Training listeners to identify the sounds of speech: II. Using SPATS software

By James D. Miller, Charles S. Watson, Doris J. Kistler, Jill E. Preminger, and David J. Wark

In the first section of this two-part article, several lines of research on auditory perceptual learning were shown to support the proposition that users of hearing aids and cochlear implants would benefit from systematic training on the new "speech code" represented by the *modified sounds* they experience through those devices. Research also suggests that improved abilities to use the speech code can improve the recognition of meaningful sentences. We proposed several criteria for training systems designed to take advantage of what has been learned over the past half century about auditory perceptual learning of both speech and other complex sounds. In this second part we describe such a system and summarize the results of a validation study conducted with it.

The Speech Perception Assessment and Training System (SPATS)* is designed to improve a listener's perception of natural everyday speech. It consists of two independent speech-recognition testing and training programs, one for the constituents of speech, referred to earlier as "the code," and the other for sentences. The purpose of the constituent training program is to sharpen the listener's attentional focus on those spectral-temporal properties that specify the elements of syllables, onsets (single consonants and clusters), nuclei (vowels and vowel-like sounds), and codas (final consonants and clusters). A sentence module trains a combination of bottom-up and top-down processing in a novel auto-scoring identification task in which listeners are encouraged to make strong use of linguistic context.

SYLLABLE CONSTITUENT TRAINING

There are more than 212 different syllable constituents in spoken English, including at least 68 onsets, 28 nuclei, and 116 codas. Analysis of a textual database led us to select 109 of these (45 onsets, 28 nuclei, and 36 codas) as most important for the perception of English. Importance is jointly determined by an item's frequency of occurrence in running text and the number of different English words in which the constituent occurs.

For testing and training, each constituent type is subdivided into four levels. Level I includes the most important 25% of those selected; Level II the 50% most important; Level III the 75% most important; and Level IV all (100%) of the selected constituents. To provide a variety of phonetic contexts, onsets were combined with four different nuclei, codas were attached to five stems, and nuclei were placed in an h(nucleus)d context. The resulting

388 syllables were recorded by each of eight speakers with Middle-American accents. For example, the onset /pl/ was combined with four nuclei resulting in "plee," "plah," "ploo," and "pler." Each of these four onset-nucleus pairs was spoken by eight speakers resulting in 32 different recordings of the onset /pl/.

A typical training screen, in this case for Level 3 (75% most important) syllable onsets, is displayed in Figure 1. The onsets are arranged by the place and manner of their production. The rows are color-coded to call attention to properties shared by all elements in a given row; for example, onsets containing "l-sounds" are always in a lemon-colored row.

On each trial the listener hears one of the constituents and is asked first to imitate the sound and then to identify it by clicking on one of the buttons in the display. Correct and incorrect responses are signaled by changing both letter colors and background colors. For example, in Figure 1 the correct response was /sk-/, but the participant selected /g-/. The listener is encouraged to click on the highlighted buttons to "rehear" the correct onset and to hear the incorrectly identified onset. This is called "post-trial rehearing" and its availability speeds perceptual learning.

A novel "adaptive item selection" algorithm is used to control how often each constituent is presented. The greatest amount of training is devoted to constituents that are moderately difficult for the client and most likely to be
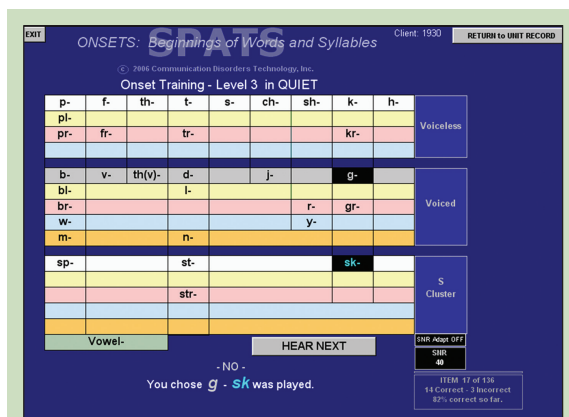
Figure 1. *A response screen (Onsets Level 3) used in SPATS. The listener initiates the trial, imitates the sound heard, and clicks on the appropriate response button. The screen immediately shows a correct response (background turns black, letters turns cyan). In the sample above an error was made. The correct onset (sk-) is shown, as is the error (g-).*

---

* Patent pending

learnable. As individual constituents are learned and become recognized more often, they are presented less often. This contributes to efficient learning, as little time is wasted with items that are too easy or too difficult. Focusing on a moderate level of difficulty reduces the participants' frustration and increases their motivation as they realize that they have become able to recognize constituents that were initially quite difficult.

The SPATS system implicitly teaches the phonetic structure of English and the relations between production (articulation) and perception. Observing their errors in a phonetically meaningful space not only seems to help listeners recognize the constituents of speech more accurately, but also helps them learn their likely errors. This knowledge can be very helpful in decoding everyday speech.

One of the positives that users frequently report is that for the first time they appreciate what their hearing loss really means in terms of their ability to identify certain sounds of speech but not others. This awareness may benefit both the participant and the audiologist. Participants may learn that despite being fitted with hearing aids, they still find certain speech sounds inaudible or indistinguishable from others. The audiologist may be able to use this information in the selection of a hearing aid or cochlear implant or in the programming of either device.

Some of the features of SPATS are illustrated below with results from a very successful hearing aid user, Client 520, whose audiogram is shown in Figure 2.

Although Client 520 has used hearing aids successfully for many years, this client nonetheless demonstrated improvement with about 24 hours of SPATS training—2 hours a week for 12 weeks on onsets, nuclei, and sentences (the sentence task is descri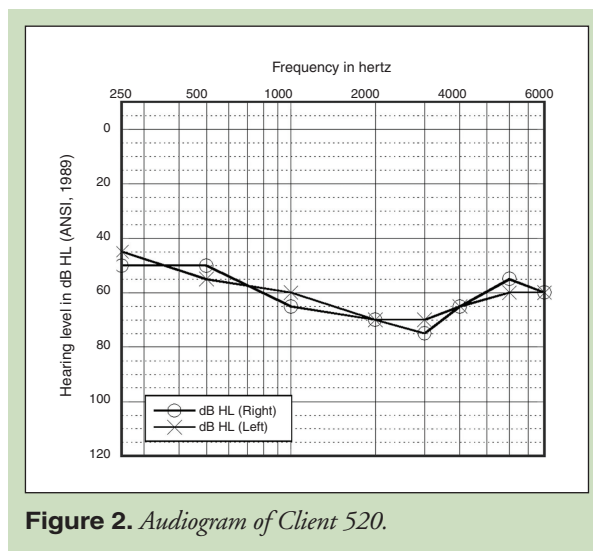bed below). The progress of constituent training can be evaluated in two ways: by monitoring overall signal-to-noise ratio (SNR) and by monitoring performance in quiet.

Training to hear syllable constituents in noise (multi-talker babble) is accomplished by adapting the SNR, using an algorithm that converges on a specific target percent correct (e.g., 70%). The training gradually enables the listener to tolerate

Figure 2. *Audiogram of Client 520.*

higher levels of noise while achieving the same percent correct. Prior to training, Client 520 was able to identify 70% of the onsets correctly when the SNR was at 6 dB. After training, she could maintain a 70% correct performance with an SNR of only 3 dB.

The normal SNR for this condition (based on the performance of listeners with normal audiograms) is about -4 dB. Expressed in percentage of norm,* this client improved from 77% to 84% of norm. Clients with similar training have gained an average of 16% for all trained constituent types. These and other observations support the conclusion that hearing aid users can learn to "listen down" into noise and improve their SNRs from about 60% of norm to 76% of norm. This is also consistent with the results of training listeners with normal hearing to detect weak components in a masking background, as discussed in our previous article.

Participant performance in quiet may also be evaluated by measuring the "mastery category" assignment for each constituent during pre-training and post-training tests. SPATS testing and training automatically and objectively assigns each item in a list of constituents to one of five mastery categories (100, 75, 50, 25, or 0). Items in Category 100 (Very Easy) are almost always correctly heard. Items in Category 50 (Medium) are correctly heard about half the time. SPATS training moves constituents from lower to higher mastery categories. The results of such training are shown for Client 520 in Figure 3.

Each item is placed in a row based on the mastery score obtained in a pre-training test. The number in each of the cells shows the mastery score after training. For example, the onset (g-) is in the middle row and had a mastery score of 50-Medium on the pre-training test, but a mastery score of 100-Very Easy on the post-training test as indicated within its cell. Green backgrounds indicate improved mastery with training. Yellow backgrounds indicate no change. Red backgrounds indicate a decrease in mastery.

Notice that performance remained at high levels (yellow backgrounds in the top row) or improved (green backgrounds) for 28 of the 34 items tested. Performance declined (red backgrounds) or remained at low levels (yellow backgrounds in the bottom rows) for only 6 items. This client shows a clear average increase in the mastery of onsets and the average improvement on this scale was 21 points. Our experience to date indicates that clients with a similar course of training will gain an average of 14 scale points for all constituent types.

The recognition of syllable onsets and syllable nuclei is very important in word recognition. Therefore, the reported gains in their recognition in quiet and in noise imply important improvements for everyday listening, as they are achieved despite a variety of phonetic contexts and talkers. To ensure transfer from constituent training and to motivate such training, SPATS intermixes constituent training with training on spoken sentences presented in multitalker babble. Sentence training is described below.

---

* %Norm = [(SNR$_{client}$ − 40)/SNR$_{norm}$ − 40)]*100

Mastery categories		Syllable onsets (Level 3) for client # 520											
Very easy	100	bl-100	d-100	gr-100	l-100	m-100	r-100	v-100	str-100	w-100	p-75	pr-75	
Easy	75	br-100	f-100	fr-100	h-100	kr-100	sk-100	t-100	b-0	sp-50			
Medium	50	g-100	k-100	n-100	s-100	tr-100	y-100						
Difficult	25	Vwl-100	sh-75	th(v)-50	v-25								
Very difficult	0	ch-100	j-50	pl-75	th-0	Ave. gain = 21 points							

Figure 3. *Changes in onset mastery categories (see text for explanation).*

SPATS SENTENCE TRAINING

The sentence module consists of 1000 recorded sentences, four to seven words in length. The sentences have simple syntax and common everyday words. Many of the sentences are chosen as examples of statements or questions that a hearing-impaired person would want to be able to hear, such as "We are under a tornado alert" or "What would you like for dinner?" The response screen format is shown in Figure 4.

The words in the sentence appear in their proper positions in the line above the response alternatives after they are correctly identified. The alphabetical list of alternative responses includes three foils for each of the words in the sentence (the target words), plus the target words themselves. The foils are selected to share one or more syllable constituents with the target words, either their onsets, nuclei, or codas. Thus the list will contain 28 words for a seven-word sentence. The listener first hears the sentence and then sees the list. She then selects any words that she hears by clicking on them, in any order. If a foil is selected, it changes color to red, and the sentence is replayed. This process continues until all of the target words are correctly identified.

The task includes a very strong cognitive component, since the first word or two identified greatly reduce the subset of likely target words within the remaining options. Listeners are encouraged to use that contextual information and to attempt to correctly identify all of the words as rapidly as possible.

Listeners in our training studies on the average decoded 360 sentences naturally spoken by 10 different talkers. To accomplish this each listener had to identify about 2000 naturally spoken words. Sentences are presented at a variety of SNRs during training in the range from +15 to –15 dB. Pre- and post-training sets of 30 sentences were presented to eight listeners. Analysis of the data indicates that performance on the sentence task improved with training and that such improvement amounted to the equivalent of about a 5-dB reduction in the SNR required to achieve a targeted level of performance.

SUMMARY OF SPATS

SPATS trains all of the significant constituents of English syllables. The system uses *adaptive item selection* to automatically focus training on learnable items. Constituent training is organized to emphasize items in their order of importance. The client's performance in noise is presented in terms of percent of normal performance, a metric that can be understood by clients more easily than the abstract concept of signal-to-noise ratio.

In addition, clients can be given a detailed listing of their speech-perception problems separately for syllable onsets, nuclei, and codas. In this way, hearing aid and cochlear implant users can begin to truly understand their hearing loss. For example, one SPATS user had been told by his audiologist that he had a "moderate-to-severe high-frequency sloping loss." Yet he said that he never truly understood the implications of the audiogram for everyday life. Participating in SPATS provided both this patient and his audiologist with objective evidence of the entire spectrum of speech elements that could and could not be recognized.

Seeing such evidence will give audiologists a clear rationale for providing training on the problem sounds, a rationale that is easy to communicate to their clients. The client also becomes informed about the role of top-down processing by the sentence task. SPATS intermixes training with fluent, naturally spoken sentences with syllable-constituent drills. Its sentence training is objectively scored and it trains all of the bottom-up listening skills in combination with the top-down use of linguistic context. Although not discussed here, SPATS offers the option of constructing a programmed curriculum tailored to an individual client's needs, which then automatically guides training.

Training results

Preliminary evaluations of the SPATS training program have been reported at national meetings and will not be detailed here. However, these evaluations have demonstrated improvements in the identification of syllable constituents and sentences in quiet and noise by users of hearing aids and cochlear implants. In addition, these gains in performance have been shown to generalize to other standard tests of word and sentence recognition, such as the W22 word lists, the CNC word lists, and the HINT sentences.
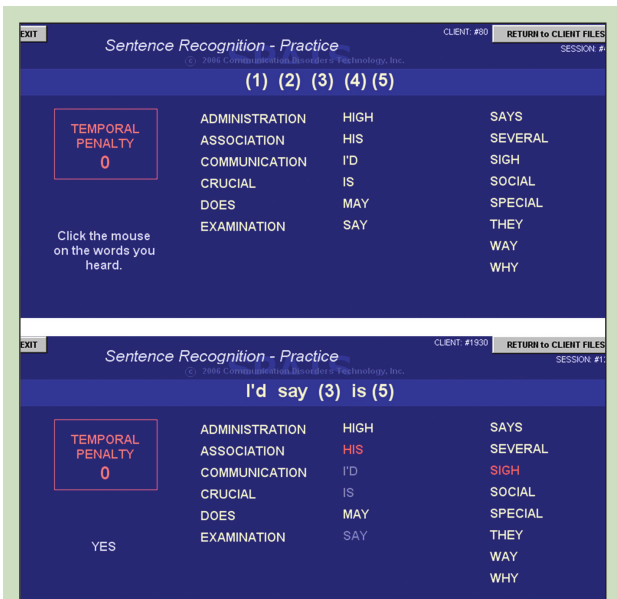
Figure 4. *Sentence screens for the sentence "I'd say communication is crucial." The top screen is seen immediately after the first hearing of the sentence. The bottom screen is seen when the sentence has been partially recognized. The client works until all words have been recognized and appear in the top row. Correct words turn light blue when clicked. Incorrect words turn red when clicked.*

SPATS, LACE, and other computerized systems

Many readers will recognize that this system is designed to achieve goals similar to those of the LACE system developed by Sweetow and now marketed by NeuroTone, Inc., although the two systems use different general approaches to speech training.

SPATS was developed during essentially the same period of time as LACE. The initial proposal describing it was submitted to NIH in 2002; thus neither system is likely to have benefited from knowledge of the other. From the data collected on SPATS and those obtained with LACE, it appears that both programs result in enhanced speech recognition by hearing aid users.

The most significant differences between the two systems are the training of syllable constituents in SPATS, which is not done with LACE, and the requirement in SPATS that listeners follow a systematic training curriculum intermixing constituent and sentence training. While sentence training may have greater face validity, the adaptive item strategy used in SPATS identifies the specific speech contrasts with which listeners have difficulty and then provides intensive training to focus the listener's auditory attention on the critical acoustic properties of those specific sounds. The contribution of that narrowly targeted training

to sentence recognition remains to be evaluated in controlled experiments in which some listeners have constituent training and others do not.

It also remains to be seen whether the SPATS and LACE training systems are complementary or redundant. Other computerized training systems are available for purchase or free from their developers (see Table 1). Each has positive features and can be useful in an aural rehabilitation program.

A FINAL WORD

This is an exciting time for rehabilitative audiology. There is growing evidence that computer-based auditory training is beneficial and can be accomplished in the audiologist's office or the patient's home.[2-12]

It is time for audiologists and individuals with hearing loss to recognize the value of auditory training. For many users it could represent the difference between satisfaction with amplification and a return for credit. Speech-perception training for users of hearing aids or cochlear implants should become as common as physical therapy for people learning to use artificial limbs. For at least some users it is equally important.

When effective training is used in conjunction with a new device (aid or implant), it is very likely that device benefit, satisfaction, and use will improve. While the structures of the damaged cochlea are unlikely to improve as a result of training, there is evidence to suggest that

through mechanisms of brain plasticity the impacts of training include changes in the brain that improve or increase the resources available for the task of speech recognition by hearing-impaired persons.[13,14]

James D. Miller, PhD, is the Principal Scientist at Communication Disorders Technology, Inc., (CDT) of Bloomington, IN, and an Adjunct Professor of Speech and Hearing Sciences at Indiana University and Director of Research Emeritus at Central Institute for the Deaf, St. Louis. **Charles S. Watson**, PhD, is President of CDT and Professor Emeritus in Speech and Hearing Sciences at Indiana University. **Doris J. Kistler**, PhD, is a Research Professor and **Jill E. Preminger**, PhD, is an Associate Professor of Audiology in the Department of Surgery, both at the University of Louisville. **David J. Wark**, PhD, is an Associate Professor of Audiology in the School of Audiology and Speech-Language Pathology of the University of Memphis. Readers are invited to contact Dr. Watson at watson@indiana.edu or Dr. Miller at jamdmill@indiana.edu.

REFERENCES

1. Miller JD, Watson CS, Kewley-Port D, et al.: SPATS: Speech perception assessment and training system. *J Acoust Soc Am* 2007;122(5):3063 (Abstract).
2. Burk M, Humes LE, Amos NE, Strauser LE: Effect of training on word recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear Hear* 2006;27:263-278.
3. Burk M, Humes LE: Effects of training on speech recognition performance in noise using lexically hard words. *J Sp Lang Hear Res* 2007;50:25-40.
4. Burk M, Humes LE: (In press) Effect of training on speech-recognition performance in noise using lexically easy and hard words in older adults with hearing impairment. *J Sp Lang Hear Res*
5. Fu Q-J, Galvin JJ: Perceptual learning and auditory training in cochlear implant recipients. *Trends Amplif* 2007;11:193-205.
6. Miller JD, Dalby JM, Watson CS, Burleson DF: Training experienced hearing-aid users to identify syllable-initial consonants in quiet and noise. *J Acoust Soc Am* 2004;115(5):2387.
7. Miller JD, Dalby JM, Watson CS, Burleson DF: Training experienced hearing-aid users to identify syllable constituents in quiet and noise. Presentation at ISCA Workshop on Plasticity in Speech Perception (PSP2005), June 2005, London: A46.
8. Miller JD, Watson CS, Kistler DJ, et al.: Preliminary evaluation of the speech perception assessment and training system (SPATS) with hearing-aid and cochlear-implant users. *J Acoust Soc Am* 2007;122(5):3063 (Abstract).
9. Stecker GC, Bowman GA, Yund EW, et al.: Perceptual training improves syllable identification in new and experienced hearing aid users. *J Rehab Res Dev* 2006;43:537-552.
10. Sweetow R, Henderson-Sabes J: The case for LACE: Listening and auditory communication enhancement training. *Hear J* 2004;57:32-38.
11. Sweetow R, Palmer CV: Efficacy of individual auditory training in adults: A systematic review of the evidence. *JAAA* 2005;16:494-504.
12. Sweetow RW, Sabes JH: The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program. *JAAA* 2006;17:538-558.
13. Kraus N, Banai K: Auditory processing malleability: Focus on language and music. *Curr Dir Psychol Sci* 2007;16:105-109.
14. Bacon S, ed: ASHA 2006 Research Symposium: Issues in the Development and Plasticity of the Auditory System. *J Comm Disord* 2007;40(6):433-536.

Table 1. *Several computerized auditory training systems.*

Product name	Company/URL
Speech Perception Assessment and Training System (SPATS)	Communication Disorders Technology, Inc. www.comdistec.com
Computer Assisted Speech Training (CAST)	Tiger Speech Technology www.tigerspeech.com
Internet Based Open-Set Speech Recognition I-Star	Tiger Speech Technology www.tigerspeech. com
Listening And Communication Enhancement (LACE)	Neurotone www.neurotone.com
Listening Room	Advanced Bionics www.bionicear.com
Seeing And Hearing Speech	Sensimetrics www.sens.com
Sound And Beyond	Cochlear www.cochlearamericas.com