Recognition of natural-rate, time-compressed, and natural fast-rate sentences by younger and older listeners

Danielle Zion, Carol Espy-Wilson, & Sandra Gordon-Salant

University of Maryland, College Park, MD

Introduction

Previous studies have compared natural-rate speech to "clear" speech, which is the term used to describe a natural effort to slow the rate of speech and over-articulate speech to improve intelligibility (Picheny, Durlach, & Braida, 1986; Krause & Braida, 2002), but few studies have investigated the changes that occur naturally when increasing the rate of speech. Instead, many studies use uniform time compression (UTC), which is useful for investigating the effects of rapid speech presentation, but it may not necessarily simulate the acoustic changes that occur naturally with rapid articulation. It is well established in the literature that older listeners, even those with normal hearing, experience difficulty with rapid speech rates (e.g., Gordon-Salant & Fitzgibbons 1993, 2004), but differences in performance with natural fast speech are not as well understood. In this study, both natural fast and UTC speech were used to determine if recognition of UTC speech is comparable to that of natural fast speech, and if recognition performance for these forms of rapid speech is influenced by contextual information, background noise, and age.

Method

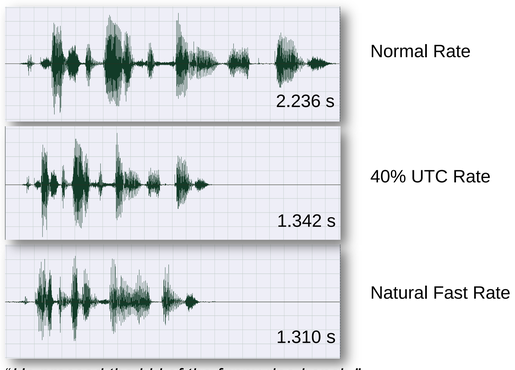
Participants:

- Two listener groups (all native speakers of English):
  - Young Normal Hearing (YNH): 19-22 years (n=13); $\overline{X}$= 20 yrs
  - Older Normal Hearing (ONH): 66-76 years (n=12); $\overline{X}$= 69 yrs
- Pure tone thresholds ≤25 dB HL, 250-4000 Hz (ANSI, 2010)

Stimuli:

- IEEE Sentences (1969); two contexts: High Probability (HP) and Anomalous Probability (AP) (n = 60 sentences/ea)
  - HP (original IEEE sentences), e.g., "*Four hours of steady work faced us*."
  - AP (Herman & Pisoni, 2000): syntactically correct but semantically meaningless, e.g., "*He pressed the bid of the funny ripe bench*."
- Sentences recorded by a male talker at a normal rate and natural fast rate
- A set of 40% UTC sentences was generated by modifying the normal rate recordings using Praat (version 4.3; Boersma & Weenick, 2004).
  - Natural fast rate sentences approximated, on average, 40% UTC.

Figure 1. Sample waveforms of a sentence in each of the three rates.



"*He pressed the bid of the funny ripe bench.*"

Procedure:

- Sentence lists (n = 10 sentences/list) presented at 65 dB SPL either in quiet or noise (4-talker speech babble, +10 dB SNR)
- Each listener tested in 12 conditions: 2 contexts (HP, AP) x 3 rates (Normal, 40UTC, Fast) x 2 environments (Quiet , Noise)
- List order and condition to list assignment randomized for each listener

Results

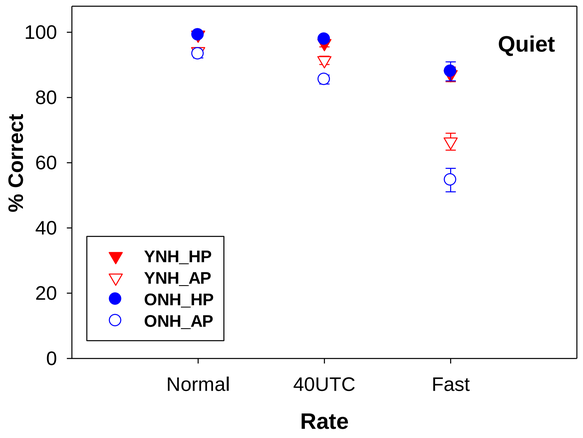


Figure 2. Mean recognition scores of the two groups for the HP and AP sentences at three rates in quiet. ANOVA revealed significant main effects of context (p<.001) and rate (p<.001) and significant context x group (p=.003) and context x rate (p<.001) interactions.

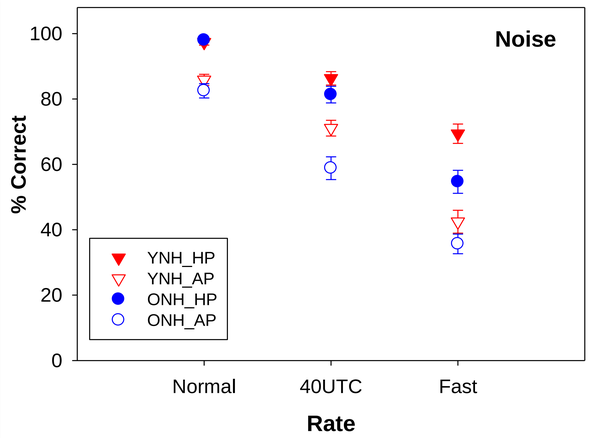


Figure 3. Mean recognition scores of the two groups for the HP and AP sentences at three rates in noise. ANOVA revealed significant main effects of context (p<.001), rate (p<.001), and group (p=.003) and a significant rate x group (p<.05) interaction.

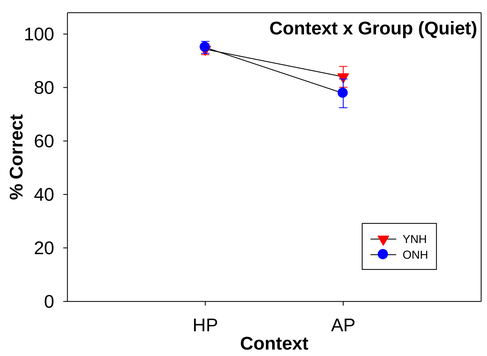


Figure 2a. Older listeners performed significantly worse (p=.001) than younger listeners for AP but not HP sentences. Context effect significant (p<.001) for both groups.

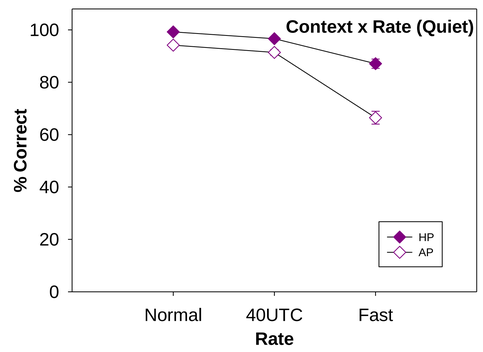


Figure 2b. Context was significant (p<.001) at each rate. Rate effect significant (p≤.006) for all paired comparisons except HP normal rate vs. 40UTC.

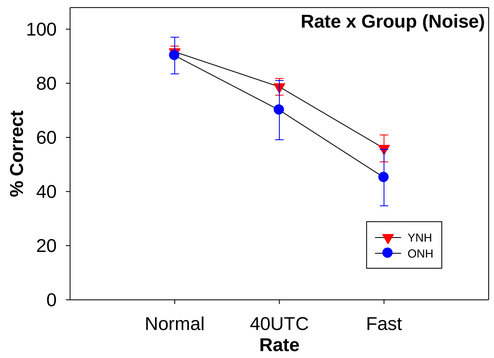


Figure 3a. Mean recognition scores of the two subject groups, collapsed across context, for the three rates in noise. Group effect significant for 40UTC (p=.012) and fast (p=.005), but no difference between groups for the normal rate (p=.52). Rate effect significant (p<.001) for both groups.

Conclusions

The main findings in this study can be summarized as follows:

Rate effect

- Differences in recognition of UTC and natural fast speech are apparent.
  - In quiet, performance deteriorates at the natural fast rate, even when context and environment are favorable.
  - In noise, the rate effect is prominent, with poorest recognition performance for natural fast speech.
- Preliminary acoustic analyses of the natural fast recordings show:
  - shorter vowel durations, omissions and distortions of consonants, including lenition, relative to natural-rate speech.
- Although the UTC speech materials approximated the average duration of the natural fast sentences, they appear to underestimate the listening difficulty encountered with natural fast speech.
  - This difference may depend on the amount of UTC applied: Brungart et al. (2007) used UTC to accelerate slow and conversational audio-visual speech to a fast rate and found performance was equivalent with natural fast speech and 30UTC conversational speech, but worse with 66UTC slow speech.
- The perceptual and acoustic differences in natural fast versus UTC speech suggest that different strategies may be employed by listeners in processing these two types of rapid speech (Janse, 2004).

Age effect

- Age effects were observed in quiet and noise conditions, but varied as a function of context and rate.
  - Older listeners performed worse than younger listeners when the context cue was unavailable in quiet (Figure 2a).
    - Thus, in favorable listening conditions, older listeners take advantage of context to perform similarly to younger listeners
  - Younger listeners outperformed older listeners in noise at the two fast rates, but not the normal rate. (Figure 3a).
    - These findings generally reflect the widely accepted notion that older listeners have more difficulty processing rapid speech.

Context effect

- All listeners generally perform better with HP than AP sentences, for all rate conditions in quiet and in noise.
  - In quiet, context effects are more apparent for rapid rate anomalous sentences than for normal rate speech.
  - Additionally, the magnitude of the difference between recognition of HP vs. AP sentences is greater for natural fast speech than for UTC speech (Figure 2b).

Taken together, the results of this study suggest that even when overall duration is equal, listening to natural fast speech is more difficult than listening to artificially time compressed speech. This is true even when listening conditions are favorable, with no background noise and use of speech materials that provide contextual cues. In less favorable conditions, older listeners, who are known to have difficulty processing rapid speech information, may be at an even greater disadvantage despite normal hearing sensitivity.

References

1. Brungart, D. S., van Wassenhove, V., Brandewie, E., & Romigh, G. (2007). The effects of temporal acceleration and deceleration on AV speech perception. *Auditory-Visual Speech Processing (AVSP 2007), Kasteel Groenendaal, Hilvarenbeek, The Netherlands, 2007*, 27.
Gordon-Salant, S. & Fitzgibbons, P.J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Hear Res, 36*, 1276-1285.
2. Gordon-Salant, S. & Fitzgibbons, P.J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *JASA, 115(4)*, 1808-1817.
3. Herman, R., & Pisoni, D.B. (2000). Perception of elliptical speech by an adult hearing-impaired listener with a cochlear implant: Some preliminary findings on coarse-coding in speech perception. In *Research on Spoken Language Processing Progress Report No 24* pp. 87-112 Bloomington IN: Speech Research Laboratory, Indiana University.
4. IEEE. (1969). IEEE recommended practice for speech quality measurements, *IEEE Report* No. 297.
5. Janse, E. (2004). Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication, 42(2)*, 155-173.
6. Krause, J.C. & Braida, L.D. (2002) Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *JASA, 112(5)*, 2165-2172.
7. Picheny, M.A., Durlach, N.I., & Braida, L.D. (1985). Intelligibility differences between clear and conversational speech. *J Speech Hear Res, 28*, 96-103.

Acknowledgements

This work was supported by a grant from the UMD-ADVANCE Program (funded by NSF).