Cross-Rate Variation in the Intelligibility of Dual-Rate Gated Speech in Older Listeners

Valeriy Shafiro, Stanley Sheft, Robert Risley

Dept. of Communication Disorders & Sciences, Rush University Medical Center, Chicago, USA

Valeriy_Shafiro@rush.edu, Stanley_Sheft@rush.edu, Robert_Risley@rush.edu

Abstract

Intelligibility of sentences gated with a single primary rate (0.5-8 Hz, 25-75% duty cycle) or gated with an additional concurrent rate of 24 Hz and a 50% duty cycle was examined in older normal-hearing and hearing-impaired listeners. With a stronger effect of age than hearing loss, intelligibility tended to increase with primary rate and duty cycle, but varied for dual-rate gating. Reduction in the total amount of speech due to concurrent 24 Hz gating had little effect on the intelligibility for the lowest and highest primary rates, but was detrimental for rates between 2 to 4 Hz, mimicking the pattern previously obtained from young normal-hearing listeners. The dual-rate intelligibility decrement with a 2 Hz primary rate significantly correlated with speech intelligibility in multi-talker babble, suggesting overlap of perceptual processes. Overall, findings reflect interaction of central and peripheral processing of speech occurring on different time scales.

Index Terms: gating, aging, hearing loss, speech in noise

1. Introduction

Speech perception involves integration of continuous time-varying acoustic input into discrete higher order categories such as phonemes, syllables, and words. Perceptually relevant information, temporally distributed in the speech signal, is processed simultaneously on several time scales [1],[2]. For instance, (sub)segmental acoustic cues to phonemes may be based on spectro-temporal variation on the order of tens of milliseconds (e.g., formant transitions), while perception of intonation or post-perceptual processing of syntactic and semantic information may require a longer processing time on the order of hundreds of milliseconds [3].

Redundant acoustic specification of underlying perceptual categories ensures robust speech perception under adverse listening conditions when speech signals are reduced through signal transmission losses or masked by noise. Miller & Licklider [4] first demonstrated that with adequate sampling of the signal, speech perception can remain accurate even with as much as 50-75% of the original signal discarded or masked by noise. In their work, speech gated at a rate of 6 Hz or higher with a 50% duty cycle was as intelligible as the original signal. At slower gating rates, however, intelligibility was reduced, as explained by the authors, because of insufficient frequency of sampling of each word. Later known as "glimpsing" or "multiple looks," Miller & Licklider's explanation for intelligibility based on sampling of transmitted speech information has been widely embraced [5]. Though effective in accounting for many specific experimental results, glimpsing models do not consider possible effects due to the processing of various linguistic units on multiple time scales.

An extension of Miller & Licklider's method designed to examine involvement of multiple time scales in speech perception was recently introduced by Shafiro and colleagues [6]. Speech perception at slow gating rates of about 0.5 – 1 Hz, where the duration of the remaining and discarded speech approaches that of whole words, could be expected to involve more centrally based contextual and semantic processing to fill in the missing information, with the perception of retained words requiring relatively little effort. On the other hand, perception of speech interrupted at faster rates (e.g., 8 Hz or above), which samples each phonemic segment within a word, involves greater reliance on temporal smoothing over small gaps in the signal. However, for speech gated at a single fast rate, the contributions of perceptual processing on a shorter time scale overlaps with the longer temporal scale of context-based processes resulting in high overall intelligibility.

To avoid this potential confound and differentially assess the contribution of perceptual processing on the shorter time scale, the researchers compared speech gated with a single rate and 50% duty cycle to that obtained with concurrent gating at two rates, with fast gating applied to speech previously gated at a slower rate (Fig. 1). Although dual-rate gating reduced the total amount of the preserved speech signal from 50% to 25%, intelligibility was higher than with single-rate gating at a 25% duty cycle, indicating that listeners were able to integrate information in the remaining speech gated with a concurrent fast rate. Shafiro's et al. [6] results, obtained from the young normal-hearing (YNH) listeners, also revealed a complex nonmonotonic pattern of interactions between the two concurrent rates, indicating that the ability to benefit from faster gating rates varied as a function of the slower primary rate (Fig. 2). Concurrent gating with a fast rate was least detrimental for sentences previously gated with either slow or fast rates that produced word and (sub)phonemic-duration intervals, but was least effective for intermediate rates of 2 – 4 Hz that produced syllable-size intervals.

The ability to utilize information remaining after additional fast-rate gating, however, may also be affected by a number of listener-specific factors. Previous research with temporally modulated maskers demonstrated considerable reduction in speech masking release for both elderly and hearing-impaired (HI) individuals [7]. In addition to presbycusis, reduced speech masking release in the elderly has been associated with a more general reduction in a number of spectral and temporal processing abilities [8]. In the context of speech gating, these factors could have a negative effect on the processing of temporal information subsequent to rapid stimulus gating.

The present study investigated the effects of age and hearing impairment on the perception of sentences gated with either a single rate or two concurrent rates, with the difference between the two indicating ability to incorporate into longer time-scale processing the information derived from a shorter time scale. A second question was whether performance with two-rate gated speech is associated with speech-in-noise intelligibility, as expected from models of the two based on the same spectral and temporal processing abilities.

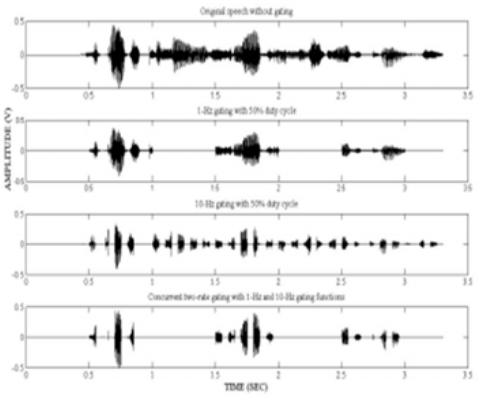

Figure 1: *An example of single- and two-rate gating methods. The top panel shows the continuous waveform of a sentence; the two middle panels show this sentence gated at a rate of 10 or 1 Hz (50% duty cycle). The bottom panel shows the sentence interrupted at the two rates concurrently. When concurrent, the faster gating rate of 10 Hz affects only the speech remaining after gating at the slower primary rate of 1 Hz, leaving 25% of the original speech signal.*

2. Method

2.1. Stimuli, Design, Procedure

Speech stimuli were HINT sentences produced by a male talker gated with either a single rate or two concurrent rates as illustrated in Fig. 1. All sentences were gated at one of the four single primary rates (0.5, 2, 4, 8 Hz) presented alone or with subsequent gating at 24 Hz. To avoid floor effects that might result for HI listeners, a 75% duty cycle was used with the four primary rates in addition to the 25 and 50% duty cycles of Shafiro et al. [6]. The concurrent faster gating rate, which always had the same starting phase as the slower rate, was applied only to sentences gated with 50 and 75% duty cycles. The order of the 20 gating conditions was randomized across sentence lists with stimuli presented diotically in quiet through Sennheiser 250 II headphones at 70 dB SPL. Listeners were asked to repeat what they heard after the presentation of every sentence. One HINT list of 10 sentences (about 50 words) was used for every gating condition. Each test sentence list was preceded by an unscored practice list of five IEEE sentences for that specific gating condition. In addition, for all listeners, speech perception in noise was evaluated following QuickSIN protocol [9] using sentences presented in the presence of a four-talker babble.

2.2. Subjects

There were two groups of older listeners with groups distinguished by pure-tone average (PTA) audiometric thresholds at 0.5, 1.0, and 2.0 kHz as either normal hearing (ENH) or hearing impaired (EHI). The eight participants in the ENH group (5 women; age range: 64-87 yrs; mean:74.5) in each ear had a PTA of 25 dB HL or less. For the EHI group, the seven listeners (6 women; age range: 60-77 yrs; mean: 66.7 yrs) had a mild-to-moderate sloping hearing loss confirmed as sensorineural by bone-conduction thresholds and tympanometry. All study participants spoke English as their first and primary language and obtained nearly perfect intelligibility on a single random list of HINT sentences presented without gating.

3. Results

Rate-intelligibility functions for ENH (Fig. 3, top panel) and EHI (Fig. 3, bottom panel) show considerable effect of duty cycle with single-rate gating (filled symbols). A 2 (group) x 3 (duty cycle) x 4 (gating rate) repeated-measures ANOVA on the arcsine-transformed intelligibility scores revealed main effects of duty cycle [$F(2, 12) = 794.38$, $p < 0.001$] and rate [$F(3, 18)=43.57$, $p < 0.001$], and a significant interaction between rate and duty cycle [$F(6, 36)=34.74$, $p < 0.001$], but no significant effect of group (i.e., hearing loss). For both groups, performance generally improved with gating rate (but see below) and worsened with reduction in duty cycle. However, across gating rates, intelligibility was not equally affected by reduction in duty cycle, which was most detrimental in the 2 – 4 Hz region, compared to lower or higher gating rates.

To evaluate the differences between single- and two- rate performance for the ENH and EHI listeners, two additional 2 (group) x 2 (gating method: single- vs. two-rate gating) x 4 (primary rate) repeated-measures ANOVAs were conducted separately for 50% and 75% duty-cycle conditions. Both ANOVAs revealed significant main effects of rate [$F (3, 18) = 87.7$ & 93, $p < 0.001$], gating method (single vs. two-rate) [$F (1, 6) = 50.95$ & 62.86, $p < 0.001$], and their significant interaction [$F (3, 18) = 7.16$ & 10.75, $p < 0.002$]. As before, there were no significant group differences, although with a 50% duty cycle, differences approached significance ($p < 0.07$), indicating a possible minor effect of hearing loss on performance with either single- or two-rate gating.

Post-hoc pairwise comparisons for each of the five rate-intelligibility functions further revealed significant local minima in the 2 – 4 Hz region, most pronounced for EHI listeners. For the ENH, the minima were found in the dual-rate conditions based off of a 50% duty cycle of the primary rate and in the single-rate conditions that had a 25% duty cycle (Fig. 3, top panel, open circles and filled triangles, respectively). For the EHI, the minima were found for all intelligibility functions except the one derived with single-rate gating at a 75% duty cycle (Fig. 3, bottom panel).

Remarkably, for both ENH and EHI groups, performance with dual-rate concurrent gating was always better than would be expected based on the total amount of the original speech remaining after gating. That is, application of a 24-Hz secondary rate to previously gated speech further reduced the amount of the original speech to 25% from the 50% that remained after initial gating with a 50% duty cycle, or to 38-42% from the 75% that remained after single-rate gating with a 75% duty cycle. (Due to differences in the duty cycles of the primary and secondary rates, the reduction in the amount of original speech varied slightly across primary gating rates for the two-rate sentences first gated with a 75% duty cycle). The intelligibility of two-rate gated sentences containing 25% of the original signal was always better than that obtained with single-rate gating with a 25% duty cycle. Furthermore, the intelligibility of two-rate gated sentences containing 38-42% of the original signal considerably surpassed that obtained with single rates and a 50% duty cycle. Both results indicate ability of the listeners in both groups to utilize speech cues from concurrent gating with a fast secondary rate. In fact, for the lowest primary rate of 0.5 Hz and the highest rate of 8 Hz, the performance with two-rate gating closely approximated that obtained with the primary rate alone.

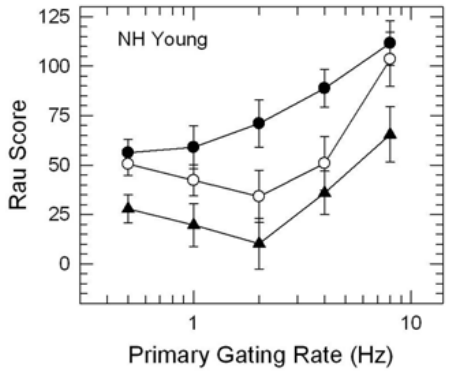

Figure 2: *Subset of rate-intelligibility functions for YNH listeners from Shafiro et al.[6]. Filled circles and triangles represent conditions with a single gating rate and a 50 or 25% duty cycle, respectively, and open circles represent dual-rate gating with subsequent 24-Hz gating with a 50% duty cycle following primary gating.*

On the other hand, the intelligibility decrement with dual-rate gating was greater for 2- and 4-Hz primary rates with both 50 and 75% duty cycles, indicating difficulty in perceiving speech fragments distributed on a finer temporal scale at these primary rates.

To evaluate potential effects of age on the ability to utilize information remaining following fast secondary gating, the performance of the ENH listeners was compared to the YNH listeners tested by Shafiro et al. [6] who participated in identical single-rate 50% duty cycle and the corresponding dual-rate conditions. The 2 x 2 x 4 repeated-measures ANOVA on the arcsine transformed intelligibility scores revealed a significant effect of age [$F(1, 7) = 48.21$, $p < 0.001$], gating method (single or two-rate) [$F(1,7) = 194.11$, $p < 0.001$] and rate [$F(3,21) = 112.05$, $p < 0.001$]. Age also interacted with rate [$F(3, 21) = 7.1$, $p < 0.01$], but not with gating method, indicating a similar pattern of cross-rate variation for both young and older NH listeners.

Finally, for all older listeners, there was a significant correlation ($r = 0.45$, $p < 0.05$) between the decrement in intelligibility with two-rate gating based off of a 2-Hz primary rate and a 75% duty cycle and the ability to perceive speech in multi-talker babble noise. This gating condition was chosen for correlation analysis as presenting the maximum range of individual variation across listeners. Significant correlation indicates that listeners who were more detrimentally affected by the fast secondary gating also tended to have more difficulty in perceiving speech in the presence of a complex masker.

4. Discussion

The present findings demonstrate the negative effects of aging on the intelligibility of sentences gated with either a single rate or two concurrent rates. On the other hand, hearing loss, while approaching significance in some conditions, did not have a strong influence on intelligibility. This might be due to the only mild-to-moderate hearing losses of the EHI listeners (all of whom demonstrated near perfect scores for intact sentences) and indicate that perceptual processing of small temporal fragments of speech produced by fast secondary gating is influenced more by age than hearing loss.

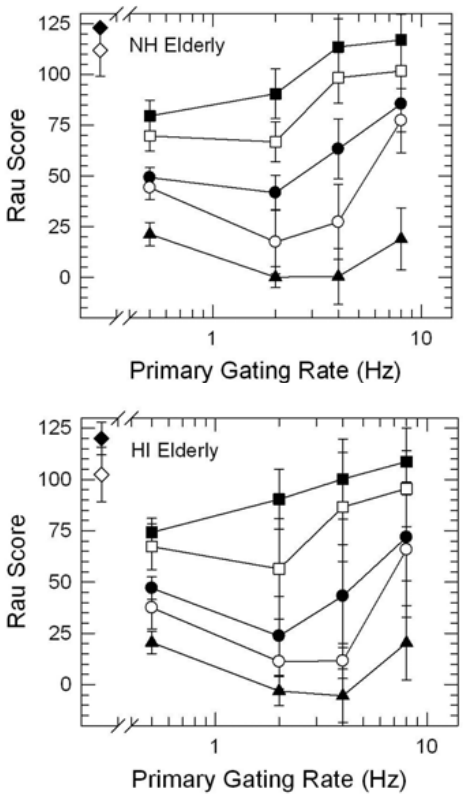

Figure 3: *Rate-intelligibility functions of the ENH (top) and EHI listeners (bottom). Legend is the same as in Fig. 2, with the addition of filled squares for single-rate gating with a 75% duty cycle, and open squares for the corresponding two-rate gating function. The filled and open diamonds in the upper left indicate baseline intelligibility with intact speech and 24-Hz single-rate gating with a 50% duty cycle, respectively.*

The finding of a strong age effect, and only a minor role of hearing loss, mirrors that of Sheft at al. [10] who tested comparable groups of YNH, ENH, and EHI listeners on discrimination of stochastic temporal fine-structure patterns. These researchers found moderate-to-strong correlations between speech intelligibility in multi-talker babble assessed with QuickSIN and both temporal fine-structure and spectral processing of stochastic modulation patterns. In the current work, the finding of a moderate correlation between intelligibility of speech in multi-talker babble and the ability to utilize information available after secondary gating with a fast rate further suggests potential overlap in the underlying processes examined by Sheft et al. [10] with those involved in the processing of speech fragments distributed on a brief time scale.

Similar to earlier findings with YNH by Shafiro et al. [6], the similar cross-rate nonmonotonic variation in intelligibility of two-rate gated speech in ENH and EHI listeners indicates that speech intelligibility was determined neither by the total amount of original speech present after secondary gating nor simply by the frequency of sampling of the speech cues. At the lowest and highest primary rates, the intelligibility of two-rate gated speech approached that obtained with a single rate, despite a substantial reduction in the amount of the remaining speech. However, for the intermediate primary rates of 2 – 4 Hz, listeners were not able to effectively utilize the information in the fast secondary rate to the same degree as at the lowest and highest rates.

This cross-rate variation in intelligibility suggests a complex pattern of interactions between speech perception processes that occur on different time scales [1], [2]. For single-rate gating at 0.5 Hz, speech-on times may contain one or more undistorted words, but also large speech-off silent intervals where no information is available. Under these conditions, speech perception would be driven mostly by more central and slower processes that fill in the missing information by using syntactic, semantic, and other higher order contextual cues, when available. Thus, secondary gating of word-sized speech intervals with a fast rate might tap into a qualitatively different set of perceptual processes that operate on a shorter time scale. In contrast, with a high primary rate of 8 Hz, both speech-on and speech-off times are short enough to sample most of the perceptually salient phonemic speech cues, so that full words can be reconstructed completely based primarily on low-level cues with relatively little uncertainty. In this case, a fast secondary rate would have minimal effect, comparable to that observed with fast primary rates alone. On the other hand, for syllable-size intermediate primary rates of 2 – 4 Hz, the duration of the speech-on intervals is less than the average word duration so that, on average, more than half of the word may be omitted during the speech-off times. In this case, the speech cues remaining after primary gating cannot effectively constrain word identification, possibly failing to provide higher order perceptual templates against which the remaining lower level speech cues can be verified.

The reduced performance obtained with fast secondary gating of primary rates of 2 – 4 Hz may thus indicate a processing bottleneck created by the need to conduct a more exhaustive higher order lexical search, while simultaneously integrating finer low-level temporal cues that remain following 24-Hz secondary gating. This explanation is consistent with the Reverse Hierarchy Theory account (RHT) of speech perception that predicts performance penalties for simultaneous perceptual access to finely grained low-level stimulus details and higher order information (such as sentence-level cues). According to RHT, listeners can access either low-level stimulus details or to high-level cues, but not to both [11].

The presence of local minima around 2 – 4 Hz is also consistent with the view that emphasizes the importance of the low-rate modulation spectrum [12]. These gating rates produce periodic interruptions that affect the region of the speech modulation spectrum that has been previously shown to be critical for speech perception [13]. The rate-intelligibility functions of the present work that show local minima resemble the inverse of the typical speech modulation spectrum. The effect of speech gating on intelligibility may in part relate to modulation masking, that is, gating rates that coincide with the dominant region of the speech modulation spectrum introduce the greatest interference. Decreasing duty cycle from 75 to 50 to 25% does increase the peak in the modulation spectrum at the gating rate. The effect of duty cycle on intelligibility is thus consistent with involvement of modulation masking. However, the introduction of a secondary gating rate has little effect on the modulation spectrum beyond the region proximal to the secondary rate. In this case, involvement of modulation masking would not predict the local minima around 2 – 4 Hz. Thus, masking in the modulation domain appears unable to fully account for the perception of temporally distributed speech, indicating that a better understanding of perceptual processes, time scales, and underlying linguistic units may be necessary to explain the pattern of cross-rate variation obtained with two-rate gating.

5. Conclusions

In line with previous work with YNH listeners, the present findings in ENH and EHI listeners reveal a cross-rate variation in intelligibility of speech concurrently gated with a fast secondary rate which cannot be accounted for by either total amount of the remaining speech or frequency of speech sampling. With a strong effect of age, but not hearing loss, the results suggest a complex rate-dependent interaction between central and peripheral perceptual processing of gated speech occurring on different time scales. A significant correlation in the ability to utilize information in the fast secondary rate and perception of speech in noise suggests a potential clinical use for the concurrent gating method.

6. Acknowledgements

Research support provided by NIH-NIDCD (R03DC008676).

7. References

[1] Rosen, S., “Temporal information in speech, acoustic, auditory, and linguistic aspects,” Phil. Trans. R. Soc. B. 336: 367-373 1992.

[2] Poeppel, D., Idsardi, W.J., and van Wassenhove, V., “Speech perception at the interface of neurobiology and linguistics,” Phil. Trans. R. Soc. B. 363:1071-1086 2008.

[3] Mattys, S.L., “The use of time during lexical processing and segmentation: A review,” Psychonomic Bulletin & Review 4:310-329 1997.

[4] Miller, G.A., and Licklider, J.C.R., “The intelligibility of interrupted speech”. J. Acoust. Soc. Am. 22:167-173 1950.

[5] Cooke, M., “A glimpsing model of speech perception in noise,” J. Acoust. Soc. Am. 119: 1562-1573 2006.

[6] Shafiro, V., Sheft, S., and Risley, R., “Perception of temporally interrupted speech: Effects of two concurrent gating rates on intelligibility,” J. Acoust. Soc. Am. 127:1991 2010.

[7] Festen, J., and Plomp. R., “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,” J. Acoust. Soc. Am. 88: 1725–1736 1990.

[8] Jin S.H., and Nelson P.G., “Interrupted speech perception: The effects of hearing sensitivity and frequency resolution,” J. Acoust. Soc. Am. 128: 881-889 2010.

[9] Killion, M.C., Niquette, P.A., Gudmundsen, G.I., et al. “ Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners, “ J Acoust Soc Am, 116: 2395-2405 2004.

[10] Sheft, S., Shafiro, V., Lorenzi, C., et al. “Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception,” Ear and Hear.: submitted.

[11] Nahum, M., Nelken, I., and Ahissar, M., “Low-level information and high-level perception: The case of speech in noise.” PLOS Biology 6, e126 (2008).

[12] Greenberg. S. “Understanding speech understanding,” Proceedings of the ESCA Tutorial and advanced research workshop on the auditory basis of speech perception, Keele, England.

[13] Drullman, R., Festen, J.M., and Plomp, R., “Effect of temporal envelope smearing on speech reception,” J. Acoust. Soc. Am. 95:1053-1064 1994.