1 General Introduction

1.1 Overview of the thesis

1.2 Theories of language comprehension

1.2.1 Predictive language processing

1.3 Speech degradation

Speech can be distorted by variability in speakers' production, like, accented speech, soft/rapid speech, or it can arise from listener-related factors like, hearing loss, auditory processing disorder. It can also be a result of noise from transmission, like ambient noise, or distortion in the transmission (e.g., telephone line). All these sources of distortion make listening condition adverse. In laboratory setup, the effect of speech distortion, and the mechanism of listening in adverse listening condition is studied using artificial distortion of speech. For example, algorithms like PSOLA and WSOLA are used to compress or elongate an auditory signal so as to change the rate of speech. White or pink noise signal is superimposed on the top of speech signal so as to add a background noise.

In the same vein, noise-vocoding is used to remove the spectral detail of the speech signal only leaving its temporal and periodicity cues (see Section X.X.X).

Noise-vocoding was initially developed as a means to reduce the information in speech signal to be transmitted through the telephone line [1, 2] — [Re-read this thoroughly]. Shannon and colleagues later used the same technique as an analogue to cochlear implant [3–5] – number of channels used in a cochlear implant are similar to the number of noise-vocoding channels in terms of their speech output and intelligibility [... cite probably Wagner et al. ...].

1.4 Comprehension of degraded speech

The first factor that determines the intelligibility of noise-vocoded speech is the number of channels. With an increase in noise-vocoding channels, speech intelligibility increases [e.g., 3, 6]. For example, speech processed through 8 channels noise vocoding is more intelligible than the speech processed through 4 channels noise vocoding [4]. Participant related variables (age, vocabulary), test materials (words, sentences, accented speech), and listening conditions (quiet, background noise) also influence the intelligibility of noise vocoding speech. Accuracy is higher for sentences than for words in isolation (CITE). Compared to quiet, response accuracy was reduced when listeners were presented with vocoded speech in the presence of a background noise. At the same level of degradation, accuracy is higher for younger adults than for older adults, i.e., with age, keeping all other variables constant, comprehension of degraded speech decreases. Other factors like sentence context and vocabulary also play a role, which will be discussed below. In sum, comprehension of degraded speech is a not only the amount of spectral details available, but also other listener and speaker related factors. How a listener utilizes available context information to 'make-up' for the impoverished auditory information is the critical factor determining intelligibility and comprehension of degraded speech.

1.4.1 Role of sentence context

Literature from sentence reading provides us with an insight how readers use the information available as the words are presented to them to make predictions about

what word they'll see next. In visual world paradigm, Altmann and Kamide [7] showed that a listeners predict upcoming word of a sentence using the cue provide by the sentence context. For example, they presented participants a picture of four objects: cake, XXX, XXX, and XXX while the participants were listening to the sentence '*The boy will eat the ...*'. Even before hearing *cake*, participants fixated at the picture of cake. This finding has been replicated multiple times [Kamide, Altmann, and Haywood [8]; Altmann and Kamide [9]; CITE other recent papers] in different languages [**MISHRA**]. This is observable in behavioral measures as well as electrophysiological measures.

Kutas and Hillyard [10] reported smaller N400 amplitude for highly probable sentence endings than for less probable sentence endings given different sentence contexts. They found that the N400 amplitude was more sensitive to sentence ending than to the constrain imposed by the preceding words. – Check new studies on this line – DeLong, Urbach, and Kutas [11] showed that listeners form probabilistic predictions about upcoming words in a sentence. In a highly predictable sentence context like 'The day was breezy so the boy went outside to fly ...', N400 amplitude was much smaller for an expected continuation 'a kite' than for an unexpected continuation 'an airplane'. This study has been further replicated in Spanish-English bilinguals [CITE] showing that when presented with ... ... Similarly, it has been observed that readers tend to skip the predictable words more than unpredictable words while reading, and predictable words are read faster than unpredictable words [12, 13]. Semantic context already provides listeners information about what the predictable word is going to be, therefore their fixation time on the predictable words is lesser than unpredictable words, and it takes lesser time to read the predictable words.

Taken together, these studies show that as the sentence unfolds, a human comprehender forms the meaning representation of the available context information, and generates prediction about what linguistic input is going to come next.

In a noisy environment, however, it is difficult to understand the context itself. While reading text that is visually degraded, readers rely on the available context information [e.g., 14]. Listeners generally rely on sentence context moreso in an adverse listening condition than in a clear listening environment. For example, Sheldon, Pichora-Fuller, and Schneider [15] showed that when the speech signal is degraded, word recognition is improved by sentence context in both younger and older adults. They presented listeners with senteces with high and low context information, noise-vocoded at different levels of spectral degradation. They found that response accuracy was higher for sentences with high context information than for the sentences with low context information. In cochlear implantees, XYZ et al., have shown that high predictability sentences result in higher accuracy than low predictability sentences in a word recognition task. Contrary to the findings of most of the studies using clean speech and reading clean text, sentence context does not *always* help in comprehension of all noise-vocoded speech. When the noise-vocoded speech is least degraded, listeners might rely mostly on the bottom-up auditory input than on top-down predictions generated from the sentence context; hence, context does not render benefit in this case. For example, in the 32-channels noise-vocoded speech in Obleser and Kotz [16], sentence context, and consecutively top-down semantic prediction does not yield any benefit in sentence comprehension when compared to 4-channels noise-vocoded speech.

In summary, sentence context provides information necessary to generate top-down predictions. Listeners use this context information and form predictions to better comprehend degraded speech.

1.4.2 Effect of aging

1.5 Research motivation

References

[1] Homer Dudley. "The vocoder". In: *Bell Laboratories Record* 18.4 (1939), pp. 122–126.

[2] "The vocoder". In: *Nature* 145.3665 (1940), p. 157.

[3] R. V. Shannon et al. "Speech Recognition with Primarily Temporal Cues". In: *Science* 270.5234 (Oct. 13, 1995), pp. 303–304. URL: http://dx.doi.org/10.1126/science.270.5234.303.

[4] Philipos C. Loizou, Michael Dorman, and Zhemin Tu. "On the number of channels needed to understand speech". In: *The Journal of the Acoustical Society of America* 106.4 (Oct. 1999), pp. 2097–2103. URL: http://dx.doi.org/10.1121/1.427954.

[5] Robert Shannon, Qian-Jie Fu, and John Galvin Iii. "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation". In: *Acta Oto-Laryngologica* 124.0 (Apr. 1, 2004), pp. 50–54. URL: http://dx.doi.org/10.1080/03655230410017562.

[6] Matthew H. Davis et al. "Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences." In: *Journal of Experimental Psychology: General* 134.2 (2005), pp. 222–241. URL: http://dx.doi.org/10.1037/0096-3445.134.2.222.

[7] Gerry T.M Altmann and Yuki Kamide. "Incremental interpretation at verbs: restricting the domain of subsequent reference". In: *Cognition* 73.3 (Dec. 1999), pp. 247–264. URL: http://dx.doi.org/10.1016/s0010-0277(99)00059-1.

[8] Yuki Kamide, Gerry T.M Altmann, and Sarah L Haywood. "The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements". In: *Journal of Memory and Language* 49.1 (July 2003), pp. 133–156. URL: http://dx.doi.org/10.1016/S0749-596X(03)00023-8.

[9] Gerry T.M. Altmann and Yuki Kamide. "The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing". In: *Journal of Memory and Language* 57.4 (2007), pp. 502–518.

[10] Marta Kutas and Steven A. Hillyard. "Brain potentials during reading reflect word expectancy and semantic association". In: *Nature* 307.5947 (Jan. 1984), pp. 161–163. URL: http://dx.doi.org/10.1038/307161a0.

[11] Katherine A DeLong, Thomas P Urbach, and Marta Kutas. "Probabilistic word pre-activation during language comprehension inferred from electrical brain activity". In: *Nature Neuroscience* 8.8 (July 10, 2005), pp. 1117–1121. URL: http://dx.doi.org/10.1038/nn1504.

[12] Steven Frisson, Keith Rayner, and Martin J. Pickering. “Effects of contextual predictability and transitional probability on eye movements during reading”. In: *Journal of Experimental Psychology: Learning Memory and Cognition* 31.5 (2005), pp. 862–877.

[13] Keith Rayner et al. “Eye movements and word skipping during reading: Effects of word length and predictability.” In: *Journal of Experimental Psychology: Human Perception and Performance* 37.2 (Apr. 2011), pp. 514–528. URL: http://dx.doi.org/10.1037/a0020990.

[14] Catherine Clark, Sara Guediche, and Marie Lallier. “Compensatory cross-modal effects of sentence context on visual word recognition in adults”. In: *Reading and Writing* 34.8 (2021), pp. 2011–2029. URL: https://doi.org/10.1007/s11145-021-10132-x.

[15] Signy Sheldon, M. Kathleen Pichora-Fuller, and Bruce A. Schneider. “Priming and sentence context support listening to noise-vocoded speech by younger and older adults”. In: *The Journal of the Acoustical Society of America* 123.1 (Feb. 2008), pp. 489–499. URL: http://dx.doi.org/10.1121/1.2783762.

[16] Jonas Obleser and S. A. Kotz. “Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network”. In: *Cerebral Cortex* 20.3 (June 26, 2009), pp. 633–640. URL: http://dx.doi.org/10.1093/cercor/bhp128.