

# Speech Perception Assessment and Training System (SPATS-ESL) for Speakers of Other Languages Learning English<sup>1</sup>

James D. Miller, Roy Sillings, Charles S. Watson, and Diane Kewley-Port

*Communication Disorders Technology, Inc.  
Indiana University Research Park  
501 N. Morton Street, Sta 215  
Bloomington, Indiana 47404*

**Abstract.** The SPATS software system (1, 2, 3, & 4), originally developed for the hearing-impaired, has been modified for use with ESL learners (5 & 6) with TOFEL (pbt) scores ranging from slightly below to well above 500. SPATS-ESL includes the identification of syllable constituents: onsets, nuclei, and codas as well as sentence recognition. The syllable constituent tasks include the progressive introduction of increasing numbers of constituents until the learner becomes adept at the identification of 45 onsets, 28 nuclei, and 36 codas presented by 8 talkers in a variety of phonetic contexts. The sentence task emphasizes increasing speed and decreasing errors in the recognition of short, meaningful sentences spoken by ten talkers. The sentences are presented in a background of multi-talker babble at five signal-to-noise ratios: +10, +5, 0, -5, and -10 dB. The syllable constituent and sentence tasks are interleaved throughout training. In constituent training, SPATS uses a proprietary training algorithm, Adaptive Item Selection (AIS), which automatically focuses training on individuals' items of intermediate difficulty and is independent of their language history. Proctored tests allow certification of a learner's English speech perception in relation to native-speaker performance.

**Keywords:** Speech, Perception, English, Testing, Training, Cross-Language, ESL  
**PACS:** 43.71.Hw

## INTRODUCTION

SPATS-ESL has five modules: **I. Sentence Module, II. The Syllable-Constituent Module, III. The Curriculum Module, IV. The Report Module, and V. The Proctored Test Module.** Each is briefly described in the sections that follow. (More detailed presentations of various aspects of the philosophy and implementation of SPATS software can be found in References 1, 2, 3, & 4.)

### The Sentence Module

The Sentence Module provides practice in top-down and combined top-down and bottom up speech perception skills. One thousand sentences have been recorded by ten different talkers. Each is spoken naturally with rate of speech, intonation patterns and stress patterns selected by the talker. Therefore, the range of phonetic accommodations that occur in everyday speech are found in this corpus. As will be explained the scoring of the sentence task is objective and entirely computer based.

---

<sup>1</sup> This some of this material was presented as a poster (5) at the Acoustical Society of America 2<sup>nd</sup> Special Workshop on Speech: Cross-Language Speech Perception and Linguistic Experience held in Portland, OR 21-23 May, 2009.

*The Basics of the Sentence Task*

A spoken sentence of three to seven words is presented. A screen then appears that shows “slots” for each word at the top and an alphabetic list of words that contains the spoken words plus three phonetically similar foils for each. The user is instructed to click on the words that they thought they heard. Correctly selected words are dimmed and appear in the appropriate slot in the header. When a foil is selected, it turns red, an error is recorded, and the sentence is replayed. Whenever the listener pauses five seconds without responding, a “temporal penalty” is assessed. Beginning and final screens for a five-word sentence are shown next in Figures 1 & 2.

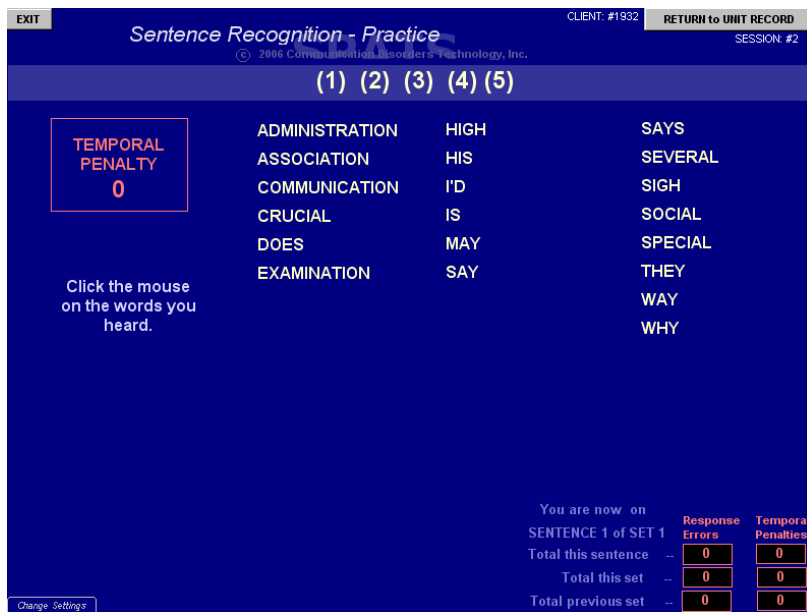


Figure 1. Sentence screen seen after hearing a five word sentence presented in multi-talker babble.

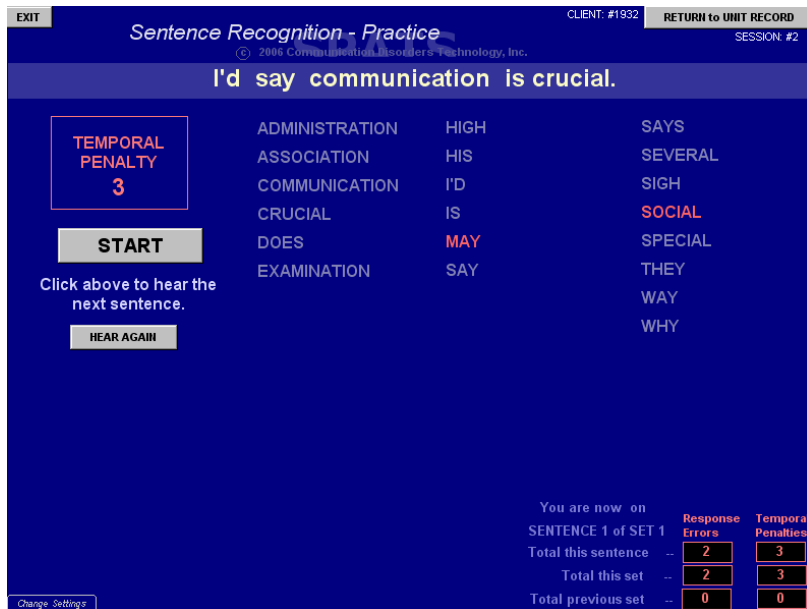


Figure 2. Sentence screen seen after completing identification of the words in the sentence. There were two errors (shown in red) and three temporal penalties. The effective percent correct for this sentence is  $(5/(5+2+3)) * 100$  or 50% as explained below.

In each group of 15 sentences, three are at each of five signal-to-noise ratios (SNRs): +10, +5, 0, -5, & -10 dB. The learner is shown his overall effective percent correct after the completion of each group of 15 sentences. The effective percent correct is the total number of words divided by the number words plus the number of errors plus the number of temporal penalties times 100. One learner's progress is shown in Table 1. Native speakers score 90 and above on this task.

**TABLE 1.** One learner's progress on the Sentence Task

<i>Sentences</i>	<i>Words</i>	<i>Errors</i>	<i>Temporal Penalties</i>	<i>Effective Percent Correct</i>	<i>Performance Category</i>
1-15	84	20	20	68	Needs Work
16-30	85	16	12	75	Fair
31-45	82	10	8	82	Fair
46-60	87	5	8	87	Very Good
61-75	85	9	9	83	Good

### *Kinds of Sentence Units*

Sentence tests consist of 15 sentences to be solved consecutively. Sentence training consists of groups of six or three sentences interspersed between syllable constituent testing and training as explained in section III, The Curriculum Module. The SNR's are randomized such that three sentences will be presented at each of the five SNR's +10, +5, 0, -5, & -10 dB in each successive group of 15 sentences whether in tests or in training.

## **II. The Syllable-Constituent Module**

### *Materials Used*

Syllables are made up of three types of constituents: onsets, nuclei and codas. Onsets are the beginnings of syllables and consist of consonants, consonant clusters, or their absence, that is, the syllable begins with a vowel. Nuclei are middles of syllables and consist of vowel and vowel-like sounds. Codas are endings of syllables and consist of final consonants, constant clusters, or their absence, that is, the syllable ends with a vowel. . Syllable constituents are chosen as the "building blocks" of spoken language rather than phonemes because the articulatory and acoustic characteristics of phonemes depend on both phonetic context and position within the syllable.

An Americanized Celex database was studied to determine the textual and lexical frequencies of occurrence of word-initial onsets, all nuclei, and word-final codas. Only word-initial onsets and word-final codas were studied because the database does not assign as onsets or codas consonants and consonant clusters falling within a word. An importance was assigned to each constituent within a type based on the average of their ranks in textual and lexical importance. Based on this study it was determined that spoken English has 45 onsets, 28 nuclei, and 36 codas of significance. Other constituents occur very rarely or only in foreign loan words and were deemed safe to omit.

Within each constituent type the items were grouped into quartiles with the first quartile containing the 25% most important items within a type, the second quartile containing with the next 25% , the third quartile containing the next 25%, and fourth quartile containing the 25% least important. Within SPATS, for each constituent type the quartiles are cumulated to form Cumulative Levels. Cumulative Level 1 includes the items in the first quartile, Cumulative Level 2 includes all of items in the first and second quartiles, Cumulative Level 3 includes all of the items in the first, second and third quartiles. Cumulative Level 4 includes all of the significant items of a constituent type, that is, 45 onsets, 28 nuclei, or 36 codas.

Charts 1, 2, and 3 below show the organization of syllable constituents used in testing and training in SPATS-ESL. In the lists of onsets and codas in Charts 1 & 3, the letters "Vwl" refer to syllables that begin or end without consonants. Nearly 20% of words have no initial consonant or consonant cluster, and, similarly, nearly 20% of words end without a consonant or consonant cluster. Chart 2, the list of Nuclei, deviates from the organization described above. It was quickly determined from early trials with ESL students that the first quartile of nuclei

(ranks 1-7) was extremely difficult for ESL learners as it contained /i/, /I/, /ε/, /æ/, /ɑ/, /Λ/, and /σ/. Of these seven only the /σ/ as in “heard” was clearly identifiable to ESL learners. Therefore, nuclei were selected for Cumulative Level 1 that represented extremes when plotted in Miller’s Auditory Perceptual Space (7). Gradually more interior nuclei are added as training progresses through Cumulative Levels 1, 2, 3, & 4. This sequence was found to be highly efficient for ESL learners in our samples. Notice that Chart 2 contains the importance ranking for each item even though the items are not arranged in importance quartiles as they are for onsets and codas. Nonetheless, the *average* importance decreases from the first to the fourth quartile.

**Chart 1.** List of syllable onsets used in SPATS-ESL.

List of Onsets				
Col. 1	Col. 2	Col. 3	Col. 4	The number next to each onset is its importance based on its average rank in lexical and textual frequency of occurrence
1 VwI	12 p-	24 thv-	35 dr-	
2 s-	13 l-	25 br-	36 kw-	Each onset is spoken by 8 talkers and combined with 4 vowels: “ee,” “ah,” “oo,” & “er.”
3 r-	14 n-	26 gr-	37 fl-	
4 k-	15 pr-	27 kr-	38 sl-	
5 b-	16 g-	28 sp-	39 kl-	
6 m-	17 v-	29 pl-	40 sw-	
7 h-	18 st-	30 th-	41 sm-	
8 d-	19 sh-	31 y-	42 gl-	
9 t-	20 fr-	32 sk-	43 sn-	
10 w-	21 j-	33 bl-	44 skr-	
11 f-	22 tr-	34 str-	45 z-	
	23 ch-			
Cumulative Level 1 = Col. 1				
Cumulative Level 2 = Col. 1 + 2				
Cumulative Level 3 = Col. 1 + 2 + 3				
Cumulative Level 4 = Col. 1 + 2 + 3 + 4				

**Chart 2.** List of syllable nuclei used in SPATS-ESL. Each was recorded in the context shown by eight talkers.

List of Nuclei					
Col. 1	Col. 2	Col. 3	Col. 4	The number next to each nucleus is its importance based on its average rank in lexical and textual frequency of occurrence	
2 heed	8 hayed	1 hid	5 head		
3 heard	12 hulled	6 hud	10 hawed		
4 had	13 hoed	14 hoard	15 how'd		
7 hod	16 hewed	18 haired	23 Hal'd		
9 hide	17 hard	19 hood	24 hold		
11 whoed	20 hilled	26 hired	25 hauled		
22 held	21 hoid	28 heeled	27 hailed		
Cumulative Level 1 = Col. 1					
Cumulative Level 2 = Col. 1 + 2					
Cumulative Level 3 = Col. 1 + 2 + 3					
Cumulative Level 4 = Col. 1 + 2 + 3 + 4					

Chart 3. List of syllable codas used in SPATS-ESL.

List of Codas			
Col. 1	Col. 2	Col. 3	Col. 4
1 Vwl	10 -m	19 -bz	28 -ch
2 -z	11 -nt	20 -ks	29 -sts
3 -n	12 -v	21 -p	30 -pt
4 -d	13 -k	22 -dz	31 -ngz
5 -t	14 -nz	23 -mz	32 -vz
6 -l	15 -ts	24 -f	33 -ps
7 -ng	16 -r	25 -kt	34 -sh
8 -s	17 -st	26 -th	35 -g
9 -nd	18 -ns	27 -j	36 -b

The number next to each coda is its importance based on its average rank in lexical and textual frequency of occurrence

Each coda is spoken by 8 talkers and combined with 5 stems: "ee," "ah," "oo," "er," & "el."

**Cumulative Level 1 = Col. 1**  
**Cumulative Level 2 = Col. 1 + 2**  
**Cumulative Level 3 = Col. 1 + 2 + 3**  
**Cumulative Level 4 = Col. 1 + 2 + 3 + 4**

In the Tables 2, 3, & 4 that follow, the basic composition of testing and training exercises are documented for each combination of constituent type and cumulative level. In practice a listener is graduated from one level to the next when a performance score reaches the minimum level expected for a native speaker of English. Notice that SPATS procedures require the learner to identify the sound presented as distinguished from all others in the set and, therefore, the learner must be able to make all of the possible pairwise discriminations in order to reach a correct decision. On tests or benchmark units each item in the set is presented twice with the order of items randomized. On training units the numbers of trials are adjusted to match the difficulty encountered by the ESL learners. As a rule, it has been found for ESL learners that the order of increasing difficulty of constituent types is onsets, codas, and nuclei. Therefore, the number of trials in training unit has been set to four times the number of items for onsets, five times the number of items for codas, and seven times the number of items for nuclei. This also has the advantage that the total numbers of trials are approximately equal across constituent types. Notice that these tables include a cumulative level called 4 Maintenance (4M). This level includes all of the items of level 4, but the number of trials is reduced to two times the number of items. When a learner reaches a criterion score at Cumulative Level 4, the learner is shifted to Level 4M.

TABLE 2. Items and trials per testing or training unit for cumulative levels of Onsets

<i>Cumulative Level</i>	<i>Number Onsets</i>	<i>Onset Contrasts</i>	<i>Trials per Item Tests</i>	<i>Total Trials Tests</i>	<i>Total Trials Training</i>
1	11	55	2	22	44
2	23	253	2	46	92
3	34	561	2	68	136
4	45	990	2	90	180
4 Maintenance	45	990	2	90	90

TABLE 3. Items and trials per testing or training unit  
for cumulative levels of Nuclei

<i>Cumulative Level</i>	<i>Number Nuclei</i>	<i>Nucleus Contrasts</i>	<i>Trials per Item Tests</i>	<i>Total Trials Tests</i>	<i>Total Trials Training</i>
1	7	21	2	14	49
2	14	91	2	28	98
3	21	210	2	42	147
4	28	378	2	56	196
4 Maintenance	28	378	2	56	56

TABLE 4. Items and trials per testing or training unit  
for cumulative levels of Codas

<i>Cumulative Level</i>	<i>Number Codas</i>	<i>Coda Contrasts</i>	<i>Trials per Item Tests</i>	<i>Total Trials Tests</i>	<i>Total Trials Training</i>
1	9	36	2	18	45
2	18	153	2	36	90
3	27	351	2	54	135
4	36	630	2	72	180
4 Maintenance	36	630	2	72	72

Table 5 below shows the total numbers of constituents and contrasts tested and trained at each cumulative level. It is believed that the requirement that an ESL student master nearly all of the 1998 contrasts induces the learning of the dimensions and the categories of the English sound system. This is consistent with Kingston's (8) view that the ESL learner must learn to attend to the dimensions and boundaries of the English sound system. This lays the foundation for rapid, accurate perception of spoken English and provides a necessary foundation for the acquisition of correct pronunciation and accent reduction.

TABLE 5. Items and trials per testing or training unit  
summed over constituent types

<i>Cumulative Level</i>	<i>Number Constituents</i>	<i>Constituent Contrasts</i>	<i>Total Trials Tests</i>	<i>Total Trials Training</i>
1	27	112	54	138
2	55	497	110	280
3	82	1122	164	418
4	109	1998	218	556
4 Maintenance	109	1998	218	218

#### *Response Screens*

The ESL learners are taught to identify the constituent heard by clicking on virtual response buttons that are labeled orthographically and phonetically arranged on a computer screen. For each constituent type and cumulative level, the client is first introduced to the screen and then proceeds with testing and training. Screens are always introduced in order from Cumulative Levels 1 to 4. By the time the trainee reaches the most complex screens at Level 4 he/she has become very familiar with previous screens. The screens for Cumulative Levels 1 & 4 are presented in Figures 3, 4, and 5, below. The response screens for syllable onsets at Cumulative Levels 1 & 4 are shown in Figure 3. Onsets are placed into three groups. The topmost group has releasing consonants that are voiceless. The middle group has releasing consonants that are voiced. The bottom group includes s-clusters. In this case the releasing consonant is always an /s/. In the voiceless and voiced groups, the columns represent the place of articulation of the releasing consonant while the rows represent the manner of articulation of the distinguishing elements in a column. For s-clusters the columns and rows similarly represent the distinguishing phonemes in the

cluster. Syllables that begin without a consonant, that is syllables that begin with a vowel, are represented by the virtual button labeled, Vowel-, at the bottom of the screen. This organization is not explicitly taught to the users. However, errors are usually simply related on the screen. For example, place of articulation errors are usually side by side in a row, manner errors are usually in the wrong row, but in the same column and group as the signal, while voicing errors have a similar location as the signal but in the wrong voiceless/voiced group. This arrangement seems to make it easy for the learner to grasp the nature of their errors such as whether they missed the l-sound or the r-sound in a cluster or whether they confused ch- with sh-, or f- with th-, and so on.

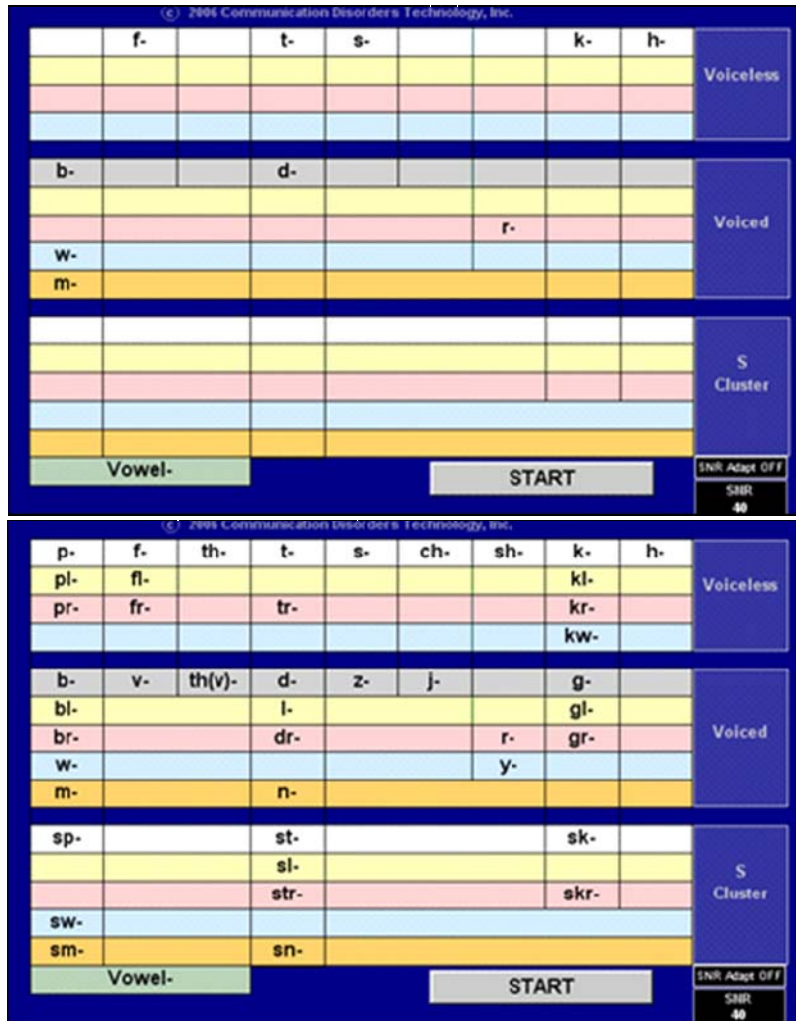


Figure 3. Response screens for onsets: Cumulative Level 1 above & Cumulative Level 4 below.

VOWELS	VOWELS with R	VOWELS with L
	heard	
had		
		held
heed		
hide		
hod		
who'd		

START

VOWELS	VOWELS with R	VOWELS with L
	heard	
had		Hal'd
hayed		hailed
head	haired	held
heed		heeled
hewed		
hid		hilled
hide	hired	
hod	hard	
hawed		hauled
hoed	hoard	hold
hold		
hood		
how'd		
hud		hulled
who'd		

START

Figure 4. Response screens for syllable nuclei: Cumulative Level 1 above & Cumulative Level 4 below.



As shown Figure 4, the virtual keys on the vowel response screens are organized into columns. The usual vowels and diphthongs are in the left hand column as they do not include a rhotic or the vowel-like sonorant lateral, dark-l. The rhotic vowel [er] and rhotic diphthongs are in the second, rose colored column. Diphthong-like nuclei formed with the sonorant lateral are located in the right-most, lemon colored column

Cumulative Level 1 above								
			-t	-s				Voiceless
Voiceless								Voiced
			-d	-z				
Voiced								Nasal Clusters
			-n				-ng	
				-nd				
-Vowel						-L		

Cumulative Level 4 below								
-p	-f	-th	-t	-s	-ch	-sh	-k	Voiceless
-pt				-st			-kt	
-ps			-ts	-sts			-ks	
Voiceless								Voiced
-b	-v		-d	-z	-j		-g	
-bz	-vz		-dz					
Voiced								Nasal Clusters
-m			-n				-ng	
			-nt	-nd				
-mz			-nts	-nz			-ngz	
-Vowel			-R			-L		

Figure 5. Response screens for syllable codas: Cumulative Level 1 above & Cumulative Level 4 below.

The response screens for codas are organized into voiceless, voiced and nasal groups. The columns represent the place of articulation of the consonant that immediately follows the nucleus. The left-most columns represent the front of the mouth while right-most columns represent the back of the mouth. Syllables that end in one of the usual vowels or diphthongs (i.e. non-rhotic or non-lateral) require a response on the virtual button labeled, “-Vowel.” Syllables that end with a rhotic or a dark-el require responses on buttons labeled “-R” and “-L”, respectively. However, if a syllable is closed by one of the items represented on the screen above the “-Vowel,” “-R,” or “-L” buttons, then the listener must respond on the appropriate button in Voiceless, Voiced, or Nasal groups. For example, if the presented syllable is “er,” a response on the “-R” button is required, but if the presented syllable is “erp,” a response on the “-p” button is required.

### Benchmark (Testing) and Training Units

Work with syllable constituents is broken into Benchmark (Testing) and Training Units. (The term Benchmark is used for tests because it is believed that some students may experience high anxiety or embarrassment during a “Test” and perhaps less negative affect during a “Benchmark” unit.) The number of trials in a unit depends on the constituent type and cumulative level as shown in right-and columns of Tables 2, 3, & 4. However, the procedures

for selecting the items to be presented are very different for Benchmark as opposed to Training Units. In Benchmark or Testing Units, items are selected randomly with the restriction that each item is presented twice. For example, a benchmark unit for syllable onsets at Cumulative Level 2 (see Table 2) would consist of 46 trials wherein each of 23 items would be presented two times. A Training Unit would consist of 92 trials, but the items would be selected by the Adaptive Item Selection (AIS) algorithm. The AIS algorithm presents items of intermediate difficulty for the individual learner with much higher probability than items that are either very easy or very difficult for that learner. For a given constituent type and cumulative level, four training units are scheduled to be run between successive benchmark units. If BU stands for Benchmark Unit and TU stands for Training Unit then the programmed sequences are always BU-TU-TU-TU-TU-BU and so on.

#### *Scoring of Benchmark (Testing) and Training Units*

At the end of every unit, two scores are calculated. These are the familiar Percent Correct (PC) and the average Item Mastery Score (IMS). The PC is simply the number of correct responses divided by the number of trials multiplied by 100. The IMS at the end of a unit is computed as follows. At the beginning of a benchmark unit all items in constituent type and cumulative level are set to an IMS of 50. Every time an item is correctly identified its IMS is increased by 25 points until it reaches a maximum of 100. Every time error occurs the IMS's of both stimulus and the erroneous response are decreased by 25 points until a minimum of 0 is reached. At the end of a unit, the mean IMS for all of the items in a set is calculated as the IMS for that unit. The IMS scores for items are carried over from unit to the next until the next benchmark unit is reached, when all of the items are again assigned a value of 50. The best measure of performance for a benchmark unit is the PC. The best measure of performance for a training unit is the end-of-unit IMS. The PC is not an informative measure of performance on training units because the PC will be influenced by the AIS algorithm which presents items of intermediate difficulty with higher probability than items that are either very easy or very difficult. The PC for a benchmark run measures the current level of performance over an entire unit. The end-of-training-unit IMS measures how well the learner can perform by the end of a training unit which may include short-term gains that are not yet consolidated into permanent memory. In general, on benchmark units PC's are higher than end-unit IMS's while on training units IMS's are higher than PC's. However, once all of the items are nearly mastered, then both the PC's and the IMS's are high for both benchmark and training units. Experience has shown that high IMS's at the end of training units are 'harbingers' of high PC's on benchmark units. These concepts are illustrated in the section on reports of progress.

#### *Actual and Adjusted Scores*

The Percent Correct and end-of-unit IMS scores can be calculated as "actual scores" or "adjusted scores." The actual scores are simply based on the items in the Constituent Type and Cumulative Level of the unit being tested or trained. The adjusted scores are corrected to reflect the proportion of items in a set relative to the number of items at Cumulative Level 4. To find the value of an adjusted score one multiplies Cumulative Level 1 scores by 0.25, Cumulative Level 2 scores by 0.50, and Cumulative Level 3 scores by 0.75, and Cumulative Level 4 scores by 1.00. In this way, adjusted scores reflect both the performance at a given cumulative level as well as progress toward Cumulative Level 4.

### **III. The Curriculum Module**

#### *Background*

The Curriculum Module allows the user to design customized sequences of training and testing with sentences, onsets, nuclei, and codas. These sequences can be simple or very complicated and designed to include a variety of contingencies.

#### *The Default Curriculum*

Through experience, a particular curriculum has been found to be effective with ESL learners. This is referred to as the SPATS-ESL *Default Curriculum*. The code for the Default Curriculum is presented in Chart 4 and explained in the associated text. This curriculum begins with an Introduction to the sentence task. Next, a 15-sentence pretest is given with sound to evaluate how well the learner can use both hearing and inference to identify the words in a sentence. Following the sentence pretest, the curriculum rotates through a repeating sequence of constituent and sentence units. A complete rotation includes an onset unit, six sentences, a nucleus unit, six sentences, a coda unit and 3 sentences. The first time a new combination of constituent type and cumulative level is encountered, the learner is given a brief Introduction to the relevant response screen. Notice that there are 15 sentences presented in each rotation. The sentences' SNR's are randomly selected with the restriction that three are at each of the SNR's of +10, +5, 0, -5, & -10 dB. The default curriculum is progressive in the sense that the student is moved to the next higher cumulative level whenever

Chart 4. *Default Curriculum*

```

Sentences INTRO
Sentences PRETEST with audio
-- Begin Rotation #1 ( on 0 of 100 )
Onsets L1 INTRO ii
Onsets L1 QUIET i
Sentences PRACTICE 2 sets of 3
Nuclei L1 INTRO ii
Nuclei L1 QUIET i
Sentences PRACTICE 2 sets of 3
Codas L1 INTRO ii
Codas L1 QUIET i
Sentences PRACTICE 1 set of 3
-- End Rotation #1 -----
-- CURRICULUM COMPLETED ---
-

```

the end-of-unit IMS reaches the value associated with the low end of the native speaker range for a particular constituent type. Once a learner reaches Cumulative Level 4 for a particular constituent type, the number of trials per item is reduced to two (see Tables 2, 3, & 4) when a criterion is met for maintenance. This criterion is met when the PC equals or exceeds the minimum native speaker score minus 5. Otherwise said, the learner must achieve a percent correct at Cumulative Level 4 is equal to or greater than 89 for onsets, 89 for nuclei, or 83 for codas to be switched to the maintenance schedule.

This curriculum works quite well in that the time spent on constituent types that are easily mastered is much less than the time spent on constituent types that are more difficult. In general it is found for ESL learners that onsets are the easiest to learn, codas intermediate, and nuclei the most difficult. Under the Default Curriculum, native speakers of English reach their asymptotic levels of performance in four to five training rotations, requiring roughly 1-2 hours to complete. ESL learners usually require many more rotations to approach native speaker accuracy.

#### IV. The Report Module and Client Feedback

Detailed reports of performance can be accessed by SPATS-ESL administrators. These reports include confusion matrices, information transmitted, IMS scores, and lists of confusions.

Clients are given feedback regarding their performance at the end of every run. They can also access a listing of the items ordered by their difficulty in the unit just completed. They can follow their progress graphically as well. For purposes of illustrating graphic feedback the progress of a Taiwanese teacher of English is illustrated. This person volunteered to work the SPATS-ESL curriculum because of interest in program and in her residual problems perceiving English as spoken by native speakers. This client's literacy in English was quite high as evidenced by her TOEFL (pbt) of 660. Her data are presented to illustrate the graphic feedback provided and to illustrate that even quite accomplished second-language users may have some significant residual perpetual problems. Figures 6, 7, & 8 show this client's progress in the identification of the constituents of English. On these Figures the adjusted PCs on Benchmark (test) units are shown in black and adjusted end-of-unit IMS scores are shown in red. Progress on the identification of words in sentences in a background of competing babble is shown for this client in Figure 9.

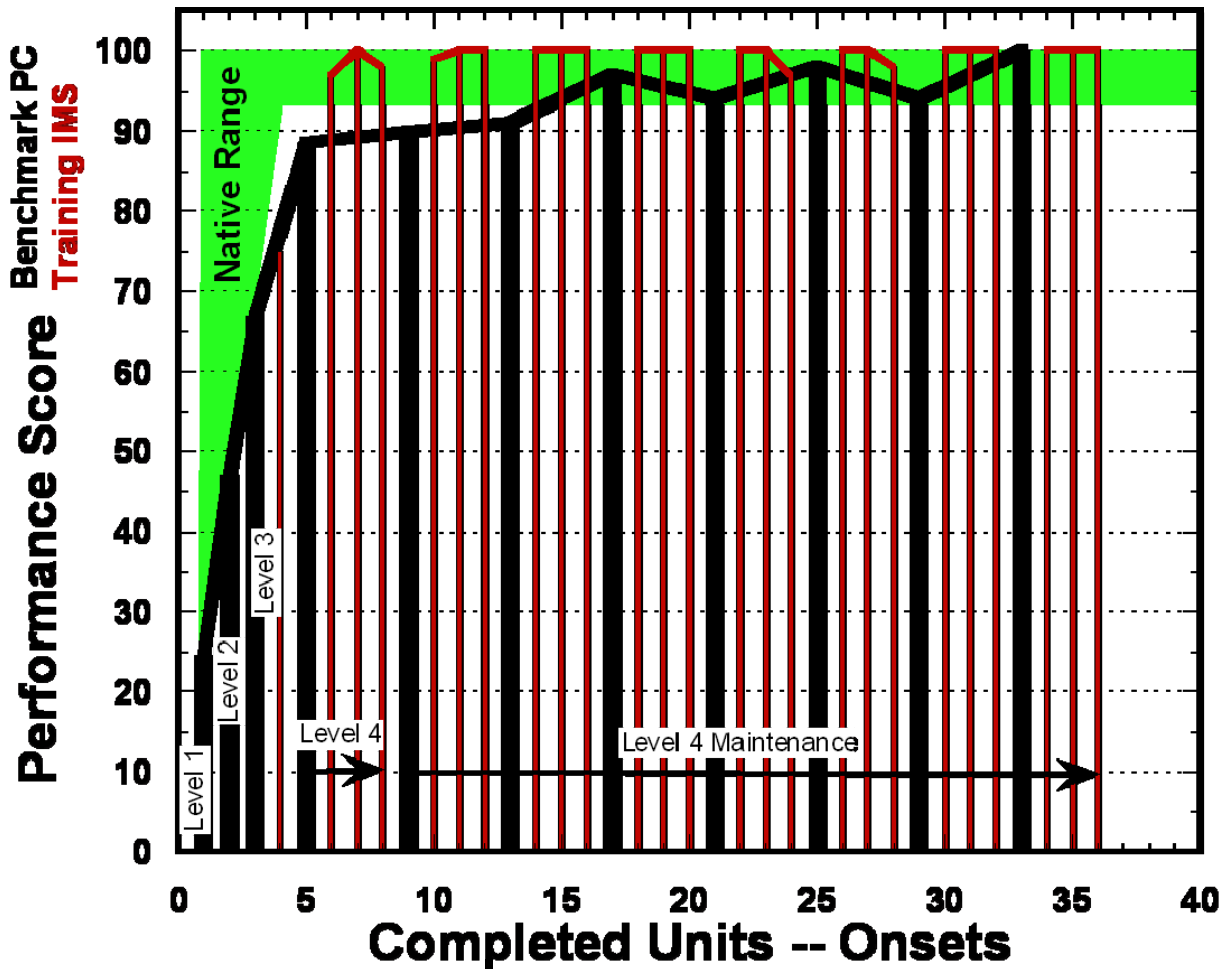


Figure 6. Showing progress in the identification of syllable onsets by an ESL student. The performance scores are “adjusted scores” as described at the end of Section II above. For this student only three training units were scheduled between benchmark units (tests).

As shown in Figure 6 above, even though this student was very good at identifying the onsets of English syllables, she still demonstrated improvement. While a native speaker of English would be expected to have reached the native speaker range of 94-100 (the green area at the top of the graph) in 4 or 5 units, this learner required nearly 15 units before the PCs on benchmark units exceeded 94%. Also notice that the adjusted end-of-unit IMS scores in red, reached high levels prior to the test PCs in black. This happens because the AIS algorithm allows that learner to focus in on subtle differences within a unit, but time is required before those short-term gains are consolidated into permanent, long-term gains. Notice that the student continued to improve during maintenance units.

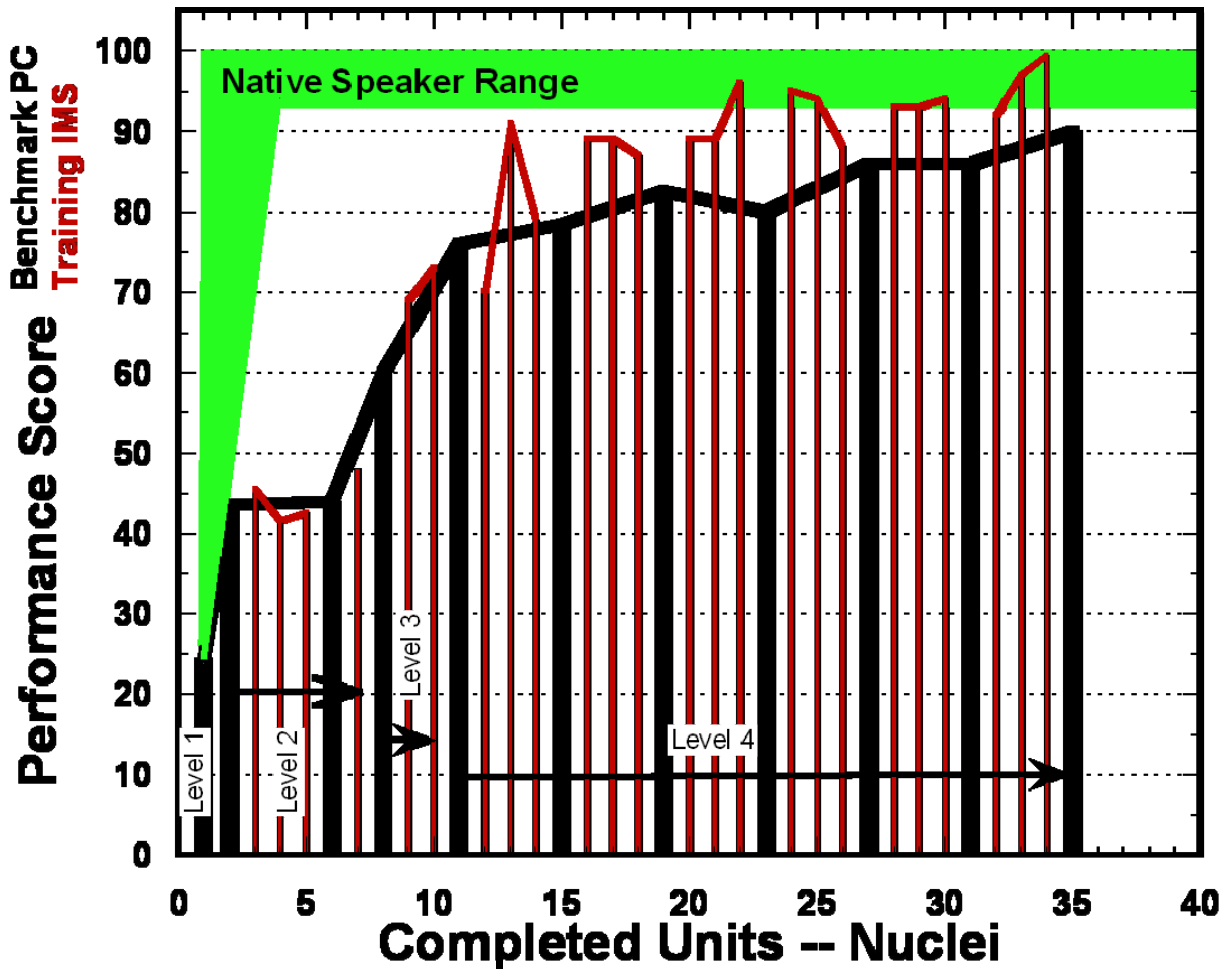


Figure 7. Showing progress in the identification of syllable nuclei by an ESL student. The performance scores are “adjusted scores” as described at the end of Section II above. For this student only three training units were scheduled between benchmark units (tests).

This same student had much more difficulty learning to identify the syllable nuclei of English as shown in Figure 7 as native speakers would reach the native speaker range in 4 to 5 units. Notice that although the ESL student approached the native speaker range, she never actually reached the native speaker range of 94 to 100 on benchmark units (tests). On the other hand, the end-of-unit IMS scores did enter the native speaker range, which indicates that with more training native speaker identification skill would be attained.

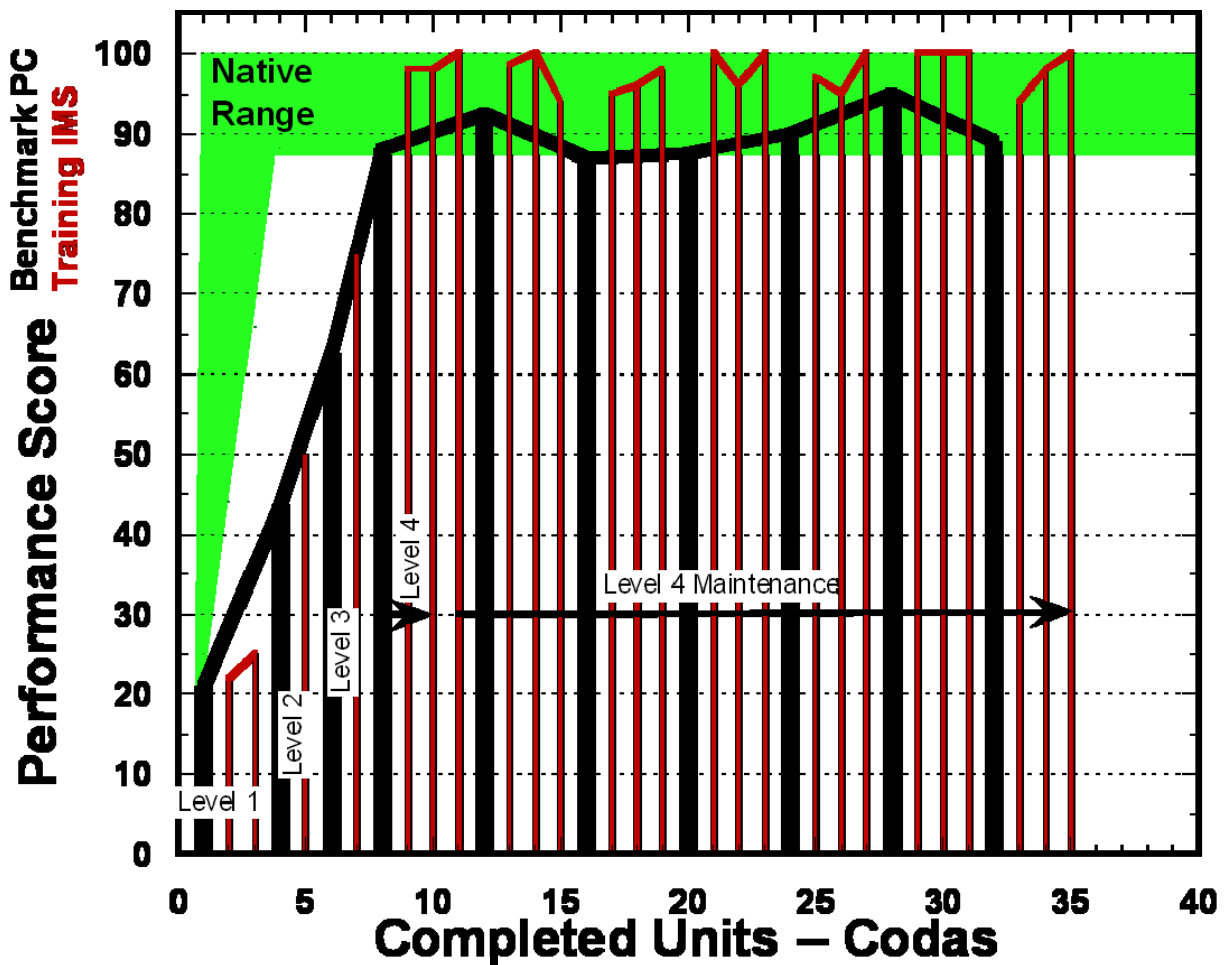


Figure 8. Showing progress in the identification of syllable codas by an ESL student. The performance scores are “adjusted scores” as described at the end of Section II above. For this student only three training units were scheduled between benchmark units (tests).

Notice that this student required a total of 8 completed units to reach a score of 88% at Cumulative Level 4. A native speaker would reach this level after 4 or 5 completed units. After reaching the normal range little additional improvement is noted. Native speakers also have similar difficulties with identification of codas, probably because codas are less precisely articulated than onsets or nuclei by native speakers.

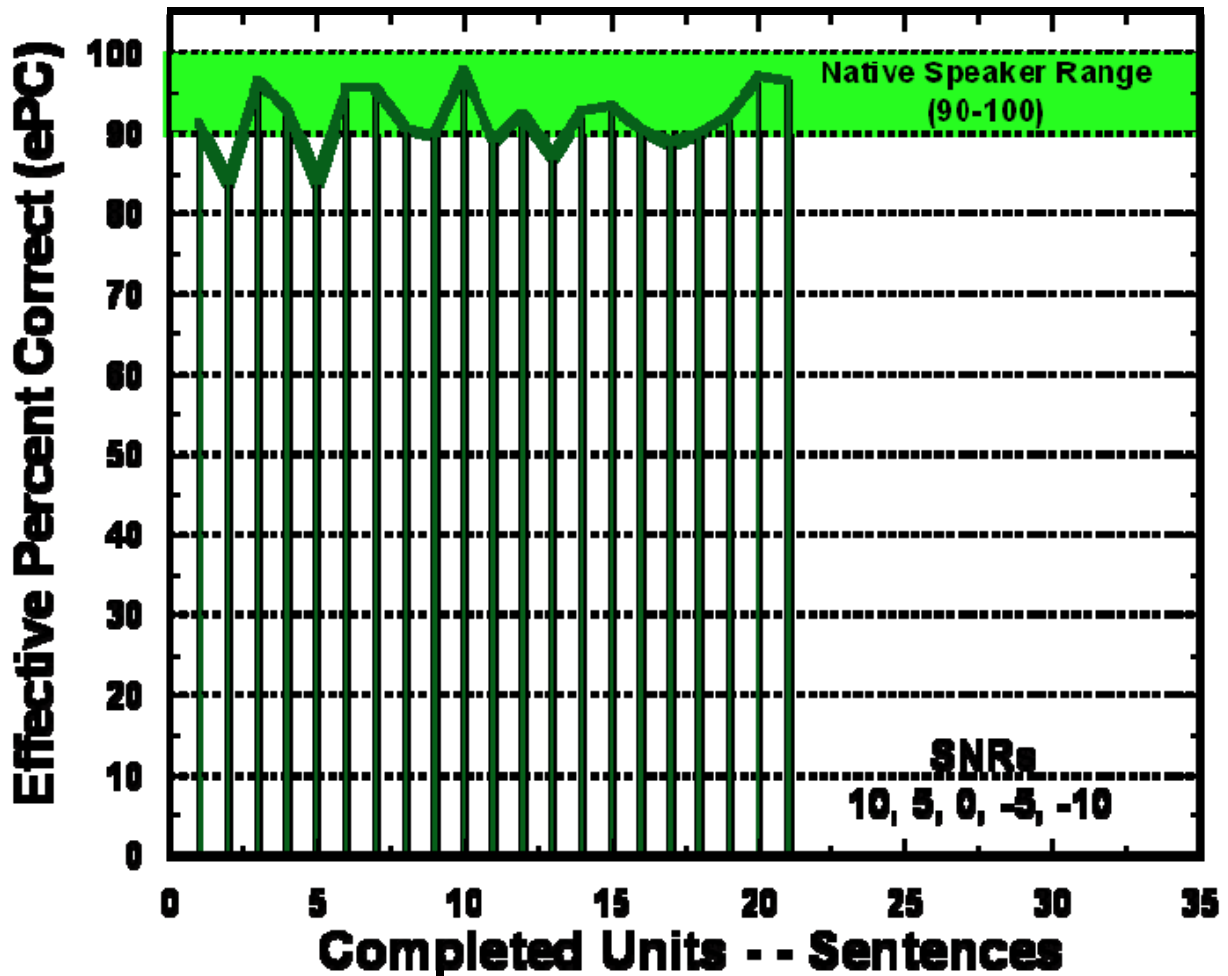


Figure 9. Showing progress in the identification of words in sentences. Each unit consists of 15 sentences with three of the sentences at each of the SNR's shown on the graph. The effective percent correct calculated by finding the ratio of the total number of words presented divided by the sum of the number words, errors, and temporal penalties and multiplying this ratio by 100. Native speakers of English score in the range from 90 to 100 after orientation to the task. Although not shown, this client only demonstrated clear improvement at the SNR of -10 dB. This graph is shown to illustrate the feedback given to all SPATS-ESL users in the latest versions of the program.

## V. The Proctored Test Module

An ESL-Learner can schedule proctored tests with a SPATS-ESL administrator. In this way a student's performance can be certified in comparison to that of native speakers of English for any combination of Constituent Type and Level and on the Sentence task.

## VI. Results

Results from 30 ESL learners using SPATS-ESL were presented at the Workshop on Cross-Language Speech Perception (6). It appears that most ESL-learners with a basic knowledge of English, pbt TOEFL scores that range from just below to well above 500, can approach the performance of native speakers of English after 15-35 hours of spaced practice on SPATS-ESL. It is hypothesized that this highly efficient perceptual learning is achieved because of the unique structure of SPATS-ESL which adapts to the individual learners needs and focuses on important perceptual skills required to accurately perceive spoken English. Moreover SPATS-ESL appears to provide the ESL learner with the skills needed to learn more English through lectures by and conversation with native speakers and to benefit from pronunciation instruction and self monitoring of speech productions.

## ACKNOWLEDGMENTS

This work was supported by Grant 5R44DC006338-03 from the National Institutes for Deafness and Other Communication Disorders to Communication Disorders Technology, Inc. (CDT) of Bloomington, IN. . Authors Kewley-Port, Miller, and Watson are stockholders in CDT and author Sillings is an employee of CDT.

## REFERENCES

- 1) Watson, C.S., Miller, J.D., Kewley-Port, D., Humes, L.E., and Wightman, F.L. (2008) The Hearing Journal, 61(9), 26-31. [http://www.audiologyonline.com/theHearingJournal/pdfs/HJ2008\\_09\\_p26-31.pdf](http://www.audiologyonline.com/theHearingJournal/pdfs/HJ2008_09_p26-31.pdf)
- 2) Miller, J.D., Watson, C.S., Kistler, D.J, Preminger, J.E., and Wark, D. J. (2008) The Hearing Journal, 61(10), 29-33. [http://www.audiologyonline.com/theHearingJournal/pdfs/HJ2008\\_10\\_p29-33.pdf](http://www.audiologyonline.com/theHearingJournal/pdfs/HJ2008_10_p29-33.pdf)
- 3) Miller, J.D., Watson, C.S., Kewley-Port, D., Sillings, R., Mills, W.B., and Burleson, D. F. (2008) Proceedings of Meetings on Acoustics, Vol. 2, 050005, 17 pp. <http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=PMARCW000002000001050005000001&idtype=cvips>
- 4) Miller, J.D., Watson, C.S., Kistler, D. J., Wightman, F.L., and Preminger, J.L. ( 2008) Proceedings of Meetings on Acoustics, Vol. 2, 050004, 9 pp. <http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=PMARCW000002000001050004000001&idtype=cvips>
- 5) Miller, JD, Sillings, R., Watson, CS, and Kewley-Port, D. (2009) J. Acoust. Soc. Am. 125(4):2755(A)
- 6) Miller, JD, Sillings, R, Watson, CS, Darcy I, and Bardovi-Harlig, K (2009) J. Acoust. Soc. Am. 125(4):2767(A)
- 7) Miller, JD. (1989) J. Acoust. Soc. Am., 85, 2114-2134.
- 8) Kingston, J. (2003) Language and Speech, 46(2-3), 295-349.