On The Structure of Phonetic Categories

Jeri J. Jaeger & John J. Ohala
University of California, Berkeley

1. Introduction

In the earliest major statement of a phonetic feature system, Jakobson, Fant & Halle (1951) proposed features which were intended to capture the parameters used by all languages to make phonological distinctions. These features were set up as binary categories, because, it was argued, speakers actually encode and decode speech by making a series of binary decisions about the classification of speech sounds. Chomsky & Halle's later system also contained binary features for much the same reasons (Chomsky & Halle 1968). Ladefoged's 1971 system consisted of primarily binary features with a few multivalued features; he argued that some phonological oppositions (such as vowel height) are not binary. However, in a recent statement, Ladefoged (1980) presented strictly dimensional phonetic features, argued to capture the actual parameters used in speech perception and production, but not necessarily to reflect any kind of phonological or classificatory reality to the speaker or hearer. He further brought into question the whole issue of the psychological status of categorical phonological features, arguing that if they exist at all, they have little to do with these phonetic parameters used in speech production and perception.

These claims about the structure and status of phonological features raise a number of important questions. First, are speech sounds really organized into a categorical system which is in some sense psychologically real, or are speech sounds simply entities made up of some value on a number of phonetic scales or dimensions? Second, if there is such a featural classification system, is it strictly binary? Third, is binary classification really the natural way humans categorize things?

This third question may be the key to the other two. There has been a great deal of recent research examining the structure of natural categories and in particular semantic categories. This research has revealed that such categories have a complex internal structure, and are generally not binary. One way of examining the categorical nature of phonological features would be to use the tools developed for testing the nature of semantic categories. If it is found that phonetic featural categories behave much like semantic categories, then we would not only have useful evidence about how phonetic features are conceptually structured, but we would also feel more confident about claiming them to be psychologically real, in that they could be considered part of a general pattern of human categorization behavior.

In this paper, then, we will first discuss briefly the issues of binariness and review data on the organization of semantic categories. Second, we will present the findings of an experiment designed to look at the structure of phonetic featural categories. And finally we will discuss these results in terms of the issues of binariness and psychological reality raised above.

2. Binariness and the Structure of Semantic Categories

The notion of 'binariness' entails three basic components (see Garner

1970, 1978). The first is that all elements to be categorized must fit into two and only two categories for some particular parameter. The second is that this division is strictly categorical; all elements to be categorized must belong to either one category or its opposite; an element may not belong to both categories, or to neither. The third component we may call 'homogeneity'; this says that within each category every element is as conceptually good a member of that category as every other.

It is clear that the first two of these components are strictly held to by the Jakobson, Fant & Halle and Chomsky & Halle systems: every feature has only two possible values, and every segment for which a feature is relevant must have either a positive or negative value of that feature. The third component is implied by Jakobson's notion of 'bipolarity'; but because there is some variation at the phonetic level within featural categories, proponents of these systems might agree that categories are not strictly homogeneous at the phonetic level.

We will now turn to the question of the structure of semantic categories. Recent research, notably by Eleanor Rosch and her associates at Berkeley (see Rosch 1973, 1978) has demonstrated through a series of experiments that semantic categories have a complex internal structure. Such categories consist of prototypical members, surrounded by members of less and less typicality, with vague boundaries. Prototypical members are those which have the most typical combination of attributes which are associated with the category; they are the members which have the least number of attributes in common with prototypical members of opposing categories at the same taxonomic level of categorization. On the other hand, a peripheral member of one category can also be a member of other categories at the same level of taxonomic abstraction, depending on the context of the categorization. Since the make-up of each category is more or less determined by the cultural practices of the people using the concept, prototypicality will be to some extent determined by the function and importance of category members within the speech community.

Figure 1 illustrates the structure of the semantic category 'vehicle'. (These data were taken from Rosch 1973, and are based on ranking experiments done with American Ss; the numbers in parentheses are mean rank. Since Rosch did not actually list attributes, these have been abstracted from Ss' responses.) A car is clearly the most prototypical 'vehicle' for these Ss, while 'horse' and 'skis' are the least prototypical. 'Car' not only has the most typical confluence of attributes (such as 'it transports you', 'it is a machine', etc.), but also is undoubtedly the most frequently used vehicle for the Ss tested, which adds to its prototypicality. Note that the peripheral members of the category can be considered to be fairly good instances of other categories at the same level of abstraction as 'vehicle'; for example, 'horse' is a good example of 'animal'. But the prototypical vehicle 'car' is unlikely to be considered a good example of any other category. 'Horse' would most likely be considered a 'vehicle' in the context: 'How do I get from here to there?', but an 'animal' when it comes to breeding, for example.

From this example it can be seen that semantic categories are not binary, strictly categorical, or homogeneous. There is quite clearly a good deal of structure within each category, and individual elements can

Figure 1: The internal structure of the semantic category 'vehicle' (adapted from Rosch 1973).

Prototypical Members			Peripheral Members	Attributes
car (1.0)	boat (2.7) motor-scooter (2.5)	tricycle (3.5)	horse (5.9) skis (5.7)	it transports you it is a machine it has an engine it goes on land you get in it

be members of different categories at the same taxonomic level. Even natural categories which lend themselves to the analysis of having clear opposites more readily than 'vehicle' does (so would be better candidates for binariness) cannot be argued to be strictly binary. For example, with respect to the category 'male vs. female', which is about as binary as a natural category gets, there are certainly some people whose categorization would be unclear, for either biological or social reasons. So it can be seen that binariness is not an inherent property of natural semantic categories.

3. The Experiment

3.1. Design and Procedure

The present experiment was designed to look at the questions of the binariness and internal structure of phonetic featural categories (see Jaeger 1980 for details). The experimental design used was that of concept formation. This standard tool of psychological research is based on the assumption that previously existing or natural categories can be brought to the consciousness of subjects by careful training and feedback techniques, and that subjects should be able to demonstrate through their behavior something about the natural structure of these categories. Nonexistent categories, or unnatural categories, should be either extremely difficult or impossible for Ss to learn or to manipulate in any consistent way.

In the current experiment, the phonetic featural categories [anterior], [sonorant], and [voice] were explored. These features were chosen primarily because for each of them there is a phoneme whose category membership is controversial: first, there has been much argument about whether [w] should be [+anterior] or [-anterior], given the fact that it has constrictions at both the lips and the velum. Second, [h] is considered a sonorant in some feature systems, but an obstruent in others. Third, word-initial /b,d,g/ in English have questionable voicing status, given that they are usually phonetically voiceless, but are nevertheless considered part of voiced phonemes. The categories taught were [+anterior], [-anterior], [+sonorant], and [+voice]; the phonemes used in each part of the experiment are shown in Table 1.

There were 28 Ss, all linguistically naive speakers of American English between the ages of 18-39, mainly students and staff at U. C. Berkeley. Each S performed two categorization tasks: they were taught either [+anterior] or [-anterior] and either [+sonorant] or [+voice].

Table 1: Phonemes used in present experiment.

A. [anterior]

positive = /p,b,t,d,f,v,s,z,m,n/

negative = /ɔ,ʌ,k,g,ʃ,y,h/

control = /θ,ð/ [+anterior]

test = /w/

B. [sonorant]

positive = /r,m,n,w,y/

negative = /p,t,k,b,d,g,f,v,θ,ð,s,ʃ,tʃ,dʒ/

control = /l/ [+sonorant] /z/ [-sonorant]

test = /h/

C. [voice]

positive = /v,ð,z,m,n,r,w,y/

negative = /p,t,k,f,θ,s,h/

control = /l/ [+voice]
/ʃ/ [-voice]

test = /b,d,g/

In addition, 10 Ss participated in the [+voice] task as part of a separate experiment, making a total of 38 Ss.

The basic procedure was as follows. Ss were run individually and were given a full set of instructions before the experiment began. They sat in a sound treated room and listened over headphones to a list of tape-recorded words spoken by a male speaker of standard Midwest American English. The first segment in each word was either a positive or a negative instance of the featural category being taught. Ss were instructed to pay attention to the way the first sounds in the words were pronounced, and in the case of 'sonorant' and 'voice', to attend to the "overall sound quality". After each word, the same voice said 'yes' or 'no', indicating to the S whether the sound just heard had been part of the category or not. Ss were instructed that after they had heard several words and had some idea of what the correct category was, they were to begin responding 'yes' or 'no' to each word, before the feedback voice gave them the correct answer. Their responses were monitored by the experimentor, and if they responded to 15 tokens in a row with two or fewer errors, they were considered to have reached criterion, and were then given a test. (They were required to hear at least 25 words before the test began, to insure that they had heard the full range of exemplars.) The test contained three types of words: 1) Words with the segments taught as being positive or negative instances of the category, half of which continued to be reinforced. 2) Control tokens; these are clear positive or negative exemplars which had not yet been presented to the S; correct categorization of these control tokens was necessary before a S was judged to have formed the category correctly. 3) Test words, whose category membership was in question. After the test was over, the Ss were asked to name the category and discuss their decision-making procedures.

3.2. Results

The general results of the experiment are shown in Table 2. A few remarks can be made about Ss' overall performance; generally, some

Table 2: General results.

Category	#Ss	% Ss Formed Category	Average # Trials to Criterion (S.D.)
[+anterior]	14	50	45 (18)
[-anterior]	14	79	33 (15)
[+sonorant]	14	50	31 (14)
[+voice]	24	38	43 (23)

categories were easier to form than others, with [-anterior] having the highest success rate, and [+voice] being the most difficult. The most likely explanation for this is that it is easier for naive Ss to attend consciously to obvious articulatory gestures than to auditory properties of sounds or to less obvious vocal tract movements or configurations. The fact that only 38% of the Ss exposed to the [+voice] category could learn it, indicates that the movement of the vocal cords is extremely difficult for naive speakers to become aware of or to be objective about.

There are several places to look for information about the internal structure of the categories tested here. First, the errors Ss make in categorizing various tokens bears on the prototypicality question, in that it would be expected that there would be fewer errors in categorizing prototypical members, and more errors on peripheral members. Second, the categorization of the test words by individual Ss, and the agreement among Ss on the test words, gives evidence about the binariness question. Third, the names given to the categories by the Ss gives data about which attributes Ss consider to be criterial for category membership.

The results for [+anterior] and [-anterior] are shown in Tables 3 and 4.[1] Comparing the correct and incorrect responses to different members of these categories, it can be seen that the categories [+anterior] and [-anterior] clearly have internal structure, which is based primarily on the expected physical dimension of 'front of the mouth' vs. 'back of the mouth'. The labials and labio-dentals are the most prototypical of the [+anterior] category (that is, they received the most correct responses), while laryngeals, then low back vowels and velars are prototypical of the [-anterior] category. The large number of errors on palatals and alveolars shows that these are peripheral members of the categories; in fact, comments from Ss indicated that the palato-alveolar area was a very unnatural place to make a 'front-back' division of the vocal tract, and in particular, any feature which divides [s] from [ʃ] articulatorily is making a division that goes against speakers' intuitions.

In general each S made a mainly binary decision about the categorization of the test segment; however, different Ss made different decisions about which category the [w] should belong to, apparently determined by which category the S was taught, and which attributes he or she was attending to. For the [+anterior] category, 6 out of 7 Ss responded mainly positively to [w] (90% positive responses), while one S rejected it (100% negative). It appeared that when Ss were attending to the forward movement of lips or tongue they could easily detect the lip

Table 3 : Results for [+anterior][2]

	%Positive Responses	%Negative Responses
Positive Tokens	(correct)	
labials	89	9
labio-dentals	92	8
alveolars	73	25
	82	16
Negative Tokens		(correct)
palatals	21	79
velars	4	94
low back vowels	3	97
laryngeal	5	95
	9	91
Control Tokens		
[θ,ð] (positive)	83	17
Test Tokens		
[w]	77	23

Table 4: Results for [-anterior]

	%Positive Responses	%Negative Responses
Positive Tokens	(correct)	
laryngeal	100	0
low back vowels	94	4
velars	89	9
palatals	75	21
	88	10
Negative Tokens		(correct)
alveolars	6	90
labio-dentals	0	98
labials	4	95
	4	94
Control Tokens		
[θ,ð] (negative)	5	95
Test Tokens		
[w]	45	55

rounding involved in the pronunciation of [w]; the one S who rejected it reported feeling her tongue drawn back, and found this sensation more salient than the lip movement. The overall inclusion rate of [w] in the [+anterior] category was 77%. It can be argued from this that [w] is a good exemplar of the category [+anterior], as responses to [w] were similar to responses to phonemes taught as being members of the category.

For the [-anterior] category, 6 Ss responded mainly negatively to the test words (90% negative responses), 4 Ss responded positively (100% positive), and one S, reporting that she felt both constrictions, split her answers (40% positive, 60% negative); [w] had an overall inclusion rate of 45%. This indicates that even when Ss have their attention drawn to gestures made at the back of the oral cavity, the labial movement can continue to be more salient, perhaps because it is more kinesthetically obvious and easier to objectify.

Ss in the [+anterior] group named the category such things as: "forward vs. back of mouth, shut vs. open mouth, teeth and lips vs. gutteral", and "closed mouth vs. closed throat." The [-anterior] category was described as "sound comes from back of throat vs. front, eg. tongue and teeth; open vs. closed mouth", and "pushing up back or middle of tongue but not front." These characterizations of the categories indicated that the attributes involved included not only the 'front of the mouth vs. back of the mouth' dimension, but also an 'open mouth vs. closed mouth' factor, and an attribute based on which specific articulators and parts of articulators were involved.[3] Clearly, then, the categorization of [w] with respect to the [anterior] feature is ambiguous across the speech community, and each instance of categorizing it as one or the other was dependent on the specific context of each Ss' experiences, that is, which category they were taught, and which attributes they considered most criterial.

Table 5 shows the results for [+sonorant]. Although there was little variation in responses to the [+sonorant] phonemes (probably due to the small number of exemplar types), there was a great deal of structure apparent in the [-sonorant], or obstruent, category. Not surprisingly, voiceless stops were considered to be the prototypical obstruents, and voiced fricatives the most peripheral obstruents. The attributes involved included continuancy, lack of turbulence, an open vocal tract position, and, for some Ss, voicing. Ss named the category "soft, vowel-like, fluid, long, drawn-out vs. gruff, harsh, clearcut, hard, abrupt", and said "the correct ones started in the back of the throat, or had a rattling sound in the back of the throat". However, apparently not all Ss considered voicing necessary, as three out of seven Ss included the test segment [h] in the [+sonorant] category (83% positive responses) while four rejected it (100% negative responses), for an overall inclusion rate of 36%. Those Ss who included it were evidently more attuned to the 'open' or 'flowing' attributes, while those who rejected it considered it not 'soft' or 'vowel-like' enough. Again, the categorization of this speech sound is not strictly binary across the speech community.

The results for the [voice] experiment are shown in Table 6. The structure evident for both the [+voice] and the [-voice] categories shows that besides the actual presence or absence of voicing, continuancy

Table 5: Results for [+sonorant]

	%Positive Responses	%Negative Responses
Positive Tokens	(correct)	
glides	93	4
nasals	95	5
[r]	91	6
	93	5
Negative Tokens		(correct)
voiced fricatives	23	73
voiced affricate	17	83
voiceless affric.	13	87
voiceless fric.	10	87
voiced stop	9	88
voiceless stop	3	94
	11	86
Control Tokens		
[l] (positive)	86	14
[z] (negative)	4	93
Test Tokens		
[h]	36	64

Table 6: Results for [+voice]

	%Positive Responses	%Negative Responses
Positive Tokens	(correct)	
[r]	100	0
nasals	98	2
voiced fricatives	91	9
glides	75	25
	90	10
Negative tokens		(correct)
voiceless fric.	14	84
voiceless stops	1	99
	8	91
Control Tokens		
[l] (positive)	74	22
[ʃ] (negative	4	96
Test Tokens		
/b,d,g/	32	68

or duration was considered an important attribute in this category, since the prototypical voiceless sounds are the stops, and the most peripheral of the voiced sounds are the glides, which are inherently rather short. The test tokens, word initial /b,d,g/, were rejected by 6 Ss (91% negative responses) and included in the category by 2 Ss (89% positive responses); one S split her responses (56% positive, 44% negative). The overall percentages were 32% positive responses, 68% negative. Ss variously named the category "a vibration, a rumble, more throat, a soft sound that could be drawn out, vs. more air, more upper, real short." Although it is clear from these names the Ss gave to the category that voicing per se was indeed the most important attribute being attended to, it is not entirely clear from these results whether the test segments were rejected due to their actual lack of voicing or their short duration (or both). Nevertheless, the same pattern of indeterminancy of boundaries held for the voicing category.

3.3. Discussion

The main points developed here about the structure of these featural categories are illustrated in Figure 2. It is clear from the preceding discussion, and from these diagrams, that phonetic featural categories have an internal structure extremely similar to that of semantic categories; it can also be argued that the correct decision with regard to the categorization of some of the controversial segments, is to consider them members of both categories, or at least ambiguous as to category membership.

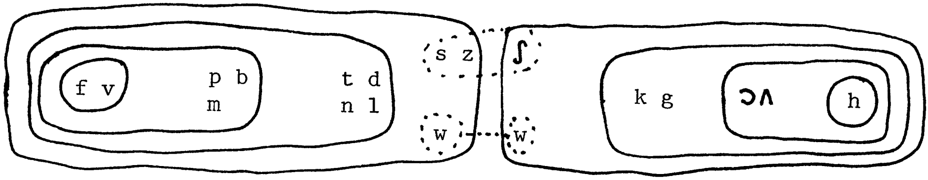
Let us now return to the questions raised earlier in this discussion, looking first at the question of binariness. One of the main problems with binary feature systems is that they force either/or decisions about every member of a phonological system. In the case of [w], Ohala and Lorentz (1977) have argued that it is not only phonetically inaccurate to call [w] primarily a labial or a velar, but it is also phonologically and cross-linguistically inaccurate, since in some languages [w] patterns phonologically with the labials, in other languages with the velars, and in some languages it patterns with both labials and velars in different phonological rules. Ohala and Lorentz present phonetic explanations for the specific patternings found. The present experiments have produced evidence that conteptually [w] may be both a labial and a velar, indicating that the feature [anterior] is not strictly binary. We can conclude, then, that in describing a particular language, a linguist may decide to call [w] [+anterior] or [-anterior] if it behaves as one or the other in that language, but this should not be taken either as a statement about the phonetic properties of [w], or as any sort of cross-linguistic generalization about [w]. In order to achieve phonetic, cross-linguistic, and conceptual accuracy, [w] must be specified as both [+anterior] and [-anterior]. The same basic arguments can be made for [h]; the question of /b,d,g/ is more complex, and will not be analyzed further here.

Secondly, what does this research have to say about the question of whether phonetic features exist as categories at all? It could be objected that since we were asking Ss to make categorical judgments in

Figure 2: The internal structure of phonetic categories.

1. [+anterior]

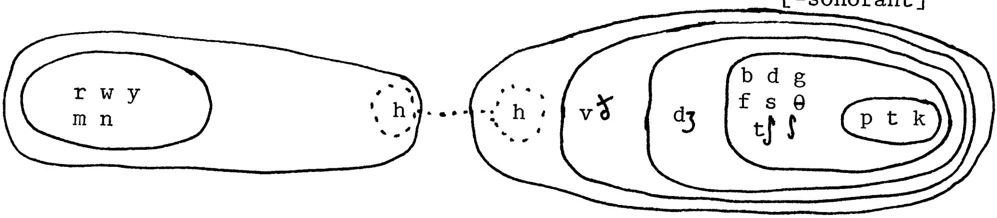
[-anterior]



Attributes: Activity in front vs. back of mouth. Closed vs. open mouth. Involvement of teeth, lips, and tongue tip vs. back of the tongue or laryngeal constriction.

2. [+sonorant]

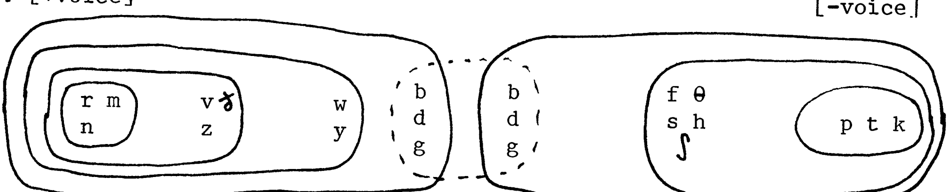
[-sonorant]



Attributes: Continuancy, lack of noise, softness, voicing, open vocal tract.

3. [+voice]

[-voice]



Attributes: Vibration in throat, continuancy, lack of noise.

this experiment, they naturally behaved categorically, as a function of the experimental design. However, we would argue that if these categories were non-existent or totally unnatural, Ss would not have been able to form them at all, and especially would not have shown such consistency in responses which revealed the inner structure of the category. Fortunately, the results of this experiment are not the only evidence we have as to the categorical nature of phonetic features. In the first place, one of the original motivating factors for the development of feature systems was the fact that phonological processes operate on classes of sounds which can be most readily explained in terms of such categorical systems. Further, Jaeger (1980) has reviewed data from perceptual confusions, speech errors, short term memory confusions, systematic misarticulations, similarity judgments, and language acquisition, and has found that data drawn from all these sources, which surely reflect speech production and perception processes as well as psychological processes, clearly point to the same sort of categorical behavior as was evident in the present experiment. In fact the categorizations inherent in speech errors and perceptual confusions led Goldstein (1980) to claim that "features describe groupings...in a multidimensional space that remain groupings, even when the space as a whole undergoes some severe transformation." We would argue, then, that a set of strictly dimensional phonetic parameters does not account for the observed behavior of speakers with regard to speech sounds any better than does a strictly binary classification system.

4. Conclusions

The above evidence and arguments support the following conclusions. Speech sounds are conceptually organized into a classification system based on their phonetic properties. This classification is neither strictly binary nor strictly categorical. Featural categories are not determined solely by a single phonetic attribute, but generally include a number of different attributes, some of which are more criterial than others; it is likely that different features covary in natural ways when they have attributes in common. And finally, phonetic featural categories have an internal structure similar to that of other natural categories. This leads us to argue that phonetic features are a categorical systematization that speakers and hearers impose on the otherwise non-discreet processes of speech production and perception, which allows them to reduce particular aspects of speech communication to an orderly, systematic, and manageable process. In this way phonetic features function to organize the world in exactly the same way that semantic categories do, and this similarity of function is reflected in the similarity of internal structure which has been discovered to exist in both semantic and phonetic categories.

Footnotes

1. The figures for the 'positive tokens' and 'negative tokens' include learning session responses made during and after reaching criterion, and all test session responses. Learning and test session responses were combined so that the largest possible number of tokens could be used for the comparison of responses to different phoneme types.

2. A possible third type of response was 'no response'; because of this, some of the positive plus negative percentages do not add up to 100%.

3. A likely reason why the labiodentals, rather than the labials, were the most typical of the [+anterior] category is that for some Ss involvement of teeth or lips was the most criterial attribute, and the labio-dentals involve both, so were clearly members of the category.

References

Chomsky, N. & M. Halle. 1968. The Sound Pattern of English. New York: Harper & Row.

Garner, W.R. 1970. The stimulus in information processing. American Psychologist 25:350-8.

Garner, W.R. 1978. Aspects of a stimulus: features, dimensions, and configurations. In: Rosch & Lloyd, 99-133.

Goldstein, L. 1980. Categorical features in speech perception and production. JASA 67:1336-1348.

Jakobson, R., G. Fant & M. Halle. 1951. Preliminaries to Speech Analysis; the Distinctive Features and their Correlates. Cambridge: MIT Press.

Jaeger, J.J. 1980. Categorization in Phonology: An Experimenal Approach. U.C. Berkeley Doctoral Dissertation.

Ladefoged, P. 1971. Preliminaries to Linguistic Phonetics. Univ. Chicago Press.

Ladefoged, P. 1980. What are linguistic sounds made of? Language 56:485-502.

Ohala, J. & J. Lorentz. 1977. The story of [w]. In: BLS 3, 577-599.

Rosch, E. 1973. On the internal structure of perceptual and semantic categories. In: T.E. Moore (ed.), Cognitive Development and the Acquisition of Language. New York: Academic, 111-144.

Rosch, E. 1978. Principles of categorization. In: Rosch & Lloyd, 27-48.

Rosch, E. & B.B. Lloyd (eds.) 1978. Cognition and Categorization. Hillsdale, N.J.: Lawrence Erlbaum Assoc. Pub.