Effects of NAL-R on Consonant Vowel Perception in Hearing Impaired Subjects

Christoph Scheidiger[1]*, Torsten Dau[1] and Jont B. Allen[2]

[1]*Centre for Applied Hearing Research, Technical University of Denmark*

[2]*Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign*

*csche@elektro.dtu.dk

Introduction

Consonant vowel (CV) identification experiments in speech weighted noise
- Two gain conditions; Flat-Gain (FG) and NAL-R (compensation for hearing loss)
- 16 hearing impaired (HI) ears are tested at 4 SNRs
- Subjects are tested at their most comfortable loudness (MCL)

Audibility of CVs (at MCL)?
- Audibility based on pure tones and long term average speech spectrum are not appropriate
- Rigorous definition of audibility based on token entropy

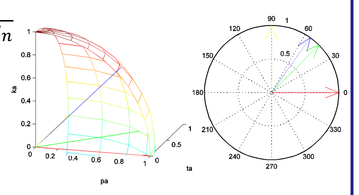
Effects of NAL-R?
- Are the CV recognition rates higher with NAL-R? Do the confusions change?

Key Findings
- *Most comfortable loudness* (MCL) testing works -> sounds are audible
- Spectral correction (i.e. NAL-R) decreases the entropy of the responses
- Spectral correction makes listeners respond more consistently on a token level

Methods

Cosine Direction: Every confusion matrix defines a vector space. The angle between responses can be calculated, with the help of the Hellinger distance metric:

$$\cos\theta_{pq} = \vec{p}\cdot\vec{q} = \sum_{n=1}^{N}\sqrt{p_n}\sqrt{q_n}$$



Entropy: Measure of randomness of response

$$\mathcal{H}(p) = \sum_{n=1}^{N} p_n \log_2\left(\frac{1}{p_i}\right)$$

- 16 HI ears (8 subjects, all with post-lingual hearing loss), age μ=74 year (σ = 7year)

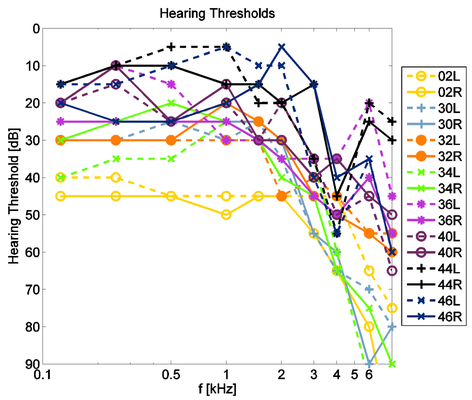


Figure 1: Pure Tone Thresholds of the 16 ears

- 14 consonants (/p,t,k,f,s, ʃ, b, d, g, v, z, ʒ, m, n/) followed by the vowel /a/
  - Two talkers per consonant (male & female)
  - All stimuli have < 3% error (i.e. 1 error /38 trials) for normal hearing (NH) ears at ≥ -2dB SNR
- Four noise levels Quiet, 12dB, 6dB, 0dB (speech weighted noise)
- Max entropy task (14 possible responses (/p, t, k, f, s, ʃ, b, d, g, v, z, ʒ, m, n/ + /a/)
- Two conditions
  - Flat-Gain (FG), no spectral correction presented at MCL
  - Spectral Correction (NAL-R), spectral correction according to hearing thresholds also presented at MCL
- For each subject, the number of trials per token and SNR varies between 5 and10, depending on the error rate of the first 5 trials

Results

- Categorization (first arrow indicates what happened to the error with NAL-R, second what happened to the entropy)

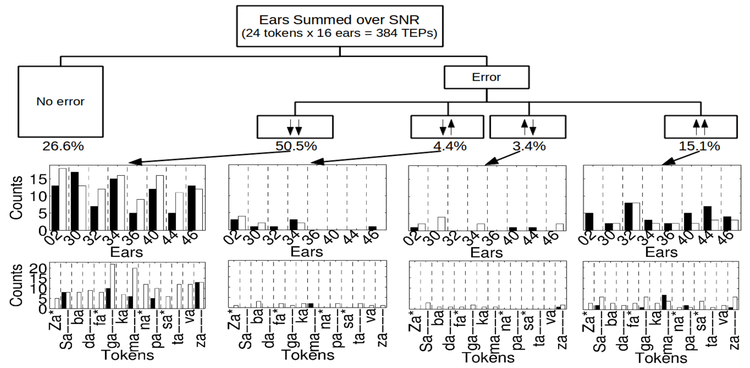


Figure 2: Categorization of token-ear pairs

- Confusion Bars collapsed over SNR (proportion correct indicated by white space, confusions indicated by color, see legend)

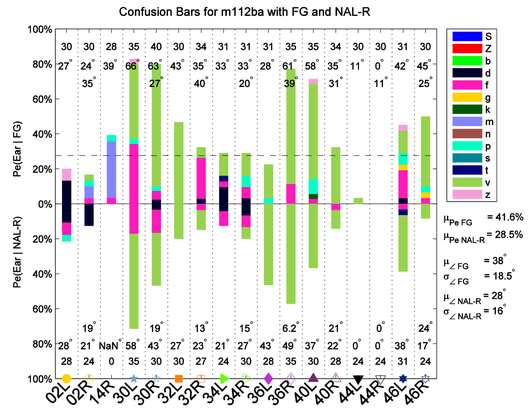


Figure 3 (a): Confusion bars for male token of /ba/

Conclusions Figure 3 (a):
- Average error decreases by 13%
- Dominated by /va/ confusions
- /fa/ confusions decrease with NAL-R
- Angle to correct answer decreases with NAL-R by 10°
- Entropy decreases with NAL-R

Conclusions Figure 3 (b):
- Average error decreases by 2%
- Dominated by /da/ confusions
- /ga/ confusions increase with NAL-R
- Angle to correct answer decreases with NAL-R by 3°
- Entropy stays high with NAL-R

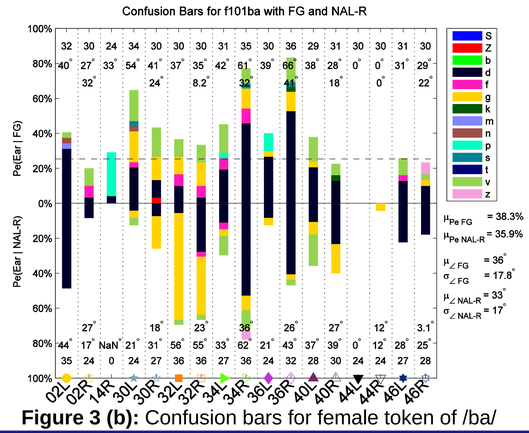


Figure 3 (b): Confusion bars for female token of /ba/

- Audibility is not an issue despite MCL testing

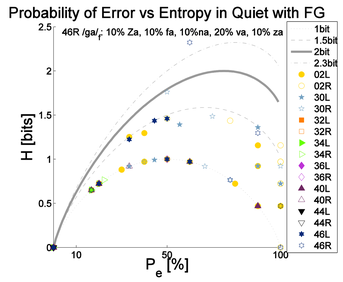


Figure 4: Error vs. Entropy in quiet, proposed audibility threshold (2-bit curve). All tokens for all ears are audible except, female /ga/ for ear 46R.

- Entropy in quiet is a rigorous measure for audibility
- 2-bit curve is a reasonable threshold , since the average Miller and Nicely confusion group is 3.

Reasons why PTTs are inappropriate:
- CV perception is binary (Singh and Allen [2012])
- Natural wide band speech is much more complex than pure tones, with non-linear effects like forward masking playing an important role.
- Average speech spectrum is irrelevant when it comes to CV perception, since it is dominated by vowel energy

- Angle between conditions (FG and NAL-R)

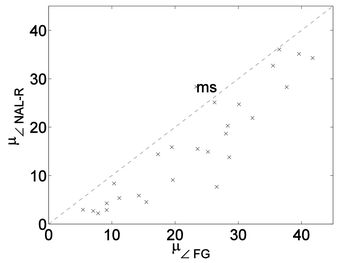


Figure 5(a): Scatter plot of mean angles between responses and correct answer. Angles with NAL-R decrease for all tokens but the male /sa/ token (ms).

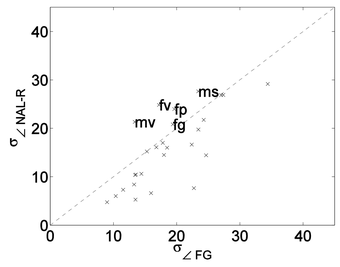


Figure 5(b): Scatter plot of standard deviation of angles between responses and correct answer. The standard deviation is a measure of listener consistency. It increases for all tokens but the ones labeled (fv: female /va/).

Conclusion

The data from the two experiments showed:

- Despite the uncommon approach of measuring CV confusions at MCL, we demonstrated with the low entropy in quiet that audibility was not an issue.
- We defined a more rigorous definition for audibility in high-entropy recognition tasks
- Hellinger Distance is a powerful tool to analyze confusion matrices
- NAL-R in general, decreases the entropy of the responses; this is true for all listeners and all tokens
- NAL-R makes the responses on a token level more consistent across all ears

Summary: NAL-R in general helped and listeners are consistent on a token basis; however, specific listeners have problems with specific tokens, implying that tailored training could help

References

Han, W. (2011). "Methods for robust characterization of consonant perception in hearing-impaired listeners," PhD thesis, University of Illinois.

Li, F. (2010). "Perceptual cues of consonant sounds and impact of sensorineural hearing loss on speech perception," PhD thesis, University of Illinois.

Li, F., Trevino, A., Menon, A., Allen, J.B. (2012). "A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise," J. Acoust. Soc. Am., 132, 2063-75.

Singh, R., Allen, J.B. (2012). "The influence of stop consonants' perceptual features on the articulation index model," J. Acoust. Soc. Am., 131, 3051-68.