Relationship of frequency-pattern training to speech perception

STANLEY SHEFT*, VALERIY SHAFIRO, AND KRISTEN CORTESE

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

Though discrimination of frequency patterns can relate to speech perception and the discrimination ability generally improves with training, the relationship between the training and speech perception is not known. Training regimens typically utilize simple repetition of the discrimination or identification trials. In the current work, the training protocol was based on interactive pattern reconstruction, increasing memory demands to accentuate learning. With either four- or five-tone patterns, the task was to assemble the constituent tones in the correct order. Tones were randomly selected from logarithmically scaled distributions (frequency: 400-1750 Hz, duration: 75-600 ms). In training but not test sessions, listeners were allowed multiple repetitions of the intact pattern to self-correct their interim response. To assess relationship to speech abilities, the same task was used in pre- and post-training measures with the tonal pattern replaced by samples of sinewave speech (SWS). Despite a high level of stimulus uncertainty, results showed a significant stimulus-specific benefit of training. Small but significant improvement in SWS intelligibility between pre- and posttest sessions was also obtained with greater relationship between results from intelligibility and pattern-reconstruction conditions post training.

INTRODUCTION

The frequency transitions and modulations of speech can serve multiple functions, not only conveying phonetic information but also enhancing signal coherence, segregation, and segmentation. Consistent with this involvement, Sheft *et al.* (2012a; 2012b) found significant age-mediated relationships in adults between speech perception in noise and the ability to discriminate random pitch patterns generated by frequency modulation (FM) of a tonal carrier with narrowband lowpass noise. Both behavioral and physiological measures show effect of auditory training on the processing of frequency patterns (Watson *et al.*, 1976; Tervaniemi *et al.*, 2001; Foxton *et al.*, 2004; Gaab *et al.*, 2006). Despite relationship between psychoacoustic and speech abilities, the effect of frequency-pattern training on speech perception is not known. The goal of the present study was to evaluate this effect as an initial step in determining the feasibility of a modified approach to frequency-pattern training in relationship to speech perception as a component of auditory rehabilitation.

*Corresponding author: stanley_sheft@rush.edu

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Training regimens in previous work have typically utilized simple repetition of the discrimination or identification trials used to assess pre- and post-training performance. The current experiment was designed to study the efficacy of a modified frequency-pattern training protocol. To make the task both more demanding and engaging for the listener, the basis of the protocol was interactive pattern reconstruction. Along with interactive participant involvement in the task through multiple self-corrected responses, the increased memory demands of the task were intended to accentuate learning.

Speech perception was primarily assessed using sinewave speech (SWS). SWS was chosen as a challenging and distorted stimulus set that would exhibit some level of direct relevance to the frequency patterns of the training protocol. In conjunction with measurement of the ability to reconstruct tonal frequency patterns, pattern reconstruction was also evaluated with segmented SWS tokens. The intent of theses SWS conditions was to use stimuli for which the intact pattern represented a form of speech in a task not requiring speech intelligibility.

METHOD

Participants were 26 normal-hearing English speakers (age: 20-28 yrs.), with 13 subjects in the experimental group which received training and 13 in the control group which did not. The experimental protocol consisted of a pretest, three training sessions, and a posttest. Six to seven days separated the pre- and posttest sessions. Speech intelligibility and psychoacoustic frequency-pattern reconstruction ability were assessed during the pre- and posttest sessions with training sessions only on the psychoacoustic task. All testing was with diotic stimulus presentation over headphones.

For the experimental group only, speech intelligibility was measured for AzBio sentences in comodulated noise (2.5-Hz lowpass noise modulator) at a −13 dB signal-to-noise ratio with speech at 70 dB SPL. The primary speech measures for both groups were intelligibility of SWS. SWS was generated with three component sinewaves continuously modulated in both frequency and amplitude with values estimated by linear prediction analysis. To set baseline intelligibility to allow for possible performance improvement and to strengthen relevance to the tonal frequency patterns of the training protocol (see below), component FM was lowpass filtered at 12 Hz with a 4th-order Butterworth filter, and amplitude variations were compressed by a factor of 0.7. SWS intelligibility was measured for vowel-consonant-vowel (VCV) tokens and consonant-nucleus-consonant (CNC) words. CNCs were scored both in terms of number of words and phonemes correct. Before scored testing, listeners were familiarized with SWS, first with sentences, and then with VCV and CNC tokens. Open-set SWS testing was at 75 dB SPL.

The initial condition of the pattern-reconstruction task used pure-tone stimuli. Working with four-tone frequency patterns, the listening task was to assemble the constituent tones in the correct order (Fig. 1). Subjects heard the target sequence only once, but could listen to both constituent tones and their interim reconstruction

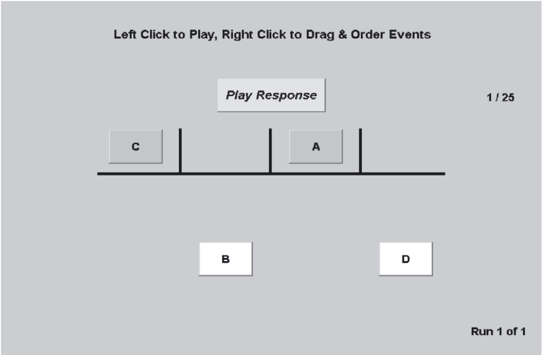

Fig. 1: Illustration of the pattern-reconstruction task with four elements. The pattern elements represented by the four boxes labeled A, B, C, and D are rearranged in the upper place holders to reconstruct the original pattern.

of the sequence as often as wanted. Correct-answer feedback for each sequence tone was provided after every trial.

Constituent tones of the patterns were randomly selected from logarithmically scaled distributions with frequency ranging from 400 to 1750 Hz and duration from 75 to 600 ms. These ranges were chosen to have a rough correspondence to the dominant regions of speech-element characteristics. To maintain discriminability of sequence components, any two component frequencies were separated by at least a factor of 1.2 with any two durations differing by at least a factor of 1.4. Constituent tones were shaped with a 50-ms rise/fall time or half the tone duration when less than 100 ms. The same task was also used in the pre- and posttest sessions with the tonal stimuli replaced by SWS VCVs and CNCs randomly segmented with exponential sampling (minimum duration: 75 ms) to create pattern elements. For each pattern-reconstruction condition (i.e., tone, VCV, CNC), data were collected from a single 25-trial block preceded by five practice trials. Stimulus level was 75 dB SPL.

Three training sessions between the pre- and posttest were completed by subjects on laptop computers at home. Each session lasted about 45 minutes. To document performance change, training sessions began with a repetition of the frequency-pattern task used in the pre- and posttest sessions. The actual training protocol required subjects to reconstruct five- instead of four-tone frequency patterns. Unlike the pre- and posttest sessions, during training, subjects were allowed multiple repetitions of the intact target sequence to self-correct their interim response. Subjects were encouraged to continue each training trial until they were confident that they had replicated the target sequence. Subjects completed two 25-trial training blocks in each session.

RESULTS

Figure 2 shows performance of the trained subject group on the five repetitions of the pattern-reconstruction task with four tonal elements. In this and all subsequent analysis, a Freeman-Tukey arcsine transform was applied to data before statistical analysis. A repeated-measures Analysis of Variance (ANOVA) showed a significant effect of repetition [$F(4,48) = 9.40$, $p < .001$, $\eta_p^2 = .44$], with Holm-Sidak *post-hoc* analysis indicating the contrasts of posttest performance with either the pretest or the first training session as the only significant differences. Across repetitions and subjects, error rate dropped by a factor of almost four with increasing stimulus duration within the 75-600 ms range. In contrast, no effect of stimulus frequency was observed within the range of frequencies used.

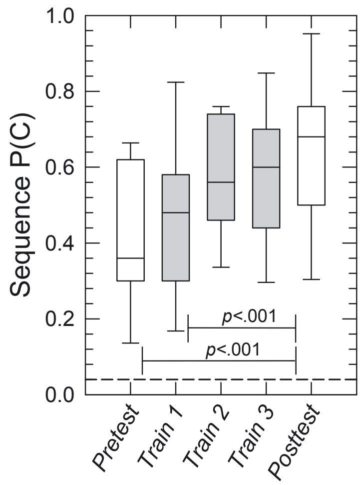

Fig. 2: For the trained subject group, box plots showing results in terms of proportion correct on the four-tone pattern-reconstruction task from pretest, the three training sessions, and posttest. The dashed line at the bottom indicates chance performance. Significant change re posttest is indicated by the horizontal line ending at the comparison condition.

Results from both subject groups comparing pre- to posttest performance with the three stimulus types in the pattern-reconstruction task are shown in Fig. 3. Across both groups, there was a significant effect of stimulus type in the pretest [$F(2,48) = 10.67$, $p < .001$, $\eta_p^2 = .31$] and posttest [$F(2,48) = 19.30$, $p < .001$, $\eta_p^2 = .45$], with *post-hoc* comparisons indicating significantly better performance ($p \leq .001$) in the VCV condition than with either CNC or tonal stimuli. Significant improvements from the pre- to posttest were found with each stimulus type for the trained group, and only in the VCV condition for the control group. For the trained subjects, the posttest change in performance obtained with tonal stimuli was significantly larger than that for VCV ($t = 2.05$, $p = .05$) or CNC ($t = 2.74$, $p = .01$) stimuli. Training effect was evaluated with a between-group Analysis of Covariance (ANCOVA) on posttest scores with the pretest as a covariate. In separate analyses for each stimulus type, a significant effect of training was obtained only with the tonal stimuli [$F(1,23) = 9.81$, $p = .005$, $\eta_p^2 = .30$].

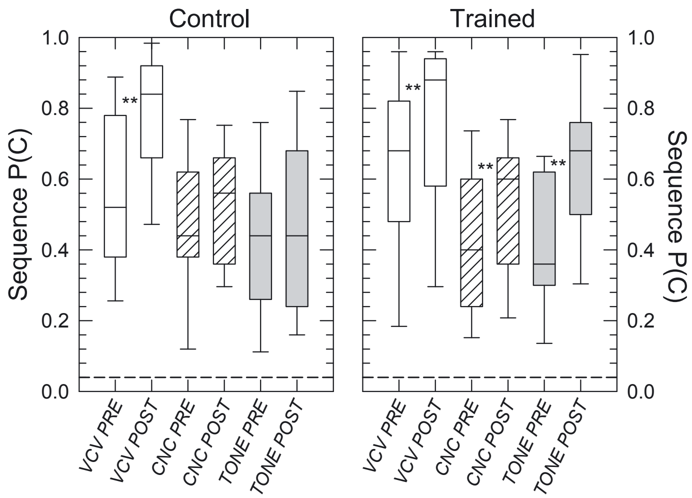

Fig. 3: For the control (left panel) and trained (right panel) subject groups, box plots showing pre- and posttest performance on the three pattern-reconstruction conditions using VCV, CNC, and tonal stimuli. Within group, significant change in performance from pre- to posttest at the $p < .01$ level is indicated by the double asterisk.

For the trained subjects, small but significant improvements in SWS intelligibility between the pre- and post-training sessions were obtained, with no effect on speech-in-noise performance as measured with AzBio sentences in modulated noise (Fig. 4). Controls showed no significant posttest improvement on any speech measure. Using an ANCOVA on posttest scores with pretest as a covariate, a significant effect of training was obtained only for VCV intelligibility [$F(1,23) = 7.95$, $p = .01$, $\eta_p^2 = .26$].

Pearson correlations are shown in Table 1 for both subject groups for pre- and posttest results from the three pattern-reconstruction conditions (i.e., VCV, CNC, and tone sequences) and two SWS measures. With the Bonferroni-Holm correction for multiple comparisons, only the two posttest relationships from the trained

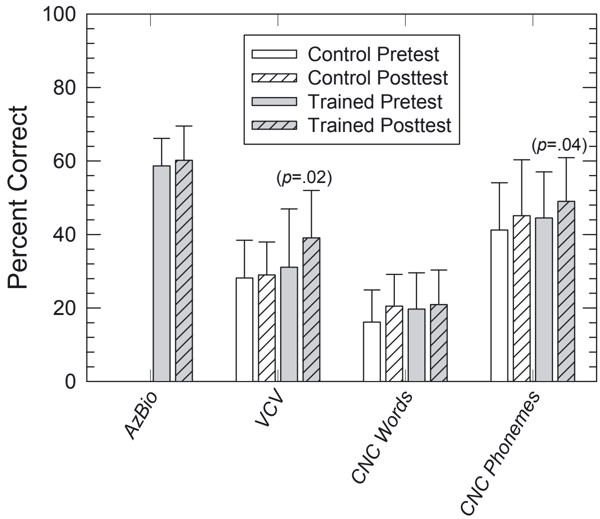

Fig. 4: For the control and trained subject groups, mean group performance on speech tests in terms of percent correct with error bars representing 1 SD. The control group was not tested with AzBio sentences. Within group, significant change from pre- to posttest is indicated by the level of significance in parentheses.

subjects involving the CNC phoneme measure and either the VCV- or CNC-sequence condition were significant in one-tailed testing. Thus, there was some indication of greater relationship post- than pre-training between pattern-reconstruction ability and speech perception as assessed by SWS intelligibility.

DISCUSSION

Watson *et al.* (1976) found that stimulus uncertainty diminished the benefit of training in learning to discriminate a local change to a single element of a tonal sequence. In contrast but consistent with results from Foxton *et al.* (2004), the current study demonstrated, despite a high level of stimulus uncertainty, significant improvement due to training on a task in which the signal was the pattern of the entire stimulus sequence. Though the current study intentionally employed an involved subject task, the effect of training was not solely procedural, as observed by dependence of both training effect and posttest correlation to speech intelligibility on stimulus type in the pattern-reconstruction conditions.

SWS stimuli were used in pattern-reconstruction task to incorporate speech structure into the signal pattern. Across all subjects, performance was best in the pattern task with the VCV stimuli in both the pre- and posttest measures, suggesting benefit from the added sequence structure. In separate pilot work, this benefit was lost when SWS modulation patterns were inverted to disrupt the speech-like structure of stimuli

Control

Pretest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	-.243	-.057	-.418
CNC P(C)	.105	.541	.408

Posttest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	-.279	.411	-.153
CNC P(C)	.283	.652	.428

Trained

Pretest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	.341	.264	.038
CNC P(C)	.355	.665	.191

Posttest	VCV Sequence	CNC Sequence	Tone Sequence
VCV P(C)	.287	.467	-.045
CNC P(C)	**.709 (.04)**	**.703 (.04)**	.182

Table 1: For the control (top) and trained (bottom) subject groups, correlations between results from the three pattern-reconstruction conditions (VCV, CNC, and tone) and two SWS measures (VCV and CNC phonemes). Significant correlations are indicated in bold with the Bonferroni-Holm adjusted one-tailed *p* value in parentheses.

without altering the modulation statistics. In the present work, a role of sequence structure is also observed with performance from only the VCV and CNC pattern-reconstruction conditions showing significant correlation to a measure of speech intelligibility (Table 1). These relationships were obtained only posttest and only after training with a different stimulus type (i.e., tonal sequences). The benefit of training may then in part relate to change in the ability to utilize the manner by which auditory information is structured.

Despite indication of greater relationship post- than pretest between pattern-reconstruction ability and speech perception, the improvement in speech intelligibility due to training was small and limited to the SWS VCV stimulus set. Several factors may have contributed to this outcome. The absence of a training effect on sentence perception in noise may relate to speech redundancy so that young normal-hearing listeners were able to rely on cues other than frequency patterns and transitions, that is, stimulus aspects that were trained. This speculation suggests potentially greater training benefit for older and hearing-impaired listeners due to fewer available, or redundant, speech cues and possibly compromised but trainable frequency-pattern processing ability. Between the two SWS metrics studied, only VCV intelligibility showed a benefit due to training. This may have been due to VCV stimuli being more reliant than CNCs on low-level auditory processing. Correct response to VCVs requires identification of a single consonant while CNCs need processing of all sounds in the stimulus to form an open-set response. Presumably this results in less involvement of high-level linguistic factors in the VCV condition.

In the pattern-reconstruction conditions, effect of training was restricted to the trained stimulus type. The current protocol utilized only three training sessions. Wright *et al.* (2010) demonstrated that generalization lags behind stimulus-specific learning on auditory tasks and may not appear until after four days of training. In terms of practical applications of the current procedure, the concern is less with generalization to other stimulus types in the pattern-reconstruction task than with benefit for speech perception. Results showing strongest posttest relationships between speech intelligibility and the SWS pattern-reconstruction conditions suggest that these stimuli may lead to greater speech benefit from training than obtained with tonal sequences. Future work will investigate the effect of training length and stimulus type with study participants including older and hearing-impaired listeners, that is, populations that may benefit from auditory rehabilitation that includes the current approach to frequency-pattern training.

ACKNOWLEDGMENTS

Work was supported by Award Number R15DC011916 from the National Institute on Deafness and Other Communication Disorders (NIDCD). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDCD or the National Institutes of Health.

REFERENCES

Foxton, J.M., Brown, A.C.B., Chambers, S., and Griffiths, T.D. (**2004**). "Training improves acoustic pattern perception," Current Biol., **14**, 322-325.

Gaab, N., Gaser, C., and Schlaug, G. (**2006**). "Improvement-related functional plasticity following pitch memory training," Neuroimage, **31**, 255-263.

Sheft, S., Risley, R., and Shafiro, V. (**2012a**). "Clinical measures of static and dynamic spectral-pattern discrimination in relationship to speech perception," in *Speech Perception and Auditory Disorders*. Edited by T. Dau, M.L. Jepsen, T. Poulsen, and J.C. Dalsgaard (Danavox Jubilee Fndn., Ballerup), pp. 481-488.

Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (**2012b**). "Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception," Ear Hearing, **33**, 709-720.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R.J., and Näätänen, R. (**2001**). "Superior formation of cortical memory traces for melodic patterns in musicians," Learn. Memory, **8**, 295-300.

Watson, C.S., Kelly, W.J., and Wroton, H.W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am., **60**, 1176-1186.

Wright, B.A., Wilson, R.M., and Sabin, A.T. (**2010**). "Generalization lags behind learning on an auditory perceptual task," J. Neurosci., **30**, 11635-11639.