PHONETICS EAR-TRAINING – DESIGN AND DURATION

Patricia D.S. Ashby

Department of English and Linguistics, University of Westminster, London, UK
ashbyp@wmin.ac.uk

ABSTRACT

Study of the product of routine phonetics ear-training revealed a number of effects. The goal/benchmark was mastery of the sounds of the IPA (including Cardinal Vowels). Students with no previous experience of phonetics were followed through a typical one year ear-training programme; their ability to recognise sounds was tested at two points during the year. Vowel identification findings confirmed that even in the non-language ear-training environment contextualized vowels are harder to identify than vowels in isolation, but they also raised questions about the contribution made by length of training to the level of achievement.

Keywords: Cardinal Vowels; ear-training; phonetics pedagogy; vowel perception.

1. INTRODUCTION

A longitudinal study of the process and product of general phonetic ear-training followed 125 UK-based undergraduates (a heterogeneous mix of nationalities, mother-tongues, etc., and none with prior knowledge of phonetics), through a one year course using established ear-training techniques (including nonsense words, substitutions, etc.) [2, 4, 7, 9, 10]. The aim was to describe the phonetic judgements made by students in a non-language learning context and, long term, to facilitate a more informed approach to practical phonetics training, materials design and assessment. Such training is still fundamental to a wide range of courses (and professions) including courses in general phonetics – practical skills account for 60% of the marks in the IPA's *Certificate* examination – speech therapy, drama, accent coaching, linguistic fieldwork, etc.

The study examined a number of factors: mother-tongue influences [1], confusability of vowels [3] and consonants, as well as the correlation between length of training and level of achievement, and the impact of contextualization on identifying sound values. Unlike language-based studies by Iverson *et al* [6], Nearey [8] or Strange *et al* [12], etc., there is no single language-specific influence here – stimuli are not real words and responses are made by a linguistically mixed audience; similarly, no overall consideration can be given to students' psychological relation to symbols [9], since they also have a heterogeneous orthographic background. Participants, in fact, constituted a typical UK undergraduate cohort and the materials were all ordinary teaching and assessment materials.

This paper examines further results for Cardinal Vowels (CVs). Language-based research suggests that isolated vowel qualities produced by a single speaker (we could call these 'simple targets' – see Strange's summary [11] of Delattre *et al* [5]) will be easier to identify than the same vowels produced by different speakers (introducing inter-speaker variability or 'elaborated targets', see also [11]) or produced in the random consonantal environments of nonsense words (in-word or 'dynamic targets', see also [11]). This paper describes responses to 'simple targets' against 'dynamic targets' and presents some initial findings on the effect of duration of training.

2. THE EVALUATION PROCESS

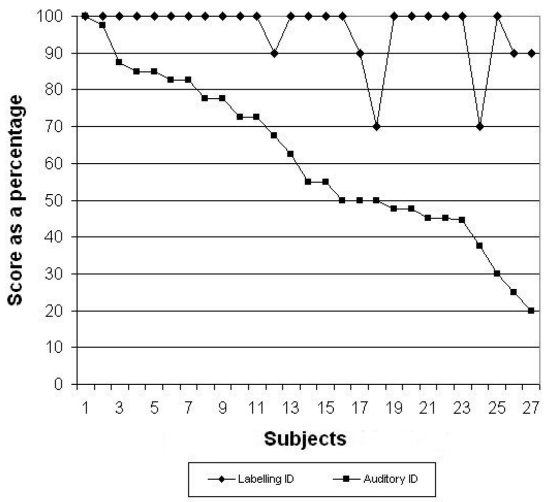
2.1. Scoring the answers

Responses were scored in two ways: (i) categorical right *vs* wrong (making no allowance for errors of height adjacency [3], etc.), and (ii) taking into account accuracy of backness, openness and rounding judgements – called here “BOR scoring” – in each transcribed response. Categorical scoring, of course, enables statements regarding how many sounds are right or wrong, but reveals very little about the nature of wrong perceptions and/or transcriptions. BOR-scoring, especially when coupled with further error-classifications for judgements which cannot even be submitted to this scoring process – the use of an English phonemic vowel representation ([eɪ] for pCV2 [e], for example), or use of as yet unlearnt secondary (sCV) values to represent primaries ([ø] for [o], etc.) – gives far more (useful) information, not only enhancing our understanding of the judgements themselves but also informing future pedagogy (which sounds and sound-relationships to focus on in practical training sessions, etc.).

Knowledge of the symbols themselves was checked in a separate test, ensuring that misrepresentations were not just a function of ignorance of symbol shapes. Confirmation of symbol knowledge is demonstrated in the results from one such check, see Figure 1. With 21 out of 27 students in this representative sample scoring 100% on symbols when prompted by BOR labels rather than auditory stimuli, misrepresentations identified in responses to the ear-training tests themselves were not regarded as a direct product of symbol-ignorance. Instead, they were described as:

- Out-of-range symbols (using sCVs before they had been officially learnt or practised; using consonant symbols instead of vowel symbols).
- Malformed symbols (invented shapes, not part of the IPA alphabet).
- English Vowel Symbol (use of cap-i, ash, upsilon, etc.).
- Insertion and deletion errors (addition of extra vowels or omission of vowels).

Figure 1: Graph showing knowledge of primary Cardinal Vowel symbols.



3. FINDINGS

3.1. Contextualization and identification success rates

It was expected that *distractions from the consonantal environment would lower the identification success rate for vowels as 'dynamic targets' compared to the rate when identifying them as 'simple targets'.*

In the nonsense materials, many of the consonants themselves were also in the process of being mastered/internalised and were therefore relatively unfamiliar (and maybe unstable) values which would contribute little to the identification of the vowels. The best opportunity for testing this first was at the 12-week assessment point, mid-way through the training programme. In the initial stages of training, students were exposed to primary Cardinal Vowels both in isolation (to help them internalise or memorise the eight actual vowel qualities) and in simple nonsense words (to help them start to respond to whatever cueing was available in a wide – and random – range of phonetic environments). At the 12-week assessment, therefore, halfway through the training period, it seemed reasonable to evaluate a) whether the values had been internalised at all (using 'simple targets') and b) how secure or robust this knowledge was (testing the same values in the face of distraction – or help – of the sort encountered in the consonantal environments of nonsense words, 'dynamic targets').

Accordingly, two exercises were presented at the first assessment: Test 1a consisted of isolated pCVs (15 values with five monophthongs and five 2-part vowel glides): 1 [u], 2 [ɔ], 3 [ɑ], 4 [e], 5 [o], 6 [ia], 7 [ɛu], 8 [ɔi], 9 [iɑ], 10 [oe], and Test 1b of nonsense words (five words each with two pCVs): 1 [dʒoɱip'], 2 [ɛɫruθ], 3 [çɔɸeŋg], 4 [fɾɑtʃeɫ], 5 [ɹuɲɠon̪].

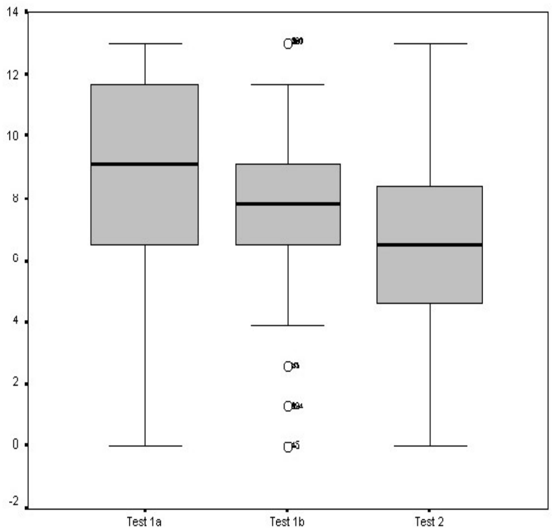
As expected, the results show that there is a big difference in identification success rate from Test 1a to Test 1b. Table 1 shows that in Test 1a, of the 1,875 vowel judgements made (125 individual responses to each of the 15 tokens), 1,470 or 78.4% were correct. In Test 1b, 10 vowel tokens in nonsense words or 1250 judgements, the identification success rate falls massively (a drop of virtually 20%), with only 734 or 58.7% being correctly identified.

Table 1: Percentage of correct judgements of Primary Cardinal Vowels in Test 1

TEST	Number of judgements	Context	Correct responses	
			Actual	As a %
1a	1,875	Isolated	1,470	78.4%
1b	1,250	In-word	734	58.7%

This deterioration is displayed in the first two boxes in Figure 2, Test 1a and Test 1b. What we can see here (as well as a few good and bad outliers) is the upper limit of the 'box' shrinking from Test 1a (isolated or 'simple target' pCV tokens) to Test 1b (in-word or 'dynamic target' pCV tokens). (Such correspondences could not be examined in Test 2 because no isolated tokens of CVs were used at the second assessment point.)

Figure 2: Boxplot illustrating performance of all subjects in all tests (identification success rates as a correlate of length of study).



3.2. Length of training and identification success rates

With regard to length of training, it was expected that *the identification success rate for in-word 'dynamic targets' would improve from the initial, mid-course assessment (Test 1b) to the final assessment (Test 2).*

Table 2: Percentage of correct judgements of all vowels (pCVs **and** sCVs) by Test

TEST	**Number of judgements**	**Context**	**Correct responses**	
			Actual	As a %
1b	1,250	In-word	734	58.7%
2	1,826	In-word	925	50.7%

Comparing the overall identification success rate for in-word vowels in Test 1 with that for in-word vowels in Test 2, Table 2 shows that there is a sizeable drop in achievement (8%), the rate dropping from 58.7% correct in Test 1 (week 12) to 50.7% (in week 24).

This overall deterioration in performance is again illustrated in Figure 2, looking now at the additional box for Test 2. Here, the downward trend observed from Test 1a to Test 1b continues – the top of the new Test 2 box is lower again. Additionally, the median drops (from 9 in Test 1a to 7 in Test 2), further strengthening the impression of deteriorating standards. However, what this image also shows is that the overall range of ability is the same – there are still subjects scoring 100% at the top end of the range and others scoring zero. This suggests that the tests themselves are ill-matched rather than that the training is ineffectual or the subjects lacking in ability.

If we disaggregate the Test 2 scores, factoring out the responses to pCVs and sCVs, the picture changes. Table 3 lists the identification success rates for pCVs alone at the two Test points. Now, the rate for pCVs can be seen not to have deteriorated at all (showing, in fact, a 0.9% increase) – at worst a plateau effect and at best, a tiny improvement.

Table 3: Disaggregated vowel identification success rates in Test 2.

Vowel type	**Number of judgements**	**Context**	**Correct responses**	
			Actual	As a %
pCV	1,392	In-word	831	59.6%
sCV	434	In-word	120	27.6%

The design of the training programme was such that sCVs were introduced into weekly practice classes only in week 13, after it was assumed the pCVs were sufficiently well-established to resist interference. Table 3 suggests this might be true, although the Test 2 results overall mask this (see below). Test 2 offers evidence of interference which could be one or more of several kinds: (i) the presence of the sCVS destabilises the established perceptions of pCV values, impeding improvement and/or (ii) with only 12 weeks of focused exposure and training, the sCV values are less well internalised by the learners and therefore attract a much lower identification success rate and/or (iii) so much effort is being expended on learning the new sCV values, that insufficient effort is being made to improve the already partly established pCV ones. There are also resonances here with the phonological impact of sound-system acquisition in the work of Iverson *et al* [6].

In spite of any interference, however, the identification success rate for the longer standing pCVs seems robust to the extent that the inclusion of sCVs has not caused it to deteriorate. In fact, the fractionally higher Test 2 score for primaries can be read two ways. It indicates some improvement has taken place, but that the visible effect of this has been neutralised by additional complications offered by the inclusion of the sCV values and by the (inevitably) more sophisticated and slightly trickier nonsense words. Alternatively, it could be read as indicating that there is a possible saturation point for contextualized vowel perception beyond which no further progress is likely to be made in the average year-long training course.

4. DISCUSSION AND CONCLUSIONS

4.1. Contextualized training materials

Perhaps unsurprisingly, this study shows that in-word or contextualized Cardinal Vowels – 'dynamic targets' – are, indeed, harder to identify correctly than isolated tokens, 'simple targets', of the same vowel qualities.

Although trivial on one level, this finding does have a message for materials design. Ear-training programmes (live or online) need to give careful thought to balancing the structure of materials and maximising the time spent on training the ability to identify vowels in (*quasi*) real language materials.

A further pointer in the possible mis-match of test materials suggests consideration might also be given to comparative levels of difficulty in test materials if results are to be a true reflection of level of achievement and a clear demonstration of progression.

4.2. Perceptual capacity threshold

It also seems that for any given learner, the identification success rate could be governed at least in part by that learner's perceptual capacity.

The fact that the total ***range*** of results (summarised in Figure 2) remained largely constant across the three different tests and two levels of difficulty – at all stages, there were individuals scoring full marks and others scoring nothing – suggests that some learners may be more predisposed towards becoming phoneticians than others. This finding hints at an innate capacity which may not be shared by everyone.

4.3. Length of training

However, it is also the case that the very small improvement detected in the pCV results after the extended training time suggests that if managed properly, more practice may indeed generate improved results even if these are subject to personal ceilings. Nonetheless, it is by no means certain, judging from this very small difference noted for the identification success rate of pCVs from week 12 to week 24, whether for general purposes any more extended or intensive training is either necessary or likely to be beneficial.

More investigation is needed to determine whether or not, with carefully controlled materials and programmed learning, any individual subject can achieve a meaningfully increased level of accuracy. (If this is not the case, the finding has implications for course design and resource management which cannot be ignored.)

5. REFERENCES

[1] Ashby. P. 2000. Phonetic judgments of Cardinal Vowels in ear-training. Paper presented to the *Colloquium of the British Association of Academic Phoneticians.* Glasgow. (Unpubl.)

[2] Ashby, P. 2002. *Practical Phonetic Training and the Nature of Phonetic Judgements*. Unpublished PhD thesis, University of London: UCL.

[3] Ashby, P. 2003. Learning Cardinal Vowels. *Proc 15th ICPhS.*. Barcelona, 3089-3092.

[4] Ashby. P. 2006. Phonetics Pedagogy. Brown, K. (ed.) *Encyclopedia of Language and Linguistics.* Oxford: Elsevier. 372-378.

[5] Delattre, P.C., Liberman, A.M., Cooper, F.S. and Gerstman, L.J. 1952 An experimental study of the acoustic determinants of vowel colour: Observations of one- and two-format vowels synthesized from spectrographic patterns. *Word* 8, 195-210.

[6] Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Ketterman, A. and Siebert, C. 2003. A perception interference account of acquisition difficulties for non-native speakers. *Cognition* 87, B47-B57.

[7] Jones, D. 1911. Phonetics and ear-training. *Die neueren Sprachen* 19, 318-319.

[8] Nearey, T.M. 1989 Static, dynamic, and relational properties in vowel perception. *J .Acoust. Soc. Am* 85, 2088-2113.

[9] SIL. 1983. *Introduction to practical phonetics*. High Wycombe: Summer Institute of Linguistics.

[10] Smalley, W. 1963. *Manual of articulatory phonetics*. New York: Manual of Anthropology.

[11] Strange, W. 1989. Evolving theories of vowel perception. *J. Acoust. Soc. Am* 85, 2081-2087.

[12] Strange, W, Akahane-Yamada, R., Kubo, R., Trent, S.A., Nishi, K. 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *J. Acoust. Soc. Am* 109, 1691-1704.