Published in final edited form as:

Ear Hear. 2010 April ; 31(2): 268–276. doi:10.1097/AUD.0b013e3181c9fb9c.

Amplification and consonant modulation spectra

Pamela Souza, PhD and

Dept. of Communication Sciences and Disorders, Northwestern University, Evanston, IL

Frederick Gallun, PhD

National Center for Rehabilitative Auditory Research, Portland VA Medical Center, Portland, OR

Clinicians and researchers who work with hearing aids are most concerned with ensuring that the acoustic frequency spectrum is audible and not distorted. For example, it is assumed that applying high-frequency gain will improve speech recognition for a listener with high-frequency hearing loss, because that will improve audibility of the acoustic frequency spectrum. Clinicians are less likely to concern themselves with the *modulation spectrum*, which represents the dynamic variations in level within each acoustic frequency band. To understand the difference between the acoustic frequency spectrum and the modulation spectrum, consider a simple example: a 1 kHz carrier tone modulated by a 4 Hz sinusoid. The acoustic frequency spectrum (level as a function of acoustic frequency) would have a component centered at 1 kHz, with the width of that component determined by the duration and onset/offset of the tone. The modulation spectrum (modulation depth as a function of modulation frequency) would have a component at 4 Hz with a modulation depth of 100% (and a modulation index of 1).

A useful description of modulation rates for speech is provided by Rosen (1992). Envelope (2-50 Hz) conveys information about consonant manner, voicing, and vowel identity. Periodicity (50-500 Hz) and aperiodicity convey information about voicing, manner, and prosodic cues such as stress and intonation. Fine-structure (600 Hz and higher) conveys information about consonant place of articulation. Thus, we can think of speech in terms of a modulation spectrum, in which a range of modulation rates is present within each carrier frequency band.

How does the modulation spectrum relate to speech recognition? Gallun and Souza (2008) related a simplified representation of the modulation spectrum to consonant errors. If a nonsense syllable had a modulation spectrum that was quite similar to another nonsense syllable, listeners were likely to confuse the two. In that study, acoustic frequency was substantially degraded to force listeners to rely more heavily on modulation spectra, so we do not know whether the same confusions would have resulted had acoustic frequency cues been readily available (i.e., in spectrally-robust speech).

This study examined the effect of clinical amplification strategies on the consonant modulation spectrum, and to determine whether there was a relationship between the modulation spectrum and consonant errors for spectrally-robust, amplified speech presented to listeners with hearing loss. Nonsense syllables were presented to the listener and recorded at the output of the hearing aid. (1) WDRC increased heterogeneity of the modulation spectrum across consonants and (2) for spectrally-robust speech processed with either compression limiting or WDRC amplification, two consonants with similar modulation spectra are more likely to be confused with one another than are two consonants with dissimilar modulation spectra.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

We also do not know whether altering the modulation spectrum for a particular syllable (e.g., making two syllables more or less similar) might have changed error patterns. However, the idea that it is important to preserve the modulation spectrum has been discussed in the research literature for some time. For example, preserving modulation characteristics at the output of a system is the basis of the modulation transfer function used to measure room acoustics (Houtgast & Steeneken, 1985). Some investigators suggested that changing the modulation spectrum could reduce speech clarity or quality (Grant, Elhilali, Shamma, Walden, Surr, Cord & Summers, 2008; Huber & Kollmeier, 2006; Payton, 1999). Completely removing some range of modulation frequencies across all carrier bands had a significant negative effect (Drullman, Festen & Plomp, 1994) but such drastic manipulations would be unlikely to occur in real life.

Regarding more realistic effects, we would expect the multichannel wide-dynamic range compression (WDRC) strategies widely used in digital hearing aids to alter modulation spectra as they change the amplitude variations of the signal in specific frequency bands. Stone and Moore (Stone & Moore, 2007; Stone & Moore, 2008) noted that fast-acting compression with a relatively high ($\sim 3:1$) compression ratio had the following negative effects on the envelope: (1) Compression reduced the similarity of slow modulations (envelope) across channels. They proposed that such an effect could reduce the ability to fuse frequency channels of a target in the presence of a distractor. (2) Compression altered the envelope of the target signal and (3) Compression caused the modulations of the target and distractor to be more similar to one another, potentially making them more difficult to separate.

Because the various measures embedded all envelope changes into a single value, Stone and Moore commented that they were unable to parse out the components of the modulation spectrum that had been most affected. Also, note that all of their data were collected for vocoded signals in which listeners were limited to use of modulations below 100 Hz (with envelope smoothing filter set at 45 Hz); that is, periodicity/aperiodicity cues were not available. That processing removed fine-structures cues, such as fundamental frequency, that would normally have been available to hearing-impaired listeners. Such cues would have aided talker separation (Arehart, Souza & Miller, submitted) and might have offset the impact of the envelope changes. Previous work (Noordhoek & Drullman, 1997) showed that as long as acoustic frequency (fine-structure) cues were available, reducing modulation by reducing peaks via amplitude compression was less detrimental than “filling” valleys via reverberation. In summary, then, it has not been shown whether using clinically-realistic amplitude compression to alter the natural modulation characteristics of spectrally-robust speech, such as would be the case with most individuals with hearing loss, will necessarily have a detrimental effect. Instead, different means of altering modulation spectra might have different effects (positive, negative, or neutral) in terms of their impact on speech recognition. Therefore, the following questions were addressed in the current study: (1) what is the effect of clinical amplification strategies on the consonant modulation spectrum? (2) Is there a relationship between the modulation spectrum and consonant errors for spectrally-robust, amplified speech presented to listeners with hearing loss?

Method

Subjects

Participants were 13 adults (mean age 67 years) with mild-to-moderate sensorineural loss (Table 1). All subjects had symmetric loss with the exception of one subject who had a severe hearing loss in his non-test ear. All were monolingual American English speakers. Three of the subjects had prior hearing aid experience. The ear with the better monosyllabic word recognition was selected as the test ear, although there was little difference for most

subjects. Mean monosyllabic word recognition (NU6 list at 30-40 dB SL) in the test ear was 93% and mean QuickSIN (Killion, Niquette, Gudmundsen, Revit & Banerjee, 2004) score in the test ear was 6.3 dB. Listeners were paid for their participation.

Hearing Aid

Hearing aid fitting and data collection were completed in a double-walled sound booth. Each listener was fit monaurally in the test ear with a 16-band, four-channel behind-the-ear hearing aid. Earmolds were custom-made Lucite skeleton molds with #13 standard tubing. No vent was used to ensure that the signal received by the listener was subject to the intended signal processing. Loudness discomfort levels (LDLs) were measured for warble tones at .5 and 3 kHz using an ascending presentation in which the listener was asked to indicate when the sound became uncomfortably loud.

One memory of the hearing aid was programmed with compression limiting (CL) amplification and one with fast-acting WDRC amplification. When programmed for CL, the hearing aid had a dual limiter. The front-end limiter had a compression threshold of 102 dB SPL. The back-end limiter compression threshold was determined by the maximum power output (MPO). Real-ear saturation response (RESR) was measured for each listener and the maximum power output was adjusted as needed to ensure that the RESR was below the listener's LDLs. The mean MPO across listeners was 121 dB SPL. Both front and back end limiters had a 10:1 compression ratio, 1 ms attack time and 90 msec release time. Below the limiting threshold, linear gain was set according to each listener's prescribed NAL-NL1 (Dillon, 1999) insertion response (REIR) targets, as follows. The target was generated using a Fonix 6500 probe-microphone system. A 65 dB SPL composite noise was presented from a speaker positioned at 45 degrees azimuth and 12 in. from the subject's ear. A probe tube was inserted to 5 mm beyond the medial tip of the earmold. The real-ear unaided response (REUR) and real-ear aided response (REAR) were measured and the calculated real-ear insertion response (REIR) compared to the target. All of the testers were also ASHA-certified audiologists and used their best judgment to adjust the aid for an appropriate match to target.

When programmed for WDRC, the hearing aid still had a dual limiter but processing was dominated by the addition of fast-acting WDRC with a broadband equivalent input compression kneepoint of 45 dB SPL. The attack and release times of the input compressor were 5 and 100 msec, respectively, in each channel. Similar to the procedure for setting the CL gain, a target was generated (in this case, the NAL-NL1 65 dB SPL input target) and the REUR and REAR were measured. The calculated real-ear insertion response (REIR) was compared to the target and frequency-gain response was adjusted as needed to obtain a match to target that the testers judged to be sufficiently close to target (Figure 1). In recommended use of the NAL-NL1 procedure, the prescribed compression ratio would normally be achieved by matching the measured REIG for low- and high-level inputs to targets generated for low- and high-level inputs. NAL-NL1-prescribed CRs for these listeners ranged from 1.2:1 to 1.9:1. A modification was made for this study such that the compression ratio was set to 2:1 in each channel. That compression ratio dictated the gain for low- and high-level inputs; no further adjustments were made to match low- or high-level targets. This was a purposeful constraint to ensure that the WDRC condition was processed differently than the CL condition for all subjects.

Feedback reduction, directional microphones, and manual volume controls were disabled for the duration of testing. The nontest ear was occluded with a foam earplug.

Stimuli

Speech materials were 22 consonant-vowel (CV) syllables (/pi/, /ti/, /ki/, /fi/, /θi/, /si/, /ʃi/, /bi/, /di/, /gi/, /vi/, /ði/, /dʒi/, /mi/, /ni/, /ʃi/, /zi/, /wi/, /ri/, /ji/, /li/) from the UCLA recording of the Nonsense Syllable Test (NST) (Resnick, Dubno, Hoffnung & Levitt, 1975), spoken by a female talker. The CVs were digitally recorded (44.1 kHz sampling rate) onto the hard drive of a computer. For presentation to the listener, CVs were converted to analog, low-pass filtered at a cut-off frequency of 10 kHz, and then output at 65 dB SPL measured at the position of the listener's head in free field. The speaker (JBL LSR25P) was positioned at head level and one meter directly in front of the listener.

In-ear recordings of the CVs were made at the output of the hearing aid using a probe microphone system (ER7C), with the probe tube extended 5 mm beyond the medial tip of each subject's earmold. That configuration has been shown to provide reliable speech recordings (Caldwell, Souza & Tremblay, 2006). Subjects were instructed to minimize body movements while the acoustic recordings were made. The amplified stimuli were low-pass filtered (10 kHz cutoff), digitized (44.1 kHz sampling rate), and recorded onto the hard drive of a computer for analysis. A calibration tone preceded all speech recordings and was available to determine actual dB SPL levels. Three repetitions of each CV were recorded for a total 66 tokens for each condition and subject. This allowed us to compare the three tokens for consistency and also avoid the occasional token recorded with a click or distortion due to unexpected body movement, coughing, etc. One exemplar was selected from the three available for analysis.

Modulation Spectra

Procedures for calculating the modulation spectra were completed using locally developed Matlab code. A modulation spectrum was obtained for each amplified CV (15 subjects $\times$ 22 CVs $\times$ 2 amplification conditions) as follows:

1. The CV was digitally filtered into five octave-wide bands with center frequencies of .25, .5, 1, 2, and 4 kHz. Although the original implantation of the SCI (Gallun & Souza, 2008) included an 8 kHz band, in this case the high-frequency roll-off of the hearing aid response effectively limited the signal bandwidth.
2. Output of each filter was half-wave rectified and low-pass filtered at 50 Hz. The decision to discard modulation frequencies above 50 Hz was based on previous data (e.g., Drullman, et al., 1994) suggesting that modulation rates greater than 50 Hz contribute relatively little to phoneme recognition. However, this simplification does not mean that no information was conveyed by faster modulation rates.
3. Each envelope was zero-padded to a total duration of five seconds then downsampled to 1000 Hz and submitted to Fast Fourier Transform (FFT). This allowed an FFT resolution of 0.2 Hz.
4. The energy value within each FFT bin was divided by the energy level in the 0 Hz bin (i.e., by the DC energy). This normalized the modulation values within each frequency bin such that maximum modulation resulted in a value of one.
5. Energy was summed across FFT bins such that the combined energy values represented the output of six octave-wide rectangular filters with center frequencies of 1, 2, 4, 8, 16, and 32 Hz.

The final modulation spectrum for each amplified CV was a vector of 30 values (six octave-band modulation filters $\times$ five octave-band carrier frequencies). All carrier bands received equal weighting. Note that this implementation of modulation spectrum did not preserve any information about the relative phases of modulation across bands.

We also calculated SCI values for the unprocessed NSTs. That calculation was the same as described above, except that the CV was digitally filtered into six octave-wide bands with center frequencies of .25, .5, 1, 2, 4, and 8 kHz.

For any pair of CVs, the Spectral Correlation Index (SCI) was the Pearson r value between the modulation values for each CV. Because only one vowel was used, the SCI quantified how similar one consonant was to another consonant. An SCI value of 1.0 indicated that two consonants had identical modulation spectra.

Consonant Recognition

Subjects were instructed to select the CV heard from an orthographic display of the 22 choices (e.g., “chee” for /tʃi/) on a touch screen. Each subject completed a set of practice blocks (22 trials per block) using the WDRC condition with 65 dB SPL input. During the practice trials, visual correct answer feedback was provided after each response. The subject repeated practice blocks until stable performance (defined as $\pm 5\%$ for three consecutive blocks) was achieved. A test block consisting of 220 trials (10 repetitions of each CV) was then completed without feedback. The data reported here were collected as part of a larger data set in which the subject was tested under six conditions (3 intensity levels each for the WDRC and CL conditions) and the order of conditions was randomized across subjects. Only limited acoustic recordings were collected at the low- and high-intensity levels, therefore only the 65 dB SPL input will be considered further in the context of the modulation spectrum.

Results

Unprocessed Tokens

Excluding the 1.0 values for each CV correlated with itself, SCI values for the unprocessed stimulus set ranged from .1472 to .9877. The least similar CVs were /li/ and /ti/ (.1472); /li/ and /tʃi/ (.1766); /ri/ and /ti/ (.2027); /bi/ and /ti/ (.2300) and /vi/ and /ti/ (.2709). The most similar CVs were /wi/ and /vi/ (.9877); /fi/ and /θi/ (.9804); /wi/ and /ji/ (.9797); /ni/ and /mi/ (.9785) and /mi/ and /vi/ (.9781). Figure 2 compares the modulation spectra for two of the dissimilar CVs, /tʃi/ and /li/. Note the substantially higher modulation values in the high-frequency carrier bands for /tʃi/.

Effect of Compression Limiting Amplification on Consonant Modulation Spectra

We first examined how CL amplification altered the unprocessed modulation spectra. Data for two of the 22 syllables are shown in Figure 2. Each point represented a comparison of the modulation spectra for two CVs, plotted as the compression limiting SCI versus the unprocessed SCI for that CV pair. For the filled circles, one of the compared CVs is always /ti/; for the open circles, one of the compared CVs is always /di/. Points falling below the diagonal showed less similarity across syllables after CL amplification (i.e., the CL SCI was smaller than the unprocessed SCI). Points falling above the diagonal showed greater similarity across syllables after CL amplification. For /ti/, the majority of the points fell above the diagonal, indicating that CL amplification altered the modulation spectra for /ti/ such that it became more like that of other consonants. The syllable /di/ showed the opposite pattern, in that modulation spectra usually became less similar to other consonants. Because previous data showed that subjects tend to confuse syllables with similar modulation spectra, a simple prediction was that CL amplification would result in more /ti/ errors, and fewer /di/ errors, compared to unprocessed speech.

A data point falling on the diagonal line indicated that the SCI for unprocessed and CL was the same; that is, that CL amplification had no effect on similarity of the modulation spectra

for those two CVs. Any change between unprocessed and CL amplification could therefore be quantified by measuring direction and distance for any data point relative to the diagonal. That distance was termed the “shift” value for that CV pair. A positive shift value (points falling above the diagonal) indicated that the modulation spectra for those two CVs became more similar after amplification. A negative shift value (points falling below the diagonal) indicated that the modulation spectra for those two CVs were more similar without amplification. Small shift values (i.e., close to 0) indicated that the relationship between the two CVs was relatively unchanged after amplification. Note that the shift value did not reflect whether two syllables are similar or dissimilar in terms of modulation characteristics, only how that relationship *changed* as a result of the CL amplification. Those shift values were summarized in Figure 3.

With a few exceptions described here, most consonants demonstrated small shift values. That is, they retained their degree of similarity or dissimilarity to other consonants after CL amplification. The exceptions were the CVs /ti/, /tɿ/ and /dZi/, for which the modulation spectra became more similar to other CVs after CL amplification, and the CVs /bi/ and /di/, for which the modulation spectra became less similar to other CVs after CL amplification. An acoustic analysis indicated that unprocessed envelopes of those syllables were characterized by abrupt level changes over time in one or more frequency bands. When the level changes occurred in the low-frequency bands, modulations in those bands were increased by CL amplification and the CV became less similar relative to others. When those level changes occurred in high-frequency bands, modulations in those bands were reduced by CL amplification and the CV became more similar to others.

A one-way analysis of variance (ANOVA) performed on the shift values indicated a main effect of CV ($F=122.94$, $p<.005$). Post-hoc means comparisons using Fisher's LSD indicated the shift for each of the five CVs noted above was different from that of every other phoneme ($p<.005$). For the remaining CVs, compression limiting amplification had a negligible effect on the SCI. These results suggested that the main impact of compression limiting was to reduce the range of rapid level variations and that the modulation spectrum was only altered in the case of phonemes characterized by such rapid variations.

Effect of Fast-Acting WDRC Amplification on Consonant Modulation Spectra

The relationship between CL and WDRC amplification was compared by plotting SCI for WDRC as a function of the SCI for CL (Figure 4). As in Figure 2, each point represented the SCI for a single subject and CV comparison. Points falling below the diagonal indicated that fast-acting WDRC resulted in less similar modulation spectra, and that was the case for all CVs including the two examples shown here. As before, the changes were quantified by calculating the “shift” distance and direction between each point and the diagonal, with the diagonal representing no difference in SCI value between the two amplification conditions. Results across all subjects and consonants are summarized in Figure 5. WDRC increased distinctiveness of modulation spectra for all consonants, with greatest change for stop consonants. A one-way ANOVA indicated a significant main effect of consonant ($F=18.54$, $p<.005$), and post-hoc analysis (Fisher's LSD) confirmed that /pi/, /ti/, /ki/, /gi/, and /tɿ/ had larger shift values than the other CVs ($p<.005$).

It was also of interest to compare the WDRC-amplified CVs to their unprocessed versions (Figures 6 and 7). As was the case when comparing WDRC amplification to CL-amplified CVs, WDRC increased distinctiveness of modulation spectra for nearly all consonants, with greatest change for stop consonants. And, as for the WDRC vs. CL analysis, a one-way ANOVA confirmed a significant main effect of consonant ($F=18.45$, $p<.005$), with post-hoc analysis (Fisher's LSD) indicating that /pi/, /ti/, /ki/, /gi/, and /tɿ/ had larger shift values than the other CVs ($p<.005$).

Relating SCI to Error Patterns

Although these listeners were fit with appropriate amplification and presented with clearly produced speech in quiet, they still made errors. For CL, scores ranged from 56.4% to 90.9%, with a group mean of 79.0%. For WDRC, scores ranged from 56.4% to 92.3%, with a group mean of 75.0%. Consonant confusions made by these subjects are provided in Appendix A (for compression limiting) and Appendix B (for WDRC). For both types of amplification, listeners scored highest for /b/ and lowest for /θ/ and /Δ/.

In previous work (Gallun & Souza, 2008), the probability of confusing one syllable with another was related to how similar the modulation spectra of the two syllables were. That is, higher SCI values were associated with higher error rates. Figures 8 and 9 show the relationship between error rates and SCI values for the current data set. SCI values are categorized on the abscissa. Values on the ordinate indicate the probability of making a specific error. Error rates are calculated as percentage of total responses, across all study participants. To understand error rates, consider an example: CL consonant pairs considered to have relatively low modulation spectrum similarity (.65-.70). There were 336 consonant pairs which had SCI values between .65 and .70. Three subjects made an error between such consonant pairs: one subject confused /Zi/ and /dZi/ 70% of the time; another subject confused /Zi/ and /gi/ 10% of the time; and a third subject confused /ti/ and /Si/ 10% of the time. Across all subjects then, syllables with SCI values between .65 and .70 were confused only 2.7% of the time $[(0.10 + 0.10 + 0.70)/336 = 0.0027]$. The general pattern for was that listeners were more likely to choose a syllable which had a modulation spectrum that was similar to the presented syllable. This effect was statistically significant for CL ($F=80.22$, $p<.005$) and WDRC ($F=84.46$, $p<.005$) amplification.

In Figures 8 and 9 there are a few values that deviate from the general pattern. These are indicated by arrows and were due to persistent errors of one or two subjects. For example, in Figure 8, the data point for SCIs between .25 and .30 reflects one subject who chose /di/ for /gi/ 70% of the time.

Discussion

The view presented here is that it is meaningful to consider any consonant in terms of how similar its modulation characteristics are to another consonant. The evidence for considering this representation is that two consonants with similar modulation spectra are more likely to be confused with one another than are two consonants with dissimilar modulation spectra, at least within an amplification type. This was shown previously for unamplified but spectrally-limited speech (Gallun & Souza, 2008). The present data extend this result to spectrally-robust speech processed with either compression limiting or WDRC amplification.

The significant relationship between the SCI and consonant errors does not mean that listeners rely exclusively on the modulation spectrum to identify syllables. Unlike Gallun and Souza (2008), listeners in the current study had access to acoustic frequency cues. To the extent that those cues covary with modulation spectra, it may not be necessary to utilize modulation cues *per se* to recognize a specific syllable. It may also be the case that the relative importance of acoustic frequency versus modulation frequency varies across situations or across individuals. As a rule, the ability to use acoustic frequency cues is determined by the listener's individual frequency resolution. The subjects tested here had mild-to-moderate hearing loss, which suggests that they would have only minimally impaired frequency resolution (Van Tasell, 1993). It's likely that a listener with severe loss who has very broad auditory filters would have less access to acoustic frequency cues and might rely more heavily on temporal cues such as those represented in the modulation

spectrum. In turn, such a listener's performance might be more strongly predicted by modulation spectral measures. Indeed, this is essentially the situation mimicked by the processing used by Gallun and Souza (2008).

Effects of Amplification

It was of interest to examine the effect of different amplification schemes on the consonant modulation spectrum. WDRC smoothes the speech envelope (Jenstad & Souza, 2005), and this reduction of slow modulations might be presumed to result in more similar modulation spectra among different consonants. Instead, we found that SCI values were reduced for WDRC relative to compression limiting or unprocessed speech. In terms of modulation spectrum similarity, then, WDRC increased heterogeneity of the modulation spectrum across consonants. To understand this acoustic effect, consider the complexity of fast-acting multichannel WDRC, which continually varies gain across different frequency regions, in combination with the SCI which characterizes modulations up to 50 Hz. Figure 10 illustrates this using the envelope for the .5 kHz carrier band for /bi/, taken from acoustic recordings for a single subject (MM10). The solid line shows the envelope for the WDRC amplification and the dashed line shows the envelope for the CL amplification. Within this carrier band, WDRC introduces an overshoot, magnifies some slight modulations, and probably introduces some spurious modulations as well. Such effects will be occurring in an unpredictable way for every phoneme. Thus, WDRC can make the envelope within one or more carrier bands dissimilar to another phoneme, resulting in a smaller SCI value.

Making consonant envelopes less similar to one another did not reduce the number of consonant confusions for WDRC amplification, nor did it result in significantly better performance for WDRC. One possibility is that WDRC introduces additional distortions, not captured by the SCI, which offset the benefit of dissimilar modulation spectra. Modulating the signal introduces spectral sidebands, and the greater the depth of the introduced modulation the greater the amplitude of the sidebands. Our data indicated that WDRC introduced a greater degree of modulation within a frequency band. The compression sidebands introduced by this process might also produce audible distortion products, and the intensity of those distortion products would be proportional to the difference in SCI. In other words, WDRC has the potential to add distortion that is highly correlated with the SCI, so changes in the modulation spectrum are correlated with other signal modifications.

Considerations of an impaired auditory system

The present data were collected for listeners with hearing loss wearing commercial hearing aids. When testing listeners with hearing loss we must consider audibility of the carrier frequency band which conveys the modulation frequency information. Modulation spectrum cues can be of no value if they are below threshold because the acoustic frequency band which carries them is below threshold. Listeners tested in this study had mild-to-moderate high-frequency loss and were provided with appropriate amplification through 4 kHz. We introduced an additional control in which the 8 kHz carrier band was not considered in the aided SCI calculation, as it would have been inaudible for most or all of the listeners. We propose that this implementation of the SCI is a reasonable approximation to the information transmitted to the listeners. However, it is possible that those individuals with more severe high-frequency thresholds could have received reduced modulation cues from the 4 kHz carrier band as well.

Even if the carrier frequency band is audible, the SCI did not consider directly the issue of listener sensitivity to modulation. The temporal smoothing mechanism (Moore, Glasberg, Plack & Biswas, 1988) within the auditory system plays a role. Loss of the cochlear active mechanism and recruitment might magnify some modulations and suppress others. Because

listeners become relatively less sensitive to modulation above 32-64 Hz (Chi, Gao, Guyton, Ru & Shamma, 1999; Kohlrausch, Fassel & Dau, 2000), some of the more rapid modulations such as those shown in Figure 10 may change the SCI, but be less relevant to perception.

There also could be individual differences in sensitivity to modulation that impact performance. There are established methods for measuring modulation sensitivity using noise bands or pure tones. In speech, modulation is occurring transiently or at variable rates, and may correlate or diverge across multiple frequency channels in a different way from modulations in conventional psychoacoustic stimuli. We expect that the ability to use modulation spectrum as a speech cue will be influenced by a listener's sensitivity to speech or speech-like modulations, but there is as yet no method to predict this in a straightforward way. Nonetheless, the fact that consonant errors were closely related to similarity of modulation spectra within each type of amplification suggests that the acoustic differences captured by the SCI are an appropriate summary of the modulation information utilized by listeners.

Details of the modulation spectrum

Our data indicate that the within-band modulation spectrum is not the same in each frequency band. For these consonant-vowel syllables, the largest differences in modulation patterns were usually seen among the higher-frequency bands, and adjacent bands tended to look more similar to one another than did widely separated bands. This is consistent with work by Crouzet and Ainsworth (2001) who found that envelope was more highly correlated across bands at low modulation frequencies and/or adjacent carrier bands. Given that the high-frequency bands (and especially the higher-frequency modulations within those bands) varied the most from one consonant to another, those bands may contain important cues to distinguish one syllable from another. Apoux and Bacon (2004) found that listeners placed the most weight on the highest-frequency carrier band in difficult listening situations (i.e., in background noise). Silipo, Greenberg and Arai (1999) also believed that modulations above 10 Hz in frequency bands above 1.5 kHz were most important to speech recognition.

The SCI is a simplified metric which has shown to be related to consonant error patterns. However, in its current implementation all modulation frequencies, including those which have very little energy, are given equal weight. We suspect that weighting frequency bands and/or modulation frequencies that provide more useful information or are less redundant with other information could improve its utility. It may also be necessary to incorporate some audibility quantification into the metric, especially for listeners with more hearing loss. Work has already begun in our laboratory to explore these issues.

Acknowledgments

This work was supported by grants R01 DC 006014 and P30 DC 04661 to the first author, and by grant VA RR&D CD2 C4963W to the second author. We thank Evelyn Davies-Venn and Marc Brennan for their assistance with data collection, and Eric Hoover for his help with the modulation spectrum analysis.

References

- Apoux F, Bacon SP. Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise. *J Acoust Soc Am* 2004;116(3):1671–80. [PubMed: 15478433]
- Arehart K, Souza P, Miller C. Effects of age on concurrent vowel perception in acoustic and simulated electro-acoustic hearing. submitted.
- Caldwell M, Souza PE, Tremblay KL. Effect of probe tube insertion depth on spectral measures of speech. *Trends in Amplification* 2006;10(3):145–154. [PubMed: 16959735]

- Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 1999;106(5):2719–32. [PubMed: 10573888]
- Dillon H. NAL-NL1: A new prescriptive fitting procedure for non-linear hearing aids. *Hearing Journal* 1999;52(4):10–16.
- Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 1994;95(2):1053–64. [PubMed: 8132899]
- Gallun F, Souza P. Exploring the role of the modulation spectrum in phoneme recognition. *Ear and Hearing* 2008;29(5):800–813. [PubMed: 18596640]
- Grant KW, et al. An objective measure for selecting microphone modes in OMNI/DIR hearing aid circuits. *Ear Hear* 2008;29(2):199–213. [PubMed: 18595186]
- Houtgast T, Steeneken H. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America* 1985;77(3):1069–1977.
- Huber R, Kollmeier B. PEMO-Q--A new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech, and Language Processing* 2006;14(6):1902–1911.
- Jenstad LM, Souza PE. Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language and Hearing Research* 2005;48(3):651–667.
- Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. 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* 2004;116(4):2395–2405. [PubMed: 15532670]
- Kohlrausch A, Fassel R, Dau T. The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J Acoust Soc Am* 2000;108(2):723–34. [PubMed: 10955639]
- Moore BC, Glasberg BR, Plack CJ, Biswas AK. The shape of the ear's temporal window. *J Acoust Soc Am* 1988;83(3):1102–16. [PubMed: 3356815]
- Noordhoek IM, Drullman R. Effect of reducing temporal intensity modulations on sentence intelligibility. *J Acoust Soc Am* 1997;101(1):498–502. [PubMed: 9000740]
- Payton KL. A method to determine the speech transmission index from speech waveforms. *Journal of the Acoustical Society of America* 1999;106(6):3637–3648. [PubMed: 10615702]
- Resnick JB, Dubno JR, Hoffnung S, Levitt H. Phoneme errors on a nonsense syllable test. *J Acoust Soc Am* 1975;58:114.
- Silipo R, Greenberg S, Arai T. Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations. *Proceedings of the 6th European Conference on Speech Communication and Technology* 1999:2687–2690.
- Stone MA, Moore BC. Quantifying the effects of fast-acting compression on the envelope of speech. *J Acoust Soc Am* 2007;121(3):1654–64. [PubMed: 17407902]
- Stone MA, Moore BC. Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing speech task. *J Acoust Soc Am* 2008;123(2):1063–1076. [PubMed: 18247908]
- Van Tasell DJ. Hearing loss, speech, and hearing aids. *Journal of Speech and Hearing Research* 1993;36(2):228–44. [PubMed: 8487516]

Figure 1. Mean prescribed NAL-NL1 real-ear insertion gain target (dashed line) and mean measured real-ear insertion gain for CL (filled circles) and WDRC (open triangles). Error bars show $\pm$ one standard deviation about the mean.

Figure 2.

Example of SCI for CL amplification as a function of SCI for unprocessed speech, for two of the presented syllables. Each of the 286 filled circles represents a comparison between /ti/ and another syllable (13 subjects $\times$ 22 consonants). Each of the 286 open circles represents a comparison between /di/ and another syllable. Higher SCI values indicate greater similarity of two consonant modulation spectra. Note the different pattern between these two syllables, such that CL amplification resulted in the modulation spectrum for /ti/ becoming more similar to other syllables, while the modulation spectrum for /di/ became less similar to other syllables.

Figure 3.

Mean “shift” values for CL versus unprocessed speech, where the shift is the direction and distance for any data point relative to the diagonal (from Figure 2). Note that the shift value did not reflect whether two syllables are similar or dissimilar in terms of modulation characteristics, only how that relationship *changed* as a result of the CL amplification. Shift=0 indicates amplification did not change the relationship between the modulation spectra for the two syllables. Shift<0 indicates that the modulation spectra became more dissimilar after amplification. Shift>0 indicates that the modulation spectra became more similar after amplification. Asterisks (*) indicate shift values that were significantly different from all other syllables.

Figure 4.

As in Figure 2 but showing SCI for WDRC as a function of SCI for compression limiting.

Figure 5.
As in Figure 2 but showing SCI for WDRC as a function of SCI for unprocessed speech.

Figure 6.
As in Figure 3 but for the relationship between WDRC and CL amplification.

Figure 7.
As in Figure 3 but for the relationship between WDRC amplification and unprocessed speech.

Figure 8.
Relationship between consonant errors and SCI for compression limiting. Error bars represent +/- one standard error about the mean.

Figure 9.

Relationship between consonant errors and SCI for WDRC. SCI values are categorized on the abscissa. Values on the ordinate indicate the probability of making a specific error. Error rates are calculated as percentage of total responses, across all study participants. Error bars represent $\pm$ one standard error about the mean.

Figure 10. Comparison of envelope for WDRC (solid line) and CL (dashed line). Data shown are for /bi/ for the .5 kHz analysis band for a single subject (MM10).

Table 1

Audiometric thresholds (in dB HL)

Frequency (kHz)	.25	.5	1	2	3	4	6	8
Mean	28.1	32.3	41.2	51.5	53.1	55.8	54.6	56.2
SD	8.8	9.0	10.6	12.3	12.3	10.4	10.9	16.0
Minimum	10	20	30	30	30	40	40	20
Maximum	40	55	60	70	70	70	70	80

Table 2

Modulation spectra for /tʃi/ and /li/. Values are amount of modulation at each modulation frequency. SCI was .1766.

Modulation frequency (Hz)	Carrier Frequency (kHz)											
	.25		.5		1		2		4		8	
	/tʃi/	/li/	/tʃi/	/li/	/tʃi/	/li/	/tʃi/	/li/	/tʃi/	/li/	/tʃi/	/li/
1	2.3	2.1	2.2	2.1	2.5	2.5	2.4	2.5	2.4	2.5	2.5	2.4
	8	7	8	8	8	3	3	3	3	8	3	9
2	1.2	1.1	1.0	1.6	1.4	1.8	1.4	1.1	1.5	1.2	2.6	0.9
	3	7	6	0	9	2	0	5	4	3	4	2
4	1.2	1.5	1.2	2.3	0.7	2.3	2.4	0.7	2.8	0.7	5.0	0.5
	5	1	5	5	2	2	1	1	7	6	4	5
8	0.5	0.7	0.5	1.1	0.4	1.0	2.2	0.3	2.9	0.3	5.3	0.2
	0	1	1	1	0	4	7	8	0	6	1	3
16	0.3	0.7	0.3	0.9	0.3	0.8	0.6	0.2	1.0	0.0	2.3	0.0
	3	1	0	2	3	7	6	3	2	9	6	7
32	0.1	0.1	0.1	0.1	0.2	0.1	0.3	0.0	0.3	0.0	0.6	0.0
	6	2	8	5	3	6	3	8	8	3	9	4