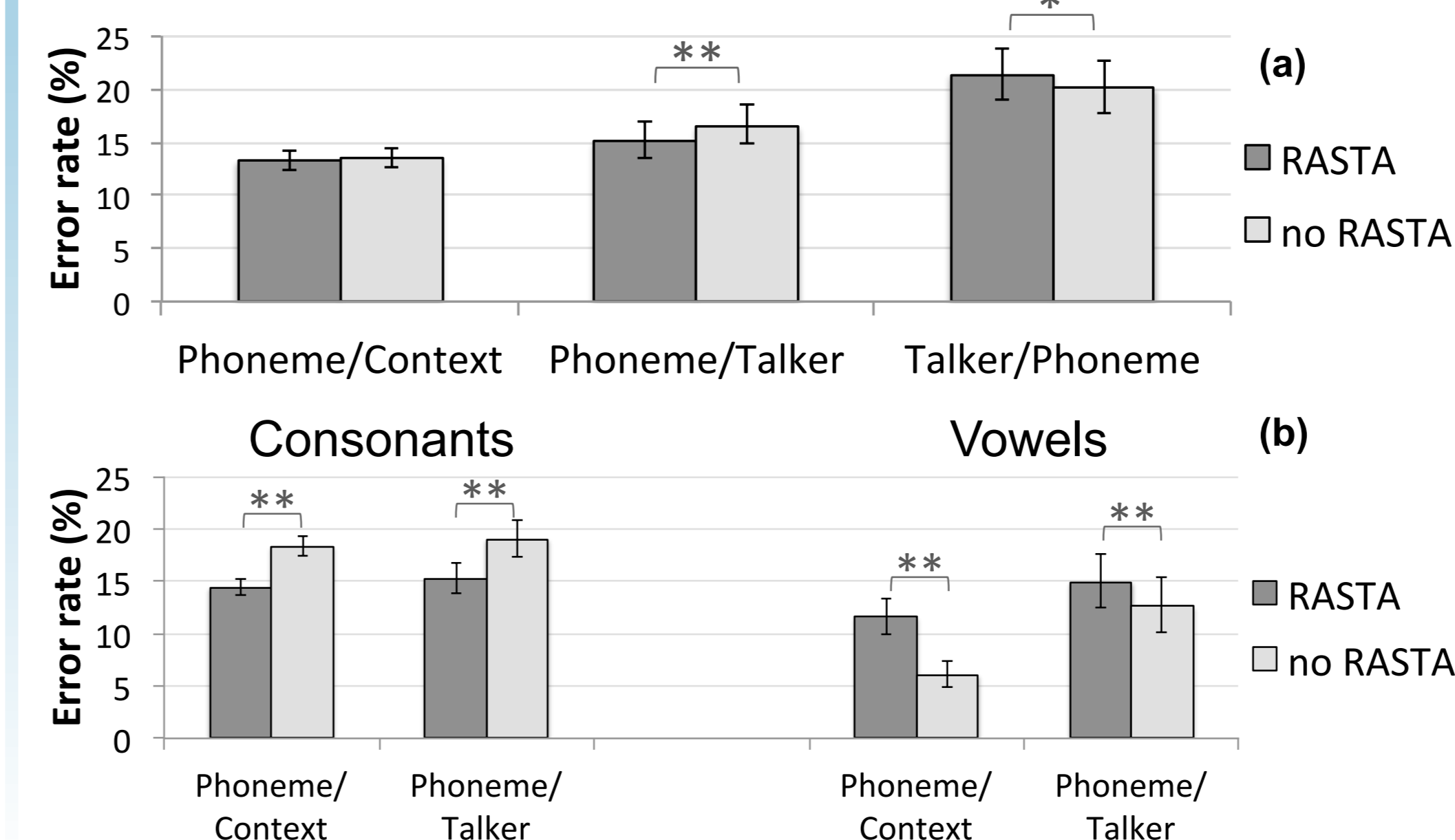
EFFECT OF THE NUMBER OF SPECTRAL CHANNELS



EFFECT OF THE NUMBER OF CEPSTRAL COEFFICIENTS



EFFECT OF RASTA FILTERING



CONCLUSION

A novel method for the evaluation of speech features:

- optimal numbers of channels correspond to classical values
- best features depend on task and type of phonetic contrast

Simple and natural and relevant to real tasks:

- appropriate for zero-resource applications [2]
- does not rely on any supervised classification technique
- enables fine-grained analysis of information contained in the features

Soon: software freely available

REFERENCES

- [1] N.A. Macmillan, C.D Creelman *Chapter 9: classification designs for discrimination* Detection theory: A user's guide, 2004
- [2] M.A. Carlin, S. Thomas, A. Jansen, H. Hermansky *Rapid evaluation of speech representations for spoken term discovery* Proceedings of Interspeech, 2011
- [3] P. Fousek, P. Svojanovsky, F. Grezl, H. Hermansky *New nonsense syllables database, analyses and preliminary ASR experiments* Proceedings of the International Conference on Spoken Language Processing (ICSLP), 2004