MEASURING SPEECH PERCEPTION WITH RECOVERED ENVELOPE CUES USING THE PERIPHERAL AUDITORY MODEL

Nursadul Mamun, Khadija Akter, Hussnain Ali, John H. L. Hansen

Center for Robust Speech Systems (CRSS): Cochlear Implant Processing (CI) Lab

Erik Jonsson School of Engineering & Computer Science, University of Texas at Dallas, Richardson, Texas, U.S.A.

(Nursadul.Mamun, Hussnain.Ali, John.Hansen)@utdallas.edu; khadija007008@gmail.com

Cochlear Implant Laboratory

1. INTRODUCTION

- **Issues:** Listening tests, which relies on subjective feedback under laboratory test conditions, can be expensive, time consuming, and may present inconsistencies with test/re-test reliability.
- **Issues:** Speech perception can be degraded at high levels for both normal & impaired listeners [2].
- **Auditory Nerve (AN) models** are useful to detect, analyze, and segregate dynamic acoustic stimuli in complex environments [1].
- **Spectral Envelope (ENV)** is sufficient for understanding speech in quiet whereas the **Temporal Fine Structure** (TFS) is necessary for speech segregation in noisy conditions [3].

2. METHODS

- **Predicting speech Intelligibility under different types of noises and distortions**
- **Comparing predicted scores with subjective scores and scores from existing metrics**

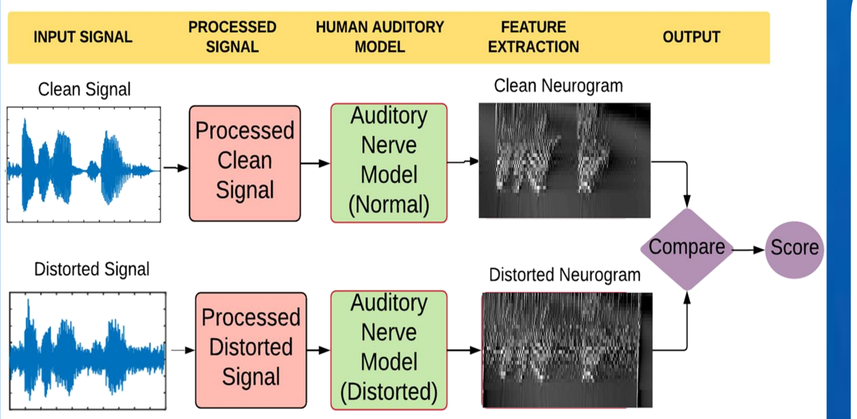


Fig.1. Basic block diagram of proposed method.

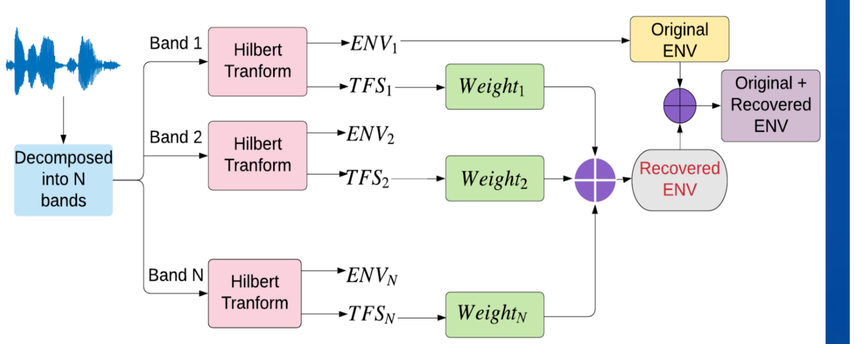


Fig. 2. Signal processing steps for generating original and recovered envelope.

3. METHODS

Test Corpora:

- Databases: IEEE sentences, NU#6 words.
- SNR: -20 to 30 dB; dB-SPL: 65~ 95 dB.
- Noise: Babble, Speech shape, Interrupted, Car.
- Distortions: Phase jitter, Peak clipping, Center Clipping.
- Subjective Data Studies: Dubno [4], Studebaker [5]

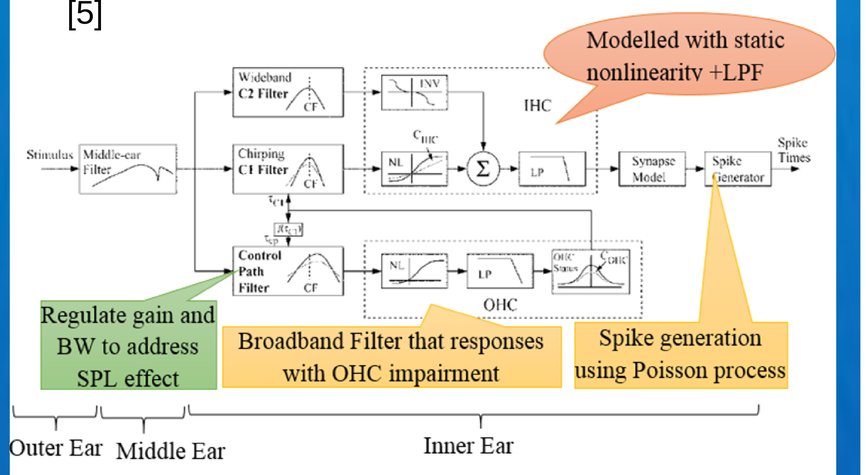
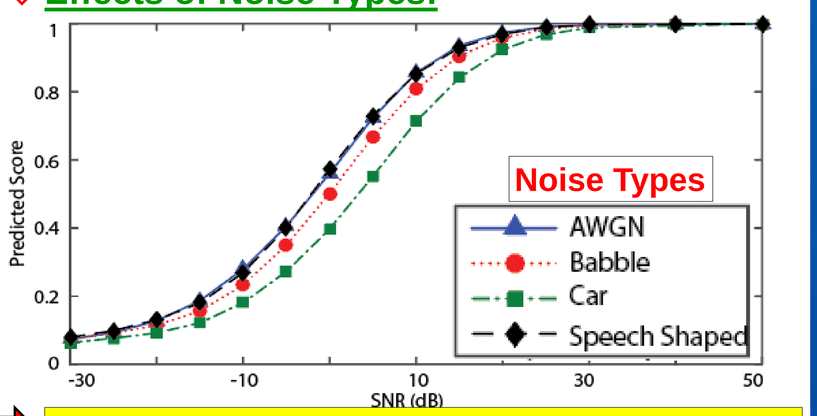


Fig. 3. Auditory Nerve Model (Zilany and Bruce et. al. 2006 [1])

4. RESULTS

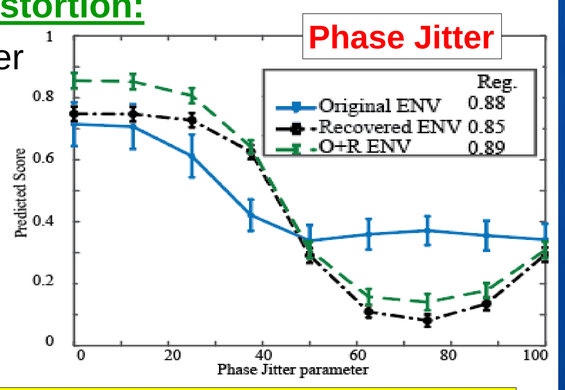
Effects of Noise Types:



Intelligibility Predicted Score increases as SNR increases.

Effects of Distortion:

1. Phase Jitter (PJ)
   - SNR: 5 dB
   - Corpus: 200 NU#6 words

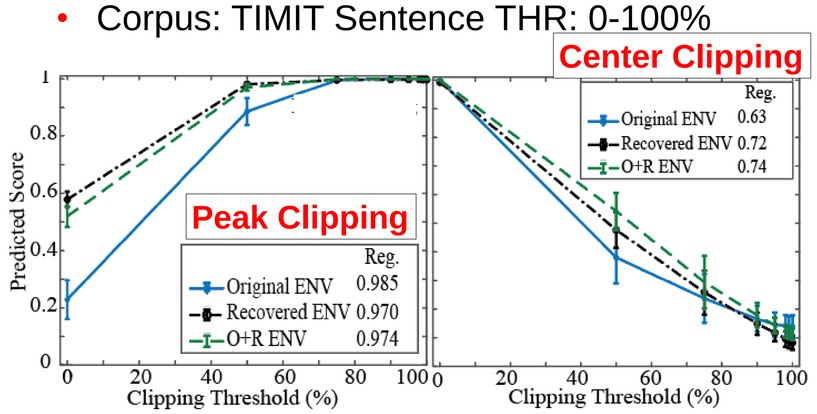


Regression Score shows Fusion of Recovered + Original Envelopes achieves highest accuracy.

5. RESULTS

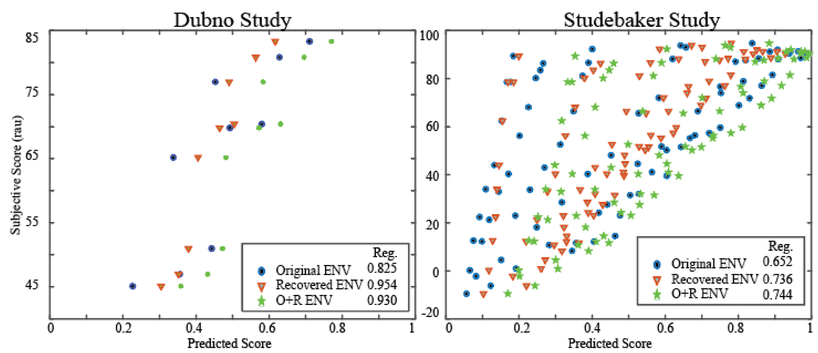
Effects of Distortion (cont.)

2. Peak & Center Clipping (C-C)
   - SNR: 5 dB
   - Corpus: TIMIT Sentence THR: 0-100%



Regression Score shows: Peak Clipping: Orig. ENV Better; Center Clipping: Fusion of Recovered + Original Envelopes best accuracy.

Subjective Study:



Final Correlation with Dubno data: Recovered ENV is best (0.95); with Studebaker data: Fusion of Recovered + Original ENV best (0.74)

6. CONCLUSIONS

- Proposed metric successfully predicts SP under noisy and distorted conditions
- Recovered ENV alone or with original ENV shows better correlation performance
- Comparison with existing metrics:

	P-J	P-C	C-C
Proposed	0.89	**0.98**	**0.74**
SRMR	0.84	0.97	0.64
SII	0.79	0.97	0.43
STOI	**0.96**	0.96	-0.30

REFERENCES

1) M.S. Zilany; I.C. Bruce (**2006**). *Journal of the Acoustical Society of America* **120**, 1446.
2) N. Mamun; W. Jassim; M.S. Zilany, (**2015**) *IEEE/ACM Transactions on Audio, Speech and Language Processing* **23**, 760.
3) B. C. Moore (**2008**) *Journal of the Association for Research in Otolaryngology* 9(4):399.
4) J.R. Dubno; A.R. Horwitz; J.B. Ahlstrom. (**2005**) *Journal of the Acoustical Society of America* **118**, 923.
5) G.A. Studebaker; R.L. Sherbecoe, D.M. McDaniel, C.A. Gwaltney (**1999**) *Journal of the Acoustical Society of America* **106**, 2111.

Acoustical Society of America
Victoria, Canada, November 5–9, 2018

This work was supported by the grant R01 DC016839 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.