considerably lower levels for the categories comfortable and loud than ACALOS. This result appears plausible as the random nature of different levels presented in ACALOS tends to let listeners "save" the categories very loud and too loud for yet to come probably even louder sounds. In this way, the same levels lead to lower loudness categories as for the suggested adjustment and combined matching methods. Although the suggested methods are working in a quite different range of the individual loudness function, they led to almost the same gain prescription based on loudness compensation strategy for narrow band noises. The major advantage is that the suggested methods only use sound levels safely under the uncomfortable level which is a strong requirement for consumer audio devices. The higher gains compared to NAL-NL2 can lead to higher speech intelligibility as shown in Kreikemeier *et al.* (2011). However, higher gains might also lead to reduced acceptance for broadband stimuli. Currently, the inclusion of a loudness summation measure which would lead to reduced gain in such conditions is under investigation.

The three suggested methods showed almost identical results. Method 2 was approx. 60 s faster than method 1 and 3 and was often rated as the easiest method.

Further steps are the refinement of an appropriate fitting rule and integration into real devices (hearing aids, other audio devices) as well as evaluation in terms of loudness, quality, and speech reception.

[This work was supported by the BMBF ("Modellbasierte Hörsysteme").]

REFERENCES

ANSI (American National Standards Institute) (**2004**) "Specification for audiometers" (S3.6–2004), New York

Brand, T., and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling" J. Acoust. Soc. Am. **112**, 1597-1604

Ewert, S.D., and Grimm, G. (**2011**). "Model-based hearing aid gain prescription rule" ISAAR 2011

Hougaard, S., and Ruf, S. (**2011**). "EuroTrak I: A Consumer Survey About Hearing Aids in Germany, France, and the UK" in *The Hearing Review*

Keidser, G., and Dillon, H. (**2006**). "What's New in Prescriptive Fittings Down Under?" in *Hearing Care for Adults*, pp. 133-142

Kreikemeier S., Latzel, M. and Kießling J. (**2011**). "Loudness-based Hearing Aid Fitting with Instantaneous Insitu-Percentile Monitoring" Z. Audiol. **50**, 62-72

Latzel, M., Margolf-Hackl, S., Denkert, J., and Kießling, J. (**2004**). "Präskriptive Hörgeräteanpassung auf Basis von NAL-NL1 im Vergleich mit einem lautheitsbasierten Verfahren" DGA.

Lecluyse, W., and Meddis, R. (**2009**). "A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners" J. Acoust. Soc. Am. **126**, 2570-2579

Clinical measures of static and dynamic spectral-pattern discrimination in relationship to speech perception

STANLEY SHEFT, ROBERT RISLEY, AND VALERIY SHAFIRO

Department of Communication Disorders & Sciences, Rush University Medical Center, 600 South Paulina Street, suite 1012, Chicago, IL 60612 USA

Two experiments evaluated discrimination ability for both static and dynamic spectral patterns. The static conditions measured the ability to detect a change in the phase of a low-rate sinusoidal spectral ripple of wideband noise. The dynamic condition determined the signal-to-noise ratio (SNR) needed to discriminate 1-kHz pure tones frequency modulated by different 5-Hz lowpass noise samples drawn from the same underlying noise distribution so that discrimination was based on the temporal pattern of fluctuation. Both procedures used a modified descending method of limits with test stimuli recorded on a CD for clinic use. Results from the first experiment showed a significant relationship of both metrics to masked speech intelligibility. Using only the static procedure, the second experiment evaluated the role of fine-structure information in the perception of masked speech through vocoding of psychoacoustic and speech stimuli. In this case, results showed significant relationship only when the psychoacoustic and speech stimuli were either both vocoded or both unprocessed, consistent with involvement of stimulus fine structure in speech perception at low SNRs. Overall, results from both experiments support clinical utility of the procedures in the context of speech processing ability.

INTRODUCTION

Due to manner of production, speech can be represented by distinctive spectral patterns that vary over time. From this basis, past work has shown relationship between the ability to discriminate spectral patterns and measures of speech intelligibility in clinical subject groups. This work has evaluated auditory processing of both static and dynamic spectral patterns. A common approach in procedures that used static patterns was to assess the ability to either detect or discriminate periodic spectral rippling of wideband stimuli (*e.g.*, Litvak *et al.*, 2007; Won *et al.*, 2011). Past evaluation of dynamic spectral patterns has measured discrimination of either the rapid spectral variations of Schroeder-phase harmonic complexes or low-rate stochastic frequency modulation (FM) of pure-tone carriers (Drennan *et al.*, 2008; Sheft *et al.*, 2011). The current study represents initial efforts at developing clinically feasible measures of both static and dynamic spectral-pattern discrimination. Past work evaluating discrimination of static spectral patterns in clinical subject groups measured performance in terms of the threshold density of spectral rippling. So that density was constant at a value consistent with involvement

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

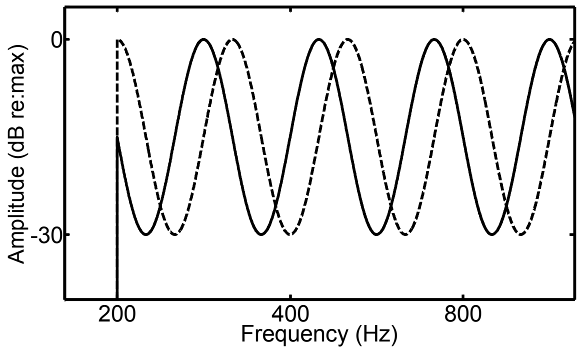

Fig. 1: Schematic illustration of the contrasting amplitude spectra of a discrimination trial with difference due to change in ripple starting phase.

in either speech or complex pitch perception, the current task modified past procedure to measure the just detectable change in phase of a fixed-frequency sinusoidal spectral profile. For assessment of processing of dynamic spectral patterns, current work continued our past study of discrimination of low-rate stochastic FM. In the context of speech processing ability, study of low-rate FM was chosen in that past work indicates that it can enhance speech coherence, aid segmentation at the word and syllable boundaries of the speech stream, and benefit speech intelligibility in the presence of competing interference.

EXPERIMENT 1

Method

Discrimination of static spectral patterns was assessed using wideband stimuli (0.2–8.0 kHz) whose amplitude spectra were sinusoidally rippled in terms of the logarithms of both frequency and amplitude. Ripple density was 1.5 cycles per octave with a peak-to-trough difference of 30 dB. In the cued two-interval forced-choice (2IFC) procedure, the phase of the sinusoidal spectral ripple of the standard stimulus was randomized each discrimination trial with the task to detect a change in ripple starting phase (Fig. 1). Using random component phases, stimuli were generated with ¼-Hz resolution of an IFFT. The 500-ms rippled stimuli were shaped with a 50-ms rise/fall time, passed through a speech-shape filter emphasizing the mid frequencies, and presented to listeners at 80 dB SPL.

Dynamic spectral-pattern discrimination was evaluated in terms of the ability to discriminate 1-kHz pure tones frequency modulated by different samples of 5-Hz lowpass noise. A consequence of the modulation is that the instantaneous frequency of the stimulus follows the amplitude pattern of the noise modulator. The bandwidth

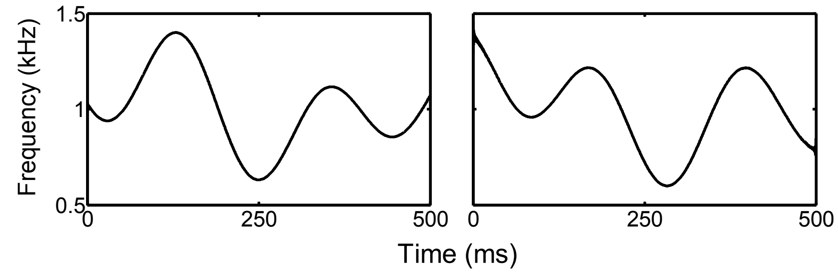

Fig. 2: Schematic illustration of stochastic FM showing the contrasting instantaneous frequency functions of two stimuli of a discrimination trial.

of the noise modulator determines the average rate of FM. For 5-Hz lowpass noise modulators, average rate is roughly 4 Hz. Due to the stochastic modulation, the long-term stimulus spectrum is continuous with a bandwidth that reflects modulator peak amplitude. This peak amplitude also determines ΔF, the maximum frequency excursion of the FM stimulus. In the present work, ΔF was fixed at 400 Hz for all stimuli. With ΔF fixed and a common sampling distribution of noise modulators, discrimination can rely on only the temporal pattern of frequency deviation (Fig. 2). The 500-ms modulated stimuli were temporally centered in 1000-ms maskers with thresholds measured in terms of the signal-to-noise ratio (SNR) needed to just discriminate pattern of frequency fluctuation. To have modulation characteristics similar to speech, maskers were speech-shaped wideband noise which was processed to include slow random variations in local fine-structure periodicities and loudness. The fine-structure periodicities were introduced through an iterative delay-add process in which delay time was dynamically varied between 0.75-3.0 ms by the time structure of 15-Hz lowpass noise. The loudness variations were achieved by comodulating the maskers with 2.5-Hz lowpass noise. Both signals and maskers were shaped with a 50-ms rise/fall time. In the task, masker level was fixed at 80 dB SPL with the level of the FM tones varied to estimate the threshold SNR.

The clinical test procedure was based on the approach of Kidd *et al.* (2007). Using a modified descending method of limits, thresholds were derived from performance on two 36-trial blocks, each cycling through six levels of the dependent variable (d.v.). D.V.'s were the logarithm of the phase delta in radians and SNR in dB. With eight levels of the d.v., the first block used the six highest levels while the second presented levels three through eight. In the cued 2IFC procedure, the cue was the second stimulus presentation with listeners verbally indicating their selection of the signal interval. For both tasks, the trial blocks were recorded on a CD for subsequent clinic use with testing of each condition requiring less than ten minutes.

With only 72 trials spread across eight levels of the d.v., accurate threshold estimation is an important concern. The 2IFC psychometric function ranges between 50 and 100% correct. Assuming a stable underlying function with function slope symmetric about threshold at 75% correct, threshold can be arithmetically derived if levels of the d.v. are evenly spaced and at least minimally bracket the threshold point. Specifically, threshold is:

$$high + step/2 - step*(2*p - num),$$

where *high* is highest level of the d.v., *step* is the decrement between successive levels of the d.v., *num* is the number of levels used, and *p* is the sum of the correct-response probabilities across all levels. A final assumption is that no response probability can be below chance performance when used in threshold derivation. With relatively few trials, response variance will be high. The threshold algorithm was tested in simulations with variance added to both the internal representation of d.v. level and the response probability. For these simulations, rather large values for variance were added as a uniformly distributed 6-dB d.v. level range and a 0.2 response-probability range. Despite added variance, the mean of the estimated thresholds was only 0.4 dB less than the underlying threshold value with a standard deviation of 2.3 dB, demonstrating feasibility of the algorithm for threshold estimation in clinical practice, and requiring only arithmetic calculations.

Speech perception was evaluated in terms of the intelligibility of sentences from the Quick Speech-in-Noise Test (QuickSIN) in the presence of four-talker babble. Presented at 70 dB HL, four lists of six sentences were used. Across each list, the SNR decreased in 5-dB steps from 25 to 0 dB. Based on the number of key words correctly repeated, results were converted to the metric SNR Loss, the estimated SNR in dB needed for 50% correct relative the performance of normal-hearing listeners.

Data were collected from 15 young adults (age range 21-29 yrs) who had normal audiometric thresholds, and 31 older participants (age range 57-87 yrs) separated as 16 normal-hearing (NH) and 15 hearing-impaired (HI) listeners. Older HI participants had audiometric pure-tone averages (PTAs) of greater than 20 dB HL and exhibited a mild-to-moderate symmetric sloping sensorineural loss. Despite the label of normal hearing, all but two of the older NH listeners exhibited at least a mild hearing loss at 4 and 8 kHz. All listeners participated in the FM and QuickSIN conditions. Developed after testing began, only 25 listeners (nine young adults, seven older NH and nine older HI) participated in the Ripple-Phase condition. All stimuli were presented diotically using Etymotic ER-3A insert earphones.

Results

Results from the three conditions are shown in separate panels of Fig. 3. For all tasks, best performance was obtained from the younger listeners. Among the older participants, the distributions of thresholds of the NH and HI listeners almost fully overlapped in the FM and QuickSIN conditions, with a trend for a performance decrement associated with hearing loss only in the Ripple-Phase results. Results from each condition were submitted to separate one-way analyses of variance

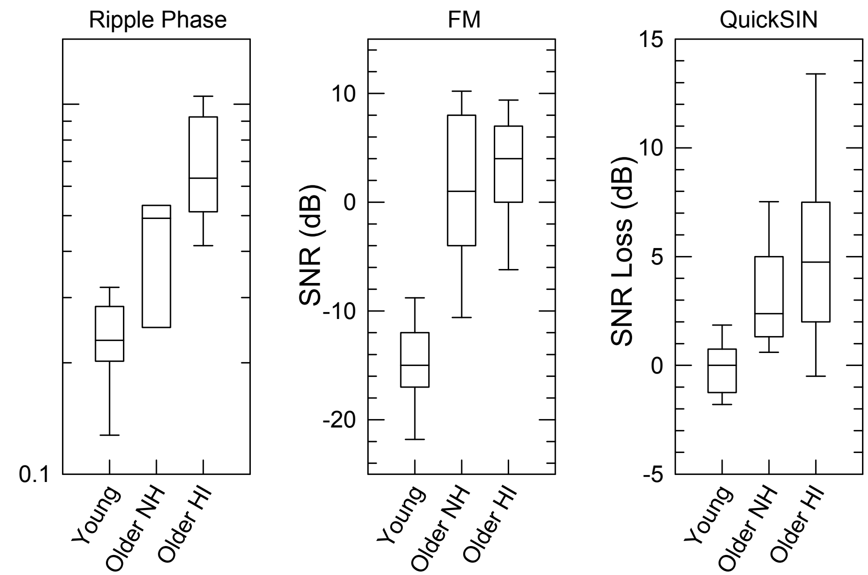

Fig. 3: Box plots showing results from each condition. The line through each box is the median threshold; the upper and lower box edges indicate the 25th and 75th percentiles with error bars showing the 10th and 90th percentiles. With only seven older NH participants in the Ripple-Phase condition, error bars were not calculated.

(ANOVA) on factor listener group. Main effects of group were significant [Ripple Phase: $F(2,22)=18.7$, $p<.001$; FM: $F(2,43)=40.8$, $p<.001$; QuickSIN: $F(2,43)=18.7$, $p<.001$]. For each ANOVA, *post-hoc* pairwise comparisons with Bonferroni corrections showed significant differences ($p<.015$) between performance of the younger listeners and both groups of older participants with the effect of hearing loss among the older participants not significant in any condition. In terms of the relationships between the psychoacoustic measures of spectral-pattern discrimination and masked speech perception, the correlations between the QuickSIN thresholds and the Ripple-Phase and FM results were 0.58 ($p=.001$) and 0.61 ($p<.001$), respectively.

EXPERIMENT 2

Recent work indicates that temporal fine-structure information can play a role in speech perception, especially at low SNRs (*e.g.*, Gnansia *et al.*, 2009; Hopkins and Moore, 2011). Despite emphasis in past work on spectral characteristic, logarithmically rippled noise exhibits fine-structure periodicities, most notable in the outputs of individual auditory channels. The second experiment utilized vocoding of psychoacoustic and speech stimuli to manipulate temporal fine-structure information. Unlike past work which vocoded spectrally rippled stimuli to approximate changes in frequency resolution associated with hearing impairment (*e.g.*, Litvak *et al.*, 2007; Won *et al.*, 2011), the present condition noise vocoded rippled stimuli with a high number of filter channels so processing would alter primarily fine structure rather than amplitude spectrum.

Method

Ripple-phase discrimination was measured as described in Experiment 1 with change of ripple density to 3 cycles per octave. This density of three spectral peaks per octave coincides with an augmented chord in music (*e.g.*, C E G#) with rippled stimuli evoking a strong musical chord percept as if played on an organ. A second condition processed the rippled stimuli through a 35-channel vocoder with center frequencies between 0.2–8.0 kHz, replacing channel fine structure with filtered noise. Channel envelopes were lowpass filtered at 160 Hz before carrier modulation. Unlike Experiment 1, thresholds were estimated with a single 42-trial block with 7 levels of the d.v.

Two conditions evaluated the intelligibility of Speech-in-Noise (SPIN) sentences, administered in a closed-set format with 50 response options. In the first, the sentences were masked at an SNR of -12.5 dB by 8-kHz lowpass noise which was comodulated by 10-Hz lowpass noise. In the second condition, the SPIN sentences were processed through a four-channel noise vocoder, again with lowpass envelope filtering at 160 Hz. For this condition, the SNR in the presence of the comodulated masker was 12.5 dB. For both the psychoacoustic and speech tests, data were collected from 22 normal-hearing young adults.

Results

Mean thresholds from the Ripple-Phase condition, expressed in terms of semitones, are shown in Fig. 4 (left panel). For the normal or unprocessed stimuli, the average just-detectable change in the fundamental of an augmented chord was 0.21 semitones. We are unaware of other data evaluating the ability to discriminate change of chord fundamental. Vocoding the rippled stimuli elevated thresholds to a mean value of 0.27 semitones, with the effect of vocoding significant in a paired-samples T test [$t(21) = -3.07$, $p=.006$]. SPIN results were broken down in terms of word predictability (Fig. 4, right panel), with effect of predictability significant for only the normal [$t(21) = -6.06$, $p<.001$] rather than vocoded speech. Correlations between overall SPIN performance (average of low & high predictability) and ripple-phase thresholds indicated significant relationship only when the psychoacoustic and speech stimuli were either both vocoded or both unprocessed (Table 1). Despite the effect of predictability for unprocessed SPIN words (see Fig. 4), predictability did not significantly affect the relationship to psychoacoustic thresholds when correcting for multiple comparisons with a bootstrap resampling procedure.

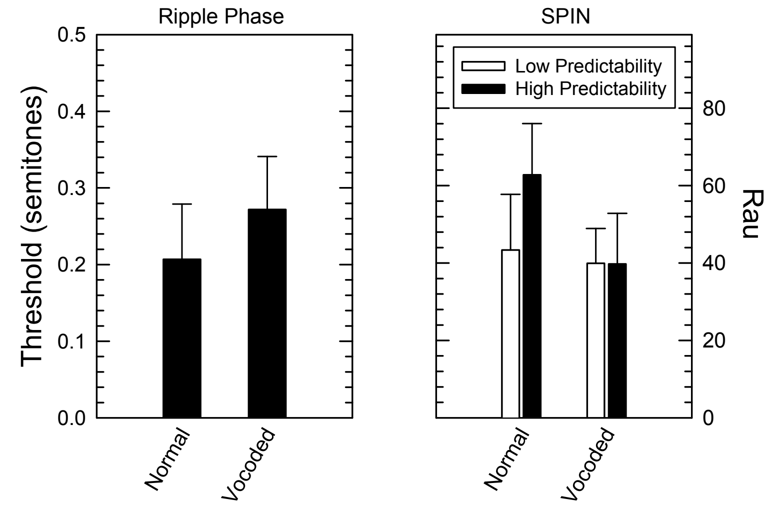

Fig. 4: Mean results from 22 young NH listeners in Experiment 2 with error bars representing 1 standard deviation.

	Normal SPIN	Vocoded SPIN
Normal Ripple	**-.631** (.001)	-.145 (.259)
Vocoded Ripple	.132 (.279)	**-.525** (.006)

Table 1: Pairwise Pearson correlations among experimental measures with *p* values of one-tailed significance tests in parentheses.

DISCUSSION

Results from the first experiment showed an effect of age on the discrimination of both static and dynamic spectral patterns. The age effect in the clinical FM condition is consistent with past laboratory measures of FM discrimination (Sheft *et al.*, 2011) and other metrics of fine-structure processing (*e.g.*, Hopkins and Moore, 2011). The fixed ΔF of 400 Hz in the FM condition extends well beyond the bandwidth of the auditory filter tuned to the 1-kHz carrier frequency, allowing for involvement of cross-channel processing with temporally distributed place coding of aspects of the dynamic spectral patterns. Ongoing work in which this effect is simulated through vocoding along with results obtained from cochlear-implant users indicate that reliance on place coding elevates thresholds for discriminating stochastic FM, suggesting that the effect of age in the present data may in part relate to a deficit in auditory temporal coding.

Results from both experiments showed significant correlations between the psychoacoustic measures and speech processing ability. In the second experiment, significant relationships were obtained only when the rippled and speech stimuli were either both vocoded or both unprocessed. If auditory processing of the intact or normal rippled stimuli is exclusively spectral, it would be expected to correlate equally with both SPIN conditions. Additionally, if processing of speech in the presence of a modulated masker at a low SNR showed no involvement of stimulus fine structure, removal of fine-structure information in the vocoded ripple condition would be expected to have no effect on relationship to performance in the normal SPIN condition. Neither prediction was met, indicating relationship of fine-structure processing to speech perception.

Intended as clinical measures, the psychoacoustic results represent neither trained nor asymptotic performance. The work is part of our ongoing efforts to develop clinically feasible measures of psychoacoustic abilities that may enhance assessment and diagnosis of patient hearing difficulties. In this context, the procedures may also be of use in determining appropriate settings for prosthetic devices and monitoring rehabilitation progress.

ACKNOWLEDGEMENTS

Work was supported by NIDCD (R15 DC011916 and R03 DC008676).

REFERENCES

Drennan, W.R., Longnion, J.K., Ruffin, C., and Rubinstein, J.T. (**2008**). “Discrimination of Schroeder-phase harmonic complexes by normal-hearing and cochlear-implant listeners” J. Assoc. Res. Otolaryngol., **9**, 138-149.

Gnansia, D., Péan, V., Meyer, B., and Lorenzi, C. (**2009**). “Effects of spectral smearing and temporal fine structure degradation on speech masking release” J. Acoust. Soc. Am., **125**, 4023-4033.

Hopkins, K., and Moore, B.C.J. (**2011**). “The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise” J. Acoust. Soc. Am., **130**, 334-349.

Kidd, G.R., Watson, C.S., and Gygi, B. (**2007**). “Individual differences in auditory abilities” J. Acoust. Soc. Am., **122**, 418-435.

Litvak, L.M., Spahr, A.J., Saoji, A.A., Fridman, G.Y. (**2007**). “Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners” J. Acoust. Soc. Am., **122**, 982-991.

Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (**2011**). “Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception” Ear Hear. (submitted).

Won, J.H., Clinard, C.G., Kwon, S., Dasika, V.K., Nie, K., Drennan, W.R., Tremblay, K.L., Rubinstein, J.T. (**2011**). “Relationship between behavioral and physiological spectral-ripple discrimination” J. Assoc. Res. Otolaryngol., **12**, 375-393.