Short and Long Compression Release Times: Speech Understanding, Real-World Preferences, and Association with Cognitive Ability

DOI: 10.3766/jaaa.21.2.6

Robyn M. Cox*
Jingjing Xu*

Abstract

Background: Several previous investigations have explored the relationship between cognitive abilities and speech understanding with short and long hearing aid compression release times in adult hearing aid wearers. Although there was consensus that such a relationship exists, the details have not been consistent across studies. This investigation was designed to further explore, extend, and generalize this topic.

Purpose: Questions addressed: (1) the association between cognitive abilities and speech understanding with short and long release times for more ecologically valid speech than in previous studies, (2) the presence or absence of release time preferences in daily life, and (3) the association between the laboratory measures and real-world release time preferences.

Research Design: The study was a blinded randomized crossover trial.

Study Sample: Twenty-four subjects with mean age of 72 yr completed the study. They were experienced users of bilateral amplification with mild to moderately-severe symmetrical hearing losses. They were healthy, English-speaking active seniors recruited using advertisements and letters.

Data Collection and Analysis: There were five test sessions. They included audiometric and cognitive testing, fitting bilateral Oticon Adapto hearing aids, a four-week trial with either short or long release time, outcome measures, a four-week trial with the other release time, further outcome measures, and a final interview.

Results: Taken together with the previous studies, the results suggest that compression processing release time is more critical for patients with lower cognitive abilities than for those with higher cognitive abilities. Further, we postulate that the best release time for listeners with lower cognitive abilities depends on the redundancy of the tested speech. Those with lower cognitive abilities might benefit from short release time when contextual speech is used, or when speech is rich in context, release time might not be important; however, when speech is low in semantic context, listeners with lower cognitive abilities might require long release time for best performance. Listeners do appear to be able to distinguish between long and short release time processing in daily life listening. However, release time preference was not predictable from cognitive abilities or aided measures of speech understanding. About two-thirds preferred long release time, and one-third preferred short release time.

Conclusion: The relationship between cognitive abilities and performance with short and long release time processing was supported and further elucidated in this research. In addition, release time was seen to be a salient variable in subjective performance with amplification in daily life. Accurate prospective prescription of release time has the potential to make a material contribution to successful amplification provision.

Key Words: Cognitive, compression, hearing aid, outcomes, speech understanding

*School of Audiology and Speech-Language Pathology, University of Memphis

Robyn M. Cox, School of Audiology and Speech-Language Pathology, University of Memphis, 807 Jefferson Ave., Memphis, TN 38105; E-mail: robyncox@memphis.edu

This work was supported by Eriksholm Research Center, Denmark.

Abbreviations: APHAB = Abbreviated Profile of Hearing Aid Benefit; AT = attack time; BKB-SIN = Bamford-Kowell-Bench Speech in Noise; DOSO = Device Oriented Subjective Outcome; HAPQ = Hearing Aid Performance Questionnaire; ICRA = International Collegium of Rehabilitative Audiology; LNS = Letter–Number Sequencing; NAL-NL1 = National Acoustics Laboratory–Nonlinear Version 1; PI = performance intensity; RT = release time; SNR = signal-to-noise ratio; SPAC = Speech Pattern Contrast test; VLM = Visual Letter Monitoring test; WCST = Wisconsin Card Sorting Test; WDRC = wide dynamic range compression

Modern hearing aids almost always include the capability for wide dynamic range compression (WDRC) processing. WDRC processing can function in many different ways based on the settings of basic parameters. These basic parameters include the attack time (AT) and the release time (RT). AT is a measure of the speed at which device gain is decreased after the input sound level is raised. RT is a measure of the speed at which device gain is increased after the input sound level is reduced. Most modern hearing aids use short AT (less than 20 msec) so that sudden loud sounds will not be amplified to uncomfortably loud levels. In the study reported here, the AT was always short. Although there are widely circulated theoretical rationales for selecting shorter or longer release times in WDRC processing (e.g., Dillon, 2001), there is still relatively little evidence about the effectiveness of these choices in terms of the success of the amplification system in addressing the everyday problems of the hearing-impaired listener. In addition, generic prescription methods such as the National Acoustics Laboratory–Nonlinear Version 1 (NAL-NL1) and the Desired Sensation Level input/output procedure do not include guidelines for choosing RT (Byrne et al, 2001; Scollie et al, 2005). Thus, practitioners and manufacturers have limited evidence-based guidance about when, or for whom, WDRC devices should be configured for short RT versus long RT.

There is no clear consensus about exactly what time intervals constitute short and long RT. However, release times in the range of 10–100 msec would generally be considered short. Release times greater than 500 msec would generally be considered long. Conventional wisdom suggests that short RTs have some potential advantages. These include improved audibility of soft consonants in ongoing speech and more normalized loudness perceptions for both low- and high-level everyday sounds. There are also putative disadvantages. The most frequently cited are an increased perceived noisiness when speech occurs in moderate background noise and reduction of intensity cues that are important for speech understanding. Long RT also is reputed to have positive and negative impacts. On the positive side, long RT is claimed to maintain the intensity relationships among speech sounds, thus preserving their naturalness, intelligibility, and localizability. On the negative side, long RT can result in brief intervals of very low gain during which important auditory events are missed. There are other proposed consequences of short and long release times as well as different combinations of AT and RT. This topic has been thoroughly explored in Dillon (2001) and Moore (2008), among others.

Based on these considerations, it would appear that selection of appropriate RT should be an essential and systematic aspect of hearing aid provision. In attempts to provide a scientific underpinning for prescription of RT, numerous investigations have been undertaken and published regarding the advantages and disadvantages of short and long release times for speech intelligibility. When studies used multichannel compressors and compression parameters (compression threshold, compression ratio, and attack and release times) that would be found in modern hearing aids, the results typically have not shown a statistically significant difference between short and long RT (e.g., Novick et al, 2001; van Toor and Verschuure, 2002; Moore et al, 2004; Jenstad and Souza, 2005; Shi and Doherty, 2008). Taken as a whole, this body of literature has not established overall superiority for either type of RT processing under any listening conditions.

Gatehouse, Naylor, and Elberling (2006a) noted three potential limitations in existing studies of the benefits of short and long release times: (1) studies tended to have low power, in that there were few subjects and relatively few outcome measures; (2) studies might not have provided sufficient acclimatization time for subjects to maximize their ability to use cues provided by new processing schemes; and (3) studies tended to seek overall superiority for either short or long RT when, in fact, the best processing might differ across individuals. Gatehouse and colleagues addressed these concerns in a study in which hearing aids with short and long release times were compared in laboratory and real-world settings. Fifty subjects with mostly mild–moderate hearing loss were fitted unilaterally and wore each hearing aid for a 10-week acclimatization period. An extensive set of speech-intelligibility scores (closed-set phoneme level) and self-report data were obtained. Results of group-level analyses indicated that (1) the hearing aids with short release time yielded slightly (but significantly) more speech-intelligibility benefit, (2) reported real-world listening comfort was significantly higher with long release times, and (3) real-world satisfaction was essentially equal with both release times. These results indicated that the superiority of short or long RT is dependent on the outcome measurement domain.

Gatehouse and colleagues (2006a) further noted that there were substantial individual differences in both the intelligibility and comfort domains of outcome. Despite the group-level advantage for short RT in terms of speech understanding, some subjects had better understanding with long RT. Likewise, despite the group-level advantage for long RT in listening comfort, some subjects clearly preferred the comfort provided by short RT processing. In an attempt to explain the bases for individual differences with short and long RT, Gatehouse and colleagues (2006b), in a separate part of the same experiment described above, explored the contribution of several subject characteristics including audiologic, ecologic, and psychoacoustic variables. One of the variables that emerged that explained roughly 10–30% of the variance in intersubject differences in benefit with long and short RT was named Cognitive Abilities. Its value was derived from a test involving memory and speed of processing for visually presented letters or digits. This test has subsequently been called the Visual Letter Monitoring Test (VLM). The Gatehouse and colleagues (2006b) data suggested that individuals with better scores on the VLM (and, by implication, higher cognitive abilities) were likely to benefit more from short release time processing in a hearing aid, whereas those with poorer VLM scores were likely to benefit more from long RT processing. For subjects with high cognitive abilities, the superiority of short RT was especially evident in listening situations with modulated maskers.

The notion of a relationship between cognitive abilities and benefit from short or long RT is appealing because it is compatible with other research that has shown a relationship between speech perception and cognitive abilities (e.g., Daneman and Merikle, 1996; Wingfield and Tun, 2001; Pichora-Fuller, 2007). It is now widely accepted that cognitive abilities such as working memory and speed of processing are related to speech perception in difficult listening situations: individuals with higher cognitive abilities obtain better speech-perception scores (e.g., Lunner, 2003; Humes, 2007; Rudner et al, 2008). In addition, if a link can be established between a measure of cognitive abilities and benefit from fast versus slow compression processing, this offers the potential of developing a clinical protocol for prescribing compression RT on an individual basis. It is not surprising, therefore, that other researchers have explored this topic further.

Lunner and Sundewall-Thorén (2007) reported a study that was designed to duplicate the Gatehouse and colleagues (2006b) results and to extend them to a different language (Danish). In the study, 23 experienced hearing aid wearers were recruited and used their own hearing aids (which were technologically the same as those used by Gatehouse et al). Twelve subjects were bilaterally aided. The subjects' own hearing aids were reprogrammed to the short and long RT conditions in a crossover design. Ten-week acclimatization periods were provided for each RT setting. Laboratory speech-recognition ability was measured using a Danish test with five-word low-context sentences presented in both modulated and steady-state noise. An adaptive procedure was used in which signal-to-noise ratio (SNR) was varied based on the subject's performance. A performance-intensity function was constructed for each test condition. The same Visual Letter Monitoring Test (adapted to Danish) was used to quantify cognitive abilities, and subjects were partitioned into low, moderate, and high cognitive performance subgroups. The results revealed that the subjects with higher cognitive scores performed better than those with lower cognitive scores with both short and long RT processing. However, speech score differences between cognitive groups were quite small with long RT processing in a steady-state masker but much larger with short RT processing in a modulated noise masker. Further, there was a significant moderate relationship between cognitive score and the advantage of fast-acting compression compared to slow-acting compression with a modulated noise masker. Overall, Lunner and Sundewall-Thorén duplicated the findings of Gatehouse and colleagues (2006b).

Foo and colleagues (2007) reported a study that further extended this line of inquiry. In this study, 32 experienced hearing aid wearers were recruited as subjects and used their own hearing aids, which were technologically identical to those in the two previous studies (Gatehouse et al, 2006a, 2006b; Lunner and Sundewall-Thorén, 2007). The hearing aids were reprogrammed to the short and long RT conditions, and subjects were immediately tested using these conditions (i.e., no acclimatization time was allowed). Performance was measured with two Swedish-language speech tests: one with five-word low-context sentences similar to those used by Lunner and Sundewall-Thorén and the other with more natural sentences. Two maskers were used with each speech test: one steady state and one modulated. Cognitive abilities were measured with a Swedish version of the Visual Letter Monitoring Test as well as with a second test. The second test (the Reading Span test) is similar to the VLM in calling for simultaneous memory storage and semantic processing. However, the Reading Span test calls for a greater quantity of memory storage and semantic processing than the VLM. Subjects were split into low and high cognitive ability groups. Data analyses revealed that cognitive ability as measured by the Reading Span test was significantly associated with performance on the speech-intelligibility tests in all listening conditions. However, cognitive ability as measured by the VLM was not significantly associated with speech recognition in noise. Further, in the eight comparisons of performance with short versus long RT processing, only two yielded significant differences. The differences between short and long release times occurred for the subjects with lower cognitive scores (not for those with higher cognitive scores), and the results were the same for steady-state and modulated maskers. In addition, categorizing subjects' cognitive ability using the Reading Span test yielded one pattern of differences between short and long RT conditions, whereas categorizing cognitive ability using the VLM yielded a different pattern. Thus, the results of Foo and colleagues (2007) provide limited support for previous findings regarding the relationship between cognitive abilities and the advantage of short or long RT and suggest a more complex relationship pattern.

Although all the studies described above are consistent with the existence of a relationship between cognitive abilities and performance with short or long release time compression processing, the details are different. The specific results for short and long RTs appear to be sensitive to several variables, including the test used to classify subjects into cognitive groups, the speech-intelligibility test used to determine subject performance with compression processing, and, perhaps, the modulation characteristics of the masker used or acclimatization to the hearing aids.

The study described in this article was designed to further explore, extend, and generalize this topic. In this research, a set of tests was used to assess subjects' cognitive abilities and a composite cognitive score was derived. The two speech-understanding tests used to quantify laboratory performance with short and long RT processing were chosen with a view to including more ecologically valid stimuli. Finally, the hearing aid worn by subjects was a different type from the single type used in all the previous studies. In addition, previous work did not explore real-world preferences for short or long RT processing. In the present study, that issue was considered: Each subject was asked whether he or she preferred to use short or long RT processing in daily life, and several questionnaires were completed to determine real-world subjective performance with short and long RT conditions.

The specific research questions were as follows:

1. What is the association between cognitive abilities and performance with short and long RT processing when cognitive abilities are determined using a composite cognitive score and speech-intelligibility tests exhibit greater ecological validity than most of those used in previous studies?
2. Do individuals have a clear preference for short or long RT in real-world listening?
3. If so, can that preference be predicted accurately based on either individual cognitive status or aided speech-intelligibility scores obtained in the laboratory?

METHOD

The overall plan of the study was to recruit experienced hearing aid wearers who responded to a battery of cognitive tests. Each subject was fitted bilaterally with new hearing aids with short AT (nominally 10 msec) and capable of both long and short RT processing. Each subject experienced both short (nominally 40 msec) and long (nominally 640 msec) RT processing in daily life in a randomized crossover design. An acclimatization period of four weeks was allowed for the first RT setting. During the fourth week the subject responded to several self-report questionnaires regarding real-world experiences with the RT setting. Next, laboratory testing of speech recognition in noise was performed. The second RT setting was then established, and the trial process was repeated. Finally, each subject was interviewed about his or her overall preference in daily life for the first or second RT condition experienced. Subjects were paid for their participation.

Power

The field trial was planned for a group of 30 subjects, which provided 81% power ($\alpha = 0.05$) to detect a small-to-medium effect of RT on laboratory speech-understanding scores. Unexpected subject attrition (see below) increased the minimum detectable effect size, but it remained within the small-to-medium range. An additional 26 subjects were recruited to provide cognitive data only. This yielded a total N of 50 to support the principal components analysis planned on the cognitive scores (see below).

Blinding

Subjects were blinded: they were not informed about the specific purpose of the study or about the order of test conditions. In addition, the scoring of laboratory speech-understanding tests was blinded in the following way: the researcher who administered the test scored it as usual. At the same time, the subject's spoken responses were recorded, and these recordings were later scored by another researcher who was blinded to subject and RT condition. The few scoring discrepancies were resolved between the two scorers by listening to the recordings.

Subjects

Volunteers were recruited via newspaper advertisements and letters to previous clinic patients or research subjects. Inclusion criteria were as follows: postlingual bilateral hearing impairment that was essentially symmetrical between ears (pure-tone average differences ≤15 dB); sensorineural hearing loss with minimal conductive overlay (mean air–bone gap ≤15 dB, or normal immittance test results); thresholds in the range of 25–80 dB HL in both ears from 500 Hz to 3 kHz; sloping (5–20 dB/octave) or flat (±5 dB/octave) audiogram configuration for both ears from 500 Hz to 3 kHz; current wearer of bilateral hearing aids (any type) with self-reported use at least 4 hr per day and at least 3 mo experience with current devices; fairly active lifestyle, based on interview and questionnaire (to provide a variety of suitable situations for self-report data); English as the first language (necessary for the speech-in-noise tests); self-rated health (physical and mental) of good or excellent; and adequate literacy and cognitive competence to complete informed consent and required questionnaires (assessed through interview and reading task). Exclusion criteria were history of otologic surgery or chronic middle ear or outer ear pathology, evidence of retrocochlear involvement, fluctuating hearing loss, and known psychiatric or neurologic disorder. About 200 individuals were evaluated as potential subjects. Of these 32 met all requirements and began the study. Eight of these dropped out due to scheduling conflicts, health problems, or inadequate hearing aid fit. Twenty-four persons completed the protocol. There were 17 men and 7 women. Ages ranged from 41 to 89 yr (mean = 71.8, SD = 11.5). Figure 1 depicts the composite audiograms for the men and the women.

There was a second group of subjects comprising 26 individuals who were used only for the purpose of exploring the relationships among the cognitive tests. All of these subjects were hearing impaired with audiograms similar to the main group. There were 18 men and 8 women with ages from 35 to 91 (mean = 65.2, SD = 10.9).

Procedure

The subjects in the second group described above were not part of the field trial. These persons were seen for one test session and responded only to the five cognitive tests described below.

For the 24 subjects in the field trial, there were five sessions of data collection. Sessions 1 and 2 were devoted to obtaining background information and audiometric tests and exploring the subject's unaided speech-recognition abilities. Unaided speech recognition was measured using the Speech Pattern Contrast (SPAC) test (see below) in quiet and in both modulated and unmodulated noises. In addition, each subject's cognitive abilities were estimated through the administration of five cognitive tests (details below).

Session 3 was devoted to bilateral hearing aid fitting using Oticon Adapto hearing aids. Details of the hearing aid fittings are given below. The RT was programmed for either short or long processing for the first trial. By random selection, 14 subjects received

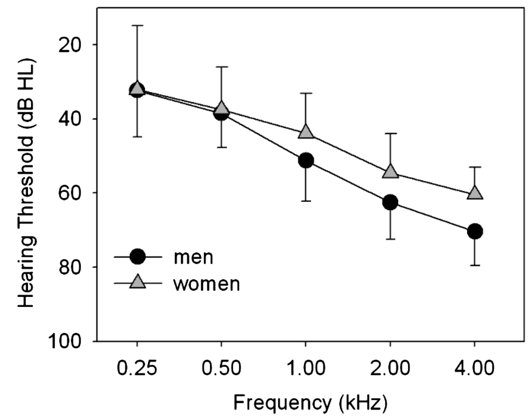

Figure 1. Composite audiograms for the men and the women who served in the field trial ($N = 24$).

the long RT for the first trial condition, and 10 subjects received the short RT. The subject was instructed about the use and care of the hearing aids and activities for the four-week real-world trial. At this time, the subject was given a booklet with the three outcome questionnaires (see below) but was instructed not to complete the questionnaires until the fourth week of the field trial. A follow-up telephone call was made to the subject two days later, and, if necessary, fine-tuning adjustments were made and the four-week trial was restarted. No further fine-tuning adjustments were allowed for the duration of the study.

In session 4, the completed self-report questionnaires regarding experiences with the first RT condition were collected. Aided speech recognition was then measured using the SPAC test and using the Bamford-Kowell-Bench Speech in Noise (BKB-SIN) test (see below). The second RT condition was then programmed into the hearing aids, and another four-week trial was begun. Once again, the subject was given a booklet with the three outcome questionnaires and was instructed not to complete the questionnaires until the fourth week of the field trial.

In session 5, the completed self-report questionnaires regarding experiences with the second RT condition were collected. Aided speech recognition was then measured using the SPAC test and the BKB-SIN test. Finally, the subject was interviewed about his or her real-world preference for the first or second hearing aid trial condition and was asked to give his or her level of confidence in that decision.

Cognitive Tests

The five cognitive tests chosen for this study involved elements of working memory, processing capacity, and/or processing speed. Consideration was given to the literacy demands of the tests with the goal that reading ability would not be a factor in test scores. Even though only one of the tests involved spoken stimuli, all tests involved spoken instructions. Subjects wore their hearing aids for all tests:

1. The Visual Letter Monitoring Test described by Gatehouse and colleagues (2003) contains elements of both memory and processing speed. The Gatehouse version of the test was used in this research with minimal changes to account for American rather than U.K. English. In this test, a stream of single letters, alternating vowels and consonants, appears on a computer screen at 1 sec (or 2 sec) intervals. The subject monitors the letters. When three consecutive letters form a word, the subject responds by pressing the space bar. To be counted as correct, the response must occur before the next letter appears. In a stream of 151 letters, there are 20 correct words. Subjects completed a 1 sec interval list and a 2 sec interval list. Scoring was determined in terms of d-prime values (to account for false alarms as well as correct responses). The scores for the two lists were combined into a single score.
2. The Wisconsin Card Sorting Test (WCST) measures cognitive flexibility and problem solving (Heaton, 1981). The subject is shown four response cards and one key card. The task is to match the key card to one of the response cards. No instructions are given, but right/wrong feedback is provided after each key card trial. The criterion for a correct match is one of three characteristics (symbol, color, number). The criterion changes as the test progresses. The test is not timed. The WCST was scored in terms of percent of perseverative responses (a perseverative response is one that is incorrect but would have been correct in the previous stage of the test).
3. The Letter–Number Sequencing (LNS) subtest from the *Wechsler Adult Intelligence Scale*, third edition (1997), quantifies working memory and processing capacity. This is the only test that involved spoken stimuli. The tester says a group of letters and numbers. The subject's task is to remember the group and repeat the items reordered in terms of numbers first (low to high) and then letters (alphabetically). The number of stimuli increases with successive trials. The score was based on the maximum number of stimuli correctly recalled and reordered.
4. The Visual Rhyme Test described by Hällgren and colleagues (2001) and Lunner (2003) measures speed of phonologic processing. An English version of the test was composed for this study. In this test, a pair of words is presented on the computer screen. There are four types of pairs: 12 pairs look similar and do rhyme (PEAK–LEAK), 13 pairs look similar and do not rhyme (GREAT–MEAT), 12 pairs look different and do rhyme (HYMN–RIM), and 13 pairs look different and do not rhyme (RING–HOT). For each pair, the subject's task is to press a "YES" key or a "NO" key to indicate whether the words rhyme or not and to make this response as quickly as possible. The 50 pairs of words are presented in random order. Scoring was in terms of the response time per correct response for the 50-pair list.
5. The Stroop (1935) test is a general measure of cognitive flexibility and control of the effects of cognitive interference. The version of the test used in this study compared the time taken in two types of tasks: (1) saying the colors of colored dots (time1) and (2) saying the colors in which color words are printed, for example, the correct response to the word *RED* printed in blue would be "blue" (time2). Because of the cognitive interference between the word and the color, the second type of stimulus is more difficult. The test was scored in terms of the relative amount of interference ([time2 – time1]/time1 [Uttl and Graf, 1997]).

To simplify interpretation, all cognitive test scores were expressed in a form in which a higher score indicated better performance. This involved transformation of scores for the Wisconsin Card Sorting, Visual Rhyme, and Stroop tests because in the customary scoring for those tests a higher score indicates poorer performance.

Speech-Recognition Tests

In previous studies of the relationships among cognitive ability, compression release times, and speech recognition, the speech-recognition tests featured speech stimuli generated by highly intelligible talkers. Thus, it could be argued (as noted by Foo et al) that the results might be different if the talkers produced speech that was ecologically more valid. In this study, we sought to use tests that were relatively closer to speech encountered in daily life.

The Speech Pattern Contrast Test

The Speech Pattern Contrast (SPAC) test was developed by Boothroyd (1985). It is a four-alternative forced-choice test using familiar monosyllabic words. Each response is scored for two different phonetic contrasts. We used three segmental subtests, each consisting of 12 test words and yielding two contrast scores. The outcome of the SPAC test was a composite score computed as the average of six consonant contrast scores: initial consonant voicing, final consonant voicing, initial consonant continuance, final consonant continuance, initial consonant place, and final consonant place.

Recordings of the SPAC test by 12 different talkers (who all produced speech that was judged to be of normal intelligibility) were generated at the University of Memphis Hearing Aid Research Laboratory for a previous study (Cox et al, 1987). By design, the recordings were made as ecologically valid as possible. In the recordings, each SPAC key word was embedded in one of 12 carrier sentences, such as "can you find ____ now," "identify ____ please," and "I'd like you to make ____ your next choice." These carrier sentences were devised to present the items in a variety of natural contexts with respect to preceding and following phonemes, position of test item in the utterance, and length of utterance. For each subtest, the 12 sentences were randomly assigned to the 12 test items. Each talker recorded a different combination of four forms. A complete form comprised 36 test items (3 subtests × 12 test words). For the present study, two talkers were used (talker #1 and talker #4). It had been empirically determined that talker #1 was less intelligible than talker #4. Two forms were used for each talker in each test condition.

The SPAC test was administered with two maskers from the International Collegium of Rehabilitative Audiology (ICRA) compact disk (Dreschler et al, 2001). The noises were an unmodulated speech-shaped noise (ICRA CD, track 1) and a modulated noise (ICRA CD, track 6), which was a speech-shaped noise modulated by two-talker babble. The speech was presented in a sound-treated room at 65 dB SPL at an SNR of 0 dB. It was determined in advance that this SNR resulted in a composite score that was about 70–80% of the subject's score in quiet. For statistical treatment, the SPAC percent correct scores were transformed into rationalized arcsine units (rau) to homogenize the variance (Studebaker, 1985). In summary, for each RT condition, the SPAC test was administered for two talkers and two maskers. Each administration comprised two test forms and yielded a composite score.

The Bamford-Kowell-Bench Speech in Noise Test

The BKB-SIN test (Etymotic Research, 1985) consists of high-predictability sentences presented with a modulated masker of four-talker babble. The portion of the test used in this study was composed of eight pairs of lists. Four pairs of lists were used in each RT condition. Each list comprises 10 sentences. The first sentence is presented at an SNR of 21 dB. Each successive sentence is presented at a 3 dB poorer SNR. The listener's task is to repeat each sentence. There are a total of 31 key words per list. The test was presented in the sound-treated room at a level of 83 dB SPL (a "loud but OK" level). Scoring was in terms of the percent correct key words at each SNR, and these data were used to produce a performance-intensity function for each RT condition.

Hearing Aid Fittings

The Oticon Adapto hearing aids used in this study were chosen because they employed more advanced signal processing than the devices used in the previous studies while retaining the ability to switch between short and long RT processing. They were behind-the-ear-style instruments fitted with vented earmolds and an active feedback cancellation system. They were two-channel wide dynamic range compression processors with input compression thresholds in both channels of 50 dB SPL. Two programs were available, but they were programmed identically. The microphone was omnidirectional. The digital noise-management algorithm was disabled to ensure that there would be no interaction between it and the RT processing in noise. There was no user volume control. For fitting, the hearing aids were programmed using the manufacturer's proprietary method. Immediate fine tuning was performed based on the subject's evaluation of own voice quality, speech clarity, and acceptability of loud environmental noises.

Figure 2 illustrates the mean fitted real-ear aided responses for soft and loud speech compared with the mean threshold and the mean prescription for loud speech from the NAL-NL1 method (Byrne et al, 2001). The figure shows results for the left ear; the right ear was essentially identical. On average, soft speech (55 dB SPL) was amplified to a level about 5 dB above the subject's thresholds, but the audibility of soft speech was very limited above 2000 Hz. Loud speech (75 dB SPL) was amplified to a level similar to the NAL-NL1 target for that level through the midfrequency region (750–2000 Hz). In addition, average maximum output levels (average of 500, 1000, and 2000 Hz) were compared with the optimum maximum output level

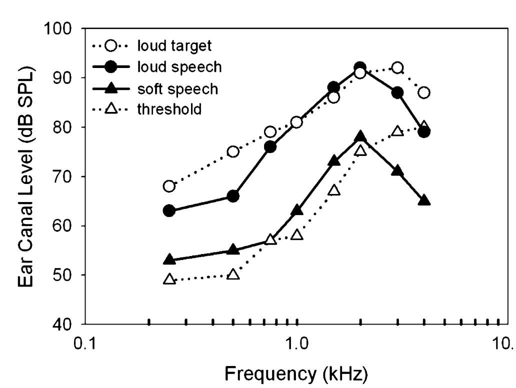

Figure 2. Mean real-ear aided responses for soft and loud speech for the hearing aids used in the field trial. Soft speech is compared with mean thresholds, and loud speech is compared with mean target values from the NAL-NL1 prescription method (Byrne et al, 2001). Data are for the left ear. The right ear data were essentially identical.

prescribed using the Dillon and Storey (1998) method. For both left and right ear fittings, the mean three-frequency average maximum power output level was 8.3 dB lower than the mean Dillon and Storey prescribed level.

For both short and long RT conditions, attack and release times were measured using the Fonix 6500 hearing aid test system. For the short RT condition (nominally AT/RT = 10/40 msec), the average measured values were AT = 10 msec, RT = 30 msec. For the long RT condition (nominally AT/RT = 10/640 msec) the average measured values were AT = 14 msec, RT = 420 msec. These test box measures were not equivalent to the engineering methods used to officially specify the circuit attack and release times. However, the test box measures verified the consistently short AT and the substantial difference between the short and long RT values.

Outcome Questionnaires

Subjects completed three outcome questionnaires at the end of each four-week trial: the Abbreviated Profile of Hearing Aid Benefit (APHAB [Cox and Alexander, 1995]), the Hearing Aid Performance Questionnaire (HAPQ [Gatehouse et al, 2006a]), and the Device Oriented Subjective Outcome (DOSO) Scale (Cox et al, 2009).

The APHAB comprises 24 items that produce scores for four subscales. Three subscales (Ease of Communication, Reverberation, Background Noise) quantify speech communication problems in different listening situations. The fourth subscale (Aversiveness) quantifies problems with environmental sounds. The APHAB was scored to measure the difference between unaided and aided scores. For the speech communication subscales, this is typically a benefit (improved score), whereas the Aversiveness subscale usually registers a penalty (worse score) because amplified sounds are more aversive than unamplified sounds.

The HAPQ is composed of 26 items yielding three subscale scores. One subscale (Speech Variations) concerns hearing aid benefit in situations where speech varies (in level, speed, person, etc). The two other subscales (Environmental Sounds and Intense Sounds) measure benefit/penalty for amplified environmental sounds and amplified intense sounds, respectively.

The DOSO consists of 28 items producing six subscale scores. One subscale (Speech Cues) quantifies improvements in speech understanding. The other subscales address other aspects of the hearing aid wearer's experiences (Listening Effort, Pleasantness, Quietness, Convenience, and Use). In addition, two of the DOSO subscales (Speech Cues and Listening Effort) have two equivalent forms. For this study, both forms were used. Thus the expanded DOSO questionnaire was 40 items in length.

RESULTS

Producing a More Comprehensive Cognitive Score

One goal of this investigation was to quantify cognitive ability in a more comprehensive manner than in the studies of Gatehouse and colleagues (2006a), Lunner and Sundewall-Thorén (2007), and Foo and colleagues (2007). In the first two of these studies, only the VLM was used to determine cognitive skills. Foo and colleagues used an additional test but did not combine the two cognitive tests in any way. Since the scores of the cognitive tests used by Foo and colleagues were only moderately related to each other, this resulted in the problematic situation in which different individuals were defined as having high and low cognitive abilities depending on the particular test used as a criterion.

To avoid this in the present study and to determine a more general definition of cognitive ability and consistent allocation to cognitive groups, preliminary analyses were performed to explore the relationships among the five cognitive tests and to determine how to combine some or all of them for a more comprehensive measure. For this purpose the data were used from the 50 subjects described above (24 from the field trial and 26 additional subjects). Table 1 depicts the linear correlation coefficients among the five cognitive tests. All statistical analyses were performed using SPSS Version 14 software. The Wisconsin Card Sorting Test, the Letter–Number Sequencing test, and the Visual Letter Monitoring Test have the closest consistent interrelationships, indicating that there is considerable commonality among them. Nevertheless, the modest magnitude of the correlations reveals that the three tests also differ to some extent in the cognitive skills that they assess. The Visual Rhyme test and the Stroop test tended to have lower correlations with other tests, suggesting that they function more in separate cognitive domains. Principal component analyses also were carried out on the cognitive data, but they did not yield any additional insights.

Consideration also was given to whether scores from each cognitive test appeared to be associated with speech-understanding ability. Although there have

Table 1. Correlation Coefficients among the Cognitive Tests ($N = 50$)

	LNS	WCST	VLM	Visual Rhyme
WCST	0.36*			
VLM	0.40**	0.45**		
Visual Rhyme	0.18	0.28*	0.35*	
Stroop	0.33*	0.05	0.11	0.20

*significant at .05 level (two tailed).

**significant at .01 level.

been numerous reports of associations between cognitive scores and speech understanding, this result is not always found (for a review, see Akeroyd, 2008). This matter was explored in the present study using data from the group of subjects in the field trial and all five SPAC scores obtained (unaided–quiet, unaided–unmodulated noise, unaided–modulated noise, aided–unmodulated noise, aided–modulated noise). It was important to establish a clear difference between cognitive groups; therefore, for each cognitive test, the eight individuals with the highest scores were compared to the eight individuals with the lowest scores (except for the LNS, for which the low cognitive score group was composed of 14 individuals due to the presence of a subset with tied scores). The procedure of discarding the subjects who lie close to the median score produced maximally different low- and high-performing cognitive groups and increased the likelihood of detecting differences due to the cognition variable. Thus, although there was a loss of power due to reduced N, there was an offsetting increase in power due to the strong separation of groups. The main effect of cognitive group was statistically significant for the VLM ($F[1, 14] = 5.25$, $p = .038$), the WCST ($F[1, 14] = 4.8$, $p = .046$) and the LNS test ($F[1, 20] = 4.78$, $p = .04$). The differences between low and high cognitive groups were not significant at the $p = .05$ level for the Visual Rhyme test or for the Stroop test. It is of interest to note that the three cognitive tests that were most closely associated with speech understanding were the ones that included a memory element as well as a cognitive processing element. This is consistent with observations of previous researchers (see Akeroyd, 2008). Figure 3 illustrates the average SPAC test scores for each cognitive test for subjects with higher and lower cognitive scores.

Based on the observation that the VLM, WCST, and LNS tests all were associated with speech-understanding ability and were moderately associated with each other (whereas the Visual Rhyme and Stroop tests were not associated with speech understanding and were weakly associated with the other tests), it was decided to combine scores from these three cognitive tests to generate a more comprehensive but still cohesive assessment of cognitive skills than any single test could yield. Scores for the three tests were combined by first converting all scores for a given test to z scores and then summing the scores across the three cognitive tests for each subject. This produced a composite cognitive score for each subject that was used in further analyses. Data for the Stroop and Visual Rhyme tests were not considered further.

Relationship between Cognitive Abilities and Speech Understanding with Long and Short RT

SPAC Test Scores

A mixed-model analysis of variance was performed to examine the relationships between cognitive grouping and speech understanding measured with the SPAC test. Low and high cognitive groups were formed from the eight lowest- and eight highest-scoring subjects for the composite cognitive score based on the rationale of maximizing the difference between cognitive groups given earlier. Within-subject variables were release time (short and long), talker (talkers #1 and #4), and masking noise (unmodulated and modulated). The main interest of this study was in the interaction between cognitive group and release time. This interaction was significant ($F[1, 14] = 7.0$, $p = .019$), and it is illustrated in Figure 4. Subjects with higher cognitive abilities yielded better speech-understanding scores than those with lower cognitive abilities. Post hoc pairwise comparisons exploring the interaction revealed that the group with lower cognitive scores performed significantly better with short RT processing than with long RT processing ($p = .014$). For the group with higher

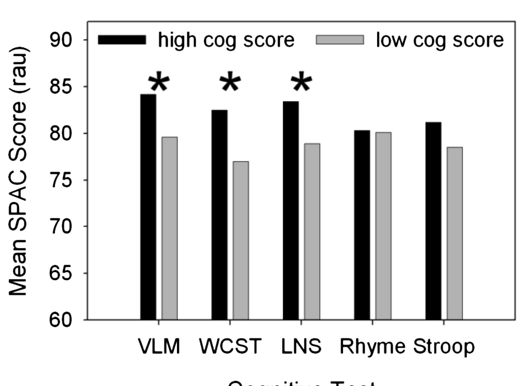

Figure 3. Average SPAC test scores for the higher-scoring and lower-scoring subjects for each cognitive test. Pairs that are significantly different are denoted by an asterisk.

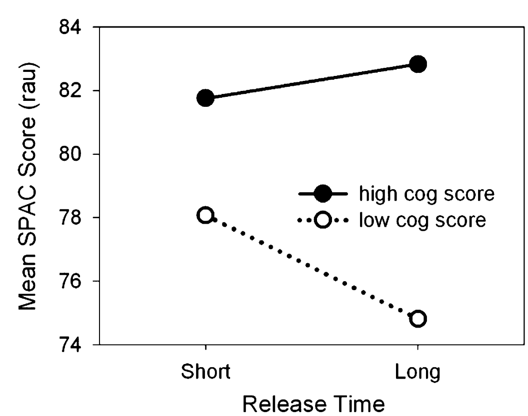

Figure 4. The interaction between cognitive group and release time for SPAC test scores. Data are averaged across both talkers and both maskers.

cognitive scores, the difference in mean SPAC scores for long RT and short RT processing was not statistically significant.

The analysis also revealed that, as in previous studies, speech understanding for Talker #1 was significantly poorer than for Talker #4 ($F[1, 14] = 24.9$, $p = .001$). However, there were no significant interactions involving talker. Further, there was not a significant difference in speech understanding when the masker noise was modulated versus unmodulated, and there were no significant interactions involving masker noise type.

BKB-SIN Test Scores

The BKB-SIN test, which employs contextually rich sentences and a modulated noise masker delivered at a series of fixed SNR conditions, was used for the purpose of generating performance intensity (PI) functions for both short and long RT conditions for the different cognitive groups. These PI functions were directly compared with those determined by Lunner and Sundewall-Thorén (2007). Their PI functions were derived from a test using context-free but somewhat predictable key words, delivered at SNRs that were adaptively varied for each subject. The study by Lunner and Sundewall-Thorén showed large differences in mean PI functions between listeners with lower and higher cognitive abilities, especially in modulated masking noise. The largest differences were seen for short release time processing.

To allow a more direct comparison with the questions posed in the present study, Lunner and Sundewall-Thorén's data for the modulated masker were extracted from their published curves and plotted together with PI functions from the present study's BKB-SIN data. The curve-fitting procedures described by Lunner and Sundewall-Thorén were followed to generate PI func-

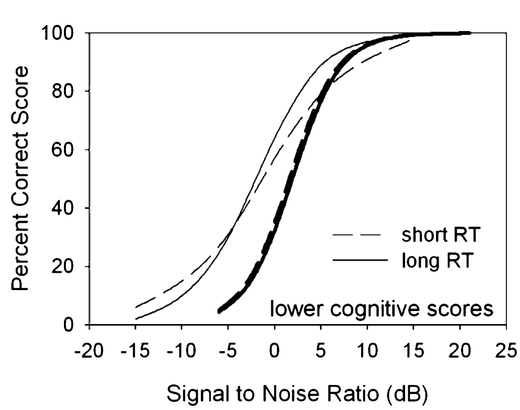

Figure 5. PI functions showing performance with short and long RT processing for subjects with lower cognitive scores. The lighter lines were developed by extracting the data from published curves in Lunner and Sundewall-Thorén (2007). The heavy lines portray data from the BKB-SIN test in the present study.

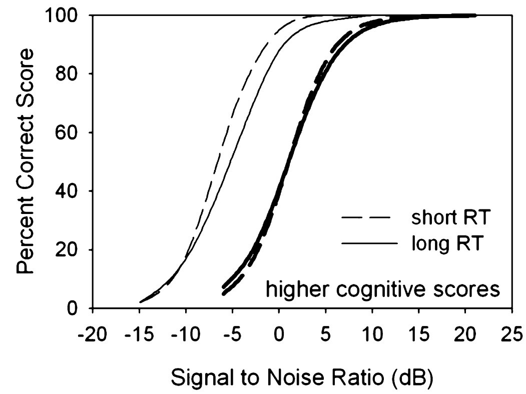

Figure 6. PI functions showing performance with short and long RT processing for subjects with higher cognitive scores. The lighter lines were developed by extracting the data from published curves in Lunner and Sundewall-Thorén (2007). The heavy lines portray data from the BKB-SIN test in the present study.

tions from the average BKB-SIN scores across subjects at each SNR. Figure 5 displays a direct contrast between performance with short and long RT processing for subjects with lower cognitive scores for both studies. Figure 6 gives the corresponding data for subjects with higher cognitive scores.

Although we could not perform a statistical evaluation of differences between the PI functions obtained in the two studies, both figures suggest the same pattern. In Figure 5, subjects with lower cognitive scores in the present study yielded PI functions with no notable differences between short and long RT processing. In contrast, the PI functions derived from Lunner and Sundewall-Thorén have somewhat different slopes for the two RT conditions. In Figure 6 the same trends are seen for subjects with higher cognitive scores.

To quantify the PI functions obtained in our study, curves were fitted to each subject's data for each RT condition. Then for each curve, the slope was computed in percent per decibel between scores of 25 and 75%. A mixed-model ANOVA was performed to examine the relationship between cognitive grouping and PI function slopes measured with the BKB-SIN test for short and long RT conditions. Neither main effects nor interaction were significant in this analysis. Thus, the differences between PI functions obtained with short and long RT processing for subjects with higher versus lower cognitive abilities reported by Lunner and Sundewall-Thorén were not reproduced in this study, which used a different speech test and a different psychometric procedure.

Advantage of Short versus Long RT Processing

Both Gatehouse and colleagues (2006b) and Lunner and Sundewall-Thorén (2007) report correlations between VLM score and the advantage of short versus long RT processing in modulated noise (speech

Table 2. Correlations between Cognitive Abilities and Speech Intelligibility Advantage of Short RT Processing over Long RT Processing with Modulated Masker

Study	Cognitive Test	Speech Test	Pearson *r*	Spearman *r*
Gatehouse et al (2006b)	VLM	Four Alternative Auditory Feature test	0.30*	0.39**
Lunner and Sundewall-Thorén (2007)	VLM	Dantale II	0.47*	0.49*
Present study	VLM	SPAC	–0.36	–0.49*
Present study	VLM	BKB-SIN	–0.29	–0.21
Present study	composite	SPAC	–0.22	–0.29
Present study	composite	BKB-SIN	–0.06	0.01

Note: Results are shown for three studies.
* significant at .05 level (two tailed).
** significant at .01 level.

understanding with short RT vs. speech understanding with long RT). This computation was repeated for the present study for the BKB-SIN data and for the SPAC test with modulated masker and Talker #4. Talker #4 was the more intelligible talker in the SPAC test and was, therefore, judged more similar to the highly intelligible talkers used in the previous studies. Computations for the present study were performed using the VLM cognitive score, to allow a strict parallel with the two previous studies, and for the composite cognitive score derived in the present study. The results of all three studies are shown in Table 2. Gatehouse and colleagues and Lunner and Sundewall-Thorén both found a positive relationship between cognitive abilities (measured with the VLM) and the benefits of short RT processing, indicating that individuals with higher cognitive abilities performed better with short RT processing. However, in the present study, using different speech-understanding tests, that result was not reproduced. When the VLM was the sole measure of cognitive ability and the SPAC test was the measure of speech understanding, the strength of the correlations in the present study generally was consistent with findings in the two previous studies (although not always statistically significant due to the smaller *N* in our study). However, the opposite relationship was seen: higher cognitive abilities were associated with an advantage for long RT processing. When the composite cognitive score was used as the measure of cognitive abilities, and when the BKB-SIN test was the measure of speech understanding, the relationship between cognitive status and the benefits of RT processing time disappeared.

Relationship between Cognitive Abilities and Real-World Release Time Preferences

After all testing was complete, subjects were asked whether they preferred the first or second RT condition they experienced in their daily life. They also were asked for the certainty of their preference on a four-point scale, as follows: 1 = uncertain, 2 = slightly uncertain, 3 = reasonably certain, 4 = very certain. Fifteen subjects reported a preference for the long RT condition, and nine reported a preference for the short RT condition. To facilitate exploration of the relationship between cognitive abilities and real-world preferences, the certainty ratings were arbitrarily scaled positively for long RT preference and negatively for short RT preference. Figure 7 depicts the relative strength of real-world preferences as a function of cognitive scores. Subjects who preferred the long RT condition are shown with circles. Subjects who preferred the short RT condition are shown with triangles. Two aspects of Figure 7 are noteworthy: (1) only one subject was uncertain about his or her preferred real-world RT condition, and (2) there was no association between RT preference and composite cognitive score. Subjects choosing the short RT condition covered the span of cognitive scores including both the lowest and the highest. Subjects who chose the long RT condition tended to have cognitive scores throughout the middle of the range. These data did not yield any evidence that RT preferences in daily life are associated with cognitive abilities as measured in this study.

Relationship between Speech Understanding and Real-World Release Time Preferences

It was also of interest to explore the relationship between laboratory measures of aided speech

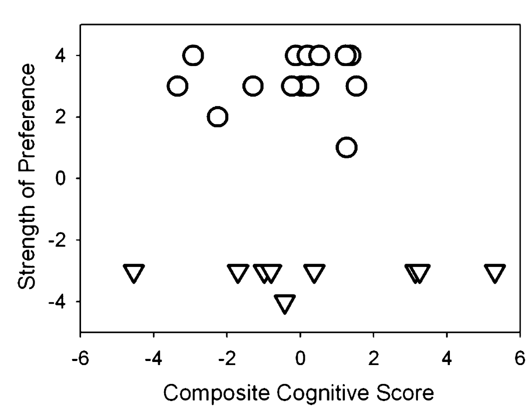

Figure 7. The relative strength of real-world preferences as a function of cognitive scores. Subjects who preferred the long RT condition are shown with circles. Subjects who preferred the short RT condition are shown with triangles.

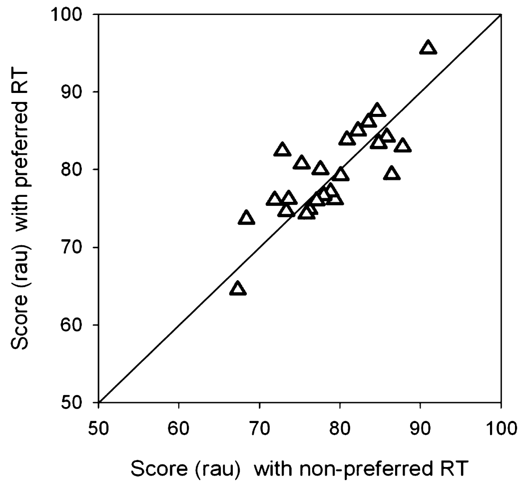

Figure 8. Mean scores obtained on the SPAC test (averaged across talkers and noises) for the preferred and nonpreferred RT conditions for each subject.

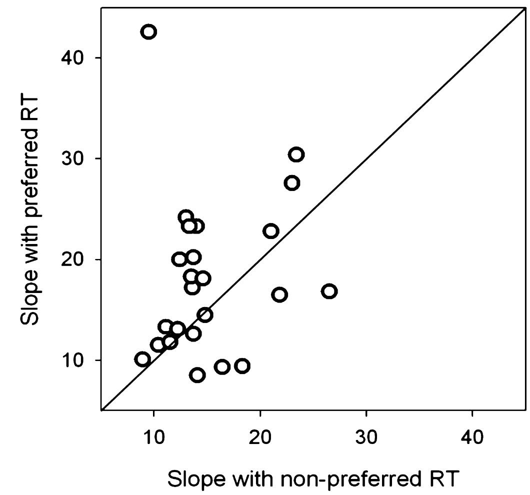

Figure 9. Slopes (percent per decibel) obtained on the BKB-SIN test for the preferred and nonpreferred RT conditions for each subject.

understanding with short and long RT processing and real-world preferences for RT processing. Figures 8 and 9 illustrate the scores obtained for the SPAC and BKB-SIN tests, respectively, for the preferred and nonpreferred RT conditions for each subject. If the test data for the preferred RT condition tended to be better than those for the nonpreferred condition, we would see the symbols tend to lie more often above the diagonal in both figures. Figure 8 shows that for the SPAC test scores this did not occur: there are as many symbols below the diagonal as above it, and all of them are close to the diagonal. Thus, there is no evidence that scores from the SPAC test for different RT conditions were related to real-world RT preferences. Figure 9 shows that for the BKB-SIN test there is a tendency for more symbols to be above the diagonal, indicating that subjects tended to prefer the RT condition that produced the steepest PI function. A *t*-test did not reveal a statistically significant difference between slopes for the preferred and nonpreferred RT conditions ($t[23] = 1.67$, $p = .108$). On the other hand, the effect size (Cohen's d) for slope was 0.47 (95% CI = 0.04–0.88), which suggests that steepness of PI function slope does impact the real-world RT preference (this result was not substantially changed when the data for the single outlying subject were adjusted). However, since the confidence interval for the effect is so wide and almost encompasses zero, it is clear that there is much uncertainty about this result.

Questions about the Validity of the Real-World Preferences

A reader might question the validity of the RT preference score obtained at the conclusion of the study. There is not general agreement on whether hearing aid wearers can subjectively distinguish between signals processed with short and long RTs in a laboratory task. For example, Gilbert and colleagues (2008) report that a substantial number of hearing-impaired listeners were not consistently able to detect any difference between signals processed with short and long RTs. On the other hand, Neuman and colleagues (1995) found that most hearing-impaired listeners were rather sensitive to quality differences between laboratory signals compressed with different release times. Since it is uncertain whether listeners can detect a difference between compressed signals in the highly controlled conditions of the laboratory, it is even less certain that these differences are salient in the uncontrolled and rapidly changing hubbub of daily life. In the present study we take the following view: if the final binary decision between short RT and long RT was consistent with judgments about the relative merits of short and long RT processing made by the subject at other times during the eight-week field trial, this would constitute rather convincing evidence that subjects were able to distinguish consistently between the two RT conditions.

The responses to the three questionnaires were evaluated to explore this matter. As noted above, the APHAB, HAPQ, and DOSO questionnaires were completed during the fourth and eighth weeks of the field trial, and the binary preference choice was made at the conclusion of the final laboratory test session, typically a few days after the second set of questionnaires was completed.

The subjective data obtained from the questionnaires were analyzed in two ways: (1) comparing scores for short RT and long RT conditions and (2) comparing scores for the preferred RT condition to those for the nonpreferred RT condition. All analyses were performed using repeated-measures ANOVA. Each of the

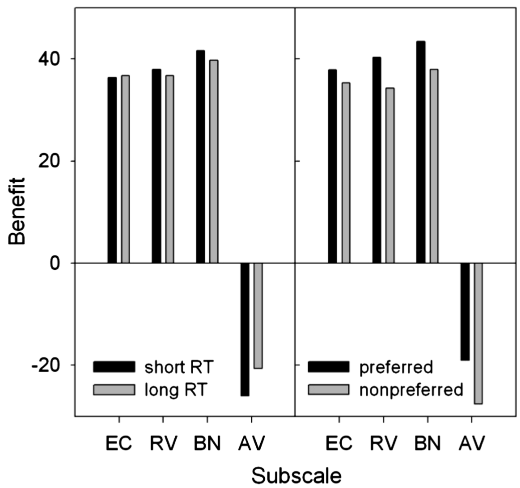

Figure 10. Mean scores for the subscales of the APHAB questionnaire for short and long RT conditions measured during the field trial and for the preferred and nonpreferred RT conditions identified in the final interview.

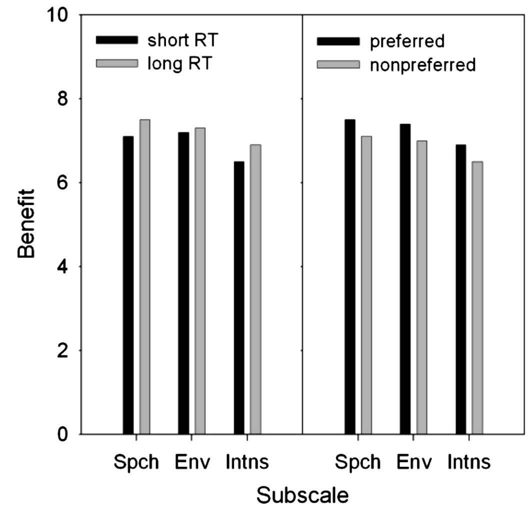

Figure 11. Mean scores for the subscales of the HAPQ questionnaire for short and long RT conditions measured during the field trial and for the preferred and nonpreferred RT conditions identified in the final interview.

three questionnaires was analyzed separately. Figures 10, 11, and 12 depict the mean scores for the subscales of the APHAB, HAPQ, and DOSO questionnaires, respectively. Each figure contrasts the mean scores for short and long RT conditions as well as the mean scores for the preferred and nonpreferred conditions. All subscales were included in the analyses except the Use subscale of the DOSO. The data for Use were strongly skewed in the direction of more subjects reporting longer use. In addition, as seen in Figure 12, Use scores did not suggest any sensitivity to differences between RT conditions. Thus, this subscale's data were omitted from further consideration.

In the figure panels comparing the long and short RT conditions, there seems to be a slight advantage for long RT in the responses to the HAPQ and DOSO, but that pattern is not seen in responses to the APHAB. In the statistical analyses comparing scores for short RT and long RT conditions, there was no statistically significant difference ($p > .05$) for the main effect of release time for any questionnaire. Further, there were no significant interactions between RT and subscale scores for any questionnaire. Thus, there was no statistical evidence for an overall better subjective impression for either the short or the long RT condition in real-world experience.

In contrast, all three figures show consistently higher scores for the RT condition that was declared to be preferred at the end of the study. Further, the statistical analyses of scores in the preferred and nonpreferred RT conditions indicated that these differences were probably not due to chance. The main effect of preference was statistically significant for the APHAB ($F[1, 23] = 5.57$, $p = .03$) and for the DOSO ($F[1, 23] = 10.44$, $p = .004$) and showed the same trends for the HAPQ ($F[1, 23] = 3.73, p = .066$). Once again, there were no significant interactions between RT and subscale scores for any questionnaire. Thus, analysis of the self-report data in terms of ultimate preference for short or long RT condition revealed that this binary decision made at the end of the eight-week field trial was consistent with responses to questionnaires that were obtained for the same RT conditions at separate earlier times. In other words, there was statistical evidence for an overall better subjective impression for one of the RT conditions in real-world experience, but the preferred condition differed across individuals.

A potential concern about the preference data centers on the possibility that subjects might have tended simply to choose as their preferred condition the one they were exposed to in the second segment of the field trial. As noted earlier, for the first trial condition 14 subjects received the long RT and 10 subjects received the short RT. Thus, although there was an overall preference for the long RT condition, the majority of subjects experienced the short RT condition in the second segment of the field trial. Nevertheless, a majority of the subjects (17 of the 24) ultimately preferred the condition they experienced in the second segment. This outcome was further explored as shown in Table 3. Table 3 is a contingency table showing the percent of subjects who preferred short versus long release time as a function of the condition they experienced in the second segment of the field trial. Of the subjects who experienced the long RT condition second, 90% preferred the long RT overall. Of the subjects who experienced the short RT condition second, 57% preferred the short RT overall. The percentages shown in the table were evaluated using the McNemar test (Siegel, 1956). The results support the following statements: (1) the proportion of subjects

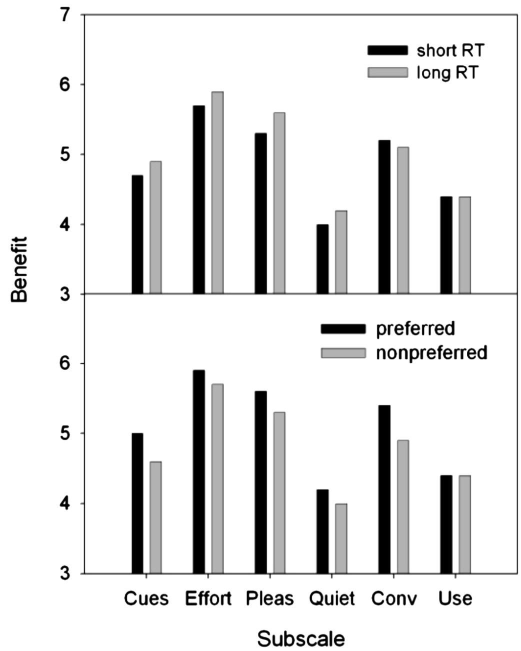

Figure 12. Mean scores for the subscales of the DOSO questionnaire for short and long RT conditions measured during the field trial and for the preferred and nonpreferred RT conditions identified in the final interview.

who preferred the first condition was significantly greater when their first condition was long than when it was short ($X^2[df = 1] = 19.3, p < .001$); (2) the proportion of subjects who preferred the second condition was significantly greater when their second condition was long than when it was short ($X^2[df = 1] = 6.95$, $p < .01$). These results suggest that subjects tended to prefer the long RT condition regardless of the order in which they experienced it. Last, in the exit interview, subjects were asked to provide up to three reasons for the choice they made between the RT conditions. The reasons given tended to center on clarity, quality, and loudness. They are not treated in detail here. However, it is of interest to note that only 6% of the reasons implied that a process of "getting used to" the hearing aid figured into the preference decision. Therefore, although we cannot rule out a predilection for the second RT condition, such a predilection, if it existed, was not a major determinant of final preferences.

DISCUSSION

Relationship between Cognitive Abilities and Speech Understanding with Long and Short RT

The present study further explored the relationship between cognitive abilities of hearing-impaired listeners and understanding of amplified compressed speech with varying release times. Three previous studies have provided somewhat conflicting views of this relationship. All three found a significant interaction between cognitive score and short versus long RT processing in at least one of the test conditions. However, the details of the relationships were not consistent. One notable factor, however, is that all of the reported statistically significant effects of RT have occurred in the results for listeners with lower cognitive abilities rather than for those with higher cognitive abilities.

Table 3. Contingency Table Showing the Percent of Subjects Who Preferred Short versus Long Release Time as a Function of the Condition They Experienced in the Second Segment of the Field Trial

	Condition for Second Segment	
Preferred Condition	Short RT	Long RT
Long RT	43	90
Short RT	57	10

In Lunner and Sundewall-Thorén (2007), individuals with lower cognitive abilities performed significantly better with long RT. This result was consistent with the less specific report of Gatehouse and colleagues (2006b) in which individuals with lower cognitive ability tended to perform better with long RT processing. In Foo and colleagues (2007), individuals with lower cognitive abilities performed significantly better with long RT for one speech-understanding test but significantly better with short RT for the other speech-understanding test. Again, there were no statistically significant differences for those with higher cognitive abilities. In our study, individuals with lower cognitive abilities performed significantly better with short RT for the SPAC test (Fig. 4) and showed no difference between short and long RT for the BKB-SIN test (Fig. 5). As in the previous studies, there were no statistically significant differences for those with higher cognitive abilities.

At first glance, the results for these studies seem puzzlingly inconsistent regarding the advantages of short and long RT processing. Nevertheless, the results do show a pattern. Among these studies, four test conditions used speech tests featuring short sentences. On the two occasions when the low cognitive group performed better with long RT processing, the sentences contained low-context but somewhat predictable key words (Lunner and Sundewall-Thorén used the Dantale II test, and Foo et al used the Hagerman Sentences). On the two occasions when the low cognitive group performed better with short RT processing, or equally with both RTs, the speech tests were both drawn from the same source of relatively natural, high-context, unpredictable sentences (Foo et al used the Hearing in Noise Test, and the present study used the BKB-SIN). Overall, this pattern suggests that when listeners with lower cognitive ability are challenged by the need to identify words based on audibility alone (i.e., without semantic context), they require long RT processing for best performance. On the other hand, when these listeners are presented with speech that is more ecologically valid (less clearly articulated but rich in context), the meaning can be largely constructed from contextual cues. In this situation, the RT processing variable might be less important, or perhaps some new cues made available by short RT processing can be a valuable supplement to context.

This interpretation is compatible with the results of Gatehouse and colleagues (2006b), in which listeners were asked to identify context-free but somewhat predictable key words. Those with lower cognitive abilities required long RT processing for their best performance. When considering the present study, it is important to note that although the Four Alternative Auditory Feature test used by Gatehouse and the SPAC test used in the present study are both four-alternative forced-choice tests, there are subtle but potentially important differences between the speech stimuli. The SPAC test items were more ecologically valid due to more natural articulation and the use of a carrier sentence that required comprehension in order to pinpoint the key word (see examples of carrier sentences given earlier). The results for the SPAC test were consistent with previous results for more ecologically valid test items, in that listeners with lower cognitive abilities performed significantly better on the SPAC test with short RT processing.

Taken together, these studies of laboratory speech understanding are in agreement that RT processing is more important for listeners with lower cognitive abilities than for those with higher cognitive abilities. They further suggest that the relationship between cognitive abilities and speech understanding with short and long RT processing depends on the characteristics of the tested speech. This last point is bolstered by the results depicted in Table 2 showing the relationships for three studies between speech-understanding and cognitive scores measured with the VLM. The most notable difference among these results is that the direction of the correlation for the present study is opposite to that for the two previous studies. This result seems most likely to be the consequence of some characteristics of the speech stimuli in the different studies.

It is also worth considering the possible effects of different psychophysical test procedures on the results of comparisons of short and long RT processing. Figures 5 and 6 both show the PI functions for the present study to be at more favorable SNRs than those for Lunner and Sundewall-Thorén (2007). Some of this discrepancy probably resulted from calibration differences between the studies, and some might be attributable to the fact that subjects in our study had, on average, a bit more hearing loss than those of Lunner and Sundewall-Thorén. Note also, however, that subjects in our study responded to sentences delivered at a fixed set of SNRs that was the same for all subjects. In contrast, Lunner and Sundewall-Thorén's subjects responded to sentences delivered at SNRs that were adaptively varied and differed across subjects. Naylor and Johannesson (2009) demonstrate that the SNR of the stimulus delivered to the subject's ear canal varies as a function of the SNR delivered to the hearing aid in combination with the settings of compression parameters used to process the sound. Thus, it is possible that some of the lack of correspondence between the results of our study and those of Lunner and Sundewall-Thorén is attributable to the impact of these different psychophysical procedures on the SNRs delivered to the subjects' ear canals. The extent to which this factor might have influenced the patterns seen in the PI function slopes in Figures 5 and 6 is not known.

Another variable that has produced inconsistencies among the four studies is the potential benefit of masker modulations and the extent to which this interacts with the value of short RT processing. For normal-hearing listeners, temporal envelope dips in the masker allow "glimpses" of speech cues that can result in improved speech understanding. It is well established that, as a group, older hearing-impaired listeners are relatively less able than normal-hearing listeners to capitalize on masker modulations to improve speech understanding (e.g., Souza et al, 2007; Wilson et al, 2007). Further, Larsby and colleagues (2008) show that modulated maskers have different effects (benefits and penalties) on speech understanding for different performance levels (80% correct vs. 50% correct), and Naylor and Johannesson (2009) demonstrate that the SNR of a processed signal depends in a complicated way on the combination of masker modulations, compression parameters, and the SNR at the hearing aid's input. Also, both Olsen and colleagues (2004) and Jenstad and Souza (2005) note that short RT processing produces intelligibility improvements for some phonemes but offsetting decrements for other phonemes. Clearly, the relationship among RT processing, masker modulations, and cognitive abilities is complex. In Gatehouse and colleagues (2006b) and Lunner and Sundewall-Thorén (2007), listeners with higher cognitive scores obtained more benefit than listeners with lower cognitive scores from short RT processing in modulated maskers. However, in the present study and in Foo and colleagues (2007) the relationship between cognitive ability and RT processing was independent of the modulation characteristics of the masker. Thus, it remains to be determined under what combinations of performance level, speech material, and masker characteristics there is a relationship between cognitive abilities and RT processing speed.

Real-World Release Time Preferences and Relationship to Laboratory Measures

Previous published studies of the relationship between cognitive abilities and compression release time have relied mostly on laboratory speech-understanding tests. These studies have consistently shown a relationship between cognitive abilities and speech understanding with short and long RT processing, although the nature of the relationship is not completely clear. Based on the considerations outlined above, it is reasonable to propose that release time processing speed is more critical for listeners with lower cognitive abilities than for those with higher cognitive abilities. Further, within this group, longer release times have been shown to provide better performance when the speech is very clear and low in context. However, in more ecologically valid high-context speech, the importance of processing time might diminish or even favor short RT processing.

These laboratory types of studies allow us to explore the *efficacy* of short and long release times in terms of their ability to facilitate speech understanding under well-defined, controlled conditions. The essential next step is to explore the *effectiveness* of short and long release times by examining the experiences of hearing-impaired individuals when they use devices with short and long RTs in their everyday lives. This was done in the present study by asking subjects which of the two field trial conditions they preferred and also by asking them to complete three outcome questionnaires during the fourth week of each segment of the field trial. The data obtained in this part of the study allowed us to examine differences between short and long RT processing in daily life.

The analyses comparing self-rated performance with short RT and long RT (Figs. 10, 11, and 12) were not able to demonstrate an overall statistically significant difference between them in any questionnaire. Thus, there is no evidence from this investigation that either short or long RT is superior overall in daily life. Two previous studies have reported field trials in which subjects wore modern compression devices in daily life for several weeks and then completed questionnaires comparing short and long RT conditions. Results from van Toor and Verschuure (2002) were consistent with those reported in the present study: overall, there were no significant differences in daily life effectiveness between short RT and long RT. On the other hand, subjects in the Gatehouse and colleagues (2006a) study did produce self-reports that revealed a pattern of significant differences between short and long RT conditions at the group level. They noted that when subjects rated comfort, the long RT tended to yield better scores, whereas when subjects rated intelligibility, the short RT received the higher scores.

In both of these prior studies it was noted that there was evidence that long RT processing appealed more to some subjects but short RT processing appealed more to other subjects. In this sense, all three investigations produced consistent outcomes. In the present study, roughly two-thirds of the subjects preferred long RT processing in their daily lives, whereas about one-third preferred short RT processing.

When the questionnaire data were compiled into preferred and nonpreferred RT conditions (as illustrated in Figures 10, 11, and 12) there were noteworthy differences in all questionnaires favoring the preferred condition. This result is consistent with a conclusion that most individual subjects did distinguish between the short and long RT conditions in daily life and held a consistent preference for one of them. Further, effect sizes (Cohen's d) computed to assess the magnitude of the subjective differences between preferred and nonpreferred RT conditions ranged from 0.27 to 0.29 for the three questionnaires. One way to interpret this finding is that a listener using his or her preferred release time has about a 58% probability of producing a higher subjective score than one using his or her nonpreferred RT (Grissom, 1994). An effect of this magnitude is probably material in contributing to the overall success of hearing aid provision. This finding bolsters our original contention that the selection of appropriate RT during hearing aid fitting should be an essential and systematic aspect of the protocol. The data obtained in this study allow a preliminary exploration of this matter.

Were real-life preferences for short RT or long RT predictable from a subject's cognitive scores? If this were the case, we would expect a pattern in which subjects with the higher cognitive scores would have preferred one RT condition while those with the lower cognitive scores preferred the other RT condition. Figure 7 indicates that this was not the case: the overall pattern of preferences did not support the hypothesis.

Were real-life preferences for short RT or long RT predictable from a subject's aided speech-understanding scores measured in the laboratory? If this were the case, we would expect subjects to prefer the RT condition that yielded the highest SPAC score or the condition that produced the steepest PI function slope on the BKB-SIN test. The distribution of subject data shown in Figure 8 indicates that scores obtained with short and long RTs on the SPAC test (as used in this study) would not be helpful in predicting the preferred RT condition in the real world. The distribution of subject data shown in Figure 9, while not definitive, hints at the possibility that a measure of PI function slope might be indicative of real-world RT preference. This potential predictor deserves further exploration.

Perhaps the RT preference for a particular individual depends on a combination of variables such as cognitive ability, speech understanding, and/or other data. For example, Gatehouse and colleagues (2006b) note that susceptibility to spectral and temporal smearing was related to the benefit of different RT values. Moore (2008) has suggested that ability to utilize temporal fine structure in the signal can influence the value of short RT processing for a given listener. Others have suggested that the ecological factors that determine an individual's hearing demands in daily life might be important determinants of the best nonlinear processing characteristics (e.g., Gatehouse et al, 1999). Further research is indicated to explore the additional variables that might contribute to the optimal release time processing for any individual. The results of our study argue that the ability to prospectively prescribe the most favorable release time processing for a given patient could produce worthwhile improvements in the quality of hearing aid fitting.

CONCLUSIONS AND LIMITATIONS

The research described here explored the effects of short and long compression release time processing in hearing aid fittings worn in a laboratory setting and in everyday life. Twenty-four hearing-impaired individuals served as subjects. In the laboratory, we measured understanding of amplified speech using two speech tests, and we determined each subject's cognitive abilities. These data added to the existing small body of literature that has explored the relationship between cognitive abilities and speech understanding with short and long release times. When interpreted in the light of previous studies, our results suggest that compression processing release time is more critical for patients with lower cognitive abilities than for those with higher cognitive abilities and that the most advantageous release time depends on characteristics of the speech signal.

When subjects wore the hearing aids in everyday life, they were able to distinguish between conditions with short and long release times and they tended to prefer one of them. About two-thirds of the subjects preferred the long release time. Neither laboratory speech understanding nor cognitive abilities was an accurate predictor of the release time that a subject preferred in everyday life.

The research had some limitations. The number of subjects was fewer than ideal, and this might have limited our ability to detect worthwhile trends in the data. The study was accomplished with a single hearing aid, so the generalizability of our findings to other similar hearing aids has not been established. It also is possible that other combinations of attack and release times might yield different results.

Acknowledgments. Graham Naylor, Ph.D., contributed to the planning stages and provided valuable comments on an earlier draft of the manuscript. Robert Sherbecoe, M. A., and Ginger Gray, M.A., collected and managed the data for the study.

REFERENCES

Akeroyd MA. (2008) Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol* 47(Suppl 2):S53–S71.

Boothroyd A. (1985) Evaluation of speech production of the hearing impaired: some benefits of forced-choice testing. *J Speech Hear Res* 28(2):185–196.

Byrne D, Dillon H, Ching T, Katsch R, Keidser G. (2001) NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. *J Am Acad Audiol* 12(1):37–51.

Cox RM, Alexander GC. (1995) The abbreviated profile of hearing aid benefit. *Ear Hear* 16(2):176–186.

Cox RM, Alexander GC, Gilmore CA. (1987) Intelligibility of average talkers in typical listening environments. *J Acoust Soc Am* 81 (5):1598–1608.

Cox RM, Alexander GC, Xu J. (2009) Development of the Device Oriented Subjective Outcome Scale (DOSO). Poster presented at the Annual Meeting of the American Auditory Society. Retrieved from http://www.memphis.edu/ausp/harl/publications.htm#posters.

Daneman M, Merikle PM. (1996) Working memory and language comprehension: a meta-analysis. *Psychon Bull Rev* 3(4):422–433.

Dillon H. (2001) Compression systems in hearing aids. In *Hearing Aids*. New York: Thieme, 159–186.

Dillon H, Storey L. (1998) The National Acoustic Laboratories' procedure for selecting the saturation sound pressure level of hearing aids: theoretical derivation. *Ear Hear* 19(4):255–266.

Dreschler WA, Verschuure H, Ludvigsen C, Westermann S. (2001) ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology. *Audiology* 40(3): 148–157.

Etymotic Research (1985) *BKB-SIN Speech in Noise Test, Version 1.03*. Elk Grove Village, IL: Author.

Foo C, Rudner M, Rönnberg J, Lunner T. (2007) Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J Am Acad Audiol* 18(7):618–631.

Gatehouse S, Elberling C, Naylor G. (1999) Aspects of auditory ecology and psychoacoustic function as determinants of benefits from and candidature for non-linear processing in hearing aids. In: Rasmussen A, Andersen P, Poulsen T, eds. *Auditory Models and Non-linear Instruments*. [Proceedings of the 18th Danavox Symposium.] Copenhagen: Holmens Trykkeri, 221–233.

Gatehouse S, Naylor G, Elberling C. (2003) Benefits from hearing aids in relation to the interaction between the user and the environment. *Int J Audiol* 42(Suppl 1):S77–S85.

Gatehouse S, Naylor G, Elberling C. (2006a) Linear and nonlinear hearing aid fittings—1. Patterns of benefit. *Int J Audiol* 45(3): 130–152.

Gatehouse S, Naylor G, Elberling C. (2006b) Linear and nonlinear hearing aid fittings—2. Patterns of candidature. *Int J Audiol* 45 (3):153–171.

Gilbert G, Akeroyd MA, Gatehouse S. (2008) Discrimination of release time constants in hearing-aid compressors. *Int J Audiol* 47(4):189–198.

Grissom RJ. (1994) Probability of the superior outcome of one treatment over another. *J Appl Psychol* 79(2):314–316.

Hällgren M, Larsby B, Lyxell B, Arlinger S. (2001) Cognitive effects in dichotic speech testing in elderly persons. *Ear Hear* 22(2):120–129.

Heaton R. (1981) *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.

Humes LE. (2007) The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol* 18(7):590–603.

Jenstad LM, Souza PE. (2005) Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *J Speech Lang Hear Res* 48(3):651–667.

Larsby B, Hällgren M, Lyxell B. (2008) The interference of different background noises on speech processing in elderly hearing impaired subjects. *Int J Audiol* 47(Suppl 2):S83–S90.

Lunner T. (2003) Cognitive function in relation to hearing aid use. *Int J Audiol* 42(Suppl 1):S49–S58.

Lunner T, Sundewall-Thorén E. (2007) Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid. *J Am Acad Audiol* 18(7):604–617.

Moore BC. (2008) The choice of compression speed in hearing AIDS: theoretical and practical considerations and the role of individual differences. *Trends Amplif* 12(2):103–112.

Moore BC, Stainsby TH, Alcántara JI, Kühnel V. (2004) The effect on speech intelligibility of varying compression time constants in a digital hearing aid. *Int J Audiol* 43(7):399–409.

Naylor G, Johannesson R. (2009) Long-term signal-to-noise ratio (SNR) at the input and output of amplitude compression systems. *J Am Acad Audiol* 20(3), online.

Neuman AC, Bakke MH, Mackersie C, Hellman S, Levitt H. (1995) Effect of release time in compression hearing aids: paired-comparison judgments of quality. *J Acoust Soc Am* 98(6):3182–3187.

Novick ML, Bentler RA, Dittberner A, Flamme GA. (2001) Effects of release time and directionality on unilateral and bilateral hearing aid fittings in complex sound fields. *J Am Acad Audiol* 12(10): 534–544.

Olsen HL, Olofsson A, Hagerman B. (2004) The effect of presentation level and compression characteristics on sentence recognition in modulated noise. *Int J Audiol* 43(5):283–294.

Pichora-Fuller MK. (2007) Audition and cognition: what audiologists needs to know about listening. In: Palmer C, Seewald R, eds. *Hearing Care for Adults*. Stafa, Switzerland: Phonak, 71–85.

Rudner M, Foo C, Sundewall-Thorén E, Lunner T, Rönnberg J. (2008) Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *Int J Audiol* 47(Suppl 2):S91–S98.

Scollie S, Seewald R, Cornelisse L, et al. (2005) The Desired Sensation Level multistage input/output algorithm. *Trends Amplif* 9(4):159–197.

Shi LF, Doherty KA. (2008) Subjective and objective effects of fast and slow compression on the perception of reverberant speech in listeners with hearing loss. *J Speech Lang Hear Res* 51(5):1328–1340.

Siegel S. (1956) *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Souza PE, Boike KT, Witherell K, Tremblay K. (2007) Prediction of speech recognition from audibility in older listeners with hearing loss: effects of age, amplification, and background noise. *J Am Acad Audiol* 18(1):54–65.

Stroop JR. (1935) Studies of interference in serial verbal reaction. *J Exp Psychol* 18:643–662.

Studebaker GA. (1985) A "rationalized" arcsine transform. *J Speech Hear Res* 28:255–262.

Uttl B, Graf P. (1997) Color–Word Stroop test performance across the adult life span. *J Clin Exp Neuropsychol* 19(3):405–420.

van Toor T, Verschuure H. (2002) Effects of high-frequency emphasis and compression time constants on speech intelligibility in noise. *Int J Audiol* 41(7):379–394.

Wechsler D. (1997) *The Wechsler Adult Intelligence Scale*. 3rd edition. San Antonio, Texas: Psychological Corporation.

Wilson RH, Carnell CS, Cleghorn AL. (2007) The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *J Am Acad Audiol* 18(6):522–529.

Wingfield A, Tun PA. (2001) Spoken language comprehension in older adults: interactions between sensory and cognitive change in normal aging. *Semin Hear* 22(3):287–301.