Modulation masking release using the Brazilian-Portuguese HINT: Psychometric functions and the effect of speech time compression

John H. Grose[a], **Silvana Griz**[b], **Fernando A. Pacífico**[b], **Karina P. Advíncula**[b], and **Denise C. Menezes**[b]

[a]Dept. Otolaryngology – Head & Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7070, U.S.A

[b]Dept. Fonoaudiologia, Universidade Federal de Pernambuco, Recife/PE, CEP 50670-420, Brazil

Abstract

Objective—The Brazilian-Portuguese Hearing In Noise Test (HINT) was used to investigate the benefit to speech recognition of listening in a fluctuating background. The goal was to determine whether modulation masking release varied as a function of the speech-to-masker ratio at threshold. Speech-to-masker ratio at threshold was manipulated using the novel approach of adjusting the time-compression of the speech.

Design—Experiment 1 measured performance-intensity functions in both a steady speech-shaped noise masker and a 10-Hz square-wave modulated masker. Experiment 2 measured speech-to-masker ratios at threshold as a function of time-compression of the speech (0, 33, and 50%) in both maskers.

Study Sample—Participants were normal-hearing adults who were native speakers of Brazilian Portuguese (Experiment 1: N = 10; Experiment 2: N = 30).

Results—The slope of the performance-intensity function was shallower in the modulated masker than in the steady masker for both words and sentences. Thresholds increased with increasing time-compression in both maskers, but more markedly in the modulated masker, resulting in reduced modulation masking release with increasing time-compression.

Conclusions—Speech-to-masker ratio at threshold varies with time-compression of speech. The results are relevant to the issue of whether degree of masker modulation benefit depends on speech-to-masker ratio at threshold.

Keywords

modulation masking release; speech recognition; time compression; psychometric function

Corresponding Author: John H. Grose, Ph.D., Dept. OHNS, University of North Carolina at Chapel Hill, G190 Physicians Office Bldg., CB# 7070, 170 Manning Drive, Chapel Hill, NC 27599-7070, U.S.A. john_grose@med.unc.edu.

Introduction

Speech recognition in noise is usually more acute if the masking noise is modulated than if it is steady (e.g., Miller & Licklider, 1950; Fullgrabe et al., 2006; Bernstein et al., 2012). This benefit can be quantified as modulation masking release [MMR], which is the difference in speech recognition threshold measured in the steady and modulated maskers. The magnitude of MMR depends on a variety of stimulus factors – expanded on below – including the type of speech material, the nature of the modulating masker, and the speech-to-masker ratio at which MMR is measured. However, we begin by noting that none of this research has been done using Brazilian-Portuguese speech. Recently, a Brazilian-Portuguese version of the Hearing In Noise Test (HINT) has become available (Bevilacqua et al., 2008), and an underlying motivation for this study was to characterize MMR using this new (non-English) test. The primary research focus, however, was to determine the dependence of MMR on speech-to-masker ratio using the novel stimulus manipulation of varying the time compression of the target speech (cf. Grose et al., 2009). The goal was to determine whether the magnitude of MMR could be systematically varied by adjusting the speech-to-masker ratio at threshold using time compression.

A variety of stimulus factors affect the magnitude of MMR. One factor is the type of speech material used. A sampling of speech material that has been used to assess the benefit of a modulated masker includes: vowel-consonant-vowel stimuli (e.g., Fullgrabe, Berthommier et al., 2006; Gnansia et al., 2008), nonsense syllables (e.g., Dubno et al., 2003; Bernstein, Summers et al., 2012), monosyllabic words (e.g., Miller & Licklider, 1950; Stuart & Phillips, 1996), spondee words (e.g., Dirks & Bower, 1971), and sentences (e.g., Jin & Nelson, 2006; Desloge et al., 2010). A second stimulus factor that influences MMR is the nature of the masker. Maskers range from noise (usually speech-shaped noise [SSN]) to single- or multi-talker competitors (e.g., Festen & Plomp, 1990; Gustafsson & Arlinger, 1994; Oxenham & Simonson, 2009; Francart et al., 2011). For modulated noise (non-talker) maskers, the modulation patterns range from single- or multi-talker speech envelopes to sinusoidal or square-wave modulators. In addition, both regular and irregular duty cycles have been incorporated into these fluctuation patterns (Stuart & Phillips, 1996; George et al., 2006). A third stimulus factor that affects the magnitude of MMR is the speech-to-masker ratio at threshold; i.e., the level of the target speech relative to the masker at which recognition performance meets the criterion for threshold. Several studies have demonstrated that, for normal-hearing listeners, the benefit of masker modulation decreases as the speech-to-masker ratio increases (Oxenham & Simonson, 2009; Christiansen & Dau, 2012; Smits & Festen, 2013). The basis for this effect is that the slopes of the psychometric functions for speech recognition in steady and modulated maskers are usually not parallel. Rather, the psychometric function for speech recognition in a modulated masker is typically shallower than that in a steady masker (e.g., Dirks & Bower, 1971; Bernstein & Grant, 2009; Oxenham & Simonson, 2009).

Bernstein and colleagues (Bernstein & Grant, 2009; Bernstein & Brungart, 2011; Bernstein, Summers et al., 2012) and Smits and Festen (2013) have hypothesized that this dependence of MMR on the speech-to-masker ratio at threshold is the predominant reason for the difference in MMR magnitude between listeners with normal hearing and listeners with

cochlear hearing loss. Several studies have demonstrated that listeners with cochlear loss exhibit reduced benefit from masker modulation (e.g., Festen & Plomp, 1990; Peters et al., 1998; Jin & Nelson, 2006; Lorenzi et al., 2006; Christiansen & Dau, 2012). A number of reasons for this have been suggested including reduced audibility (Desloge, Reed et al., 2010) and poor temporal resolution (Festen, 1987). However, Bernstein and colleagues argue that the difference is due largely to the fact that the listeners with cochlear hearing loss generally have higher masked speech thresholds than do listeners with normal hearing; i.e., they receive the target speech at a higher speech-to-masker ratio at threshold. Given the differing slopes of the psychometric functions, this leads to a reduced MMR magnitude in the impaired listeners. They support their hypothesis by demonstrating that if normal-hearing listeners receive speech under conditions that result in higher speech-to-masker ratios at threshold, the listeners' MMR concomitantly declines. Manipulations that have resulted in altered speech-to-masker ratios at threshold in normal-hearing listeners include varying response set size (Buss et al., 2009; Bernstein, Summers et al., 2012), filtering the speech (Oxenham & Simonson, 2009; Christiansen & Dau, 2012), and testing non-native speakers of the target speech (Nakamura & Gordon-Salant, 2011; Calandruccio et al., 2014). Oxenham and Simonson (2009) pointed out that any manipulation that reduces speech redundancy should result in an increased speech-to-masker ratio at threshold. If correct, a similar effect should be observed when speech redundancy is manipulated by means of time compression. The main purpose of this study is to test this novel hypothesis.

Time-compression of speech is a manipulation that has been used in a variety of contexts such as determining sources of age-related speech processing deficits (e.g., Gordon-Salant & Fitzgibbons, 2001; Jenstad & Souza, 2007; Grose, Mamo et al., 2009). Quantification of time compression is usually expressed in terms of the proportion of the time-waveform content that is excised. For example, 33%-time-compression implies that one-third of the original time waveform has been removed. Compression algorithms typically remove segments of the spoken phrase with the goal of increasing speech rate without markedly affecting pitch contour or generating other distortions (e.g., by removing complete pitch periods of voiced vowels and portions of the inter-word gaps). The net result is that the redundancy of the time-compressed speech is reduced and the speech recognition threshold is increased, particularly in masking. In the context of the current hypothesis, it is expected that this elevation of speech-to-masker ratio at threshold should be associated with a diminished magnitude of MMR. Support for this can be gleaned from the temporal envelope study of Grose et al. (2009) where the data pattern shows that time-compression of target speech elevates the speech-to-masker ratio at threshold and, concomitantly, reduces the magnitude of MMR. This finding is shown in Fig. 1 which redraws data from that study to demonstrate that the speech-to-masker ratio at threshold (in this case, 50% correct recognition of IEEE sentences presented in SSN) depends on both the masker type (steady or 16-Hz modulated) and the time-compression of the target speech (uncompressed or 33% time-compressed). Relative to the uncompressed speech, speech-to-masker ratios at threshold were elevated for the 33% time-compressed speech but markedly more so in the modulated masker than in the steady masker. As a result, the magnitude of MMR was greater for the uncompressed speech than for the compressed speech. Because the focus of that study was on temporal envelope processing, the basis of the time-compression effect in

terms of the underlying psychometric functions was not further considered. The purpose of the present study, therefore, was to specifically focus on the role of psychometric function shape in testing the hypothesis that the speech-to-masker ratio at threshold increases as a function of time compression and, concomitantly, that the magnitude of MMR decreases. This hypothesis was tested using the Brazilian-Portuguese version of the HINT. A preliminary experiment was undertaken to verify that the new Brazilian-Portuguese HINT exhibits the expected dependency of the slope of the performance-intensity function on the modulation characteristics of the masker. The main experiment (Experiment 2), manipulated time-compression.

Experiment 1. Performance-Intensity functions for Brazilian-Portuguese HINT in steady and modulated noise

The purpose of this experiment was to generate performance-intensity functions for Brazilian-Portuguese speech under conditions of steady and modulated masking. The goal was to verify the assumption underlying the main experiment (Experiment 2) that the slope of the performance-intensity function measured in a modulated masker is shallower than that measured in a steady masker and, therefore, that the magnitude of MMR differs as a function of speech-to-masker ratio.

Method

Subjects—The participants were 10 young adults (5 female) ranging in age from 19–24 yrs (mean 20.4 yrs). Previous work from our laboratory has shown that this sample size is appropriate to support the effect sizes examined here (Grose, Mamo et al., 2009). All participants were native speakers of Brazilian Portuguese and had normal hearing as documented by pure-tone thresholds   20 dB HL for the octave frequencies 250 – 8000 Hz. None reported any history of otologic or neurologic disorder. Each subject gave signed consent for participation in the study, which was approved by the Brazilian National Committee of Ethics in Research (*Comissão Nacional de Ética em Pesquisa* – CONEP - (CAAE): 02466612.2.0000.5208).

Stimuli—The test material was the Brazilian-Portuguese HINT (Bevilacqua, Banhara et al., 2008). This version of the HINT consists of 12 lists of 20 sentences per list. The sentences, originally recorded at the House Research Institute, U.S.A., were resampled to 24,414 Hz and scaled to have equal RMS levels across all sentences.

The masker was the SSN supplied with the HINT. This noise has the same spectral shape as the long-term average speech spectrum (LTASS) of the sentences comprising the test material. The masker was presented continuously under two conditions: steady and modulated. In the steady condition, the masker was presented at a constant level of 65 dB SPL. In the modulated condition, the masker oscillated between 65 dB SPL and 30 dB SPL at a rate of 10 Hz (cf. Desloge, Reed et al., 2010). For both steady and modulated conditions, the masker level of 65 dB SPL was used as the denominator for speech-to-masker ratio derivations. The modulation pattern was quasi-square-wave, with 1-ms ramps imposed on the transitions between the high and low levels.

Procedure—For each masker type, one complete HINT list (20 sentences) was presented at each of six speech-to-masker ratios. For the steady masker, the speech-to-masker ratios were −14, −11, −8, −5, −2, and 1 dB; for the modulated masker, the ratios were −23, −20, −17, −14, −11, and −8 dB. These ratios were selected based on pilot listening as being optimal for capturing the steeply sloping portion of the performance-intensity function for word-level recognition in most normal-hearing young adults. The sentences were output to the right phone of a Sennheiser HD580 headset through a digital signal processing platform (Tucker-Davis RX6) under the control of a computer running a custom Matlab™ script. The subject listened monaurally within a single-walled sound-attenuating booth and was instructed to repeat aloud as much of the perceived sentence as possible, even if the perceived sentence did not appear to make grammatical or semantic sense. Outside the booth, the experimenter monitored the oral response of the subject through headphones linked to a microphone within the booth. As each sentence was presented to the subject, the text of the sentence was simultaneously displayed on the computer screen in front of the experimenter, with each word highlighted in a position-sensitive shaded rectangle. The experimenter coded the errors by clicking the computer mouse on the words that were omitted or repeated incorrectly. The computer program registered and tracked these word-wise errors and computed a percent correct score at the end of each list. In addition to performance at the word level, performance at the sentence level was also monitored. Each subject listened to 12 randomly selected lists: one list for each of six speech-to-masker ratios for each of two masker types. Performance in the steady masker was measured first and, for each masker type, testing began at one of the higher speech-to-masker ratios.

Results and Discussion

The results for speech recognition performance are displayed in Fig. 2. Each panel shows data from one subject, plotting percent correct against speech-to-masker ratio for each of four sets of data: (1) word recognition in the steady masker (unfilled circles); (2) word recognition in the modulated masker (unfilled squares); (3) sentence recognition in the steady masker (filled circles); and (4) sentence recognition in the modulated masker (filled squares). The final panel displays the group mean for each of the four data sets. Each data set was fitted with a logistic function that minimized the sum of squares error, and these functions are shown in each panel, with solid lines denoting the fits to the steady masker data and dashed lines denoting the fits to the modulated masker data. The general pattern of these data indicates that the performance-intensity functions for word recognition are shifted leftwards on the abscissa relative to the sentence recognition functions; i.e., for a given speech-to-masker ratio performance is better at the word level than at the sentence level. Boothroyd and Nittrouer (1988) have shown that there is a predictable relationship between the probability of correctly identifying a word and the meaningfulness of the context in which it is presented, such as a sentence, and that this relationship varies with speech-to-masker ratio. Thus, as the predictability of a sentence context increases, the probability of correctly identifying individual words within that sentence also increases. The current data conform to their pattern of results showing that, for a given speech-to-masker ratio, the probability of perceiving a word correctly within a sentence is higher than the probability of perceiving the whole sentence correctly.

From the performance-intensity functions shown in Fig. 2, the speech-to-masker ratio associated with the estimated 50% correct point was identified as well as the slope of the function. For the word-level and sentence-level functions, respectively, the dB difference between the 50% points measured in the steady and modulated maskers was taken as the magnitude of MMR. Dealing first with the word-level performance, the mean speech-to-masker ratio at the 50% threshold was −8.8 dB (standard deviation [SD] = 0.86 dB) for the steady masker. This threshold is lower than that usually measured for isolated single words, such as monosyllabic or spondaic words (Wilson et al., 2008), but is similar to the −7.4 dB found for word identification in AzBio sentence context by Buss et al. (2014). For the modulated masker, the speech-to-masker ratio at threshold for word recognition was −18.9 dB (SD = 2.1 dB). This compares well with the −19.1 dB found by Buss et al. (2014) for words in AzBio sentences who also used a 10-Hz modulated masker. The dB difference between the 50% correct points measured in the steady and modulated maskers was taken as the magnitude of MMR. Across the ten subjects, the average MMR was about 10 dB (SD = 1.52 dB). A paired-sample t-test indicated that the MMR was significantly greater than zero (t[9] = 20.87; p < 0.001). Turning now to the sentence-level performance, the mean speech-to-masker ratio at the 50% threshold in the steady masker was −6.5 dB (SD = 0.76 dB). This is lower than the −4.6 dB tabulated for the 50th percentile performance for co-located speech and noise for the Brazilian-Portuguese HINT found by Bevilacqua et al. (2008) and to the −3.9 dB which represents the average across 13 language versions of the HINT (Soli & Wong, 2008). Threshold here in the modulated masker was −15.2 dB (SD = 1.5 dB), yielding an average MMR across the ten subjects of 8.6 dB (SD = 1.34 dB). A paired-sample t-test indicated that this MMR was significantly greater than zero (t[9] = 20.3; p < 0.001).

Of primary interest to this study was the finding that the slopes of the psychometric functions differed between the steady and modulated maskers. For word recognition in the steady masker, the mean slope was about 16.2%/dB and, for the modulated masker, the mean slope was about 7.3%/dB. The observed slope in the steady masker is steeper than the 9.9%/dB found by Buss et al. (2014) for word recognition within AzBio sentence context. The shallower slope in the modulated masker compares reasonably well with their slope of 6.8%/dB measured in a 10-Hz modulated masker. A paired-sample t-test indicated the slope for the steady masker was significantly steeper than that of the modulated masker (t[9] = 5.98, p < 0.001). For sentence recognition in the steady masker, the mean slope was about 14.1%/dB and, for the modulated masker, the mean slope was about 7.9%/dB. The observed slope in the steady masker is steeper than the 11.4%/dB measured during the development phase of the Brazilian-Portuguese HINT (Bevilacqua, Banhara et al., 2008), or the value of 10.3%/dB averaged across 13 language versions of the HINT (Soli & Wong, 2008). A paired-sample t-test indicated that the slope for the steady masker was significantly steeper than that of the modulated masker (t[9] = 7.56, p < 0.001). This difference in slopes between the steady and modulated maskers reinforces the dependence of the derived MMR on the criterion point on the psychometric function associated with threshold performance. Thus, the assumption underlying Experiment 2 is confirmed that the magnitude of MMR measured using the Brazilian-Portuguese HINT varies as a function of the signal-to-masker ratio at threshold at which performance is measured.

Experiment 2. Modulation masking release as a function of speech time compression

The purpose of Experiment 2 was to vary the speech-to-masker ratio at threshold by systematically manipulating the level of speech redundancy in the Brazilian-Portuguese HINT sentences using time compression. The goal was to test the hypothesis that, in normal-hearing listeners, the magnitude of MMR decreases as the speech-to-masker ratio at threshold increases.

Method

Subjects—The participants were 30 young adults (23 female) ranging in age from 17–25 yrs (mean 21.5 yrs). They were divided into three groups of 10 participants, with each group assigned to one of three conditions (see below). None of the participants had taken part in Experiment 1, and they were all native speakers of Brazilian Portuguese. All had normal hearing as documented by pure-tone thresholds 20 dB HL for the octave frequencies 250 – 8000 Hz in the test ear, except for one subject who had a threshold of 25 dB HL at 8000 Hz. None reported any history of otologic or neurologic disorder. Each subject gave signed consent for participation in the study, which was approved by the Brazilian National Committee of Ethics in Research (*Comissão Nacional de Ética em Pesquisa* – CONEP - (CAAE): 02466612.2.0000.5208).

Stimuli—The speech stimuli used in this experiment were the sentences of the Brazilian-Portuguese HINT. The sentences were either presented in their original, uncompressed format (time compression [TC] = 0%), or at two levels of compression where either one-third (TC = 33%) or half (TC = 50%) of the sentence waveform had been removed. Time compression was undertaken off-line using the proprietary iZotope Radius algorithm within Adobe Audition™ that specifies a stipulated change in waveform duration while maintaining speech realism. Pilot listening indicated that the resulting time-compressed sentences were perceived as rapid speech having the same pitch attributes as the original sentences, and were otherwise free of noticeable distortions. As with Experiment 1, the masker was a noise with the same LTASS as the original sentences. The masker was either output continuously at a level of 65 dB SPL or was square-wave modulated between 65- and 30-dB SPL at a rate of 10 Hz. Stimuli were output via a digital signal processing platform (RX6, Tucker-Davis Technologies) and presented to the right ear through a Sennheiser HD580 headphone.

Procedure—Speech recognition thresholds were measured using a two-down, one-up adaptive procedure that converged on the 71% correct point. The subject was seated inside a single-walled, sound-attenuating booth and repeated back each sentence as it was presented. As with Experiment 1, the text of the sentence was simultaneously displayed on the computer screen in front of the experimenter, with each word highlighted in a position-sensitive shaded rectangle. The experimenter used the computer mouse to tag the words that were omitted or repeated incorrectly. However, for the purposes of the adaptive procedure, the sentence was given an overall score of 'correct' or 'incorrect' wherein the complete sentence had to be repeated accurately for a score of 'correct' and any error resulted in a score of 'incorrect.' Following two correct sentences in a row, the presentation level of the

next sentence was reduced by 2 dB; following one incorrect sentence, the presentation level of the next sentence was increased by 2 dB. A threshold estimation track continued until six reversals in level direction had occurred. The threshold for that track was taken as the mean of the final four reversal levels. For each subject, three threshold estimates were obtained for a given masker condition, with an additional estimate obtained if the range of the first three exceeded 3 dB. The final threshold for that condition was taken as the mean of all estimates obtained. The first threshold estimation track was always undertaken with a steady masker, but the masker type for subsequent tracks was quasi-randomized across replications. Because sentences were presented without repetition, a single subject could participate in only one level of TC (0, 33, or 50%); i.e., 10 subjects per TC level. This was because the collection of 6 – 8 threshold estimates (up to 4 estimates in each of the steady and modulated maskers) approached the maximum number of threshold-track trials that could be obtained without the subject hearing any given sentence more than once. The adaptive procedure, including stimulus presentation, was controlled by a custom Matlab™ script.

Results and Discussion

The results are displayed in Fig. 3 where speech recognition thresholds are plotted for each level of time compression. Individual thresholds measured in the steady masker are shown as filled circles and those measured in the modulated masker are shown as unfilled circles; for each subject, a vertical bar connects the respective steady and modulated thresholds. Group mean thresholds for the steady masker are shown as filled squares while those for the modulated masker are shown as unfilled squares (1 SD error bars). Mean thresholds and derived MMR are also tabulated in Table 1 for the three levels of time compression. Inspection of the data pattern in Fig. 3 and Table 1 suggests three features: (1) for all subjects, thresholds in the modulated masker are lower than in the steady masker signifying the occurrence of a positive MMR; (2) thresholds generally worsen as the amount of TC increases; and (3) the elevation in threshold with increasing TC is greater for the modulated masker than for the steady masker, resulting in a reduction in MMR with increasing TC. To assess this, the data were submitted to an analysis of variance (ANOVA) with one between-subjects factor (*TC level*) and one within-subjects factor (*masker type*). The analysis indicated a significant effect of *TC level* (F[2,27] = 172.9; p < 0.001), a significant effect of *masker type* (F[1,27] = 313.3; p < 0.001), and a significant interaction between these factors (F[2,27] = 37.3; p< 0.001). Simple effects testing indicated that the effect of *masker type* was significant at each level of *TC level* (p < 0.001). This result indicates that the speech-to-masker ratio required for criterion speech recognition increased with increasing time compression for both the steady and modulated maskers. However, the speech-to-masker ratio increased more rapidly for the modulated masker than for the steady masker. To determine whether the derived MMR decreased with increasing time compression, a second ANOVA was undertaken on the derived MMR magnitudes across TC level. The analysis revealed a significant effect of *TC level* (F[2,27] = 37.3; p < 0.001), and linear contrasts indicated that MMR magnitude was significantly different across each of the three levels of time compression (p < 0.05).

The results of Experiment 2 therefore support the hypothesis that the magnitude of MMR for speech depends on the speech-to-masker ratio at threshold. An important issue, however, is

whether time-compression of speech affects MMR simply through a change in signal-to-masker ratio at threshold or whether it affects MMR by some other means as well. To consider this issue, it was first necessary to derive estimates of the performance-intensity functions associated with the time-compressed speech. This estimation was undertaken by deconstructing the individual adaptive tracks into performance-by-level matrices for the different TC levels and masker types. That is, although adaptive psychophysical procedures are designed to efficiently converge on target percent correct levels, the trial-by-trial performance within a threshold estimation track reflects the underlying psychometric function within the range of levels constrained by that track and therefore the shape of the function can be estimated (Leek et al., 1992). To compile the cumulative performance-by-level matrix for each of the six conditions (three TC levels X two masker types), the individual adaptive tracks associated with a particular condition were deconstructed into the trial-by-trial outcome (correct/incorrect) tabulated for each presentation level. These data were combined across all subjects participating in each condition to generate the cumulative performance-by-level matrix for that condition. From this matrix, the cumulative psychometric function was then derived. The results of this exercise are shown in Fig. 4 where each panel plots performance collapsed across subjects as a function of level for one of the three TC conditions. Within each panel, the parameter is masker type (steady masker: unfilled circles; modulated masker: filled squares) and the symbol size indicates the number of observations contributing to that point. For each data set, the best fitting logistic function is also shown (solid line for steady masker; dashed line for modulated masker). Despite the variability and limited dynamic range of the data points collapsed across subjects, the fitted functions exhibit two features. First, for each level of TC, the slope of the performance-intensity function is steeper for the steady masker than for the modulated masker, although this difference is negligible for TC = 50%. Thus, the magnitude of MMR decreases with increasing TC. Second, for both the steady and modulated maskers the slopes of the functions become increasingly shallow as the amount of time compression increases, particularly between TC = 33% and TC = 50%. This might suggest that performance-intensity functions are not independent of TC level. In support of this interpretation, Hosoi et al. (1999) found that performance-intensity functions for nonsense Japanese phrases measured in quiet overlapped for uncompressed (TC = 0%) and TC = 33% speech, but that the function for TC = 50% speech tended to be shallower.

One limitation of the performance-intensity functions shown in Fig. 4 is that the data to which they are fitted are largely clustered within one region of the function. This is expected since the adaptive procedure is designed to efficiently converge on a particular point on the psychometric function. As a result of this clustering, however, the shapes of the functions outside of this region invite more cautious interpretation. To highlight this, the performance-intensity functions fitted to the TC = 0% adaptive track data of Experiment 2 were compared to the performance-intensity functions fitted to the TC = 0% fixed level data of Experiment 1. Although the participant pool was different across the two experiments, the stimuli (uncompressed target sentences and steady/modulated maskers) were the same. From Table 1 it can be seen that the mean threshold in the steady masker for the uncompressed sentences in Experiment 2 was 61.4 dB SPL (speech-to-masker ratio = −3.6 dB). Based on the function fitted to the cumulative data collapsed across subjects for this condition, this

threshold is associated with a percent correct score of 75.1%. The mean threshold in the modulated masker for uncompressed speech, as shown in Table 1, was 52.7 dB SPL (speech-to-masker ratio = −12.3 dB). Based on the function fitted to the cumulative data collapsed across subjects for this condition, this threshold is associated with a percent correct score of 76.4%. In other words, collapsed across subjects, the point on the cumulative psychometric function that was being adaptively tracked was nominally between 75.1% and 76.4% (average = 75.7% correct). Turning now to the psychometric functions derived for sentence-level performance at fixed speech-to-masker ratios in Experiment 1, 75.7% correct performance was associated with a speech intensity of 60.8 dB SPL in the steady masker (speech-to-masker ratio = −4.2 dB); in the modulated masker, it was associated with a speech intensity of 53.7 dB SPL (speech-to-masker ratio = −11.3 dB). These speech intensity values of 60.8 dB SPL and 53.7 dB SPL for the steady and modulated maskers, respectively, derived from the functions fitted to Experiment 1 data are reasonably similar to the threshold intensities observed in Experiment 2 of 61.4 dB SPL and 52.7 dB SPL, respectively. The magnitude of MMR is therefore also reasonably similar (7.1 dB vs. 8.7 dB). This similarity confirms that the adaptive data of Experiment 2 and the fixed-level data of Experiment 1 for TC = 0% reflect the same underlying psychometric functions.

Returning now to the TC = 33% and TC = 50% time-compressed data of Experiment 2, thresholds for the sentences in the steady masker were 4.6 dB and 8.9 dB higher, respectively, than the threshold of 61.4 dB SPL for the TC = 0% sentences. If the subjects of Experiment 1 had listened to the uncompressed sentences in the steady masker at levels 4.6 dB and 8.9 dB higher than 60.8 dB SPL (the level associated with 75.7% correct), they would have been performing with a recognition accuracy of 97.1% correct at 65.4 dB SPL and 99.7% correct at 69.7 dB SPL. For these subjects to obtain the same percent correct scores in the modulated masker, the speech intensities would have had to be 61.7 dB SPL and 69.1 dB SPL, respectively. In turn, the MMR magnitudes associated with these points would have been 3.7 dB and 0.6 dB, respectively. In other words, if the subjects of Experiment 1 had listened to the TC = 0% sentences at speech-to-masker ratios associated with threshold level performance for the time-compressed sentences of Experiment 2, they would have exhibited MMR magnitudes of 3.7 dB and 0.6 dB, respectively, for the two levels of time compression. These MMR magnitudes are 1-to-2 dB smaller than the observed values of 4.3 dB and 2.6 dB measured in Experiment 2.[1] It might be argued that these differences imply that the manipulation of time compression affects masking release in ways other than simply adjusting the measurement point along the performance-intensity function. For example, given the constant modulation rate of 10 Hz used here it is possible that the reduced redundancy of compressed speech is offset by a greater proportion of the speech time waveform being accessible during the masker minima; that is, the informational value of the glimpsed speech could vary with time compression. However, this possibility is undermined by two observations. First, in the study of Grose et al. (2009) using IEEE sentences it was shown that, for the single level of time-compression tested (TC=33%), MMR did not differ between masker modulation rates of 16 Hz and 32 Hz. Second, using

---

[1] A similar outcome is found if this exercise is repeated with thresholds in the modulated masker used as reference.

uncompressed Brazilian-Portuguese HINT sentences we have also shown that recognition thresholds in modulated maskers are stable across modulation rates of 4 – 32 Hz (Advíncula et al., 2013). Further research is required to resolve this issue, perhaps by parametrically varying masker modulation rate in association with degree of time compression.

Acknowledging that questions remain about the dependence of psychometric function shape on TC level, the present findings are nonetheless in line with the notion that the reduced MMR observed in listeners with cochlear hearing loss could be due to the elevated speech-to-masker ratios at threshold associated with the impaired listeners (Bernstein & Grant, 2009; Bernstein & Brungart, 2011; Bernstein, Summers et al., 2012; Smits & Festen, 2013). Whether this is the only factor contributing to the reduced MMR in cochlear loss listeners is a matter of current debate (e.g., Christiansen & Dau, 2012). What is clear, however, is that this mechanism for reduced MMR is not applicable to all cases of reduced benefit of masker modulation. For example, in the study of Grose et al. (2009) that compared MMR in younger and older listeners with relatively normal audiograms and found reduced MMR in the older listeners, the speech-to-masker ratios at threshold did not differ between the two age groups for the steady masker. Thus, the reduced MMR was entirely due to elevated speech-to-masker thresholds in the modulated masker, suggesting that age-related reductions in speech MMR are likely to involve additional mechanisms. In this context, we are currently investigating age-related differences in temporal masking as contributing to the MMR effect.

Summary

The goal of this study was to measure MMR using the Brazilian-Portuguese HINT with a particular focus on determining the dependence of MMR on speech-to-masker ratio at threshold using the novel manipulation of varying time-compression of speech. Experiment 1 measured performance-intensity functions for word- and speech-recognition in a noise having the same LTASS as the target speech material where the masker was either steady or modulated at a rate of 10 Hz. The results demonstrated that the slopes of the functions measured in the modulated masker were shallower than those measured in the steady masker, confirming the assumption that the derived magnitude of MMR depends on the speech-to-masker ratio at threshold. The purpose of Experiment 2 was to adjust the speech-to-masker ratio at threshold using the novel stimulus manipulation of varying the time-compression of the target speech. The results confirmed that speech-to-masker ratio at threshold increased with increasing time-compression of the target speech, but more so in the modulated masker than in the steady masker. As a consequence, the magnitude of MMR decreased with increasing time-compression. This finding supports the notion that the magnitude of MMR depends upon the speech-to-masker ratio at threshold.

Acknowledgments

Acronyms/Abbreviations

dB HL	decibel hearing level
dB SPL	decibel sound pressure level
HINT	Hearing In Noise Test
Hz	Hertz
LTASS	long-term average speech spectrum
MMR	modulation masking release
SD	standard deviation
SSN	speech-shaped noise
TC	time compression
Yrs	years

Bibliography

Advíncula KP, Menezes DC, Pacífico FA, Griz SMS. Effect of modulation rate on masking release for speech. Audiol Com Res. 2013; 18:238–244.

Bernstein JG, Brungart DS. Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. J Acoust Soc Am. 2011; 130:473–488. [PubMed: 21786913]

Bernstein JG, Grant KW. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2009; 125:3358–3372. [PubMed: 19425676]

Bernstein JG, Summers V, Iyer N, Brungart DS. Set-size procedures for controlling variations in speech-reception performance with a fluctuating masker. J Acoust Soc Am. 2012; 132:2676–2689. [PubMed: 23039460]

Bevilacqua MC, Banhara MR, Da Costa EA, Vignoly AB, Alvarenga KF. The Brazilian Portuguese hearing in noise test. Int J Audiol. 2008; 47:364–365. [PubMed: 18569110]

Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. J Acoust Soc Am. 1988; 84:101–114. [PubMed: 3411038]

Buss E, Calandruccio L, Hall JW. Masked sentence recognition assessed at ascending target-to-masker ratios: Modest effects of repeating stimuli. Ear Hear. 2014 ePub ahead of print.

Buss E, Whittle LN, Grose JH, Hall JW 3rd. Masking release for words in amplitude-modulated noise as a function of modulation rate and task. J Acoust Soc Am. 2009; 126:269–280. [PubMed: 19603883]

Calandruccio L, Buss E, Hall JW 3rd. Effects of linguistic experience on the ability to benefit from temporal and spectral masker modulation. J Acoust Soc Am. 2014; 135:1335–1343. [PubMed: 24606272]

Christiansen C, Dau T. Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. J Acoust Soc Am. 2012; 132:1655–1666. [PubMed: 22978894]

Desloge JG, Reed CM, Braida LD, Perez ZD, Delhorne LA. Speech reception by listeners with real and simulated hearing impairment: effects of continuous and interrupted noise. J Acoust Soc Am. 2010; 128:342–359. [PubMed: 20649229]

Dirks DD, Bower DR. Influence of pulsed masking on spondee words. J Acoust Soc Am. 1971; 50:1204–1207. [PubMed: 5117650]

Dubno JR, Horwitz AR, Ahlstrom JB. Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. J Acoust Soc Am. 2003; 113:2084–2094. [PubMed: 12703719]

Festen, JM. Speech-reception threshold in fluctuating background sound and its possible relation to temporal auditory resolution. In: Schouten, MEH., editor. The Psychophysics of Speech Perception. Dordrecht, The Netherlands: Nijhoff; 1987.

Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am. 1990; 88:1725–1736. [PubMed: 2262629]

Francart T, van Wieringen A, Wouters J. Comparison of fluctuating maskers for speech recognition tests. Int J Audiol. 2011; 50:2–13. [PubMed: 21091261]

Fullgrabe C, Berthommier F, Lorenzi C. Masking release for consonant features in temporally fluctuating background noise. Hear Res. 2006; 211:74–84. [PubMed: 16289579]

George EL, Festen JM, Houtgast T. Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2006; 120:2295–2311. [PubMed: 17069325]

Gnansia D, Jourdes V, Lorenzi C. Effect of masker modulation depth on speech masking release. Hear Res. 2008; 239:60–68. [PubMed: 18434049]

Gordon-Salant S, Fitzgibbons PJ. Sources of age-related recognition difficulty for time-compressed speech. J Speech Lang Hear Res. 2001; 44:709–719. [PubMed: 11521766]

Grose JH, Mamo SK, Hall JW 3rd. Age effects in temporal envelope processing: speech unmasking and auditory steady state responses. Ear Hear. 2009; 30:568–575. [PubMed: 19633565]

Gustafsson HA, Arlinger SD. Masking of speech by amplitude-modulated noise. J Acoust Soc Am. 1994; 95:518–529. [PubMed: 8120263]

Hosoi H, Tsuta Y, Nishida T, Murata K, Ohta F, et al. Variable-speech-rate audiometry for hearing aid evaluation. Auris, nasus, larynx. 1999; 26:17–27. [PubMed: 10077252]

Jenstad LM, Souza PE. Temporal envelope changes of compression and speech rate: combined effects on recognition for older adults. J Speech Lang Hear Res. 2007; 50:1123–1138. [PubMed: 17905900]

Jin SH, Nelson PB. Speech perception in gated noise: the effects of temporal resolution. J Acoust Soc Am. 2006; 119:3097–3108. [PubMed: 16708964]

Leek MR, Hanna TE, Marshall L. Estimation of psychometric functions from adaptive tracking procedures. Percept Psychophys. 1992; 51:247–256. [PubMed: 1561050]

Lorenzi C, Husson M, Ardoint M, Debruille X. Speech masking release in listeners with flat hearing loss: effects of masker fluctuation rate on identification scores and phonetic feature reception. Int J Audiol. 2006; 45:487–495. [PubMed: 17005491]

Miller GA, Licklider JCR. The intelligibility of interrupted speech. J Acoust Soc Am. 1950; 22:167–173.

Nakamura K, Gordon-Salant S. Speech perception in quiet and noise using the hearing in noise test and the Japanese hearing in noise test by Japanese listeners. Ear Hear. 2011; 32:121–131. [PubMed: 20628303]

Oxenham AJ, Simonson AM. Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. J Acoust Soc Am. 2009; 125:457–468. [PubMed: 19173431]

Peters RW, Moore BC, Baer T. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J Acoust Soc Am. 1998; 103:577–587. [PubMed: 9440343]

Smits C, Festen JM. The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise. J Acoust Soc Am. 2013; 133:3004–3015. [PubMed: 23654404]

Soli SD, Wong LL. Assessment of speech intelligibility in noise with the Hearing in Noise Test. Int J Audiol. 2008; 47:356–361. [PubMed: 18569108]

Stuart A, Phillips DP. Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. Ear Hear. 1996; 17:478–489. [PubMed: 8979036]

Wilson RH, McArdle R, Roberts H. A comparison of recognition performances in speech-spectrum noise by listeners with normal hearing on PB-50, CID W-22, NU-6, W-1 spondaic words, and monosyllabic digits spoken by the same speaker. J Am Acad Audiol. 2008; 19:496–506. [PubMed: 19253782]
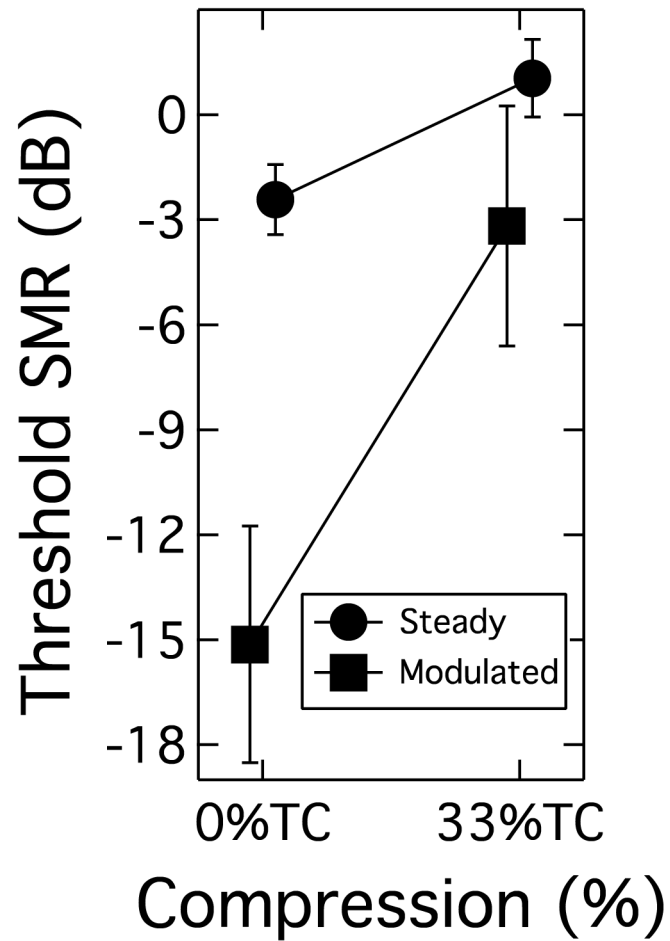
Fig. 1.
Speech-to-masker ratios at threshold for uncompressed (TC = 0%) and time-compressed (TC = 33%) IEEE sentences. The parameter is masker type: steady (circles) and 16-Hz modulated (squares). Data redrawn from Grose et al. (2009).
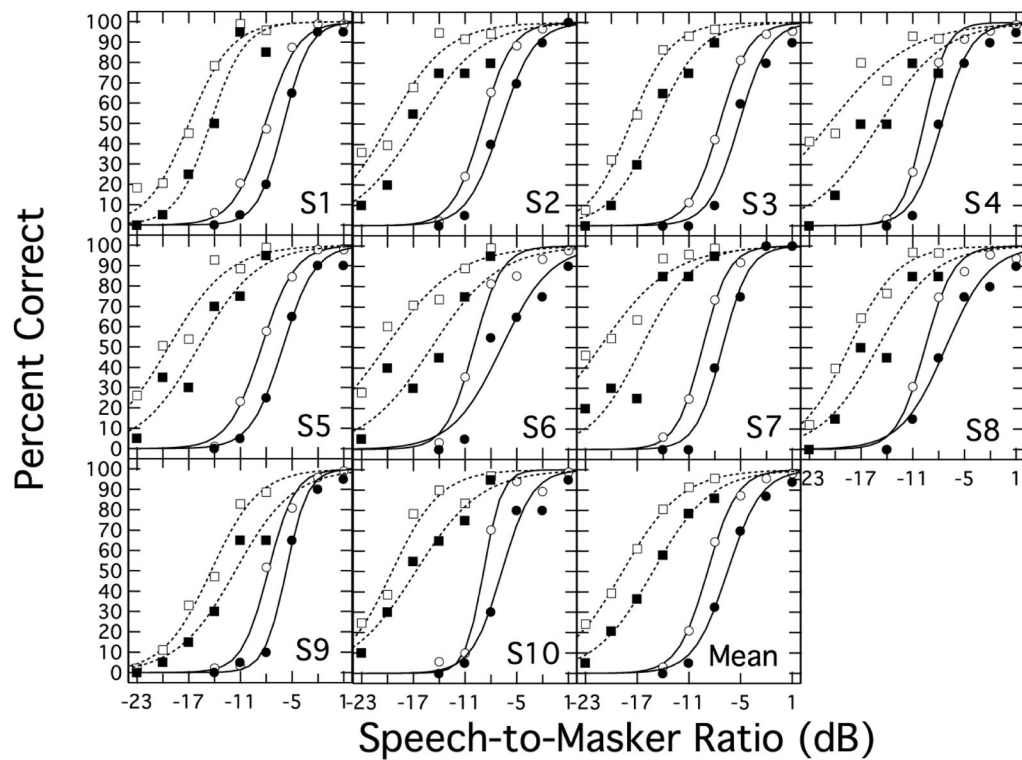
Fig. 2.
Percent correct word recognition (unfilled symbols) and sentences (filled symbols) as a function of speech-to-masker ratio measured in a steady masker (circles) and a modulated masker (squares). Also shown are best fitting logistic functions for the steady masker (solid lines) and modulated masker (dashed lines). Each panel shows data from one subject, with the final panel displaying the mean data.
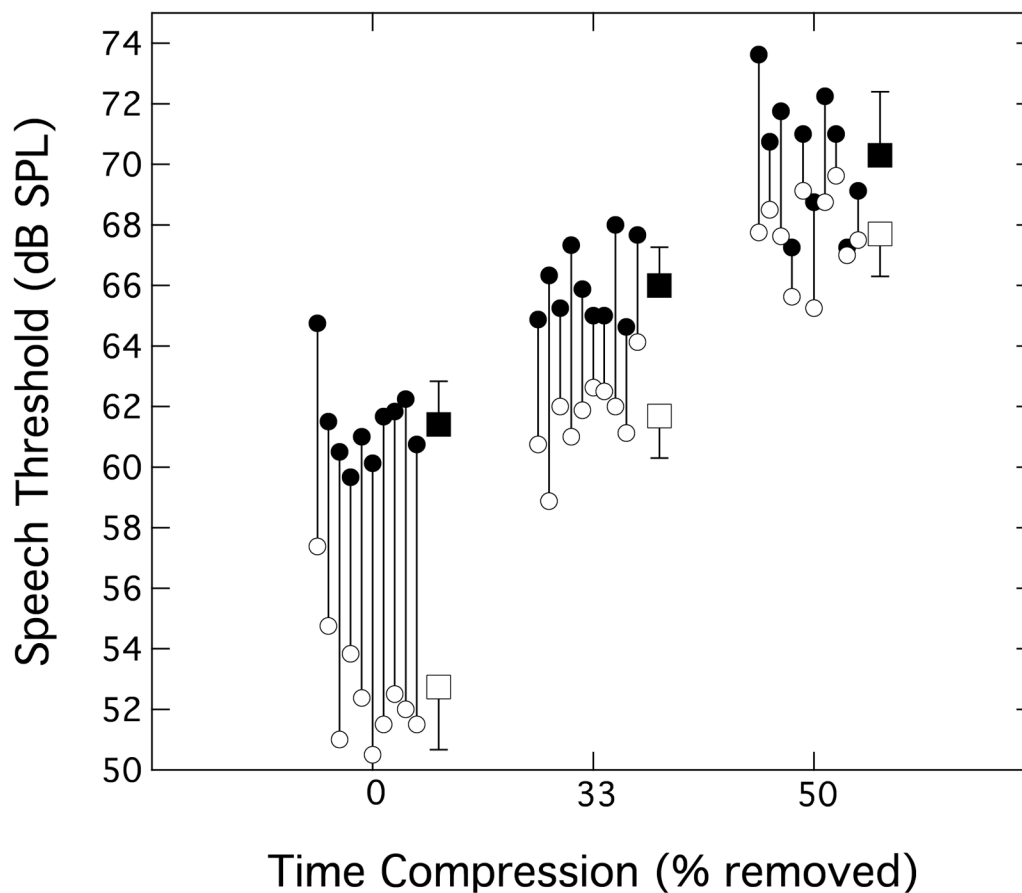
Fig. 3.
Speech recognition thresholds plotted for each level of time compression. Individual thresholds measured in the steady masker (filled circles) and modulated masker (unfilled circles) are connected with a vertical bar. Group mean thresholds are also shown for the steady masker (filled squares) and modulated masker (unfilled squares) with 1-SD error bars.
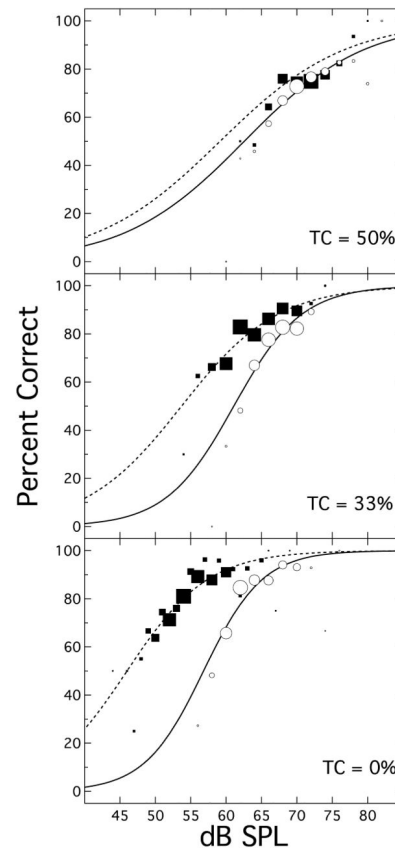
Fig. 4.
Group-level performance-intensity functions derived from individual threshold estimation tracks for TC = 50% (upper panel), TC = 33% (middle panel), and TC = 0% (lower panel). Each panel plots percent correct sentence recognition as a function of sentence level collapsed across subjects for the steady masker (open circles) and the modulated masker (filled squares). Symbol size indicates the number of observations comprising that point. Best-fitting logistic functions are also shown for each data set (solid line: steady masker; dashed line: modulated masker).

Table 1

Group average speech recognition thresholds (dB SPL) in the steady and modulated maskers, and derived MMR (dB), for each of the three levels of time compression. Standard deviations in parentheses.

	Steady Masker	Modulated Masker	Derived MMR
TC = 0%	61.4 (1.4)	52.7 (2.1)	8.7 (1.5)
TC = 33%	66.0 (1.3)	61.7 (1.4)	4.3 (1.7)
TC = 50%	70.3 (2.1)	67.7 (1.4)	2.6 (1.6)