Connected Discourse Tracking as a Hearing Aid Evaluation Tool

Jill E. Preminger
University of Louisville School of Medicine

The purpose of this project was to evaluate the usefulness of the tracking procedure as a simulated hearing aid field study technique. The purpose of the first experiment was to determine whether tracking rate was influenced by text equivalency or practice effects. The purpose of the second experiment was to determine whether the tracking procedure was sensitive to small but significant changes in a hearing aid's directional capabilities (microphone mode). With the method used in the present study, tracking can be a reliable procedure; the same tracking rate was measured on several re-tests with new passages. The results in the present study also showed significant differences in tracking rate across changes in hearing aid microphone mode for a group of 6 subjects. When looking at performance for individual subjects, however, the procedure was not sensitive enough in order to differentiate between hearing aid modes in 100% of the subjects. The tracking method allows the examiner to observe an individual's communication performance in a carefully controlled conversational setting. While this procedure is time consuming, it can be a useful technique.

There is a need for standardized testing techniques to assess the benefit provided by new hearing aid processing schemes. Walden (1997) proposed a core of evaluation techniques to be used in clinical trials of new hearing aid technologies. These evaluation techniques included objective laboratory measures and subjective field evaluations. Objective laboratory tests are useful because performance can be measured under carefully controlled conditions. This allows for reliable measures. Subjective field evaluations may be unreliable (Preminger & Cunningham, 2003). In the real world, the listening environment is constantly chang-

Correspondence concerning this article should be addressed to Jill E. Preminger, University of Louisville School of Medicine, Program in Audiology, Myers Hall, Louisville, Kentucky 40292. Phone: (502) 852-7691. Fax: (502) 852-0865. Electronic mail may be sent via Internet to jill.preminger@louisville.edu

ing, unlike the static listening environments created for laboratory studies. The changing listening environment in the real world makes it difficult to determine whether differences in field-study ratings are due to changes in hearing aid processing, changes in attitude, or simply changes in the environment.

Unfortunately, laboratory measures are not always good predictors of field study results (Punch, Robb, & Shovels, 1994; Stone, Moore, Alcántara, & Glasberg, 1999; Valente, Fabry, Potts, & Sandlin, 1998). Studies have found that objective tests have discriminated between hearing aid processing schemes when field studies have not (Valente et al., 1998); field studies have discriminated between hearing aid processing schemes when objective tests have not (Stone et al., 1999); and through the use of objective tests and field study measures, different hearing aid processing schemes have been selected as superior (Punch et al., 1994). These differing results between laboratory and field study measures may be due to true differences between these conditions, or due to unreliable field study results. Until better field study testing techniques are available, new laboratory tests are needed which can simulate real world listening in a more controlled environment.

Connected discourse tracking is a method that can mimic characteristics of a real-world conversation in the controlled laboratory environment. Tracking was first described by De Filippo and Scott (1978). With this procedure, a speaker reads from a prepared text, while a receiver must repeat everything the speaker says with 100% accuracy. If the speaker repeats a work or phrase incorrectly, the receiver (and speaker) will use a pre-determined technique in order to elicit the correct word or phrase from the speaker. The tracking procedure allows for the evaluation of successful and unsuccessful communication strategies (Owens & Raggio, 1987) as well as overall communication abilities (De Filippo & Scott; Owens & Telleen, 1981).

Connected discourse tracking has been used to evaluate the effectiveness of a tactile aid as a supplement to lipreading (De Filippo & Scott, 1978) and to compare the effectiveness of two different tactile devices (De Filippo, 1984). Tracking has also been used as an outcome measure within subjects in order to differentiate performance between: speechreading with cochlear implants compared to speechreading alone (Jenkins, Chmiel, & Jerger, 1989), speechreading with a cochlear implant compared to speechreading with a tactile aid (Skinner et al., 1988), speechreading when specific bands of electrodes were deactivated within a cochlear implant (Geier & Norton, 1992), and speechreading while listening to four different simulated cochlear implant processing schemes (Faulkner, Rosen, & Smith, 2000). There are no reports in the literature, however, of connected discourse tracking as a tool to differentiate among hearing aid processing schemes.

Care must be taken in the implementation of a tracking procedure because it can be an unreliable procedure. Differences in tracking rate can result from different speaker characteristics and different listener characteristics (Tye-Murray &

Tyler, 1988). This is not a problem if results are only evaluated within subjects using the same speaker. The tracking procedure can also be unreliable due to variations in text difficulty and text presentation (Tye-Murray & Tyler, 1988). The reliability of the procedure can be improved by controlling for text complexity (De Filippo, Lansing, Elfenbein, Kallaus-Gay, & Woodworth, 1994; Hochberg, Rosen, & Ball, 1989; Matthies & Carney, 1988), by limiting the acceptable repair strategies (Fenn & Smith, 1987; Matthies & Carney, 1988), and by limiting the number of times a phrase is repeated (and thus not requiring 100% accuracy; Fenn & Smith, 1987).

The purpose of this project was to evaluate the tracking procedure as a simulated hearing aid field study technique. This would be a useful procedure if differences in tracking rate, within a subject, are due to differences in the hearing aid and not due to differences inherent in the test. Because tracking rate has been shown to be influenced by text complexity (De Filippo et al., 1994; Hochberg et al., 1989), it is possible that some text passages will be easier to track than others. It is also possible that, within subjects, tracking rate will increase due to practice effects. As a listener becomes more familiar with the tracking procedure, the talker's voice, and the topic of each passage, a subject's performance could improve over time. This practice effect could be seen either across several test passages (tracking rate could increase as a function of passage presentation order), or within a passage (tracking rate could be slower for the first half of a passage and faster for the second half of the passage). The purpose of the first experiment was to determine whether tracking rate was influenced by passage (text) equivalency and to determine whether tracking rate was influenced by practice effects. It is also important to determine whether the tracking procedure can be used to differentiate between hearing aid processing schemes. The purpose of the second experiment was to determine whether the tracking procedure was sensitive to small but significant changes in a hearing aid's directional capabilities.

EXPERIMENT I: DEVELOPMENT OF TRACKING MATERIALS

In the first experiment tracking performance for high-context and low-context materials was evaluated to determine whether there were practice effects and to determine whether some passages were easier to track than others.

Subjects

A group of 6 young adults (83% female) with normal hearing tracked the low-context materials. A new group of 5 young adults (60% female) with normal hearing tracked the high-context materials. Across both groups of subjects, ages ranged from 22 to 31 years with a mean age of 26. Subjects were students in the audiology and speech language pathology graduate programs at the University of Louisville. Subjects with normal hearing were used in this preliminary evaluation of the tracking technique due to ease of access. Many speech materials that are used to assess listeners with hearing loss have first been evaluated by listeners with normal hearing. This was the case for the Speech Intelligibility Rating (SIR) test (Cox & McDaniel, 1984, 1989; McDaniel & Cox, 1992), which served as the basis for the high-context speech materials in the present study.

It may be a concern that the education level of the subjects used in the present study may not generalize to the typical education level of individuals with hearing loss. This concern can be evaluated by looking at the performance for both the low-context and high-context test passages.

Talker

One individual (the author) served as the talker for all data collection in order to eliminate talker variability as a procedural factor (Tye-Murray & Tyler, 1988).

Hearing Aid

Each subject was fit monaurally with a four-channel wide dynamic-range compression (WDRC) behind-the-ear hearing aid fit with a foam earmold. The hearing aid was adjusted for very low gain and a band-limited response, so that the aided audibility was actually poorer than unaided audibility. The hearing aid had two microphone modes: omnidirectional or directional; the hearing aid was set in the omnidirectional mode. All testing was completed in the sound field with the unaided ear plugged with a foam earplug.

Materials

High-context. Fourteen passages from the SIR test were used as tracking material, 12 test passages and 2 practice passages. These passages were taken from a children's encyclopedia and were all at a seventh grade reading level (Cox & McDaniel, 1984). They are equivalent in terms of length (108-110 words), subject matter (common plants, animals, and household objects), vocabulary, and sentence structure (Cox & McDaniel, 1984). The topic of each passage is always mentioned in the first sentence. These passages have been shown to have equal intelligibility in a pre-recorded format for listeners with normal hearing and with hearing loss (Cox & McDaniel, 1989; McDaniel & Cox, 1992). Because of the similarity in their semantic and syntactic structure, it was predicted that the passages would be equivalent in terms of live voice presentation required for tracking. An example of a SIR passage is shown in Appendix A.

Low-context. Seven passages were taken from the "Observatory" column in the *New York Times* Science Section. Each passage is a complete article about a recent scientific finding related to the natural sciences. The first sentence of each passage is frequently misleading and provides minimal information that can be used to predict what appears next. The language used in these passages is fairly technical and the topics are fairly obscure. The passages varied in length from

230 to 272 words. An example of a low-context passage is shown in Appendix B. Two shorter passages (approximately 100 words each) taken from the same newspaper column were used as practice passages.

Procedure

Based on the techniques used in previous research (De Filippo & Scott, 1978; Fenn & Smith, 1987; Lunato & Weisenberger, 1994; Matthies & Carney, 1988; Owens & Raggio, 1987; Schoepflin & Levitt, 1991; Tye-Murray & Tyler, 1988) the following procedures were implemented:

1. Data collection did not begin until each subject tracked two practice passages.
2. Prior to tracking, subjects were told the topic of each test passage. Subjects were also able to read the topic of the low-context passages prior to tracking, because some of these topics may have contained unfamiliar words (e.g., nematodes).
3. All pauses within a passage were pre-determined, so that decisions about phrase length would not have an impact on tracking speed.
4. The only repair strategy used was repetition of the entire missed phrase. A phrase could be repeated up to two times. After three presentations, the speaker continued to the next phrase whether or not the subject repeated the phrase correctly; 100% correct repetition was not required.
5. All passages were presented to each subject by the same talker.
6. Results were analyzed within subjects, so that subject and/or subject-talker interactions did not influence the results.

All tracking was performed in an auditory alone condition in a sound booth. The tracking materials were read with monitored live voice and transmitted through an audiometer to a loudspeaker at 0° azimuth. A multi-talker noise babble was delivered from a CD to a second loudspeaker located at 180° and presented at 65 dBA in reference to the listener's head. Each subject tracked all passages (either high-context or low-context) in one 30- to 40-min session.

High-context passages. While tracking two practice passages, the level of the speech was varied until the subject made approximately one error for every 10 words. The signal-to-noise ratio (SNR) at the end of the second practice passage was used for all subsequent testing for the individual subject. Tracking performance was then measured for 12 separate passages; each subject heard the 12 passages in a random order. If the subject missed a word or phrase, the phrase was repeated up to two times. The sender proceeded to the next phrase after the second repetition, regardless of the subject's accuracy. The time required to complete each passage was recorded, and tracking rate was calculated in words per minute (wpm). Only words that were repeated correctly were included in this calculation.

Low-context passages. The procedure for the presentation of these passages was identical to the procedure for the high-context passages, except for adjustments to account for the increased length of the low-context passages. The first two sentences of each passage (approximately 40 to 50 words) were presented as practice and not included in the final wpm calculations. Then, each passage was divided into two sections of approximately 100 words. The time required for the completion of each section within a passage was recorded and wpm was calculated. An example of how the low-context passages were divided into sections is shown in Appendix B.

Results

High-context passages. The individual subject results in terms of passage order are shown in Figure 1. A repeated measures ANOVA was performed for the factor Passage Order to determine whether there was a practice effect. The main effect of Passage Order was not significant, $F(11, 55)=0.64$, $p=.79$. Performance did not improve after repeated exposure to the task.

The data were also analyzed in order to determine if certain passages were easier to track than others. Mean tracking rate for each passage is shown in Figure 2. A repeated measures ANOVA was performed; the main effect of Topic was

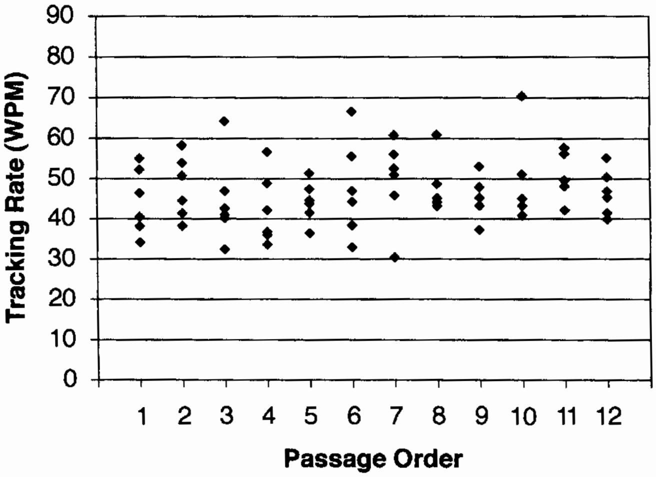

Figure 1. Individual subject tracking rate as a function of passage order for high-context passages. Each diamond symbol represents the tracking rate for one passage for an individual subject.

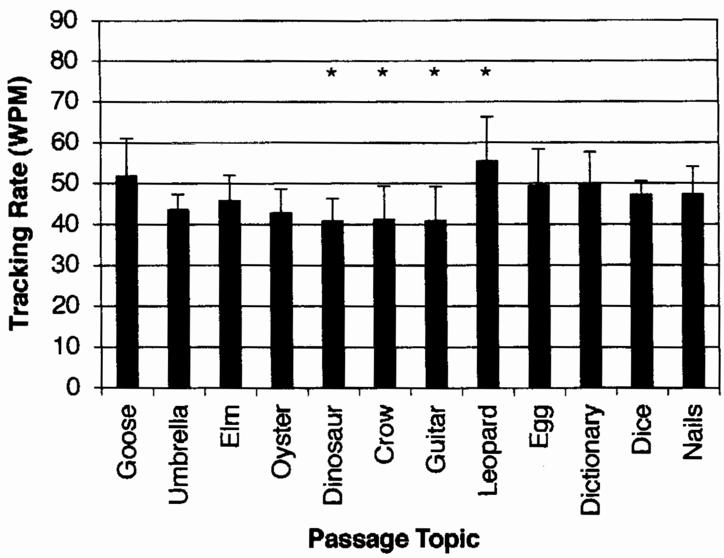

Figure 2. Mean subject tracking rate as a function of passage topic for high-context passages. Passage topics with significantly lower or higher mean tracking rate are indicated with an asterisk.

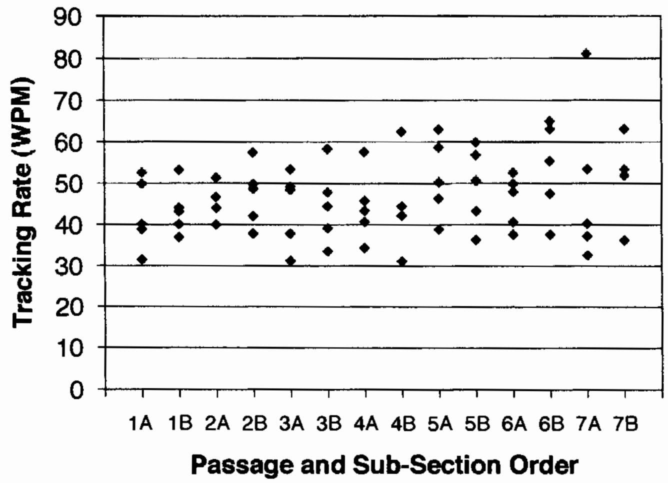

Figure 3. Individual subject tracking rate as a function of passage order (numbers 1 through 7) and as a function of sub-sections within a passage (A or B) for low-context passages. Each diamond symbol represents the tracking rate for one sub-section for an individual subject.

significant, $F(11, 55) = 3.764, p = .0013$. Follow-up testing, with the Tukey HSD test, indicated that tracking rate in wpm for the topic "leopard" was significantly better than tracking rate for topics "dinosaur," "crow," and "guitar." There were no significant differences in tracking rate for the remaining eight topics.

Low-context passages. The individual subject results in terms of passage order are shown in Figure 3 for the 14 sub-sections (2 sub-sections, A was heard first and B was heard second, within each of the 7 passages) in the order in which they were administered. Two types of practice effects could be evident for this type of data: Subjects could show improved tracking rate as a function of passage order, or subjects could show improved tracking rate within a passage as a function of sub-section order. A two-factor repeated measures ANOVA was performed for the factors Passage Order and Sub-Section Order. The effect of Passage Order was not significant, $F(6, 24) = 0.89, p = .52$. The effect of Sub-Section Order also was not significant, $F(1, 24) = 2.29, p = .20$. Subjects did not increase their tracking rate as they became more familiar with the material within a passage or as they became more familiar with the task across passages.

The data were also analyzed in order to determine if certain passages were easier to track than others; the mean tracking speed for each topic is shown in Figure 4. A partially nested ANOVA was performed for the factor Topic, with Sub-Section nested within topic, and with the factor Subject as a random variable. The effect of Topic was not significant, $F(6, 24) = 1.87, p = .13$, and the effect of

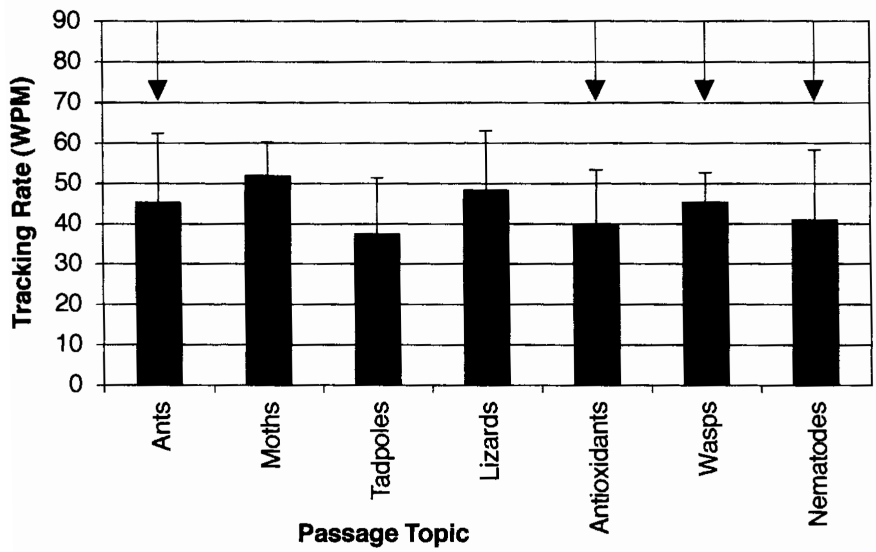

Figure 4. The same as Figure 2, for the low-context passages. The passages selected for use in Experiment II are indicated by the arrows.

SubSection also was not significant, $F(7, 28) = 0.55$, $p = .79$.

Only four of the passages (eight sections) were needed for the second experiment. The four passages with the most similar mean tracking rate and the smallest inter-subject standard deviations were selected. These are indicated with arrows in Figure 4.

Discussion

Based on these results there was no overall improvement in tracking over a 30- to 40-min time period for the high-context and the low-context materials. There were, however, significant differences in mean tracking rate, as a function of passage topic, for the high-context materials. Cox and McDaniel (1989) found that the 12 SIR passages were equivalent when used within the SIR test, a test that used subjective intelligibility ratings of the SIR passages. In the present study, however, some SIR passages were tracked faster than others. There are three reasons that may explain this discrepancy in passage equivalency across the two studies. First, the SIR test used recorded materials that have been adjusted in order to achieve equal intensity levels across the 12 passages (Cox & McDaniel, 1989), while the tracking procedure used live voice presentation. Second, both the SIR test and the tracking procedure presented the SIR passages in a background of multi-talker noise. The variability inherent in multi-talker noise has been shown to impact the intelligibility of the SIR passages (Beck & Speaks, 1993). Finally, the equivalency of the SIR passages was measured using subjective intelligibility ratings (Cox & McDaniel, 1989); the present investigation used the SIR passages within a different procedure, connected discourse tracking. The use of different procedures may result in differences in intelligibility (Beck & Speaks, 1993). It was expected that the tracking procedure would be more variable than the SIR test due to the live-voice presentation. It is encouraging, however, that the tracking procedure did result in equivalent tracking rates across 8 of the 12 SIR (high-context) passages and across all 7 low-context passages.

In order to use the high-context and low-context passages to differentiate among hearing aids, the differences in tracking rate must be controlled for. In the next experiment, tracking performance was measured for only the eight high-context passages that were found to be equivalent in the first experiment. For the low-context materials, no significant differences in tracking speed were measured across the seven passages. The four passages with the smallest variability in tracking speeds across the 5 subjects (these are indicated in Figure 4) were selected for use in the second experiment.

EXPERIMENT II: EVALUATION OF TRACKING MATERIALS

The purpose of this experiment was to determine if the tracking procedure was able to discriminate between two different hearing aid microphone conditions.

Previous research has demonstrated that passage difficulty influences tracking speed (De Filippo et al., 1994; Hochberg et al., 1989). Both low-context and high-context materials were evaluated because it was hypothesized that test sensitivity would improve with low-context materials.

Subjects

A new group of 6 young adults (83% female) with normal hearing participated in this experiment. Subjects were students in the audiology and speech language pathology graduate programs at the University of Louisville. Their ages ranged from 23 to 29 years with a mean age of 25 years.

Talker

Two graduate students served as the talkers in this experiment. Each talker tested 3 subjects. All statistical analyses were performed within subjects; talker effects should not influence the test results.

Hearing Aid

Each subject was fit with the same hearing aid used in Experiment I, using the same procedure to adjust the frequency gain characteristic. Tracking performance was measured for two different hearing aid microphone modes, omnidirectional and directional. Within the directional mode the hearing aid could be adjusted among three polar plots: diffuse, side, or rear.

The Hearing In Noise Test (HINT) was the "gold standard" for this evaluation. The HINT is a commonly used laboratory sentence test that has shown good intra-subject test reliability (Nilsson, Soli, & Sullivan, 1994). HINT performance was measured for each directional microphone condition (diffuse, side, and rear) and the polar plot was selected that resulted in a small but significant improvement in HINT scores for the directional mode versus the omnidirectional mode. (Nilsson et al. reported a 95% confidence limit of 1.5 dB.) Subjects' HINT scores improved by at least 2 dB in the directional mode, with an average improvement of 3.29 dB, compared to performance in the omnidirectional mode.

Materials

High-context materials. The eight passages, which were found to be equivalent in Experiment I, were used in the present experiment. Each subject heard the eight different topics in a random order, alternating between the omnidirectional and directional microphone conditions.

Low-context materials. The four passages marked in Figure 4 were used in the present experiment. As in Experiment I, each passage was divided into three sections. The first two sentences comprised the first section, which was used as practice. The second and third sections were used to evaluate tracking speed. Each subject heard the four passage topics in a random order. The subjects switched between omnidirectional mode and directional mode before each section; the hearing aid mode for the first passage was selected randomly across subjects.

Procedure

Each subject was evaluated with the high-context materials first and then evaluated with the low-context materials during a second test session. The tracking procedure was the same as in Experiment I. The SNR was determined while subjects listened to the practice passages in the omnidirectional mode. The SNR was determined separately for the high-context and low-context passages.

Results

High-context materials. The data were analyzed in two ways in order to determine if presentation order, passage topic, or hearing aid mode affected tracking speed. In order to look at topic, a repeated measures ANOVA was performed. There was no significant effect for the factor passage Topic, $F(7, 35) = 0.67$, $p = .69$. A second repeated measures ANOVA was performed for the factors Presentation Order and Hearing Aid Mode. The factor Presentation Order was not significant, $F(3, 15) = 0.36$, $p = .79$), but the factor Hearing Aid Mode was significant, $F(1, 5) = 47.93$, $p < .00$. As in Experiment I, the subjects tracked across topic and across presentation order with similar speed. There were, however, large differences in tracking speed in omnidirectional ($M = 42.13$ wpm) versus directional ($M = 61.98$ wpm) modes.

It is important to consider whether the tracking procedure can differentiate between hearing aid modes within an individual subject. The mean tracking performance for each subject and for each hearing aid mode are shown in Figure 5 and Table 1. Individual *t* tests were performed in order to determine whether

Table 1

Mean Intra-Subject Tracking Rate in WPM (and Standard Deviations) for Each Hearing Aid Mode for High- and Low-Context Speech Materials. Probability Levels from the Individual *t* Tests are Also Shown.

	High-context materials					Low-context materials				
	Omni		Directional			Omni		Directional		
Subject	*M*	*(SD)*	*M*	*(SD)*	*p*	*M*	*(SD)*	*M*	*(SD)*	*p*
1	43.48	(6.9)	63.81	(7.2)	<.01	39.24	(9.2)	59.93	(6.8)	.06
2	36.84	(5.4)	62.70	(5.2)	<.01	31.75	(5.4)	53.89	(6.5)	.03
3	55.45	(8.9)	66.16	(5.8)	.09	33.48	(9.3)	49.73	(6.7)	.01
4	30.78	(7.2)	60.64	(6.1)	<.01	31.03	(12.2)	55.38	(9.7)	.02
5	36.00	(3.3)	53.45	(6.8)	<.01	36.26	(8.7)	58.05	(7.2)	.02
6	50.22	(1.6)	65.39	(3.9)	<.01	32.42	(7.0)	61.93	(2.9)	<.01

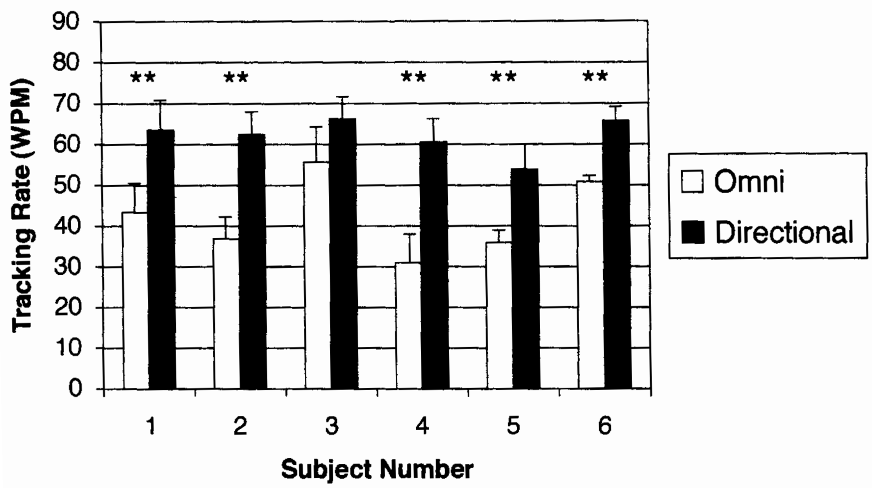

Figure 5. Individual subject mean tracking rate for high-context passages heard under two hearing-aid microphone conditions. Mean tracking rate in the omnidirectional microphone mode is shown with the unfilled bars and mean tracking rate in the directional mode is shown with the black bars. A significant difference between the omnidirectional and directional microphone modes at the .05 level is shown with a single asterisk; and at the .01 level, with two asterisks.

there were significant differences in tracking speed between the microphone modes. The individual *t* tests showed significant differences for 5 of the 6 subjects at the .01 level. Subject 3 was the only subject who did not track with significantly different wpm rates between the directional and omnidirectional modes; this subject did have a significant difference in HINT scores (3.29 dB) between the two microphone modes.

Low-context materials. The data were analyzed in two ways in order to determine if presentation order, presentation sub-section, passage topic, or hearing aid mode affected tracking speed. In order to look at topic and sub-sections within a topic, an ANOVA was performed with Sub-Section being a nested factor within Topic and with Subject as a random factor. There was no significant effect for passage Topic, $F(3, 15) = 2.77$, $p = .78$. There was also no significant effect for topic Sub-Section, $F(4, 20) = 0.13$, $p = .97$; as in Experiment I, performance did not improve as a function of sub-section within a topic and performance was similar across the four topics. A repeated measures ANOVA was performed for the factors Presentation Order and Hearing Aid Mode. Presentation Order was not significant, $F(3, 15) = 0.46$, $p = .71$, but the factor Hearing Aid Mode was significant, $F(1, 5) = 157.33$, $p < .00$. This indicates that the 6 subjects tracked passages across topics, across presentation order, and across presentation sub-sections with

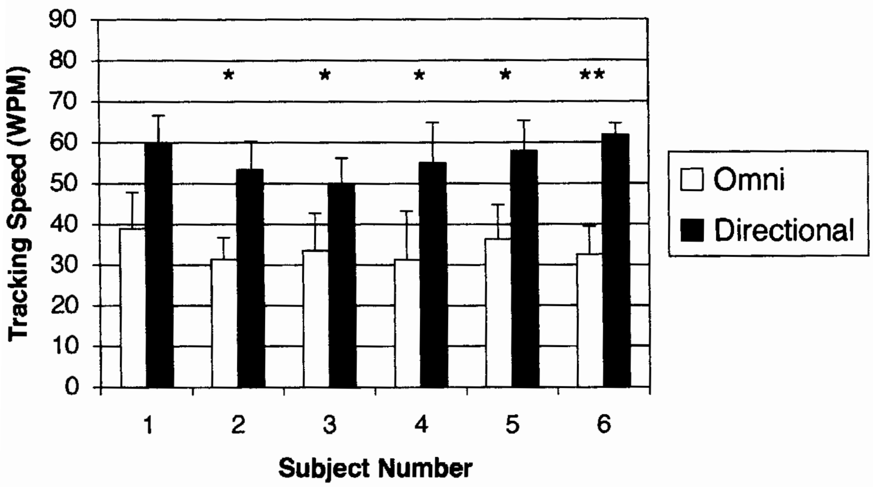

Figure 6. The same as Figure 5, for low-context passages.

similar speed. There were, however, large differences in tracking speed in omnidirectional ($M = 34.03$ wpm) versus directional ($M = 56.48$ wpm) modes.

Again, it is also important to consider whether the tracking procedure can differentiate among hearing aid modes within an individual subject. The mean tracking performance for each subject and for each hearing aid mode are shown in Figure 6 and Table 1. Individual t tests were performed in order to determine whether there were significant differences in tracking speed across the microphone modes. The individual t tests showed significant differences for 4 of the 6 subjects at the .05 level and for 1 subject at the .01 level. Subject 1 was the only subject who did not track with significantly different wpm rates between the directional and omnidirectional modes, but the difference in tracking rate did approach significance, $t(3) = 2.93$, $p = .06$. This subject did have a significant difference in HINT scores (4.47 dB) between the two microphone modes.

Discussion

Intra-subject sensitivity. Previous research has shown that the tracking procedure was able to differentiate among presentation conditions (Geier & Norton, 1992) or cochlear implant processing strategies (Faulkner et al., 2000; Fujiki et al., 1998). These reported results were for groups of subjects; performance was not evaluated for individual subjects. The results in the present study also showed significant differences in tracking rate across two hearing aid microphone modes for a group of 6 subjects. When looking at performance for individual subjects, connected discourse tracking was able to differentiate between hearing aid modes for 5 of the 6 subjects both for high-context and low-context speech materials.

The procedure was not sensitive enough in order to differentiate between hearing aid modes in 100% of the subjects.

It is possible that talker effects could explain the limited sensitivity of the tracking procedure. In other words, individual talker characteristics could have overridden the microphone effects. As it turns out, subject number 3 and number 1 (the 2 subjects who did not show significant differences in tracking speeds across the two microphone modes), listened to different speakers; however, due to the small number of subjects and talkers in the present experiment, the importance of talker effects cannot be ruled out.

It is likely that individual subject sensitivity would improve through the use of additional test passages. It took approximately 30 min to administer the high-context passages and 40 min to administer the low-context passages. While additional passages may improve the sensitivity of the test, the long testing time may not be desirable. The time to perform connected discourse tracking in the laboratory, however, would be less than the time it takes to administer a field trial. In a typical field trial, subjects rate hearing aid performance on a daily basis over a 2- to 10-week period. Also, field trials typically include multiple laboratory visits for hearing aid adjustments and subjective reporting of hearing aid benefits (Preminger & Cunningham, 2003).

High-context versus low-context materials. It was hypothesized that the low-context materials would have improved sensitivity over the high-context materials. This was not the case. Mean tracking rates for both materials were significantly different between the hearing aid modes for 5 of the 6 subjects. The SNR was adjusted separately prior to tracking the high-context and then the low-context test passages. The mean SNRs across the 6 subjects in Experiment II were -15 dB for the high-context passages and -12.8 dB for the low-context passages; this was not a significant difference, $t(5) = 2.29$, $p = .071$. Five of the 6 subjects tracked the high-context passages with a more difficult SNR than for the low-context passages. In the first experiment subjects tracked the high-context passages at an average rate of 47.25 wpm and the low-context passages at an average rate of 47.45 wpm. Two different sets of subjects tracked the high- and low-context passages, so it is difficult to make comparisons across type of passage. Nevertheless, it appears that text complexity did not influence the test sensitivity.

Based on the fact that text complexity did not influence the test sensitivity it is recommended that future studies use high-context materials. The high-context materials used in the present study were at a seventh grade reading level; these should be appropriate for subjects with a wide range of educational experience. In addition, the results of the individual t tests (see Figures 5 and 6 and Table 1) showed increased sensitivity for the high-context materials as compared to the low-context materials.

Areas for future research. Walden (1982) described the "direct observation" field study technique, which entails the training of observers to tally appropriate and inappropriate communication behaviors and strategies used by listeners who wear hearing aids in a variety of communication situations. The connected discourse tracking method could be evaluated as a pseudo direct observation method. For example, instead of adjusting the SNR to achieve a constant error rate (as was done in the present experiment), listeners with hearing loss can be tested with the same SNR. An SNR can be selected that is representative of a typical communication setting (e.g., Pearsons, Bennett, & Fidell, 1976, found the SNR of conversation in public places averages between 5 and 10 dB). The tracking procedure in the sound booth would allow the examiner to observe the subject's overall communication abilities in a simulated communication environment, including the subject's use of repair strategies. This is information that is not typically measured in a laboratory investigation.

Results of the present investigation indicate that this pseudo direct observation method holds promise. Subjects with hearing loss could be evaluated with the high-context materials that were used in the present experiment, given future research to determine passage equivalency and test sensitivity for individuals with hearing loss who wear hearing aids.

Conclusions. Researchers and clinicians continue to search for sensitive and reliable field study evaluation techniques. Hearing aid users wear their hearing aids in the real world and this is where we would like to evaluate them. Until better field study methods are available, connected discourse tracking can be used as a pseudo field study evaluation technique. The tracking method allows the examiner to observe an individual's communication performance in a carefully controlled conversational setting. While this procedure is time consuming, it can be a sensitive and useful technique.

ACKNOWLEDGMENTS

This project was funded by the Mary and Mason Rudd Surgical Research Fund, Jewish Hospital Foundation, Louisville, KY. The author would like to thank Amanda Jones and Lauren Welch for assistance with data collection and Tim Trine at Starkey Laboratories for providing the experimental hearing aid. Portions of this paper were presented at the Academy of Rehabilitative Audiology Summer Institute, June 2003.

REFERENCES

Beck, W.G., & Speaks, C. (1993). Intelligibility of selected passages from the Speech Intelligibility Rating (SIR) test. *Journal of Speech and Hearing Research, 36,* 1075-1082.

Cox, R.M., & McDaniel, D.M. (1984). Intelligibility ratings of continuous discourse: Application to hearing aid selection. *Journal of the Acoustical Society of America, 76,* 758-766.

Cox, R.M., & McDaniel, D.M. (1989). Development of the Speech Intelligibility Rating (SIR) test for hearing aid comparisons. *Journal of Speech and Hearing Research, 32,* 347-352.

De Filippo, C.L. (1984). Laboratory projects in tactile aids to lipreading. *Ear and Hearing, 5,* 211-227.

De Filippo. C.L., Lansing, C.R., Elfenbein, J.L., Kallaus-Gay, A., & Woodworth, G.G. (1994). Deriving passage difficulties for a tracking study via the Cloze technique. *Journal of the American Academy of Audiology, 5,* 366-378.

De Filippo, C.L., & Scott, B.L. (1978). A method for training and evaluating the reception of ongoing speech. *Journal of the Acoustical Society of America, 63,* 1186-1192.

Faulkner, A., Rosen, S., & Smith, C. (2000). Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 108,* 1877-1887.

Fenn, G., & Smith, B.Z.D. (1987). The assessment of lipreading ability: Some practical considerations in the use of the tracking procedure. *British Journal of Audiology, 21,* 253-258.

Fujiki, N., Naito, Y., Hirano, S., Kojima, H., Kamoto, Y., Nishizawa, S., Konishi, J., & Honjo, I. (1998). Influence of a speech-coding strategy on cortical activity in cochlear implant users: A positron emission tomographic study. *Acta Oto-Laryngologica, 118,* 797-802.

Geier, L.L., & Norton, S.J. (1992). The effects of limiting the number of Nucleus 22 cochlear implant electrodes programmed on speech perception. *Ear and Hearing, 13,* 340-348.

Hochberg, I., Rosen, S., & Ball, V. (1989). Effect of text complexity on connected discourse tracking rate. *Ear and Hearing, 10,* 192-199.

Jenkins, H., Chmiel, R., & Jerger, J. (1989). Speech tracking in the evaluation of a multichannel cochlear prosthesis. *Laryngoscope, 99,* 245-251.

Lunato, K.E., & Weisenberger, J.M. (1994). Comparative effectiveness of correction strategies in connected discourse tracking. *Ear and Hearing, 15,* 362-370.

Matthies, M.L., & Carney, A.E. (1988). A modified speech tracking procedure as a communicative performance measure. *Journal of Speech and Hearing Research, 41,* 394-404.

McDaniel, D.M., & Cox, R.M. (1992). Evaluation of the Speech Intelligibility Rating (SIR) test for hearing aid comparisons. *Journal of Speech and Hearing Research, 35,* 686-693.

Nilsson, M., Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95,* 1085-1099.

Owens, E., & Raggio, M. (1987). The UCSF tracking procedure for evaluation and training of speech reception by hearing-impaired adults. *Journal of Speech and Hearing Disorders, 52,* 120-128.

Owens, E., & Telleen, C.C. (1981). Tracking as an aural rehabilitative process. *Journal of the Academy of Rehabilitative Audiology, 14,* 252-258.

Pearsons, K.S., Bennett, R.R., & Fidell, S. (1976). *Speech levels in various environments* (Report No. 3281). Cambridge, MA: Bolt, Beranek and Newman, Inc.

Preminger, J.E., & Cunningham, D.R. (2003). A case-study analysis of various field study measures. *Journal of the American Academy of Audiology, 14,* 39-55.

Punch, J.L., Robb, R., & Shovels, A.H. (1994). Aided listener preferences in laboratory versus real-world environments. *Ear and Hearing, 15,* 50-61.

Schoepflin, J.R., & Levitt, H. (1991). Continuous discourse tracking: An analysis of the procedure. *Journal of Communication Disorders, 24,* 237-249.

Skinner, M.W., Binzer, S.M., Fredrickson, J.M., Smith, P.G., Holden, T.A., Holden, L.K., Juelich, M.F., & Turner, B.A. (1988). Comparison of benefit from vibrotactile aid and cochlear implant for postlingually deaf adults. *Laryngoscope, 98,* 1092-1099.

Stone, M.A., Moore, B.C.J., Alcántara, J.I., & Glasberg, B.R. (1999). Comparison of different forms of compression using wearable digital hearing aids. *Journal of the Acoustical Society of America, 106,* 3603-3619.

Tye-Murray, N., & Tyler, R.S. (1988). A critique of continuous discourse tracking as a test procedure. *Journal of Speech and Hearing Disorders, 53,* 226-231.

Valente, M., Fabry, D.A., Potts, L.G., & Sandlin, R.E. (1998). Comparing the performance of the Widex SENSO digital hearing aid with analog aids. *Journal of the American Academy of Audiology, 9,* 342-360.

Walden, B.E. (1982). Validating measures for hearing aid success. In G. Studebaker & F.H. Bess (Eds.), The Vanderbilt Hearing Aid Report. *Monographs in Contemporary Audiology,* 188-192.
Walden, B.E. (1997). Toward a model clinical-trials protocol for substantiating hearing aid user-benefit claims. *American Journal of Audiology, 6,* 13-24.

APPENDIX A

AN EXAMPLE OF A HIGH-CONTEXT PASSAGE[1]

***Goose* (108 words)**

The goose is a/ web-footed bird/ which is related to/ the duck and the swan./ A goose is/ larger than a duck/ but smaller/ than a swan./ About forty different/ varieties of geese/ live in different/ parts of the world./ Seventeen kinds/ of wild geese/ live in the/ United States and Canada./ These geese are/ twenty to forty/ inches long,/ with wide wings./ They are all/ migratory birds,/ flying north in summer/ and south in the winter./ Some of them/ fly as far north/ as the Arctic Circle,/ and some as/ far south/ as Mexico./ The goose/ has a neck/ a little longer/ than that of a duck.

APPENDIX B

AN EXAMPLE OF A LOW-CONTEXT PASSAGE

***Case of Female Mimicry: Smelling Like a Queen* (255 words total)**

As insect colonies go,/ a nest of/ tropical ants/ can be a/ real snake pit,/ a hotbed of/ aggressive male behavior./ Sexual competition/ is so intense/ that wingless males/ fight one another/ to the death/ for the privilege of/ mating with queens. (42)

But there is/ more than one/ type of rooster/ in this henhouse./ There are winged/ males, too,/ and they are/ much more docile./ One would think/ that the wingless males/ would run roughshod/ over them,/ doing the ant equivalent/ of kicking sand/ in their faces,/ and prevent them/ from mating./ But a team of scientists from/ Germany, Denmark and Italy/ has now shown/ that this is/ not the case./ Winged males,/ they say,/ are not attacked/ by their/ more aggressive brethren/ – and they still/ get their share of/ romance./ Writing in/ the journal *Nature,*/ the researchers/ also describe why/ this happens./ They display a/ remarkable example/ of female mimicry. (109)

The winged ants/ produce chemicals/ that are very similar/ to those produced/ by the queens./ The scent of/ the chemicals,/ which are on/ the surface of/ the ants' bodies,/ trick the more/ aggressive males/ into thinking that/ the winged ones/ are female./ The wingless ants/ not only/ do not fight/ the winged ants,/ they often try/ to mate with them./ The researchers say/ most examples of/ female mimicry/ within a species/ involve physical/ or behavioral similarities./ Only a few examples/ involving chemicals/ are known./ In these cases,/ the males doing/ the mimicking are weaker./ They may survive,/ but do not have/ much luck in love. (105)

---

[1]Adapted from "SIR Test," on *The HARL Speech Intelligibility Tests* [CD], the Hearing Aid Research Laboratory at the University of Memphis. http://www.ausp.memphis.edu/harl/sir.html