Modeling the time course of cue weighting angle calculations

5aPP5

Sara Ng[1], Gregory Ellis[2], Pamela Souza[2], Frederick Gallun[3], Richard Wright[1]

[1]Linguistics, University of Washington
[2]Communication Sciences and Disorders, Northwestern University
[3]Oregon Hearing Research Center, Oregon Health and Science University

Overview

- cue profile test assesses a hearing-impaired listener's use of specific speech cues
  - "weighting angle" quantifies how well individual listeners can use spectral & temporal cues
- **previous iterations of cue profile are too time-consuming**
- we created an **inter-trial stability metric** to determine whether early stopping was possible
  - **testing time can be shortened significantly** without much loss information
- a preliminary **machine learning classifier** shows promise in predicting 3-way weighting angle class from cue profile data, while preliminary, are encouraging

Background

Pure Tone Audiograms (PTA) are inadequate to describe hearing loss

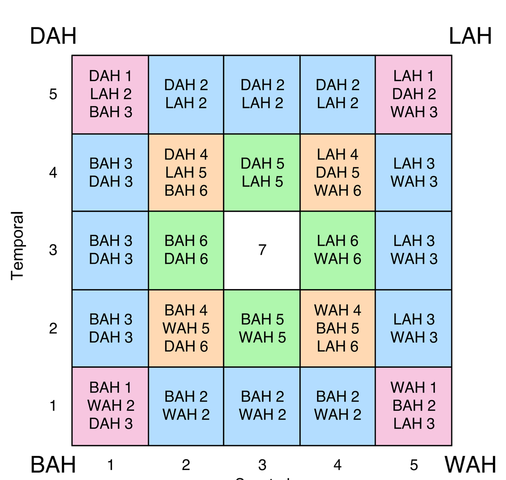
- Ability to recover speech info. is influenced by etiology and the individual [3]
- Sensitivity to pure tones is not a reliable indicator of other kinds of auditory ability [2]

Cue profile [3,4,5,6] is an alternative that provides useful information about the kinds of speech cues listeners attend to

- It can be used to compute a **cue weighting angle**, which shows whether the listener attends more to spectral information or the amplitude envelope [6]
  - Amount of hearing loss according to PTA is not indicative of the weighting angle
- A listener's cue profile is **reliable and repeatable** over several months [5]

Cue Profile Procedure

- familiarization, training, and test phases
- for each trial:
  1. listener hears a synthesized syllable from a set of 25 where F2 & F3 transitions and amplitude envelope varied along continua
  2. they label the stimulus as either "BAH," "DAH," "LAH," or "WAH"



Pink stimuli are considered unambiguous. Blue stimuli are ambiguous in a single dimension, and the inner squares (orange, green, white) are more ambiguous still. The (syllable, #) text is used in classification experiments and indicates the dimension of the accuracy vector that is incremented for a given response.

profile stage	# trials	stimuli
familiarization	40	Endpoints (in pink) with correct response highlighted
training	40/blk	Endpoints with correct response shown after trial, repeated until 80% accuracy attained
test	375	All stimuli in randomized order without feedback

A linear discriminant analysis categorizes each test (stimulus, response) pair into one of four groups using 2 discriminants. The first discriminant's coefficients are used to compute a cue weighting angle ranging 0-90°.

$$\theta = tan^{-1}\left(\frac{spectral\ coef}{temporal\ coef}\right)$$

Listener Demographics

- 26 listeners with 375 trials, 1 listener with 360, collected in [5] and [6]
- mild to moderately-severe sensorineural hearing loss
- 63-89 years (mean 73.6)

Methods: Stability Measure

Compute rolling averages for listener's data:

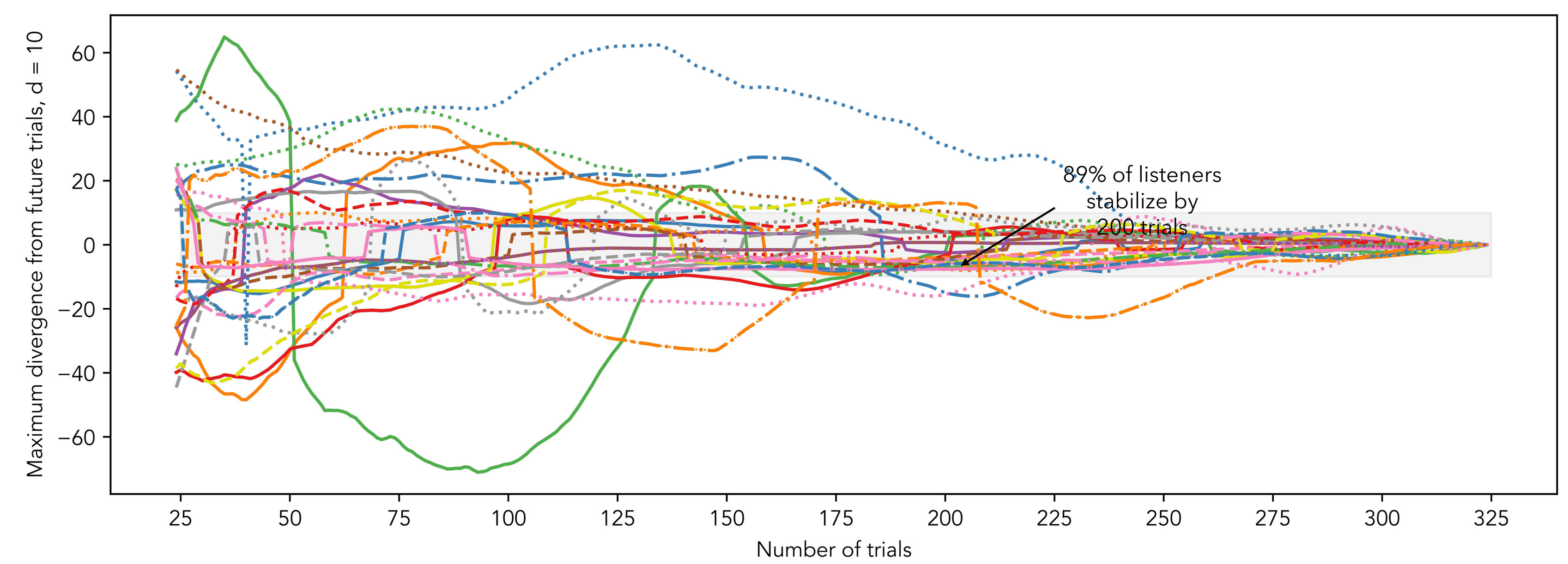
1. compute angle using all trial data up to that point.
   - produces a sequence of 375 predicted angles
2. Rolling average θ_{avg_i} over a window size w at trial i is calculated as

$$\theta_{avg_i} = \frac{\sum_{max(0,i-w)}^{i} \theta_{p_i}}{w}$$

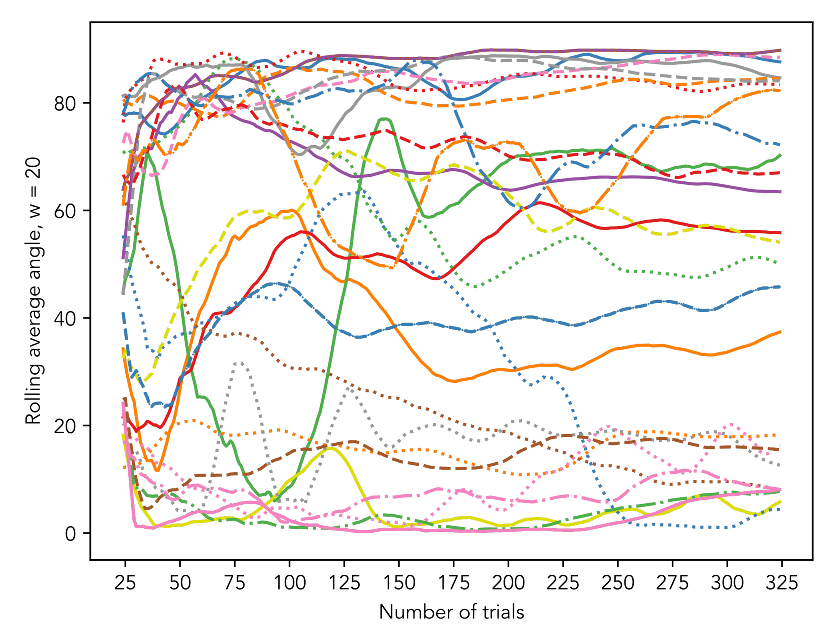
Find **stability point**:
for divergence tolerance d, a listener is stable by trial i if, for all trials j, $i < j \leq 325$

$$|\theta_{avg_i} - \theta_{avg_j}| < d.$$

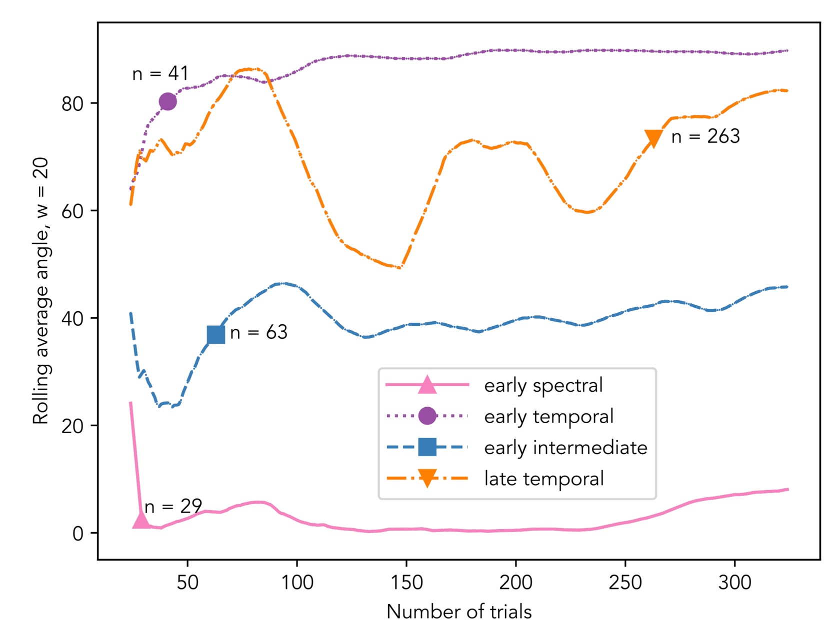
Results: Angle Stability



Maximum divergence between the moving average cue profile angle at a given trial number and all subsequent angles for a listener. Values within the grey rectangle are within the stability window for divergence of 10°.



moving average angle for each subject



four exemplar listeners and their stability points labeled

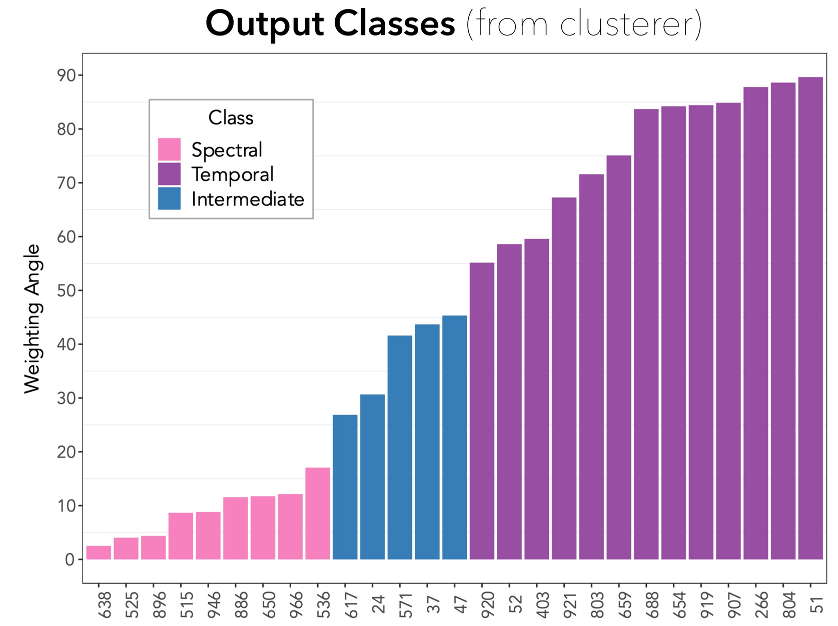
- At $t > 325$, there is **evidence of subject fatigue,** skewing angles temporally & not indicative of their actual cue preference. Because of this, we only used the first 325 trials when determining stability.
- We default to a window size $w=20$ and **maximum divergence d=10°**, and compare to $d=5°$.
  - For $d=5°$, 77% of listeners reach stability within **200 trials**.
  - For $d=10°$, **89% of listeners reach stability** within the same time frame.

Methods: Clustering and Classification

Step 1) with coarse-angle categorization as a clinical motivation, we use an **automatic clusterer** to define the boundaries between spectral, temporal, and balanced cue weighting.

Step 2) train an LSTM, a neural network that handles sequential information, to **predict the angle class using the cue profile and demographic data**

Output Classes (from clusterer)



Input Features (for LSTM)

	training data type	dim.	values
sequential	Running angle: predicted using trials up to $t=i$	1	0 - 90
	Stimulus type and response	25	1 - 4
	Grid accuracy	7	0 - # trials
nonsequential	QuickSIN	1	float
	PTA score, per ear, 500 1000 and 2000 Hz	6	int
	Age	1	int 63-89
	Sex	1	M or F

Results: Machine Learning Classification

Train and Eval pipeline:

1. Augment training data by creating copies of each listener with first *n* trials, using step size of 5
2. Configure network with 2 LSTM [1] layers, hidden dimension 32, dropout rate 0.1
3. Select input features for model, with each subsequent models adding one feature to the previous best combination
4. Train each model using 3 random seeds for 100 epochs with early stopping
5. For each input configuration, report the seed which gave best test performance

model type	input features used	test acc.
random forest	stimulus type & resp. + grid accuracy	0.57
LSTM	**PTA**	**0.83**
	PTA + Running angle	0.73
	PTA + Running angle + Age	**0.83**
	PTA + Running angle + Age + Gender	0.81
	PTA + Running angle + Age + Gender + QuickSIN	**0.83**
	PTA + Running angle + Age + Gender + QuickSIN + stimulus & resp.	0.73

Discussion

The cue profile can be simplified.

- for many listeners, testing during the cue profile **could be shortened by 175 trials** (46%) with reasonable fidelity

Simplifications should consider individual differences.

- **listeners with strongly spectral or strongly temporal angles often stabilize very quickly**
- listeners with medial angles can stabilize quickly, but it's less common
- depending on parameters selected, some listeners require the full cue profile test, or may not stabilize at all

Machine learning shows promise as a mechanism to shorten the cue profile

- a simple neural network can fit well to observed input features
- naïve data augmentation strategies improve performance
- generalization to unseen data is less reliable
  - cause is likely the small dataset size

Acknowledgments

This work was funded by NIH NIDCD R01 DC006014 (PS) and R01 DC 015051 (FJG). The views expressed are those of the authors and do not represent the views of the NIH or the Department of Veterans Affairs. We would like to thank Lauren Balmert (Northwestern) and Mari Ostendorf (University of Washington) for their guidance in developing the stability metric. Thanks to Agatha Downey (University of Washington) for her help debugging the LSTM.

References

[1] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[2] Musiek, F. E., Shinn, J., Chermak, G. D., & Bamiou, D. E. (2017). Perspectives on the pure-tone audiogram. *Journal of the American Academy of Audiology*, *28*(07), 655-671.

[3] Souza, P. E., Ellis, G., Marks, K., Wright, R., & Gallun, F. (2021). Does the Speech Cue Profile Affect Response to Amplitude Envelope Distortion?. *Journal of Speech, Language, and Hearing Research*, *64*(6), 2053-2069.

[4] Souza, P., Gallun, F., & Wright, R. (2020). Contributions to speech-cue weighting in older adults with impaired hearing. *Journal of Speech, Language, and Hearing Research*, *63*(1), 334-344.

[5] Souza, P., Wright, R., Gallun, F., & Reinhart, P. (2018). Reliability and repeatability of the speech cue profile. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2126-2137.

[6] Souza, P. E., Wright, R. A., Blackburn, M. C., Tatman, R., & Gallun, F. J. (2015). Individual sensitivity to spectral and temporal cues in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 58, 520-534.

get this poster and more info about the project