HOW SUCCESSFULLY DOES VISUAL FEEDBACK TRAIN LISTENERS TO PRODUCE AND PERCEIVE NON-NATIVE PHONOLOGICAL CONTRASTS?

M. Ziolkowski and K. Landahl
University of Chicago, Chicago, IL, USA

ABSTRACT

This study evaluates the efficacy of three techniques for teaching students of a second language to phonologize natively allophonic variants. Each technique—"listen-repeat-compare" with a tape recorder, tutoring with a native speaker, and computer-driven visual feedback—was used to train ten English speakers in producing phonemic vowel and consonant length and pitch accent in Japanese. Though results were mixed, visual feedback proved its utility.

THE PROBLEM

A native speaker and instructor of the Japanese language issued a plea for help in her effort to assist her second-year students, native English speakers, overcome deficiencies in producing and perceiving properties of Japanese phonology with even rudimentary proficiency. Particularly salient were her students' difficulties with Japanese vowel and consonant quantity, and pitch accent.

All three of these phonological properties are phonemic in Japanese—*ojiisan* (long [i]) means 'grandfather', *ojisan* (short [i]) 'uncle'; *aka* (short consonant) denotes 'dirt', *akka* (long consonant) 'dirty money'; *kaME* (low-high pitch pattern) refers to a 'large-mouthed jar', *KAme* (high-low) to 'tortoise'. So, respecting these properties is crucial not just for achieving more native-like Japanese pronunciation and listening skills, but is a necessity for understanding Japanese and being understood correctly when speaking it.

What is interesting here from a linguistic point of view is that none of these Japanese phonological properties is wholly absent from the English learners' native system. Instead, each serves a more or less active role in English allophony. Vowel length has been observed to vary systematically with the voicing of a following consonant; consonants geminate when like segments meet at morpheme or word boundaries. Pitch may figure in the realization of English stress where it is one of the constellation of relevant features, but it bears the functional load solo in Japanese pitch accent.

The challenge for native English speakers learning Japanese appears to be one of changing the functionality of these properties. The challenge to the instructor is increasing the salience of these properties, hoping they may be put towards a phonemic end. We responded to the challenge with visual feedback.

THE SOLUTION?

We decided that a computer-based visual feedback system provided the appropriate means for increasing the salience of the acoustic cues to these properties, and selected the Visi-Pitch Model 6095/6097, manufactured by Kay Elemetrics of Lincoln Park, NJ. It affords real-time display of fundamental frequency and intensity contours for input utterances, and allows these data as well as their graphical representation to be stored on a personal computer for later analysis and display. Stored model utterances can then be redisplayed, and contours for student renditions overlaid on native models in real-time.

Figure 1 shows the fundamental frequency contours (f_0 traces) associated with a pair of words, uttered by a male native speaker, which contrasts all three Japanese properties. The upper half of the display traces the f_0 contour for *GAchcho* 'joint authorship', and the lower half, the trace for *gaCHOO* 'goose'.

Each of the phonological properties is iconically displayed in the figure, where time runs horizontally and f_0 vertically

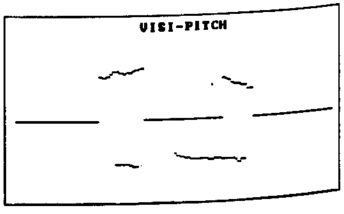

Figure 1. GAchcho *and* gaCHOO.

with greater displacements of the trace from the baseline associated with higher f_0. In both halves of the screen there is a break in the f_0 trace for the voiceless, medial consonant "ch", and the longer drop to the baseline in the upper half correlates with the longer "ch" in *GAchcho* compared to the short affricate in *gaCHOO*. Consonant quantity, then, is cued by the extent of baseline drops. The length contrast for [o] at the end of these two words is depicted by the lengthier last "chunk" of f_0 trace in the lower contour, and pitch accent patterns "high-low" versus "low-high" by the relatively higher vertical displacement of the first and second "chunks" of f_0 in the upper and lower halves of the screen, respectively.

In Figure 2, the Visi-Pitch's potential for training phonological properties one at a time becomes apparent. A female native speaker's f_0 traces for *itai* 'dead body' and *iitai* 'one body', a minimal pair distinguished by length of the word-initial vowels, appear in the two halves of the display. When the two contours are more similar in overall appearance, as they are here, the acoustic correlate of the feature differentiating the two words, vowel quantity, is more visibly transparent. The first "chunk" of the f_0 trace is clearly shorter in the upper half of the screen than in the lower—the first vowel in *itai* is shorter than the first vowel in *iitai*.

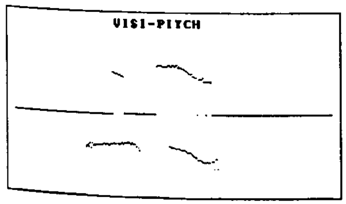

Figure 2. itai *and* iitai.

Thus, visual feedback exercises can be constructed offering learners straightforward, graphical cues to the phonological properties which have proven difficult for them to master. Visual feedback presents such properties much more saliently than do traditional strategies for teaching these distinctions. The only feedback offered, for example, by a Level III tape recorder may be compromised by students' probable deficiency in perceiving differences in the first place. Individual instruction with a native speaker of the language may be preferable, but the superior feedback associated with the greater cost typically is limited to a ready source of native examples and to the tutor's "affective" support for successful productions. Neither tape nor tutor affords salient presentation of the properties which cue non-native contrasts as well as visual feedback does.

But is visual feedback more effective than these other techniques? Does extracting relevant acoustic properties hone learners' abilities to perceive and produce non-native phonological properties? Common sense seems to dictate visual feedback's superiority to a tape recorder certainly, and to a tutor possibly, and pedagogues [1] have encouraged its use for fine-tuning phonetic distinctions in second language learning under the assumption that it works. Previous research (see [2]) has found increases in the perceived nativeness of utterances produced after visual feedback training on sentential intonation. But, to our knowledge, no systematic study of the extent of visual feedback's efficacy in fine-tuning phonetic capabilities or phonologizing the allophonic has ever been undertaken. Nor has visual feedback been rigorously compared to instruction with a tutor. The following study begins to fill these gaps.

THE STUDY

We employed the three teaching strategies discussed above—tape, tutor, visual feedback—in production-directed training on the three Japanese phonological properties mentioned there—vowel and consonant length, and pitch accent. Our goal was to quantify each strategy's effect on learners' abilities to produce these properties directly, through acoustic analysis of their productions. Also, we wondered whether production-directed training enhanced subjects' perception, and examined subjects' perceptual acuity with the distinctions before and after treatment.

Subjects

Thirty University of Chicago students participated in the study. Half were completing their first year of Japanese language study, and thus were familiar with the three phonological features to be trained; half were Japanese-naive. No subject had studied any other language that phonemically exploits segmental quantity, pitch accent or tone. The experiment demanded about four hours over three weeks and subjects were compensated for their participation.

Stimuli

Minimal pair lists were constructed illustrative of each of the three phonological properties of interest. 9 pairs represented the vowel quantity contrast, 11 pairs constituted the list for consonant length, and 8 pairs appeared on the pitch accent list. Words were chosen to facilitate identifying segment boundaries on the Visi-Pitch in later acoustic analyses.

These 56 words were presented to 10 native speakers of Japanese for elicitation. The utterances of 2 speakers—one male, one female—were selected as models on the criterion that they were judged to be the best examples of the Tokyo dialect by the native speaker/second-year instructor.

Design

A third of the learners (5 Japanese-naive, 5 experienced) were randomly assigned to one of three treatment groups, each of which invoked a different training method (tape, tutor, visual feedback). Training lasted for one hour; twenty minutes spent on each of the three phonological properties of Japanese. After pretesting, subjects in the visual feedback condition received a twenty minute "crash course" in operation of the Visi-Pitch, which served as the visual feedback tool, before using it in training.

Subjects completed a battery of speaking and listening tests 1) immediately before training to measure baseline performance, 2) immediately after, 3) one week after, and 4) three weeks following instruction.

Within each testing session, subjects completed three different tasks, always administered in the following order. A "read-aloud" production test was followed by a "listen-and-repeat" mimicry task, and the session concluded with a forced-choice perception test. In the reading task, subjects were given a randomized list of stimulus words and asked to pronounce them "cold" to prompts by an experimenter at a rate determined by a light-flashing metronome. In the mimicry task, subjects were asked to repeat after a tape consisting of the male and female model speaker utterances presented in a randomized order. In the perception test subjects were played a tape with another randomization of the stimulus set and asked to circle the member of the minimal pair they thought each token represented. For example, *itai* appeared next to *iitai* on the response sheet, and subjects had to circle which word they thought the native speaker was saying.

A subset of the full stimulus set was used in the training and pre-treatment testing sessions. The stimulus sets expanded from 28 words (2 x 5 vowel length pairs, 2 x 5 consonant dyads, 2 x 4 pitch accent pairs) to 56 words from a subject's first session to last. We will determine, in later analysis, if training on a phonological property generalizes to items not explicitly trained upon.

Analysis

All subject data in production tests were acoustically analyzed for vowel and consonant length, and pitch accent as relevant. (Due to measurement difficulties we excluded data from three subjects: one Japanese tape, one naive tape, one naive tutor.) For quantity, the absolute segment durations were determined and used to compute the dependent measures for quantity described below. For pitch accent, analysis was restricted to that subset of stimuli for which the Visi-Pitch allowed word segmentation into syllables. In these cases the average f_0 for the two syllables "bearing" the pitch accent (high-low or low-high) was measured. Since each word must be self-normalizing (native speakers can determine the accent pattern of words spoken in isolation), the relevant value for the dependent measure for pitch accent was the difference in average f_0 between the two syllables, divided by whichever average was smaller. (A two-way ANOVA on the model speakers utterances showed a main effect of accent-type but not speaker identity for this metric.)

The dependent measure of subject performance in the reading and mimicry production tasks was different for the quantity and pitch accent tokens. For each quantity distinction, a regression equation was fit for each subject in each test session for reading and for mimicry. The equation consisted of μ, the average speaking rate, plus α, quantifying how much longer long segments are and how much shorter short segments are than μ,

plus an error term, ε. α was the dependent measure for both vowel and consonant length performance. For pitch accent the dependent measure was the difference between the average values of the metric described above for the two accent types.

For the perception tests the percentage correct was the dependent measure.

Results

Repeated-measures analyses of variance were run for each of the production and perception measures with language experience (Japanese or naive) and training type (tape, tutor, visual feedback) as grouping factors.

For analysis of subjects' performance on the three phonological properties, exercise type (reading, mimicry) was an additional grouping factor in the ANOVA. Session (pre-, post-, week 1, week 3) was the repeated measure.

For vowel duration three significant values obtained: a main effect for session (Pr>F=0.0001), a main effect for exercise type, (Pr>F=0.0376), and an interaction between exercise type, language experience and training type (Pr>F= 0.0791). For consonants there was a main effect for session (Pr>F=0.0001), an interaction of exercise and session (Pr>F=0.0351), and a marginal interaction of exercise, session and treatment type (Pr>F=0.0228). For pitch accent main effects of exercise type and session were observed (Pr>F=0.0001 for both).

For the perception data, baseline performance was a covariate and the last three sessions were the repeated measures for a two-way ANOVA with language experience and training type as grouping factors. The analysis revealed no statistically significant results, although a main effect of training type did approach significance (Pr>F=0.088).

Discussion

The results for production of the contrasts have been discussed in greater detail elsewhere [3], but those of relevance for evaluation of the efficacy of visual feedback in comparison to tape and tutor training derive from the two interactions in which training type is implicated. Although robust effects for training type were not obtained, it is not surprising they were not, given the small number of subjects in each of the groups and the extent of individual performance variation we have observed in the data.

Our interpretation of the interactions obtained for vowel and consonant quantity performance is the following. For all subjects, mimicry generally improves to native-like following training, but those trained using the tape recorder tend to exaggerate the length distinctions when reading. Visual feedback subjects, then, more closely resemble the tutored subjects by showing more or less native-like durations when reading without a native example.

The perception results are intriguing. Remember that the main effect for training type approached statistical significance for the perception data as a whole. A robust finding for perceptual acuity is that naive visual feedback subjects and Japanese-experienced tutored subjects perform best, followed by naive tape and Japanese visual feedback subjects with Japanese tape and naive tutored learners performing worst. This generalization holds for perceptual ability as a whole, as well as for perception of each phonological feature individually. We can think of no principled reason why this strict ordinal pair-wise ranking should have obtained. We suspect the best explanation to be "luck of the draw"—the naive visual feedback condition happened to be populated by subjects with exceptional perceptual skills.

ACKNOWLEDGEMENT

Thanks go to Mayumi Usami, Brenda Tunnock, Sema Yetim, the Consortium for Language Teaching and Learning, and the University of Chicago Language Laboratories and Archives.

REFERENCES

[1] Chun, D. (1989), "Teaching tone and intonation with microcomputers", *CALICO Journal*, vol. 7, pp. 21-46.

[2] de Bot, K. (1983), "Visual feedback of intonation I: Effectiveness and induced practice behavior", *Language and Speech*, vol. 26, pp. 331-350.

[3] Landahl, K., M. Ziolkowski, M. Usami, and B. Tunnock (1992), "Interactive articulation: Improving accent through visual feedback", *Proceedings of the Second International Conference on Foreign Language Education and Technology*, pp. 283-292.