Article published online: 2020-09-02

Effect of Competition, Signal-to-Noise Ratio, Race, and Sex on Southern American English Dialect Talkers' Sentence Recognition

Andrew Stuart[1] Yolanda F. Holt[1] Alyssa N. Kerls[1] Madeline R. Smith[1]

[1] East Carolina University, Greenville, NC

J Am Acad Audiol 2020;31:521–530.

Address for correspondence Andrew Stuart, Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC 27858-4353 (e-mail: stuarta@ecu.edu).

Abstract

Background Although numerous studies have examined regional and racial–ethnic labeling of talker identity, few have evaluated speech perception skills of listeners from the southern United States.

Purpose The objective of the study was to examine the effect of competition, signal-to-noise ratio (SNR), race, and sex on sentence recognition performance in talkers from the Southern American English dialect region.

Research Design A four-factor mixed-measures design was used.

Study Sample Forty-eight normal-hearing young African American and White adults participated.

Data Collection and Analyses The Perceptually Robust English Sentence Test Open-set was used in quiet and in continuous and interrupted noise and multitalker babble at SNRs of −10, −5, 0, and 5 dB.

Results Significant main effects of competition ($p < 0.001$) and SNR ($p < 0.001$) and a competition by SNR interaction were found ($p < 0.001$). Performance improved with increasing SNRs. Performance was also greater in the interrupted broadband noise at poorer SNRs, relative to the other competitors. Multitalker babble performance was significantly poorer than the continuous noise at the poorer SNRs. There was no effect of race or sex on performance in quiet or competition.

Conclusions Although African American English and White American English talkers living in the same geographic region demonstrate differences in speech production, their speech perception in noise does not appear to differ under the conditions examined in this study.

Key words
- dialect
- multitalker babble
- noise
- sentence recognition

Introduction

Beyond regional and racial-ethnic labeling of talker identity, limited research has evaluated speech perception skills for African Americans and White Americans from the southern United States. Regional differences in speech perception have been previously documented. For example, Willis[60] asked listeners from Buffalo, New York, and Fort Erie, Ontario, to categorize synthetic vowels, finding substantial differences in the listener's category boundaries for /ɛ-æ/ and /æ-ɑ/ tokens that directly corresponded to their different regional production of those tokens. Labov and colleagues examined regionally diverse North American English listeners' recognization of vowels (Labov et al[30]), excised vowels, words, and sentences (Labov[27]; Labov and Ash[28]). The listeners had difficulty in correctly recognizing vowels, words, and sentences produced by talkers from distant geographical regions and sometimes by talkers from their own ethnically diverse communities.

Copyright © 2020 by the American Academy of Audiology. All rights reserved. Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 760-0888.

DOI https://doi.org/10.3766/jaaa.19029.
ISSN 1050-0545.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

A listener's difficulty in accurately recognizing speech tokens (i.e., vowels, words, or sentences) produced by talkers from distant communities and sometimes speech tokens produced by diverse talkers from their own community may be related to a difference in cue weighting strategies used within a group. As discussed by Holt and Lotto[23] listener strategies for the use of cues may be language- or dialect-specific along four dimensions: (a) acoustic informativeness for category identity (i.e., the parameter is useful for distinguishing categories), (b) variance (i.e., the amount of pattern variation within a category and between categories), (c) perceptual weight (i.e., the robust nature of the auditory representation of the token to the auditory system where equal physical changes do not result in equivalent changes in perception), and (d) auditory task (i.e., phonetic categorization without a lexical component may be more difficult to perceive with accuracy).

Thomas[57] provides an example of the implementation of different cue weighting strategies based on group. He evaluated White talkers from Ohio and Mexican-American talkers from Texas on their production and perception of the vowel height in the words "tight," "tide," and "tie." Both groups produced higher glides (i.e., lower F_1 and higher F_2) preceding the voiceless consonant in "tight," but their perceptual use of the spectral change for "tight," "tide," and "tie" varied. The White Ohio participants used the height difference as a cue to discriminate "tight" from "tide"—production of final /d/ versus /t/. This perceptual alignment was concomitant with the low rate of final stop release for the Ohio group. The Mexican-American Texas participants used the spectral cue to distinguish between the /d/ as in "tide" and null as in "tie." The Texas group, unlike those from Ohio, had a high rate of final stop release in the productions of "tight" and "tide." The results of this experiment suggest the two groups of talkers assign different values to the same spectral cue and may have paired the spectral change with secondary cues on final release burst to make their judgments. One group used the height variation to cue the presence or absence of voicing, whereas the other group used it to cue the presence or absence of the phonetic and phonological final stop.

Similarly, Labov[27] and Labov and Ash[28] reported listeners' performance when presented with unfamiliar regionally distinctive productions of the tokens /u/ as in "boot," /o/ as in "boat," and /a͡ʊ/ as in "out" from lifelong members of the Pamlico Sound region of North Carolina. The listeners who were from outside this region were unsuccessful in recognizing the tokens with above chance accuracy. These results suggest that listeners who are less familiar with a regional dialect of North American English may not be able to parse the cues present in the speech token to accurately identify the intended speech target. As described by Holt and Lotto[23] auditory tasks that require phonetic categorization outside a lexical frame may be more difficult for a listener.

Speech discrimination and identification tasks require the listeners to align their knowledge and experience with the presented tokens. Cue weighting strategies, particularly for vowels, but also for consonants, are influenced by temporal relationships (duration), acoustic spectral variance (transition and glide height), and burst release for syllable final stops. The self-same auditory stimulus may be identified as different targets based on the linguistic experiences and expectations of the listener. For longer utterances presented under ideal listening conditions, with syntactic and lexical cues present to support a listener's identification accuracy, it is possible listeners with varied backgrounds and linguistic experiences would demonstrate equivalent skills when completing a speech recognition task. However, under challenging conditions, because of variation in cue weighting strategies that may be influenced by regional and racial-ethnic linguistic experiences, it is unknown how a diverse group of monolingual listeners with varying dialects might perform on a speech recognition task.

African American English (AAE) talkers living in the same geographic region as White American English (WAE) peers have been noted to use a distinct dialect termed AAE. The phonological and morphosyntactic variation in AAE is well described (e.g., Green[19]). More recently, attempts have been made to complete acoustic phonetic descriptions of AAE and to catalog the regional variation within the dialect. In southern versions of AAE, as spoken in eastern North Carolina, similarities and differences have been noted in the vowel space area for AAE and WAE talkers. The AAE members of the community raise the vowels /ɪ, ɛ, æ/ as produced in the words "hid," "head," and "had," producing tokens with lower F_1 values than WAE speaking peers. For the vowels /u/ as in "whod" and /o/ as in "hoed," the AAE talkers produce tokens with lower F_2 than WAE peers (Holt[22]). These differences in speech production may result in speech perception differences for the members of this racial-ethnic diverse group in a manner consistent with that of Willis,[60] where group differences in vowel production were observed as group differences in the perception of vowel token boundaries.

Differences in vowel perception and production, although intriguing, do little to enhance our understanding of the functional use of cue weighting strategies by listeners engaged in tasks consistent with real-world listening difficulties. For example, a listener's ability to accurately extract a speech signal in a noisy environment is of particular theoretical and practical interest. Does a listener's racial-ethnic background and dialect exposure significantly alter the ability to extract and encode speech stimuli? The primary variable contributing to the ability to perceive speech is audibility. During situations of daily living, a myriad of additional factors affect speech perception. This is due in part to the variability encountered in everyday listening situations. Performance in a speech perception task is predicated on talker internal, previously described, and talker external variables. Additional variables include background competition (e.g., type and level), additional listener variables (e.g., cognitive status, age, and hearing sensitivity), and response task goal (e.g., discrimination, identification, or recognition; Gilbert et al[18]; Theunissen et al[56]). Speech recognition performance can differentiate normal listeners and those with disordered speech perception. However, differences in speech recognition performance between and among groups of listeners do not identify what cue weighting strategies listeners' use. Cue weighting strategies

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

are likely to be different because of, in part, the wide variety of talker internal variables related to typical exposure to any variety of North American English dialects. For example, a listener's ability to discriminate the vowels in "boot," "boat," and "out" or to distinguish the words "tight," "tied," and "tight" or "head" and "had" should be predicated solely on the listener's ability to discriminate the tokens, not on their knowledge and experience with the regional dialect of the talker. Incorporating speech from a variety of North American English dialects in speech perception assessments is one way to address this shortcoming, yet the performance of racial-ethnic diverse listeners on such a task is still unknown.

To the best of our knowledge, there are no studies examining speech perception in noise skills for African Americans and White Americans from similar regional dialects. When choosing test materials, traditional word and sentence recognition materials are problematic because they typically have been designed to remove sources of variability using single talkers with indistinct regional dialect speaking at slow articulation rates (Bess[4]). The face validity of such traditional tests is questionable. The recently developed Perceptually Robust English Sentence Test Open-set (PRESTO) is an appealing alternative (Gilbert et al[18]). It was principally created to investigate the effect of variability in speech on listeners' speech recognition performance (Gilbert et al[18]). The target talker variability included dialect, sex, and multiple talkers. The PRESTO sentences were created from the Texas Instruments/Massachusetts Institute of Technology Acoustic-Phonetic Continuous Speech Corpus (Garofolo et al[17]). This database includes recordings of sentence lists from >600 talkers representing eight major dialects of American English (i.e., New England, New York City, Northern, North Midland, South Midland, Southern, Western, and Army Brat—individuals who moved frequently during their childhood). Sentence lists, containing 18 utterances, were constructed for the PRESTO corpus such that no talker was repeated in a list and the sex of the talker was balanced. Each sentence list contained 5–10 words with 3–6 key words, with a total of 76 key words per list (e.g., "We always *thought* we would *die* with our *boots* on" and "*Brush fires* are *common* in the *dry underbrush* of *Nevada*," with key words in italics). Sentences were mixed with random samples of six-talker babble of three women and three men with General American dialects. Each sentence having the same root-mean-square amplitude level was calibrated to be presented at 64 dB SPL, and the signal-to-noise ratio (SNR) was manipulated (i.e., −5, −3, 0, and 3 dB). Performance on the PRESTO was assessed with correct key word identification from participants' typed responses. The PRESTO is more difficult than conventional sentence recognition tests as the variability in target talker burdens the listener with an added attention-processing load to perceptually orient to each new talker (Gilbert et al[18]). That is, with typical speech recognition tests, the listener hears the same talker with multiple utterances. In the PRESTO, the listener hears multiple talkers, from multiple regional dialect backgrounds producing different utterances. This variability is likely to require the listener to pay more attention to the task and is also likely to remove any advantage a listener from the same dialect background as the presenting talker has in test performance. Individuals with regional dialects should, therefore, not be disadvantaged when completing the speech perception tasks in the PRESTO. This assumption, however, has not been demonstrated empirically, and the effect of talker/listener dialect on PRESTO performance (and other speech-in-noise tests) has been advocated (Gilbert et al[18]).

In previous reports, listeners with normal hearing sensitivity demonstrated large variability in performance on the PRESTO. Gilbert et al[18] reported an overall mean sentence recognition performance (i.e., averaged across four SNRs) of 63%, with a range of 40–76% in 121 young adults aged 18–39 years. Listeners were native talkers of American English with residency in the United States before 18 years of age. Tamati et al[54] speculated that the variability Gilbert et al[18] observed in listener performance was related to listeners' background and linguistic experience. Toward that end, Tamati et al[54] examined participants from the original Gilbert et al[18] study, whose performance fell in the lower or upper quartile of the original distribution of scores, on a number of factors that could contribute to performance variability. They evaluated perceived real-world listening (i.e., self-report questionnaire on situational hearing and hearing health history), indexical processing abilities (i.e., talker and sex discrimination tasks and a regional dialect categorization task), and neurocognitive abilities (i.e., verbal short-term memory, verbal working memory, attention/inhibition, vocabulary size, executive function, and nonverbal intelligence) of listeners. Tamati et al[54] found that low and high PRESTO performers did not differ in real-world listening. However, participants did evidence differences in indexical processing, short-term and working memory, vocabulary size, and some domains of executive functioning. Tamati et al[54] concluded that these differences might contribute to the variability observed in PRESTO performance.

All the original and subsequent studies examining listener performance on the PRESTO recruited participants from the midwest or midland dialect region (Gilbert et al[18]; Tamati et al[54]; Faulkner et al[13]; Plotkowski and Alexander[36]). It is unknown if AAE and WAE talkers from a southern dialect region would demonstrate similar performance on the PRESTO test. Furthermore, the influence of racial-ethnic dialect variation has not been evaluated on this assessment, and limited data have been collected on the effect of different types of maskers on listener performance in the presence of the multiple regional dialects presented by the PRESTO talkers.

In previous studies (Gilbert et al[18]; Tamati and Pisoni[55]; Faulkner et al[13]; Plotkowski and Alexander[36]), the PRESTO was presented in a background of competing multitalker babble at SNRs ranging from −5 to 3 dB. The effectiveness of any masker in obscuring the speech signal is dependent on the level of the masker relative to the signal, the acoustic spectrum of the noise with respect to the signal, and its temporal continuity (Miller[33]). Maskers are generally classified as energetic or informational. Energetic masking involves spectral overlap of speech and noise signals. Informational masking involves spectral overlap of speech and competitor, as well as the competitor perceptually interfering with the speech signal. Broadband noises and

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

competing speech/babble are typical energetic and informational maskers, respectively. When energetic and informational maskers are similar in spectral and temporal content, informational maskers are more detrimental to speech perception at equivalent SNRs (Brungart et al[6]; Helfer and Freyman[20]; Rosen et al[39]). The greater masking effect seen with informational masking has been attributed to a difficulty separating the talker from competing background talkers that cannot be fully attributed to the auditory periphery. It has been suggested that talker and speech masker similarity results in increased attention uncertainty (Kidd et al[26]; Brungart et al[6]; Freyman et al[15]; Durlach et al[12]; Freyman et al[16]; Hoen et al[21]). That is, the listener cannot separate the target talker and the competing speech into separate perceptual streams or objects, resulting in errors in correct recognition of the target talker speech. The talker/speech competitor similarity can be attributed to similar acoustic characteristics (e.g., same talker/competitor voice and same sex) or relevance of the contextual information (e.g., semantic similarity). Informational masking is also affected by the ratio of energetic and informational masking in the multitalker babble and the number of talkers in the multitalker babble (Freyman et al[16]; Hoen et al[21]; Rosen et al[39]).

When background competitors decrease in level, increase in intermittency, or both, listeners' speech perception improves; that is, the effectiveness of the masker is reduced and listeners experience a perceptual advantage or release from masking while listening to speech in these competitors. Stuart and colleagues (Stuart[45,46]; Stuart and Phillips[50]; Stuart et al[49,52,53]) have used speech perception in stationary and nonstationary energetic maskers (i.e., continuous and interrupted broadband noises) to examine temporal resolution abilities in many groups including those with hearing loss, simulated hearing loss, older individuals, young children, and bilingual talkers. The competing interrupted noise consists of noise bursts and silent periods between them, with durations of both noise and silence varying randomly from 5 to 95 msec. The temporal structure of the noise was selected to mimic the acoustic elements of speech (i.e., several milliseconds to tens of milliseconds, representative of the durations of consonantal bursts to steady-state vowels, respectively). The noise–time fraction (i.e., the proportion of time occupied by the noise) for the interrupted noise is one-half. Both continuous and interrupted noises have identical long-term average spectra differing only in their temporal structure. In one paradigm, speech stimuli are fixed in level and the noises are presented at equal and varying SNRs. In a second paradigm, the noises are fixed in level and thresholds for speech stimuli are determined. Listeners evidence enhanced speech recognition and lower speech thresholds in the interrupted relative to the continuous noise. The perceptual advantage seen in the interrupted noise has been attributed to the listener's temporal ability to resolve speech fragments or get glimpses or looks of speech between the gaps of noise and to patch the information together to identify the specific speech stimuli (Miller[33]; Miller and Licklider[34]; Pollack[37]; Dirks et al[10]). The effect of stationary and nonstationary energetic maskers on PRESTO performance is unknown.

The purpose of the present study was to examine performance on the PRESTO by listeners who were lifelong or near lifelong residents of the broadly defined Southern American English dialect region. Specifically, the effect of competition, SNR, race, and sex on Southern American English dialect talkers' sentence recognition was examined. Both AAE and WAE normal-hearing listeners (male and female) were included. Performance was evaluated in competing energetic and informational masking (i.e., continuous broadband, interrupted broadband, and multitalker babble) at four SNRs (i.e., −10, −5, 0, and 5 dB). Performance was also evaluated in a baseline quiet condition. It was important to first establish that there was no performance difference on the PRESTO in quiet as a function of race and sex such that any differences found in noise could not be attributed to differences in baseline performance. It was hypothesized that there would be no difference in performance in quiet as a function of race and sex as none have been reported previously in the literature for normal-hearing listeners. It was hypothesized that performance in the interrupted noise condition would be better than the performance in both the continuous noise and the multitalker babble condition. This improvement for interrupted noise is expected because of the intermittent nature of the noise, which allows the listener to have interrupted auditory glances at the signal. The poorest performance was expected in multitalker babble—the informational masker. It was also hypothesized that performance would improve with an increasing SNR—a well-accepted effect. It was further hypothesized that no effect of racial-ethnic group membership or sex would be observed. A sex effect has not been previously demonstrated with normal-hearing adolescent and young adults on other speech recognition in noise tasks (e.g., Ribera[38]; Sbompato et al[41]; Jacobi et al[24]). The effect of race has not been explored previously in the literature. Although speech production differences by racial-ethnic group membership have been observed in the eastern North Carolina (Holt[22]), differences between AAE and WAE speech perception are not expected if demographic and educational backgrounds and linguistic and dialect exposure are similar.

Methods

Participants

Participants were two groups of 24 African American ($M = 21.8$ years, standard deviation [SD] = 2.3) and 24 White ($M = 22.5$ years, SD = 1.5) young adults. Each group contained 12 women and 12 men. Both groups of participants presented with normal pure-tone thresholds defined as $\leq$25 dB HL (ANSI[2]) at octave frequencies from 250 to 8000 Hz. Participants also presented normal middle-ear function indices of peak compensated static acoustic admittance, tympanometric width, tympanometric peak pressure, and equivalent ear canal volume within the 90% range of sex-specific normative data (Roup et al[40]). Participants reported a negative history of hearing loss and speech, language, cognitive, neurological, and otological disorders. All participants self-identified as being lifelong or near lifelong residents of the broadly defined the Southern American English dialect region as defined in the

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Atlas of North American English: Phonetics, Phonology and Sound Change (Labov et al[29]).

We were interested in assuring that the groups of participants shared similar demographic and linguistic and dialect exposure such that neither variable could be attributed to group differences (if found) in speech recognition performance. Toward that end, participants were required to complete a questionnaire probing demographic variables plus linguistic and dialect exposure. All participants reported their ethnicity as non-Hispanic or non-Latino except one White man who reported ethnicity as Hispanic or Latino. There were no statistical differences in the number of years reported living in the south, as defined by the U.S. Census Bureau regions, as a function of race or sex ($p > 0.05$). The highest level of reported education did not differ between African Americans and Whites ($p > 0.05$); however, women tended to have a higher proportion completing bachelor's degrees and men a higher proportion with some college education ($p < 0.05$). Approximately 15% (n = 7) reported studying abroad, and 70% (n = 34) reported studying a second spoken language. There were no differences between the proportions of these participants as a function of race or sex ($p < 0.05$).

Apparatus and Stimuli

A double-wall sound-treated audiometric suite (IAC Acoustics, North Aurora, IL) served as the test environment. The recorded stimuli were routed from a dual-disc compact disc player (Phillips Model CDR 765 K02, Andover, MA) to a clinical audiometer (Grason-Stadler GSI 61 Model 1761-9780XXE, Eden Prairie, MN) and an insert earphone (Etymotic Research Model ER-3A, Elk Grove Village, IL).

The sentence stimuli were derived from the compact disc recordings of the PRESTO (Gilbert et al[18]). Twelve PRESTO sentence lists (i.e., 2, 3, 4, 7, 8, 10, 11, 13, 14, 15, 17, and 23) were chosen from the corpus originally used by Faulkner et al.[13] These lists were selected for list equivalency and listener consistency under multiple conditions described by Faulkner et al.[13] To familiarize participants with the sentence materials and task, an additional list (#1) served as a practice list (Gilbert et al[18]). Sentence lists ranged in length from 134 to 144 s, with a mean of 137 s. The competing noises were compact disc recordings of PRESTO multitalker babble (Gilbert et al[18]) and continuous and interrupted broadband noises. The multitalker babble was constructed from six-talker babble of three male talkers and three female talkers using General American English dialect. The continuous and interrupted broadband noises have been described in detail elsewhere (Stuart and Phillips[50,51]; Stuart[44]). Both the continuous and interrupted broadband noises had equivalent long-term average spectra and were normalized to have equal power (for a detailed description of the construction and calibration of the noise stimuli, see Stuart[44]). They differed only in their temporal structure. The interrupted noise had a rectangular on/off envelope with randomized gating. The noise duty cycle was 0.50. The interrupted noise was characterized with noise bursts and silent periods between them with durations of both varying randomly from 5 to 95 msec. The spectra of the noise competitors

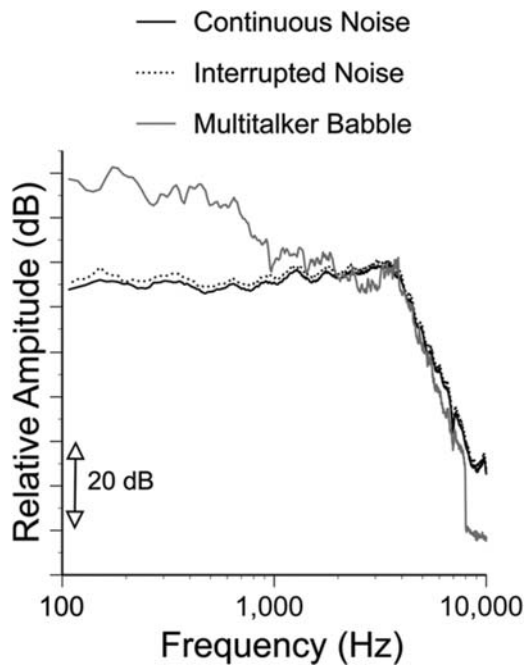

Fig. 1 Spectra of competing noises and multitalker babble transduced through an insert earphone to a 2-cm^3 coupler.

are presented in ►**Figure 1**. (To measure the spectra of the noise competitors, 30-second tokens of each noise were routed through the same compact disc player, audiometer, and insert earphone, noted earlier, to a handheld sound level meter [Brüel and Kjær Type 2250, Nærum, Denmark] with a 1-inch pressure microphone [Brüel and Kjær Type 4144] and a-2 cm^3 coupler [Brüel and Kjær Type DB 0138]. The microphone was calibrated with a Brüel and Kjær Type 4231 Sound Calibrator. The output from the sound level meter was then routed to a signal acquisition interface [Sound Technology Dynamic Signal Acquisition System, Poulsbo, WA] and to a laptop computer [Lenovo Model X1 Carbon, Morrisville, NC]. Stimuli were then recorded with SpectraPlus-SC FFT Spectral Analysis System software [version 5.1.0.33, Pioneer Hill Software, Poulsbo, WA] with a sampling rate of 22050 Hz and 16-bit sampling. Fast Fourier Transforms [FFTs] were performed on these tokens using SpectraPRO-FFT software using a Hanning window, FFT size of 2048, and a decimation ratio of 1.)

Procedure

The University and Medical Center Institutional Review Board at East Carolina University approved all experimental procedures including recruitment and acquisition of informed consent before data collection. All participants provided voluntary informed consent before data collection.

Stimuli were presented monaurally to the right ear of each participant. Before experimental testing, the participants received the practice list for familiarization of the test stimuli. Each participant received 13 PRESTO lists—a list in quiet followed by 12 sentence lists in the competing stimuli (i.e., three noises with four SNRs). The list presented in quiet was intended as a baseline condition to ensure no baseline differences in speech recognition between groups. As 12

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

PRESTO lists were used, the list presented in quiet was repeated in one noise condition for each participant. The presentation order of all lists was randomized across participants. All 12 sentence lists presented in competition were counterbalanced with a digram-balanced Latin squares design (Wagenaar[59]). As such, the 12 lists (including the repeated practice list) were equally presented across the three competitors and four SNRs. The PRESTO sentences were presented at 65 dB SPL. The competing stimuli were presented at four different SNRs (i.e., −10, −5, 0, and 5 dB). Each competitor was calibrated to the same root-mean-square amplitude level. All stimuli were calibrated with the insert earphone coupled to a 2-cm^3 coupler (Brüel and Kjær Type DB 0138), 1-inch pressure microphone (Brüel and Kjær Type 4144), and sound level meter (Brüel and Kjær Type 2250). The participants were instructed to repeat the sentences aloud following presentation. Scoring of key word percent correct was determined for all conditions (Gilbert et al[18]) by normal-hearing research assistants following participants' responses.

Results

Sentence recognition percent correct performance in quiet, as a function of race (i.e., African American and White) and sex (i.e., female and male) was examined with a two-factor analysis of variance (ANOVA). Prior to inferential analyses, this and subsequent percent correct data underwent an arcsine transformation. The main effects of race [$F_{(1,\ 44)} = 0.53$, $p = 0.47$, $= 0.01$] and sex [$F_{(1,\ 44)} = 0.68$, $p = 0.41$, $= 0.01$] were not statistically significant. The race by sex interaction was also not statistically significant [$F_{(1,44)} = 0.75$, $p = 0.39$, $= 0.02$]. (Effect sizes are indexed by partial eta squared. Cohen[8] classifies small, medium, and large effect size values as 0.10, 0.25, and 0.40, respectively.) Average performance in quiet was 94.4% (SD = 5.4).

Sentence recognition percent correct performance as a function of competition (i.e., continuous noise, interrupted noise, and multitalker babble), SNR (i.e., −10, −5, 0, and 5 dB), race (i.e., African American and White), and sex (i.e., female and male) was examined with a four-factor mixed ANOVA. Mauchly's test of sphericity was used to test the compound symmetry assumption for the repeated measures variables. The degrees of freedom and p values were adjusted, and Greenhouse–Geisser values were reported for instances in which the sphericity assumption was not satisfied. The summary of the analysis is presented in ►**Table 1**. Statistically significant main effects of competition and SNR ($p < 0.001$) were found. The competition by SNR interaction was also statistically significant ($p < 0.001$). In general, performance was better in the interrupted noise. Performance also improved with increasing SNRs. Performance differences were most pronounced at the poorer SNRs. Mean sentence recognition percent correct performance as a function of competition and SNR is presented in ►**Table 2** and ►**Figure 2**. The largest variance in scores occurred at SNRs where scores were ~50%, whereas variances are greatly reduced when scores approach 0% and 100% (Thornton and Raffin[58]).

Table 1 Summary of the Four-Factor Mixed ANOVA Comparing Sentence Recognition Performance as a Function of Competition, SNR, Race, and Sex

Source	*df*	*F*	*p* Value	
Competition	1.75	287.67	<0.001*†	0.87
SNR	3	1,311.14	<0.001*	0.97
Race	1	0.76	0.39	0.03
Sex	1	0.39	0.54	0.01
Competition × SNR	4.41	116.69	<0.001*†	0.73
Competition × race	2	1.51	0.23	0.03
Competition × sex	2	0.21	0.83	0.01
SNR × race	3	0.37	0.77	0.01
SNR × sex	3	1.64	0.18	0.04
Race × sex	1	1.87	0.18	0.04
Competition × SNR × race	6	1.14	0.31	0.02
Competition × SNR × sex	6	0.19	0.98	0.00
Competition × race × sex	2	0.83	0.44	0.02
SNR × race × sex	3	2.20	0.09	0.05
Competition × SNR × race × sex	6	1.14	0.34	0.02

*Statistically significant at $p < 0.05$.
†Greenhouse–Geisser *p* value.

Table 2 Mean Sentence Recognition Percent Correct Performance and SDs (in parentheses) as a Function of Competition and SNR

Competition	SNR (dB)			
	−10	−5	0	5
Multitalker babble	1.7 (2.4)	24.1 (11.6)	65.6 (12.1)	86.8 (6.8)
Continuous noise	6.9 (11.1)	32.0 (11.8)	68.4 (12.6)	82.9 (8.7)
Interrupted noise	56.6 (18.6)	68.8 (13.0)	79.8 (11.8)	84.7 (9.8)

It was of interest to find the source of the competition by SNR interaction. Toward that end, four sets of two orthogonal single-*df* comparisons (Keppel and Wickens[25]; Maxwell et al[32]) were undertaken to examine differences in performance among the three noise competitors at each SNR. At 5 dB SNR, performance in the multitalker babble was significantly better than the continuous noise ($p = 0.008$, $= 0.14$), and there was no statistically significant difference between the interrupted noise and a linear combination of the continuous noise and multitalker babble ($p = 0.95$, $= 0.00$). At 0 dB SNR, there was no statistically significant difference in the performance in the multitalker babble and continuous noise ($p = 0.24$, $= 0.03$). However, there was a statistically significant difference between the interrupted noise and a linear combination of the continuous noise and multitalker

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

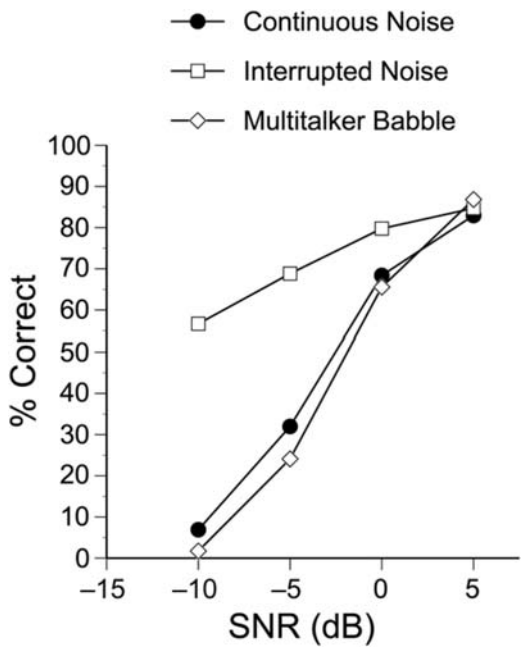

Fig. 2 Mean sentence recognition percent correct as a function of competition and SNR.

babble ($p < 0.0001$, $= 0.54$). At −5 dB SNR, performance in the multitalker babble was significantly poorer than the continuous noise ($p < 0.0001$, $= 0.25$), and there was a statistically significant difference between the interrupted noise and a linear combination of the continuous noise and multitalker babble ($p < 0.0001$, $= 0.92$). Likewise, at −10 dB SNR, performance in the multitalker babble was significantly poorer than the continuous noise ($p < 0.0001$, $= 0.44$). There was a statistically significant difference between the interrupted noise and a linear combination of the continuous noise and multitalker babble ($p < 0.0001$, $= 0.89$).

Discussion

The effect of competition, SNR, race, and sex on Southern American English dialect talkers' PRESTO performance was examined. Four hypotheses were posited: an effect of noise would be evident with best performance in the interrupted noise followed by that in the continuous noise and multitalker babble, performance would improve with an increasing SNR, racial-ethnic group membership would not affect performance, and sex would not affect performance. The statistically significant effects of the competition and SNR on PRESTO performance confirmed the first two hypotheses. Improving performance with increasing SNRs is well established (Miller[33]). The improved scores in interrupted noise are consistent with previous findings in sentence recognition relative to continuous noise at equivalent SNRs (Stuart et al[49,52,53]; Stuart and Phillips[50]; Stuart[45,46]; Stuart and Butler[47,48]) attributed to a release from masking (Miller[33]; Miller and Licklider[34]; Pollack[37]; Dirks et al[10]).

Performance differences were seen between the continuous noise and multitalker babble at 5 dB SNR and the two poorest SNRs (i.e., −5 and −10 dB). The opposite effect of the two competitors at the high versus the two low SNRs was observed. At the positive SNR, better listening performance was observed in the multitalker babble. Although this is a small difference (cf. 86.8% versus 82.9%) and the effect size is trivial to small (= 0.14, Cohen[8]), this is somewhat unexpected as listeners were expected to have poorer performance in the informational masker. At the positive SNR, the listener can likely more easily separate the target talker and the competing speech into separate perceptual streams or objects and consequently identify the target speech (Kidd et al[26]; Brungart et al[6]; Durlach et al[12]; Freyman et al[15,16]). It may also be the case that the multitalker babble is more of an energetic masker than an informational masker, as it was constructed from six talkers. At poorer SNRs, separation of the target talker and the competing speech is more difficult, and performance significantly decreases relative to the continuous noise masker. Although the amount of informational masking in background decreases with increasing the number of talkers greater than two (e.g., Freyman et al[16]; Hoen et al[21]; Rosen et al[39]), competitors with six talkers (as used in this study) still provide greater masking efficiency relative to continuous broadband noise at poor SNRs (e.g., Miller[33]; Carhart et al[7]; Danhauer and Leppler[9]; Duquesnoy[11]; Festen and Plomp[14]; Simpson and Cooke[43]). In the case of the PRESTO sentences and competing multitalker babble, the similarity could be attributed to comparable acoustic characteristics (i.e., speech) and contextual information (i.e., semantic sentences). It may also be the case that differences in spectral content between the continuous noise and multitalker babble (i.e., greater low-frequency spectra <1000 Hz in the latter; see ►**Figure 1**) may provide more energetic masking at lower SNRs.

For normal-hearing young adult listeners with a Southern American English dialect, race and sex do not appear to affect PRESTO performance, in quiet or with background competition, at least for this sample of Southern AAE and WAE participants from North Carolina. The absence of a sex effect on speech perception in noise is consistent with previous reports (Ribera[38]; Sbompato et al[41]; Jacobi et al[24]). Tamati et al[54] found differences in indexical processing, short-term and working memory, vocabulary size, and some domains of executive functioning in listeners who perform above average versus those that perform below average. Tamati et al[54] posited that differences in indexical processing and neurocognition contribute to PRESTO performance variability among listeners. It is likely that the same factors contribute to performance variability in the cohort from this study. Furthermore, participants in this study have similar demographic variables and linguistic and dialect exposure. That is, their self-reported years of residence in the south, education, study abroad experiences, second language study, and exposure to regional accents of American English were not significantly different across race and sex groups. Comparison of performance variability from the participants in this study with those from previous studies using PRESTO materials (at least in multitalker babble) is difficult because of differences in lists used and different SNRs. We did not observe differences in variability across test conditions as a function of sex or race

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

(Levene's test of variance for equality, $p > 0.05$). Variability in scores varied across SNRs. It was largest with scores near 50% and was reduced when scores approached 0% and 100%. This is consistent with speech recognition scores modeled by a binomial distribution (Thornton and Raffin[58]).

Participants with a Southern American English dialect had similar PRESTO performance (i.e., scores within 5–15%) in multitalker babble and continuous broadband noise, with previously reported talkers residing in a geographically different region of the United States. For example, Gilbert et al[18] reported mean PRESTO performance scores of 70.8% and 37.6% in multitalker babble with 0 dB and −5 dB SNR, respectively. Participants demonstrated mean performance scores of 65.6% and 24.1% at the equivalent SNRs, herein. Slightly lower scores in this study may be attributed to differences in PRESTO sentences used between studies. In Phase I of their study, Gilbert et al[18] used ten PRESTO lists. The lists used in this study were 12 from Faulkner et al,[13] and it is likely that these were not the same as those used by Gilbert et al.[18] Faulkner et al[13] demonstrated that list equivalency (i.e., equal performance across lists) of the PRESTO materials is not evident under some conditions (e.g., multitalker babble at 0 dB SNR). Gilbert et al[18] reported an overall mean score (i.e., averaged across −5, −3, 0, and 3 dB SNRs) of 62.7%, with a range of 40.3–76.2%. Collapsed across all SNRs (i.e., −10, −5, 0, and 5 dB SNR), the mean score in this study was 44.6%, with a range of 29.5–54.2%. The lower mean score in the present study reflects the fact that the lowest SNR in this study significantly depressed overall mean performance. Plotkowski and Alexander[36] used 19 PRESTO lists presented in a competing background of speech-shaped steady-state noise at an SNR of −3 dB. Mean performance of 16 normal-hearing young adults also residing in Indiana, US, was 65.1%. This value falls between the mean scores of 32.0% and 68.4% in continuous noise at SNRs of −5 and 0 dB found in the present study reflecting similar performance across studies in steady-state energetic maskers.

Speech recognition in noise is an everyday occurrence and affects listening performance in a myriad of environments including school, work, and social environments. Identifying speech recognition disorders in noise is clinically important. In fact, the AAA,[1] ASHA,[3] and BSA[5] best practice guidelines recommend speech-in-noise testing as an essential component of a thorough adult audiologic evaluation. Speakers of different dialects perform differently on speech audiometry in quiet (Nissen et al[35]) and with tests of speech-in-noise when listening to recorded speech materials in a dialect of their own language that is not familiar (Liu and Shi[31]; Shi and Canizales[42]). Clinical audiologists, therefore, must be aware of the effect of patient dialect on speech recognition performance in noise. The findings of this study and others suggest that the PRESTO may be an attractive clinical alternative to traditional sentence recognition materials. The variability in target talker (i.e., dialect, sex, and multiple talkers) may create a level playing field for listeners with different dialects. Larger normative database studies with specific competition and SNRs would be warranted before any clinical implementation.

In conclusion, we examined PRESTO performance among listeners who were lifelong or near lifelong residents of a Southern American English dialect region from North Carolina. Specifically, the effect of noise competition, SNR, race, and sex was examined. As expected, performance improved with increasing SNRs. Different competing background noises and multitalker babble differentially affect PRESTO performance at equivalent SNRs. Performance was generally superior in interrupted noise and poorer in multitalker babble. Listener's race and sex do not affect PRESTO performance, at least for talkers with a Southern American English dialect residing in North Carolina examined in this study. Further research is warranted to see if the same findings are consistent with talkers from other homogenous dialect regions and if regional dialect of a talker predicts PRESTO performance in native speakers of American English. That is, performance between listeners from homogenous dialect regions should be compared to confirm that listeners with different regional dialects have similar PRESTO performance. Further refinement and standardization of test parameters (i.e., noise competitors and SNRs) are needed before any clinical application of the PRESTO.

Abbreviations

AAE	African American English
ANOVA	analysis of variance
FFTs	fast Fourier transforms
PRESTO	Perceptually Robust English Sentence Test Open-set
SD	standard deviation
SNR	signal-to-noise ratio
WAE	White American English

Notes
This work was presented in part at the 2018 American Speech-Language-Hearing Association Annual Convention, Boston, MA.

Conflict of Interest
None declared.

Acknowledgments
We are grateful for the provision of the PRESTO materials by Dr. David Pisoni and colleagues at the Speech Research Laboratory, Indiana University.

References

1 American Academy of Audiology (AAA). American Academy of Audiology Clinical Practice Guidelines: Diagnosis, Treatment and Management of Children and Adults with Central Auditory Processing Disorder. 2010. http://audiology-web.s3.amazonaws.com/migrated/CAPD%20Guidelines%208-2010.pdf_539952af956c79.73897613.pdf. Accessed September 14, 2019
2 American National Standards Institute (ANSI). Specification for Audiometers. (ANSI S3.6-2010). New York, NY: ANSI; 2010
3 American Speech-Language-Hearing Association (ASHA). Hearing Aids for Adults. https://www.asha.org/PRPSpecificTopic.aspx?folderid58589935381&section5Key_Issues#Assessment. Accessed September 14, 2019

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

4 Bess FH. Clinical assessment of speech recognition. In: Konkle DF, Rintlemann WF, eds. Principles of Speech Audiometry. Baltimore, MD: University of Park Press; 1983:125–201
5 British Society of Audiology (BSA). Practice Guidance Assessment of Speech Understanding in Noise in Adults with Hearing Difficulties. 2019. https://www.thebsa.org.uk/wp-content/uploads/2019/04/OD104-80-BSA-Practice-Guidance-Speech-in-Noise-FINAL.Feb-2019.pdf. Accessed September 14, 2019
6 Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am 2001;110:2527–2538
7 Carhart R, Johnson C, Goodman J. Perceptual masking of spondees by combinations of talkers. J Acoust Soc Am 1975;58:S35
8 Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, MI: Lawrence Erlbaum Associates; 1988
9 Danhauer JL, Leppler JG. Effects of four noise competitors on the California consonant test. J Speech Hear Disord 1979;44:354–362
10 Dirks DD, Wilson RH, Bower DR. Effect of pulsed masking on selected speech materials. J Acoust Soc Am 1969;46:898–906
11 Duquesnoy AJ. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. J Acoust Soc Am 1983;74:739–743
12 Durlach NI, Mason CR, Shinn-Cunningham BG, Arbogast TL, Colburn HS, Kidd G Jr.. Informational masking: counteracting the effects of stimulus uncertainty by decreasing targetmasker similarity. J Acoust Soc Am 2003;114:368–379
13 Faulkner KF, Tamati TN, Gilbert JL, Pisoni DB. List equivalency of PRESTO for the evaluation of speech recognition. J Am Acad Audiol 2015;26:582–594
14 Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am 1990;88:1725–1736
15 Freyman RL, Balakrishnan U, Helfer KS. Spatial release from informational masking in speech recognition. J Acoust Soc Am 2001;109:2112–2122
16 Freyman RL, Balakrishnan U, Helfer KS. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J Acoust Soc Am 2004;115:2246–2256
17 Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS, Dahlgren NL. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Philadelphia, PA: Linguistic Data Consortium; 1993
18 Gilbert JL, Tamati TN, Pisoni DB. Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. J Am Acad Audiol 2013;24:26–36
19 Green LJ. African American English: A Linguistic Introduction. New York, NY: Cambridge University Press; 2002
20 Helfer KS, Freyman RL. Aging and speech-on-speech masking. Ear Hear 2008;29:87–98
21 Hoen M, Meunier F, Grataloup CL, Pellegrino F, Grimault N, Perrin F, Perrot X, Collet L. Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Commun 2007;49:905–916
22 Holt YF. Mechanisms of vowel variation in African American English. J Speech Lang Hear Res 2018;61:197–209
23 Holt LL, Lotto AJ. Cue weighting in auditory categorization: implications for first and second language acquisition. J Acoust Soc Am 2006;119:3059–3071
24 Jacobi I, Sheikh MRashid, de Laat JAPM, Dreschler WA. Age dependence of thresholds for speech in noise in normal-hearing adolescents. Trends Hear 2017;21:2331216517743641
25 Keppel G, Wickens TD. Analytical comparisons among treatment means. In: Design and Analysis: A Researcher's Handbook. 4th ed. Upper Saddle River, NJ: Pearson Education; 2004:60–87
26 Kidd G Jr., Mason CR, Rohtla TL, Deliwala PS. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J Acoust Soc Am 1998;104:422–431
27 Labov W. Principles of Linguistic Change Volume 1: Internal Factors (Language in Society 20). Oxford, UK: Blackwell; 1994
28 Labov W, Ash S. Understanding Birmingham. In: Bernstein C, Nunnally T, Sabino R, eds. Language Variety in the South Revisited. Tuscaloosa, AL: University of Alabama Press; 1997:508–573
29 Labov W, Ash S, Boberg C. The Atlas of North American English: Phonetics, Phonology and Sound Change. Berlin, Germany: Walter de Gruyter GmbH; 2006
30 Labov W, Yaeger M, Steiner R. A Quantitative Study of Sound Change in Progress. Philadelphia, PA: US Regional Survey; 1972
31 Liu D, Shi LF. Performance-intensity functions of Mandarin word recognition tests in noise: test dialect and listener language effects. Am J Audiol 2013;22:147–156
32 Maxwell SE, Delaney HD, Kelley K. Individual comparisons of means. In: Designing Experiments and Analyzing Data: A Model Comparison Perspective. 3rd ed. New York, NY: Routledge; 2018:170–215
33 Miller GA. The masking of speech. Psychol Bull 1947;44:105–129
34 Miller GA, Licklider CR. The intelligibility of interrupted speech. J Acoust Soc Am 1950;22:167–173
35 Nissen SL, Harris RW, Channell RW, Richardson NE, Garlick JA, Eggett DL. The effect of dialect on speech audiometry testing. Am J Audiol 2013;22:233–240
36 Plotkowski AR, Alexander JM. A sequential sentence paradigm using revised PRESTO sentence lists. J Am Acad Audiol 2016;27:647–660
37 Pollack I. Masking by periodically interrupted noise. J Acoust Soc Am 1955;27:353–355
38 Ribera JE. Functional hearing in noise: insert earphones vs supra-aural headphones. Aviat Space Environ Med 2007;78:1159–1161
39 Rosen S, Souza P, Ekelund C, Majeed AA. Listening to speech in a background of other talkers: effects of talker number and noise vocoding. J Acoust Soc Am 2013;133:2431–2443
40 Roup CM, Wiley TL, Safady SH, Stoppenbach DT. Tympanometric screening norms for adults. Am J Audiol 1998;7:55–60
41 Sbompato AF, Corteletti LC, Moret LAde. Hearing in noise test Brazil: standardization for young adults with normal hearing. Rev Bras Otorrinolaringol 2015;81:384–388
42 Shi LF, Canizales LA. Dialectal effects on a clinical Spanish word recognition test. Am J Audiol 2013;22:74–83
43 Simpson SA, Cooke M. Consonant identification in N-talker babble is a nonmonotonic function of N. J Acoust Soc Am 2005;118:2775–2778
44 Stuart A. An investigation of list equivalency of the north-western university auditory test no 6 in interrupted broadband noise. Am J Audiol 2004;13:23–28
45 Stuart A. Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. Ear Hear 2005;26:78–88
46 Stuart A. Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children. J Am Acad Audiol 2008;19:135–146
47 Stuart A, Butler AK. Contralateral suppression of transient otoacoustic emissions and sentence recognition in noise in young adults. J Am Acad Audiol 2012;23:686–696
48 Stuart A, Butler AK. No learning effect observed for reception thresholds for sentences in noise. Am J Audiol 2014;23:227–231
49 Stuart A, Givens GD, Walker LJ, Elangovan S. Auditory temporal resolution in normal hearing preschool children revealed by word recognition in continuous and interrupted noise. J Acoust Soc Am 2006;119:1946–1949
50 Stuart A, Phillips DP. Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. Ear Hear 1996;17:478–489
51 Stuart A, Phillips DP. Deficits in auditory temporal resolution revealed by a comparison of word recognition under interrupted and continuous noise masking. Semin Hear 1998;19:333–344

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

52 Stuart A, Phillips DP, Green WB. Word recognition performance in continuous and interrupted broad-band noise by normal-hearing and simulated hearing-impaired listeners. Am J Otol 1995; 16:658–663
53 Stuart A, Zhang J, Swink S. Reception thresholds for sentences in quiet and noise for monolingual English and bilingual Mandarin-English listeners. J Am Acad Audiol 2010;21:239–248
54 Tamati TN, Gilbert JL, Pisoni DB. Some factors underlying individual differences in speech recognition on PRESTO: a first report. J Am Acad Audiol 2013;24:616–634
55 Tamati TN, Pisoni DB. Non-native listeners' recognition of high-variability speech using PRESTO. J Am Acad Audiol 2014; 25:869–892
56 Theunissen M, Swanepoel de W, Hanekom J. Sentence recognition in noise: variables in compilation and interpretation of tests. Int J Audiol 2009;48:743–757
57 Thomas ER. Spectral differences in/ai/offsets conditioned by voicing of the following consonant. J Phonetics 2000;28:1–25
58 Thornton AR, Raffin MJ. Speech-discrimination scores modeled as a binomial variable. J Speech Lang Hear Res 1978;21:507–518
59 Wagenaar VA. Note on the construction of digram-balanced Latin squares. Psychol Bull 1969;72(06):384–386
60 Willis C. Perception of vowel phonemes in Fort Erie, Ontario, Canada, and Buffalo, New York: an application of synthetic vowel categorization tests to dialectology. J Speech Lang Hear Res 1972; 15:246–255

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.