

Miller, J., Watson, C., & Sillings, R. (2011). Internet accessible, computerized training of ESL learners to perceive the sounds and words of spoken English. In S. Huffman & V. Hegelheimer (Eds.), *The role of CALL in hybrid and online language courses*. Ames, IA: Iowa State University.

## **Internet Accessible, Computerized Training of ESL Learners to Perceive the Sounds and Words of Spoken English**

**James D. Miller**

**Charles S. Watson**

**Roy Sillings**

Communication Disorders Technology, Inc

Perception of the sounds and words of English is a significant problem for ESL learners that can be efficiently addressed through the use of CALL. The Speech Perception Assessment and Training System for ESL (**SPATS-ESL**) can be used in hybrid courses or as an extracurricular adjunct to ESL programs. This system adapts to each student's specific perceptual problems and provides systematic intensive practice on their deficient perceptual skills. Since each student's password-protected files are kept on an internet accessible server, the student can use any computer with internet access. Thus SPATS training is not restricted to a particular location, computer, or language laboratory. The system tests and trains the perception of the 109 most important syllable constituents (onsets, nuclei, & codas) of spoken English and the ability to identify words in naturally spoken, everyday sentences presented in multi-talker babble. A curriculum evolved as the system has been used by ESL learners and is constantly being improved to enhance its ease of use, to provide effective feedback, and to maintain student interest. Results indicate that near-native perceptual skills in the identification of the sounds of English and of words in sentences can be achieved on average with about 24 hours of practice. The benefits for classroom ESL instruction and for immersion are discussed.

### **INTRODUCTION**

Numerous studies have demonstrated that the speech-perception system becomes "tuned" to one's native language during the first few years of life (Werker & Tees, 1984, 2005). This tuning includes the cues and boundaries for speech sounds and the cues for the segmentation of the speech stream into words, as well as the recognition of such supra-segmental factors as intonation and rhythm. The perceptual problems of adult learners of a new language seem to be ubiquitous as evidenced by the frequent complaints of literate second language learners that native speakers "talk too fast." The perceptual problems of ESL learners are "hidden problems." While problems with vocabulary, idioms, syntax, and pronunciation are obvious to the second language teacher, serious perceptual problems often are not obvious to either the teacher or the learner. The perceptual problems can interfere with classroom instruction, resulting in failures to clearly perceive the instructor's words or to accurately parse the instructor's sentences. Obviously, perceptual failures can diminish the effectiveness of many forms of second-language instruction including immersion experiences. Perceptual problems can greatly interfere with spontaneous and instructed acquisition of correct pronunciation, as it is difficult to correctly

---

pronounce distinctions that one does not perceive. Fortunately, the perceptual problems of second language learners can be largely overcome by focused, deliberate training and practice.

## **HISTORY OF SPATS-ESL**

The development of SPATS was originally funded by the NIH/NIDCD as a training system for hearing-impaired users of hearing aids and cochlear implants. Papers by Watson et al. (2008) and Miller, Watson, Kistler, Preminger, and Wark (2008) provide the background for this development. Papers by Miller, Watson, Kewley-Port et al. (2008) and Miller, Watson, Kistler, Wightman, & Preminger (2008) describe the details of SPATS for the Hearing Impaired (SPATS-HI) and provide results obtained with hearing-aid and cochlear-implant users. (All of these papers are available for download at [www.comdistec.com/new/HI.html](http://www.comdistec.com/new/HI.html)) A few years into that development, it occurred to the authors that a similar system for ESL learners might be very effective. With the cooperation of the Intensive English Program (IEP) program at Indiana University and later through the cooperation of Queens and Waterloo Universities in Canada, SPATS-ESL has been used to train more than ninety ESL learners and has undergone constant improvement based on the results of those trials.

## **SUMMARY DESCRIPTION OF SPATS-ESL**

SPATS-ESL trains the perception of the elemental sounds and syllable constituents of English and also of words spoken in naturally produced short sentences and presented in a background of multi-talker babble. The program is described in detail in Miller et al. (2010), which also includes descriptions and illustrations of the stimulus sets, response screens, and progress charts provided to the learner.

### **Constituent Training**

The elemental sounds of English are considered to be *syllable constituents* rather than phonemes, because properly pronounced phonemes vary greatly depending on their positions in speech-sound sequences. Statistical analyses of English corpora revealed that 109 such constituents are crucial for the accurate perception of English. These include 45 syllable onsets, 28 vowel and vowel-like syllable nuclei, and 36 codas. Each constituent category, onset, nucleus, and coda is trained separately in SPATS-ESL, generally with each constituent class included in each training session. Training starts with a reduced set, about one-fourth of the total constituents in a category, and when these are mastered, the next fourth is added, and then the next, until the learner is dealing with all 45 onsets, 28 nuclei, and 36 codas.

A unique algorithm, Adaptive Item Selection (AIS), focuses training on those items that are currently of moderate difficulty for the individual learner. Because the program automatically concentrates instruction on those items in need of training for each individual learner it is suitable for all L1's. All training is done with multiple talkers (four men and four women) and multiple phonetic contexts.

## Sentence Training

SPATS-ESL also includes a unique sentence perception training component. The learner hears a naturally spoken, short meaningful sentence (4-7 words) such as “There is a tornado warning” and must select the words from an alphabetical listing that includes the spoken words plus three phonetically similar foils for each word in the sentence. The sentences are presented in background babble at five different signal-to-noise ratios: 10, 5, 0, -5 and -10 db. Ten talkers recorded over 1000 sentences for this purpose.

Scoring is based on both response time and errors. To perform well on this task the ESL learners must be able to segment the sentence stream into words and to quickly identify them using a combination of hearing and contextual inferences. The use of context to disambiguate words not otherwise clearly perceived is a skill constantly used by native speakers and its importance is stressed to students undergoing SPATS training by the demands and structure of the sentence training task.

## Other features of SPATS-ESL: The Quick Test and Proctored Tests

The **Quick Test of Constituent Perception** is a ten-minute test designed to assess the perception of English speech sounds. This test is a very efficient way to demonstrate the need for perceptual training and to document improvement with training. During each trial, the learner hears a single constituent and is required to identify it from a set of four response alternatives consisting of the target sound and the three sounds most commonly confused with that sound.

**Proctored Tests** allow certification of perception of English as *native-like*, *near native* or as a percentage of the average level of performance by native listeners. Certification can be for onsets, nuclei, codas, sentences and for overall ability to recognize spoken English. The provision of separate proctored tests is included to enable instructors to validly assess the achievement of SPATS students who may have done much of their SPATS training on home computers or at other unsupervised times or locations, thus allowing the possibility that scores in the student’s file may represent the performance of someone other than the student.

## Internet Access to SPATS-ESL

All user files are kept on a virtual private server (VPS) on the internet. Using her/his ID and passcode, the student may practice on any SPATS-ESL equipped computer with internet access. In this way the student can utilize any SPATS-ESL equipped computer in a language laboratory or her/his own computer at home or at any other location with internet access.

## SPATS Administrators

Each facility that uses SPATS-ESL needs to appoint a SPATS Administrator. The duties of an Administrator are: 1) to download the program to computers; 2) to assign user IDs and passcodes to the students; 3) to advise students on technical details such as the selection and use of earphones; 4) to advise students about their progress; 5) to administer proctored tests and award associated certification; and 6) to contact the SPATS developers for any needed help or

program support. The SPATS developers, Communication Disorders Technology, Inc. (CDT), are equipped to train and assist administrators through the use of internet conferencing without additional charge.

## **INTEGRATION OF SPATS-ESL INTO AN ESL CLASS OR AN INTENSIVE ENGLISH PROGRAM**

Experience suggests that for a student to be successfully trained to distinguish sounds of spoken English a total of 15 to 35 hours of program use in sessions (30 to 90 minutes each) spaced out over 8 to 16 weeks is required. After orientation to the program, a student may work in a language lab or anywhere she/he has access to a laptop or desktop computer with internet and the appropriate equipment (earphones or loud speakers). Thus, practice with SPATS-ESL can be conducted as part of a language lab or, after orientation, as independent homework. SPATS-ESL training can be assigned as a prerequisite for or as an integral part of a pronunciation class. It can also be used as a prerequisite for an immersion experience or as a requirement for certification as an Associate Instructor at the university level.

### **Benefits of SPATS-ESL Training**

Successful completion of SPATS-ESL training enables the student to accurately identify the basic sounds of English and to be able to identify words in naturally produced sentences presented in moderate amounts of background multi-talker babble. Success at these tasks requires that the learner acquire the dimensions and boundaries of the phonetic categories of English while also learning to segment spoken sentences into words, even when the student may not know or be familiar with some of the words in the sentence. Successful users of SPATS-ESL report that they have improved ability to follow class room instruction of all kinds in English, improved understanding of native speakers of English, and improved understanding of English in difficult situations, such as in background noise or over cell phones. Successful users also develop the ability to isolate and repeat unfamiliar, new words and in this way improve their vocabulary. Learning the phonetic categories of English enables them to better monitor and correct their own speech and should increase the benefits of pronunciation instruction. Since there are large numbers of non-native speakers of English who provide instruction in English, those instructors would find their teaching greatly enhanced if they themselves had completed SPATS training. Teaching the perception of sounds that you cannot identify yourself is a very difficult challenge.

### **EXPERIENCE WITH SPATS-ESL**

The authors' experience has primarily been with users who have achieved basic literacy in English. Typically, TOEFL (pbt) scores have ranged from a little below to well above 500. The L1s of the users have included Arabic, Cantonese, Japanese, Kazakh, Korean, Mandarin, Mongolian, Portuguese, Spanish, Taiwanese, Turkish, and Vietnamese. All of these users demonstrated numerous problems identifying the sounds of English in spite of their basic English literacy. The exact difficulties that each exhibited appeared to depend not only on their L1's, but also on individual idiosyncrasies in perception.

## Current Procedures

SPATS-ESL introduces new students to the sentence task followed by sentence tests (15 sentences each). This is followed by training rotation sequences that begin with the identification of syllable onsets, followed by sentence identification (6 sentences), identification of syllable nuclei, sentence training (6 sentences), the identification of syllable codas, and concluding with sentence identification (3 sentences). Thus a single rotation includes testing and training of the three constituent classes and of sentences. Typically a single rotation is completed in three-fourths to one and one-half hours. With regard to the syllable constituents, whenever a performance criterion is met for a constituent type, the number of items is increased for that type. For each type the learner begins with one-fourth of the total number (Level 1), then graduates to one-half of the items (Level 2), then to three-fourths of the items (Level 3), and finally to all of the items (Level 4). In general, it has been found that this schedule results in roughly 20% of the training time spent on onsets, 20% on codas, 20% on sentences, and 40% on nuclei. So, for example, a learner who spends 25 hours in SPATS training would typically have spent 5 hours on onsets, 5 hours on codas, 5 hours on sentences, and 10 hours on nuclei. The details of exactly how this progression works are given in Miller et al. (2010).

## Progress Charts

Examples of how progress is monitored and charted are provided in the following section. Constituent training and testing scores are adjusted to reflect the number of items correctly perceived in relation to the total number of items to be eventually mastered. Therefore raw scores at Level 1 are multiplied by 0.25, those at Level 2 by 0.50, those at Level 3 by 0.75, and those at Level 4 by 1.00. In this way a score reflects both the Training Level reached and the accuracy of performance. These progress graphs are well fitted by exponential learning curves. An example is shown in Figure 1 on the following page. (The black bars show the results for tests. Training sessions between the bars are not shown.)

Similar graphs can be constructed for the perception of syllable onsets and codas and for the identification of words in sentences. From such graphs, the times to near-native performance for the participants at Indiana University have been estimated. Native speakers achieve scores at or above 94% correct for syllable onsets, at or above 94% correct for syllable nuclei, at or above 88% correct for codas, and at or above 90% on the sentence task. “Near native performance” has for training purposes been defined as 90% correct for onsets and nuclei, 85% correct for codas, and 90% correct on the sentence task.

In Table 1 and Figure 2 below, the estimated times to near-native performance are shown for those students who produced sufficient data to allow direct observation of the time to near-native performance or estimation by reasonable extrapolation of fitted exponential progress graphs. More details of this procedure can be found in Miller et al. (2009).

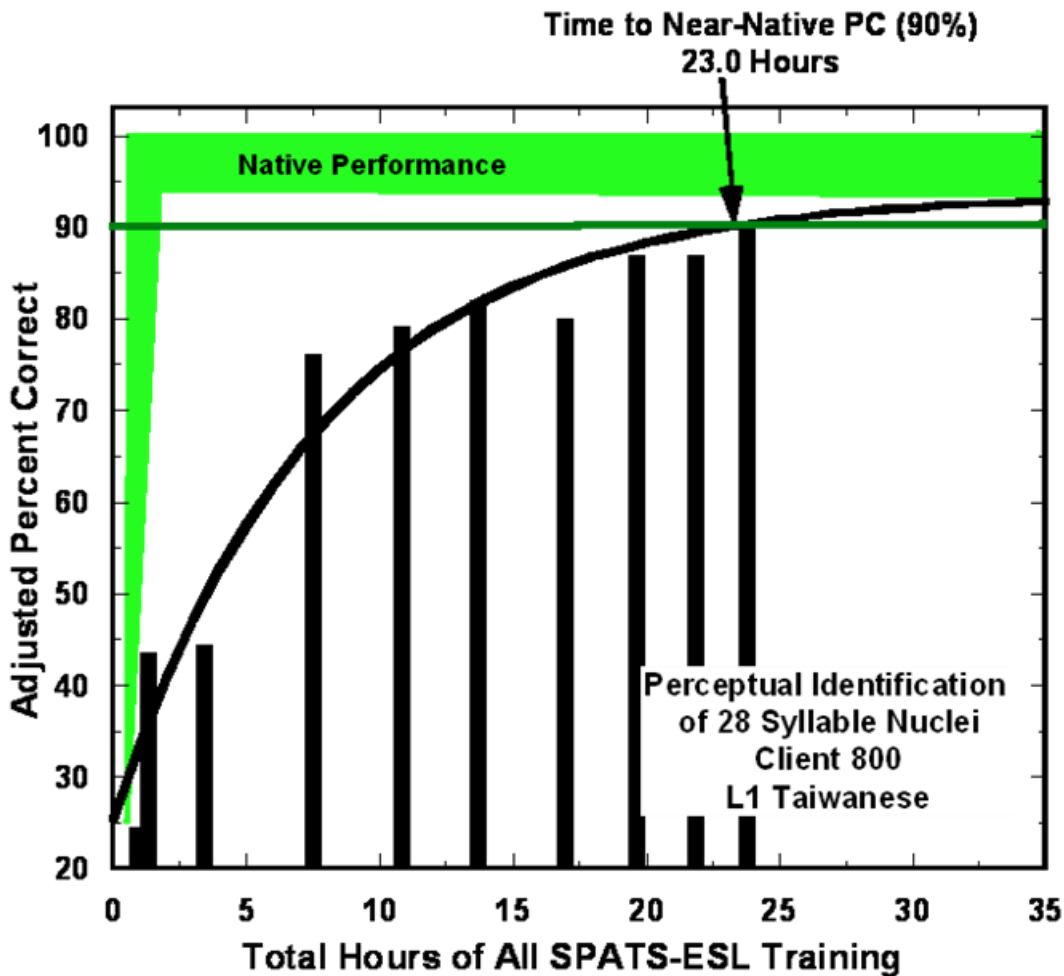


Figure 1. Progress chart and fitted curve showing a student's progress in perception of the vowels and vowel-like sounds of English.

The total training time required to approach near-native performance on all four tasks ranged from 12 to 40 hours with a mean of 24.4. While much more data is needed to confirm these estimates, these values are consistent with our observations to date from 90+ trained students. These results are summarized in Table 1.

In interpreting Table 1 and Figure 2 it is important to remember that the actual times spent on each of four tasks are about 20% of the totals shown for onsets, codas, and sentences and about 40% of the total shown for nuclei.

Table 1. Estimated total SPATS training times (Hours) to near-native performance

Tasks	Onsets	Codas	Sentences	Nuclei	All
Mean(Hrs)	9.3	11.4	15	18.4	24.4
Median(Hrs)	6.2	8.2	15.1	17.0	23.0
N	19	18	10	19	10

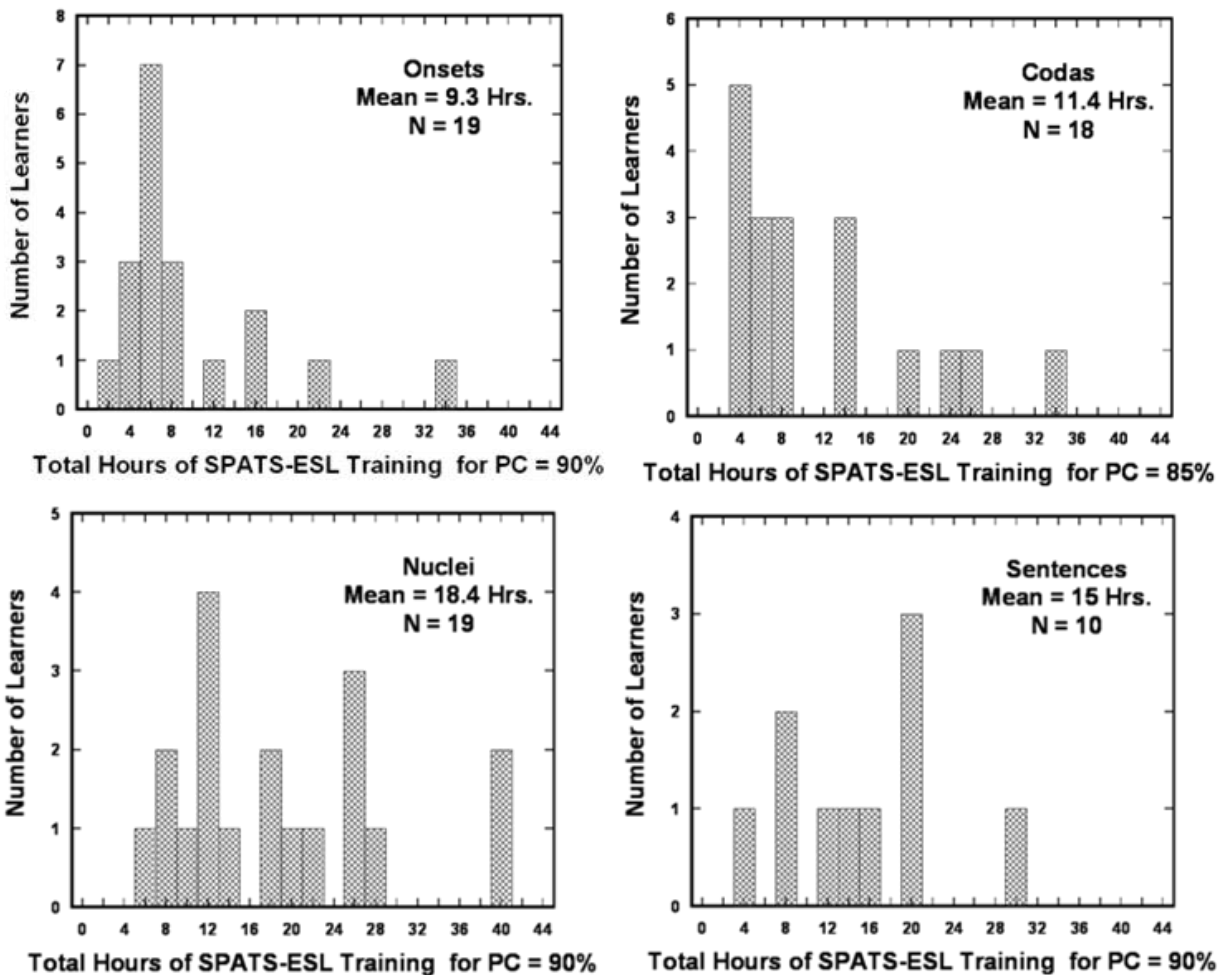


Figure 2. Histograms showing the total SPATS-ESL training times to approach near-native performance on the SPATS task identified in each panel.

### Onset confusions made by 15 Asian ESL students at Level 4 near the end of SPATS training

SPATS-ESL routinely collects data regarding the confusions made by ESL learners. These can be tracked for individual learners as well as groups. Table 2, below, is an example of one such analysis. This table lists the significant confusions (greater than 5%) remaining near the end of SPATS-ESL training for a group of 15 Asian ESL students. Notice that while some *r/l* confusions remain, especially in onset clusters, this confusion has been largely resolved. The confusions of syllables that begin with a vowel (Vw1) with those beginning with *w* is interesting and not observed in the case of native speakers of English. It is striking, however, that, for the most part, the remaining confusions shown at the bottom of the right-hand column are similar to those made by native speakers such as confusions of the voiceless *th* sound with the *f* sound and of the voiced *th* [*th*(v)] with *v*, *z*, and its voiceless counterpart.

Table 2. Onset confusions remaining near the end of SPATS training

% Correct	Onset	Confusions	% Correct	Onset	Confusions
1.000	sm-		0.888	ch-	tr-
0.988	g-		0.888	tr-	kr-
0.988	m-		0.875	bl-	l-
0.988	sp-		0.875	gl-	gr-
0.975	sk-		0.875	r-	l-
0.975	st-		0.875	w-	vwl-, v-
0.963	d-		0.863	fr-	fl-
0.963	str-		0.863	k-	kw-
0.963	t-		0.850	kr-	kw-
0.950	dr-		0.838	fl-	fr-
0.950	sh-	s-	0.825	pl-	kl-, pr-
0.950	sw-		0.825	z-	th(v)-
0.938	h-	p-	0.813	b-	m-
0.938	j-		0.813	kl-	kr-
0.925	n-		0.813	sl-	fl-
0.925	p-	h-	0.788	Vwl-	w-
0.925	sn-	sl-	0.775	br-	bl-
0.913	gr-		0.713	v-	th(v)-
0.913	y-	j-	0.700	kw-	kr-, kl-
0.900	f-	th-	0.663	l-	bl-
0.900	pr-		0.613	th-	f-
0.900	s-	th-	0.600	th(v)-	v-, z-, th-
0.900	skr-	str-	Mean % Correct = 0.874		

### Subjective Evaluations by SPATS-ESL Users

Eighteen of the participants in SPATS-ESL trials completed questionnaires regarding their subjective impressions of the value of the program. The questions and the percent answering “yes” are listed below.

*Do you think that other ESL programs should use SPATS?* 90% yes

*Would you recommend SPATS to friends or relatives learning English?* 90% yes

*Did sentence work improve your participation in every day conversation?* 85% yes

*Did drill on speech sounds improve your participation in everyday conversation?*  
80% yes

*Will SPATS training improve your pronunciation of English?* 90% yes



## **SUMMARY**

A software training program has been developed that provides systematic, individualized training designed to improve ESL learners' abilities to perceive spoken English. Specifically, this program trains the ability to identify the onsets, nuclei, and codas of English syllables and to identify words in naturally spoken sentences presented in a background of multi-talker babble. Through the use of computer and internet technology SPATS-ESL can be conveniently used as part of IEP programs or ESL classes and can be accessed by the students on any computer in a language laboratory or on his or her own home or work computer.

## REFERENCES

- Miller, J.D., Watson, C.S., Kistler, D.J., Preminger, J.E., & Wark, D.J. (2008). Training listeners to identify the sounds of speech: II. Using SPATS software. *The Hearing Journal* 61(10), 29-33. Retrieved from [www.comdistec.com/new/HI.html](http://www.comdistec.com/new/HI.html).
- Miller, J.D., Watson, C.S., Kewley-Port, D., Sillings, R., Mills, WF, & Burleson, D.F. (2008). SPATS: Speech Perception Training and Assessment System. *Proceedings of Meetings on Acoustics* 2. Retrieved from [www.comdistec.com/new/HI.html](http://www.comdistec.com/new/HI.html).
- Miller, J.D., Watson, C.S., Kistler, D.J., Wightman, F.L., & Preminger, J.E. (2008). Preliminary evaluation of the Speech Perception Assessment and Training System (SPATS) with hearing-aid and cochlear-implant users. *Proceedings of Meetings on Acoustics* 2. Retrieved from [www.comdistec.com/new/HI.html](http://www.comdistec.com/new/HI.html).
- Miller, J.D., Sillings, R, Watson, C.S., & Kewley-Port, D. (2009). Speech perception Assessment and Training System (SPATS-ESL) for speakers of other languages learning English” *J. Acoustics Society of America* 125(4): 2755.
- Miller, J.D., Sillings, R., Watson, C.S., Darcy, I., & Bardovi-Harlig, K. (2009). Experience with computerized speech-perception training (SPATS-ESL) for speakers of other languages learning English. *J. Acoustics Society of America* 125(4):2767.
- Miller, J.D., Sillings, R., Watson, C.S., Darcy, I., & Bardovi-Harlig, K. (2009). Experience with computerized speech-perception training (SPATS-ESL) for speakers of other languages learning English. Power point retrieved from [www.comdistec.com/new/ESL.html](http://www.comdistec.com/new/ESL.html).
- Miller, J.D., Sillings, R., Watson, C.S., & Kewley-Port, D. (2010). *Speech Perception Assessment and Training System (SPATS-ESL) for speakers of other languages learning English*. Retrieved from [www.comdistec.com/new/ESL.html](http://www.comdistec.com/new/ESL.html).
- Watson, C.S., Miller, J.D., Kewley-Port, D., Humes, L.E., & Wightman, F.L. (2008). Training listeners to identify the sounds of speech: I. A review of past studies. *The Hearing Journal* 61(9), 26-31. Retrieved from [www.comdistec.com/new/HI.html](http://www.comdistec.com/new/HI.html).
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7, 49–63
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46(3), 233-251.

## ACKNOWLEDGEMENTS

The work was supported in part by NIH/NIDCD SBIR Grant R44DC006338. Authors James D. Miller and Charles S. Watson are stockholders in Communication Disorders Technology, Inc.