Effects of Frequency Fine-tuning in Hearing Impaired Speech Recognition

Ali Abavisani, Jont Allen
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign

Introduction

- Current technology tunes hearing aids based on Pure-Tone Thresholds (PTT) of hearing impaired (HI) ear, along with gain rules developed based on statistics on HI speech intelligibility.
- At Human Speech Recognition (HSR) group, we are seeking a speech based test to identify and characterize HI ears and recommend hearing aid tuning, based on such speech test.
- To control the large amount of natural variability of speech, we used low-context speech segments (e.g. Consonant-Vowel), that are determined to be robust to noise from our previous experiments.
- All tokens have been recognized at least 90% correct on average for 30 NH listeners, at the highest noise available in the experiments (i.e., $SNR_{90} < 0$ dB)
- We want to examine several hypotheses on HI ear phone recognition strategies, and observe the similarities and differences between HI and NH ears.
- One hypothesis is whether changing the talker to more noise-robust one, will change the nature of confusions.
- Other hypothesis is whether changes in frequency components of primary cue region, can improve phone recognition for HI ears

Method

Subjects
4 subjects (8 HI ears), aged 40-64, with mild to moderate HL

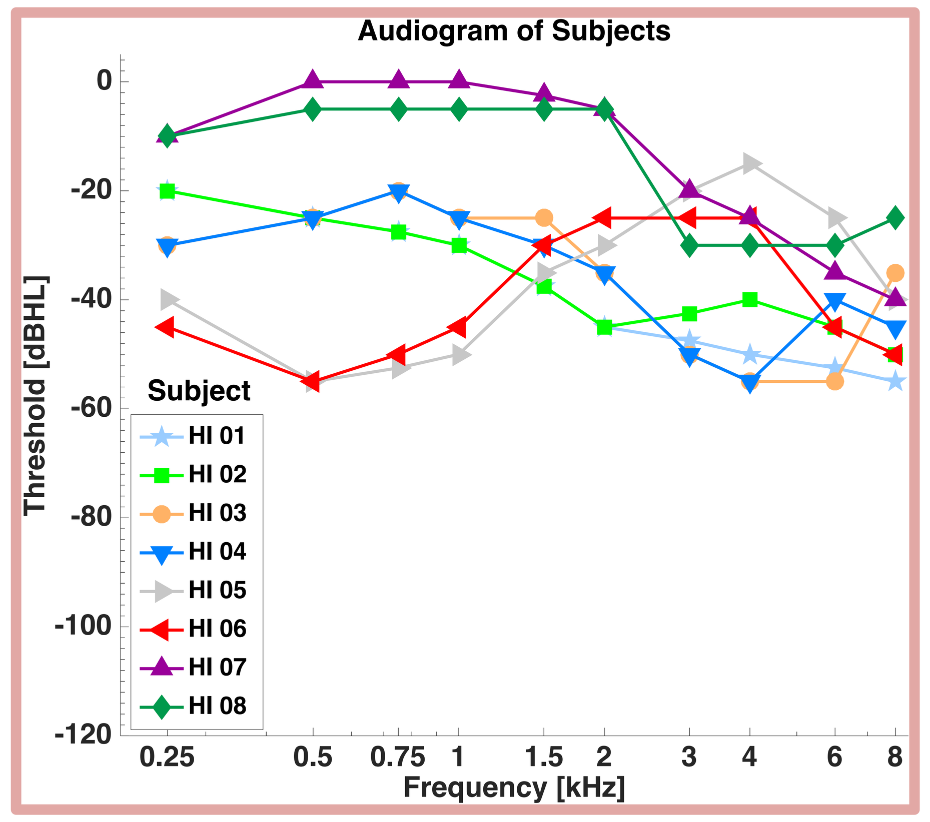
Speech material
- 14 consonants: /p, t, k, f, s, ʃ, b, d, g, v, z, ʒ, m, n/
- 4 vowels: /ɑ, æ, ə, eɪ/
- 2 male + 2 female talker for each CV, selected from pre-evaluated pool of 10 male and 10 female talkers, for noise robustness (i.e., SNR_{90})
- Talker change provides means of change in the intensity of primary cue region as $|\Delta SNR_{90}| > 6$ [dB]
- Vowel change provides means for frequency fine-tuning (small shift) in the primary cue region, where tokens with different vowels have $|\Delta SNR_{90}| < 3$ [dB]
- Presented at the most comfortable level (MCL)

SNR
- Four SNR conditions consisting of 0, 6, 12 dB and Quiet
- All tokens have SNR_{90} well below the lowest test SNR (0 dB)

List 1: Screening
- All Consonant+/a/ tokens randomly presented at highest noise
- Errorful tokens are copied into List 2 for evaluation

List 2: Evaluation
- Tokens tested three more times at SNR = 0 and 6 dB
- If the total error for a token is above 50%, the token will be copied into List 3 for complete testing.
- As soon as a token included in List 3, variation of that token will be added to List 2 for evaluation
- Different talker (same gender) with $|\Delta SNR_{90}| > 6$ [dB]
- Various vowels (same gender) with $|\Delta SNR_{90}| < 3$ [dB]

List 3: Testing
- All tokens presented 8 times at each test SNR (total of 32 presentation)



Confusion Matrix Data Analysis

- Convert "count matrix" to probability distribution ***P(i,j)*** matrix by dividing each element of the matrix by the row sum.
- **Entropy** of each token was calculated from the probability distribution of each row in the probability distribution matrix:

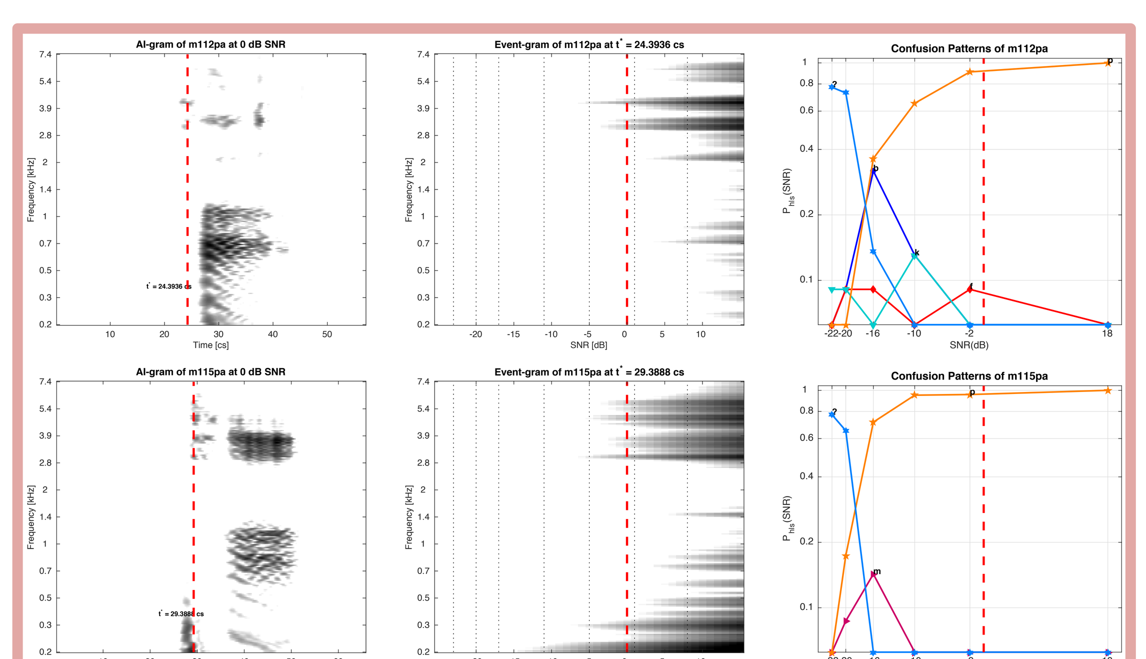
$$H_{CV}(i) = \sum_{j=1}^{14} P(i,j)\log P(i,j)$$

- **Probability of error** for each token was calculated by subtracting the diagonal element of probability distribution matrix from 1:

$$Pe_{CV}(i) = 1 - P(i,i)$$

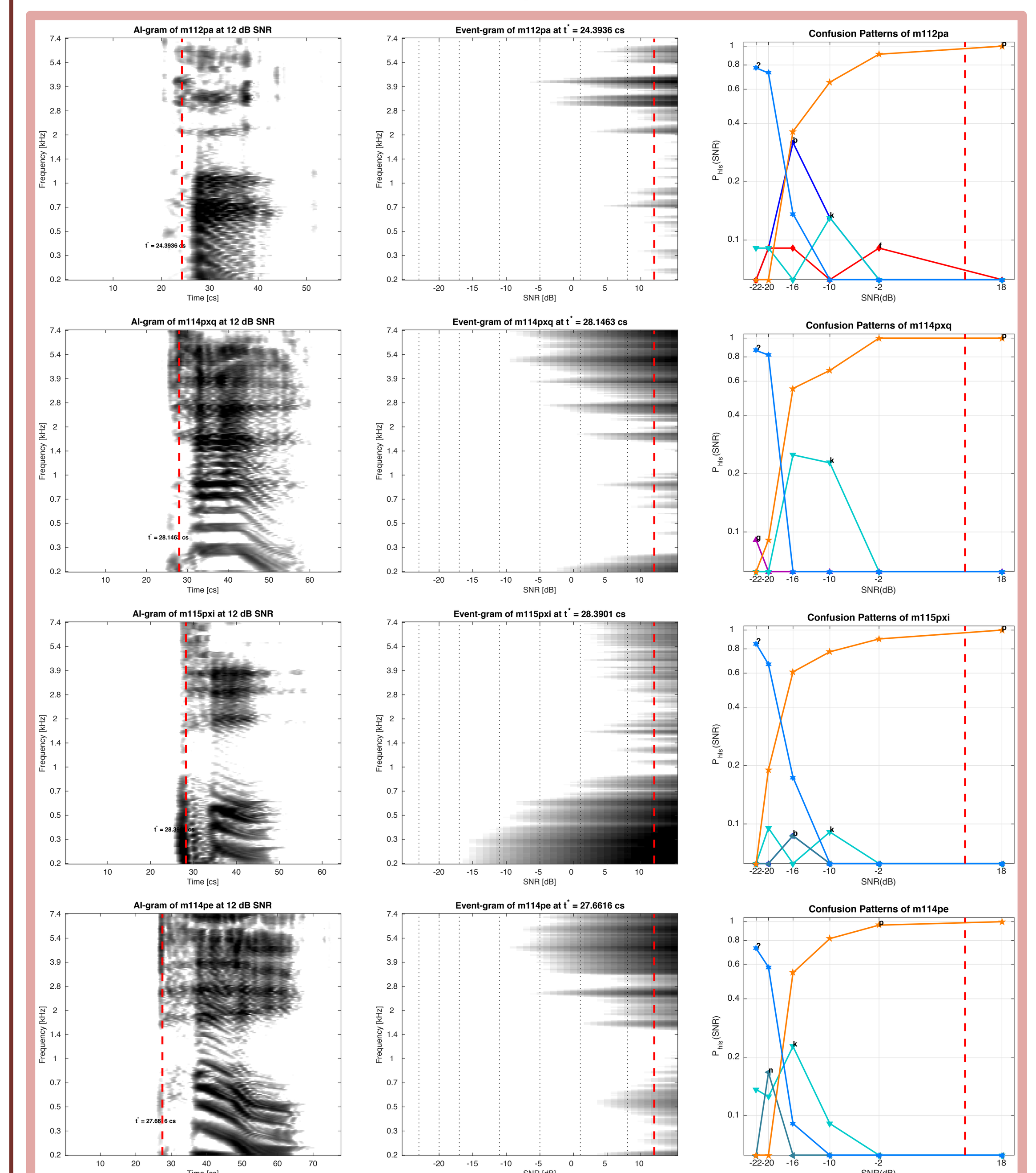
- **Confusion pattern** is the plot of one row of the confusion matrix, showing the probability of response for each consonant.
- **AI-gram** is an interpretation of the spectrogram of the token, considering the SNR at human auditory frequency bands.
- **Event-gram** is the illustration of a cross section of AI-gram on specific time stamp.

Example of talker change for /pa/ at 0 dB SNR



- First token (top row) has SNR_{90} = -4 dB, and second token (bottom row) has SNR_{90} = -10 dB.
- Confusion patterns show both tokens have been recognized almost 100% correct by NH listeners at this SNR.
- Event-grams show slightly stronger primary cue feature of /pa/ (high frequency burst), for second token.

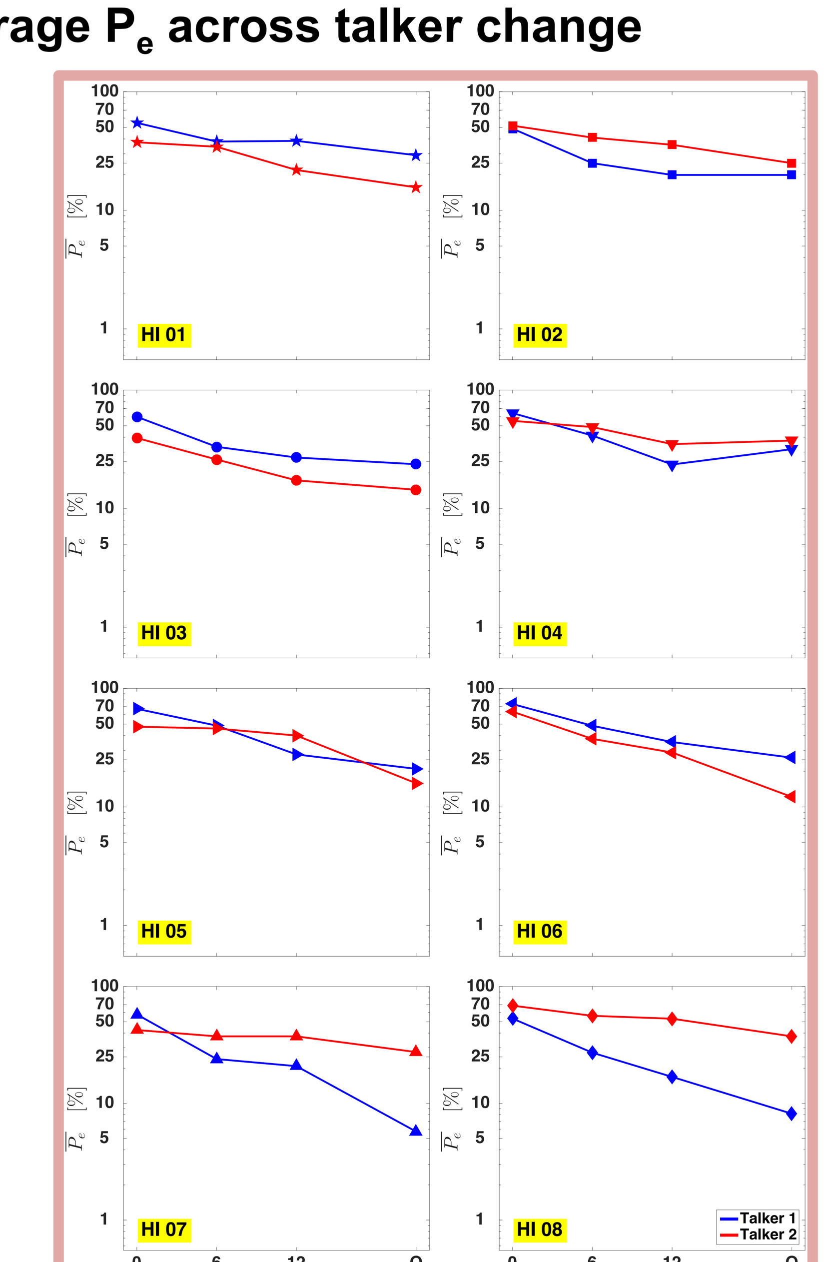
Example of vowel change for /pa/ at 12 dB SNR



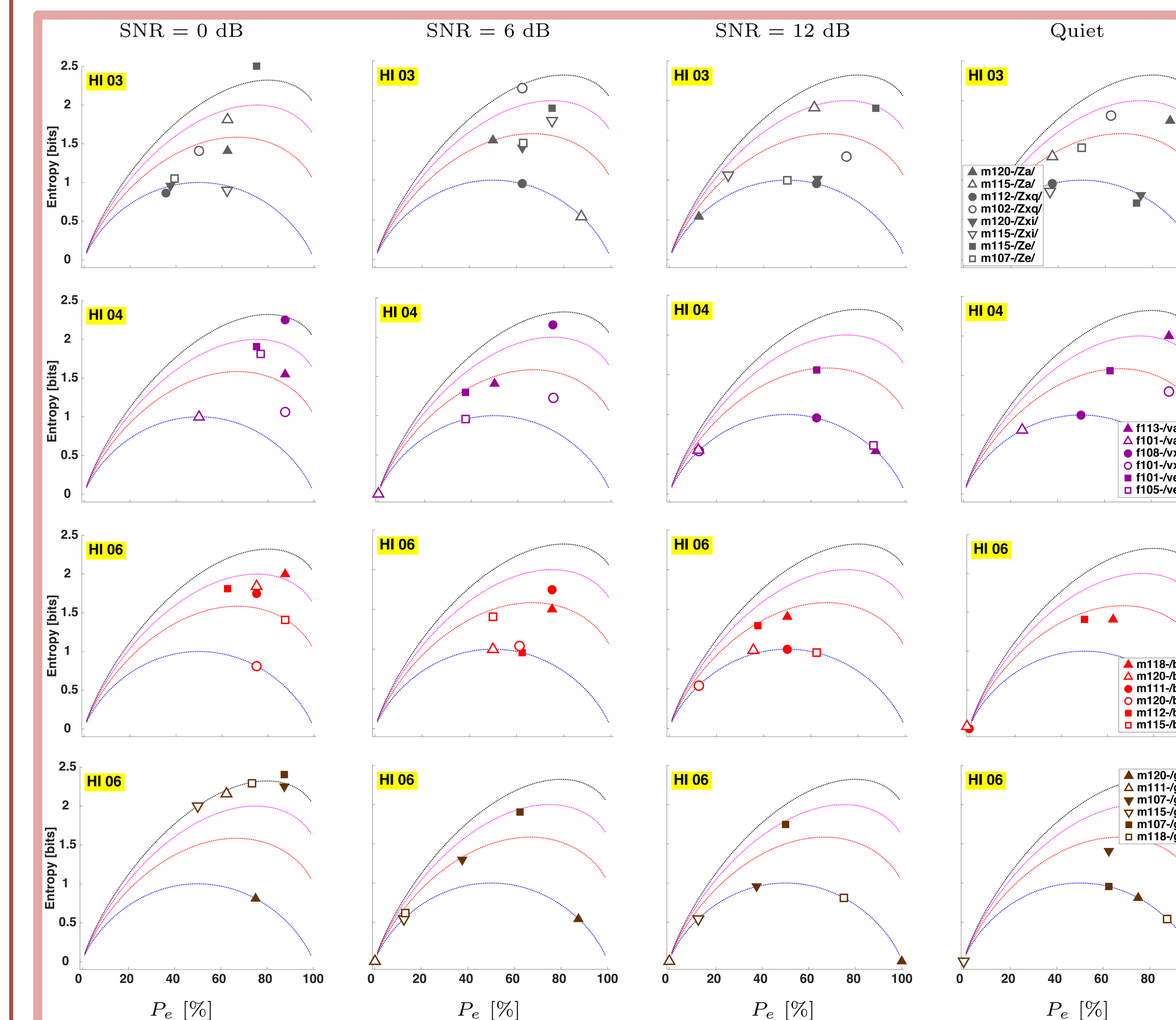
- All tokens have similar SNR_{90} ~ -3 dB.
- Event-grams show the variations along primary cue frequency region (high frequency burst) across different vowels, where changing the vowel changes the consonant articulation.
- Confusion patterns show that in all cases, the primary cue is intelligible for NH ear.
- Does change in vowel formants have effect on HI ear?

Results

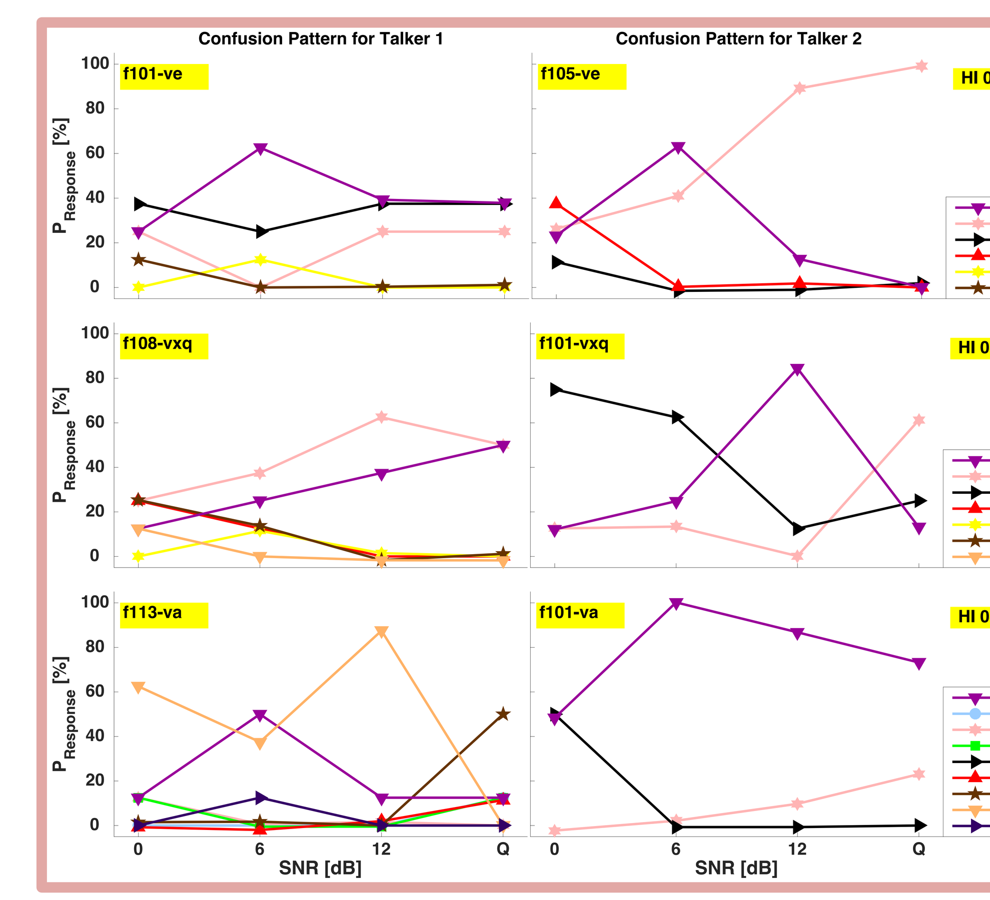
Average P_e across talker change



Examples of entropy vs. probability of error



Example of confusion patterns when vowel/ talker change



Conclusion

- Talker change to more robust tokens resulted in small changes in average error at high noise, but demonstrated various outcomes for low noise/quiet for different HI ears. This reminds the idiosyncratic nature of HI speech recognition.
- Confusion patterns along with entropy analysis show various improved/ degraded scores when vowel or talker changed for each ear.
- Vowel and/or talker change can introduce new confusions as conflicting cues can become available for HI ear by these changes.
- Control of SNR_{90} for tokens is essential in generalization of primary cue manipulations when we provide new talker for same CV.

References

[1] Abavisani, A., Allen, J. B., Evaluating hearing aid amplification using idiosyncratic consonant errors, Journal of Acoustical Society of America, Vol. 142(6), Dec. 2017
[2] Li, F., Allen, J. B., Manipulation of consonants in natural speech, IEEE Transactions on Audio, Speech and Language Processing, Vol. 19(3), 2011
[3] Regnier, M. S., Allen, J. B., A method to identify noise-robust perceptual features: Applications for consonant /t/, Journal of Acoustical Society of America, Vol. 123(5), 2008