Multiple Looks in Speech Sound Discrimination in Adults

Rachael Frush Holt and **Arlene Earley Carney**
University of Minnesota, Minneapolis

Abstract

N. F. Viemeister and G. H. Wakefield's (1991) multiple looks hypothesis is a theoretical approach from the psychoacoustic literature that has promise for bridging the gap between results from speech perception research and results from psychoacoustic research. This hypothesis accounts for sensory detection data and predicts that if the "looks" at a stimulus are independent and information is combined optimally, sensitivity should increase for 2 pulses relative to 1 pulse. Specifically, d' (a bias-free measure of sensitivity) for 2 pulses should be larger than d' for 1 pulse. One speech discrimination paradigm that presents stimuli with multiple presentations is the change/no-change procedure. On a change trial, the standard and comparison stimuli differ; on a no-change trial, they are the same. Normal-hearing adults were tested using the change/no-change procedure with 3 consonant–vowel minimal pairs in combinations of 1, 2, and 4 repetitions of standard and comparison stimuli at various signal-to-noise ratios. If multiple looks extend to this procedure, performance should increase with higher repetition numbers. Performance increased with more presentations of the speech contrasts tested. The multiple looks hypothesis predicted performance better at low repetition numbers when performance was near d' values of 1.0 than at higher repetition numbers and higher performance levels.

There is a long history of attempting to relate results from psycho-acoustic experiments to those from speech perception experiments for normal-hearing listeners, listeners with hearing loss, and individuals with cochlear implants. Several researchers have investigated the relation between psychophysical tuning curves and speech perception in normal-hearing listeners and listeners with hearing loss (e.g., Faulkner, Rosen, & Moore, 1990; Stelmachowicz, Jesteadt, Gorga, & Mott, 1985). Some investigators have assessed the effects of stimulus bandwidth on speech perception in normal-hearing and hearing-impaired individuals (e.g., Skinner, 1980; Stelmachowicz, Pittman, Hoover, & Lewis, 2001), while others also have varied the stimulus bandwidth but examined the effects in individuals with cochlear dead regions (where there are believed to be no functioning inner hair cells and/or neurons). These investigators have used psychophysical methods to identify dead regions in adults with high-frequency sensorineural hearing loss and have proposed that identifying dead regions has important implications for fitting amplification (e.g., Vickers, Moore, & Baer, 2001). Although it is certain that psychoacoutic properties of the auditory system have a major role in human speech perception, the relationship is still not well understood.

One theoretical approach that has promise for speech perception research is Viemeister and Wakefield's (1991) multiple looks hypothesis. The multiple looks hypothesis was developed as an alternative to the nonparsimonious conclusion that there are two time constants involved in temporal integration or summation. Temporal integration refers to the well-known fact that auditory thresholds decrease with increasing signal durations. The

Contact author: Rachael Frush Holt, now at Speech and Hearing Sciences, 200 South Jordan Avenue, Bloomington, IN 47405. raholt@indiana.edu.

phenomenon has been reported under varying conditions and for a range of auditory signals (Gerken, Bhat, & Hutchinson-Clutter, 1990; Scharf, 1978). Previous reports have suggested that the auditory system behaves like a perfect power integrator over relatively long signal durations (e.g., a few hundred milliseconds); that is, it displays a 3-dB decrease in detection threshold for every doubling of signal duration. The models that have been developed to account for temporal integration can be separated into two classes based on the time constants proposed by each. The power summation models, such as that described above, propose long time constants over which the system integrates the auditory input, whereas other models propose much shorter time constants. The models based on shorter time constants account for temporal resolution data, such as those from modulation and gap detection experiments. The models proposing longer time constants do a poor job predicting the temporal resolution data, whereas the shorter time constant models do a poor job predicting the temporal summation data. This discrepancy has been termed the integration-resolution paradox (de Boer, 1975; Green, 1985). The mechanism(s) underlying temporal integration are not yet well understood. However, based on the results of two experiments with normal-hearing participants aimed at clarifying this paradox, Viemeister and Wakefield suggested that the process of long-term integration does not occur and proposed a mechanism, the multiple looks hypothesis, to account for their observations. In their investigation, they tested normal-hearing listeners' detection of either a single 1-kHz, 200-μs tone pulse or two such pulses separated by 100 ms. They demonstrated that listeners' detection thresholds for two pulses were about 2.5 dB lower than their thresholds for a single pulse. Further, thresholds for the pulse pairs were unaffected by intervening noise presented over a 12-dB range in level during the gap between the pulse pairs. This finding demonstrated that listeners do not integrate power over the entire length of the auditory signal, because if they had, the level of the intervening noise should have influenced performance.

The multiple looks hypothesis proposes that the auditory system does not work as a long-term integrator, but rather that auditory filters are sampled at a relatively fast rate and that "these samples or 'looks' are stored in memory and can be accessed and processed selectively" (Viemeister & Wakefield, 1991, p. 864). The longer the duration of the signal, the more looks available to the listener and the increased probability that at least one look will be above detection threshold, resulting in improved performance with longer signal durations. Each look is on the order of 3 to 5 ms in duration. For stimulus separations greater than this time window, the looks at each stimulus are independent and information from the looks is combined. Specifically, the hypothesis predicts that if the looks at the signal are independent and the information is combined optimally, d' (a bias-free measure of sensitivity) for two pulses should be larger than d' for one pulse by a factor of the square root of two (or 1.4).

Recently, Moore (2003) addressed the issue of multiple looks in describing his spectro-temporal excitation pattern (STEP). Moore proposed that the internal representation of a stimulus can be thought of in terms of three dimensions: (a) an axis of center frequencies believed to correspond to the array of auditory filters, (b) an axis representing time, and (c) an axis corresponding to output magnitude. This internal representation is stored in working memory and "can be thought of as a vector of samples or 'looks' which are available for various kinds of computations and comparisons" (Moore, 2003, p. 567). Further, the looks are assumed to be treated intelligently, such that looks may be differentially weighted depending on the task. Moore suggested that forming a representation of a stimulus in working memory might result in a template to which further stimuli can be compared or, in the case of speech identification, could result in templates that form in long-term memory and, as such, are more permanent. The nature of the internal representation or template may change depending on the auditory task (whether it is detection, discrimination, or

identification) and/or the stimulus (whether it is speech or a simpler auditory event). Therefore, in a task such as discrimination of speech sounds, the template could potentially be a combination of the representation stored in long-term memory based on previous experience with similar speech stimuli and one formed by repetitive "looks" at the incoming signal formed in short-term memory. In this model, listeners are assumed to compare intelligently new looks with their internal representation formed by both recent looks and previous experience with similar stimuli in order to detect a change in the auditory signal. It is possible that in the process of forming a template, repetition of a stimulus may increase the number of opportunities to form an internal representation and facilitate the template-forming process. In this way, more presentations of a stimulus could serve to enhance the representation in working memory.

One procedure used in speech perception research and in testing speech perception clinically involves presentation of speech stimuli with multiple presentations or "looks"—the change/no-change procedure (Sussman & Carney, 1989). The procedure was adapted from the visually reinforced, repeating background paradigm used with infants by Eilers, Wilson, and Moore (1977), in which infants were required to turn their head in response to a change in a repeating speech stimulus (e.g., /ba/ changing to /da/). Rather than a head turn, the change/no-change procedure uses a motor response appropriate for the developmental age and ability of the listener in response to a change in the stimulus array, such that linguistic labels of "same" or "different" are not used, although the concepts of same and different must be understood.

The change/no-change procedure was developed as an alternative to more traditional speech perception tests that use word recognition paradigms. The procedure tests speech *discrimination* with standard and comparison stimuli during change trials (standard and comparison differ) and no-change trials (standard and comparison are the same). For example, if the standard stimulus were the nonsense syllable /sa/ and the comparison stimulus were /ʃa/, during a change trial the participant would hear, "/sa/ /sa/ /ʃa/ /ʃa/" and during a no-change trial, she/he would hear, "/sa/ /sa/ /sa/ /sa/." In the procedure, the sensory information remains constant, while more presentations of standard and comparison stimuli might provide more opportunities to create a memory trace or internal perceptual representation of the stimuli. If memory or internal perceptual representation processes are enhanced with more repetitions of standard and comparison stimuli, will performance improve in a manner consistent with the predictions of the multiple looks hypothesis? Specifically, speech discrimination performance might improve as the number of possible stimulus presentations increases. For example, the multiple looks hypothesis predicts that a listener would have better discrimination ability with four presentations of a standard followed by four presentations of a comparison stimulus than with two presentations of a standard followed by two presentations of a comparison. If the multiple looks hypothesis is applicable only to basic sensory tasks, such as auditory detection, the number of repetitions in the change/ no-change procedure for speech discrimination should not matter. However, if the multiple looks hypothesis has broader applicability (i.e., works over longer durations with more complicated stimuli presented above threshold), one might see better speech discrimination with more looks at the stimuli. In other words, the tenets of the multiple looks hypothesis may extend to other types of tasks, stimuli, and work over longer durations.

The purpose of this investigation was to determine whether speech discrimination performance, as measured in the change/no-change procedure, improves with the addition of multiple looks at contrasting stimuli. Two aims were addressed. The first was to determine if the presentation of multiple looks provides an advantage in the change/no-change procedure. The second aim was to determine what kind of multiple looks influence the procedure, that is, does the number of standard stimuli and the number of comparison stimuli influence

performance equally or is one more important than the other? Using procedures similar to those described in this experiment, pilot data from Bunnell (2000) suggest that the number of standard repetitions is more important than the number of comparison repetitions. This may be because the representation of the standard is enhanced with multiple presentations and, once the listener detects a change in the stimulus array, more presentations of the comparison stimulus are superfluous to completing the task.

We hypothesized that (a) multiple looks would result in improved discrimination performance, (b) the number of looks at the standard stimulus would be more important than the number of looks at the comparison stimulus, and (c) the advantage of multiple looks would apply to multiple speech contrasts.

Method

Participants

A total of 29 adults participated in this investigation. Fifteen participated in a preliminary stimulus-labeling portion and 14 participated in the change/ no-change portion. The listeners who participated in the labeling condition ranged in age from 18 to 47 years, with a mean of 23.9 years. Twelve were female and 3 were male. The listeners who participated in the change/no-change condition ranged in age from 19 to 37 years, with a mean of 24.5 years. Of these, 9 were female and 5 were male. Every participant passed a bilateral hearing screening at 20 dB HL (re: ANSI, 1989) at 0.5, 1, 2, 4, and 8 kHz and 25 dB HL at 0.25 kHz and was a native English speaker.

Stimuli

Three pairs of nonsense consonant–vowel (CV) stimuli were used: /pa/ versus /ta/, /ra/ versus /la/, and /sa/ versus /ʃa/. The fricatives, /s/ and /ʃ/, were chosen as stimuli for several reasons. First, Nittrouer and colleagues have used these fricatives extensively in their research on the developmental weighting shift (Nittrouer, 1992, 1996a, 1996b; Nittrouer, Manning, & Meyer, 1993; Nittrouer & Miller, 1997). These investigators have demonstrated that these fricatives, in particular, are sensitive to developmental changes in speech perception. This finding was important to our investigation, because data obtained from adults in this experiment will be used as a comparison for future experiments with children. Second, the important acoustic difference between /s/ and /ʃ/ results from the place in the oral cavity in which each are articulated. Place of articulation errors are common in recognition and discrimination of CV syllables, even in those with normal hearing (Miller & Nicely, 1955). The other stimulus pairs were also chosen because the stimuli within each pair vary in place of articulation; each pair tests an additional manner of production (stops and semivowels). Bilger and Wang (1976) have shown that listeners with hearing loss have difficulty distinguishing place of articulation and manner of articulation. In addition to using these stimuli with children with normal hearing, these tokens will be used in future investigations with individuals with hearing loss, so it was important to have baseline data from adults with normal hearing on these tokens. Finally, to assess speech discrimination, we chose to use nonsense syllables rather than words, because we did not want lexical knowledge to influence the results.

Synthesized exemplars were used for four of the six syllables and for the vowels of the remaining two stimuli. Synthesized stimuli are preferred because the acoustic difference between minimal pairs can be controlled more precisely than with natural productions. Although this was not necessary for testing multiple looks per se, having control over the stimuli is important for our series of experiments that will use the change/no-change procedure for studying the development of speech perception in normal-hearing and

hearing-impaired populations. The four syllables that were synthesized were /pa/, /ta/, /ra/, and /la/. The syllables /sa/ and /ʃa/ were created by splicing a synthetic /a/ to digitized natural tokens of /s/ and /ʃ/. This was done because it was very difficult to synthesize natural-sounding fricative-vowel stimuli. The fricative-vowel stimuli were recorded from an adult male in a sound-treated room with a Shure 809 microphone. The syllables were digitized with the Computerized Speech Research Environment 4.5 (CSRE) software (AVAAZ Innovations, Inc., 1995) at a 20000 Hz sampling rate and low-pass filtered at 10000 Hz. Only the fricative portions of the digitized natural stimuli were used. The synthetic speech syllables were created using the KLSYN speech synthesis program (Klatt, 1980) and a Pentium III laboratory computer. Within each stimulus pair, the vowel was identical and the only sound that differed was the consonant. Formant frequency values for /a/ were based on Hillenbrand, Getty, Clark, and Wheeler (1995) and formant frequency, bandwidth, and amplitude values for /r/, /l/, /p/, and /t/ from Klatt were used as guidelines for creating the stimuli, because they have been used successfully to create synthetic speech. Waveforms, energy spectra, and frequency spectra for each CV stimulus are displayed in the Appendix and were created using Specto 2.41 (Boys Town National Research Hospital, 2002).

/pa/ Versus /ta/—The 498-ms duration voiceless stop CV stimuli, /pa/ and /ta/, consisted of a burst (20 ms), followed by aspiration (60 ms) and the vowel (418 ms). The primary acoustic differences between /pa/ and /ta/ during their respective bursts are the lower F_2 and F_3 onsets for /pa/ relative to /ta/ and the resulting increasing formant transitions for /pa/ and the decreasing transitions for /ta/. During the aspiration, the formants for /pa/ increase during the transition to the steady-state values for the vowel /a/, while they continue to decrease for /ta/.

/ra/ Versus /la/—The 498-ms duration semivowel CV stimuli, /ra/ and /la/, consisted of the semivowel (150 ms), /r/ or /l/, followed by the vowel (348 ms). The primary acoustic differences between /ra/ and /la/ are that /r/ has a much lower F_3 (near F_2) relative to /l/ and that the F_1 transition tends to be faster for /l/ than for /r/. This results in a longer early steady state for /l/ relative to /r/.

/sa/ Versus /ʃa/—The fricatives used were those CV tokens recorded from an adult male. Although the naturally produced vowel was not used, the formant frequency values from the speaker for the vowel /a/ were used to create the synthetic /a/ in the KLSYN software, rather than using Hillenbrand et al.'s (1995) values. This helped make the syllables sound more natural. Using Cool Edit Pro (Syntrillium Software Corp., 1997), the digitized natural fricatives, /s/ and /ʃ/, were appended to the synthetically created /a/. The total length of each syllable was 533 ms. The vowel was 412 ms in duration, leaving the fricatives 121 ms in length. The total root-mean-square (RMS) power of these six nonsense syllables was equalized.

Assessing the Intelligibility of Each Stimulus—Prior to testing with the change/no-change procedure, normal-hearing adults listened to the six CV stimuli and labeled each as either the stimulus presented or something other than the stimulus presented. This was done to ensure that the stimuli were readily identified as the intended targets, because Logan, Greene, and Pisoni (1989) pointed out that, based on intelligibility scores, not all synthetic speech is perceptually equivalent to all listeners. Participants listened to a total of 100 presentations of each stimulus at 68 dB SPL in quiet in a double-walled sound booth. Within each of the six stimulus conditions, participants listened to 50 presentations of each stimulus in a stimulus pair. Their task was to label each token heard as either the intended stimulus or something other than that stimulus by touching the screen of a touch monitor. This

procedure was based on one used by Eilers, Ozdamar, Oller, Miskiel, and Urbano (1988). Half the screen was labeled one of the following, "PA," "TA," "RA," "LA," "SA," or "SHA," while the other half was labeled, "Not-PA," "Not-TA," "Not-RA," "Not-LA," "Not-SA," or "Not-SHA," respectively. For example, in the condition in which participants identified /ra/, each listened to 50 presentations of /ra/ and 50 of /la/. After each presentation, the participant was to touch the side of the screen labeled "RA" if she/he perceived /ra/ and to touch the side of the screen labeled "Not-RA" if she/he perceived something other than /ra/. The stimuli were presented in a random order and the side of the touch monitor on which the labels appeared switched randomly to keep participants' attention as high as possible.

Within each condition, half of the presentations were not scored, specifically those presentations that corresponded to the "Not-Syllable" option. For example, in the "RA" versus "Not-RA" condition discussed earlier, only the responses to presentations of /ra/ were scored. This procedure was followed because the "Not-RA" option does not necessarily mean the listener heard /la/; it simply indicates she/he did not perceive /ra/. Therefore, although "Not-RA" responses to /la/ were technically correct, they were not used to calculate the percentage correct identification for /la/. Only responses to /la/ stimuli in the "LA" versus "Not-LA" condition could be used for that purpose.

The mean percentage correct identification (and standard deviation in parentheses) for /pa/, /ta/, /ra/, /la/, /sa/, and /ʃa/ were 99 (1.5), 98 (6.2), 100 (0.7), 99 (1.8), 94 (7.1), and 98 (4.1), respectively, indicating that the stimuli were sufficiently good exemplars of the intended targets.

Creation of the Speech in Noise Stimuli—Adults with normal hearing would all score near ceiling if tested in quiet on nonsense speech contrasts in their native language, regardless of the number of stimulus repetitions. Therefore, we tested performance in different levels of background noise. To create this noise, the long-term spectrum for each of the three minimal pair stimuli was used to shape white noise in Cool Edit Pro (Syntrillium Software Corp., 1997), resulting in three syllable-shaped masking noises. For example, the spectrum for /ra–la/ was used to shape white noise and this noise was used as masking noise for the /ra/ versus /la/ comparison conditions. Each syllable-shaped noise was 10 s long and was then amplified relative to the overall RMS of the speech to achieve the desired signal-to-noise ratios (SNRs) used in the investigation. A different portion of the 10 s syllable-shaped noises was selected for each condition when mixing the noise with the speech stimuli.

The repetition conditions tested were 4:4, 4:2, 4:1, 2:4, 2:2, 2:1, 1:4, 1:2, and 1:1, where the number preceding the colon indicates the number of standard stimulus repetitions and the number following it represents the number of comparison stimulus repetitions. The interstimulus interval within each trial was 100 ms. Each repetition pairing was then mixed with the syllable-shaped noise corresponding to the syllable pair under test to achieve the SNRs used in the investigation, −10, −8, −6, and −4 dB. In two pilot experiments (one of which was Bunnell (2000)), we found that performance asymptoted for /sa/ versus /ʃa/ at about −4 dB SNR. Therefore, we selected SNRs between and including −10 and −4 dB, because these span the range over which performance changed with SNR in the pilot experiments. This allowed us to create detailed performance-intensity functions for each listener.

Equipment

E-Prime software 1.1 (Psychology Software Tools, 2002) on a Pentium III computer was used to run both the identification task and the change/no-change procedure and record

responses from the participants. The signal from the computer's sound card was routed to a Crown D-75 amplifier and then to two GSI speakers placed at +45 and −45 degrees azimuth relative to the listener. The listener was seated in a double-walled sound booth. Responses were made on a touch screen monitor placed directly in front of the listener. The overall level of the speech was 68 dB SPL at the location of the listener's head, with the level of the noise varying with SNR. Calibration was checked each day of testing.

We elected to test in the sound field because this procedure will be used with children and individuals with hearing loss in future studies. Asking young children to wear insert earphones for sessions as long as those described in our procedures is unrealistic. Therefore, we wanted the same variability due to head position as possible in the adult data as in the future child data. To reduce level effects, calibration was checked at the level of the listener's head and each listener was seated such that her/his head was situated in the calibrated position. Further, because this was a suprathreshold discrimination task, small level variations at the listener's ear due to head movement were not likely to have a significant influence on performance.

Procedure

Participants were tested using the change/no-change procedure in a repeated measures design. Performance was measured for nine repetition conditions (4:4, 4:2, 4:1, 2:4, 2:2, 2:1, 1:4, 1:2, and 1:1), with three synthetic nonsense CV syllable comparisons (/pa/ vs. /ta/, /ra/ vs. /la/, and /sa/ vs. /ʃa/), at four SNRs (−10, −8, −6, and −4 dB). This resulted in a total of 108 conditions, which were pseudorandomly presented to each listener. The order in which participants received the syllable comparisons was randomized. Within each syllable comparison, the order of SNR was randomized and within each SNR, the order of repetition condition was randomized. Within each syllable comparison, half of the listeners heard one of the syllables as the standard stimulus, while the other half heard the other as the standard. Each condition consisted of 50 trials: 25 change trials and 25 no-change trials. Participants completed all 108 conditions in either six or seven visits of approximately 2 hr each and were paid for their time.

At the beginning of each visit, participants listened to ten 2:2 trials in quiet to familiarize them with the stimuli and to ensure they understood the task. Five trials were change trials and five were no-change trials. The syllables used as the standard and the comparison for the familiarization task were the same as those assigned to that particular participant for the actual testing. For example, if a participant were being tested on the /ra/ versus /la/ condition that particular day and she/he were randomly assigned to have /la/ as her/his standard, the practice trials would be presented with /la/ as the standard and /ra/ as the comparison.

Participants were instructed to listen to a string of syllables in which they would hear a specified syllable followed by a different specified syllable or by the same specified syllable. They were informed of how many repetitions they would hear of both the standard and the comparison stimuli in order to reduce any auditory uncertainty. Listeners were told to touch one side of the touch monitor labeled "Change" if they perceived a change in the string of syllables; the side of the touch monitor labeled "No-Change" was to be touched if they did not perceive a change in the string. These labels did not switch sides of the touch monitor screen as they did in the identification portion of the study. No feedback was given as to response correctness. Participants were given a break after approximately 30–40 min of listening, or sooner if one was requested.

Results

Performance was measured in d' because it is bias-free and the multiple looks hypothesis' prediction of an increase in performance with longer stimulus durations is made in d'. This measure is calculated by subtracting the z-score for a listener's false alarm rate from the z-score for her/his hit rate. Some participants had a hit rate of 1.0 and a false alarm rate of 0.0 in certain conditions. Other listeners had 25 out of 25 hits and some false alarms or 25 out of 25 correct rejections with many hits. These results require corrections in the calculation of d'. In this study, a value of 0.01 was added to all 0.0 hit and false alarm rates and a value of 0.01 was subtracted from all 1.0 hit and false alarm rates. Thus, for perfect performance, the maximum d' in this study is 4.65. Psychometric functions for each syllable pair contrast were determined for every participant. Figures 1, 2, and 3 display mean group performance by SNR for each syllable comparison. Figure 1 displays data for the /pa/ versus /ta/ comparison, Figure 2 displays those for the /ra/ versus /la/ comparison, and Figure 3 shows data for the /sa/ versus /ʃa/ comparison. The top, middle, and bottom panels in each figure display performance for conditions in which the standard number of repetitions is one, two, and four, respectively. The open squares, open triangles, and filled circles each represent a standard number of one, two, and four repetitions, respectively, and the dashed, connected, and dotted lines indicate a comparison number of one, two, and four repetitions, respectively.

In general, the best discrimination performance occurred with higher numbers of stimulus repetitions (e.g., 4:4, 2:4, and 4:2) for all speech contrasts. In both the /ra/ versus /la/ and /pa/ versus /ta/ conditions, performance was poorest in the 1:1 condition at presentation levels above the noise floor (above about –8 dB SNR). For /sa/ versus /ʃa/, the repetition condition that resulted in the poorest performance above the noise floor (approximately –6 dB SNR) was 2:1, followed by 1:1.

Discrimination performance also varied across syllable pair contrast. Mean performance for the /pa/ versus /ta/ comparison was lower across repetition conditions (except for 4:4) than for the other two pairs of contrasts and, in general, mean performance was lower for the /ra/ versus /la/ contrast than for the /sa/ versus /ʃa/ contrast. Table 1 shows mean group performance for each syllable contrast collapsed across repetition conditions at each SNR (–10, –8, –6, and –4 dB). This table shows that at a given SNR, average performance was best for the fricative pair, followed by the semivowel pair, and finally, the poorest performance was for the stop-consonant pair. For example, mean performance across repetition conditions in d' for /sa/ versus /ʃa/ at –8 dB SNR was 1.54, while for /ra/ versus /la/ and /pa/ versus /ta/ it was 1.03 and 0.45, respectively. To attain a mean d' of at least 1.54 in the /ra/ versus /la/ condition, participants required an SNR of nearly –6 dB; for the /pa/ versus /ta/ condition, participants required an even better SNR, nearly –4 dB. By –4 dB SNR, every participant in the /sa/ versus /ʃa/ condition was performing near ceiling levels, whereas for /pa/ versus /ta/ there was always at least 1 participant performing at or just above chance even at this easiest SNR. For /ra/ versus /la/, the only condition in which at least 1 participant performed near chance at –4 dB SNR was in the lowest repetition number condition, 1:1. These findings lend further support to the observation that more repetitions of a stimulus lead to enhanced discrimination, because even at the easiest SNRs, some listeners required more than a single repetition of specific contrasts to perform above chance.

There was also an interaction between syllable pair and SNR; specifically, the slopes of the psycho-metric functions across syllable pairs varied. The slopes for the /pa/ versus /ta/ contrast were shallower than those for the /ra/ versus /la/ contrast. In turn, the slopes for the /ra/ versus /la/ contrast were shallower than those for the /sa/ versus /ʃa/ contrast.

Finally, the variability across participants was less at the lowest and the highest SNRs than at the midlevel SNRs, which was likely due to ceiling and floor effects, although this was less true for /pa/ versus /ta/ than for the other pair of contrasts. Furthermore, at a given midlevel SNR, variability tended to decrease as the number of repetitions increased, especially for the /sa/ versus /ʃa/ and /ra/ versus /la/ contrasts.

The data were entered into a four-way analysis of variance (ANOVA) with repeated measures (manner of articulation/syllable pair [stop consonant, semivowel, fricative] × standard repetition number [one, two, four] × comparison repetition number [one, two, four] × SNR [−10, −8, −6, −4]). Table 2 displays the results of the ANOVA. The first section displays the significant effects (p .05) and the second section shows the nonsignificant effects ($p > .05$). All the main effects (syllable pair, standard repetition number, comparison repetition number, and SNR) significantly influenced performance, indicating that the differences in performance observed across standard and comparison stimulus-repetition conditions and syllable-pair contrasts in Figures 1, 2 and 3 were statistically significant. Further, a number of interactions were significant, as well (p .05). First, the interaction between syllable-pair contrast and SNR was significant, indicating that the slopes of psychometric functions were indeed significantly different across syllable-pair contrast. Second, although both the number of standard and the number of comparison stimuli significantly influenced discrimination performance, there was an interaction between the number of standards and the number of comparisons. This finding is explored further in the following section on their relative contribution to speech discrimination. Finally, there were interactions between the number of standard stimuli, the number of comparison stimuli, SNR, and syllable pair, suggesting that with more advantageous listening conditions (e.g., easier SNRs), listeners require fewer presentations of the stimuli to successfully discriminate them and that the changes in performance vary across syllable pair.

To examine further the effects of the number of repetitions, we performed post hoc analyses. The data set was reduced to contain only those variables in which both the number of standard and comparison repetitions was the same across the change (i.e., it included the 4:4, 2:2, and 1:1 conditions for each syllable-pair contrast at all four SNRs). This data reduction allowed us to examine the differences between one, two, and four repetitions while maintaining a manageable number of comparisons during post hoc hypothesis testing. Again, a three-way ANOVA with repeated measures (manner of articulation/syllable pair [stop consonant, semivowel, fricative] × repetition number [one, two, four] × SNR [−10, −8, −6, −4]) revealed that all three main effects of manner, repetition number, and SNR significantly influenced performance, $F(2, 26) = 19.838, p < .001$; $F(2, 26) = 23.327, p < .001$; and $F(3, 39) = 221.417, p < .001$, respectively. Hypothesis testing on the different numbers of repetitions revealed a significant difference in performance between one repetition and two repetitions, $F(1, 13) = 26.855, p < .001$, even under a conservative Bonferroni adjustment, but no significant difference between two and four repetitions. This suggests that the significant differences in performance for repetition number are primarily due to the improvement in discrimination with doubling the number of repetitions of stimuli from one to two, rather than from two to four.

The Relative Importance of Standard and Comparison Stimulus Repetitions

To further analyze the effects of multiple looks on performance in the change/no-change procedure, we examined performance at a specific SNR for which performance was above the noise floor, yet had not reached ceiling levels. This occurred at −6 dB SNR for /pa/ versus /ta/ and /ra/ versus /la/ and at −8 dB SNR for /sa/ versus /ʃa/. To determine the relative effects of the number of standard and comparison stimuli, we examined the effects of doubling the number of standard repetitions separately from doubling the number of comparisons. Using the mean d' data, we predicted results that would be expected based on

Viemeister and Wakefield's (1991) multiple looks hypothesis in each condition. For example, we analyzed the effects of doubling the number of standards by multiplying the mean d' data from the 1:1, 1:2, and 1:4 conditions each by a factor of 1.4 to predict performance in the 2:1, 2:2, and 2:4 conditions, respectively. In turn, we multiplied the mean d' data from the 2:1, 2:2, and 2:4 conditions each by a factor of 1.4 to predict performance at 4:1, 4:2, and 4:4, respectively. The same procedure was carried out to look at the effects of doubling the number of comparisons. For these predictions, however, the mean d' data for the 1:1, 2:1, and 4:1 conditions were multiplied by a factor of 1.4 to predict performance in the 1:2, 2:2, and 4:2 conditions, respectively. In turn, we multiplied the mean d' data from the 1:2, 2:2, and 4:2 conditions by a factor of 1.4 to predict performance at 1:4, 2:4, and 4:4, respectively. We recognize that there are a number of ways to go about predicting performance based on the multiple looks hypothesis. For example, performance for the 4:1 condition could be predicted from actual performance at 2:1 or predicted performance at 2:1 (which was initially predicted from actual performance in the 1:1 condition). We elected to predict performance from actual data based on a doubling of the number of standard and the number of comparison stimulus repetitions. With no real precedent for how to carry out this procedure, we reasoned that if there were any idiosyncrasies in performance in the 1:1 condition, they would not be carried through all the subsequent predictions using this method. Also, variability in performance decreased at midlevel SNRs with increased repetition numbers. Therefore, we concluded that predicting performance from less variable data was more desirable than predicting all the data from more variable conditions (such as the 1:1 conditions).

Figure 4 displays the effects of doubling the number of standard repetitions while keeping the number of comparisons constant. The actual mean d' values (filled circles connected by lines) and the predicted values (open circles) at each repetition condition are grouped by the number of comparison stimulus repetitions. Note that there are no data points for the predicted mean d' at 1:1, 1:2, and 1:4, because these are the data points from which the predictions for 2:1, 2:2, and 2:4 were made. The top, middle, and bottom panels display data for the stop consonant, semivowel, and fricative pairs, respectively. Each series of three data points along the abscissa represent a serial doubling of the number of standard repetitions, while keeping the number of comparison repetitions constant. Figure 5 displays similar data; however, the results are grouped by the number of standard stimulus repetitions, such that the effect of doubling the number of comparisons is shown while keeping the number of standards constant. The difference between the closed and the open symbols in Figures 4 and 5 is the amount by which the data do not fit the multiple looks hypothesis' prediction. The numerical values on the figures (that look like data labels) are actually the ratio between predicted and actual performance. We added this additional method of examining the difference between predicted and actual performance because d' does not reflect performance on a linear scale. For example, a difference in d' of 0.5 reflects a larger difference in performance around values of 1.0 than it does around values of 3.0. By including the ratio of predicted to actual performance, we better address this issue. Figure 6 displays the mean ratio of predicted to actual performance at the repetition conditions in which performance was predicted. The top panel displays the effects of a doubling of the number of comparison stimuli and the bottom panel displays the effects of a doubling of the number of standard stimuli. The filled circles represent the /pa/ verusu /ta/ comparison at −6 dB SNR, the /ra/ versus /la/ comparison at −6 dB SNR is represented by "X"s, and the open squares represent the /sa/ versus /ʃa/ comparison at −8 dB SNR. If performance were predicted solely by the multiple looks hypothesis, all the data points would fall along the horizontal line at a mean ratio of 1.0. The data do not perfectly fit the prediction. However, for low repetition numbers (e.g., 1:2, 2:1, 1:4, 4:1, and 2:2), actual performance is only slightly different from the predicted performance for all the syllable comparisons. For higher repetition numbers, performance generally falls below that predicted by the multiple looks

hypothesis (except for some speech contrasts in the 4:4 condition). For lower repetition numbers, the number of standard and comparison repetitions seems to be equally important. For higher repetition numbers, doubling the number of comparison repetitions better predicts performance based on the multiple looks hypothesis than does doubling the number of standard repetitions (e.g., 4:2, 2:4, and 4:4 in Figure 6). This finding is consistent with the significant interaction between the number of standard and the number of comparison repetitions found in the ANOVA.

When similar comparisons are made at other locations along the psychometric functions (such as at more advantageous SNRs), an interesting observation emerges. The absolute difference between actual performance and that which is predicted by the multiple looks hypothesis is greater at better SNRs than at poorer SNRs. However, the ratio between predicted and actual performance at better SNRs is similar to the ratio observed at poorer SNRs. Figures 7 and 8 display the same information as Figures 4 and 5, predicted and actual performance for each syllable pair, grouped by number of comparisons and number of standards, respectively. The difference between these sets of figures is that Figures 7 and 8 display predicted and actual performance at SNRs 2 dB higher than those in Figures 4 and 5, or equivalently, at higher locations along the psychometric functions. The difference between actual and predicted performance is greater in Figures 7 and 8 than in Figures 4 and 5. However, similar to Figure 6, Figure 9 displays the mean ratio of predicted to actual performance, but at these more advantageous SNRs. The top panel displays the effects of a doubling of the number of comparison stimuli and the bottom panel displays the effects of a doubling of the number of standard stimuli. Although the absolute difference in predicted and actual performance is quite large at higher SNRs, the ratio between actual and predicted performance is similar to those at lower SNRs. This finding suggests that the multiple looks hypothesis' prediction of an increase in d' by a factor of 1.4 for every doubling in looks at the signal holds at lower stimulus repetition numbers and at d' values around 1.0 and extends to this speech discrimination task. Once participants were performing above d' of about 1.0, the multiple looks hypothesis did not predict absolute discrimination performance well in this task regardless of whether the number of comparison or number of standard repetitions doubled. However, the ratio of predicted to actual performance in d' was similar across performance levels, which suggests that once performance reaches a high enough level, increasing the number of repetitions of the stimuli does not improve discrimination by a full factor of 1.4, but that the relative degree of improvement is similar across performance levels.

Discussion

The results suggest that the addition of more looks at the stimuli in the change/no-change procedure improves performance in a discrimination task. Although the strict prediction of an increase in d' by a factor of 1.4 for a doubling in the number of looks was not supported, at least for high repetition numbers, there was a trend for improvement with more looks at the stimuli. Both "types" of looks—number of standards and number of comparisons— significantly influenced discrimination performance. For lower repetition numbers, the number of standard and comparison repetitions seemed to be equally important. For higher repetition numbers, the number of comparisons better predicted performance based on our extension of the multiple looks hypothesis than did the number of standard repetitions. This result is somewhat different than what we hypothesized based on pilot data indicating that the number of standard repetitions would be more important than the number of comparisons. One potential explanation for this finding is that listeners must keep a large amount of auditory information in memory during the higher repetition number conditions and the larger memory load might force them to rely more on the percept formed from the most recently heard stimuli (the comparisons) than on the stimuli heard first (the standards).

Post hoc analyses indicated that discrimination performance significantly improved with more looks at the stimuli for lower repetition numbers (i.e., increasing from one repetition to two repetitions), but not for higher repetition numbers (i.e., increasing from two repetitions to four repetitions), lending further support to the observation that performance did not increase by as much as would be expected by the multiple looks hypothesis at higher repetition numbers.

These results are similar to those found by Viemeister and Wakefield (1991) in their investigation in which they proposed the multiple looks hypothesis. In their investigation, they compared the threshold for detection (near a d' of 1.0) for one pulse versus two pulses. In terms of the current investigation, they used low repetition numbers at threshold in their detection task. When similar conditions were used in the change/ no-change discrimination procedure even with longer stimuli presented at suprathreshold levels, the predictions made by the multiple looks hypothesis were supported.

The effect of increasing the number of stimulus presentations applies to multiple speech contrasts. There was a significant effect for manner of articulation, suggesting that performance differs across syllable pair contrasts. Inspection of Figures 1 through 3 supports this finding, as do the data displayed in Table 1: Performance was higher at lower SNRs for the fricative pair /sa/ versus /ʃa/ than for the semivowel or stop consonant pairs, with the stop consonant pair showing the poorest performance at any given SNR. Additionally, the shapes of the psychometric functions for the different manners of articulation were significantly different, indicating that not only did performance across manner of articulation differ, but so did the interaction between manner and SNR. This was not necessarily unpredicted. Work by Boothroyd, Erickson, and Medwetsky (1994) on the audibility of individual consonants suggested this finding based on the audibility of the six stimuli used in this investigation. Finally, the advantage of multiple presentations of stimuli was greater at more difficult SNRs than at easier SNRs.

The present experiment demonstrates that the multiple looks hypothesis not only applies to detection tasks, but also applies to a discrimination procedure, at least for lower numbers of stimulus repetitions at midlevel performance. Further, the results extend the applicability of the hypothesis to longer stimuli consisting of CV combinations presented at suprathres-hold levels. Multiple looks at the stimuli improved stimulus discriminability, especially under more adverse listening conditions (e.g., SNRs between −8 and −6 dB). The results do not necessarily suggest that the neural mechanism behind the multiple looks hypothesis is at work in this task, but they do suggest that similar to Viemeister and Wakefield's (1991) investigation using short duration stimuli, performance on this speech discrimination task is enhanced with more opportunities to hear the stimuli. These results also demonstrate that speech sound discrimination is more than a solely sensory and speech acoustics phenomenon. Speech perception, even in a discrimination task in which no vocabulary or word identification ability is tapped, involves more than a compilation of acoustic features that lead to a percept. Repetition results in more opportunities to form an internal perceptual representation and leads to increased discriminability for this group of normally hearing adults. It is possible that in more difficult information processing tasks where there is more auditory uncertainty, enhancement of the perceptual representation may be even more important than in the current procedure where only two sounds are contrasted and their context is fixed. These findings have implications for models of speech perception, because a complete model will need to take into account, among other things, the strength of the internal perceptual representation. In particular, they would need to account for the variable internal representation of speech under degraded conditions, such as in background noise or when a listener has hearing loss.

These results also have practical implications regarding the use of the change/no-change procedure. The procedure has been used in previous experiments (e.g., Carney et al., 1993; Sussman & Carney, 1989) with the assumption that the number of stimulus presentations does not influence performance. Our results suggest otherwise and raise the question of what the optimal number of overall presentations is. This investigation does not address this question directly, but does suggest that little is gained from increasing the number of repetitions from two to four, at least for these contrasts. Furthermore, although the change/no-change procedure can be used with adults, it is primarily intended for use with children. Children's perception of speech is developing through at least age 10 years (Elliott, 1986; Elliott, Longinotti, Meyer, Raz, & Zucker, 1981; Sussman & Carney, 1989) and possibly through the teenage years (Johnson, 2000). Therefore, it remains to be seen whether this enhanced speech discrimination with multiple presentations of stimuli experienced by adults will benefit children in the same way.

Acknowledgments

References

American National Standards Institute. Specification for audiometers (ANSI S3.6-1989). New York: Author; 1989.

AVAAZ Innovations, Inc. Computerized Speech Research Environment 4.5 (Computer software). Ontario, Canada: Author; 1995.

Bilger RC, Wang MD. Consonant confusions in patients with sensorineural hearing loss. Journal of Speech and Hearing Research. 1976; 19:718–748. [PubMed: 1003952]

de Boer, E. Auditory time constants: A paradox?. In: Michelsen, A., editor. Time resolution in auditory systems. Berlin, Germany: Springer-Verlag; 1975. p. 141-158.

Boothroyd A, Erickson FN, Medwetsky L. The hearing aid input: A phonemic approach to assessing the spectral distribution of speech. Ear and Hearing. 1994; 15:432–442. [PubMed: 7895939]

Boys Town National Research Hospital. Specto 2.41 (Computer software). Omaha, NE: Author; 2002.

Bunnell, S. Unpublished master's thesis. University of Minnesota; Minneapolis: 2000. The effects of multiple looks in a discrimination task.

Carney AE, Osberger MJ, Carney E, Robbins AM, Renshaw J, Miyamoto RT. A comparison of speech discrimination with cochlear implants and tactile aids. Journal of Acoustical Society of America. 1993; 94:2036–2049.

Eilers RE, Ozdamar O, Oller DK, Miskiel E, Urbano R. Similarities between tactual and auditory speech perception. Journal of Speech and Hearing Research. 1988; 31:124–131. [PubMed: 3352249]

Eilers RE, Wilson WR, Moore JM. Developmental changes in speech discrimination in infants. Journal of Speech and Hearing Research. 1977; 20:766–780. [PubMed: 604689]

Elliott LL. Discrimination and response bias for CV syllables differing in voice onset time among children and adults. Journal of the Acoustical Society of America. 1986; 80:1250–1255. [PubMed: 3771932]

Elliott LL, Longinotti C, Meyer D, Raz I, Zucker K. Developmental differences in identifying and discriminating CV syllables. Journal of the Acoustical Society of America. 1981; 70:669–677. [PubMed: 7288030]

Faulkner A, Rosen S, Moore BD. Residual frequency selectivity in the profoundly hearing-impaired listener. British Journal of Audiology. 1990; 24:381–392. [PubMed: 2279196]

Gerken GM, Bhat VKH, Hutchinson-Clutter MH. Auditory temporal integration and the power-function model. Journal of the Acoustical Society of America. 1990; 88:767–778. [PubMed: 2212302]

Green, DM. Temporal factors in psychoacoustics. In: Michelsen, A., editor. Time resolution in auditory systems. Berlin, Germany: Springer-Verlag; 1985. p. 122-140.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97:3099–3111. [PubMed: 7759650]

Johnson CE. Children's phoneme identification in reverberation and noise. Journal of Speech, Language, and Hearing Research. 2000; 43:144–157.

Klatt DH. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America. 1980; 67:971–995.

Logan JS, Greene BG, Pisoni DB. Segmental intelligibility of synthetic speech produced by rule. Journal of the Acoustical Society of America. 1989; 86:566–581. [PubMed: 2527884]

Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America. 1955; 27:338–352.

Moore BCJ. Temporal integration and context effects in hearing. Journal of Phonetics. 2003; 31:563–574.

Nittrouer S. Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. Journal of Phonetics. 1992; 20:1–32.

Nittrouer S. Discriminability and perceptual weighting of some perceptual cues to speech perception by 3-year-olds. Journal of Speech and Hearing Research. 1996a; 39:278–287. [PubMed: 8729917]

Nittrouer S. The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM. Journal of Speech and Hearing Research. 1996b; 39:1059–1070. [PubMed: 8898258]

Nittrouer S, Manning D, Meyer G. The perceptual weighting of acoustic changes with linguistic experience. Journal of the Acoustical Society of America. 1993; 94:S1865.

Nittrouer S, Miller ME. Predicting developmental shifts in perceptual weighting schemes. Journal of the Acoustical Society of America. 1997; 101:2253–2266. [PubMed: 9104027]

Psychology Software Tools. E-Prime 1.1 [Computer software]. Pittsburgh, PA: Author; 2002.

Scharf, B. Loudness. In: Carterette, EC.; Friedman, MP., editors. Handbook of perception. New York: Academic Press; 1978. p. 187-242.

Skinner MW. Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation. Journal of the Acoustical Society of America. 1980; 67:306–317. [PubMed: 7354198]

Stelmachowicz PG, Jesteadt W, Gorga MP, Mott J. Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. Journal of the Acoustical Society of America. 1985; 77:620–627. [PubMed: 3973233]

Stelmachowicz PG, Pittman AL, Hoover BM, Lewis DE. Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults. Journal of the Acoustical Society of America. 2001; 110:2183–2190. [PubMed: 11681394]

Sussman JE, Carney AE. Effects of transition length on the perception of stop consonants by children and adults. Journal of Speech and Hearing Research. 1989; 36:380–395. [PubMed: 8487530]

Syntrillium Software Corp. Cool Edit Pro 1.1 [Computer software]. Phoenix, AZ: Author; 1997.

Vickers DA, Moore BCJ, Baer T. Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. Journal of the Acoustical Society of America. 2001; 110:1164–1175. [PubMed: 11519583]

Viemeister NF, Wakefield GH. Temporal integration and multiple looks. Journal of the Acoustical Society of America. 1991; 90:858–865. [PubMed: 1939890]
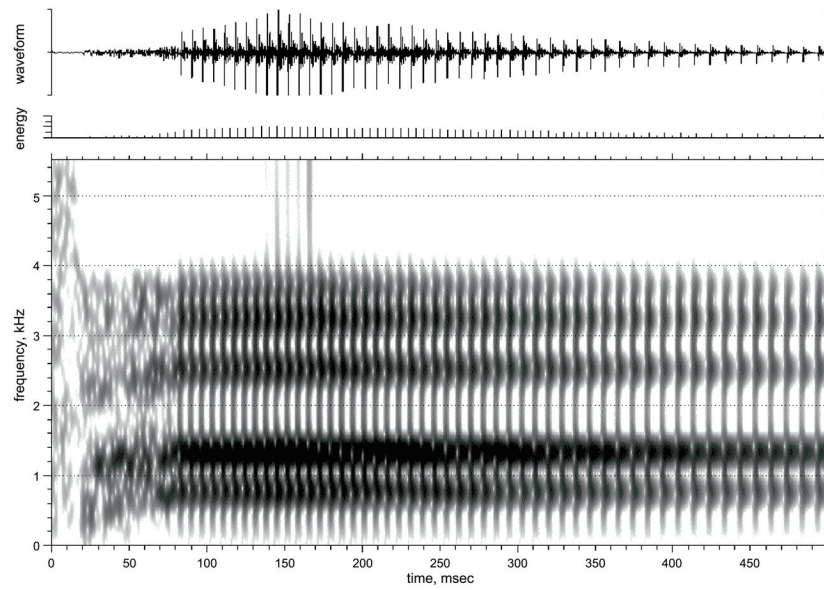
Appendix

Figure A1.
Waveform (top panel), energy spectrum (middle panel), and frequency spectrum (bottom panel) of /pa/.
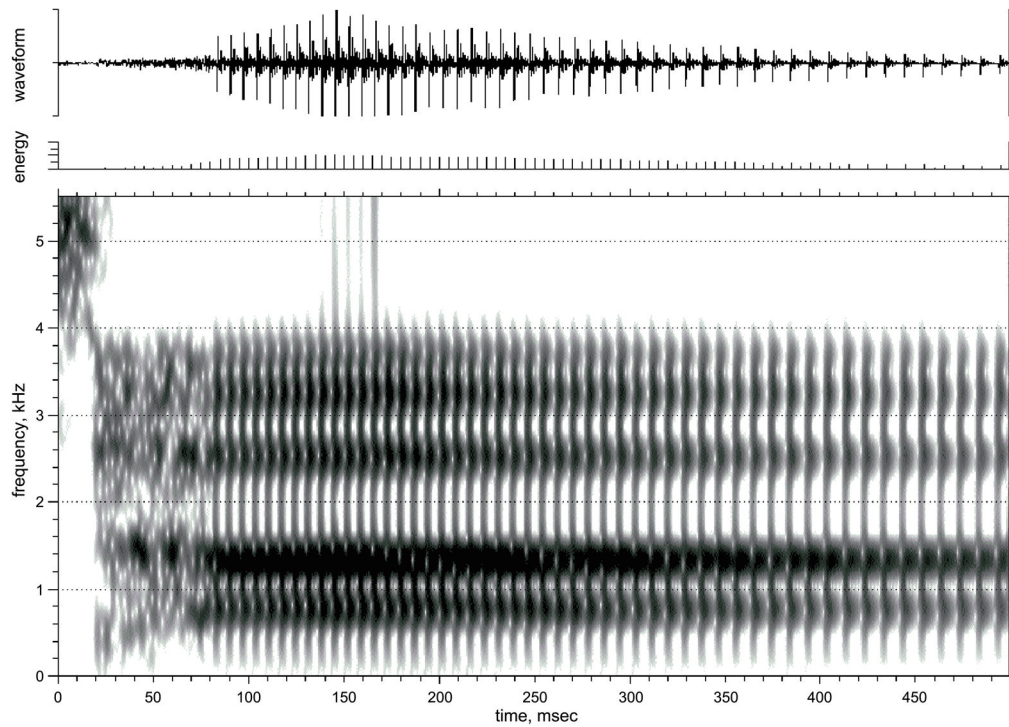

Figure A2.
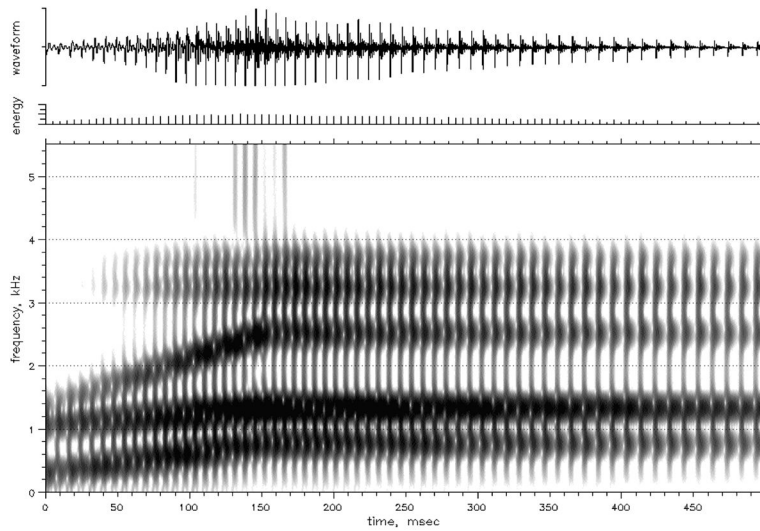Waveform (top panel), energy spectrum (middle panel), and frequency spectrum (bottom panel) of /ta/.

Figure A3.
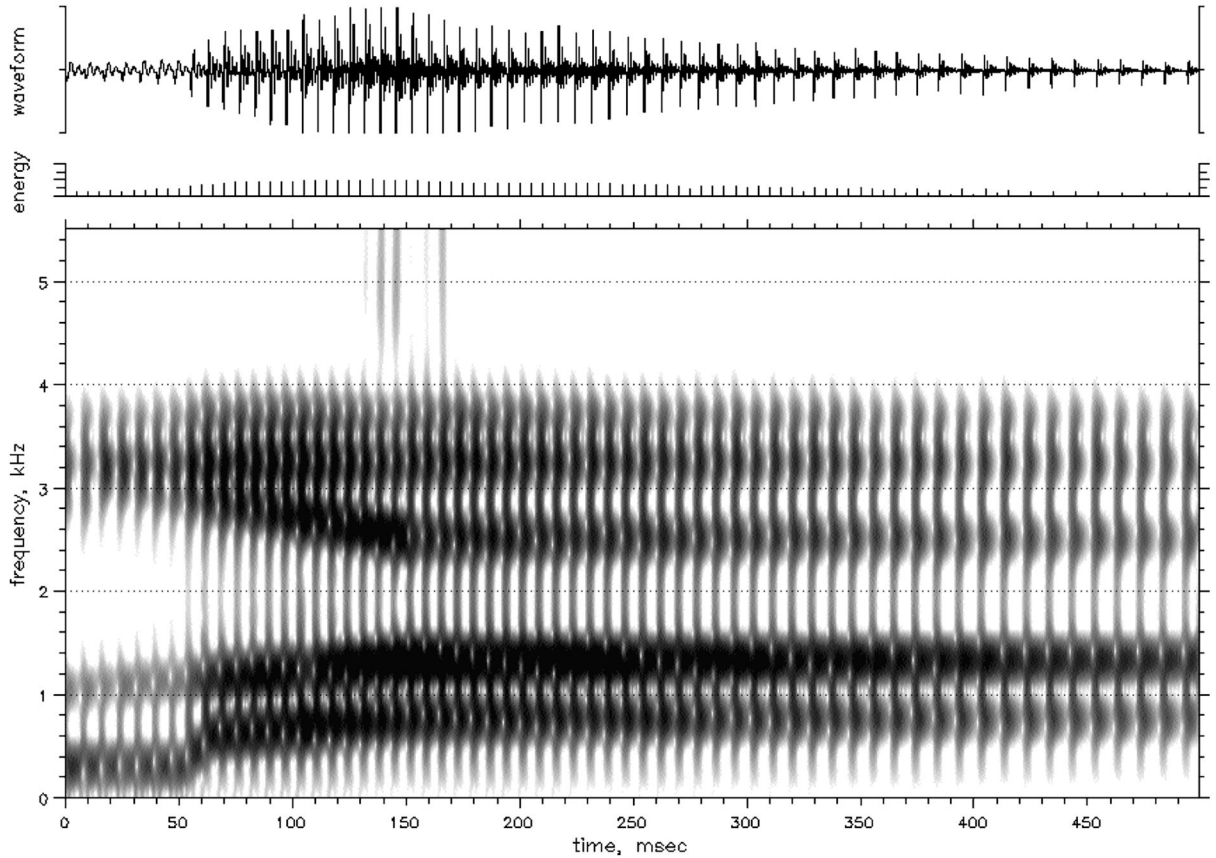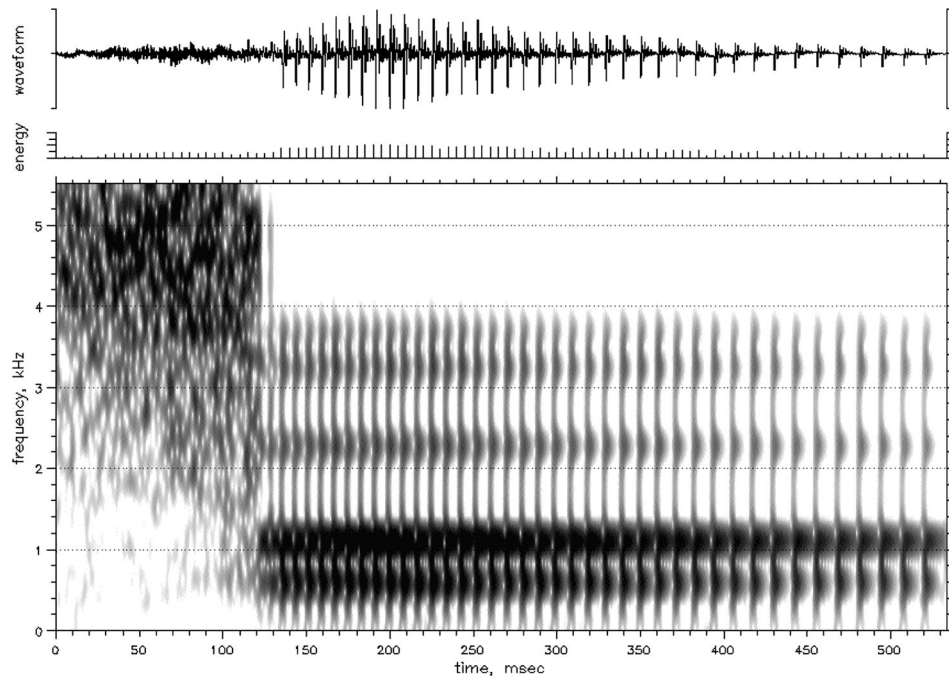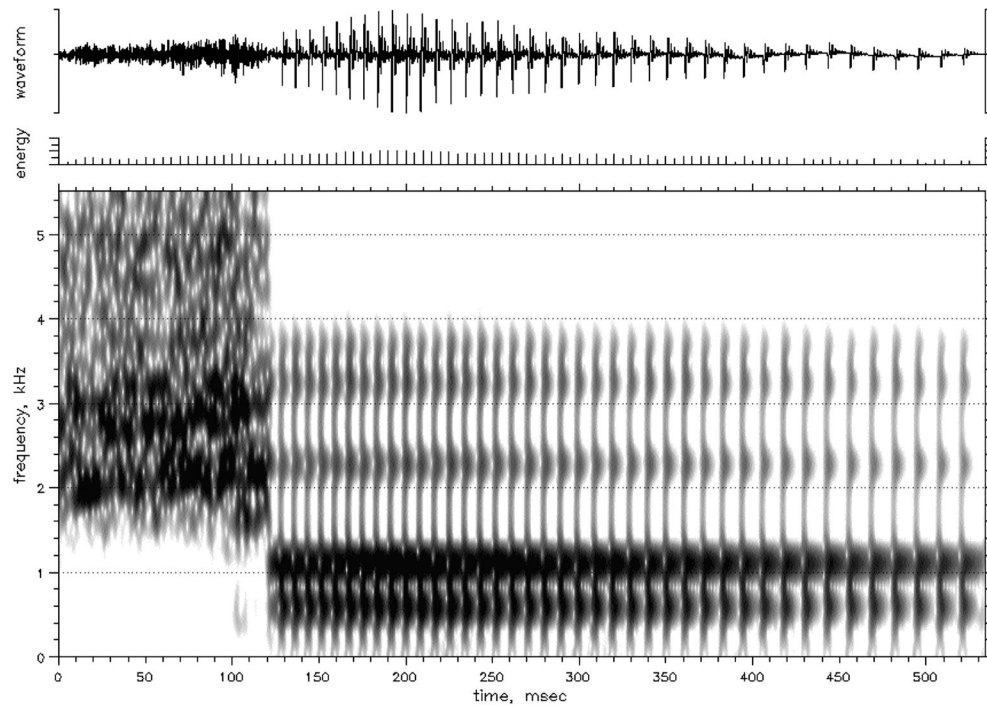Waveform (top panel), energy spectrum (middle panel), and frequency spectrum (bottom panel) of /ra/.

Figure A4.
Waveform (top panel), energy spectrum (middle panel), and frequency spectrum (bottom panel) of /la/.

Figure A5.
Waveform (top panel), energy spectrum (middle panel), and frequency spectrum (bottom panel) of /sa/.

Figure A6.
Waveform (top panel), energy spectrum (middle panel), and frequency spectrum (bottom panel) of /ʃa/.
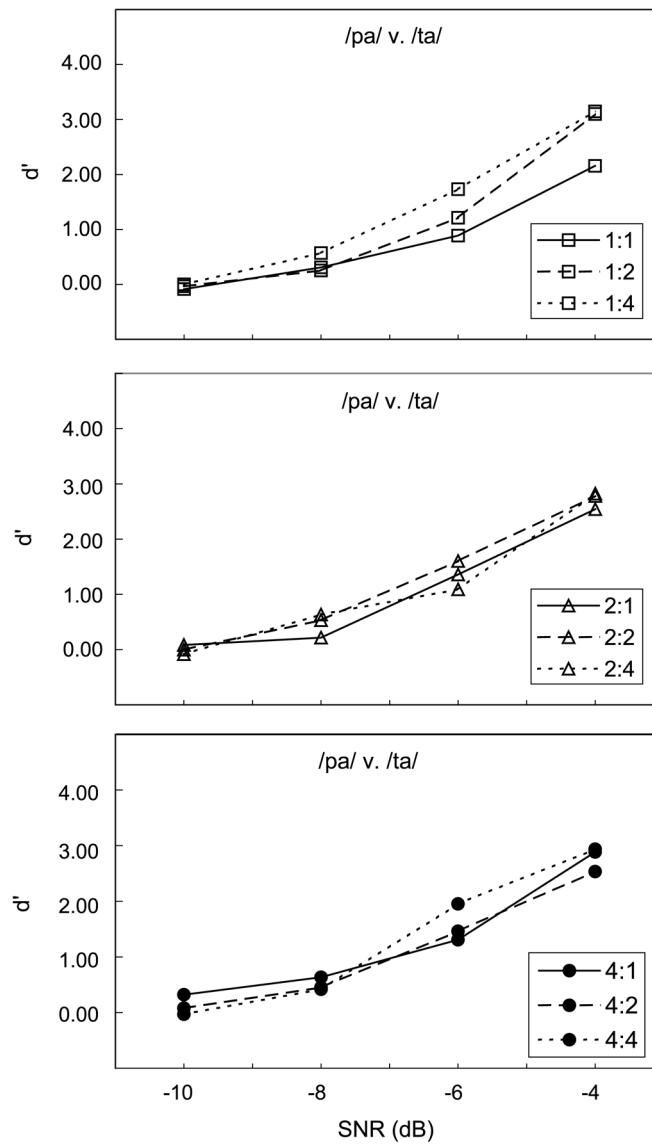
Figure 1.
Mean performance on /pa/ versus /ta/ for standard repetition number of one (top panel), standard repetition number of two (middle panel), and standard repetition number of four (bottom panel). Squares, triangles, and circles indicate a standard number of one, two, and four repetitions, respectively. Connected, dashed, and dotted lines indicate a comparison number of one, two, and four repetitions, respectively.
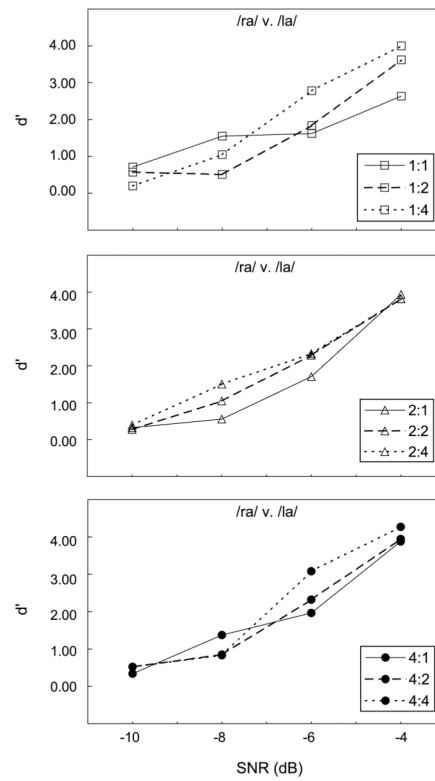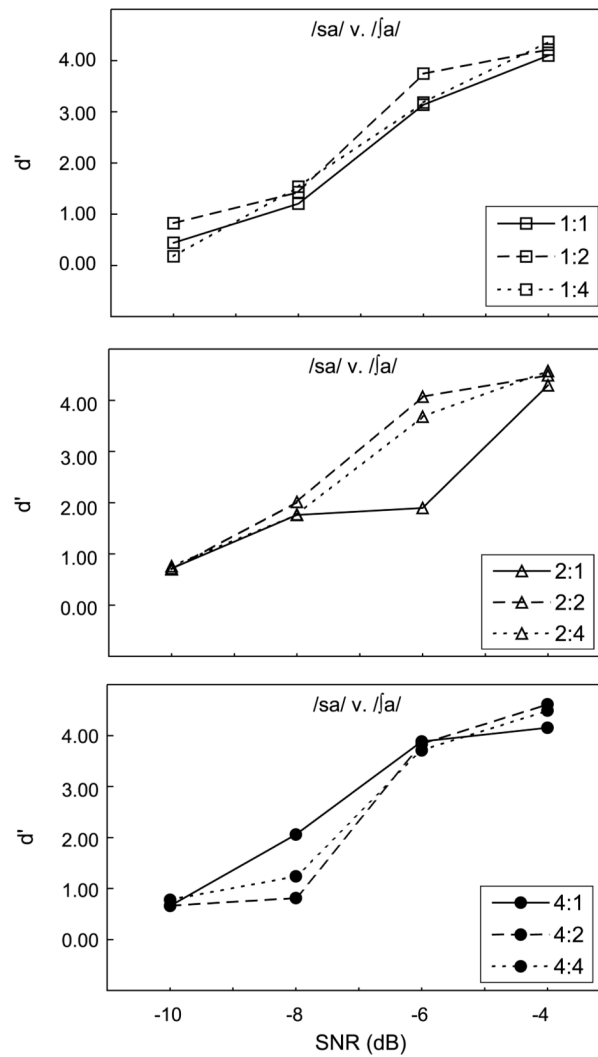
Figure 2.
Mean performance on /ra/ versus /la/ for standard repetition number of one (top panel), standard repetition number of two (middle panel), and standard repetition number of four (bottom panel). Squares, triangles, and circles indicate a standard number of one, two, and four repetitions, respectively. Connected, dashed, and dotted lines indicate a comparison number of one, two, and four repetitions, respectively.
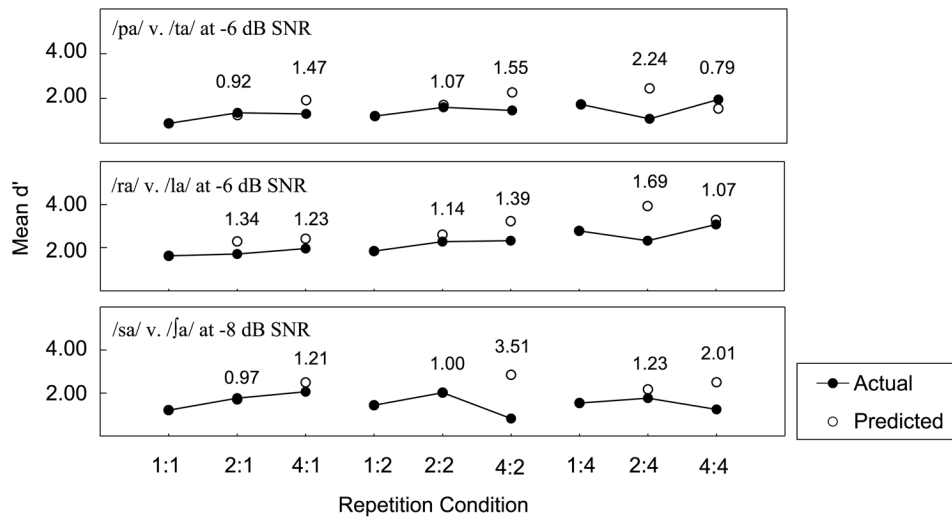
Figure 3.
Mean performance on /sa/ versus /ʃa/ for standard repetition number of one (top panel), standard repetition number of two (middle panel), and standard repetition number of four (bottom panel). Squares, triangles, and circles indicate a standard number of one, two, and four repetitions, respectively. Connected, dashed, and dotted lines indicate a comparison number of one, two, and four repetitions, respectively.

Figure 4.
Actual (filled circles) versus predicted (open circles) performance (based on multiple looks) grouped by number of comparison stimulus repetitions. The top panel shows performance for /pa/ versus /ta/ at −6 dB signal-to-noise ratio (SNR), the middle panel shows that for /ra/ versus /la/ at −6 dB SNR, and the bottom panel shows performance for /sa/ versus /ʃa/ at −8 dB SNR. The numerical values indicate the ratio of predicted to actual performance at each stimulus repetition condition in which predictions were made.
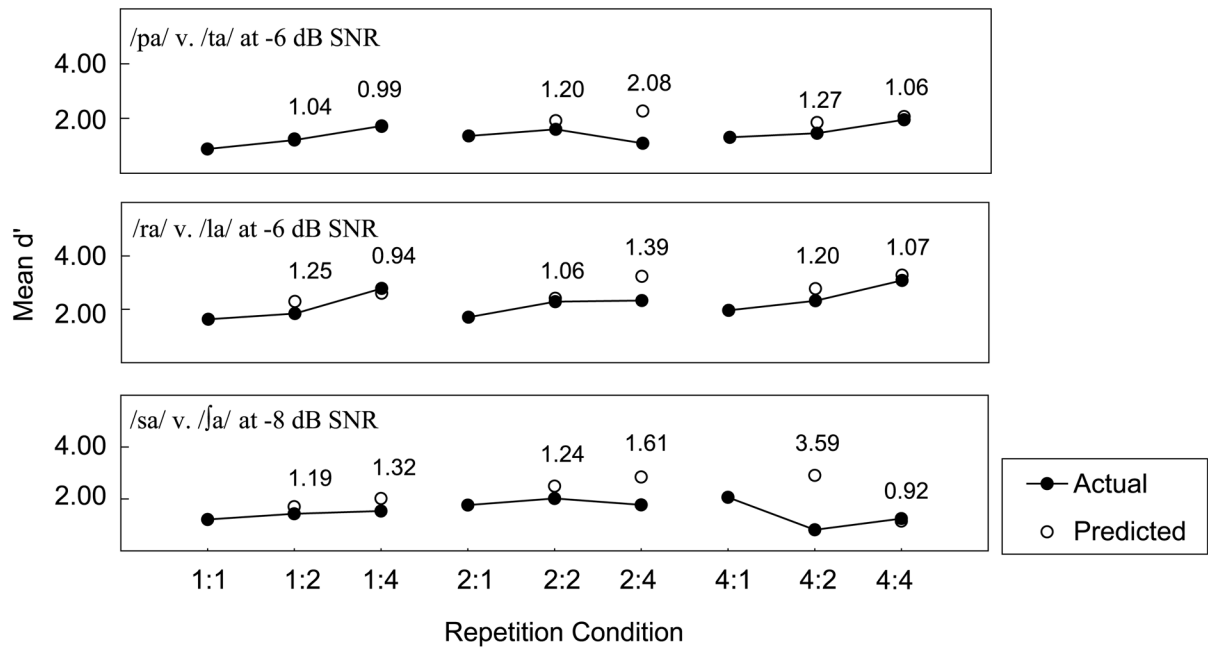
Figure 5.
Actual (filled circles) versus predicted (open circles) performance (based on multiple looks) grouped by number of standard stimulus repetitions. The top panel shows performance for /pa/ versus /ta/ at −6 dB SNR, the middle panel shows that for /ra/ versus /la/ at −6 dB SNR, and the bottom panel shows performance for /sa/ versus /∫a/ at −8 dB SNR. The numerical values indicate the ratio of predicted to actual performance at each stimulus repetition condition in which predictions were made.
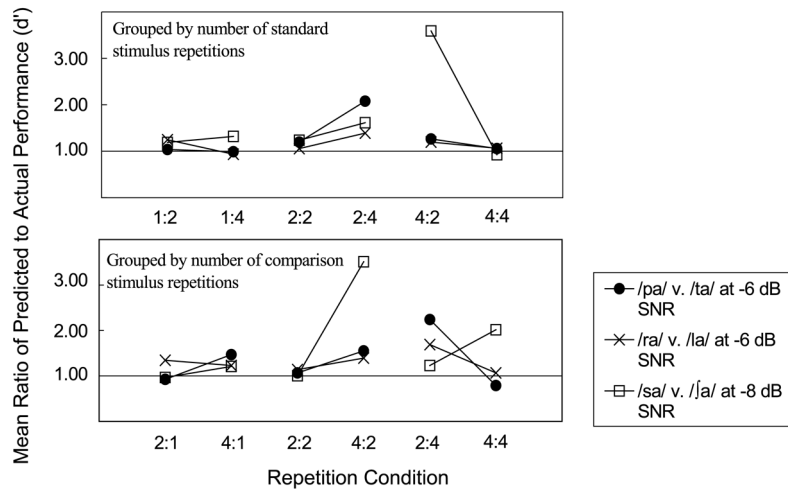
Figure 6.
Mean ratio of predicted to actual performance (based on multiple looks). The top panel displays the effects of doubling the number of comparisons. The bottom panel shows the effects of doubling the number of standards. The filled circles represent the /pa/ versus /ta/ comparison at −6 dB SNR, the /ra/ versus /la/ comparison at −6 dB SNR is represented by "X"s, and the open squares represent the /sa/ versus /ʃa/ comparison at −8 dB SNR.
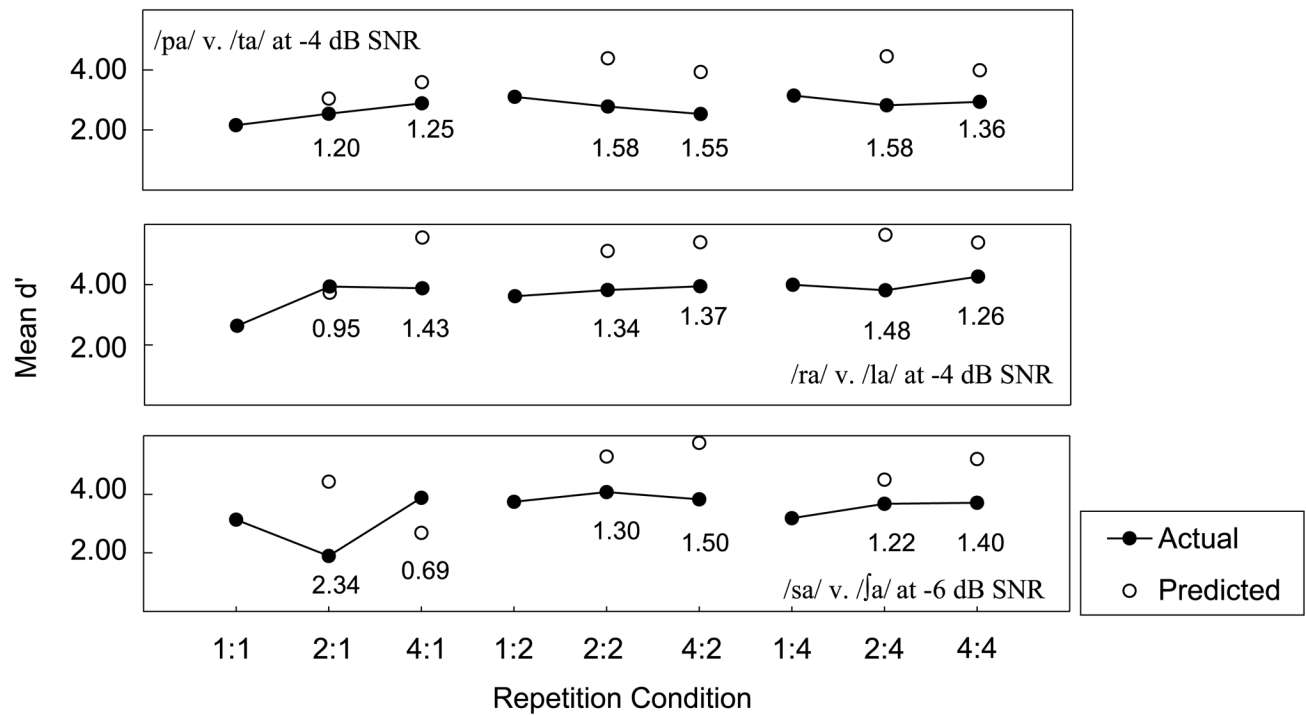
Figure 7.
Actual (filled circles) versus predicted (open circles) performance (based on multiple looks) grouped by number of comparison stimulus repetitions. The top panel shows performance for /pa/ versus /ta/ at −4 dB SNR, the middle panel shows that for /ra/ versus /la/ at −4 dB SNR, and the bottom panel shows performance for /sa/ versus /ʃa/ at −6 dB SNR. The numerical values indicate the ratio of predicted to actual performance at each stimulus repetition condition in which predictions were made.
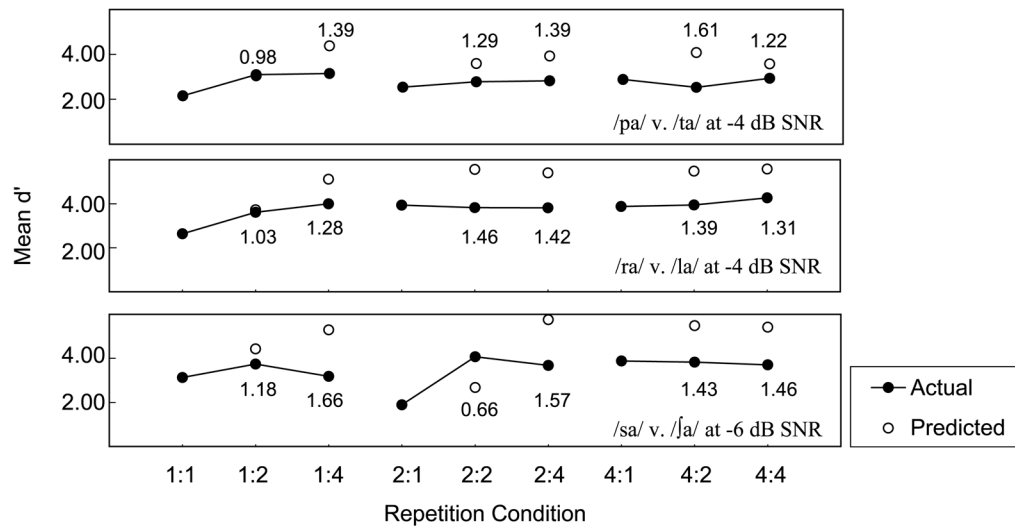
Figure 8.
Actual (filled circles) versus predicted (open circles) performance (based on multiple looks) grouped by number of standard stimulus repetitions. The top panel shows performance for /pa/ versus /ta/ at −4 dB SNR, the middle panel shows that for /ra/ versus /la/ at −4 dB SNR, and the bottom panel shows performance for /sa/ versus /ʃa/ at −6 dB SNR. The numerical values indicate the ratio of predicted to actual performance at each stimulus repetition condition in which predictions were made.
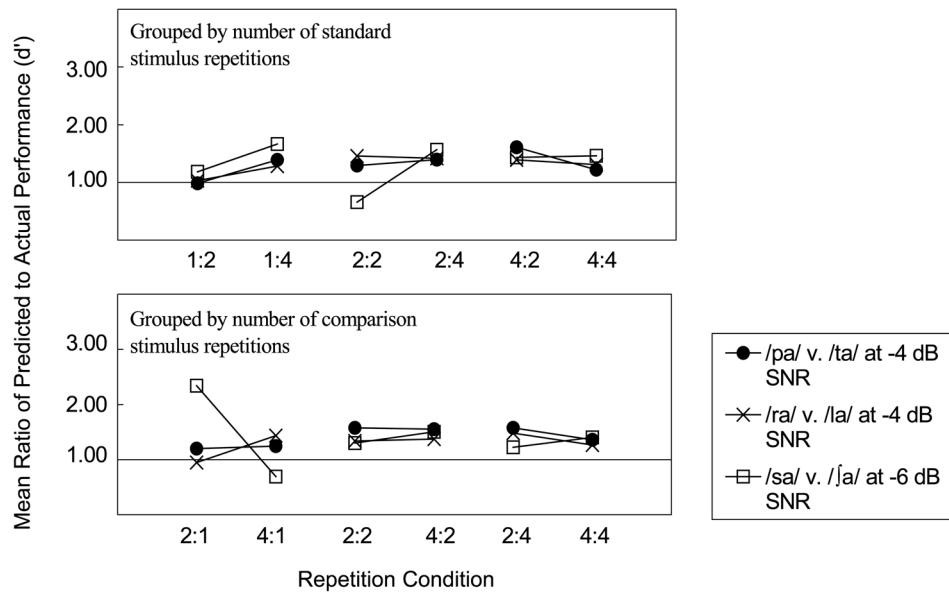
Figure 9.
Mean ratio of predicted to actual performance (based on multiple looks). The top panel displays the effects of doubling the number of comparisons. The bottom panel shows the effects of doubling the number of standards. The filled circles represent the /pa/ versus /ta/ comparison at –4 dB SNR, the /ra/ versus /la/ comparison at –4 dB SNR is represented by "X"s, and the open squares represent the /sa/ versus /ʃa/ comparison at –6 dB SNR.

Table 1

Mean group performance in d' collapsed across repetition conditions at each signal-to-noise ratio (−10, −8, −6, and −4 dB) for each pair of syllable contrasts (/pa/ vs. /ta/, /ra/ vs. /la/, and /sa/ vs. /ʃa/).

	Signal-to-Noise Ratio (dB)			
	−10	−8	−6	−4
/pa/ vs. /ta/	0.03	0.45	1.40	2.77
/ra/ vs. /la/	0.43	1.03	2.21	3.77
/sa/ vs. /ʃa/	0.63	1.54	3.46	4.36

Table 2

Analysis of variance with repeated measures results.

Main Effects and Interactions	*F*	*p*
Significant Effects		
Manner/syllable pair	$F(2, 26) = 17.193$	<.001
Standard repetition number	$F(2, 26) = 7.309$	.003
Comparison repetition number	$F(2, 26) = 17.546$	<.001
Signal-to-noise ratio (SNR)	$F(3, 39) = 272.574$	<.001
Manner × SNR	$F(6, 78) = 4.721$	<.001
Standard Repetition Number × Comparison Repetition Number	$F(4, 52) = 4.134$	.006
Standard Repetition Number × SNR	$F(6, 78) = 3.834$	.002
Comparison Repetition Number × SNR	$F(6, 78) = 7.673$	<.001
Standard Repetition Number × Comparison Repetition Number × SNR	$F(12, 156) = 2.464$	.006
Manner × Standard Repetition Number × Comparison Repetition Number × SNR	$F(24, 312) = 2.992$	<.001
Nonsignificant Effects		
Manner × Standard Repetition Number	$F(4, 52) = 0.555$	.696
Manner × Comparison Repetition Number	$F(4, 52) = 1.892$	.126
Manner × Standard Repetition Number × Comparison Repetition Number	$F(8, 104) = 1.937$	.062
Manner × Standard Repetition Number × SNR	$F(12, 156) = 1.529$	.119
Manner × Comparison Repetition Number × SNR	$F(12, 156) = 1.240$	.260