

## Phonemes, Contrast, and Phonetic Detail

### Why were phonemes invented? (is there a thing in reality that the concept ‘phoneme’ is about?)

The motivation for the concept “phoneme” is related to the theme of the first chapter of the book, namely the problem of representing language sound, for example in a grammar or scholarly article. “Representing” implies something different from “recording” in a physically-faithful way, in that a representation reduces an object to symbols of some kind. In particular, “representing” means changing a thing from a continuum to a discrete category.

For millenia, people would use whatever writing system was conventionally used for the language when talking about a language, and readers were required to know the meanings of letters in the grammar. Thus a grammar of Hebrew written in 1528 would use the Hebrew alphabet which looks something like [חֲתָרְכָּ לְ סִמְסִמָּ] (that is not real Hebrew, b.t.w.), and a grammar of Norwegian written in 1821 could contain data looking something like [har bøndene kjøpte pær?]. This orthography-bound practice makes it impossible to report facts about an unwritten language such as Shona, which didn’t have writing conventions. Linguists took to using letters based on the spelling system of languages they knew, in order to write down a new language that they ran into, so if you’re Danish or Norwegian and you hear a language with a vowel that sounds like the one in *rød* then you might write <ø>; if you’re Finnish, Swedish, German, Turkish, or Hungarian, you may write <ö>; perhaps you would write <eu> if French is your native language.

Some people objected, perhaps rightly so, that this amounted to defining unwritten languages in terms of the major European languages, rather than defining them on their own terms. That is, by this method, the sounds of Shona could only be understood by reference to the sounds of English, German or Italian. Especially under the influence of Franz Boas, American linguistic field workers took to using systems of analysis which they took to be defined “in terms of the language itself”, rather than by reference to some other language. The practical problem with explaining sounds of one language in terms of sounds of another (better-known) language is that the unknown-language sounds are often only somewhat like the better-known sounds, meaning that real differences in language sounds could not be represented by reference to the major Indo-European languages. This era coincided with the frenzied birth of logical positivism and behaviorism, so great emphasis was placed on developing objective, mechanical criteria for analysis, and eliminating any bias introduced by the analyst.

A major stumbling block to the program of objective scientific analysis of language is how to render speech symbolically. Recording technology in those days was nearly non-existent, and was also almost completely useless for communicating information about a language to others. There are a few archived (phonographic) recordings of language materials from the early era of fieldwork — most are of such poor quality as to be unusable for serious acoustic work, and in their day had no scientific use, because they could not be published or (for economic reasons) copied and distributed to other scientists. The only practical solution was to describe sounds via an experiential approximation, based on inferred physiological states, for example “[φ] is a bilabial voiceless frictionless spirant”. The viability of that system (then, as now) depends on spe-

cific experiences, where you hear examples of breathy vowels like [a̤] — performed by a trained phonetician with knowledge of vocal tract anatomy, one such as Daniel Jones — and if you later hear a similar vowel in a language like Dinka, you know to write [a̤]. The symbol used to indicate sounds are arbitrary, but by convention they are given relatively specific meanings that ideally do not vary across languages.

The technology for studying speech physiology has improved over the past century, but it still remains tricky enough that, even now, it is rarely used in studying unknown languages, is generally used in very limited doses (in order to answer one very specific question), and is primarily carried out only by professional phoneticians and not general descriptive linguists and fieldworkers. Lots of information can be obtained about the action of the articulators via introspection about tactile sensations in the vocal tract, knowledge of anatomy, and a few tricks involving toothpicks. Very low-tech introspection about what’s moving around the lips, coupled with visual information, shows that [o] involves rounding the lips. The implicit (and reasonable) assumption is that one can figure out a probable vocal tract configuration that results in a particular sound if you have adequate training in this technique, and therefore you can convey some objective information about the sound in terms of how it is physically produced, even if the descriptions are somewhat prone to error.<sup>1</sup> Fieldworkers are well aware of the quasi-subjective, “art” nature of impressionistic transcriptions. Phonetic transcriptions were thus at odds with the desideratum of total objectivity, since they almost always involve personal introspection about a present event (hearing a native speaker of language utter a given word) and its relation to your previous experiences (your training in field phonetics). It is not well known how actual individual field workers of the period coped with this philosophical problem — most probably, the field worker did not care that much about mechanical reduction of speech sounds. Narrow phonetic transcriptions were considered to be convenient fictions which ‘stand in’ for actual speech, necessary because there was no alternative method of discussing actual speech.<sup>2</sup>

Speech-analysis tools have since the late 1990’s become available in useful form to people doing language description, and correspondingly there has been some increase in the practice of giving acoustic data for languages when they are being described. Nevertheless, it still remains impossible to learn the grammatical structure of a language, or indeed to investigate any question scientifically, by just looking at an acoustic waveform. A scientific question needs to be posed as a testable hypothesis, which implies that variables can be controlled and the test can be repeated, to see if you get the same results whenever you perform the test. This means that there has to be a classification of data into groups that are “same” vs. “different” — even though the items in the “same” group are literally not “the same”. For instance, you could look at the duration of the stop consonant in a hypothetical word [lintu] “bird” pronounced many times by one speaker, and compute a mean duration. These multiple pronunciations would be “the same” in

---

<sup>1</sup> This approach tends to fall flat on its face when applied to utterly unfamiliar sounds, such as creaky or breathy vowels, or “micro-variations” such as the acoustic and articulatory differences between [ð] in English or Kikamba versus Danish or Hawrami, when one does *not* know the physiological difference between the two kinds of [ð].

<sup>2</sup> The sound spectrograph was invented in 1944, thus measurements of acoustic properties of speech could not solve this problem in the era when descriptive linguistic practices were established, between the end of the 19th century and WWII. With the invention of the spectrograph, efforts were made to identify the intangible reality that a transcription stands for.

the sense that that are the same word, pronounced by the same person, and we don't expect there to be any important differences in pronunciation depending on whether the word was pronounced at 10:32 AM on September 21, 2004 versus 11:00 AM on October 1, 2004.

This “word comparison” method assumes an extremely concrete view of language sound, one lacking the ability to equate the stop of [lintu] with the stop “t” of a similar hypothetical form [flintu] “gun”. Given a collection of multiple pronunciations of the same word, we could presume (though not uncontroversially) that the *t*-like thing in each pronunciation of [lintu] is a token of a particular “type”, namely “the *t* of the word ‘bird’.” But what justifies making broader comparisons between the *t*-like thing of [lintu] and that of [flintu], calling them “the same”, while not including the *s*-like thing of hypothetical [lisu] “trip” in that same group? Surely we would not want to rely on spelling, especially if we were dealing with an unwritten language. In addition, this word-based approach fails to distinguish systematic variations in word-pronunciation dependent on the sentential context where a word appears: an example would be the difference between the final “t” of English [sɛt] “set” versus the “t” of the same word in a phrase like [sɛr ɔskɹ] “set Oscar”. The very short answer is that the “t” of [lintu] and the “t” of [flintu] are the same “phoneme”, and the “s” of [lisu] is a separate “phoneme”; the “t” of [sɛt] and [sɛr ɔskɹ] are also the same “phoneme” but are physically distinct rule-governed “allophones”. This immediately leads to some basic questions: how does one know that in some instance “t” and “d” are separate phonemes, and “t” and “r” are allophones of a single phoneme? what proof is there that these sound relations are actually different?

The invention of the term ‘phoneme’ is generally credited to Baudoin de Courtenay (1895). Concepts that amount to what we now call the phoneme have been attributed to the Indic grammarian Patanjali (appr. 3rd century BC), the anonymous author of *Fyrsta Málfrøðiritgerðin* (“The First Grammatical Treatise”) which is a treatise on Icelandic written around 1150, and to King Sejong, the inventor of the Korean alphabet in the 15th century. The conceptual content of this term has been the subject of tremendous controversy, but all versions have in common the goal of expressing the idea “the same sound” as opposed to “a different sound”. Sapir viewed the phoneme as a mental entity, stating (1925) “This is the inner configuration of the sound system of a language, the intuitive ‘placing’ of the sounds with reference to one another”. In contrast, Twaddell (1935) rejected this mentalist definition of the phoneme, stating “Such a definition is invalid because (i) we have no right to guess about the linguistic workings of an inaccessible ‘mind’...”. The operational versus psychological characterizations of the phoneme have played a major role in phonological theory.

### **The phoneme: taxonomic version.**

The conception of “phoneme” which played the most important (negative) role in the development of contemporary phonology is that of the so-called “taxonomic phoneme”, the view prevalent in American structuralist linguistics, that is, the position held by Twaddell and colleagues, prior to the 1960’s. The historical importance of this conception of phoneme is that it was the target of attacks by Halle (1959), Chomsky (1964), Chomsky & Halle (1965) and Postal (1968), inter alii, in works which established the basis for generative phonology. The taxonomists viewed the phoneme as a convenient analytical tool for organising data, with no other significance. As behaviorists, they attributed no mental status to phonemes (indeed, the concept ‘mind’

was outside the purview of behaviorism). In that school of linguistics, special emphasis is put on analytic procedures and formal analysis (befitting its roots in logical positivism, which swept science in the 1930's). Under this view, emphasis was placed on defining analytic levels, stating an alphabet of symbols to be used at a given level of analysis, and stating the mathematical relation between levels. It was assumed that, given technological progress, a phonemic analysis could be arrived at automatically from a large enough corpus. Bloch (1948, fn.7) states "Given a sufficient sample of the dialect — say twenty or thirty hours of connected speech by an informant, recorded either on a high-fidelity machine or in a minutely accurate phonetic transcription — a linguist could probably work out the phonemic system without knowing what any part of the sample meant, or even whether any two parts meant the same things of different things".

The lowest level of analysis for the taxonomist is the level of the 'phone': this object, which one finds aplenty in narrow phonetic transcriptions, involves minimal analysis of a physical sound wave. A given 'phone' is a symbol standing for a particular class of physical sounds that are "the same" within the limits of human perception.<sup>3</sup> The important step of reducing continuous sound to a string of symbols is, essentially, presupposed via basic ear-training. A 'phoneme' is a separate type of symbolic object which stands for a particular class of phones encountered in a language, following certain analytic rules. Strict separation of analytical levels was a primary principle of the theory, which was necessitated by the need to avoid logical circularity while maintaining (the air of) scientific rigor. In terms of what we recognise as being phonology in some sense, the three taxonomic levels of representation are:

Phonetic level, containing phones	[p <sup>h</sup> lɪŋks]
Phonemic level, containing phonemes	/plinks/
Morph(ophon)emic level, containing morphophonemes	pliNk-Z

Each level maps to the adjacent level by unordered correspondence rules, such as a version of the English aspiration rule /p/ → [p<sup>h</sup>] / . \_\_ meaning "a /p/ phoneme corresponds to a [p<sup>h</sup>] phone in syllable-initial position" (which can also be understood as "a [p<sup>h</sup>] phone corresponds to a /p/ phoneme in syllable-initial position"). The alphabet of each level is logically autonomous, and thus the letter <p> at the phonemic level is not the same entity as the letter <p> at the phonetic level, even though they aren't visually distinguished. Elements at the phonemic level have no phonetic value: "phonetic value" is a property of phonetic representations. In principle, this distinction could be reinforced by selecting entirely different symbols for phones and allophones, e.g. the /p/ phoneme and the [π] phone, but typographic considerations make that a very impractical proposal, therefore the burden of differentiation is shifted to bracketing — the /p/ phoneme and the [p] phone.

'Phonemes' are objects at the second most concrete level of analysis, and are seen as being a set of phones. Thus Hockett (1942) says that "A phoneme is a class of phones determined by six criteria", using a logician's term "class" that is similar to the concept "set". A phoneme is a label *standing for* or *organizing* a set of phones, and the specific member of the set which the

---

<sup>3</sup> The ontology of this "sameness" is not generally discussed, and need not be understood in psychological terms such as by reference to perception.

phoneme maps to is determined by the context in which it occurs (expressed via context-sensitive *phonological rules*). The relation between phonemes and phones is (for some analysts) said to be subject to the formal condition of *biuniqueness* (Harris 1944), which guarantees free interchangeability of letters across levels. This condition requires that any phoneme in a given context be represented by a unique phone, and requires that a given phone be a member of exactly one phoneme: phonemes must not intersect (Bloch 1948).<sup>4</sup> This requirement on phonemic analysis guarantees that you can correctly predict the pronunciation of a word from the phonemic transcription, aided with knowledge of the allophonic rules in the language. And, just based on hearing a sound and rendering it correctly on paper in terms of phones, and knowing the mapping between phones and phonemes in a language, you can determine what phoneme it is a member of. Accordingly, the phoneme /t/ in English in word-initial position always maps to an aspirated [t<sup>h</sup>], and the phone [t<sup>h</sup>] always maps to the phoneme /t/.

An analytical litmus test for determining phonemic status, then as now, is the minimal pair. Thus [p] and [p<sup>h</sup>] are not distinct phonemes in English: the contexts where [p] appears are the complement of those where [p<sup>h</sup>] appears; therefore, the phones [p] and [p<sup>h</sup>] are members of one phoneme, /p/ (and [p], [p<sup>h</sup>] are the allophones of /p/). The pair /p/ and /b/ are separate phonemes, as shown by minimal pairs such as /pɪl/ vs. /bɪl/, distinguished only by selection of /p/ versus /b/. The reason why minimal pairs are absolutely probative of phonemic status is that they hold all variables constant, and present a distinction between words signaled by just one phone versus another — for instance English [bɪt] ‘bit’ versus [bɪd] ‘bid’.<sup>5</sup> Such examples make a knock-down argument for contrast (hence different-phoneme status) because if they did involve the same *phoneme*, there can (definitionally) only be one *phone* that appears in that context (they are not in free variation because selection of [p] versus [b] is not free, but instead indicates selection of a specific word). If two phones ever appear in the same environment, they cannot be allophones of the same phoneme, since their distribution is not complementary (there is not a 1-to-1 mapping between phoneme and phone in the given context). Instead, such words illustrate the property of *contrast* (and, allophones of the same phoneme by definition do not contrast).

If minimal pairs are not available, a plausible but non-probative argument can be made based on near-minimal pairs. For example, if the closest you could get to minimal pairs for [p] vs. [b] in Gwambomambo is the collection of words:

[palk] [apo] [kepsi] [gilap] [uport] [poit]  
[bark] [abe] [kebši] [gülab] [ubork] [bout]

<sup>4</sup> An unpleasant complication is introduced by free variation, where a given phoneme can be pronounced freely in any of two or more ways, in a given environment. When you have free variants, for example if [s] and [θ] are interchangeable, the principle is that you can always select either one phone or the other, and that the choice of one versus the other never signals a different word.

<sup>5</sup> This presupposes the correctness of the transcriptions [bɪt] and [bɪd]. If, on the other hand, the correct transcription were [btɪ] and [brɪd], these words would *not* constitute a minimal pair. It is common to find such false minimal pairs via transcriptional presumptions. By the standards of taxonomic analysis, [bɪt] and [bɪd] would not be a correct rendering of these words as surface phones. Something like [brɪt] and [brɪd] would be closer to correct, depending on the details of an individual’s pronunciation.

then the claim of phonemic status is not automatically rejected, but allophonic status also cannot be conclusively disproven because the subtle contextual differences could be included in the rule, i.e. “/p/ becomes [b] before \_\_ark#”. Before dismissing this possibility as beyond reasonable consideration, note that the only concern of the taxonomist is stating objective and mechanical procedures of analysis, which could in principle be performed by a machine. Plausibility is an undefined concept for them (since it implies a subjective and mechanically-undefined judgement on the part of the analyst).

One further fact about phonemic analysis is important: it is not supposed to be optional. That is, we are considering both the necessary and the sufficient conditions for partitioning phones into phonemes. An imaginable analysis of English phones, which maintains the contextual 1-to-1 mapping between phone and phoneme, could partition the phones English  $\{[p],[t],[k],[p^h],[t^h],[k^h]\}$  into the phonemes  $\{/p/,/t/,/k/,/p^h/,/t^h/,/k^h/\}$ . These phones can also be partitioned into the sets  $\{/p/,/t/,/k/\}$ . Only the latter solution is actually allowed by taxonomic standards; this is guaranteed by the principle of economy of phonemes, which requires an analysis to minimize the number of phonemes.

### Morphophonemics

The morphophonemic level is another level of analysis, one which bundles together phonemes. The requirement of biuniqueness is not imposed on the relationship of this level to the phonemic level, and this is how cases of phonological neutralization are dealt with. Regarding German final devoicing, where generative underlying *d* becomes [t], /t/ and /d/ are separate phonemes in the language, and therefore a mapping between underlying *d* ( $|d|$ , using the bracket for a morphophoneme) and surface [t] cannot be accomplished by a phonological rule — under the taxonomic conception — since phonological rules produce allophones from phonemes. A (taxonomic) phonological rule cannot have a phoneme as an output, it only produces (maps to) an allophone. Since you cannot know from the physical environment whether [bunt] is from /bund/ or /bunt/, the rule handling devoicing in German cannot be phonological, and it is *morphophonemic*. The German adjective “federal” would be morphophonemically  $|bund|$  and the adjective “colorful” would be  $|bunt|$ . There is a morphophonemic rule  $|d| \rightarrow /t/ / \_ \#$  in German, and there are also rules something along the lines of  $|z| \rightarrow /s/ / |p,t,k| \_$  and  $|z| \rightarrow /z/ / |b,d,g,m,n,l,r,a| \_$  in English. Morphophonemes, especially “process morphophonemes”, can be any symbol. Thus,  $|snV^*-\underline{r}-dek-a|$  in Klamath is phonemically /snedadka/ which is pronounced [snetət<sup>h</sup>a]: “V\*” means “copy of the next vowel” and “ $\underline{r}$ ” means “copy the following syllable”. These symbols are often purely arbitrary conventions created by the author, and can be replaced with any symbol that one wishes.<sup>6</sup>

The importance of the morphophoneme (and morphophonemic rules) for our understanding of the concept of “phoneme” is that mappings between sounds took place at two separate levels, in this theory of the phoneme.

---

<sup>6</sup> One could also express regular morphophonemic rules with different symbols, e.g.  $|D| \rightarrow /t/ \dots$  and  $|Z| \rightarrow /z/$ .

### Problems with the concept “phoneme”

Problems arose in attempting to define the analytic procedures for phonemic versus allophonic status. Chao (1934) noted that in Beijing Mandarin, the palatalized alveopalatal phones [č<sup>y</sup> č<sup>hy</sup> š<sup>y</sup>] only appear before the high front vowels [i,ü], whereas the velar [k k<sup>h</sup> x], alveolar [t<sup>s</sup> t<sup>sh</sup> s] and plain alveopalatal phones [č č<sup>h</sup> š] do not appear before these vowels: you can have [č<sup>y</sup>i], [ku], [t<sup>s</sup>u] and [ču], but not \*[č<sup>y</sup>u], \*[ki], \*[t<sup>s</sup>i] or \*[či]. In other words, [č<sup>y</sup> č<sup>hy</sup> š<sup>y</sup>] are in complementary distribution with [k k<sup>h</sup> x], but also in complementary distribution with [t<sup>s</sup> t<sup>sh</sup> s] and [č č<sup>h</sup> š]. Grouping of phones in complementary distribution into a single phoneme was not an optional analytical step within the taxonomic framework. Reduction of the size of the phonemic inventory is mandatory given the principle of phoneme economy (an assumption which has, for better or worse, been largely carried over into contemporary phonology). Therefore, the option of granting [č<sup>y</sup> č<sup>hy</sup> š<sup>y</sup>] the status of being phonemes themselves was unavailable, just as we also cannot say that English has both /t<sup>h</sup>/ and /t/ as phonemes. Yet the theory provides no other means of deciding *which* phoneme [č<sup>y</sup> č<sup>hy</sup> š<sup>y</sup>] belong to, so an arbitrary decision has to be made as to what phonemes these sounds represent: but arbitrary decisions are as much anathema to science as unwanted complexity is.

The distribution of [h] and [ŋ] in English also posed a problem. The fact is that [h] only appears syllable-initially and not in the context [ʷ\_\_v], and [ŋ] only appears postvocally at the end of a syllable or in the context [ʷ\_\_v]. Therefore *h* and *ŋ* are in complementary distribution which, given the principle of economy, means they would be allophones of one phoneme, represented with the symbol /h/ (given by Chao, which has come to be known as “heng”). This reduction of [h] and [ŋ] is generally rejected on the grounds of “insufficient phonetic similarity”, though it is not at all clear in what way [k] and [č<sup>y</sup>], or [t] and [ʔ] are phonetically similar enough that they can be considered allophones, unlike [h] vs. [ŋ]. The glottal [h] has a known phonetic affinity for nasalization (a phenomenon known as “rhinoglottophilia”, following Matisoff 1975) and in some Bantu languages such as Rufiji, vowels may be nasalized after *h*, suggesting that the phonetic disconnection between *ŋ* and *h* has been overstated. Another reason why the imaginable reduction of /h/ and /ŋ/ was generally rejected may be that there is no active relationship between the allophones of the supposed phoneme, whereby the pronunciation of a segment within a morpheme actually changes as a function of the physical context that the sound appears in.

A major problem pointed out by Halle (1959) is that this view of phonemes leads to a disunified treatment of phonological processes which we intuitively know should be unified. The system of Russian obstruent phonemes has some interesting gaps in voicing. There are phonemes /t/, /d/, /p/, /b/, /s/, /z/ and so on, as proven by minimal pairs. Some phonemes are unpaired in terms of voicing: there is /x/, but no contrasting \*/ɣ/; there is /č/ and no \*/ǰ/; there is /t<sup>s</sup>/ but no \*/d<sup>z</sup>/. Russian also has a regressive voicing process where obstruents agree in voicing with the following obstruent, so /kd/ → [gd], /bk/ → [pk], /sd/ → [zd]. Since /s/ and /z/, /k/ and /g/, or /p/ and /b/ are contrasting phonemes of Russian, by taxonomic analytic procedures, the rule mapping the morphophonemes |s|, |k|, |p| to the phonemes /z/, /g/, /b/ before voiced sounds and |z|, |g|, |b| to /s/, /k/, /p/ before voiceless sounds is a morphophonemic (neutralizing) rule. But since there are no independent phonemes \*/ɣ/, \*/ǰ/, the fact that [ɣ], [ǰ] *do* appear phonetically would have to be due to a separate voicing assimilation: these sounds arise *only* through of this rule, and these sounds are in complementary distribution with their voiced brethren. The rule accomplishing the

mapping from the phonemes /x, č, t<sup>s</sup>/ to [ɣ ʃ d<sup>z</sup>] must be an allophonic rule. The voicing process applied to /x, č, t<sup>s</sup>/ must be stated on a separate level, with no connection to the morphophonemic rules of voicing assimilation. Thus a theory maintaining strict separation of the phonemic and morphophonemic levels of analysis should be rejected, because it fails to express this voicing process, which common sense tells us is one single process, *as* one rule. This is the “celebrated Halle argument against the taxonomic phoneme”.

Another problem with this concept of the phoneme is English flapping. It is generally held that the flap [ɾ] is not a phoneme of English. However, flapping affects both *t* and *d*, so that (in some dialects) “writer” and “rider” are homophonous. The principle of biuniqueness disallows a surface phone from being a member of one phoneme (*t*) in some words but another phoneme (*d*) in other words: many-to-one mapping are allowed only between morphophonemes and phonemes. Since there is neutralization of the medial consonant in “rider” and “writer”, the rule responsible for this cannot be phonological (in the taxonomic sense). Biuniqueness fails: you cannot know from the pronunciation [ɾaɪtɾ] whether the intended word is “rider” or “writer”, therefore, there must exist a flap phoneme, /ɾ/, and a morphophonemic rule which turns the morphophonemes |d| and |t| into /ɾ/ in the relevant context. This flies in the face of common beliefs that there is no flap phoneme in English — although that belief might be mistaken.

Joos 1957 contains a collection of essential readings covering more than 30 years of the American structuralist tradition, and original sources should be consulted to understand the development of the phoneme, since “American Structuralism” is not a single monolithic theory. The problem of Russian voicing would probably not have been so serious for Bloomfield’s view of the phoneme as for Bloch’s.

### Other views of the phoneme

Other conceptions of phoneme existed as well, exemplified by Trubetzkoy 1939, Firth 1939, Jakobson 1949, Jones 1950, Martinet 1952 and others. As defined in Jones (1950), ‘A phoneme is a family of sounds, in a given language which are related in character and are used in such a way that no one member ever occurs in a word in the same phonetic context as any other member’: thus like the taxonomic linguists, Jones considered a phoneme to be a collection (‘family’) of sounds with non-overlapping distribution. Or, the phoneme is defined by Swadesh (1934) in ways that are similar to Sapir’s conceptualization of the phoneme (emphasis added):

The phonemic principle is that there are in each language a limited number of elemental types of speech sounds, called phonemes, peculiar to that language; that all sounds produced in the employment of the given language are referable to its set of phonemes; that only its own phonemes are at all significant in the given language. The phonemes of a language are, in a sense, percepts to the native speaker of the given language who ordinarily hear speech entirely in terms of these percept...  
The phoneme is the smallest potential unit of difference between similar words **recognizable as different to the native**. [p. 118]

The most important alternative view is this one, that the phoneme has an independent mental status needing to be discovered, as opposed to it being an arbitrary analytical construct assigned

to behavior. In post-structuralist phonology, the mental sense of “phoneme” has generally been considered to be the only one worth considering seriously: what unifies [t] and [t<sup>h</sup>] in English is that speakers “think of them as being the same”. The main problem with this theory is that it imposes an impractical burden on the language analyst, that is, there is no practical way for an analyst to reliably determine if two phonetic objects *are* unified as a phoneme or are separate phonemes, except to perform some kind of psychological test on speakers, and this is typically not possible.

Generative phonology largely rejected the phoneme as a formally significant unit of grammar, replacing it instead with concepts like the “bundle of distinctive feature values” which defines the symbolic sound unit as distinct from physical sound waves (see Chapter 6), and “underlying form” (see Chapter 4) which covers the distinction between sounds that can form lexical distinctions such as /t/ versus /d/, and those that do not ([t] versus [t<sup>h</sup>], the latter always arising through the operation of a rule in English). Some scholars working in the generative tradition have suggested that the phoneme might have some utility, for example Schane (1971), but it remains unclear exactly which concept of phoneme will turn out to be most useful, and to what extent “phoneme” needs to be a part of grammatical theory. No distinct grammatical principles have been uncovered that hold for rules creating phonemes versus allophones. More recently, there has been a resurgence of interest in the question of phonemes and contrasts, where phonemes achieve an element of theoretical legitimacy in some versions of Lexical Phonology; phonemes also pop up in terms of “Contrast Preservation” in Optimality Theory, but the exact nature of the concept “contrast” is still uncertain. See Dresher (2009) for a theory of feature assignment which takes the concept of “contrast” to be central.

### **Allophonic rule versus phonetic implementation.**

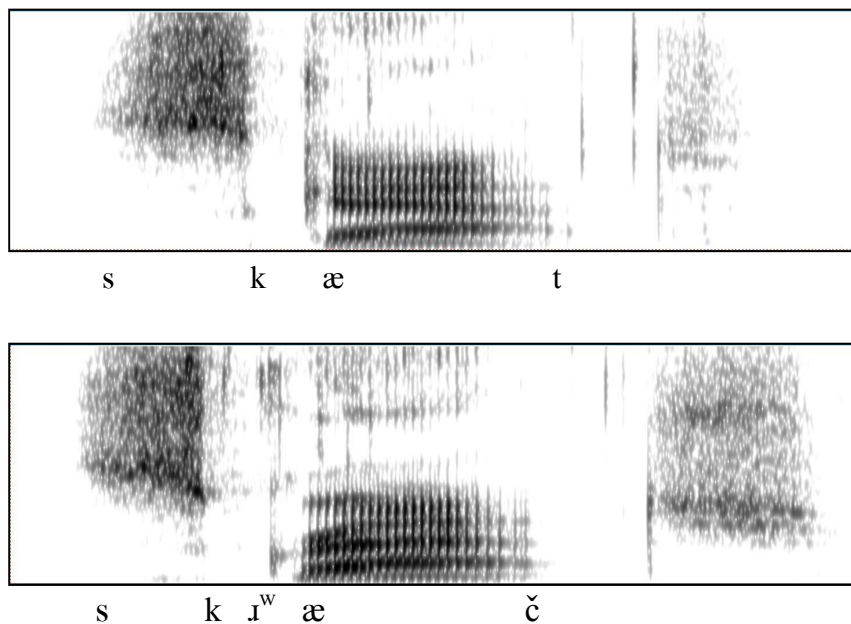
In many cases, it is not obvious that an observed contextual fact about pronunciation is due to such a thing as an “allophonic rule”, which is part of a phonology meaning sound-symbolic grammar. An example would be the claim found in some descriptions that the vowels /i,u/ have the allophones [e,o] before uvular consonants, so /iq/ → [eq]. It is possible that some language does have such a rule, but it is equally possible that the grammar simply produces [iq] and unavoidable physical constraints on production yield the apparent output [eq]. The basic problem is that language analysts only have access to a remote sound wave, not the physiological events creating the sound wave or the end of the phonological computation that is acted on by the articulatory apparatus. See Hale & Reiss (2009: ch. 5) for discussion.

This brings us to the question of distinguishing phonetics from phonology. In some versions of generative phonology, especially Chomsky & Halle (1968), no line is drawn between what would now be called “rules of phonetic implementation” and phonological rules. In the SPE conception of phonetics, phonetics *is* just the set of unavoidable universal physically-driven facts of pronunciation that derive from exigencies of human anatomy. Most so-called low-level phonetic rules would be seen as just another kind of *phonological* rule — remember that the primary import of Halle’s argument against the taxonomic phoneme was that *there is no fundamental difference* between phonetics and phonology, in contrast to the taxonomists who held that they were completely different objects. To understand how these views could both be held by rational people, consider the pronunciation of the word “school”. Phonetic descriptions may say

that the *k* is rounded, so phonetically you have [sk<sup>w</sup>uwl]. A more fine-grained description reveals that the initial *s* is also somewhat rounded, not as much as the *k*, so we could encode these observations in the transcription [s<sup>o</sup>k<sup>w</sup>uwl], with “degree of rounding” marked via different diacritics. These facts can be expressed in a grammar by rules of the type “*k* → *k*<sup>w</sup>/\_\_*u*”, “*s* → *s*<sup>o</sup>/ \_\_*C*<sup>w</sup>”, and at least to the extent that these facts are not automatic consequences of human physiology, they would be the result of grammatical rules in the SPE account. The important question to answer in deciding whether this fact of pronunciation falls within the ambit of “universal phonetic necessity” versus “language-specific detail” is whether languages can systematically vary in how such outputs are pronounced, thus can there be a language with a word [skul] without any rounding of [s], providing a crosslinguistic contrast with English?

Another description of the facts, not following the SPE uniformity hypothesis, is that process at work in the pronunciation of “school” is the result of a rule of phonetic implementation, which is the interface component responsible for implementing symbolic representations physically, but this component is a component of a grammar and thus can vary between languages. Under that theory, we could describe the process in the phonetic component by saying that the onset of lip protrusion required for the vowel *u* precedes the actual onset of the vowel *u* in time, beginning somewhere around where *s* begins, and that anticipatory lip protrusion is a continuous physical function.

One important difference between these accounts is that the phonological characterization produces two categorically-distinct segments, not a continuum over time: a phonological rule produces only a finite set of discrete categories, not a continuum. In support of the phonetic theory of *s*-rounding, it can be observed in spectrograms that the change in formant frequency that corresponds to lip rounding is actually continuous. Compare the spectrograms of “scat” and “scratch”, where *r*, which is rounded in this dialect of English.



It can be seen that rounding before rounded *r* is a physically continuous process, not an abrupt one as implied by a categorial analysis

Another differentiating feature requires reference to the theory of phonological properties (features, to be discussed in Ch. 6), and requires determining whether the concept “semi-round” is legitimately part of symbolic phonology. The question that we must ask is whether such a distinction can be represented, given the nature of those features. Distinctive feature theory only provides the oppositions [ $\pm$ round]. Thus if [u] is [+round] and [k<sup>w</sup>] is [+round], what could half-round *s* be, other than [-round] and therefore not distinct from normal [s]? The binary representation of features thereby limits the possible phonological segments; so, given that theory, it would be impossible to describe this partial rounding process phonologically.<sup>7</sup>

Similarly, it has been noted that velars become fronted in English before front vowels, as in *keep*, *kill* etc. A fact which is not widely commented on is that the degree of fronting varies as a function of the vowel, so that fronting is maximal with *keep* and minimal with *cap*. The best way to describe this situation is to simply say that the tongue body is anticipatorily placed in the actual position of the following vowel; this clearly is not a symbolic phonological description, in the same sense of “phonological description” as used in the textbook. The approach assumed here is that language sound systems are characterised by at least two components, one being the symbolic phonological component dealing with discrete units, and the other being a phonetic component which physically interprets those phonological units and operates in terms of (something best modeled by) continuous mathematical notions defined over concrete acoustic and articulatory units such as the tongue, or formants.

Numerous examples of contextual variations of “keep”/“kill” or “school” variety can be found, and they account for a considerable portion of the set of allophonic rules assumed to exist in languages. Other examples of allophonic variations in English which might be attributed to phonetic implementation include supposed devoicing of sonorants after voiceless obstruents, glottalization of syllable-final voiceless stops, vowel nasalization, and aspiration. If we abandon the SPE view that phonetics is just the universal, physically-mandatory stuff, by what criterion do we conclude that vowel nasalization in English is part of the phonetic component, or the phonological component? What are the necessary and sufficient conditions for a process to be deemed phonological — “dealt with by categorical symbolic operation” — versus phonetic — “dealt with in terms of continuous physical functions”? Skirmishes over this issue continue, but certain territory can clearly be carved out, if you assume that there is a difference between phonetics and phonology in the first place.<sup>8</sup> For instance, there is very little power to refer to concepts of “time” in phonology, except via precedence (and, as we will see later, domain of association and number of skeletal positions). The concept “halfway into the vowel” is beyond the reach of phonology, as would be “40 msc”. So too would be any reference to actual physical states (thus the actual fact of vocal fold vibration, as opposed to the feature [+voice] regardless

<sup>7</sup> There is a very important escape clause here. SPE feature theory actually anticipated this problem, and had both binary plus-minus values as well as continuous scalar feature values. Accordingly, fully round [u] could be [1round], “fairly round” [kw] could be [2round] and “half-round” [s<sup>o</sup>] could be [3round]. This idea was pursued in the 70’s to a limited extent, but scalar feature values are no longer discussed as a theoretical possibility.

<sup>8</sup> SPE denies such a distinction, rendering all phonetics part of the phonology, and certain phoneticians e.g. Articulatory Phonologists seem to deny it by subsuming all phonology under phonetics.

of physical manifestation). No feature informs us about the jaw, therefore any operation which must be understood in terms of movement of the jaw must be phonetic.<sup>9</sup> On the other side, it is at least widely assumed that phonetic implementation is “blind” to abstract grammatical structure in the sense that it cannot refer to lexical or morphological properties, or anything other than the actual physical material around a given sound. It is also assumed that phonetic implementation cannot neutralise distinctions between sounds, and thus a rule neutralising /zap/ and /zab/ to [zap] could not be phonetic. This restriction is arbitrary, and does not follow from the nature of “phonetic implementation”, largely because a fully developed theory of phonetic implementation has not yet been created. It is not clear that there is anything that *can't* be handled by phonetic implementation, and if we can't objectively partition the class of processes into relatively neat piles, we're back where we started from.

Under the conception of the phonology / phonetics distinction proposed here, a process which operates in terms of continuous values and actual physical properties must be in the domain of phonetics, and one which operates in terms of discrete symbolic units is in the domain of phonology. A rule which converts the voiceless units *p*, *t*, *k* into voiced units *b*, *d*, *g* between vowels would be a phonological process; but a process which decreases the size of a glottal opening gesture and increases the speed of a constriction-forming gesture would be a phonetic process. The difficulty lies in determining which is a better characterisation of a given process, since the phonological process of intervocalic voicing could also be described in terms of the physical consequences of symbol substitution.

One implication behind a phonological account of a sound process is that the grammar definitively selects one of two different outputs — [t] or [t<sup>h</sup>] — befitting the discrete nature of phonological processes (there might be two discrete outputs, if the phonological rule in question were optional). Since the physical output is a continuum, and since it is phonetics that is concerned with how discrete symbols are realised in continuous physical reality, the discovery that the proper description of a process accesses a continuum of values precludes the process from being treated phonologically. As emphasised in the preceding chapters, phonology considers language sounds as being drawn from a small inventory of possibilities: a phonological description does not countenance the postulation of an unlimited set of entities, such as [t] versus [t<sup>h</sup>], [t<sup>h</sup>], [t<sup>h</sup>], [t<sup>h</sup>] and so on, with different sized *h*'s iconically representing greater degrees of aspiration (See Chapter 6 for the theoretical basis for restricting the number of distinctions in a phonological account).

One impediment to resolving the question of how to distinguish phonetic versus phonological processes is that processes which have been represented as categorical choices between one segment versus another are sometimes shown to result from an analyst's imposition of categoricity on an actually continuous process. It has often been presumed that English has an allophonic rule of regressive nasalization, whereby /dɛn/ becomes [dɛ̃n], changing an oral vowel into a nasal vowel. However Cohn (1993) shows that the timing of nasal airflow in English /dɛn/ is quite different from the timing of nasal airflow in French, which has phonemically contrastive

---

<sup>9</sup> One has to be careful not to engage in sophistry when arguing for jaw-sensitivity. You might construct an argument that the jaw *could* be the crucial factor, but this does not automatically mean that the process *must* be described in terms of the jaw.

oral and nasal vowels — where nasality is uncontroversially phonological. The prediction of a phonological account of English vowel nasalisation is that the vowel should be completely nasal; but Cohn’s results show that the English vowel (unlike the French vowel) is only nasal on the second half. The limited precision of impressionistic transcriptions has caused scholars to conclude that a pre-nasal vowel in English has “become nasal”, in a categorical sense. The perceptual tendency to categorize language sounds is so strong that continuous phonetic variation probably cannot be distinguished from binary categorial distinction using ordinary auditory procedures.

One analysis of the Marshallese vowel system (Bender 1963) assumes a large inventory of vowels, including [i e ε u ʌ a u o ɔ] plus various diphthongs. Bender (1968) reanalyses the system as having just a set of phonological central unrounded vowels, deriving the remaining surface vowels by allophonic rules, whereby surface vowel quality is parasitic off of secondary articulations of the preceding and following consonant (thus, [e] only appears between palatalised consonants, [o] only appears between rounded consonants). Choi (1992) in a phonetic study of Marshallese argues that the most accurate description derives the supposed vowels [e], [o] and various diphthongs from the *surface* vowel (symbolically speaking) [ʌ]. Thus, [e], [o] and so on are not actually part of the phonological grammar of Marshallese, and the quality perceived as [e] arises because of the pattern of phonetic timing of articulators. Under Choi’s view, all that the grammar of Marshallese (i.e. the symbolic, categorical computations) contains is a set of palatalized, rounded and velarized consonants, and the vowels [u ʌ a] (not “definitely unrounded and central”, rather “neither definitively round nor non-round” etc. — a set of “vague” vowels lacking conventional symbols). In the phonetic-implementation approach to Marshallese vowels, as far as the grammar is concerned what sounds like *l<sup>w</sup>ɔŋ<sup>w</sup>* “housefly” is surface phonologically [l<sup>w</sup>ʌŋ<sup>w</sup>], but because of the way articulator movement works in this language, this ‘sounds like’ [l<sup>w</sup>ɔŋ<sup>w</sup>]. Hale (2000) and Hale & Reiss (2008) highlight this conceptual difference by using a novel bracketting symbol, thus something like *l<sup>w</sup>ɔŋ<sup>w</sup>*, to indicate “an idealization of the physical output of the body”, distinct from [l<sup>w</sup>ʌŋ<sup>w</sup>] which is “the terminal representation of the grammar”.

### English consonant allophones.

Kahn (1976) is a basic reference on the relevance of the syllable to aspiration, flapping and glottalization in English. Since the point of our discussion of consonant allophones in section 1 is to familiarize the student with the notion of allophonic variation, the full range of conditions on these rules are not investigated. For example, aspirated and flapped /t/ actually contrast in ‘capitalistic’ [k<sup>h</sup>æpərəlístɪk] versus ‘militaristic’ [mɪlətɹístɪk], the reasons presumably having to do with the difference between stress (hence aspiration) in ‘cápital’ versus ‘mílitary’. Other contemporary references on English consonant allophony include Kiparsky (1979), Selkirk (1982), Withgott (1982), Harris (1994), and de Jong (1998). There are numerous traditional descriptions of English pronunciation, including Jones (1956), Trager & Smith (1951), and Gimson (1962).

## Readings

- Baudouin de Courtenay, Jan. 1895. An attempt at a theory of phonetic alternations: a chapter from psychophonetics.' *A Baudouin de Courtenay Anthology: The beginnings of structural linguistics*. Ed. and trans by Edward Stankiewicz. Bloomington, IUP. 1972. 144-212.
- Bloch, Bernard. 1948. A set of postulates for phonemic analysis, *Language* 24.
- Bender, Byron. 1963. Marshallese phonemics: labialization of palatalization. *Word* 19: 335-341.
- Bender, Byron. 1968. Marshallese phonology. *Oceanic linguistics* 7: 16-35.
- Chao, Y. R. 1934. The non-uniqueness of phonemic solutions of phonetic systems. Bulletin of the Institute of History and Philology, Academia Sinica, vol 4, part 4: 363-97.
- Choi, John. 1992. *Phonetic underspecification and target-interpolation: an acoustic study of Marshallese vowel allophony*. Doctoral dissertation, UCLA: Los Angeles.
- Chomsky, Noam & Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of linguistics* 2: 95-138.
- Chomsky, Noam. 1964. Current issues in linguistics theory. J. Fodor & J. Katz (eds.) *The structure of language*, 50-118. Prentice-Hall: Englewood Cliffs, NJ.
- Cohn, Abigail. 1993. Nasalization in English: phonology or phonetics. *Phonology* 10: 43-81.
- de Jong, Kenneth. 1998. Stress related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of phonetics* 26: 283-310.
- Dresher, D. Elan. 2009. *The Contrastive Hierarchy in Phonology*. Cambridge Studies in Linguistics. New York: Cambridge University Press.
- Firth, J. R. 1934. The word 'phoneme'. *Le maître phonétique*, No 46.
- Gimson, A. C. 1962. *An introduction to the pronunciation of English*. London: Edward Arnold.
- Hale, Mark. 2000. Marshallese phonology, the phonetics-phonology interface and historical linguistics. *The Linguistic Review* 17: 241-257.
- Hale, Mark & Charles Reiss. 2008. *The Phonological Enterprise*. Oxford University Press: Oxford.
- Halle, Morris. 1959. *The sound pattern of Russian*. Mouton: The Hague.
- Harris, John. 1994. *English sound structure*. Blackwells: Oxford.
- Harris, Zelig. 1944. Simultaneous components in phonology. *Language* 20: 181-205.
- Hockett, Charles. 1942. A system of descriptive phonology. *Language* 18: 3-21.
- Jakobson, Roman. 1949. On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague* 5: 205-213
- Jones, Daniel. The phoneme, its nature and use. Cambridge: Heffer
- Jones, Daniel. 1956. *The pronunciation of English*. Cambridge: Cambridge University Press.
- Joos, Martin. 1957. *Readings in linguistics I*. University of Chicago Press: Chicago.

- Kahn, Daniel. 1976. *Syllable based generalizations in English phonology*. Doctoral dissertation, MIT. Distributed by Indiana University Linguistics club: Bloomington.
- Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic inquiry* 10: 421-441.
- Matisoff, J. 1975. Rhinoglottophilia: the mysterious connection between nasality and glottality. Ferguson, C., Hyman, L. & J. Ohala (eds.). *Nasálfest: papers from a symposium on nasals and nasalization*, 265-287. Language Universals Project, Stanford. Martinet, André. 1952. Function, structure and sound change. *Word* 8: 1-32
- Postal, Paul. 1968. *Aspects of phonological theory*. Harper & Row: New York.
- Sapir, Edward. 1925. Sound patterns in language. *Language* 1:37-51.
- Schane, Sanford. The phoneme revisited. *Language* 47: 503-521.
- Selkirk, Elizabeth O. 1982. The syllable. in H. van der Hulst & N. Smith (eds.) *Structure of phonological representations, vol 2*, 337-384. Foris: Dordrecht.
- Swadesh, Morris. 1934. "The Phonemic Principle." *Language* 10: 117-129.
- Trager, George & H. Smith. 1951. *An outline of English structure*. American Council of Learned Societies: Washington.
- Trubetzkoy, N.S. 1939. *Principles of phonology*. translated by C. Baltaxe, 1969. U. California Press: Berkeley.
- Twadell, W. Freeman. 1935. *On defining the phoneme*. Language monograph 16.
- Withgott, M. 1982. *Segmental evidence for phonological constituents*. Doctoral dissertation, U. Texas: Austin.