

Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task^{a)}

Matthew B. Fitzgerald^{b)}

Department of Otolaryngology, New York University School of Medicine, 550 First Avenue, NBV-5E5,
New York, New York 10016

Beverly A. Wright

Department of Communication Sciences and Disorders, Northwestern University, 2240 North Campus Drive,
Evanston, Illinois 60208

(Received 23 December 2009; revised 23 November 2010; accepted 28 November 2010)

Fluctuations in sound amplitude provide important cues to the identity of many sounds including speech. Of interest here was whether the ability to detect these fluctuations can be improved with practice, and if so whether this learning generalizes to untrained cases. To address these issues, normal-hearing adults ($n=9$) were trained to detect sinusoidal amplitude modulation (SAM; 80-Hz rate, 3–4 kHz bandpass carrier) 720 trials/day for 6–7 days and were tested before and after training on related SAM-detection and SAM-rate-discrimination conditions. Controls ($n=9$) only participated in the pre- and post-tests. The trained listeners improved more than the controls on the trained condition between the pre- and post-tests, but different subgroups of trained listeners required different amounts of practice to reach asymptotic performance, ranging from 1 ($n=6$) to 4–6 ($n=3$) sessions. This training-induced learning did not generalize to detection with two untrained carrier spectra (5 kHz low-pass and 0.5–1.5 kHz bandpass) or to rate discrimination with the trained rate and carrier spectrum, but there was some indication that it generalized to detection with two untrained rates (30 and 150 Hz). Thus, practice improved the ability to detect amplitude modulation, but the generalization of this learning to untrained cases was somewhat limited.

© 2011 Acoustical Society of America. [DOI: 10.1121/1.3531841]

PACS number(s): 43.66.Mk [MW]

Pages: 898–906

I. INTRODUCTION

The ability to adequately detect fluctuations, or modulation, in sound amplitude contributes to the accurate perception of many real-world sounds, including speech (e.g., Steeneken and Houtgast, 1980; Plomp, 1983; Rosen, 1992; Drullman *et al.*, 1994a,b; Shannon *et al.*, 1995). It would therefore seem that enhancements in this ability might aid overall hearing performance, particularly for certain clinical populations known to have difficulty with amplitude-modulation (AM) detection (e.g., individuals with dyslexia; Menell *et al.*, 1999; Lorenzi *et al.*, 2000; Rocheron *et al.*, 2002; Witton *et al.*, 2002) or to be particularly reliant on AM cues for communication (e.g., users of cochlear implants; Cazals *et al.*, 1994; Fu, 2002). However, while improvements in the capacity to detect AM have been documented during the course of development (Hall and Grose, 1994), it is not known to what extent this ability can be improved in adulthood. To begin to address this question, we investigated the influence of multiple-session training on AM detection in normal-hearing adults.

We are not aware of any published investigations in which listeners were specifically trained to detect AM, but

there is some evidence that training can affect, both positively and negatively, the perception of AM sounds. In several previous reports, listeners were trained to discriminate between either different values of interaural sound-localization cues using amplitude-modulated signals (Zhang and Wright, 2007, 2009) or different rates of AM (Grimault *et al.*, 2003; Fitzgerald and Wright, 2005). For both tasks, discrimination performance on the single trained condition improved with multiple sessions of practice. These results indicate that the processing of modulated sounds is malleable and suggest that training also could be effective for improving the detection of modulation. However, AM detection actually got worse in the only investigation we know of in which AM detection was examined in a training context. In that case, practice on AM-rate discrimination with a trained rate that elicits a pitch percept (150 Hz) led to decrements in the ability to detect AM at that rate, possibly due to the suboptimal use of a pitch cue in the detection task (Fitzgerald and Wright, 2005). Thus, not only is there a lack of direct evidence that practice can lead to improved AM detection but also the negative generalization from discrimination to detection suggests that different cues may be used for the best performance on these two tasks and therefore that training may affect the two tasks differently.

Here we asked whether directly training AM detection can improve the capacity to detect AM and, if so, whether those improvements generalize to untrained cases. To determine whether multiple-hour training could aid detection performance, we trained normal-hearing adults to detect a

^{a)}Portions of this work were presented in “The influence of practice on the detectability of auditory sinusoidal amplitude modulation,” at the 143rd Meeting of the Acoustical Society of America.

^{b)}Author to whom correspondence should be addressed. Electronic mail: matthew.fitzgerald@nyumc.org

single rate of AM in a fixed carrier over six to seven daily training sessions (using the same training regimen as Fitzgerald and Wright, 2005). Further, because the particular untrained conditions to which any learning generalizes or fails to generalize supplies useful information for understanding the theoretical and practical implications of learning on a given task (Wright and Zhang, 2009), we also examined whether the multiple-hour training influenced performance on five untrained conditions. Four of these untrained conditions employed a detection task, but with different stimuli than in the trained condition. These stimuli included untrained modulation rates and untrained carrier spectra. The fifth untrained condition employed an untrained task, AM-rate discrimination with the trained rate and carrier spectrum, and thus provided an inverse test of the negative generalization from rate discrimination to detection.

II. METHODS

A. Listeners

Eighteen young adults (11 females and seven males; aged 18–23 yr) with no previous experience in psychoacoustic experiments served as listeners. All listeners reported no history of hearing impairment. All were paid an hourly wage, and the trained listeners received a 20% bonus upon completion of the experiment.

B. Experiment organization and procedure

We used the same paradigm as in our previous investigation into sinusoidal amplitude modulation (SAM)-rate-discrimination learning (Fitzgerald and Wright, 2005). Listeners were randomly placed into one of two groups: a trained group ($n = 9$) and a control group ($n = 9$). The trained group completed a pre-test, a training phase, and a post-test. The control group completed only the pre- and post-tests.

The pre- and post-test sessions consisted of five SAM-detection conditions and one SAM-rate-discrimination condition. We obtained five threshold estimates in each of the six conditions (300 total trials per condition). The condition order was randomized across listeners, but each individual listener used the same order in the pre- and post-tests. These ~2-h test sessions were separated by an average of 8.4 days for the trained listeners and 8.7 days for the controls. The training phase occurred between the pre- and post-tests and consisted of six to seven daily ~1-h sessions. In each training session, listeners completed 12 threshold estimates (720 total trials per session) on a single condition (80-Hz SAM rate, 3–4 kHz bandpass carrier). These training sessions occurred on consecutive days, excluding weekends.

In each two-interval forced choice (2IFC) trial, a standard sound was presented in one interval, and a target sound in the other. Listeners indicated which interval contained the target sound by pressing a key on a computer keyboard. In the five SAM-detection conditions, the standard sound was an unmodulated noise, and the target sound a SAM noise. In these conditions, the target sound was a 3–4 kHz bandpass carrier modulated at SAM rates of 30, 80, or 150 Hz, or a 0.5–1.5 kHz bandpass or 5 kHz low-pass carrier modulated at 80 Hz.

In the SAM-rate-discrimination condition, the standard sound was a 3–4 kHz bandpass carrier that was SAM at 80 Hz with a 100% modulation depth, and the target sound was the same carrier with a faster modulation rate.

In each condition, we manipulated either the modulation depth (expressed in decibels relative to 100% modulation depth) or SAM rate of the target sound (in hertz) to determine the modulation detection or rate-discrimination threshold for each listener. These thresholds were obtained by decreasing the modulation depth or rate after three consecutive correct responses and increasing it after each incorrect response. When the depth or rate changed from decreasing to increasing, or vice versa, its value, a reversal, was noted. The first three of these reversal values in each 60-trial block were discarded, and the mean of the largest remaining even number of reversal values was computed to estimate the modulation depth or rate that yielded 79.4% correct performance (Levitt, 1971). We defined this value as threshold. To ensure accuracy, no threshold was computed if there were fewer than four remaining reversal values; this occurred on less than 1% of blocks. For the five SAM-detection conditions, the starting modulation depth (m) was 1 (100% modulation). The modulation depth was varied in units of $20 \log(m)$; the step size was 4 dB until the third reversal and 2 dB thereafter. For the SAM-rate-discrimination condition, the starting difference between the standard and target SAM rates was typically 15 Hz, and the step size was 3 Hz until the third reversal and 1 Hz thereafter. Listeners received feedback after each trial in every measurement phase. Custom-developed software was used to generate the stimuli, control stimulus presentation, and gather the responses.

C. Stimuli

All sounds were generated digitally [Tucker Davis Technologies (TDT) APOS]. To create the SAM noises, we multiplied a broadband Gaussian noise by a DC-shifted sinusoid of 30, 80, or 150 Hz and filtered the waveform after modulation (e.g., Viemeister, 1979). The Gaussian noise was generated on an interval-by-interval basis. The starting phase of the sinusoid was always zero degrees. The amplitude of the SAM noises was reduced by $1 + m^2/2$ to control for the increase in power resulting from AM (e.g., Viemeister, 1979). To make the unmodulated sounds, we filtered a broadband Gaussian noise to the desired bandwidth. All sounds were presented at a spectrum level of 40 dB sound pressure level. The duration of each sound was 400 ms, measured from onset to offset, including 10-ms cosine-squared rise/fall envelopes. The inter-stimulus interval was 600 ms for every condition.

The sounds were played through a 16-bit digital-to-analog converter (TDT DD1) at a sampling rate of 25 kHz, followed by an anti-aliasing filter set to low-pass at 8500 Hz (TDT FT6-2), a programmable attenuator (TDT PA4), a sound mixer (TDT SM3), and a headphone driver (TDT HB6). They were presented through the left earpiece of Sennheiser HD265 headphones. Testing was conducted in a double-walled sound-attenuating booth.

D. Data analysis

We assessed the influence of multiple-hour practice on SAM detection as follows. First, for each condition

separately, we removed the pre- and post-test data of listeners whose pre-test thresholds were greater than two standard deviations from the mean of all ($n = 18$ per condition) of the pre-test values in that condition. This practice helped ensure that the data that were included in the analyses were representative of the general population. It resulted in the removal of the data of one trained listener in the 150-Hz condition who had an unusually low pre-test threshold and one control in the 30-Hz condition whose pre-test threshold was aberrantly high. Second, for each condition separately, we determined whether multiple-hour training influenced performance by comparing the post-test thresholds of the trained listeners and controls using an analysis of covariance (ANCOVA) with pre-test threshold as the covariate. If the homogeneity-of-regression requirement for ANCOVA was violated, we instead performed a two group (trained vs control) \times two time (pre vs post) analysis of variance (ANOVA) with repeated measures on the time factor. In this case, a significant group \times time interaction indicated that the multiple-hour training affected performance. Third, to gain insight into the effect of training at the individual level, we determined the relationship between pre- and post-test performance across individual listeners by computing the linear regression of the post-test threshold on the pre-test threshold for the trained listeners and controls and comparing the regression-line slopes between groups. We also examined the learning curves on the trained condition of the individual trained listeners using both ANOVA and linear regression.

III. RESULTS

A. Learning on the trained condition

On the trained condition (80-Hz SAM rate, 3–4 kHz carrier; Fig. 1, first column), the trained listeners (squares) as a group learned significantly more than the controls (triangles), indicating that multiple-hour practice facilitated the ability to detect AM (ANOVA: group \times time interaction, $F_{1,16} = 12.16$; $p = 0.003$; ANCOVA precluded due to a significant heterogeneity of regression-line slopes: $F_{1,14} = 6.87$;

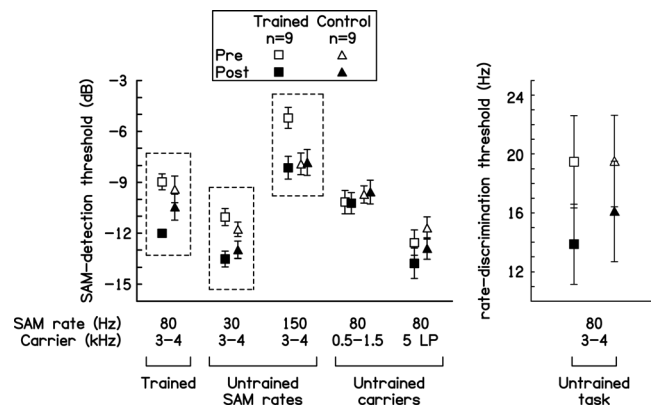


FIG. 1. Mean pre-test (open symbols) and post-test (filled symbols) thresholds of the trained listeners ($n = 9$; squares) and controls ($n = 9$; triangles) for each of the six conditions. Threshold refers to the modulation depth needed to distinguish an amplitude-modulated from an unmodulated noise (SAM-detection) or to the difference in modulation rate needed to distinguish a faster from a slower rate (rate discrimination) on 79% of trials. Error bars reflect \pm one standard error of the mean.

$p = 0.02$). This training-induced learning was also evident at the individual level. Figure 2(A) depicts the relationship between the pre- and post-test thresholds of the individual listeners. For the trained listeners (filled squares), the points all fell below the positive diagonal (solid black no-improvement line), indicating improvement between the pre- and post-tests. The slope of the regression line fitted to these data did not significantly differ from zero (slope: 0.16; $r^2 = 0.11$; $F_{1,7} = 0.90$; $p = 0.37$)¹ and was quite shallow, suggesting that the trained listeners all finished with similar post-test performance, despite the variation in their pre-test thresholds. Thus, the trained listeners with the highest pre-test thresholds tended to show the largest amount of improvement. The data points of all but two of the controls (open triangles) also fell below the diagonal line, indicating improvement. Unlike for the trained listeners, the regression line fitted to these data was significantly different from zero (slope: 0.86; $r^2 = 0.80$; $F_{1,7} = 28.14$; $p = 0.001$) and had a slope approaching 1, suggesting that there was a strong relationship between the pre- and post-test thresholds of these

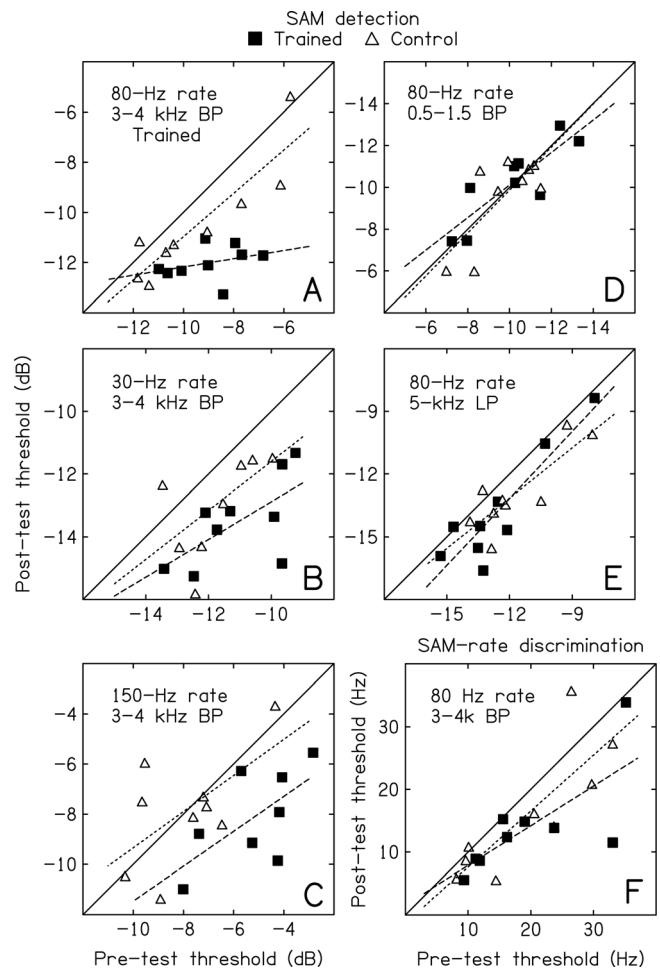


FIG. 2. For each of the six conditions (panels), pre-test (x axis) and post-test (y axis) thresholds are shown for the trained listeners (filled squares) and controls (open triangles). The linear regression of the post-test thresholds on the pre-test thresholds was determined for each data set. Separate lines were estimated for trained listeners (long dashes) and controls (short dashes). The solid diagonal line in each panel indicates a regression-line slope of 1; points below this solid line reflect improvement between the pre- and post-tests.

SAM detection
80-Hz rate
3-4 kHz carrier

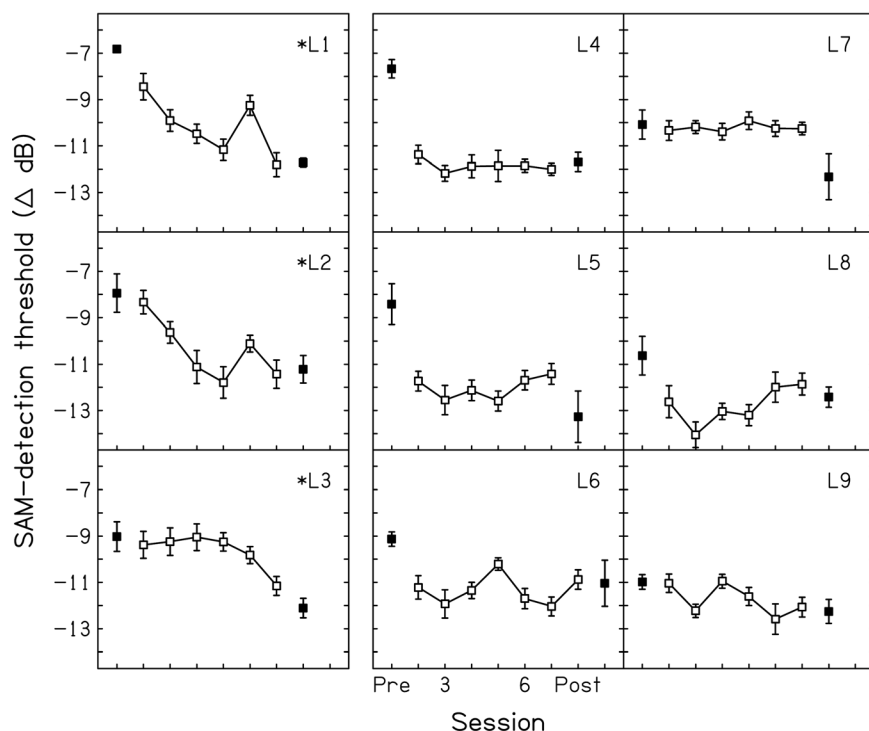


FIG. 3. Thresholds on the trained condition from the pre- and post-tests (filled squares) and during the training phase (open squares) are shown for all nine trained listeners (panels). Error bars indicate \pm one standard error of the mean within a given listener. Three listeners improved across the six training sessions (left column; asterisks next to the listener label). The remaining six listeners also improved between the pre- and post-tests, but not across the six training sessions (middle and right columns).

listeners. Thus, the controls improved, but by a relatively constant amount regardless of pre-test threshold. Finally, the regression-line slopes differed significantly between the trained listeners and controls, further substantiating the differences in behavior between these groups (heterogeneity of regression, $F_{1,14} = 6.87$; $p = 0.02$).

While the trained listeners learned significantly more than the controls between the pre- and post-tests, some trained listeners took markedly longer than others to reach asymptotic performance during the training phase. All nine trained listeners finished with thresholds similar to those previously reported for highly trained listeners tested with similar stimuli (Eddins, 1993). However, of these listeners, only three met our previous criteria for having learned during the training phase itself (L1–L3; Fig. 3, left column) in that (1) their thresholds changed significantly across the training sessions as determined by a one-way ANOVA on the training-phase data ($p < 0.05$ in all cases), and (2) a regression line fitted to their data had a significant negative slope ($p < 0.05$ in all cases; see also Wright *et al.*, 1997; Wright and Fitzgerald, 2001; Fitzgerald and Wright, 2005). These three listeners had among the highest pre-test thresholds for SAM detection on the trained condition as well as on the two untrained rates, and on the SAM-rate-discrimination condition. However, a high pre-test threshold did not guarantee that learning would extend over multiple training sessions, as two trained listeners who reached asymptotic performance at the end of the first training session (see below) also had high pre-test thresholds.

The six listeners (L4–L9; Fig. 3, middle and right columns) who did not meet our previous criteria for having learned during the training phase nevertheless showed improvement attributable to the training phase. On average,

these listeners learned significantly and reached asymptotic performance *within* the first training session. Figure 4 depicts the thresholds from the pre-test, each training session, and the post-test for the three trained listeners who met our previous criteria for training-phase learning (squares), the six

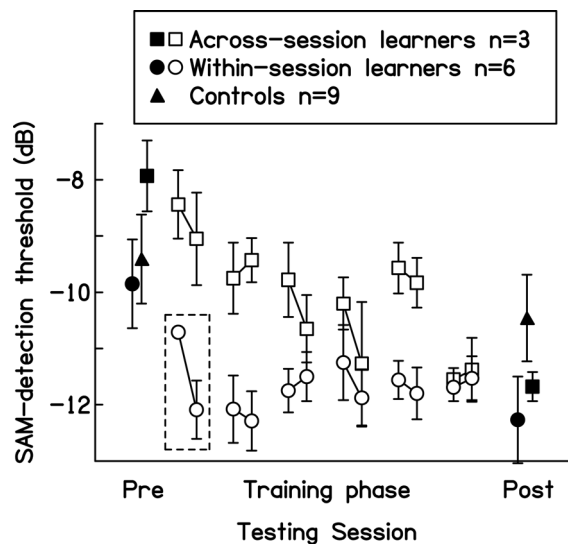


FIG. 4. Mean thresholds on the trained condition from the pre- and post-tests (filled symbols) as well as from the first and second halves of each training session (open symbols). Data are shown separately for the trained listeners who improved significantly across the six training sessions (across-session learners, $n = 3$; squares), for the trained listeners who improved between the pre- and post-tests, but not across the six training sessions (within-session learners, $n = 6$; circles), and for controls (triangles, $n = 9$; triangles). Error bars indicate \pm one standard error of the mean across listeners. Thresholds that differed significantly between the first and second halves of a training session are enclosed by a dashed rectangle.

trained listeners who did not (circles), and the controls (triangles). For the two subgroups of trained listeners, mean thresholds are shown for the first six (first half) and for the last six (second half) estimates from each training session to facilitate the examination of within-session learning. Such learning was assessed in the six trained listeners who did not meet our criteria for across-session learning. A two time (first vs second half of each training session) \times six session ANOVA with repeated measures on both factors conducted on these data yielded a significant time \times session interaction ($F_{5,40} = 3.48$; $p = 0.01$). *Post-hoc* analyses indicated that the thresholds of these listeners differed within a session only on the first of the six sessions [$t_{(5)} = 3.27$; $p = 0.02$]. Notably, their thresholds in the first half of that session tended to be lower than in the pre-test [$t_{(5)} = 2.19$; $p = 0.08$] and did not differ from the post-test thresholds of the controls [$t_{(5)} = 0.26$; $p = 0.80$]. This result, in combination with the significant improvement shown by the controls between the pre- and post-tests [$t_{(8)} = 2.94$; $p = 0.019$], suggests that pre-test exposure alone yielded learning that could be maintained without additional training (see also Mossbridge *et al.*, 2006, 2008). However, by the second half of the first training session, the thresholds of these six trained listeners were lower than their own first-half thresholds [$t_{(5)} = 3.27$; $p = 0.022$] and did not differ from their post-test thresholds [$t_{(5)} = 0.15$; $p = 0.89$]. Thereafter, there were no additional within- or across-session improvements. Thus, after completing the pre-test, it appears that the majority of listeners required more than 360, but fewer than 720 practice trials (20–45 min) to reach asymptotic performance for the detection of 80-Hz modulation. The apparent retention of learning induced only by pre-test exposure raises the possibility that these six trained listeners would have maintained their improvement even had they received no training beyond the first \sim 720 training trials. The small size of the subgroup of listeners who met our criteria for significant training-phase learning ($n = 3$) precludes statistical analysis of whether these listeners also showed within-session learning in addition to their across-session improvements.

B. Generalization to untrained conditions

The training-induced learning did not generalize to either of the two untrained carrier spectra [5 kHz low-pass: ($F_{1,13} = 0.005$; $p = 0.95$); 0.5–1.5 kHz bandpass: ($F_{1,13} = 0.19$; $p = 0.67$)] or to the untrained SAM-rate-discrimination task which shared the same trained rate and carrier spectrum ($F_{1,13} = 0.59$; $p = 0.45$) (Fig. 1, three right-most column). For the 5 kHz low-pass noise [Fig. 2(E)], the data points depicting the relationship between the pre- and post-test performance of the individual listeners were intermixed between trained listeners and controls and mostly fell below the diagonal, implying learning for both groups. The slopes of the regression lines fitted to these data were significantly different from zero and steep for the trained listeners (slope: 1.07; $r^2 = 0.81$; $F_{1,8} = 30.06$; $p < 0.001$) as well as for the controls (slope: 0.80; $r^2 = 0.70$; $F_{1,8} = 16.12$; $p = 0.005$). Moreover, neither the slopes (heterogeneity of regression, $F_{1,14} = 0.91$; $p = 0.36$) nor the y -intercepts (ANCOVA, see above) of these lines dif-

fered significantly between the two groups. The individual data and regression-line analyses for the rate-discrimination condition [Fig. 2(F)] followed a similar pattern (trained: slope: 0.63; $r^2 = 0.53$; $F_{1,8} = 7.91$; $p = 0.033$; control: slope: 0.89; $r^2 = 0.66$; $F_{1,8} = 13.78$; $p = 0.008$; heterogeneity of regression, $F_{1,14} = 0.62$; $p = 0.44$; ANCOVA, see above). These results suggest that for these two conditions, both the trained listeners and controls improved, but by similar amounts, and that the magnitude of this improvement was relatively constant regardless of the pre-test threshold. For the 0.5–1.5 kHz bandpass condition [Fig. 2(D)], the data points of the two groups and corresponding regression lines were essentially superimposed on the diagonal line, indicating no improvement by either group and no difference between the two (trained: slope: 0.78; $r^2 = 0.72$; $F_{1,8} = 18.18$; $p = 0.004$; control: slope: 1.03; $r^2 = 0.56$; $F_{1,8} = 8.90$; $p = 0.02$; heterogeneity of regression, $F_{1,14} = 0.45$; $p = 0.513$; ANCOVA: see above).

Though the results are less clear, there are some indications that learning generalized to the two untrained modulation rates. For the untrained 30-Hz rate, there was only a trend for the post-test thresholds of the trained listeners to be lower than those of the controls when the pre-test threshold was a covariate ($F_{1,14} = 3.23$; $p = 0.094$) (Fig. 1, second column). However, this trend in combination with other aspects of the data provides moderate evidence that training-induced learning generalized to this condition. Specifically, at the group level, while the pre-test thresholds did not differ between the groups [$t_{(15)} = -1.07$; $p = 0.300$], the improvement between the pre- and post-test thresholds was significantly greater in the trained listeners than the controls (ANOVA time \times group interaction, $F_{1,15} = 4.81$; $p = 0.044$). At the individual level, the data points depicting the relationship between the pre- and post-test performance fell below the diagonal for all but one of the individual listeners, denoting learning in both groups [Fig. 2(B)]. Yet, for similar pre-test values, the points of the trained listeners were generally lower than those of the controls. The slopes of the regression lines fitted to the data of each group did not differ significantly between the groups ($F_{1,13} = 0.16$; $p = 0.699$), but the y -intercept for the trained listeners was nearly significantly different from, and less than, zero [$t_{(7)} = 2.29$; $p = 0.056$], while that for the controls was not [$t_{(7)} = 0.87$; $p = 0.419$], again suggesting that the trained listeners may have improved more than the controls.

For the 150-Hz condition, the difficulty in determining whether learning generalized arises because the pre-test thresholds of the trained listeners were significantly higher than those of the controls [$t_{(15)} = 3.03$; $p = 0.008$], while the post-test thresholds did not differ significantly between the two groups [$t_{(14)} = 0.30$; $p = 0.771$; ANOVA group \times time interaction, $F_{1,15} = 8.64$; $p = 0.011$, ANCOVA not conducted due to the group differences in the pre-test thresholds] (Fig. 1, third column). Thus, these group-level analyses cannot rule out the possibility that the controls started at a performance floor and that the trained listeners, who by chance started more poorly,² reached that same floor simply through exposure to the pre-test rather than through the generalization of training-phase learning. However, the individual data points depicting the relationship between the

pre- and post-test performance provide some support for the idea that multiple-hour training contributed to performance on this condition [Fig. 2(C)]. The points of all of the trained listeners fell below the diagonal, indicating learning between the pre- and post-tests, while those of the controls were distributed around the diagonal, indicating a lack of learning. Further, for similar pre-test values, the points of the trained listeners were generally lower than those of the controls. The regression lines fitted to the data of each group did not differ significantly in slope ($F_{1,13} = 0.002$; $p = 0.961$), but the y-intercept differed significantly from, and was less than, zero for the trained listeners [$t_{(7)} = 2.45$; $p = 0.050$], but not for the controls [$t_{(7)} = 0.73$; $p = 0.489$]. This configuration of results suggests that multiple-hour training may have generalized to the untrained 150-Hz rate.

Finally, it is notable that for each of the two untrained modulation rates, the slopes of the regression lines fitted to the data were almost significantly different from zero and relatively steep for both the trained listeners (30 Hz, slope: 0.6; $r^2 = 0.41$; $F_{1,7} = 4.95$; $p = 0.061$; 150 Hz, slope: 0.70; $r^2 = 0.41$; $F_{1,7} = 4.23$; $p = 0.086$) and controls (30 Hz, slope: 0.78; $r^2 = 0.43$; $F_{1,7} = 4.44$; $p = 0.08$; 150 Hz, slope: 0.72; $r^2 = 0.36$; $F_{1,7} = 3.98$; $p = 0.086$) and, as mentioned above, did not differ between those groups (heterogeneity of regression, 30 Hz: $F_{1,13} = 0.16$; $p = 0.699$; 150 Hz, $F_{1,13} = 0.002$; $p = 0.961$). Thus, for both the trained listeners and controls, the magnitude of any improvement appeared to be relatively independent of the pre-test threshold in these conditions. This pattern differs from that on the trained condition, in which the regression-line slope of the trained listeners was quite shallow and differed significantly from the steeper slope of the controls. Summarized another way, for the controls, the regression lines had similar slopes (steep) for all of these conditions, but for the trained listeners, the slopes differed between the trained condition (quite shallow) and the two untrained rates (steep). The difference in the slopes for the trained listeners between the trained and untrained rates may be an indication that the generalization to the untrained rates, to the extent that it occurred, was not complete. This possibility arises from the assumption that the slopes for the untrained rates would have been shallow had those conditions been the trained ones, and therefore, could have been, but just were not, shallow following practice on the current trained condition. However, it is also possible that the slopes for the untrained rates would have remained steep even with direct training, in which case, the generalization could be considered complete, at least by this measure (see Wright and Zhang, 2009, for a review of generalization).

IV. DISCUSSION

The present data demonstrate that training can help normal-hearing adults to better detect SAM. Exposure to the pre-test alone led to improved performance over a week later at the post-test in controls who did not receive any intervening training. Additional training between those two tests led to even greater improvements. Listeners who received six to seven daily sessions of training on a single modulation-detection condition (80-Hz rate, 3–4 kHz carrier spectrum)

improved significantly more on that condition than did controls. However, different listeners required different amounts of training to reach their best performance. Most needed only ~ 1 h of additional practice after the ~ 2 -h pre-test, but others required 4–6 h of training. The learning of the trained listeners did not generalize to untrained carrier spectra (0.5–1.5 kHz, 5 kHz low-pass noise) or to a modulation-rate-discrimination task that shared the trained rate and carrier spectrum, but there was some indication that it aided performance on untrained modulation rates (30 and 150 Hz).

It appears that at least two different types of learning—stimulus and task learning—contributed to the improved performance on SAM detection shown by the trained listeners. We define stimulus learning as learning associated with specific feature values of the stimulus used during training and task learning as learning associated with the particular perceptual judgment to be made. The inference that each of these types of learning contributed to the overall improvements observed is based on the pattern of conditions to which learning on the trained SAM-detection condition did and did not generalize (for a review of definitions of learning type and this general approach, see Ortiz and Wright, 2009). Stimulus learning is inferred from a lack of generalization to untrained stimulus features on the trained task. The trained listeners showed evidence of this type of learning at the end of training because their learning on the trained SAM-detection task did not generalize to untrained carrier spectra. Task learning is inferred instead from a lack of generalization to untrained tasks that utilize the trained stimulus. It is implicated in the current investigation because the learning on SAM detection did not generalize to SAM discrimination even though the modulation rate and carrier spectrum were the same in both cases.

It is not clear what types of learning contributed to the improvements shown by the controls. The controls learned between the pre- and post-tests on four of six conditions, and on one of the conditions on which they did not learn, they may have already been at or near a performance floor (150-Hz rate). Unfortunately, the current experimental paradigm precludes the use of generalization patterns to determine what type(s) of learning played a role in their improvements. This is because the controls were exposed to each condition on the pre-test. Thus, we cannot determine whether their learning on any particular condition was specific to that condition, or whether it resulted from generalization from another condition. Rapid improvements that were observed in the controls sometimes have been attributed to procedural learning (e.g., Recanzone *et al.*, 1993; Wright and Fitzgerald, 2001), which we define as learning associated with any components of the training experience outside of the trained stimulus and task (such as the experimental setting, testing method, and response demands). However, by our definitions, they can also arise from stimulus (e.g., Demany, 1985; Amitay *et al.*, 2006) and task (e.g., Hawkey *et al.*, 2004) learning (for a review, see Ortiz and Wright, 2009). Because the trained listeners took part in the same pre-test as the controls, whatever types of learning contributed to the improvements of the controls also affected the performance of the trained listeners.

In addition to providing insight into the types of learning that contributed to the improvements, the pattern of generalization can also be used to make more detailed inferences about the processes that were modified by training. Improvements in performance on perceptual tasks have been attributed to refinements in one or more processing stages, ranging from the initial sensory representation (e.g., Karni and Sagi, 1991; Poggio *et al.*, 1992; Ahissar and Hochstein, 2004; Fahle, 2004) to the weighting and interpretation of that sensory information (e.g., Mollon and Danilova, 1996; Doshier and Lu, 1998; Petrov *et al.*, 2005). Depending on the view adopted, the pattern of generalization is attributed either to the tuning characteristics of the particular circuitry that was modified or to the distribution of weights given to particular aspects of the stimulus, task, and procedure used during training. Whichever of these possibilities ultimately proves to be the case for learning on SAM detection, the basic assumption regarding generalization is the same: Improvements on a trained condition result from refinements somewhere along the processing pathway and these improvements spread to untrained conditions only if performance on those conditions is mediated by the processing that was refined by the training (e.g., Ahissar and Hochstein, 1996; Wright and Fitzgerald, 2001; Demany and Semal, 2002; for review, see Wright and Zhang, 2009).

Based on this assumption, the apparent generalization of learning to untrained SAM rates, but not to untrained carrier spectra suggests that training modified processes in which different modulation rates are treated similarly, but different carrier spectra differently. Processes with these characteristics have been proposed as components of two psychophysical models of the detection of AM. Versions of both the low-pass filter model (Viemeister, 1979) and the modulation-filterbank model (e.g., Dau *et al.*, 1997a,b) include a low-pass filter, which is a mechanism that could account for common processing of multiple modulation rates [low-pass cutoff: ~ 64 Hz (Viemeister, 1979), ~ 150 Hz (Ewert and Dau, 2000; Kohlrausch *et al.*, 2000; Jepsen *et al.*, 2008)]. Both models also assume some degree of carrier specificity, whether via a pre-detection filter (Viemeister, 1979) or by only considering information within a particular carrier (Dau *et al.*, 1997a,b). There is also physiological evidence of these characteristics in numerous reports of single AM-sensitive neurons that phase lock to many different modulation rates and that are sharply tuned for carrier frequency (see Joris *et al.*, 2004, for a review). However, given the tentative nature of the conclusion that learning generalized to untrained rates, it remains possible that the learning modified processes that were specific to both rate and carrier. Such processes have been implicated at the behavioral (e.g., modulation-masking; Houtgast, 1989; Bacon and Grantham, 1989; Yost *et al.*, 1989) and physiological (see Joris *et al.*, 2004, for a review) levels and incorporated in the modulation-filterbank model (e.g., Dau *et al.*, 1997a,b).

Using the same assumption, the lack of generalization from SAM detection to SAM-rate-discrimination suggests that training on detection modified different processes from those used for optimal rate discrimination. This result echoes the previous observation that training on rate discrimination

actually led to decrements in the ability to detect modulation (Fitzgerald and Wright, 2005) and thus shows that the lack of generalization of improvement between these two tasks is bidirectional. The finding that training on neither task aided performance on the other suggests that the processes governing optimal performance on SAM detection and SAM discrimination tasks are separable, regardless of whether they are instantiated by different neural substrates or the differential weighting of cues in a common decision process. In this context, it is interesting to note that there were differences in the rate of learning and the generalization patterns between SAM detection (80-Hz rate; here) and SAM-rate discrimination (150-Hz rate; Fitzgerald and Wright, 2005), despite the use of the same training regimen for both. In terms of learning rate, only three of nine listeners required more than a single training session beyond the pre-test to reach asymptotic performance on modulation detection with an 80-Hz rate, while nine of nine listeners required multiple training sessions on rate discrimination with a 150-Hz rate. Similarly, exposure to the pre-test alone yielded considerable improvements in detection, such that the post-test thresholds approached those of highly trained listeners, but this did not occur for rate discrimination. With regard to generalization patterns, SAM-detection training with an 80-Hz rate yielded learning that appears to have generalized to rates of 30 and 150 Hz, while SAM-discrimination training at 150 Hz generalized only partially to a 300-Hz rate and not at all to a 30-Hz rate. While these differences could be a consequence of the different modulation rates used in the two investigations, they also could have resulted, at least in part, from the different tasks.

The proposed separation of the processes underlying optimal modulation detection and rate discrimination is consistent with the lack of a direct relationship between modulation rate-discrimination thresholds with a 100% modulation depth and modulation detection thresholds in individual listeners (Grant *et al.*, 1998). We note however that this separation does not indicate that the processes underlying performance on these two tasks are entirely independent. For example, there are several reports showing that one factor that limits modulation-rate-discrimination performance is the ability to detect modulation. Reductions in modulation depth can lead to poorer rate-discrimination thresholds (Burns and Viemeister, 1976; Patterson and Johnson-Davies, 1978), and the ability to detect modulation and to discriminate modulation rate appear to deteriorate in parallel with increasing modulation rate (Patterson and Johnson-Davies, 1978). Conversely, once a certain modulation depth is reached, further increases in modulation depth do not aid rate discrimination, which is in accord with the current lack of generalization between detection and rate discrimination.

The idea that separable processes subserve modulation detection and rate discrimination, coupled with the known influence of modulation depth on rate discrimination, suggests the following potential explanations for the mutual lack of benefit received from training on the other task. The negative generalization from discrimination to detection could be accounted for if the discrimination-trained listeners used the processes that typically underlie discrimination

performance suboptimally for detection (see also [Fitzgerald and Wright, 2005](#)). They may have done so because multiple-hour training on discrimination taught them to focus on the discrimination processes and those processes provided a suboptimal cue for detection (most likely a pitch cue, because the modulation rate was 150 Hz; [Fitzgerald and Wright, 2005](#)). This listening strategy would result in poorer modulation detection thresholds because deeper modulation depths are required to detect the presence of pitch than to detect the modulation itself ([Burns and Viemeister, 1976](#); [Patterson and Johnson-Davies, 1978](#)). In turn, the lack of generalization from detection to discrimination could be accounted for by assuming that listeners monitored the processes typically underlying detection performance during their training on detection but accessed those processes that typically underlie discrimination performance during their post-testing on discrimination. They may have used this strategy because the detection processes offered no cue for discrimination. In this scenario, detection training did not aid discrimination performance because the 100% modulation depth used in the discrimination task far exceeded the depth needed to obtain the best possible rate-discrimination threshold (e.g., [Patterson and Johnson-Davies, 1978](#)). Thus, the increases in sensitivity to the presence of much shallower modulation depths induced by modulation detection training would have no further effect on modulation rate-discrimination ability.

Finally, it is interesting to note that the AM detection and rate-discrimination tasks may have placed different demands on short-term memory. The modulation detection task potentially could have been performed by monitoring the percept elicited by a single interval to see if the cue used for detection was present (potentially a change in stimulus level or quality). In contrast, discrimination of modulation rate more likely required that the percept elicited by the first interval be held in short-term memory and compared to the percept elicited by the second interval so that the cue monitored in each interval (potentially a pitch cue) could be compared. Such apparent differences in the short-term memory demands for these two tasks may have contributed to the lack of generalization between them as well as to the other differences in the influence of training on AM detection and discrimination.

V. SUMMARY

In summary, practice led to significant improvements in the ability to detect AM. These improvements did not generalize to modulation detection with two untrained carriers or to modulation-rate discrimination, but there was some indication that they did generalize to the detection of two untrained modulation rates. This learning may have arisen from modifications in circuitry specialized for modulation encoding or from the reweighting of different cues in a decision process. In either event, it appears that the present training on modulation detection modified processes in which different carrier spectra are treated differently, but different modulation rates seem to be treated similarly. It also appears that these processes are separate from those used for optimal

rate discrimination. These conclusions are consistent with some previous data regarding the processing of AM and the relationship between modulation detection and rate discrimination.

ACKNOWLEDGMENTS

Karen Banai, Julia Huyck, Nicole Marrone, Julia Mossbridge, Jeanette Ortiz, Andrew Sabin, Stan Sheft, Steve Zecker, Yuxuan Zhang, and two anonymous reviewers provided insightful comments on earlier drafts of this paper. This research was funded by the National Institutes of Health/National Institute for Deafness and other Communication Disorders.

¹Of all of the regression-line slopes described in the results, the only one that differed significantly from 1 was for the trained listeners on the trained condition [$t(7) = 4.86$; $p = 0.0018$; for all others $p \geq 0.15$].

²The higher pre-test thresholds of the trained listeners than controls on the 150-Hz condition appears to be due to chance, because the trained listeners did not perform more poorly than the controls on the pre-test of any of the other five conditions, and the randomization of the condition order did not differ between the groups.

- Ahissar, M., and Hochstein, S. (1996). "Learning pop-out detection," *Vision Res.* **36**, 3487–3500.
- Ahissar, M., and Hochstein, S. (2004). "The reverse hierarchy theory of visual perceptual learning," *Trends Cogn. Sci.* **8**, 457–464.
- Amitay, S., Irwin, A., and Moore, D. R. (2006). "Discrimination learning induced by training with identical stimuli," *Nat. Neurosci.* **9**, 1446–1448.
- Bacon, S. P., and Grantham, D. W. (1989). "Modulation masking: Effects of modulation frequency, depth, and phase," *J. Acoust. Soc. Am.* **85**, 2575–2580.
- Burns, E. M., and Viemeister, N. F. (1976). "Nonspectral pitch," *J. Acoust. Soc. Am.* **60**, 863–869.
- Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (1994). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," *J. Acoust. Soc. Am.* **96**, 2048–2054.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.* **102**, 2892–2905.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2905–2919.
- Demany, L. (1985). "Perceptual learning in frequency discrimination," *J. Acoust. Soc. Am.* **78**, 1118–1120.
- Demany, L., and Semal, C. (2002). "Learning to perceive pitch differences," *J. Acoust. Soc. Am.* **111**, 1377–1387.
- Dosher, B. A., and Lu, Z. L. (1998). "Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting," *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13988–13993.
- Drullman, R., Festen, J. M., and Plomp, R. (1994a). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Drullman, R., Festen, J. M., and Plomp, R. (1994b). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Eddins, D. A. (1993). "Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region," *J. Acoust. Soc. Am.* **93**, 470–479.
- Ewert, S. D., and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* **108**, 1181–1196.
- Fahle, M. (2004). "Perceptual learning: A case for early selection," *J. Vision* **4**, 879–890.
- Fitzgerald, M. B., and Wright, B. A. (2005). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," *J. Acoust. Soc. Am.* **118**, 3794–3803.
- Fu, Q. J. (2002). "Temporal processing and speech recognition in cochlear implant users," *NeuroReport* **13**, 1635–1639.
- Grant, K. W., Summers, V., and Leek, M. R. (1998). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **104**, 1051–1060.

- Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., and Collet, L. (2003). "Learning in discrimination of frequency or modulation rate: Generalization to fundamental frequency discrimination." *Hear. Res.* **184**, 41–50.
- Hall, J. W. III, and Grose, J. H. (1994). "Development of temporal resolution in children as measured by the temporal modulation transfer function." *J. Acoust. Soc. Am.* **96**, 150–154.
- Hawkey, D. J., Amitay, S., and Moore, D. R. (2004). "Early and rapid perceptual learning." *Nat. Neurosci.* **7**(10), 1055–1056.
- Houtgast, T. (1989). "Frequency selectivity in amplitude-modulation detection." *J. Acoust. Soc. Am.* **85**, 1676–1680.
- Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). "A computational model of human auditory signal processing and perception." *J. Acoust. Soc. Am.* **124**, 422–438.
- Joris, P. X., Schreiner, C. E., and Rees, A. (2004). "Neural processing of amplitude-modulated sounds." *Physiol. Rev.* **84**, 541–577.
- Karni, A., and Sagi, D. (1991). "Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity." *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4966–4970.
- Kohlrausch, A., Fassel, R., and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers." *J. Acoust. Soc. Am.* **108**, 723–734.
- Levitt, H. (1971). "Transformed up-down procedures in psychoacoustics." *J. Acoust. Soc. Am.* **49**, 467–477.
- Lorenzi, C., Dumont, A., and Fullgrabe, C. (2000). "Use of temporal envelope cues by children with developmental dyslexia." *J. Speech Lang. Hear. Res.* **43**, 1367–1379.
- Menell, P., McNally, K. I., and Stein, J. F. (1999). "Psychophysical sensitivity and physiological response to amplitude modulation in adult dyslexic listeners." *J. Speech Lang. Hear. Res.* **42**, 797–803.
- Mollon, J. D., and Danilova, M. V. (1996). "Three remarks on perceptual learning." *Spatial Vis.* **10**, 51–58.
- Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., and Wright, B. A. (2006). "Perceptual-learning evidence for separate processing of asynchrony and order tasks." *J. Neurosci.* **26**(49), 12708–12716.
- Mossbridge, J. A., Scissors, B. N., and Wright, B. A. (2008). "Learning and generalization on asynchrony order tasks at sound offset: Implications for underlying neural circuitry." *Learn. Memory* **15**(1), 13–20.
- Ortiz, J. A., and Wright, B. A. (2009). "Contributions of procedure and stimulus learning to early, rapid perceptual improvements." *J. Exp. Psychol. Hum. Percept. Perform.* **35**(1), 188–194.
- Patterson, R. D., Johnson-Davies, D., and Milroy, R. (1978). "Amplitude-modulated noise: The detection of modulation versus the detection of modulation rate." *J. Acoust. Soc. Am.* **63**, 1904–1911.
- Petrov, A. A., Doshier, B. A., and Lu, Z.-L. (2005). "The dynamics of perceptual learning: An incremental reweighting model." *Psychol. Rev.* **112**(4), 715–743.
- Plomp, R. (1983). "The role of modulation in hearing," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman (Springer-Verlag, Berlin), pp. 270–276.
- Poggio, T., Fahle, M., and Edelman, S. (1992). "Fast perceptual learning in visual hyperacuity." *Science* **256**, 1018–1021.
- Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (1993). "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult monkeys." *J. Neurosci.* **13**, 87–103.
- Rocheron, I., Lorenzi, C., Fullgrabe, C., and Dumont, A. (2002). "Temporal envelope perception in dyslexic children." *NeuroReport* **63**, 1904–1911.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory, and linguistic aspects." *Philos. Trans. R. Soc. London, Ser. B* **336**, 367–373.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wyganski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues." *Science* **270**(5324), 303–304.
- Steeneken, H. J., and Houtgast, T. (1980). "A physical method for measuring speech-transmission quality." *J. Acoust. Soc. Am.* **67**, 318–326.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds." *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Witton, C., Stein, J. F., Stoodley, C. J., Rosner, B. S., and Talcott, J. B. (2002). "Separate influences of acoustic AM and FM sensitivity on the phonological decoding skills of impaired and normal readers." *J. Cogn. Neurosci.* **14**, 866–874.
- Wright, B. A., Buonomano, D. V., Mahncke, H. W., and Merzenich, M. M. (1997). "Learning and generalization of auditory temporal-interval discrimination in humans." *J. Neurosci.* **17**(10), 3956–3963.
- Wright, B. A., and Fitzgerald, M. B. (2001). "Different patterns of human discrimination learning for two interaural cues to sound-source location." *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12307–12312.
- Wright, B. A., and Zhang, Y. (2009). "A review of the generalization of auditory learning." *Philos. Trans. R. Soc. London, Ser. B* **364**(1515), 301–311.
- Yost, W. A., Sheft, S., and Opie, J. (1989). "Modulation interference in detection and discrimination of amplitude modulation." *J. Acoust. Soc. Am.* **86**, 2138–2147.
- Zhang, Y., and Wright, B. A. (2007). "Similar patterns of learning and performance variability for human discrimination of interaural time difference at high and low frequencies." *J. Acoust. Soc. Am.* **121**, 2207–2216.
- Zhang, Y., and Wright, B. A. (2009). "An influence of amplitude modulation on interaural level difference processing as suggested by learning patterns of human adults." *J. Acoust. Soc. Am.* **126**, 1349–1358.