Is variation encoded in phonology?

Maria-Josep Solé
Universitat Autònoma de Barcelona, Spain
E-mail: mariajosep.sole@uab.es

ABSTRACT

This paper reviews evidence on quantitative knowledge of lexical entries and explores the view that the phonological representation stores sets of experiences instances of the category and its frequency distribution, as claimed by exemplar models. Implications of such models for some areas of phonology (phonemic categories, neutralization, allophonic and lexical split, language acquisition, and sound change) are considered.

1. INTRODUCTION

Much of the work in phonetics has been directed to provide a physical interpretation (e.g., acoustic and articulatory correlates, perceptual data, psycholinguistic evidence) of the categories suggested by phonologists as given (features, phonemes, syllables, boundaries, etc). In particular a great amount of effort has been directed to find the invariant articulatory and acoustic correlates of a predefined set of discrete units, such as features or segments. The quest for invariant acoustic or articulatory correlates of features has assumed a single, abstract phonological representation, made up of sets of feature specifications, which would have invariant correlates in the physical domain. We have made a great deal of progress in the acoustic and articulatory characterization of speech features and segments, and in mapping the discrete representations to the varying acoustic result through universal or language-specific rules of phonetic implementation with complex mapping functions, but the observed variability, due to contextual, prosodic, social, stylistic and rate factors, has made it almost impossible to find invariant correlates in the physical domain[1].

The lack of invariant physical correlates of speech features and segments has questioned the premise that inspired them, *i.e.*, that mental representations are invariant, categorical and abstract. Thus in the last ten years the focus of the phonetics-phonology interface has moved away from discrete and invariant internal representations, stripped off the phonetic details and has turned to prototype, e.g., [1], usage-based and experience-based probabilistic models [2, 3, 4], which assume that variation is part of the internal representation. These models allow us to incorporate the statistical structure of phonetic variation in the underlying phonological categories. If memories for sounds and words are in any way like memory for other types of sensory input, then we would not expect that words are stored with phonetic details removed, but rather that phonetic detail is stored and contributes to the shaping of abstract averages or templates computed from statistical evidence [5], a template-like prototype structure which includes instances from which the template is computed [1], or a set of actual instances of the phonological category (TRACE model). Probabilistic models have been used for a long time in speech processing, in psycholinguistic models of lexical access, and in other domains of cognitive modelling (*e.g.*, visual, pattern recognition, decision making).

Whereas in classical formal phonology the focus is on category labels and systematic processes, much of the work in phonetics and in laboratory phonology in the last years has emphasized the gradient and quantitative structure of phonological categories, which suggests (i) that phonological categories are experientially-based, the sum of all experienced instances of the category, and (ii) that phonetic variability itself may be part of the make-up of the phonological categories [1, 4, 6]. Thus, each phonological category is associated with an experientially-based frequency distribution of phonetic realizations, with a central tendency (or prototypical values), that may differ cross-linguistically and cross-dialectally, and a certain range of variability which may also vary due, for example, to neighbourhood density. By exposure to multiple instances of the category we learn the modal values and the permissible parameters of variation. Thus the observed statistical variability may have a cognitive status, that is, mental knowledge of language may itself be statistical as a result of exposure to and use of language.

We will now explore some areas of phonology where quantitative information about speech allows a better modelling of the phenomena.

2. RELATING PRODUCTION AND PERCEPTION

One main advantage of experientially-based models, which describe underlying phonological units and patterns in terms of the statistical variability in speech data, is that the model readily relates to on-line production and perception. An early attempt to relate speech perception data to the distribution of phonetic outputs was presented by Lisker and Abramson [7]. They compared the results of perceptual categorization to measurements of VOT in various languages. Their results are shown in Figure 1, where the

 ISBN 1-876346-nn-n © 2003 UAB

frequency distribution of VOT values for word-initial /d/, /t/ and /tʰ/ in Thai, are shown by vertical bars and the identification curves are shown by lines. The correspondence between the two sets of data is clear and has been found in other studies. The crossover points of the identification curves tend to fall between the clusters of production data in the histogram. This simple descriptive technique allows us to view phonological categories as frequency distributions of phonetic data along a single dimension. It also allows us to predict the number and nature of the phonemic categories from the clustering of production data. Obviously, the tripartite picture in Fig. 1 requires multiple dimensions to account for the phonetic realization of contrasts in different prosodic, positional and speaking rate conditions. Elaborating on Lisker and Abramson's work, Nearey and Hogan [8] devised a quantitative technique for assigning a new stimulus with a given VOT value to the category for which it has the highest probability for group membership, estimated from the frequency distribution of each category. This is a model of perceptual categorization based on the distribution curves of phonetic outputs, which can be generalized to multidimensional continua. Statistical techniques of this type have been used for a long time by categorizers modelling speech or pattern recognition. More recently, Pierrehumbert [4] extended this perceptual classification model to production. The modelling of phonological categories and processes using primitives and procedures utilized to model other cognitive abilities anchors phonology in the physical world.

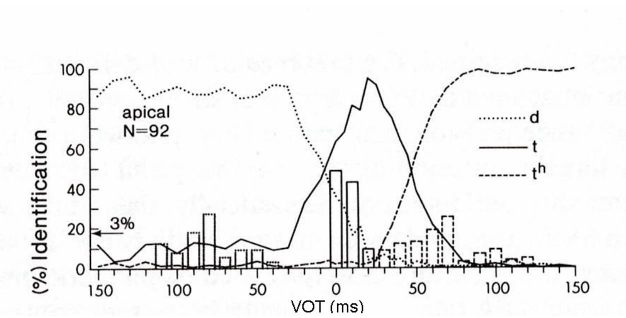

Fig. 1. Histograms of production data and identification functions for Thai /d, t, tʰ/ (from Lisker & Abramson 1970)

3. FIRST AND SECOND LANGUAGE ACQUISITION

It is well known that language acquisition is dependent on exposure to language input and that more frequent forms are acquired earlier, identified earlier, and produced accurately earlier. Beckman and Edwards [9] analyzed imitations of novel words by children and found that familiar sequences were produced more accurately than unfamiliar sequences. Derwing and Baker [10] provide evidence that frequency of forms predicts order of acquisition of different plural allomorphs, which illustrates the role of token frequency on morphological acquisition. These results suggest that the frequency of the sequence or word reinforces the storage in the long-term memory.

In second language acquisition, exemplar models allow us to represent the restructuring of the phonetic properties associated with certain categories into a different set of categories. In the early stages of second language acquisition, the first connection between the L2 sound and the L1 sound is at the level of the category label, such that L2 sounds are identified as realizations of an existing L1 category [11], *e.g.*, target English /t/ and /d/ are assimilated to Spanish /t, d/ and their phonetic outputs. This is represented by arrow (1) in Figure 2. Over time learners start making the connection between the L2 label and the distribution of phonetic properties that characterize the new category, e.g., modal values for VOT around +70 ms and 0 ms for English /t/ and /d/, respectively; a short gap of acoustic energy *vs* continuant realization for Spanish /d/ (postvocalically, following consonants that allow the free flow of air through the center of the oral cavity and optionally utterance initially); effects on preceding vowel length, etc., as illustrated by arrow (2). The production of an L2 sound will eventually correspond to the phonetic properties represented in the new categories. Maybe the most compelling evidence for the restructuring of the distribution of phonetic values and dimensions in the same portion of the phonological space for new L2 categories is when phonetic transfer from L2 to L1 occurs in proficient L2 speakers. This has been observed in bilinguals' production of stops in their L1 with VOT values resembling those typical for stops in the L2 [12]. This is in accord with the notion that speakers incorporate phonetic knowledge all the time and reflects a dynamic rearrangement of the categories as claimed by exemplar models.

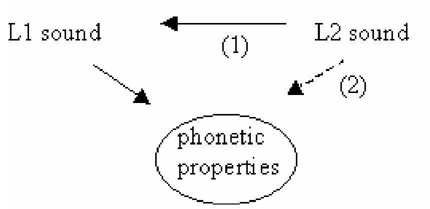

Fig. 2. Diagram of the relationship between L1 and L2 category labels and phonetic properties in various stages of second language acquisition.

Such models make interesting predictions about degree of language exposure and production/perception. Acquiring a fully native accent involves building up probability distributions for all the phonological elements in their various contexts which requires hearing and recording a large amount of speech. Perceptual training experiments have shown that robust and long-term foreign language categories are learnt only if learners are exposed to highly variable stimuli and multiple speakers [13], which suggests that speakers encode and retain instances or exemplars from the speech stimuli to form the complex category structure and permissible variation. Along similar lines, Shockey [14] attributes the failure of speakers from Honk Kong, proficient in English, to correctly identify a gated sample of reduced English speech (correctly identified by English speakers) to the lack of adequate early exposure to L2 variability and, consequently, lesser tolerance to variation.

4. PHONEME CLASSES AND NEUTRALIZATION

An interesting feature of exemplar models is that they readily relate phonological categories, or labels, to the statistical structure of phonetic variation along certain parameters. Thus, for example, voiceless and voiced stops in English distribute around different values along the VOT continuum (around 70ms for voiceless stops and 0-20ms, with a scattering of values towards realizations with voicing lead, for voiced stops [15]), such that they present two frequency distributions with little overlap. (This is in line with the principle of 'sufficient perceptual separation' or 'distinctiveness' between contrasting phonetic units.) Unaspirated stops following /s/, on the other hand, which reflect the neutralization of the voiced/voiceless contrast, as in *speak*, present VOT values in the range of 0-30 ms, similar to those for voiced stops but without the negative skewing towards negative values, congruent with the notion that the contrast between the two distributions merges into a single set of equidistant values. Other examples of the result of neutralization distributing between the values of the two contrasting categories along a certain phonetic dimension are English final unstressed vowels [i] and [u], in *coffee* and *to*, the result of neutralization of the /i:/-/ɪ/, /u:/-/ʊ/ contrasts. Though not all types of neutralization result in phonetically intermediate sounds (*e.g.*, English flapped /t, d/), they can be described in terms of the perceptual merging of overlapping distributions resulting from prosodic or contextual effects, or time constraints.

Phonological mergers (such as dialectal American English *pin/pen*) can be viewed as highly overlapping phonetic distributions which collapse into a single category.

5. ALLOPHONIC SPLIT

A great deal of allophonic variation can be attributed to coarticulation with adjacent sounds, *i.e.*, the variation can be described in terms of a numerical model which takes into account the degree and extent of coarticulation and the production constraints of the sounds involved, e.g., [16]. However, some of the allophonic variation seems to be under the control of the speaker. Such variants would have different target properties since these differences are made deliberately and are not simply ascribable to coarticulation. These have traditionally been called 'extrinsic allophones'. Solé [17] presented evidence of an allophonic split (or extrinsic allophone) in American English. She observed that the extensive vowel nasalization found in American English in vowels followed by a tautosyllabic nasal cannot be explained in terms of anticipatory nasal coarticulation. She measured the oral and nasalized portions of vowels followed by oral and nasal consonants (CVC, CVN) across different speaking rates. In coarticulatory nasalization, the onset of velum lowering in CVN sequences should be timed relative to the following nasal consonant, and therefore the nasalized portion of the vowel should be more or less constant in spite of durational differences in the vowel across speaking rates. In CVN sequences in American English, however, velum lowering is timed relative to the onset of the vowel, and thus the vowel is nasalized throughout across different rates. Thus the temporal extent of vowel nasalization has been dissociated from the conditioning environment and has become part of the programming instructions for the vowel, that is, an actual nasalized allophone. An allophonic split between oral and nasalized vowels in American English (as a phonemic split) can be adequately accounted for by exemplar models where each category is stored as a cluster of instances of the category or exemplars. A nasal consonant context will produce vowel exemplars which are biased toward the nasalized end of the continuum whereas an oral context will bias the outputs towards the oral end of the continuum. In time this will result in a bimodal frequency distribution along the nasalized-oral dimension, such that the exemplar cluster of vowels in a nasal context will consist of heavily nasalized allophones. The allophonic split could thus begin with a differentiation of pronunciations that was at first purely contextual. If the contextual differences became phonologized, this would lead to two distinct allophones, with the contextual effects encoded in the vowel.

6. SOUND CHANGE

Finally, quantitative effects are important because they tell us something about how experience changes the system. Bybee [6, 18] accounts for frequency effects in historical sound change. She observes that vowel weakening and vowel dropping – due to prosodic reduction, gestural economy and predictability – is more advanced in high-frequency words (e.g., *memory*, with a syllabic r) than in low-frequency words (*mammary*, with schwa and r). Since high-frequency words, by definition, occur more often, every time a frequent word is used, it adds a reduced exemplar to the distribution and the modal value will shift toward the reduced pronunciation. Thus, leniting changes will first become evident in high-frequency words. This model assumes that the lexical representation of a word is continuously updated as more instances are heard. Some words belonging to a more formal register (e.g., *humility* vs *humble*, *chaise* vs *chair*, *recuperate* vs *recover*) would resist sound change because they are mostly used in a formal style or by older speakers, thus resulting in a lexical split as the cluster of phonetic realizations of frequent words moves towards more reduced pronunciations. Such interpretation of sound change was advanced by Ohala [19] who suggests that diachronic change has its sources in synchronic phonetic variation. Frequency effects are also evident in frequent words that have been in the language for a long time and show vowel weakening, such as *postman, England* whereas more recent and/or less frequent words retain a full unstressed vowel (e.g., *mailman, Disneyland*). Thus, frequency appears to be recorded in long-term memory.

Ohala [19] suggests that some sound changes are due to failure on the part of the listener to discount coarticulatory effects which result in acoustic/perceptual similarity with another sound category (e.g., /kj/ > [tʃ], due to failure to

ISBN 1-876346-nn-n © 2003 UAB

discount the effects of the palatal glide on the realization of the velar stop – fronted constriction, spectral peak of the burst, long and intense noisy release, etc.). If contextual and stylistic variants are part of the phonological representation, as exemplar models suggest, how could sound changes due to listeners' failure to discount the effect of the context occur? The answer might be that these sound changes result from interpreting the most common realization of a certain category precisely because it has a higher level of activation (e.g., the auditory properties of a palatalized stop, with a long and intense fricative release cause a high activation of /tʃ/ and a lower activation of /k/, hence original /kj/ is reinterpreted as /tʃ/).

7. CONCLUSIONS

We have reviewed evidence that provide support to the view that the phonological representation stores phonetic and statistical details of the experienced instances of segment and word categories. Exemplar models provide an architecture that is suited to modelling the observed data. By way of conclusion we may note that the long-standing split between discrete abstract accounts of systematic and regular phonological variation on the one hand – which were the focus of Neogrammarians and abstract phonology – and the gradient and variable nature of dialectal, diachronic, social and stylistic data – the focus of variationists, dialectologists, historical linguists (such as Y. Malkiel, see his 1967 paper 'Each word has its own history'), and most recent research in phonetics and laboratory phonology – can be bridged by exemplar models in which segments and words are represented as frequency distributions of phonetic realization. Regular, systematic effects which even apply to novel words can be viewed as generalizations speakers are able to make using the most common pattern in the distribution.

NOTES

1. In gestural phonology invariance is shifted to the abstract 'gestural' level, with variation being part of the execution program.

ACKNOWLEDGEMENTS

Work supported by grantsBFF2001-2498, BFF2000-0075-C02-01 from the Ministry of Education, Spain and by the research group 2001SGR 00425 of the Catalan Government.

REFERENCES

[1] K. Johnson, “Speech perception without speaker normalization,” in *Talker Variability in Speech Processing*, J. W. Mullennix, Ed., pp. 145–166. San Diego: Academic Press, 1997.

[2] M. Barlow and S. Kemmer, Eds., *Usage-Based Models of Language*. Stanford: CSLI, 2000.

[3] J.Bybee and P. Hopper, Eds., *Frequency Effects and the Emergence of Linguistic Structure*, Amsterdam: John Benjamins, 2001.

[4] J. B. Pierrehumbert, “Word-specific phonetics,” in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds., pp. 101–139. Berlin: Mouton, 2002.

[5] K.N. Stevens and S.E. Blumstein, “The search for invariant acoustic correlates of phonetic features,” in *Perspectives on the Study of Speech*, P.D. Eimas and J.L. Miller, Eds., pp. 1038. Hillsdale NJ: Erlbaum, 1981.

[6] J. Bybee, “Lexicalization of sound change and alternating environments,” in *Papers in Laboratory Phonology V*, M.Broe and J.Pierrehumbert, Eds., pp. 250–269. Cambridge UK: CUP, 2000.

[7] L.Lisker and A. Abramson, “The voicing dimension: Some experiments in comparative phonetics,” in *Proceedings 6th ICPhS*, B. Hala, M. Romportl, and P. Janota, Eds., pp. 563–567. Prague: Academia, 1970.

[8] T.M. Nearey and J.T. Hogan, “Phonological contrast in experimental phonetics: relating distributions of production data to perceptual categorization curves,” in *Experimental Phonology*, J. J. Ohala and J.J. Jaeger, Eds., pp. 141–162. Orlando: Academic Press, 1986.

[9] M.A. Beckman and J. Edwards, “Lexical frequency effects on young children's imitative production,” in *Papers in Laboratory Phonology V*, M.Broe and J.Pierrehumbert, Eds., pp. 208–218. Cambridge UK: CUP, 2000.

[10] B.L. Derwing and W.J. Baker, “Rule learning and the English inflections,” in *Experimental Linguistics: Integration of Theories and Applications*, G.D. Prideaux, B.L. Derwing and W.J.Baker, Eds., pp. 248–272. Ghent: E. Story-Scientia, 1980

[11] C.T.Best, G.W. McRoberts & N.M. Sithole, “Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants,” *J. of Exp. Psychology*, 143, pp. 345-360, 1988.

[12] E. Flege, “The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification,” *Journal of Phonetics*, 15, pp. 47–65, 1987.

[13] D.B. Pisoni, S.E. Lively & J.S. Logan, “Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception,” in *The Development of Speech Perception*, J. C. Goodman & H.C. Nusbaum, Eds., pp. 121–166. Cambridge MA: MIT Press, 1994.

[14] L.Shockey “Perception of reduced forms by native and non-native speakers of English,” in *New Sounds 97*, J. Leather and A. James, Eds., pp. 314–329. Klagenfurt, 1997.

[15] L. Lisker and A. Abramson, “A cross-language study of voicing in initial stops: Acoustical measurements,” *Word,* 20, pp. 384–422, 1964.

[16] D. Recasens, M.D. Pallarès & J. Fontdevila, “A model of lingual coarticulation based on articulatory constraints,” *J. of the Acoustical Society of America,* 102.1, pp. 544–561, 1997.

[17] M.J.Solé “Phonetic and phonological processes: the case of nasalization,” *Language and Speech,* 35, pp. 29–43, 1992.

[18] J. Bybee, “The phonology of the lexicon: evidence from lexical diffusion,” in *Usage-Based Models of Language*, M.Barlow and S.Kemmer, Eds., pp. 65–85. Stanford: CSLI, 2000.

[19] J.J.Ohala “What's cognitive, what's not, in sound change,” in *Diachrony within Synchrony*, M. Morrissey and G. Hellermann, Eds.,pp. 309–355. Frankfurt: Peter Verlag, 1992.