Learn to Listen (L2L): Perception training system for learners of English as a second language

Diane Kewley-Port[*], Kanae Nishi[+], Hanyong Park[*], James D. Miller[*] and Charles S. Watson[*]

Communication Disorders Technology, Inc., Bloomington, IN and [*]*Indiana University* and [+] *Boys Town National research Hospital*

Introduction

Computer software, Learn to Listen (L2L), is being developed for comprehensive perception training of American English (AE) by second language learners. The first goal of the study is to extend the training of large sets of phonemes shown to be successful in Nishi & Kewley-Port (2007) for vowels to consonants. Vowel set here had 11 monophthongal vowels. Consonant onsets were selected as somewhat easier than codas to investigate two large sets of confusable consonants in our novel approach. Two different consonant sets were used, one with single consonants (N=16), and the other with stop-liquid clusters (N=8). The second goal is to demonstrate generalization of improvement of phoneme perception to the perception of running speech in sentences. Two languages with very different sound structures were investigated, Korean and Spanish.

Research Questions

Exp. 1) What sets of AE phonemes are perceptually confused by Korean and Spanish listeners?

Exp. 2) Can the success of large set perceptual training for vowels be extended to large set of training for consonants?

Exp. 2 & 3) Can syllable training be extended to confusable words in a sentence frame?

Exp. 1 Assessment: Methods & Results

- Assessment software, SPATS (Miller et al., 2008), was used to determine confusable subsets of vowels (N=28 including /Vr/ & /Vl/) and syllable onset consonants (N=44)
- Listeners: Korean (N=6); Spanish (N=5) [Same criteria as Trainees]
- Examined full confusion matrices. Confusable sets of consonants identified separately for Korean (Kr) and Spanish (Sp) listeners were found to have similar error patterns.

Error Rate in Vowel Nuclei

Category (N)	Examples	Kr	Sp
Monophthongs (11)	/i I eI E ae u U/	0.49	0.61
True Diphthongs (3)	/aI aU oI/	0.08	0.11
Rhoticized V (5)	haired, heard	0.28	0.28
Others (9)	hewed, hilled	0.42	0.53
Overall		0.39	0.49

Error rate in Consonant Onsets

Category (N)	Examples	Kr	Sp
Fricatives (8)	/f v θ ð s z ʃ h/	0.26	0.34
Affricates (2)	/tʃ ʤ/	0.34	0.35
r/l clusters (4)	/kl gl bɹ pɹ/	0.20	0.20
Stops (6)	/p b t d k g/	0.04	0.04
Other (24)	/w m fl kɹ kw sp sm/	0.10	0.14
Overall		0.14	0.18

Training Methods

Trainees: Criteria for participation: (1) less than 2 years in USA, (2) between 18 and 40 years old and (3) pass beginners reading test, hearing screening & PB production test. Recruited 5 Korean (Kr) and 5 Spanish (Sp) volunteers.

Stimuli: Choice of stimulus sets based on outcome of Exp. 1. Selected 3 large stimulus sets presented in consonantal context for bottom-up training.

Vowels (CVC): N=11 both Kr & Sp (see vowels words below)

Consonant onsets (CV): N = 16 Kr /p b t d k g f v θ ð s z tʃ ʃ ʤ h/ (Kr-16)

(CCV): N = 8 Sp /br pr gl kl b p g k/ (Sp-8)

Software: Testing using SPATS. For vowels used only 11 vowels in /hVd/ words in pre- & post-tests. For consonants, all 44 onsets used.

Software: Training. New vowel interface for L2L project used words in CVC context, 7 real words per vowel spoken by 12 AE talkers. Four examples words for each of the 11 vowels are:

i	I	e	ɛ	æ	a	ʌ	aw	o	U	u
heed	hid	hayed	head	had	hod	hud	hawed	hoed	hood	who'd
beef	bid	bake	bed	back	cop	bud	bought	boat	book	boot
feet	dig	fate	peck	gap	pod	cut	talk	goat	foot	loop
keep	dip	gate	set	pack	pop	duck	cough	soap	good	rude

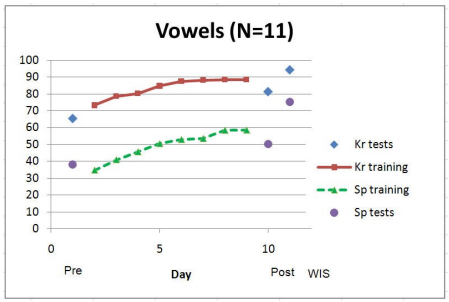
Syllable training for consonant onsets, used SPATS software. Each onset presented a CV, with four vowels, /i, a, u, er/ and 8 AE talkers.

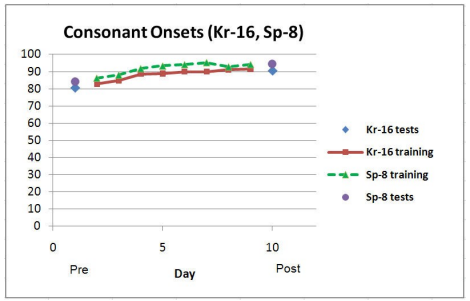
Word-in-Sentence (WIS) Task (Used in Exp. 2 & 3): To develop listening skills in fluent sentences, the WIS task was designed to require primarily bottom-up processing. There was one sentence frame with 3 words from the vowel task as follows: **"The first word is XX, the second is YY, and the third is ZZ".** Example: "The first word is **bake**, the second is **hid**, and the third is **feet**". A 4-item forced choice display presented the word and 3 foils after the sentence was heard (see below). The listeners then indicated the 3 words heard, and feedback was given. 99 sentences spoken by two talkers were used in WIS.

XX	YY	ZZ
bake	hood	fate
back	heed	fit
beak	head	feet
beck	hid	fight

Procedures Exp. 2: 11 days/listener [Pre-tests, 8 Training days, Post-tests, WIS], 1 hour each. Sounds were presented over Sennheiser headphones on PC or laptop in quiet room. Identification task with responses using keywords shown on computer display.

Exp. 2: Results & Discussion





Main effects tested by Anova's, language always between factor Kr versus Sp.

1) **PRE/POST TESTS**: 3-way repeated measures anova, language X segment (vowel vs con) X test (pre– vs. post-). All 3 factors were significant, with no interaction.

- Koreans % correct >> Spanish, particularly for vowels.
- Consonant % correct >> vowel (by 29%).
  - While AE vowels are generally thought to be more difficult for second language learners, the large difference observed here was not expected given our Exp. 1 assessment data.
- Post-test identification improved by about 11%, similar for V & C, Kr & Sp.

2) **TRAINING:** 2-way repeated measures anova, language X day, done separately for vowel and consonant training).

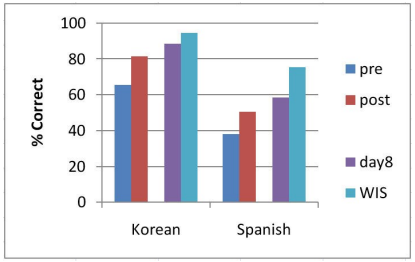
- Vowel training significantly improved identification, with 19% improvement from Day 1 to Day 8. Koreans were 34% better overall at identifying vowels across all days than Spanish, although the rate improvement similar.
  - Note that Spanish and Korean listeners in Exp. 1 assessment appeared to be at similar levels of AE proficiency, that is trainee selection criteria were the same and the errors for most vowel and consonant categories were similar. The exception was for monophthongs, and in Exp. 2 a large (34%) difference in absolute identification throughout vowel training was seen.
  - Apparently the structure of the 5-vowel space for Sp interfered with perceiving the structure of AE monophthongs more than the 9-vowel Kr space, even though learning rate was the same for both groups.

- Consonant training significantly improved identification, with 9% improvement from Day 1 to Day 8. No difference between Korean and Spanish in onset training, achieving > 90% correct).
  - Examined a subset of the more difficult sounds (i.e. no stops) at pretest. Kr subset improved more with training than did the Sp (18% vs 7%). Therefore the larger set training for Kr-16 did not interfere with training compared to the smaller set, Sp-8. Thus Nishi & Kewley-Port (2007) vowel training was successfully extended to consonants.

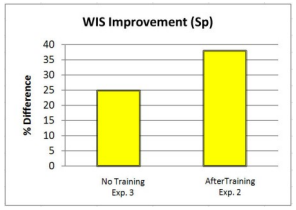
Word-in-Sentence Task (WIS)

The Word-in-Sentence (WIS) task was designed as the first step towards understanding words in running speech. Only words from the vowel task were selected and placed in a context-free sentence frame such that the listening task required primarily bottom-up skills. In Exp. 2 WIS was run for 1 hour on last day (**after training**). In Exp. 3, a pilot control study, Sp listeners only participated in the pre-test followed by WIS (**no training**).

Exp. 2. Identification of words in sentences was very good. Compared to the same words alone (Day 8), or /hVd/ words in pre- & post– tests, the WIS performance was best. Note for Sp, WIS improvement over words on Day 8 was a surprising 17%.



Exp. 3. This pilot experiment demonstrated that even with no training the WIS task was easier than identifying /hVd/ words alone. However, training in Exp. 2 showed additional improvement of 13%. Apparently WIS is an easy task using running speech that can be introduced early in training.



Conclusions

1) Learners of American English from two languages with very different sound systems, Korean and Spanish, were assessed for confusions in vowels and consonant onsets. They had similar error patterns overall in identifying AE vowels and consonants, although Spanish were much poorer at identifying the AE monophthongs.

2) Large set syllable training was shown to improve listening significantly for a set of 11 vowels (Kr & Sp), a set of 16 stops and fricatives (Kr), and a set of 8 stop-liquid clusters (Sp).
   - This suggests new and perhaps more efficient ways to train speech perception

3) A novel sentence task designed for bottom-up training of words in sentences was shown to be considerably easier than expected, much easier than analogous single-word identification tasks.
   - This suggests a path for developing a series of training exercises for systematically increasing the amount of linguistic context and the difficulty of the task.

References

Nishi, K. and Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *J. Speech-Language-Hearing Res.*, **50**, 1496-1509.

Miller, JD, Watson, CS, Kewley-Port, D, Sillings, R, Mills, WF, and Burleson, DF. (2008) "SPATS: Speech Perception Training and Assessment System" *Proceedings of Meetings on Acoustics 2*. http://scitation.aip.org/POMA, 1-17.

Acknowledgement: Thanks to our excellent research assistants, Tanya Flores and Sherne-Marie McMillan. **Work supported by:** NICHD R43-55071