"Distinctive phones" in surface representation*

D. Robert Ladd

An analogy to alphabetic handwriting suggests that it is a priori reasonable to view the relation between "phonology" and "phonetics" as involving a mapping from symbolic categories to continuous phonetic parameters. This in turn implies the existence of an interface representation consisting of categorically distinct symbolic elements. Recent evidence suggests that "systematic phonetic" transcriptions are best seen as an informal shorthand rather than a linguistically principled representation of speech; consequently, such transcriptions are not suitable for use as a formal "level of representation" at the phonology-phonetics interface. Instead, the most plausible interface representation – and still the most widely used, forty years after the emergence of generative phonology – appears to resemble a classical phonemic transcription. It is therefore appropriate to try to bring definitions of the classical phoneme into line with current knowledge of cognitive categories generally. Phonemes may be seen as language-specific phonetic categories, which need not invariably contrast with each other absolutely, but may exhibit relations of "partial similarity" or "quasi-contrast". Such cases could be treated as involving distinct subcategories of a single higher level category.

1. Phonetic abstraction and the phonology-phonetics interface

Every utterance is literally unique, different in physical detail from every other. Though this statement is clearly true, it is generally held to be irrelevant to linguistics, because many of the differences that are involved in the literal uniqueness of utterances are linguistically irrelevant, even imperceptible. Nevertheless, all of linguistics is ultimately based on being able to abstract away from the uniqueness of utterances and to identify certain speech phenomena as being "the same". This paper discusses the basis of that process of abstraction.

Specifically with regard to phonetics and phonology, the abstractions that many of us most commonly deal with are transcriptions. Transcriptions

are *symbolic abstractions,* representations to which operations of *discrete mathematics*, such as substitution and permutation, can apply. Any description of phonological phenomena that involves grammar-like operations on strings of transcription symbols is couched in terms of discrete mathematics. By comparison, the "gestures" of Browman and Goldstein's articulatory phonology (e.g. Browman and Goldstein 1986) are *quantitative* abstractions, which need to be dealt with in terms of *continuous mathematics*. Browman and Goldstein have been at pains to point out that, in a great many classic cases of assimilation and deletion, gestures are not actually deleted or substituted, only modified along continuous dimensions such as amplitude and duration.

Although *as a theory of phonology* Browman and Goldstein's ideas remain the preserve of a minority, virtually everyone agrees that at some level, scientific description of physical activity such as speaking must be expressed in quantitative terms. This leaves us with a problem. If speaking must be described in terms of continuous mathematics, then either we must banish symbolic abstractions and operations of discrete mathematics from phonology altogether, or we must accept that in some way the study of speech sounds involves an *interface* or a mapping of some sort between symbolic abstractions ("phonology") and continuous ones ("phonetics"). The first view is apparently that of Browman and Goldstein, and has been espoused in various recent work by e.g. Coleman and his colleagues (e.g. Coleman 1998), Pierrehumbert and her colleagues (e.g. Hay, Pierrehumbert and Beckman 2003), and others. The second view – the interface view – is conservative, in the sense that it dates back at least half a century to Joos (1948), and was central to the Sound Pattern of English (SPE) formalism (Chomsky and Halle 1968). Yet it is still current, and still seems defensible. For example, it is taken for granted in a paper aimed at mathematicians by András Kornai (1994). The idea of the interface is also at least part of the reason for Halle and Bromberger's insistence (1989) that "phonology is different".

In a separate paper (Ladd in preparation), I have argued the case for the conservative conception of the phonology-phonetics interface on the basis of an extended analogy to alphabetic handwriting. In written language we can generally distinguish clearly between analogues to phonology – spelling rules – and analogues to phonetics – physical realization as handwriting. Analogues to phonology include things like
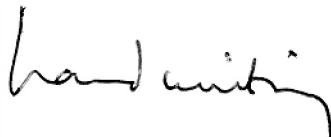
- *allographs* (e.g. the positionally-determined difference between <s> and <ſ> in older forms of the Roman alphabet);

- *complex segments* (e.g. Dutch <ij> as free variant of <y>);
- *alternations* (e.g. *happy/happier* or *carry/carried* in English);
- *post-lexical processes* (e.g. sentence-initial capitalization; e.g. German respelling of <ck> as <k-k> when a word is divided into syllables in choral scores or at the end of a line of type[1]).

Analogues to phonetics include things like coarticulation (subtle modifications of the shape of letters depending on the location of the connection to an immediately adjacent letter) and the trade-off between effort and intelligibility (the difference between fast sloppy writing and slow careful writing). More fundamentally, alphabetic handwriting demonstrates the essential plausibility of the conservative interface view: a handwritten word as a physical signal exhibits a lack of segmentability remarkably like what we see in speech, yet we know that it is appropriately idealized as a string of letters – that is, as a string of discrete elements chosen from a paradigmatic set.

The written language analogy thus gives considerable comfort to traditional phonologists. It makes clear that certain phenomena (like the alternation seen in *happy* and *happier*) involve operations on *representations*, which can and should be described independently of the act of writing. This makes it plausible that the analogous phenomena of spoken language phonology should be treated in the same way. For example, an abstract symbolic representation not only seems useful but may actually be necessary for describing things like the relation between *perfect* and *perfection* or *memory* and *memorial*. Proposals to do away with the phonology-phonetics interface have tended not to deal with those problems.

On the other hand, the written language analogy should make traditional phonologists fairly *un*comfortable in other ways. In particular, it makes clear that there are plenty of phenomena that are often treated in terms of operations on representations but which should actually be described as part of the physical realization of an abstract representation – as Browman and Goldstein have long maintained. For example, I may write the word *handwriting* as in (1):
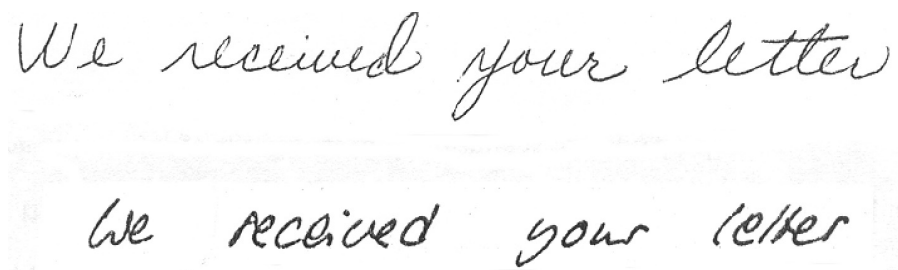
(1)

and in fact I commonly reduce *-ing* sequences to the dotted downward squiggle seen in (1), which is separated off in (2):

(2)

However, no one would say that the abstract representation of the word – the spelling – has changed. We would just say that the abstract sequence *-ing* is realized in a particular way. We would not describe (2) as a new letter, nor even think of it as a surface segment. And this fact gives less comfort to traditional non-laboratory phonologists, because it casts doubt on the legitimacy of symbolic descriptions for many phenomena of assimilation and reduction. It is clear that traditional descriptions of these phenomena – e.g. transcribing *ten past nine* as [ˌtʰɛmpæsˈnain] – treat the phonetic detail in a way that is quite comparable to treating (2) as a surface segment. Note, incidentally, that there is nothing about the physical nature of (2) that prevents it from being regarded as a segment of a written signal: any "systematic graphetics" of the writing systems of the world will have to allow for segments very much like (2), e.g. the Arabic letters ض and خ.

Moreover, the written language analogy suggests that we must give a quantitative rather than a "grammatical" treatment to many features of spoken language that are clearly language-specific and clearly communicatively significant. Consider the difference between these two written versions of the same English sentence.

(3)

Anyone with any experience of different national styles of handwriting will have no difficulty telling which of these was written by an American
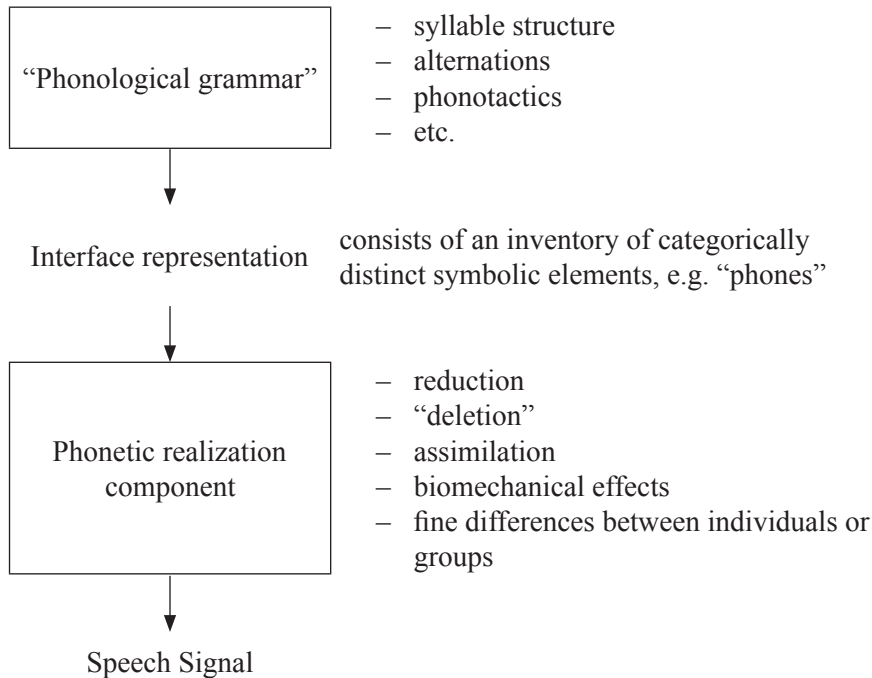
and which by a German. But with the possible exception of the form of the lower-case <r>, the interface representation is exactly the same in both cases – these are both realizations of the same string of letters. The difference lies in fine details of the way the elements of that string are realized. This may seem uncontroversial when applied to handwriting, but the implications of this analogy for spoken language are startling. The differences between "typical American handwriting" and "typical German handwriting" are precisely analogous to many social and regional differences of pronunciation – for example, the fine phonetic distinctions of vowel quality that British English puts to such vicious social use. This means that such distinctions should not be described in symbolic terms: just as we would not say that Germans and Americans use different letters, so we should not say that upper class and middle class speakers of Southern British English use different segments. Instead, we will simply say that, as a matter of group-specific realization rules, different groups produce "the same" letter or "the same" segment type in slightly different ways.

In short, if we accept the allegedly conservative phonology-phonetics interface idea, it seems to lead to the rather less conservative conclusion that *there is no appropriate symbolic or grammatical characterization of many meaningful language-specific differences.* For a variety of reasons, I believe that this conclusion is correct. We need to describe phonetic realization consistently in quantitative terms, and, fully pursuing the implications of the written language analogy, we need to treat as "phonetic realization" many phenomena that have in the past been treated using grammar-like operations of deletion, substitution, and reduction. We need to accept – really accept – Browman and Goldstein's claim that there is no appropriate segmental representation of the output of the phonetic component: we cannot say that in one production of the phrase *perfect memory* the [t] is realized and in another it is deleted, and nor can we say that some sort of intermediate category is produced. The validity of this point emerges clearly from a variety of laboratory phonology work over the past 15 years (e.g. Nolan 1992), and it is embodied in the notion of "language-specific phonetic rules" proposed by Keating (1985), Port and O'Dell (1985), Pierrehumbert (1990), Cohn (1993), and others.

The foregoing considerations give us a version of the interface view that can be roughly sketched as in (4). The architecture shown in (4) is compatible with quite a range of views. For example, it is silent on the question of whether the grammar is a set of procedural rules as in SPE and Lexical Phonology or whether it is some kind of constraint-satisfaction system like Op-

timality Theory. By the same token, the details of the quantitative realization model are not important to the overall picture, although I regard Browman and Goldstein's work as an eminently plausible candidate for such a model. I should also make clear that I take segmental "phones" as one of a number of possible types of abstract element at the interface level. That is, nothing in (4) is intended to rule out adding things like metrical structure or feature tiers or other kinds of non-segmental devices to the interface representation; my point is that the elements of the representation are *categorically distinct symbols*.[2] The principal claim diagrammed in (4) is that, in order to describe spoken language adequately, we need to think of phonetic realization as a mapping between a categorical symbolic representation and a quantitative physical signal. Given that premise, my concern here is with how we establish the elements of the representation in the first place, and what we think about the relation between those elements and the quantitative physical realities that they represent. Exploring these questions is the goal of this paper.

(4)   Phonology-phonetics interface:

| "Phonological grammar" | – syllable structure
– alternations
– phonotactics
– etc. |

↓

Interface representation — consists of an inventory of categorically distinct symbolic elements, e.g. "phones"

↓

| Phonetic realization component | – reduction
– "deletion"
– assimilation
– biomechanical effects
– fine differences between individuals or groups |

↓

Speech Signal

2. The interface representation and the classical phoneme

Traditionally in segmental phonology and phonetics there have been two main sources of authority for positing symbolic abstractions like [t] and [m] – for saying that two physical speech events which we know to be different in physical detail count as "the same". One is the authority of the expert phonetician: A and B are the same because the phonetician hears them as the same and transcribes them as the same. Roughly speaking, this is the basis of the abstraction we call a systematic phonetic transcription. The other kind of authority for positing symbolic abstractions is the authority of the native speaker: A and B are the same because the native speaker perceives no difference between them and/or uses them as if they were the same. This, of course, is the basis of the abstraction we call a taxonomic or classical phonemic transcription.

Systematic phonetic transcription is a useful shorthand for rough observations. Indeed, this seems to be the way the IPA alphabet was conceived of in its early days – as a substitute for quantitative description, *faute de mieux* (Joos 1948). Somewhere along the line, however, systematic phonetics became one of the big ideas of twentieth century phonology, promoted by Pike (1943), Abercrombie (e.g. 1967) and others (the definitive statement of this view seems to be Laver (1994)). The central idea of systematic phonetics is that there is a principled universal set of phonetic categories that abstract away from the infinite variability of the physics of speech. This in turn implies that it should be possible to give a symbolic or non-quantitative description that *fully characterizes* the linguistically significant detail of any utterance of any language. This assumption is the basis on which Chomsky (1964) characterized systematic phonetic transcription as a level of representation in his formalization of phonology: as has been pointed out by Pierrehumbert and Beckman (1988: ch. 1), the *SPE* model (Chomsky and Halle 1968) assumes that only universal biomechanical properties of speech need to be treated in terms of continuous mathematics. As we just saw, however, there is a great deal of recent laboratory research showing that many *language-specific* phonological phenomena can only be fully described in quantitative terms. Systematic phonetic transcription appears incapable in principle of serving as a formal level of representation of speech.

If we insist on the "informal shorthand" status of systematic phonetics, yet retain the notion that there is an interface representation, then the other traditional type of symbolic representation that suggests itself for our pur-
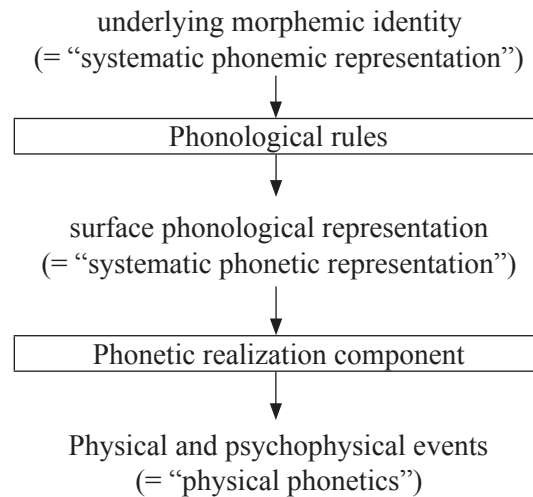
poses is the "classical phonemic" representation. On the face of it this does not look like a very promising idea. After all, the crushing critiques by Halle (1959) and Chomsky (1964) revealed fundamental theoretical problems with the classical phoneme, and their critiques have never been answered. But there are good reasons for considering this approach anyway.

The most obvious reason to take the classical phoneme seriously is that it still plays a large *de facto* role in phonological discussion, notwithstanding Halle and Chomsky. In practice – in speech therapy, in speech technology, in orthography design, in studies of language acquisition, and for that matter in most descriptive work within Optimality Theory – the phoneme notion continues to be applied as if there were no serious problem with it. For a theoretical construct that was discredited forty years ago, the classical phoneme is actually still doing pretty well. This has been true throughout the history of generative phonology. Only three years after the publication of SPE, Schane (1971) was already pointing out that whenever generative phonologists ignored so-called "low-level phonetic detail" – which was most of the time – their analyses generated a surface representation "almost amazingly identical to a classical phonemic representation (1971: 520)". At the very least, this means that many linguists of varying theoretical persuasions over many years have found an abstraction like the classical phonemic representation useful.
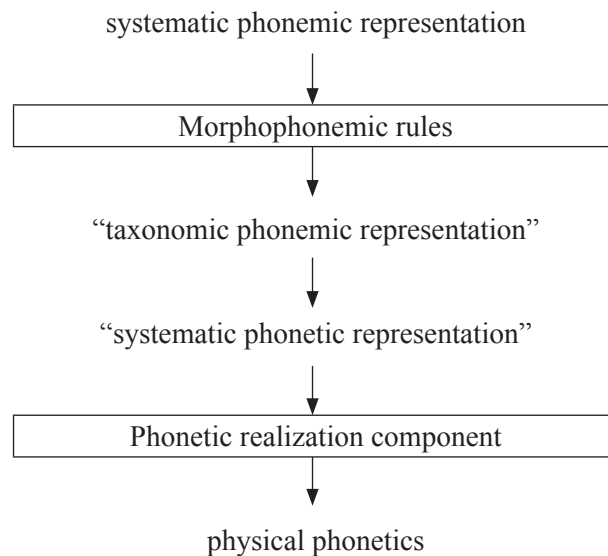
This being the case, let us go back and reexamine Halle's and Chomsky's evidence for discrediting the classical phoneme in the first place. If we do that, we see that there is an Achilles' Heel to their critiques – namely, they depend crucially on the idea that systematic phonetics is a level of representation. Note that SPE phonology assumes an interface view of phonology and phonetics that is more or less identical in overall structure to the view sketched in (4) above. Specifically, the SPE model, like the one in (4), assumes an abstract level showing the morphemic identity of phonologically conditioned allomorphs, which Chomsky called "systematic phonemic representation"; a concrete "physical phonetic" level that characterizes speech events in quantitative terms; and an intermediate level that expresses linguistically relevant phonetic distinctions in terms of abstract symbolic categories that can be manipulated by a grammar. The SPE version of (4) is shown in (5).

What Halle and Chomsky criticized about classical phonemic theory was that, in their view, it was actually a *four*-level model, with *two* representations intermediate between the morphophonological and the physical, as shown in (6).

(5)   Levels of phonological representation in SPE:

<div align="center">

underlying morphemic identity
(= "systematic phonemic representation")

↓

Phonological rules

↓

surface phonological representation
(= "systematic phonetic representation")

↓

Phonetic realization component

↓

Physical and psychophysical events
(= "physical phonetics")

</div>

(6)   Classical phonemic theory according to Chomsky 1964:

<div align="center">

systematic phonemic representation

↓

Morphophonemic rules

↓

"taxonomic phonemic representation"

↓

"systematic phonetic representation"

↓

Phonetic realization component

↓

physical phonetics

</div>

The fundamental problem that both Halle and Chomsky identified in classical phonemic theory, as represented in (6), was that it involves a map-

ping between one intermediate symbolic representation and another. Given the four-level picture in (6), Chomsky argued that various problems arise from positing two intermediate symbolic representations, and proposed to do away with the taxonomic phonemic level. Halle's famous argument based on Russian obstruent voicing assimilation (Halle 1959) reached the same conclusion on similar grounds.

However, the Halle-Chomsky argument works only if systematic phonetics is a formal "level of representation". If we go back to the original IPA conception of systematic phonetics as a shorthand for things that ought to be described quantitatively, and if we take seriously all the recent laboratory evidence about the non-categorical nature of assimilation and deletion and so forth, then the Halle-Chomsky argument – given the similarity of (5) to (4) – actually favors *retaining* the classical phonemic level. In effect, Halle and Chomsky were right to say that one of the intermediate representations in (6) is superfluous; with the benefit of forty years' hindsight, however, we can see that they eliminated the wrong one. I do not, of course, mean to suggest that we should simply equate the interface representation in (4) with a 1950s phonemic transcription and pick up where we left off half a century ago. My point is just that, if there is no systematic phonetics, then Halle's and Chomsky's arguments against the classical phoneme lose much of their force.[3]

3. On refining the phoneme

We may summarize the argument so far as follows. First, there are good reasons for an approach to phonology that involves an interface between a grammar-like component that deals in phonotactic structures and constraints, and a physical realization component that deals in quantitative parameters. Second, there are good reasons for suspecting that the interface representation is going to look something like a classical phonemic representation, i.e. based on an inventory of language-specific elements or categories. This means that we now need to consider how we might put the classical phoneme notion on a firmer empirical and theoretical basis.

3.1. Phonemes as phonetic categories

In my view, the key to refining the phoneme concept is to take seriously the idea of phonetic abstraction with which the paper began, and to see pho-

nemes primarily as *language-specific phonetic categories* – language-specific abstractions away from phonetic detail. These are the "distinctive phones" of my title. Phonemic distinctions are in the first instance about *sounding different* to the native speaker, not about signaling different messages. If we see things in this light, then we are in a position to deal with some of the classic theoretical problems of the phoneme on the basis of what we know about cognitive categories more generally.

Although as a matter of theory many adherents of the classical phoneme would presumably have endorsed the proposition that phonemes must "sound different" to native speakers of a language, in practice the basis of this proposition was the ability of sound substitutions to signal differences of *meaning*. The minimal pair test was the methodological gold standard of phonemic theory: if substituting one sound for another yields a different word or sentence, then the two sounds are deemed to sound different. In this respect, traditional phonemic theory has distinctly Whorfian overtones: the minimal pair procedure in defining phonemes suggests that lexical contrast will dictate which phonetic distinctions a speaker of the language will be capable of perceiving. There are obviously good reasons for assuming that lexical contrast is *relevant* to phonological systems, but that is no reason to accept the stronger view that differences of meaning (in some fairly narrow sense) define the status of differences of sound (in some fairly loose sense defined in practice by the IPA symbol system). The problems inherent in the stronger view were extensively discussed in the heyday of classical phonemic theory in the 1940s and 1950s; I outline them briefly in the following paragraph.

One problem is what we might call *allophonic awareness*. In cases of this sort, native speakers of a language consistently hear the difference between two phones that are supposed to be allophones of the same phoneme. The distinction between German *ich-Laut* and *ach-Laut* and pseudo-minimal pairs like *Kuchen* 'cake' vs. *Kuhchen* 'cow (diminutive)' is probably the classic example (Leopold 1948; MacFarland and Pierrehumbert 1991; Moulton 1947), but there are others. Another problem (which was actually rarely recognized as a problem until the advent of variationist sociolinguistics) is the notion of *free variation*. This is a problem because close inspection almost always finds that the choice of "free" variants is communicatively significant in some way. An example here would be the difference between apical and uvular /r/ in several languages of Western Europe, which never makes a lexical difference but certainly makes a sociolinguistic one, and which is recognized even in lay phonetic terminology. Still another long-standing problem

for classical phonemic theorists was the existence of *marginal phonemes*, for example the use of a velar fricative in the name *Bach* by speakers of English who in other respects "don't have a phoneme /x/". Do we count /x/ as a phoneme for those speakers or not? A more serious version of the same problem is what Fries and Pike (1949) referred to as *coexistent phonemic systems*. This problem is exemplified in a number of indigenous languages of Mexico, for example, whose native stop systems have allophonically determined voicing in stops but which have begun to acquire voicing contrasts as a result of massive borrowing from Spanish. Instances of these problems are widely attested in the phonology of virtually every well-studied language, but they have always been relegated to the status of interesting residual issues or have been attributed to incipient sound change or code switching, and they have never been integrated into any theory of surface representation.

These problems largely disappear if we see phonemes as phonetic types or categories, and if we assume that the formation of phonetic categories is a consequence of the whole language environment, not merely lexical contrast. Lexical contrast is obviously a major source of evidence for a language's phonetic taxonomy, but it is not the only one. All kinds of factors can play a role in sensitizing speakers of a language to differences between two phonetically distinct types: sociolinguistic distinctions, folk descriptions of sound types, exposure to neighboring languages or dialects, and paralinguistic usage of potentially lexically contrastive sounds (like the difference between an apical and a lateral click in English – the sounds of "tut-tutting" and of encouraging a horse, respectively).[4]

If we remove the obligatory lexical basis for phonetic categorization, we have no trouble accounting for the fact that babies begin to acquire the phonetic categories of the ambient language before they really have a lexicon. We have no trouble accommodating the otherwise surprising finding by Tees and Werker (1984) that English speakers who had been exposed to Hindi as babies but not subsequently were able to distinguish Hindi dental and retroflex stops as adults, even though they spoke no Hindi at all. We have no trouble understanding why English speakers are perfectly capable of distinguishing apical from lateral clicks but are likely to have trouble with the distinction between, say, voiced, nasal and aspirated lateral clicks, which are lexically distinctive in some languages.

The ideas just sketched are similar to those underlying Pierrehumbert's notion of a "fast phonological preprocessor" (Pierrehumbert 2002), which is intended to account for the categorical nature of many speech perception phenomena. In a paper that did not come to my attention until after I had

prepared a final version of this paper, Pierrehumbert (2003) suggests that the likely basis for such categorical preprocessing is the statistical distribution of phonetic tokens in the input available to a listener. Distributional facts will be influenced by lexical contrast, of course, but statistical distribution is primary; in fact, Pierrehumbert explicitly suggests that the phonetic categories of a given language correspond closely to the "positional allophones" of traditional phonemic theory. These are the "distinctive phones" proposed here.

3.2. Partial similarity between categories

In addition to widening the basis on which native speakers are assumed to establish their inventory of phonetic categories, we need to address the assumption that phonemic identity is sharply defined and the same in all contexts. This assumption was summed up in the heyday of the classical phoneme in the phrase "once a phoneme, always a phoneme". According to this doctrine, if two phones contrast lexically in any context, they must always be regarded as distinct phonemes. The implication of this principle is that either native speakers will treat two sounds as the same, in which case we are dealing with instances of a single phoneme, or they will treat them as different, in which case the two sounds are utterly and absolutely different in all contexts.[5] No allowance is made for sounds being somewhat different, or sometimes different and sometimes not, or different but still somehow also the same. Yet all these kinds of uncertainties can easily be found in the way real native speakers treat real sounds.

The cases I wish to discuss here are what we can call *partial similarity* between phonemes (or, looked at the other way, cases of *quasi-contrast* between phones). A good example involves the realizations of the /ai/ diphthong in Scottish Standard English (Scobbie, Turk and Hewlett 1999). Here we find minimal pairs like *side/sighed* and *tide/tied* that arise from the use of the longer open-syllable allophone (conventionally transcribed [ae]) in the morphologically complex forms *sighed* and *tied* and the shorter pre-voiced-stop allophone (conventionally transcribed [ʌi]) in *side* and *tide*. The uncertain phonemic status of the two variants of /ai/ can also be seen in the fact that in monomorphemic but disyllabic words like *spider* and *Bible* speakers may differ among themselves which allophone they use, and any given speaker may use one allophone in some such words and the other in others. Are we dealing with one phoneme or two? In my experience, first-year linguistics students who are native speakers of Scottish Standard

English are often puzzled by the status of the variants of the /ai/ diphthong when they are introduced to the phoneme concept. There are a number of other similar cases of partial similarity or quasi-contrast that are fairly well known, including: the American English *cot/caught* contrast; the East Coast American distinction between *can* 'be able to' and *can* 'metal container' (e.g. Bloch 1948: 20); the distinction between voiced stops with and without preceding nasal in Modern Greek (e.g. Arvaniti and Joseph 2000); the marginal status of the difference between long /e/ (orthographic *e, ee, eh*) and long /ɛ/ (orthographic *ä* or *äh*) in German; and of course, the distinction between German *ich-Laut* and *ach-Laut* discussed earlier.

Let us consider in detail one specific case of partial similarity, namely the mid vowels in French and Italian. According to a traditional phonemic analysis, both languages show a contrast between higher and lower mid vowels (in Italian /e – ɛ/ and /o – ɔ/; in French /e – ɛ/, /o – ɔ/, and /ø – œ/). Some minimal pairs are given in (7).

(7)   a.   Italian: [ˈpɛska] 'peach' vs. [ˈpeska] 'fishing'; [ˈfɔɾɔ] 'forum' vs. [ˈfoɾɔ] 'hole',
       b.   French: [ʒœn] 'young' vs. [ʒøn] 'fasting'; [etɛ] 'was' vs. [ete] 'summer'; [sɔt] 'stupid (fem.)' vs. [sot] 'jumps'.

However, this contrast applies only in lexically stressed syllables. In pretonic syllables (and, in Italian, also posttonic), the contrast between higher and lower mid vowels is neutralized, and the phonetic quality in those cases is variable or indeterminate.

Even in stressed syllables, though, there is a special relation of partial similarity between the higher and lower mid vowels. Somehow these vowels do not contrast with each other as completely as most other pairs of phonemes. Trubetzkoy discussed this with respect to French in *Grundzüge* (Trubetzkoy 1969:78):

> [The members of these oppositions] are often felt only as two meaning-differentiating nuances, that is, as two distinct yet closely related phonic entities.… From a purely phonetic point of view, the difference between French i and e is not greater than the difference between e and ɛ. But the closeness of the relationship between e and ɛ is apparent to any Frenchman, while in the case of i and e there can be no question of any particular closeness.

Trubetzkoy's observations about the psychological link between the mid vowels are astute, and I think they apply equally well to Italian.

Several factors provide evidence for the "particular closeness" ("*beson-dere Intimität*" in the original) that Trubetzkoy talks about. First, in both languages we find speakers who do not make a distinction between the members of the minimal pairs, or who make a distinction but make it in the opposite direction. Second, especially in French, there is also some degree of predictability from the phonetic environment whether we get the higher or lower vowel – we might call this quasi-complementary distribution. Third – and this partially contradicts the predictability of the quasi-complementary distribution – the lexical distribution of the higher and lower vowels is in some cases quite variable from speaker to speaker: some words consistently have one or the other, but many other words may have either. It seems especially telling that in both languages, manuals of good usage *written for native speakers* frequently devote space to the issue of which vowel is used in which words.

This whole situation seems to demand some comment. Rigorously applied classical phonemic theory would have nothing to say except "once a phoneme always a phoneme": the existence of even one minimal pair is enough to establish the existence of the contrast. One thing that seems to argue in favor of this position is the fact that native speakers of both languages seem to have no trouble distinguishing phonetically between the higher and lower variants. And yet at the same time native speakers are not supposed to need manuals of good usage to tell them which phoneme occurs in which word: no speaker of French or Italian is in any doubt about which words contain, say, /i/ rather than /e/. There is some special relationship between the higher and lower mid vowels that somehow manages to coexist with their phonemic distinctness.

Trubetzkoy's explanation of the special relationship between the French mid-vowels was based on the fact that the distinction between them is neutralized in lexically unstressed syllables. He distinguished between "constant" and "neutralizable" oppositions, and proposed that neutralizable oppositions would have a different psychological status. This can't be right. First, some of the cases of quasi-contrast, like the Scottish *side/sighed* cases, do not involve neutralization at all. Second and more important, if the neutralizability of the opposition were the explanation for these observations, then we would expect to find the same link in any case of neutralization. For example, we would expect speakers of American English to be aware of some special relationship between /t/ and /d/ because this distinction is neutralized intervocalically by so-called "flapping"; we would expect speakers of any language with final devoicing to be aware of some special relationship between the

members of voiced-voiceless phoneme pairs. I know of no evidence that this is true.

If we rule out neutralizability as the source of the "particular closeness" between the Romance mid vowels, can we find any other explanation? If pressed on this point, most phonologists would probably suggest that this is a case of sound change in progress: specifically, the pairs of higher and lower mid vowels are merging into single phonemes. Similarly, in the case of the Scottish Standard English /ai/ diphthong, we could say that the two allophones are splitting to form two separate phonemes. Yet there are obvious problems with such an explanation. First, the pace of these sound changes is glacial, apparently spanning many generations. Moreover, the situation for any given speaker is relatively clearly defined: the two sounds are different, and are used in specifiable ways, and at the same time exhibit a "particular closeness". The sound change explanation presupposes systemic instability, but in these cases the supposed instability of the system is remarkably stable. The distinction is consistently indeterminate.
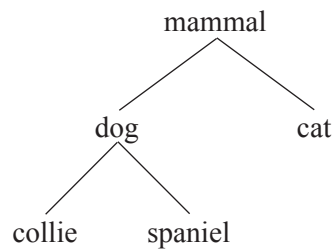
3.3. Categories and subcategories

Such phenomena of stable partial similarity or quasi-contrast can be accommodated in a theory of surface representations if we assume that, like any other system of cognitive categories, phonetic taxonomy can involve multiple levels of organization and/or meaningful within-category distinctions of various kinds. In other cognitive domains no one is surprised to find complex relations between categories, including notions like basic level categories (*cat, dog*), subcategories (*collie, spaniel*), and superordinate categories (*animal, mammal, quadruped*). I believe such notions can usefully be extended to phonetic categories as well.

Suppose we treat the classical allophone relationship as a category/subcategory relation, which we might diagram as in (8).

The diagram in (8b) shows two phonetic types (phones) as subcategories (allophones) of a single category (phoneme). In classical phonemic theory, this category-subcategory relationship can hold if and only if the difference between the two phones never gives rise to lexical contrast (either complementary distribution or free variation). What I suggest instead is that this relationship in (8b) can hold in a much greater variety of circumstances, even including cases where the difference between the two phones is sometimes lexically contrastive.
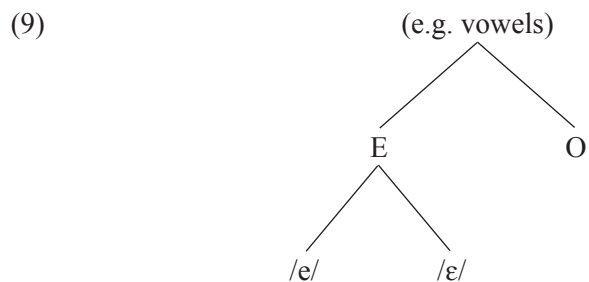
(8) a. Categories and subcategories generally:

```
                         mammal
                        /      \
                     dog        cat
                    /   \
                collie  spaniel
```

b. Allophonic variation as a category/subcategory relation:

```
                        phonemes
                        /       \
                     /A/         /B/
                    /   \
                 [a1]    [a2]
```
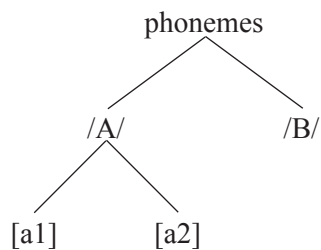
For example, the French and Italian mid-vowels could be seen as instances of just such a category-subcategory relation, as shown in (9):

(9)   (e.g. vowels)
                                    /   \
                                  E   O
                                /   \
                             /e/   /ɛ/

What classical phonemic theory says, in effect, is that native speakers are supposed to be aware only of the phoneme-level categories. In (8b), for example, they are supposed to be aware only of /A/, not [a1] and [a2]. Yet awareness of at least some allophonic differences is plainly a fact. If we think of the phoneme-allophone relation as a normal category-subcategory relation, this fact is not the least bit surprising. There is no paradox, for example, in the ability of speakers of English to regard collies and spaniels both as

instances of the category *dog* and at the same time as clearly distinct breeds. That is, there is nothing odd in assuming that we can perceive categorical distinctions at several levels. The higher and lower mid vowels of French and Italian can be seen as distinctive phones and at the same time as subtypes of a higher category.

The other point to make in this connection is that category-subcategory relations can be stable**.** Again, in classical phonemic theory, either two phones are in contrast or they are not; either two phonemes have merged or they haven't. This being the case, any observed quasi-contrast phenomenon (like the French mid-vowels, or the Scottish English *side/sighed* distinction) can only be seen as a transitory stage from one phonemic status to another. Yet nothing about the relation shown in (8b) necessarily implies instability or transition. There is nothing to preclude a prolonged period of several generations where two sound types are distinct yet linked together in a category-subcategory relation.[6]

This suggests that we need to adjust our expectations of what native speakers should be able to do. In several of these cases of partial similarity, it appears that native speakers are aware of the distinction between the two phonetic types, but at the same time they apply the distinction quite variably in the lexicon, because somehow the distinction is felt to be only a "meaning-differentiating nuance", in Trubetzkoy's phrase. This is something we want our interface representation, and our theory of phonetic categories, to express. More generally, we want our interface representation to be able to reflect the variety of statuses that sounds can have for native speakers. Just how complicated this may turn out to be is still not clear, but it is an obvious target for empirical work.

To give an idea of what such empirical work might involve, I close this section with a brief description of a recent small-scale survey I made of the distribution of the mid vowels in Italian, based on self-reports by 15 educated native speakers of Standard Italian. For both the front and back mid vowels I emailed the respondents separate lists of approximately 75 ordinary words, including nouns and adjectives, some inflected verb forms, several function words, and a handful of productive derivational suffixes. The front vowel list was sent first, followed a week or two later by the back vowel list. Respondents were simply asked to mark each word with *a* (for *aperto,* 'open') or *c* (for *chiuso,* 'closed'), according to their own usage. The responses showed a striking asymmetry between the front and back vowels: the front vowels show great inter-speaker disagreement, while for the back vowels the speakers mostly agree which words have the higher vowel and which have the

lower. Interpreting this finding via the traditional idea of phonemic contrast, we might suggest that the merger of /e/ and /ɛ/ is underway while the contrast between /o/ and /ɔ/ is still unaffected. But the native speakers I surveyed were unanimous that their intuitions were *clearer* about the front vowels than about the back vowels; that is, they found it easy to decide whether a word contained /e/ or /ɛ/, but hard to decide whether a word contained /o/ or /ɔ/, even though they gave markedly different responses in the case of the front vowels and highly consistent responses for the back vowels. An account based on classical phonemic theory and invoking incipient sound change would presumably predict that the consistent judgments should be easy and the variable ones difficult. I intend to follow up this preliminary work soon, with laboratory evidence on perception and production to supplement the self-reports.

4. Conclusion

One of the important theoretical works on phonology of the last two decades was Anderson's magisterial *Phonology in the Twentieth Century* (1985). Anderson's subtitle was *Theories of Rules and Theories of Representations,* and his central thesis was that twentieth century phonological theories could usefully be looked at according to how much descriptive work they got their rules to do and how much they got their representations to do. Classical phonemic theory, on this view, was very much a theory of representations, while SPE phonology was a theory of rules.

   Superficially, the arrival of Optimality Theory as the twentieth century drew to a close suggests that Anderson was not only premature in his choice of title, but more importantly wrong in his basic idea. The real dichotomy, it might appear, is between theories of rules and theories of constraints. Instead, I would argue that at a slightly deeper level Anderson was right: one of the things that characterizes twentieth century phonology is a dichotomy between "theories of representations" and "theories of things you can do with representations" (cf. also Goldsmith 1993). "Theories of representations" include (1) classical phonemic theory, (2) systematic phonetics, (3) distinctive feature theory from Trubetzkoy and Jakobson to SPE to Clements and Hume, and (4) theories of phonological structure like autosegmental and metrical phonology and their many offshoots. "Theories of things you can do with representations" are (1) Bloomfieldian process morphophonemics, (2) much of classical SPE phonology, and now of course (3) Optimality Theory.

What OT shows us is that you can argue about what kind of things your theory allows you to do with representations, but what the entire history of twentieth century phonology suggests is that *either way you have to have a theory of representations*. The goal of this paper has been to point out some of the phenomena that such a theory will have to take into account.

Notes

* The ideas in this paper have been developing over a long period of time and consequently owe much to many colleagues. I specifically thank Jim Scobbie for long discussions and fruitful collaboration (e.g. Ladd and Scobbie 2003) that have influenced my thinking in important ways. I also thank the members of the Cognitive Science Department at Johns Hopkins University, where I spent four months in 2000 and found both the freedom and the stimulating environment to develop this line of research.

1. This practice of respelling <ck> at syllable breaks does not appear to be sanctioned by the new German orthography, but was well-established for several decades and can still sometimes be seen even in materials printed according to the new orthographic rules.

2. At least two very disparate recent lines of work (e.g. Atterer and Ladd 2004; Gafos 2002; Ladd, Mennen and Schepman 2000) have suggested that the fine temporal coordination of gesturally distinct aspects of the speech signal is phonologically significant and/or phonologically controlled. Gafos argues that this necessitates reference to quantitative abstractions in the phonological grammar; Ladd et al.'s very much more limited theoretical point concerns the nature of the phonological entities that might be relevant to the autosegmental association of intonation with the segmental structure. Here I assume that, contra Gafos, a suitably enriched notion of autosegmental association can allow us to describe these phenomena in an abstract symbolic representation that keeps the interface distinct from the actual coordination of gestures in phonetic implementation. However, I acknowledge that future work along these lines may make the interface view untenable.

3. When I presented this paper at the conference, Steve Anderson (personal communication) suggested to me that the essence of the classical phoneme – and the focus of Chomsky's critique – is the biunique correspondence between the phonemic and the systematic phonetic levels. He argued that it therefore makes no sense to talk about refining the classical phoneme if I am not also rehabilitating biuniqueness (which of course I am not, since Chomsky's critique of biuniqueness seems beyond debate, and a fortiori since I am questioning the formal status of the systematic phonetic level altogether). However, I see my proposal here as refining the basic intuition behind the classical phoneme concept, not reviving

the entire theoretical superstructure of the 1940s and 1950s. I see no contradiction in trying to profit from the intuition while at the same time getting rid of formal theoretical ideas that are manifestly based on an outdated and empirically unsatisfactory understanding of phonetics.

4. This proposition can be tested empirically. For example, one might attempt to demonstrate the existence of different perceptual boundaries in the general phonetic area of dorsal fricatives in speakers of English, Dutch, and German. If sociolinguistic significance, allophonic awareness, and exposure to other dialects and languages are relevant, one might make the following predictions. (1) English speakers will have a single category "dorsal fricative", with no obvious perceptual discontinuities; (2) German speakers will show a relatively sharp boundary between two categories corresponding to *ich-Laut* and *ach-Laut,* with the boundary relatively far forward (e.g. between [x] and [ç]); (3) Dutch speakers will have a sharp boundary between two categories, but in a different place from the boundary in German, based on the sociolinguistic distinction between so-called "soft G" (roughly /x/) and "hard G" (roughly /χ/). Note that none of those languages has any lexical contrasts among dorsal fricatives at all.

5. Obviously, this characterization of classical phonemic theory makes no allowance for neutralization. Some versions of the theory did attempt to deal with neutralization (e.g. the notion of the "archiphoneme"), but it seems fair to say that neutralization and related phenomena posed a fundamental problem for the theory. See further the discussion of examples (7a) and (7b).

6. If we allow a richer set of relations between phonetic categories, we may have a natural mechanism for splits and mergers, through the dynamics of language acquisition. We can see how it might be possible for children to acquire a subtly different grammar from that of their parents without any superficially obvious difference in language behavior. In a phonemic split, the parent would have two contrastive phones that are sub-phonemes of a single category, and the child would acquire and reproduce the categorical difference between the two contrastive phones without also acquiring the superordinate category. In a merger, the parent would have two distinct phonemes, which the child would acquire as distinct phones but would regard as sub-categories of a single higher category that the parent did not have. Again, there would be no immediately observable behavioral consequence. If subsequent generations acquired the two as sub-phonemes, this could set the stage for a more radical loss of distinction; at the same time, it would also leave open the possibility that the merger might reverse itself. The historical reversal of splits and mergers is otherwise quite mysterious.

References

Abercrombie, David
    1967   *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
Anderson, Stephen R.
    1985   *Phonology in the Twentieth Century*. Chicago: University of Chicago Press.
Arvaniti, Amalia and Brian D. Joseph
    2000   Variation in voiced stop prenasalisation in Greek. *Glossologia* 11–12: 131–166.
Atterer, Michaela and D. R. Ladd
    2004   On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32: 177–197.
Bloch, Bernard
    1948   A set of postulates for phonemic analysis. *Language* 24: 3–46.
Browman, Catherine P. and Louis M. Goldstein
    1986   Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.
Bybee, Joan
    2001   *Phonology and Language Use*. Cambridge: Cambridge University Press.
Chomsky, Noam
    1964   The nature of structural descriptions. In: *Current Issues in Linguistic Theory*. The Hague: Mouton.
Chomsky, Noam and Morris Halle
    1968   *The Sound Pattern of English*. New York: Harper and Row.
Cohn, Abigail
    1993   Nasalization in English: phonology or phonetics. *Phonology* 10: 43–81.
Coleman, John S.
    1998   Cognitive reality and the phonological lexicon: a review. *Journal of Neurolinguistics* 11: 295–320.
Fries, Charles C., and Kenneth L. Pike
    1949   Coexistent phonemic systems. *Language* 25: 29–50.
Gafos, Adamantios
    2002   A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20: 269–337.
Goldsmith, John A.
    1993   Introduction. In: John Goldsmith (ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations*, 1–20. Chicago: University of Chicago Press.

Halle, Morris
   1959   *The Sound Pattern of Russian*. The Hague: Mouton.
Halle, Morris and Sylvain Bromberger
   1989   Why phonology is different. *Linguistic Inquiry* 20: 5–70.
Hay, Jennifer, Janet Pierrehumbert and Mary E. Beckman
   2003   Speech perception, well-formedness, and the statistics of the lexicon.
             In: John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in
             Laboratory Phonology VI*. Cambridge: Cambridge University Press.
Joos, Martin
   1948   *Acoustic Phonetics (Language monograph no. 23)*. Baltimore: Waver-
             ley Press.
Keating, Patricia
   1985   Universal phonetics and the organization of grammars. In: Victoria
             Fromkin (ed.), *Phonetic Linguistics*, 115–132. New York: Academic
             Press.
Kornai, András
   1994   Relating phonetic and phonological categories. In: E. S. Ristad (ed.),
             Language Computations, Volume 17, 21–36. Philadelphia: American
             Mathematical Society.
Ladd, D. R.
   in prep. Could there be a phonology-phonetics interface?
Ladd, D. R., Ineke Mennen and Astrid Schepman
   2000   Phonological conditioning of peak alignment of rising pitch accents
             in Dutch. *Journal of the Acoustical Society of America* 107: 2685–
             2696.
Ladd, D. R. and James M. Scobbie
   2003   External sandhi as gestural overlap? Counterevidence from Sardinian.
             In: John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in
             Laboratory Phonology VI*. Cambridge: Cambridge University Press.
Laver, John
   1994   *Principles of Phonetics*. Cambridge: Cambridge University Press.
Leopold, Werner
   1948   German ch. *Language* 24: 179–180.
MacFarland, Talke and Janet Pierrehumbert
   1991   On *ich*-laut, *ach*-laut, and structure preservation. *Phonology* 8: 171–
             180.
Moulton, William G.
   1947   Juncture in Modern Standard German. *Language* 23: 212–226.
Nolan, Francis
   1992   The descriptive role of segments: evidence from assimiliation. In: G.
             J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology
             II*, 261–280. Cambridge: Cambridge University Press.

Pierrehumbert, Janet

  1990   Phonological and phonetic representation. *Journal of Phonetics* 28: 375–394.

  2002   Word-specific phonetics. In: Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology VII*. Berlin/New York: Mouton de Gruyter.

  2003   Probabilistic Phonology: Discrimination and Robustness. In Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.): *Probabilistic Linguistics*, 177–228. Cambridge, MA: MIT Press.

Pierrehumbert, Janet and Mary E. Beckman

  1988   *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pike, Kenneth L.

  1943   *Phonetics*. Ann Arbor: University of Michigan Press.

Port, Robert F. and Michael L. O'Dell

  1985   Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455–471.

Schane, Sanford A.

  1971   The phoneme revisited. *Language* 47: 503–521.

Scobbie, James M., Alice E. Turk and Nigel Hewlett

  1999   Morphemes, phonetics and lexical items: The case of the Scottish Vowel Length Rule. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, Volume 2, 1617–1620. San Francisco, CA.

Tees, Richard and Janet Werker

  1984   Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38: 579–590.

Trubetzkoy, N. S.

  1969   *Principles of Phonology (translated by Christiane A. M. Baltaxe)*. Berkeley: University of California Press.