INTERRELATION OF PERCEPTION AND PRODUCTION IN INITIAL LEARNING OF SECOND-LANGUAGE LEXICAL TONE

Jonathan Leather

University of Amsterdam

ABSTRACT
One group of native Dutch-speaking subjects were trained to perceive the lexical tones of standard Chinese (Putonghua) and were then tested on their ability to produce them, while a second group were trained (using a visual display, and without auditory exemplification) to produce the tones, and subsequently tested on their ability to perceive them. From longitudinal records kept of subjects' perceptual decisions and of acoustic parameters of their productions, the interrelation of learners' evolving perceptual and productive abilities was examined - and found for both groups to be significantly correlated.

1. INTRODUCTION

This paper reports on experiments designed to explore the interrelation between perceptual and productive abilities in the initial learning of a new sound (sub-) system. The phonetic proficiency goal was the citation-form tone system of Putonghua - i.e. standard Chinese or "Mandarin". These tones are phonetically realized as time-varying patterns of voice fundamental frequency: level, rising, dipping and falling contours respectively (see [1] for a review).

2. METHOD

The subjects were two groups of native Dutch speakers aged between 19 and 28, and with no knowledge of any tone language. One group underwent computer-managed tone perception training involving: (i) presentation of (digitized) tokens of the four tonally-distinguished words /ȳ/, /ý/, /y̌/ and /ỳ/ ('mud', 'fish', 'rain' and 'jade') produced by one Beijing native; (ii) presentation of tone tokens of first one female, then two male and two female speakers, for tone labelling. All labelling responses (collected through keyboard input) were logged under program control. Information feedback was provided in L1 in the form "Yes, correct." or "No, it was X" (X being the intended word). Trials continued until a proficiency criterion was satisfied of at least 80% correct identification of randomly-ordered tone tokens from all four speakers, at which time production ability was tested. In the production test the subject was asked to say the Putonghua word (again one of the four minimal quadruplets) corresponding to the L1 gloss displayed on the screen. The tones thus elicited were in a fixed random order (24 trials in all). In both this production test, and the production training of the second group of learners described below, larynx period data were recorded with a Laryngograph and phonetically assessed using a speaker normalization procedure and assessment algorithm described elsewhere [2, 1].

To enable the second group of learners to acquire tone production ability without prior experience in tone perception, it was necessary to use alternative means of exemplifying for them the F0 contour shapes of the tones, and to provide them with external information feedback on the phonetic consequences of their production attempts. The training system designed for this purpose collected larynx period data and provided on the screen visual and verbal feedback on the learner's F0 contours (the signal processing and task

control routines used are detailed in [1]). A real-time plot of smoothed, speaker-normalized learner F0 was displayed beneath, or (if the learner preferred) superimposed upon, a similar plot of the exemplar F0 contour. The learner could thus make a visual appraisal of the match between his/her own tone production attempt and the model. Each production was phonetically assessed, and a record kept -again under program control - of all assessment parameters: the pitch level at 20%, 50% and 85% of the F0 contour, and its duration. Screen messages were provided for each parameter that in any trial was assessed as unsatisfactory. In the first stage of training, learners were presented on the screen with the F0 contours of the same single-speaker exemplars as were heard by the perceptually-trained learners. Subsequent stages of the training required subjects to produce the tone words in response to L1 glosses presented (on the screen) embedded in a question frame meaning "What is the word for ...?" No target F0 contour was displayed in this stage of the training until a learner's production attempt for a word elicited was unsatisfactory. Subjects continued this training until, if they satisfied the proficiency criterion by producing two consecutive "good" tokens for each of twenty randomly-ordered tone type elicitations, their ability to perceive the tones was tested. The tone stimuli in this perception test were the same digitized natural speech tokens of four Beijing natives (2 male and 2 female) used with the perceptually-trained group. Subjects were first presented with tone tokens of one (female) speaker, and then heard all four speakers; throughout, the tones were varied in a fixed random order.

3. RESULTS

3.1 Perceptually-trained group

Of the 17 learners who attained proficiency in tone perception, 9 were able without any productive training to produce tones with acceptable F0 contours in a fully contrastive system. Since these subjects' only experience of the tones was auditory, their targets for tone production must have been derived from representations developed for tone perception. Moreover, a consistency in subjects' productions suggests that these representations were fairly stable -as might be expected of the hypothetical tone prototypes they had established by the end of the training. Detailed analysis of the production assessment records reveals a fairly high degree of acoustic-phonetic invariance: a particular error type was recorded significantly more often in all, or only one of, the five test productions per tone than in two, three or four of them. In other words, those productions not assessed as "good" tended to be characterized, for individual subjects and tones, by certain recurrent patterns of deviance.

To investigate the correlation of learners' perceptual abilities with their production performances, both were quantified. A general measure of perceptual ability, M1, over the duration of the perceptual training was provided by multiplying the number of trials required to reach criterion by the mean number of misidentifications per tone type. The total number of detail errors in the production test was taken as a measure of production ability, M2. Over all 17 subjects who satisfied the proficiency criterion there is a significant positive correlation of M1 with M2 ($r = .65$, $p < .01$), indicating that subjects' productive skill was generally commensurate with their ability in perceptual learning. In greater detail: by subject, and by tone, the correlation of (i) the proportion of tokens misidentified in the final stage of the perceptual training with (ii) the proportion of production trials characterized by errors other than of duration is moderate ($r = .42$) but again significant ($p < .01$). To a noteworthy extent, then, learners produced more accurately the tones which they could more accurately label.

3.2 Productively-trained group

Some subjects were able after only production training to make correct identifications of all or some of the tone tokens in the perception test. To investigate the correlation of these subjects' productive abilities and perceptual performances, the number of production trials resulting in an 'unsatisfactory' assessment was taken as an inverse measure of productive learning, and the number of types of identification error recorded in the perception test as a measure of perceptual ability. The correlation of these measures is noteworthy among the 15 subjects who satisfied the production proficiency criterion: $r = .70$ ($p < .05$). Among these subjects there is a similar correlation between the number of **types** of error observed in production attempts on the one hand, and in perceptual decisions on the other ($r = .69$, $p < .05$). These positive correlations between measures of tone-productive ability and tone-perceptual accuracy would suggest that F0 patterns learned for the direction of tone production were referred to for the perceptual categorization of tone tokens heard from other speakers.

In a comparison of the performances of the perceptually-trained and productively-trained groups, it appears that the perceptually-trained learners enjoyed some overall advantage. By the sign test, they made fewer errors in tone perception yet no more in tone production (in each case $p < .05$). This is perhaps not surprising, considering the comparatively unnatural learning conditions of the productively-trained group. Secondly, for each of the respective tones the performances of the two groups prove to be significantly correlated in both the perceptual and productive modalities (for perceptual confusions Pearson's $r = .72$, $p < .01$, and for production errors, $r = .76$, $p < .001$).

4. DISCUSSION

The present learners did not, it appears, need to be trained in production to be able to produce, or in perception to be able to perceive, the F0 patterns of the target phonetic system: training in one modality tended to be sufficient to enable a learner to perform in the other. A learner who, after perceptual training, was able - in some cases from the first attempt - to produce the tones correctly, must have correctly inferred the requisite acoustic targets, and drawn upon his knowledge of the articulatory-acoustic characteristics of his own speech to attain them in production. A learner who, after production training, was able to identify correctly the tokens of other speakers, presumably exercised the ability to map the acoustic output of others into the phonetic space of his own. Theoretically, this could have been accomplished by fitting time- and range-normalized candidate F0 contours to learned contour prototypes -by means, for instance, of fuzzy logical pattern-matching (see e.g. [3, 4]). While the outcomes of an experimental study should not be too freely generalized to natural learning situations, the present findings would support a model of L2 speech pattern learning in which the learner's primary goal is the construction of phonetic prototypes to which the operations of both perception and production may be geared. These prototypes capture the central acoustic tendencies of 'good' tokens, and serve both perceptual decisions (cf. [5, 6]) and production activity (cf. [7, 8]). However, since there may be no simple correspondence between learners' perceptual and productive values for acoustic-phonetic parameters (e.g. [9, 10]), these prototypes, it may be postulated, are operationalized by means of schemata (i.e. structured plans) which define the serial and hierarchical orderings of cognitive and motor activities, providing the algorithmic bases for psycho-acoustic decisions (in perception) and feedback-based adjustments (in production). An association of independent perceptual and productive schemata with each phonetic prototype will account for any divergences that may be observed between a learner's perception and production of a phone (see [9, 10, 11]). With operational schemata mediating between it and the phonetic events in relation to which it is defined, the prototype would theoretically satisfy one of the classic requirements of a phonological unit ([12]): that it be neutral with respect to the activities of production and perception.

REFERENCES

[1] LEATHER J. H. (1988) Speech pattern elements in second-language acquisition. Doctoral dissertation, University of London.

[2] LEATHER J.H. (1987) Automatic recognition of Chinese word tone from F0, with and without amplitude and speaker information. Proceedings of the European Conference on Speech Technology (Edinburgh) 1, 327-330.

[3] MASSARO D.W, COHEN M.M. and TSENG C-Y. (1985) The evaluation and integration of pitch height and contour in lexical tone in Mandarin Chinese. Journal of Chinese Linguistics 13, 267-289.

[4] TSENG C-Y, MASSARO D.W. and COHEN M.M. (1986) Lexical tone perception in Mandarin Chinese: evaluation and integration of acoustic features. In H.S.R. Kao and R. Hoosain (eds), Linguistics, Psychology and the Chinese Language. Centre of Asian Studies, University of Hong Kong, 91-104.

[5] KLATT D. (1979) Speech perception: a model of acoustic-phonetic variability and lexical access. Journal of Phonetics 7, 279 -312.

[6] BLUMSTEIN S. E. and STEVENS K. N. (1979) Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. JASA 66, 1001-1017.

[7] LINDBLOM B. E. F., LUBKER J. and GAY T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics 7, 147-161.

[8] LADEFOGED P., DECLERK J., LINDAU M. and PAPCUN G. (1972) An auditory-motor theory of speech production. Working Papers Phonetics (UCLA) 22, 48-75.

[9] SHELDON A. (1985) The relationship between production and perception of the /r/-/l/ contrast in Korean adults learning English: A reply to Borden, Gerber, and Milsark. Language Learning 35, 107-113.

[10] BOHN O. and FLEGE J.E. (1990) Perception and production of a new vowel category by adult second-language learners. In J. Leather & A. James (eds) Proceedings of NEW SOUNDS 90, University of Amsterdam.

[11] SHELDON A. and STRANGE W. (1982) The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. Applied Psycholinguistics 3, 243-261.

[12] LINELL P. (1982) The concept of phonological form and the activities of speech production and perception. Journal of Phonetics 10, 37-72.