

The Relation Between Perceptual and Production Categories in Acquisition

Ian Watson

Christ Church, Oxford and Oxford University Phonetics Laboratory

ABSTRACT

Results from two experiments on acquisition of the word initial voicing contrast are compared, one on perception, one on production. Both involved monolingual and bilingual (French/English) subjects. Monolinguals are shown to have frequent disparities between their use of parameters in production and perception. Furthermore, bilingualism is found to affect subjects' production and their perception differentially, in a way which suggests that there cannot be isomorphism between their production and perceptual representations. However, the ongoing interference between the phonetic categories in bilinguals' two languages also argues that at the phonological level there is identity between cognates in their two languages. This pattern of phonological identity but phonetic difference is taken as support for models of speech processing in which distinct representations are attributed to phonology and phonetics, and to production and perception.

1. REPRESENTATIONS IN SPEECH

Accounts of the human speech faculty differ fundamentally in the way they model the relationship between speech production and perception; these differences furthermore often reflect equally deep-seated disagreements concerning the relation between phonetics and phonology. In both cases, a crucial question concerns the number of different representations postulated. For the production/perception dichotomy, the division is between approaches which attribute independent representations to production and perception (e.g. Keating's model [7]) and those which deny the existence of a dichotomy, postulating a single (usually articulatory) representation for both [2,3,4,9]. The latter approach favours perceptual models such as Direct perception and the revised Motor Theory [3,9]. For the phonetics/phonology relation, the essential contrast is between models which postulate independent phonetic and phonological representations (termed "translation models" by Nolan, [10]) and those in which there are no independent phonetic representations, the phonetic realisation of phonological units being seen rather as an automatic consequence of their very specification (hereinafter "unity models"). A recent example of a translation model would be that presented by Keating [7]; unity models are those proposed or implied by approaches such as action theory and articulatory phonology [4,2]. Unity models are often associated with views of phonetics involving a single representation, that is, with perception based on articulation.

In this paper, evidence from two experiments on phonetic acquisition is used to argue, for both dichotomies, in favour of models with more rather than fewer distinct representations. The arguments come from three related sources (i) Disparities in the patterns of acquisition by monolinguals in perception and production; (ii) differences between monolinguals and bilinguals attributable to bilinguals' identifying two different phonetic categories (e.g. French [k] and English [k^h]) with each other at the phonological level; (iii) disparities in the effect of bilingualism on perception and production.

Single and dual-representation models of phonetic processing may be distinguished according to the predictions they make concerning acquisition in general, and in particular, concerning the effect of bilingualism on processing. Single representation models would predict parallelism in the development of categories in both production and perception, and thus that both should be affected equally by bilingualism; dual representation models would allow for disparities between the two.

Data from bilinguals can also throw light on the phonetics-phonology relation. Logically, bilinguals could show one of three conditions: (i) completely merged phonetics **and** phonology; (ii) completely distinct phonetics and phonology; (iii) divergent phonetic realisation in their languages but evidence that cognate phonological categories are identified with each other cross-linguistically. The first two conditions would not allow us to adjudicate between translation and unity theories, but the latter would be embarrassed by the third type, which implies a distinction between phonological and phonetic representations: if phonetic objects differ only because their phonological specifications differ, how can different phonetic objects be identified with the same underlying phonological representation?

2. THE EXPERIMENTS

Both experiments involved monolingual and bilingual child subjects, and monolingual adult controls. The child subjects, but not the adults, were the same in the two experiments. The children were in three age groups, 6, 8 and 10 years. For each age group there were French and English monolinguals, bilinguals resident in Paris and bilinguals resident in London. All the bilinguals were simultaneous primary bilinguals who had been exposed to both languages from birth, through having one parent speaking each to them. All were receiving education in both languages. Numbers in groups were as shown in [table 1](#).

	English Mono	Bilingual English French		French Mono
Adult	5			5
10 years	6	9	10	7
8 years	5	7	6	7
6 years	5	5	6	8

Table 1: Number of speakers in each group

English and French differ in the realisation of the voicing contrast. In word-initial plosives, English has a long vs. short lag contrast, French a prevoiced vs. voiceless unaspirated contrast. Furthermore, the languages vary in the degree to which other acoustic correlates of the voicing contrast are produced and used in perception. In particular, it is known that the F0 onset frequency parameter [5,13] tends to be more salient in French than English. Both VOT and F0 onset frequency are parameters whose precise use in both production and perception is developed during childhood [1,14]. They may thus be used to investigate the extent of the influence of bilingualism on language processing, and hence, the nature of the relationships investigated here

2.1 Production experiment

Full details of the experiment are in [15]. Briefly, subjects produced one word utterances in response to pictures they were shown of common objects or activities. After a training period to make sure they knew the relevant words, they were recorded, producing five tokens each of the words 'cash' and 'gash' in English, 'cache' and 'gâche' in French. The tokens were embedded in a list with numerous fillers, in randomised order. Bilingual subjects recorded both lists, in counterbalanced order, the two recordings being on different days in most cases. The recorded /kaf/ and /gaf/ tokens were then analysed acoustically for voice onset time and F0 onset frequency. Statistical analysis was carried out using the GLIM general linear modelling package [6]. This showed that bilingualism, and its interactions with language and voicing of token (i.e. /k/ or /g/) were among the most significant factors ($p < .01$). Mean VOTs are shown in Table 2, with the voiceless tokens uppermost in each cell. Planned comparisons showed that the bilinguals' mean VOTs (which were significantly different in their two languages) were not distinct from monolinguals' for the /k/ utterances, but were for /g/ for five groups, marked with an asterisk in Table 2. For F0, only three groups showed a distinction in F0, based on voicing category, these being the monolingual adults and 10 year olds, and the 10 year old French bilinguals. In two other cases, the F0 difference just failed to reach significance, these being the 8-year-old French monolinguals and the French-resident 10-year-old bilinguals speaking English. For F0 production, then, this last case is the only one showing a divergence from monolingual norms, and a transfer to English of a pattern appropriate in French.

	Eng mono	Fr mono	Fr bi Eng	Fr bi Fr	Eng bi Eng	Eng bi Fr
6	106 23	49 -67	91 -19*	37 -58*	91 -11*	49 -17*
8	98 21	30 -73	118 7	47 -62	84 -29*	52 -73
10	99 33	33 -84	85 -46	41 -82	90 4	51 -64
Ad	71 20	35 -79				

Table 2: Mean VOT values in production

2.2 Perception experiment

Details of the perception experiment are in [16]. Two synthetic VOT continua were created using the KLSYN formant synthesiser, such that extreme tokens were heard as [gaf] (-30 ms VOT) or [k^haf] (+50 ms VOT), with 5 ms. intermediate steps. In the first continuum F0 onset was at 128 Hz, in the second 103 Hz. French and English versions were made of both continua, by appending the carrier phrases 'je dis' and 'I say' respectively. Six versions of each were recorded onto tape in pseudo-random order and used in a forced-choice identification task. Bilingual subjects responded in both language conditions, in counterbalanced order. The last five responses from all subjects in each condition were pooled within groups, and subjected to analysis of variance using GLIM. The expectations were that (i) the VOT boundary would be lower in French than in English (ii) the lower F0 onset would cause an upward shift in VOT category boundary, but more so in French than English. Again, these known patterns were then used to test the effect of bilingualism on processing in both languages.

With respect to VOT, the bilinguals had distinct category boundaries in their two languages, but, relative to monolinguals, at all age groups they had higher boundaries in French and lower boundaries in English, suggesting a significant level of interference (see Table 3). This interference varied with country of residence, as the French residents have lower mean VOT boundaries in both languages than the English residents, and with age, as the rapprochement of category boundaries increases as the subjects get older (see Table 4).

For the F0 parameter, the evidence of transfer between languages is still stronger. Whereas none of the English, but all the French monolinguals show a statistically significant boundary shift for the lower F0 onset, all the bilingual groups but one show a shift in both languages, although that in English is lower than that in French. The one bilingual group *not* to show a shift in English was the 10 year old English residents, again suggesting a role for both age and country of residence in the pattern of acquisition.

	Eng Mono	Fr Mono	Fr bi Eng	Fr bi Fr	Eng bi Eng	Eng bi Fr
6	25	8	20	8.5	21.8	15.5
8	22.5	-3	20	12	22	16.5
10	21.5	5	17	11	19	14.5
Ad	18.5	5				

Table 3: Mean Category Boundaries

AGE	Monoling	French Bi	Engl Bi
6	17	11.5	6.3
8	25.5	8	5.5
10	16.5	6	4.5

Table 4: Difference between French and English category boundaries

3. COMPARISON OF EXPERIMENTS

Three main differences emerge between the perceptual and production categories. Most obviously, there are clear disparities between the mean VOT values observed in production and the category boundaries found in perception. This is most marked for the English monolinguals; comparison of tables 2 and 3 shows that for all four age groups, the perceptual category boundary is situated below their mean VOTs for voiced tokens. This disparity is in part attributable to the properties of the synthetic stimuli used; possibly the intensity of the burst [12] or the rapid formant transitions. Nevertheless, it does attest to the fluidity of the perception/production relation.

The second difference concerns the effect of bilingualism on the production and perception of VOT. In production, no bilingual group showed any significant effect on the production of voiceless stops in either language, and there was a significant effect on voiced stops for four groups (out of six) in English and for one out of six in French. Furthermore, these disparities may be of little significance for bilinguals [11], as both prevoiced and short lag realisations of voiced tokens are found in monolingual utterances in both languages. In perception, in contrast, there was a strong effect of bilingualism in all groups for both languages, an effect which grew as age increased.

The third difference is that the bilinguals transfer use of the F0 parameter from French to English to a much greater degree in perception than in production. Furthermore, amongst the French monolinguals, there is a clear pattern of early acquisition of this parameter in perception (it is already used by the youngest subjects) but not in production (until the age of 10).

4. DISCUSSION

The above comparison shows that the perceptual responses of some groups show cue-trading relations which have no equivalent in production; that bilingualism has different effects on production and perception. and that production norms for VOT appear incompatible with patterns of perceptual response. All of these disparities suggest that the representations used in production and perception cannot simply be isomorphic

It is clear, indeed, that not all of the variables involved in a subject's perceptual representation will figure in the input to production. For a 6-year-old monolingual French speaker, the F0 parameter is included in the perceptual representation but is not specified as a production variable. For a 10-year-old French monolingual, the F0 parameter is presumably specified in production as well as in perception. There must also be other parameters which are achieved automatically in production, but whose perceptual salience has to be learnt. An example would be the role of F1 onset frequency in the voicing contrast. In English, and other languages with a short/long lag contrast, F1 onset frequency will covary inevitably with VOT [15]; thus it is presumably not specified as a separate parameter in speakers' production routines. For perception, however, the value of this acoustic feature has to be learned [14, 15]; the fact that it is inevitably achieved in production does not guarantee its availability in perceptual representation.

That perception and production involve distinct representations is confirmed by the effects of bilingualism on acquisition. The F0 parameter is highly salient in perception, yet not found at all in production in some groups. For VOT, the bilingual subjects showed a greater divergence from the monolinguals in perception. This divergence increases over time. The cause of this pattern, whereby bilingual perceptual categories get less distinct in their two languages, is presumably an inability to filter out entirely the effects of their exposure to French in building perceptual categories in English and vice versa [16]; this is not directly transferred to production.

These findings argue strongly against single representation models of phonetic processing. Direct Perception [3] suggests that perception involves access to the gestures underlying articulation. It is difficult to see, then, why the same signal should elicit different processing strategies from the same subjects, with, for example, F0 producing a much larger trading relation in French than in English. The same results offer different objections to the Motor Theory in its revised form [9]. If perception involves cognitively mediated knowledge of production invariants, why are different parameters salient in production from perception? Again, a definition of the unit to be recovered as a more abstract symbol associated

by experience with prototypical values, but not constituted by those values, overcomes these problems.

5. PHONETICS AND PHONOLOGY

The interference shown by the bilingual subjects in perception and production suggests that there must be a level of representation, presumably phonological, at which cognate sounds in their two languages are identified as equivalent, despite their phonetic differences. If phonological identity can be established despite phonetic difference, then there must be a clear distinction between the phonetic and phonological levels. These results thus fail to support an approach such as articulatory phonology, in which phonological units are essentially gestural specifications of the articulation of segments - in other words, production routines. If this were correct, then what is conventionally thought of as the /k/ phoneme in French must be quite different from the /k/ phoneme in English; the units used by a bilingual in the production of the two sounds are distinct from one another, as are the units used in the corresponding percepts. In this case, it is difficult to see why bilinguals should allow productions of one to affect the other, as it has been argued they must to explain this study's findings (cf. also [8,11]). Allowing a more fluid relationship between the phonological units and their realisations eliminates this problem.

6. CONCLUSION

The studies compared here thus suggest a model in which, in each language, distinct perceptual and production representations are associated with a phonological representation. Cognate sounds in different languages may be identified with each other at this phonological level, and this identification may lead, through time, to a rapprochement of the phonetic categories which realise them. Of the types of model discussed in section 1, it is thus clearly the translation type which is favoured by these findings.

7. REFERENCES

- Bernstein, L.E. (1983) Perceptual development for labelling words varying in voice onset time and fundamental frequency. *Journal of Phonetics* 11, 383-393.
- Browman, C.P. & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology* 6, 201-251.
- Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14, 3-28.
- Fowler, C., Rubin, P., Remez, R.E. & Turvey, M.T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (ed.) *Speech Production*, (Volume 1). New York: Academic Press.
- Haggard, M., Summerfield, Q., & Roberts, M. (1981) Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading F0 cues in the voiced- voiceless distinction. *Journal of Phonetics* 9, 49-62.
- Healy, M.J.R. (1988) *GLIM, an introduction*. Oxford: Clarendon Press.
- Keating, P. (1990). Phonetic representation in a generative grammar. *Journal of Phonetics* 18, 321-334.
- Leopold, W. (1939-1949). *Speech development of a bilingual child*. Evanston, Illinois: Northwestern University Press, 4 volumes.
- Lieberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revisited. *Cognition* 21, 1-36.
- Nolan, F.J. (1990). Who do phoneticians represent? *Journal of Phonetics* 18, 453-464.
- Obler, L. (1982) The parsimonious bilingual. In L.Obler and L Menn (eds.) *Exceptional language and linguistics*. 339-346. New York: Academic Press.
- Repp, B.H. (1979). Relative amplitude of aspiration noise as a voicing cue to syllable initial stop consonants. *Language and Speech* 22, 173-189.
- Serniclaes, W. (1987) Etude expérimentale de la perception du trait de voisement des occlusives du français. *Doctoral thesis*. Faculté des Sciences Psychologiques et Pédagogiques, Université Libre de Bruxelles.
- Simon, C. & Fourcin, A.J. (1978) Cross-language study of speech pattern learning. *Journal of the Acoustic Society of America* 63, 925-935.
- Watson, I.M.C. (forthcoming^a) The effect of bilingualism on the production of the voicing contrast in English/French bilingual children. *Journal of Phonetics*
- Watson, I.M.C. (forthcoming^b). The development of perceptual correlates of the voicing contrast in primary bilingual children. *Journal of Phonetics*.