Phonological Abstractness in the Mental Lexicon

Baković, Heinz, Rawski

March 27, 2020

1 Talking and Listening

When we try to understand how we speak, and how we listen, an unavoidable fact is that we use the same vocabulary items again and again. So, not only are these stored in memory, but they are learned on a language-specific basis. In this way, we directly confront the issue of mental representations of speech; that is, something about how we store, retrieve and remember units of speech. What are these units? What are these mental representations? These questions are central to generative phonology, a fact reflected in the title of a collection of important papers by Morris Halle, a founding father of the field: *From Memory to Speech and Back Again* (Halle, 2003).

In this chapter we review some of the most basic — and most fascinating — conclusions and open questions in phonology regarding how abstract these mental representations of speech are. In an era of big data, and data mining, the prevailing attitude in science and scientific observation seems to be to store every last detail, to never assume any detail is irrelevant, no matter how small, because you never know when it may in fact make a difference somewhere. This attitude is also present in the psychology of language, where it has been shown that people are sensitive to astounding details of aspects of speech reflecting people's age, gender, social status, neighborhood, and health (Coleman, 2002; Pierrehumbert, 2002; Johnson, 2006). Pierrehumbert (2016) reviews this literature and highlights some theoretical accounts.

In the context of this prevailing attitude, the theory of generative phonology makes some surprising claims. First, it claims it is *necessary* to abstract away from much of this detail to explain systematic aspects of the pronunciations of the morphemes, words, and phrases we utter. In other words, while a person is sensitive to subtle phonetic details which are informative about various aspects of a speaker's condition, the theory of generative phonology claims those very same details are *irrelevant* to the very same persons' mental representations of the *pronunciations* of morphemes and words. As such the mental representations of these pronunciations are particular abstractions. We review the arguments for this position in section 3. Section 4 shows how these ideas lead to the *phoneme*, a necessarily abstract category of speech units, which is said to be the fundamental unit out of which the pronunciation of vocabulary items are built.

We then turn to a question posed by Paul Kiparsky: *how abstract is phonology?* (Kiparsky, 1968) — or, more specifically, *how abstract are the mental representations of speech?* A vigorous back and forth debate over this question followed, but now, over fifty years later, there is still no consensus regarding the answer. A basic problem is that some languages present good evidence that morphemes are specified with phonological content that is never realized as such in any surface manifestation of those morphemes. At issue is precisely what kind of evidence justifies such abstract phonological content. In our minds, this question, and others related to it, are particularly important and remain among the most interesting, and understudied, questions in phonology today.

But first, we begin this chapter with some thoughts on the nature of abstraction and idealization.

2 What is abstraction?

The first thing to point out about abstraction is not how odd it is, but how common it is. Orthographic letters are abstractions. The capital letter "A" and the lowercase letter "a" are quite different, and yet at some level of abstraction they are referring to the same thing. Money is an abstraction. Whole numbers and fractions are also abstractions. For example, there is no such thing as "three." There are only examples of collections of three items. Such abstractions are not just taken for granted, but they are valued: the earlier toddlers and young children learn such abstractions, the more we marvel at their intelligence.

A very common approach to problem solving one learns in grade school is shown in the diagram in Figure 1.

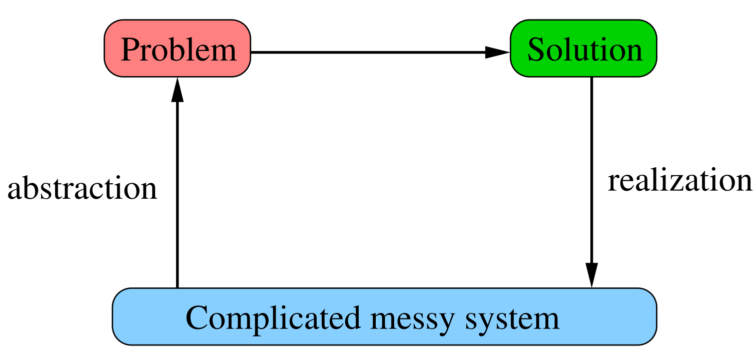

Figure 1: Abstraction, Problem-Solving, and Realization

For example, in an elementary-level math class, students are often given a problem expressed in plain language. Their job is to extract the relevant information, organize it into an equation, solve the equation, and report back the answer. Figure 2 shows a 4th grade math exercise.

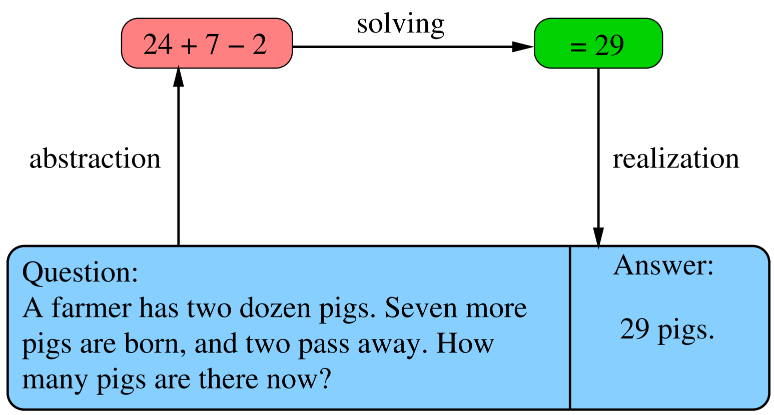

Figure 2: A 4th grade math exercise illustrating the process in Figure 1.

Similarly, when responding to challenges that his law of falling bodies did not correspond to the real world, Galileo used an analogy to accounting:

> what happens in the concrete [...] happens in the abstract. It would be novel indeed if computations and ratios made in abstract numbers should not thereafter correspond to concrete gold and silver coins and merchandise [...] (quoted in Wootton (2015, pp. 23–24))

Here Galileo mentions an important feature: one can make computations with abstract numbers without directly referencing the particular objects that implement them. Given a proper abstraction, one can make an inference or calculation at the level of numbers with results that then correspond to some specific physical effect or process. This property, that a given abstraction can have many implementations, is known as **multiple realizability**. An abstract object is multiply-realizable by a number of concrete objects. The concrete objects might differ from each other in various ways, but the ways in which they differ are irrelevant to the abstraction. We suppose orthographic letters, and the mental representations of units of speech, are abstract in this way, too.

The second important point about abstraction is that there are degrees of abstraction. Even the question in Figure 2 as written in plain language is abstract since, for example, it uses numbers to explain the situation, and not for instance a photograph of the two dozen pigs and/or video of pigs birthing and dying. Figure 3 below illustrates the concept of degrees of abstraction. The only change we would make to it would be to add question marks after each of the subcaptions in the figure. The question is always "Is the level of abstraction too realistic? just right? or too abstract?"

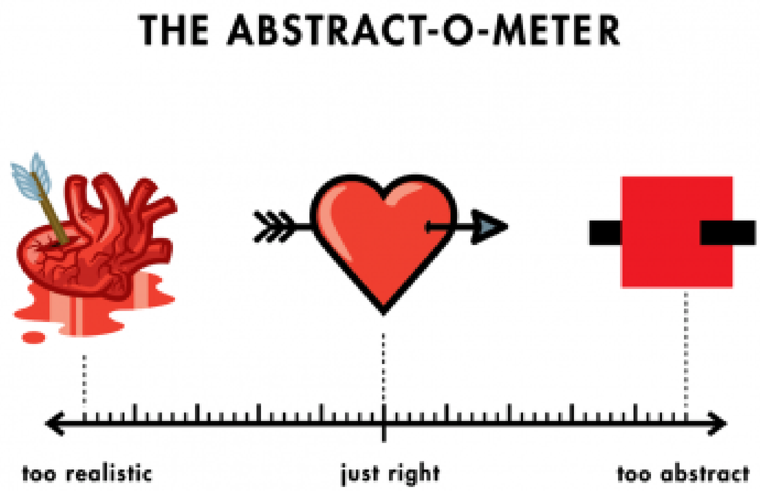

Figure 3: Degrees of Abstraction (from `https://computersciencewiki.org/index.php/Abstraction`).

Answering this question is not easy. It is not at all obvious what the right level of abstraction is. It is appropriate to argue about whether a particular degree of abstraction is appropriate or not.

We are in an age of big data and the milieu of the age seems to be to err on the side of "too realistic" and to avoid analyses that are "too abstract." Our own view is that this is shortsighted. Many things were at one time considered to be "too abstract" but are now recognized as being perfectly reasonable and in fact essential and useful. When one studies the history of mathematics, it is interesting how notions once considered to be "too abstract" were considered crazy or useless. This includes things like the number 0, real numbers, $\sqrt{-1}$, uncountable infinity, number theory, and so on. When Pythagoras and his group realized that the $\sqrt{2}$ could not be expressed as a fraction of whole numbers, they called it "irrational," literally "unreasonable." The term sticks today. Cantor went mad after realizing that infinitely-sized sets had different degrees of cardinality. Today this fact underlies the Church-Turing thesis of computability. The development of the complex numbers caused much costernation in the 17th and 18th centuries but are routinely used today to understand complex physical systems. Developments in number theory in the early part of the 20th century had no purpose other than to satisfy some strange mathematical aesthetic, and now underlie secure cryptographic communications, protecting business operations, journalists, and other vital communucations.

In short, we are sympathetic to the view put forth by Cheng (2015, p. 22): "Abstraction can appear to take you further and further away from reality, but really you're getting closer and closer to the heart of the matter".[1]

In this chapter we show that abstractness is a property exploited both by speakers of natural languages and by scientists describing the linguistic knowledge of those speakers.

3 Phonology and the Mental Lexicon

The pronunciation of a word is a sequence of events in time. From the perspective of speech production, these events are defined articulatorily. From the perspective of speech perception, these events can be defined acoustically and perceptually. The International Phonetic Alphabet (IPA) defines categories of speech sounds articulatorily and provides a symbol for each categorized speech sound. For consonants, these symbols specify three aspects of articulation: the place of articulation, the manner of articulation, and the activity of the vocal folds. Symbols representing vowels primarily specify the degree of jaw aperture, how forward the positioning of the root of the tongue is, and any rounding of the lips. For example, the pronunciation of the word 'math' is transcribed in the IPA as [mæθ], as there are three distinct sounds in sequence: [m], which is articulated by stopping airflow at the lips but releasing it through the nose, with the vocal folds held together such that they vibrate; [æ], which is an open front unrounded vowel; and [θ], which is articulated by constricting but not stopping airflow with the blade of the tongue between the teeth with the vocal folds spread apart. For most speech acts, these articulations are similar across speakers of the same idiolect, despite individual physiological variation.

For a word like 'tune', one transcription of it using the IPA is [tun], often referred to as a broad description. In contrast, a narrow transcription of this word using the IPA is [tʰũːn]. The difference between the broad and narrow transcriptions is the degree of abstraction. Both transcriptions reveal systematic aspects of standard American English speech. However, the broad transcription only includes so-called *constrastive* information and the narrow transcription includes some *non-contrastive* information as well, indicated in this case with various diacritic marks: the aspiration on the [t], indicated with [ʰ], and the nasalization [˜] and extra duration [ː] of the vowel [u]. Both kinds of transcriptions can be used as abstract representation of the pronunciation of this word stored in memory, and one could ask whether the long term memory representation of the pronunciation of the word 'tune' is more like the broad transcription, the narrow transcription, or something else. One can ask if there is only one long-term memory representation of the pronunciation of the word 'tune', or if there are multiple, possibly partially redundant representations. All of these possibilities are open to study. In the remainder of this chapter, we use broad IPA transcriptions as we discuss

[1] One of the clearest recent expositions on abstractness and its virtues is in Ch. 2 of Cheng's book, *How to Bake* π, which we encourage readers of the present chapter to read.

representations of the pronunciations of words. The degree of abstractness chosen is not critical, but we settle here to facilitate additional discussion.

With this in mind, we ask the question: what does modern generative phonological theory say about the mental representations of speech? The central empirical fact that informs this question is that, in many languages of the world, morphemes are pronounced in different ways depending on context. To illustrate with an example, Odden (2014) draws attention to the pattern exhibited by the different verb forms in Kerewe shown in Table 1.

Infinitive	*1sg habitual*	*3sg habitual*	*Imperative*	*gloss*
[kupaamba]	[mpaamba]	[apaamba]	[paamba]	'adorn'
[kupaaŋga]	[mpaaŋga]	[apaaŋga]	[paaŋga]	'line up'
[kupima]	[mpima]	[apima]	[pima]	'measure'
[kupuupa]	[mpuupa]	[apuupa]	[puupa]	'be light'
[kupeketʃa]	[mpeketʃa]	[apeketʃa]	[peketʃa]	'make fire w/ stick'
[kupiinda]	[mpiinda]	[apiinda]	[piinda]	'be bent'
[kuhiiga]	[mpiiga]	[ahiiga]	[hiiga]	'hunt'
[kuheeka]	[mpeeka]	[aheeka]	[heeka]	'carry'
[kuhaaŋga]	[mpaaŋga]	[ahaaŋga]	[haaŋga]	'create'
[kuheeba]	[mpeeba]	[aheeba]	[heeba]	'guide'
[kuhiima]	[mpiima]	[ahiima]	[hiima]	'gasp'
[kuhuuha]	[mpuuha]	[ahuuha]	[huuha]	'breathe into'

Table 1: Kerewe verbs, from Odden (2014, pp. 88–89)

There is an interesting difference between the first group of verb forms and the second group of verb forms. In the first group, for example, the pronunciation for the verb stem 'adorn' is consistently [paamba] regardless of whether it is in the infinitival (prefixed with [ku]), 1sg habitual (prefixed with [m]), 3sg habitual (prefixed with [a]), or imperative form (not prefixed). As such it makes sense to assume that /paamba/ represents in Kerewe speaker's mental lexicon the major features of the pronunciation of the verb stem 'adorn'.[2] However, when the same kind of morphological analysis is applied to the forms of the verb 'hunt', we find that the verb stem's pronunciation in the 1sg habitual form is [piiga] whereas in the other forms it is [hiiga]. So the question naturally arises, what long-term memory representation of the pronunciation of the verb stem 'hunt' do speakers of Kerewe have?

One possibility is that both /piiga/ and /hiiga/ are stored, along with the knowledge that [piiga] is used in the 1sg habitual form and that [hiiga] is used otherwise.[3] This is fine as far as it goes, but it is of interest that the

[2]Here and elsewhere in this chapter we follow traditional phonological notation of transcribing the mental representation of speech between slashes and the actual pronunciaton within square brackets. When the distinction is immaterial, we use italics.

[3]More and more detailed arguments against this 'morpheme alternant theory' can be

other verbs in this second group pattern exactly the same way: they begin with [p] in the 1sg habitual form, and with [h] otherwise. Furthermore, there are no verb stems in Kerewe which begin with [h] in the 1sg habitual form. Taken together, these observations suggest there is something *systematic* about the variation in the pronunciation of the various forms of this group of verbs in Kerewe, a systematicity that simple storage of all pronunciations of the verb stems does not readily or insightfully capture.

The methods of modern generative phonology lead analysts to posit that the long-term memory representation of the pronunciation of the verb stem 'hunt' that speakers of Kerewe have is /hiiga/, and that h representations are transformed to p representations immediately after m representations. Consequently, the 1sg habitual form of 'hunt' is [mpiiga] because it derives from /m+hiiga/ by application of this phonological transformation.[4] This phonological analysis explains the systematic variation observed because it predicts that every /h/-initial verb stem ought to be realized as [p]-initial when it follows any prefix ending with m, such as the 1sg habitual.

There are two key reasons why the mental representation of 'hunt' cannot instead be /piiga/, with p transformed to h after not-m. First, it is assumed that phonological transformations cannot make direct reference to the negation of a class of speech sounds; in this case, to 'not-m'. Second, and more significantly, it has been independently determined that the members of the first group of verb stems in Table 1 all begin with /p/, and yet the putative p-to-h-after-not-m transformation does not apply to those verb forms. Positing that members of the first group of verb stems begin with /p/ and that members of the second group begin with /h/ (transformed to p after m) succinctly and systematically distinguishes the two groups.

One argument against this position is that there are cases where it does seem like two distinct pronunciations of a morpheme are stored. A well-known example is the nominative suffix in Korean, which has two pronunciations: [ka] and [i]. The [ka] occurs with vowel-final words, and the [i] occurs with consonant-final words. This alternation is phonologically conditioned, but it is nevertheless strange to have a rule which converts /ka/ to [i], or /i/ to [ka], or some other machination. The strangeness is compounded by the fact that there are no other places in the Korean lexicon or language which exemplify any $k \sim \varnothing$ or $i \sim a$ alternation. In other words, a single posited underlying form and rules (whatever they are) would only account for the alternation between [ka] and [i] in the nominative suffix.

So in Korean, the best analysis of the nominative appears to be a long-term memory representation of the pronunciation of the nominative as $\{$/i/, /ka/$\}$ with the choice of which one to select being based on the phonological properties of the stem the suffix attaches to. Given that such examples

found in Kenstowicz and Kisseberth (1979, pp. 180–196).

[4]The '+' indicates the boundary between the 1sg habitual prefix /m/ and the verb stem /hiiga/.

exist in the world's languages, the question is: why don't we just use the same kind of analysis for Kerewe? The answer is the one given above. The systematicity observed in the realization of Kerewe verb stems points to the straightforward phonological analysis in that case, and the lack of systematicity observed in the realization of the Korean nominative suffix points to the alternant selection analysis in that case.

The argument just made is the basic one for the position that morphemes are stored in long-term memory with a single representation, known as the *underlying form*. The fact that the same morpheme can be pronounced in different ways depending on context is due to phonological transformations of this underlying form into its various *surface forms*. The fact that these transformations are systematically applied explains the systematicity in the alternations of the morphemes across the vocabulary of the language.

The phonological analysis in the generative tradition thus comes with two symbiotic parts. The first posits, where systematicity demands it, a single mental representation in long-term memory of the pronunciation of a lexical item — that item's underlying form. The second is the transformational part of a phonological grammar which defines how underlying forms are mapped to surface forms, which are more concrete representations of the pronunciation of the lexical item in the particular context in which it is realized. For example, in the case of Kerewe, the underlying form for 'hunt' is /hiiga/, the underlying form for the 1sg habitual form of 'hunt' is /m+hiiga/, there is a phonological transformation changing *h* to *p* after *m*, and so the surface form of this verb form is [mpiiga].

Once underlying forms and transformations making them distinct from some of their corresponding surface forms are posited, a natural question arises: *how distinct can underlying forms be from surface forms?* This is the question of abstractness in phonology. We approach this question first from the perspective of two interrelated, fundamental concepts in phonological theory, phonemes and distinctive features, in section 4, setting the stage for discussion of more specific examples of evidence for different types of abstractness in analyses of phonological patterns in section 5.

4 Phonemes and Features

Much of the question of abstractness in phonology concerns the individual units of speech whose concatenation makes up the pronunciation of lexical items. For example, there is a clear sense that the sounds transcribed as *k* in the Kerewe verb stems *heeka* 'carry' and *peketʃa* 'make fire with stick' are the same at some level of abstraction; they are just being utilized in different places in different lexical items. More strikingly, there is an equally clear sense that the [t^{h}] in [t^{h}ɪk] 'tick', the [t] in [stɪk] 'stick', and the [ɾ] in [ˈæɾɪk] 'attic' are also the same in English, and at the same level of abstraction.

The idea that people's mental lexicons make use of a mental alphabet of abstract speech sounds in more or less the same way that a physical dictionary of English makes use of Roman letters has a very long tradition in linguistics, dating back at least to Pāṇini's grammar of Sanskrit, likely written in the 6th–5th century BCE.[5] Details of individual theories aside, the speech sounds of this abstract mental alphabet are referred to as *phonemes*. Phonemes are obviously abstractions: they represent groups or categories of speech sounds, abstracting away from many phonetic particulars, such as the relatively easily perceptible differences between [t^h], [t], and [ɾ] in the English example above. In this section we trace the evidence for the psychological reality of phonemes and highlight two prominent views of phonemes: one as an indivisible alphabetic symbol, and the other as a local confluence of more basic units or properties known as *distinctive features*.

Readers interested in a more comprehensive review of the phoneme are directed to Dresher (2011), who provides an excellent review and discussion. Dresher distinguishes three views of the phoneme as an entity: the phoneme as physical reality, the phoneme as psychological concept, and the phoneme as a theoretical fiction. After considering the behaviorist underpinnings of the physical view, and dismissing the fictional view due to the realism that dominated linguistics since the 1950s, Dresher (2011, p. 245) concludes that "once we abandon empiricist assumptions about science and psychology, there is no obstacle to considering the phoneme to be a psychological entity." This is an important framing, because it means that any linguistic work describing the nature and content of the phoneme is by necessity a statement about mental representation, in much the same way as the concept of underlying forms discussed above. So the question becomes, how much abstraction or idealization is involved in such a psychological entity?

Edward Sapir was the first to present evidence for the psychological reality of phonemes. Perhaps Sapir's better-known article, "The Psychological Reality of Phonemes" (Sapir, 1933), was preceded by his article in the first issue of the journal *Language*, "Sound Patterns of Language" (Sapir, 1925). Together, these two papers establish the perspective that (1) the phoneme is a psychological unit of speech, and (2) the character of the phoneme requires taking into account how it functions in the larger phonological context of the language; it cannot be understood solely from investigation of the articulatory or acoustic properties of its surface realizations. Sapir (1925) argued that languages with the same surface inventory of speech sounds could be organized differently at the underlying, phonemic level. Sapir (1933) argued that the same underlying phonemic organization can explain the behavior of native speakers of a language, whether it be with respect to errors they make in the perception or production of speech or in terms of the choices they make when devising or using a novel writing system.

Nearly a century later, psychological evidence for phonemes continues

[5]A much more comprehensive historical context can be gotten from Anderson (1985); see in particular pp. 270–276.

to be found and presented in the literature. The first issue of *Language* in 2020, 95 years after the publication of Sapir's article in the very first issue of the same journal, includes an article by William Labov arguing that the regularity of a sound change in Philadelphia is understandable to the extent that speakers of this dialect of English have an abstract mental representation of the front unrounded mid vowel (Labov, 2020).

Jumping back in time to observations by Bloomfield (1933) and Bloch (1941), among others: English unstressed vowels reduce to schwa [ə] in many contexts, making [ə] an allophone of every English vowel phoneme. Thus we have [ˈfo̲ɾəˌɡɹæ̲f] 'photograph', with a primary-stressed [o], an unstressed [ə], and a secondary-stressed [æ], alternating with [fə̲ˈtʰɑ̲ɡɹə̲ˌfi] 'photography', with primary-stressed [ɑ] flanked by unstressed [ə]s. It follows that phonemic representations of English morphemes must include unreduced vowels only, with reduction to [ə] being due to a phonological transformation, and moreover that the underlying form of many morphemes, such as /fotɑɡɹæf/ 'photograph', will not directly correspond to *any* of their complete surface manifestations due to the nature of stress assignment in English.[6]

Phonemic analysis is the set of methods by which a language's underlying inventory of phonemes is induced from the distribution and behavior of its surface speech sounds, which can of course differ from language to language. Borrowing an example from Hayes (2009, pp. 31–34), English and Spanish have a set of surface speech sounds that can be broadly transcribed as [t d ð ɾ],[7] but their distribution and behavior in each language is such that:

- /ɾ/ is a phoneme distinct from /t/ and /d/ in Spanish, but not in English, where [ɾ] is sometimes a surface realization of /t/ (e.g. [bæt] 'bat', [ˈbæɾɚ] 'batter') and other times a surface realization of /d/ (e.g. [sæd] 'sad', [ˈsæɾɚ] 'sadder'), and

- /ð/ is a phoneme distinct from /t/ and /d/ in English, but not in Spanish, where [ð] is always a surface realization of /d/ (e.g. [unˈdisko] 'a record', [losˈðiskos] 'the records').

From the same surface inventory of speech sounds [t d ð ɾ], then, we arrive a different phonemic inventory in each language: /t d ð/ in English and /t d ɾ/ in Spanish, with [ɾ] a conditioned variant of /t d/ in English and [ð] a conditioned variant of /d/ in Spanish.[8]

[6]Bloch (1941, pp. 281–283) observed that this represents a potential learning challenge, because arriving at the 'right' underlying form for a morpheme requires exposure to a sufficient variety of its surface manifestations.

[7]Narrower differences include the precise tongue tip position for the articulation of [t d] (alveolar in English, dental in Spanish) and the degree of constriction of [ð] (more close in English, more open in Spanish).

[8]Whether /d/ or /ð/ is the right representation of this phoneme in Spanish is a matter of some debate. Compare Harris (1969) with Baković (1994), with Lozano (1979) opting for underspecification of the difference between the two speech sounds.

A phonemic inventory is often presented as a stand-alone entity in the context of a phonological analysis, but phonologists recognize that individual phonemes more often than not have restrictions on their own distributions, much like their various surface manifestations do (Hall, 2013). For example, some argue /ŋ/ is a phoneme of English, but it is only found word-finally (e.g. [sɪŋ] 'sing'), before word-level suffixes (e.g. [ˈsɪŋɚ] 'singer', [ˈsɪŋɪŋ] 'singing'), or when followed by /k/ or /g/ ([sɪŋk] 'sink', [fɪŋgɚ] 'finger'). Similarly, /ð/ is a phoneme of English, but its distribution is heavily restricted. It is found at the beginning of a handful of function words, mostly determiners ([ðɪs], [ðæt], [ðiz], [ðoz], [ðɛm], [ði], [ðaɪ], [ðə], [ðɛn] 'this, that, these, those, them, thee, thy, the, then'), in a handful of words ending in [ðɚ] ([ˈʌðɚ], [ˈbɑðɚ], [ˈwɛðɚ], [ˈfɛðɚ], [ˈmʌðɚ], [ˈbɹʌðɚ], [ˈfɑðɚ] 'other, bother, weather, feather, mother, brother, father'), and at the end of a handful of verbs, mostly denominal ([bɹið], [beɪð], [ʃið], [ɹaɪð] 'breathe, bathe, sheathe, writhe').

An important development in 20th Century linguistics is the idea that the phoneme is not the minimal unit after all, and that instead there are subphonemic units — *distinctive features* — out of which phonemes are built. This idea was developed further by phonologists of the Prague School, notably Roman Jakobson and Nikolai Trubetzkoy, for whom *contrast* between two phonemes was a central theoretical premise. Analyzing the nature and systematicity of these oppositions required viewing phonemes as possessing features. The particular features necessary to distinguish one phoneme from others contribute to the content of that phoneme. In this sense,

> [a]ny minimal distinction carried by the message confronts the listener with a two-choice situation. Within a given language each of these oppositions has a specific property which differentiates it from all the others. The listener is obliged to choose either between two polar qualities of the same category [...] or between the presence and absence of a certain quality [...]. The choice between the two opposites may be termed *distinctive feature*. The distinctive features are the ultimate distinctive entities of language since no one of them can be broken down into smaller linguistic units. The distinctive features combined into one simultaneous or [...] concurrent bundle form a *phoneme*. (Jakobson et al., 1952, p. 3)

Phonological features were intended to be the cognitive connection between the articulatory and perceptual speech systems.

> [T]he distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems. In speech perception detectors sensitive to the property detectors [...] are activated, and appropriate information is provided to centers corresponding to the distinctive feature[s] [...]. This information is forwarded to higher centers in the nervous system where identification of the utterance takes place. In producing speech, instructions are sent from higher centers in the nervous system to the

> different feature[s] [...] about the utterance to be produced. The features then activate muscles that produce the states and configurations of different articulators[.] (Halle, 1983, p. 95)

Over a quarter century later, the same idea informs the neurolinguistics and biolinguistics literatures.

> The [...] featurally specified representation constitutes the format that is both the endpoint of perception — but which is also the set of instructions for articulation. (Poeppel and Idsardi, 2011, p. 179)

> Features serve as the cognitive basis of the bi-directional translation between speech production and perception, and are part of the long-term memory representation for the phonological content of morphemes, thus forming a memory-action-perception loop [...] at the lowest conceptual level. (Volenec and Reiss, 2017, p. 270)

Just as a phonemic analysis of a language's phonological patterns reveals its phonemic inventory, so does a featural analysis reveal its distinctive feature inventory. Because a phoneme consists of an individual combination of distinctive feature values, and because phoneme inventories are language-particular and potentially asymmetrical, there will be individual combinations of distinctive feature values that do not correspond to phonemes in a given language.[9] For example, the phonemic inventory of Russian includes voiceless oral stops /p t k/ and voiced oral stops /b d g/ at the same three places of articulation (labial, coronal, dorsal), but voiced nasal stops /m n/ at only two of these — meaning that the combination of distinctive feature values [+nasal, +dorsal] does not correspond to a phoneme in Russian.

Thus, the distribution (= recombination) of a distinctive feature can be restricted much like the distribution of a phoneme can be. These twin facts open the door to the use of restricted phonemes and distinctive feature combinations in the kinds of abstract analyses that command particular attention in phonology, to which we now turn.

5 How Abstract are Mental Representations?

The question posed in this section title has not attracted much attention in recent work in generative phonology, but it is very much alive today and lurks under the surface in many modern debates, such as the extent to which phonology can be reduced to physiological principles governing articulation and perception (Ohala, 1981, 1997; Hale and Reiss, 2000; Hayes et al., 2004; Blevins, 2004b; Heinz and Idsardi, 2013; Reiss, 2018).

Early work in generative phonology admitted the possibility of very deep derivations, with phonological forms potentially undergoing many phonological rules applied in crucial sequence such that the specifications of a given

[9]We put aside physically incompatible combinations, such as [+high, +low].

underlying representation could be quite a bit different from those of its eventual surface representation.[10] This potential led to concerns about the possibility of 'excessive abstractness', though different types of abstractness and how one is to measure them with respect to one another were often more matters of opinion than principle. When Kiparsky (1968) asked *how abstract is phonology?*, his proposed answer was thus not so much a line drawn in the sand as it was the curtailment of one particular type of abstractness. In other words, theoretical proposals may limit the *possible types of differences* that may hold between an underlying representation and its various surface manifestations; they do not somehow limit the *'distance'* between these representations, as has often been claimed.

The form of the evidence for an abstract analysis is usually one where what otherwise appears to be the same phoneme exhibits two distinctive forms of behavior, one expected (call this one A) and one unexpected (call this one B). The more strands of evidence of this sort that exist in any given analysis, the more compelling the case for it. Yawelmani (Kisseberth, 1969a), Nupe (Hyman, 1970), and Dida (Kaye, 1980) are particularly interesting cases. The unexpected behavior of B can be understood in one of three ways, all of them involving abstractness.

One way is to posit that B is a phoneme with abstract properties distinct from A. Relevant phonological transformations sensitive to the A/B distinction apply, and B later neutralizes on the surface with A. We call this the *neutralized abstract phoneme* approach.

The second way is to posit that A and B are phonemically identical, but that there is some other, abstract phoneme C, present in forms with B but not those with A, to which relevant phonological transformations are sensitive. Typically, C is later deleted. We call this the *deleted abstract phoneme* approach.

The third way is to posit that A and B are phonemically identical, but that there is some abstract lexical marking X, present in forms with B but not those with A, to which relevant phonological transformations are sensitive. Typically, X has no other function. We call this the *abstract lexical marking* approach.

We begin with the result that Kiparsky's proposed principle was meant to ensure: that every phoneme specified as a particular combination of distinctive features in a given morpheme's underlying form will be realized with that precise set of specifications in some surface form of that morpheme.[11] This principle excludes both of the abstract phoneme approaches defined above, with each of these being of two types.

One type is what we we will here call *absolute abstractness*: a combi-

[10] Or very similar, in some respects, by way of a there-and-back-again sequence of changes known as a *Duke of York derivation* (see Pullum 1976; McCarthy 2003; Baković 2013 for discussion).

[11] We put aside the (im)precise formulation of Kiparsky's principle, as well as its ensuing revisions (Kiparsky, 1982, 1993), and focus instead on its intended function.

nation of distinctive feature values posited as a phoneme in some set of morphemes but that is not realized as such in the surface forms of *any morphemes in the language*; Crothers (1971, p. 3) thus refers to absolutely abstract phonemes as 'imaginary segments'. These can be divided further into absolutely abstract phonemes that are deleted and those whose distinctive feature specifications are altered, in both cases after having some crucial effect in the derivation along the way toward being deleted or neutralized.

An example of a deleted absolutely abstract phoneme appears in Chomsky and Halle (1968, p. 234): the absolutely abstract phoneme /x/ is posited to be present in a set of English morphemes that unexpectedly fail to undergo an otherwise regular vowel shortening transformation; after successfully blocking shortening, the absolutely abstract phoneme deletes.

An example of a neutralized absolutely abstract phoneme appears in the well-known analysis of Yokuts vowels (Kuroda, 1967; Kisseberth, 1969a,b; Kenstowicz and Kisseberth, 1979): the absolutely abstract phoneme /uː/ is posited to be present in a set of morphemes in which the vowel ultimately surfaces as [oː] (or [o], due to independently conditioned vowel shortening). This phoneme both spreads lip rounding to following *i* and fails to spread it to following *a*, which is the expected behavior of /uː/ but not /oː/.[12]

Another type of abstractness disallowed by Kiparsky's principle is what we will call *relative abstractness*: a combination of distinctive feature values posited as a phoneme in some set of morphemes but that is not realized as such in the surface forms of *any morphemes in that set*. These can also be divided further into relatively abstract phonemes that are deleted and those whose distinctive feature specifications are neutralized, again in both cases after having some crucial effect in the derivation.

An example of a deleted relatively abstract phoneme appears in an analysis of exceptions to vowel coalescence in Mushunguli (Hout, 2017). The relatively abstract phonemes /j w/ are posited to begin a set of stems that in all of their surface forms begin with /i u/ (respectively). These stems unexpectedly block otherwise regular coalescence with preceding /a/, after which these relatively abstract phonemes are deleted. The speech sounds [j w] otherwise exist in Mushunguli, but never before [i u] (again, respectively), so the transformation deleting the relatively abstract phonemes can be stated in regular, phonologically-conditioned terms.

A somewhat more complex example, involving both a neutralized relatively abstract phoneme *and* a deleted relatively abstract phoneme, appears in the analysis of Spanish rhotics of Harris (1969, 1983, 2001, 2002); for a review, see Baković (2009). There are two surface rhotics in Spanish, a tap [ɾ] and a trill [r]. Harris adduces several arguments for representing the trill as a sequence of two taps underlyingly, with a transformation converting

[12]Some have argued that the Yokuts case is, at best, an example of a neutralized *relatively* abstract phoneme (Hockett, 1967, 1973; Blevins, 2004a; Weigel, 2005), as defined below in the text.

the second tap to a trill and another transformation deleting the first tap. In some cases the abstract sequences of two taps are derived by the concatenation of a morpheme ending in a single tap with another morpheme beginning with a single tap; in these cases, other surface realizations of both morphemes will have taps. Otherwise, all surface realizations of morphemes posited to contain sequences of two taps underlyingly will have trills.

For any given example like those sketched above, there is an alternative analysis that does not involve positing an abstract phoneme at all, relying instead on abstract lexical marking — otherwise contentless specifications that distinguish those morphemes exhibiting unexpected behavior from those that don't. Kiparsky (1968) usefully distinguishes "the diacritic use of phonological features" (= abstract phonemes) and "the phonological use of diacritic features" (= abstract lexical marking). He is clearly in favor of the latter, and his proposed principle strongly curtails the former, but both clearly involve abstractness of some kind.

We would be remiss if we didn't mention here something that all abstract phoneme analyses crucially rely on: *opaque interactions* between phonological transformations (Kiparsky, 1973). In each of the examples sketched above, the deletion or neutralization transformation that voids the representation of its abstract phoneme crucially must not apply before application of the relevant transformation(s) sensitive to the abstract phoneme.[13]

In some cases, the abstract phoneme is needed in the representation to prevent the application of an otherwise applicable transformation (e.g., English /x/, Mushunguli /j u/). These involve the type of opaque interaction known as *counterfeeding*. In other cases, the abstract phoneme is needed in the representation to ensure the application of an otherwise inapplicable transformation (e.g., the first of a sequence of taps in Spanish). These involve the type of opaque interaction known as *counterbleeding*.[14]

The question of abstractness is thus intimately intertwined with the question of opacity: to the extent that a theoretical framework (dis)allows opaque interactions, it (dis)allows abstract phoneme analyses.

Finally, this section and the last focused on the issue of phonological abstractness in the mental lexicon at the level of the phoneme and of the distinctive feature. This is primarily because this is where most research has focused, with the possible exception of the syllable (Goldsmith, 2011; Strother-Garcia, 2019). However, other representational levels and systems also have been argued to play a role in the mental representations of words,

[13]The 'must not apply before' as opposed to 'must apply after' wording recognizes that opaque interactions can be had with simultaneous as opposed to ordered application of phonological transformations (Kenstowicz and Kisseberth, 1979; Joshi and Kiparsky, 1979, 2006; Kiparsky, 2015).

[14]Note that in the Yokuts case, the abstract phoneme /uː/ is needed in the representation to both counterfeed *and* counterbleed a relevant transformation: it both prevents spreading of lip rounding to a (counterfeeding) and ensures spreading of lip rounding to *i* before neutralizing with /oː/ (counterbleeding).

notably tonal representations (Yip, 2002; Jardine, 2016) and metrical stress (Hayes, 1995; van der Hulst, 2013). These representations have also been the subject of intense study, including the psycholinguistic and neurolinguistic literature, to which we turn in the next section.

6 Types of Evidence

6.1 Synchronic and Diachronic patterning

The preceding sections discussed the perspective of phonological abstractness using structural linguistic arguments from the typology of spoken language. However, it is often claimed that structural arguments are far from a definitive proof, constituting one type of evidence. This is especially true for the psychological reality of these abstract forms. Ohala (1974) makes this point:

> It seems to me that the important question should not be whether phonology is abstract or concrete, but rather what evidence there is for the psychological reality of a particular posited underlying form. If our aim is simply a descriptive one, then of course abstract forms can be posited without any further justification than the fact that they make it easier to state certain regularities. However, if we are interested in reflecting the competence of a native speaker, then we must provide evidence that what we are claiming as part of a speaker's knowledge is indeed such.

What other types of evidence are there? This section describes various other types of evidence bearing on the question of abstractness. The first type of evidence comes from expanding the typology to consider non-spoken language, that is, signed and tactile language. This evidence is still structural, but provides an important window into the lexicon. How dependent is the long-term memory representation of a word on the physical system which externalizes it? At the same time, Ohala and many others take behavioral and psychological tests as providing additional necessary, if not sufficient, evidence for the reality of these forms. Recent advances in neuroimaging and increasing ease of use allow for intricate looks into the biological inderpinnings of phonological representations, combined with computational models and simulations. Howeverm literature in all these topics is vast, especially with regard to experimental work. Here we provide a sample of work that bears on the questions described earlier.

6.2 Evidence from Signed and Tactile Phonology

An important source of evidence for the existence and nature of phonological abstraction comes from languages without a spoken modality - namely, signed and tactile languages. Sign languages arise spontaneously in deaf communities, are acquired during childhood through normal exposure without instruction, and exhibit all of the facets and complexity found in spoken languages (see Sandler and Lillo-Martin (2006) for a groundbreaking overview). However, if human language evolved without respect to modality, we should find hearing communities that just happen to use sign language instead of spoken language, and we do not. Sign languages are thus "an adaptation of existing physical and cognitive systems for the purpose of communication among people for whom the auditory channel is not available" (Sandler, 1993). This holds equally true for language expressed by the DeafBlind through the tactile modality, often called tactile or pro-tactile sign languages (see (Edwards, 2014) for a recent phonological analysis).

Sign languages offer, as Sandler (1993) puts it, "a unique natural laboratory for testing theories of linguistic universals and of cognitive organization." They give insight into the contents of phonological form, and conditions on which aspects of grammar are amodal and which are tied to the modality. They also offer unique opportunities to study the emergence of phonology under various conditions: where home-signers (Goldin-Meadow, 2005) are congregated and a sign language emerges among themselves, as in Nicaragua (Senghas et al., 2004); and where a localized hereditary deaf population lives among hearers who also sign, as in Bali (Marsaja, 2008) or in a Bedouin group in Israel (Sandler et al., 2005).

One crucial contribution of non-spoken phonology is that it switches the issue of "how abstract is phonology?" to "where does abstraction lie, and to what extent is it independent of the modality?" There are generally two directions answers can take. To the extent that a given phonological form or constraint is present across modalities, one may make the case that it is truly abstract, in the sense that it exists without regard to the articulatory system which realizes it. On the other hand, to the extent that a given form or constraint differs, one can ascribe that difference to the modality, to the nature of the articulatory system. In this way, non-spoken phonology provides nuance into the relationship between the pressures of abstraction and the pressures of realization in the mental lexicon.

Up until 1960, sign languages were simply not considered as fully-fledged natural languages possessing morpho-syntactic structure, let alone an independent level of phonological structure (see (van der Hulst, pear) for the history of sign phonology). Recognition of phonological compositionality came from the groundbreaking work of (Stokoe, 1960). His proposed segmentation of signs into meaningless chunks implied that sign languages display both morpho-syntactic and phonological levels, a "duality of patterning" or "double articulation" which had long been identified as unique properties of spoken languages (Martinet, 1960; Hockett, 1960).

Stokoe's phonological system contained a finite number of abstractions for what he perceived as separate parts or 'aspects' of the sign: the handshape, the movement of the hand, and the location in front of or on the body. Insisting that the study of sign language form should not be seen through the prism of spoken language phonology, Stokoe named the study of the perceptible form of signs CHEROLOGY, and called the abstract basic units (handshape, movement, etc.) CHEREMES, from the Greek root *cher* meaning 'hand'. From one angle, Stokoe's partitioning of the form of signs was purely phonetic, designed to allow a compositional space-saving system of phonetic transcription that would replace holistic drawings. However, Stokoe also regarded the symbols as representations of the discrete units that characterize the signs.

Groundbreaking work on the psychological reality of signed abstractions came from the Salk Institute during the 1970s (see (Klima and Bellugi, 1979) for an accessible overview, and Emmorey (2001) for more recent developments). Performing recall experiments like those done in spoken languages, they showed that percepts of signs in short-term memory are compositional. Studying production errors or 'slips of the hands', they argued for the compositionality in the articulatory phase. Comparing American and Japanese Sign Language, they showed that ASL signers make judgments about what they considered well-formed or ill-formed. Such judgements presuppose knowledge of how smaller units can be combined. This knowledge is captured in phonotactic constraints listing the basic units as well as their permissible combinations.

While Stokoe's groundbreaking work stressed the simultaneity of the units that constitute a sign, later researchers (Liddell and Johnson, 1989; Liddell, 1984; Newkirk, 1981; Supalla and Newport, 1978) argued that it was necessary to reference the beginning and end points of the movement of signs, for example for inflectional purposes (seen in the ASL signs I-GIVE-YOU as opposed to YOU-GIVE-ME, where the referents are marked by discrete locations), or to express assimilations involving a switch in the beginning and end point of the movement (see Sandler (1989) and van der Hulst (1993) for discussion).

These considerations resulted in signed words containing abstract timing units, forming a temporal skeleton to which the content units (features) of the sign associate in an many-to-one fashion (often called autosegmental in linguistics (Goldsmith, 1976a). Most researchers (Liddell, 1984; Liddell and Johnson, 1989; Sandler, 1989, 1993; Perlmutter, 1993; Brentari, 1998) proposed an abstract sequence that temporally represented both initial and final location, as well as an intermediary movement unit.

The notion of simultaneous and compositional structure in a sign, as well as sequential structure, leaves a big question: how modality-dependent are these properties? Languages in both modalities have sequential and simultaneous structure, but there are striking differences in the degree and centrality of such structure. Spoken languages vary in syllable structure, word length, and stress patterns among syllables. Sign languages appear limited in all these aspects. They are overwhelmingly monosyllabic, have no clusters, and show extremely simple stress patterns, due to few polysyllabic words apart from fully reduplicated forms (Wilbur, 2011).

Additionally, the structural organization in signed or spoken language has a direct effect on the phonology. Strikingly few phonological rules that are not morphosyntactically triggered have been discovered in sign languages, mainly due to sign's lack of sequential structure. Significant sequential structure in sign mainly appears under morphosyntactic operations that concatenate morphemes and words (affixation, compounding, and cliticization) (Aronoff et al., 2005). Predictably, when this occurs, a smorgasbord of phonology arises.

In general, sequential affixation is rare across sign languages (Sandler, 1996; Aronoff et al., 2005), and sign exhibits a strong tendency to express concatenative morphology through compounding (Meir, 2012). Aronoff et al. (2005) show that affixation usually results from grammaticalization of free words, via a series of diachronic changes concerning phonological and semantic factors. They cite the relative youth of sign languages as causing their lack of affixes compared with verbal languages. No known sign languages are over 300 years old, with some like Nicaraguan Sign Language, as young as 40 (Woll et al., 2001).

The lack of sequential structure in sign languages does not imply structural simplicity, however. Sign languages routinely employ nonconcatenative morphology (Emmorey, 2001; Meier, 2002), incorporating morphological material simultaneously in the phonology with restricted sequential form. Of course, simultaneous phonological structure exists in all languages, but differ across modalities in the amount. Very few sign features actually become sequenced, while in spoken language features are overwhelmingly sequenced, rarely simultaneous. Comparing hand configuration and place autosegments to autosegments in spoken language shows further differences. Unlike spoken autosegments for tone or harmony patterns, which typically consist of one or two features, the hand configuration autosegment in sign languages is extremely complex, containing almost half of sign features organized in an intricate feature geometry (van der Hulst, 1995; Sandler, 1996).

In sum, the phonological module may leverage the representational abilities of the particular articulatory/perceptual system. van der Hulst and van der Kooij (2018) cite Brentari (2002) and Emmorey (2001) that visual perception of signs (even with sequential properties) is more "instantaneous" than auditory speech perception, and adapt Goldsmith (1976b)'s division of phonology in terms of the notions of "vertical and horizontal slicing of the signal". They state

> an incoming speech signal is first spliced into vertical slices, which gives rise to a linear sequence of segments. Horizontal slicing then partitions segments into co-temporal feature classes and features. In the perception of sign language, however, the horizontal slicing takes precedence, which gives rise to the simultaneous class nodes that we call handshape, movement, and place. Then, a subsequent vertical slicing of each of these can give rise to a linear organization

Of course, the phonetics and phonology of sign language differ in many ways, and this isn't surprising. Lillo-Martin (1997) cites Blakemore (1974)'s result that exposure to vertical and horizontal lines in the environment affects development of feline visual perception, and asks "why shouldn't exposure to the special acoustic properties of the modality affect perception, especially auditory perception?" Sandler and Lillo-Martin (2006) note that unlike spoken syllables in many languages, sign language syllables prohibit location clusters comparable to consonant clusters, or diphthong-like movement clusters, and there must be a movement between locations due to the physiology of the system. Additionally, sign syllables do not have onset-rhyme asymmetries, which affects syllable structure and stress assignment, they typically align intonation, conveyed by facial expression, with phonological/intonational phrases, not syllables inside those phrases, where spoken languages usually do (Nespor and Sandler, 1999). The similarities and differences in phonological abstraction across modalities means signed languages will have an important role as evidence.

6.3 Psycholinguistic and neurolinguistic evidence

As mentioned, behavioral testing has long been argued as necessary evidence for the mental reality of phonological abstraction, in addition to typological evidence. The introduction and improvement of neuroimaging methods enabled links between behavioral tasks and gross neural excitation levels associated with them. In addition, recent simulation tools allow for modeling of phonological representations in an idealized *in silico* setting. These methods led to an explosion of results in uncovering the roots of phonological abstraction. Here we overview several results bearing on the segmentation of speech into different abstract timescales, as well as the organization of speech by its salient features into abstract phonemes.

Speech production engages a vastbuted neuronal ensemble. Eickhoff et al. (2009) report speech-related activation in left inferior frontal gyrus (IFG), ventral precentral gyrus (motor and premotor cortex), ventral postcentral gyrus (somatosensory cortex), superior temporal gyrus (STG), supplementary motor area (SMA), anterior insula, superior paravermal cerebellum, basal ganglia and thalamus. A review of data from 82 neuroimaging experiments (Indefrey and Levelt, 2004) determined that phonological operations occur within the average time window of 205 ms, followed by an average of 145 ms of phonetic processing. Important structures for this temporal organization of speech include the cerebellum and basal ganglia. Insula information is pushed to the cerebellum and basal ganglia, well-established constituents of cortical-subcortical loops for movement preparation (Jueptner and Krukenberg, 2001).

More specifically, the sequencing of motor programs for articulation is filtered through through basal ganglia, while the cerebellum converts these discrete sequences into a fluent, temporally distributed action (Eickhoff et al., 2009). Cerebellar dysfunction affects temporal aspects of speech production, resulting in dysarthria characterized by improper timing of cognitively discrete elements, substantial aberrations in their total and relative duration, disrupted coordination of orofacial and laryngeal movements, as well as slowed/delayed execution of articulatory movements. (Ackermann et al., 2007).

One issue regards the biological mechanism underlying the temporal segmentation of the signal produced and perceived. Neurobiological studies on speech perception have uncovered that the human perceptual system consistently uses two time scales to analyze a continuous speech signal, a segmental time-frame of roughly 10–80 ms, and a syllabic time-frame of 100–500 ms (Poeppel and Hackl, 2008; Poeppel and Idsardi, 2011; Chait et al., 2015):

> These two critically important windows that appear instantiated in spoken languages: segments and syllables. Temporal coordination of distinctive features overlapping for relatively brief amounts of time (10–80 ms) comprise segments; longer coordinated movements (100–500 ms) constitute syllabic prosodies. (Poeppel and Idsardi, 2011, 182)

A more fundamental question concerns the neural mechanism which drives these windows and their coordination. Oscillatory activity within and between neural populations has been posited (Giraud and Poeppel, 2012). Neural populations which comprise a certain type of neuron may show a stable neural oscillation at certain frequency bands which varies depending on their excitatory and inhibitory properties (see (Buzsaki, 2006) for an accessible overview). Evidence suggests that pyramidal interneuron gamma oscillations, as well as theta oscillations, comprise the segmental vs. syllabic time distinction. These two oscillations funnel an incoming speech signal into time windows of different sizes, computationally represented by the waveform of the population activity.

In silico modeling work reveals another interesting property. While these oscillations do track the signal, a crucial feature is that rhythms of distinct frequencies show specific coupling properties, termed cross-frequency coupling (Hyafil et al., 2015b). Briefly, this means that the stable oscillatory populations innervate each other, allowing different timescales to track one another, effectively parsing a temporally complex signal efficiently. Specifically for speech perception, Hyafil et al. (2015a) showed that when a network showing gamma oscillations coupled to a network showing theta osciallations, it was able to effectively segment a corpus of phonological words much better than a network where this coupling was absent.

Another salient question concerns neural evidence for abstract phonemic and featural represnetations of words. Kazanina et al. (2006) show using magnetoencephalography (MEG) that Russian and Korean speakers react differently to tokens of [d] vs. [t]. In Russian, these sounds are contrastive, members of different phonemes, /d/ and /t/. In Korean, the sounds are not contrastive and map into a single phoneme /T/. Russian speakers separated the sounds into two categories, whereas Korean speakers did not. Kazanina et al. (2006) conclude from this that perceptual space is shaped not only by the phonetic distribution of sounds, but also by a more abstract phonemic analysis of speech sounds.

Much evidence suggests concludes that various portions of the superior temporal sulcus (STS) directly encode phonological representations. Mesgarani et al. (2014) showed that acoustic phonetic information is represented in the STS, distributed along five distinct areas, each roughly corresponding to a general 'manner of articulation' parameter of speech sounds. Using implanted electrical cortical grids placed along the temporal gyrus, they found one electrode responded selectively to stops, one to sibilant fricatives, one to low back vowels, one to high front vowels and a palatal glide, and one to nasals. Similarly, Bouchard et al. (2013) constructed an auditory-based 'place of articulation' cortical map in the STG, confirming labial, coronal and dorsal place features, cutting across various manner classifications. Scharinger et al. (2012) used MEG to localize three vowel feature variables (height, frontness and roundness comprised by the first three formants) to the superior temporal gyrus. Intriguingly, Scharinger et al. (2011) used a combination of MEG imaging and statistical modeling to mapping the entire vowel space of a language (Turkish) onto three-dimensional cortical space described by lateral-medial, anterior–posterior, and inferior-superior axes. Their statistical model comparisons showed that, whereas cortical vowel maps reflect acoustic properties of the speech signal, articulator-based and featural speech sound information "warps the acoustic space toward linguistically relevant categories". Hickok and Poeppel (2007) hypothesize that "the crucial portion of the STS that is involved in phonological-level processes is bounded anteriorly by the most anterolateral aspect of Heschl's gyrus and posteriorly by the posterior-most extent of the Sylvian fissure".

There is also interesting recent evidence from Buchwald and Miozzo (2011, 2012) using aphasic patients that also supports a level of abstract phonological processing . The English words *pill* and *spill* are assumed to contain the segment /p/ in their underlying representations. In the surface representation *pill* has [p^h] and *spill* has [p]. Buchwald and Miozzo (2011) measured voice-onset-time (VOT) productions of two aphasic patients unable to realize /s/ in clusters like /sp/, /st/, /sk/, comparing them with correctly produced consonants. The question was, when an aphasic simplifies the cluster in *spill*, will they aspirate the /p/, consistent with the deletion of /s/ occurring phonologically, or will it be produced without aspiration, meaning that the phonological mapping /sp/ $\rightarrow$ [sp] was intact and the fricative deleted post-phonologically?

Their results showed two different patterns. One patient's initial stop consonant had a long VOT ([p^h]), while the other had a short VOT ([p]). This evidence suggests that first patient's errors were phonological, and the other's were phonetic, "consistent with an account of spoken production containing at least two processing levels that can be selectively impaired by brain damage: one processing stage with context independent representations and another with context-specific representations"(Buchwald and Miozzo, 2011). Another study showed similar patterns for durational properties of nasal consonants which deleted in /sn/ and /sm/ clusters (Buchwald and Miozzo, 2012).

These results reflect an explosion of work using neurolinguistic and psycholinguistic tests to describe the sort of representations speakers have. For further work on the particular biological substrate of speech perception and production, refer to (Phillips et al., 2000; Binder et al., 2000; Hickok and Poeppel, 2004, 2007; Indefrey and Levelt, 2004; Obleser et al., 2004; Mesgarani et al., 2008, 2014; Idsardi and Raimy, 2013; Monahan et al., 2013; Idsardi and Monahan, 2016; Hickok and Poeppel, 2016; Poeppel and Hackl, 2008; Poeppel and Monahan, 2008; Poeppel et al., 2007; Poeppel and Idsardi, 2011; Blumstein and Baum, 2016; Guenther and Hickok, 2016; Tremblay et al., 2016). The intersection of experimental results with theory promises many new insights into the mental content of the lexicon.

7 Conclusion

In this chapter, we have endeavored to motivate and review a central idea of modern generative phonology that the fundamental representational units of words in languages are abstract and psychologically real. In particular the systematic patterning of the pronunciation of morphemes motivates an abstract mental representation of a morpheme's pronunciation. These underlying forms are largely regarded as sequences of phonemes, which are themselves abstractions, and which are organized along featural dimensions. These abstract mental representations find support not only from the patterning in morpho-phonological paradigms, but also from language change, from sign language linguistics, and from psycholinguistic and neurolinguistic study.

There are many open and fascinating questions regarding the nature of abstract mental representations of words. Many are among the most basic and fundamental. How abstract can they be? How are they learned? How are they realized in the brain? The fact that we simultaneously know both so much and so little about phonological abstractness in the mental lexicon sends a clear message that this will continue to be a fertile and exciting area of research for many years to come.

To conclude this chapter, we can do no better than to repeat the concluding sentences of Labov (2020, pp. 57, emphasis added): "We have a common ground in our understanding of what it means to know a language. It involves knowing a vast number of particular things. *But at bottom it is reaching down to something very deep, very abstract, and very satisfying.*"

References

Ackermann, H., Mathiak, K., and Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *The Cerebellum*, 6(3):202–213.

Anderson, S. (1985). *Phonology in the Twentieth Century*. The University of Chicago Press.

Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language*, 81(2):301.

Baković, E. (1994). Strong onsets and Spanish fortition. In Giordano, C. and Ardron, D., editors, *Proceedings of the 6th Student Conference in Linguistics*, pages 21–39. MIT Working Papers in Linguistics, Cambridge, MA. ROA-96, Rutgers Optimality Archive, http://roa.rutgers.edu.

Baković, E. (2009). Abstractness and motivation in phonological theory. *Studies in Hispanic and Lusophone Linguistics*, 2:183–198.

Baković, E. (2013). *Blocking and Complementarity in Phonological Theory*. Equinox, London.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral cortex*, 10(5):512–528.

Blakemore, C. (1974). Developmental factors in the formation of feature extracting neurons. *The neurosciences: Third study program*, pages 105–113.

Blevins, J. (2004a). A Reconsideration of Yokuts Vowels. *International Journal of American Linguistics*, 70(1):33–51.

Blevins, J. (2004b). *Evolutionary Phonology*. Cambridge University Press.

Bloch, B. (1941). Phonemic overlapping. *American Speech*, 16(4):278–284.

Bloomfield, L. (1933). *Language*. New York: H. Holt and Co.

Blumstein, S. E. and Baum, S. R. (2016). Neurobiology of speech production: Perspective from neuropsychology and neurolinguistics. In *Neurobiology of language*, pages 689–699. Elsevier.

Bouchard, K. E., Mesgarani, N., Johnson, K., and Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441):327.

Brentari, D. (1998). *A prosodic model of sign language phonology*. Mit Press.

Brentari, D. (2002). Modality differences in sign language phonology and morphophonemics. *Modality and structure in signed and spoken languages*, pages 35–64.

Buchwald, A. and Miozzo, M. (2011). Finding levels of abstraction in speech production: Evidence from sound-production impairment. *Psychological Science*, 22(9):1113–1119.

Buchwald, A. and Miozzo, M. (2012). Phonological and motor errors in individuals with acquired sound production impairment. *Journal of Speech, Language, and Hearing Research*.

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Chait, M., Greenberg, S., Arai, T., Simon, J. Z., and Poeppel, D. (2015). Multi-time resolution analysis of speech: evidence from psychophysics. *Frontiers in neuroscience*, 9:214.

Cheng, E. (2015). *How to Bake Pi: An Edible Exploration of the Mathematics of Mathematics*. Basic Books, New York.

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

Coleman, J. (2002). Phonetic representations in the mental lexicon. In Durand, J. and Laks, B., editors, *Phonetics, phonology and cognition*, pages 96–130. Oxford University Press, Oxford.

Crothers, J. (1971). On the abstractness controversy. UC Berkeley Department of Linguistics, Project on Linguistic Analysis Reports, Second Series, No. 12, CR1-CR29.

Dresher, E. B. (2011). The phoneme. In Marc van Oostendorp, Colin J. Ewen, E. H. and Rice, K., editors, *The Blackwell Companion to Phonology*, volume 1, pages 241–266. Wiley-Blackwell, Malden, MA & Oxford.

Edwards, T. (2014). *Language Emergence in the Seattle DeafBlind Community*. PhD thesis, University of California, Berkeley.

Eickhoff, S. B., Heim, S., Zilles, K., and Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1896):2399–2421.

Emmorey, K. (2001). *Language, cognition, and the brain: Insights from sign language research*. Psychology Press.

Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15(4):511.

Goldin-Meadow, S. (2005). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language.* Psychology Press.

Goldsmith, J. (1976a). *Autosegmental Phonology.* PhD thesis, MIT, Cambridge, MA.

Goldsmith, J. (2011). The syllable. In Goldsmith, J. A., Riggle, J., and Yu, A. C. L., editors, *The Blackwell Handbook of Phonological Theory*, pages 164–196. Wiley-Blackwell.

Goldsmith, J. A. (1976b). *Autosegmental phonology*, volume 159. Indiana University Linguistics Club Bloomington.

Guenther, F. H. and Hickok, G. (2016). Neural models of motor speech control. In *Neurobiology of Language*, pages 725–740. Elsevier.

Hale, M. and Reiss, C. (2000). Substance abuse and dysfunctionalism: Current trends in phonology. *Linguistic Inquiry*, 31:157–169.

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, 30(2):215–275.

Halle, M. (1983). On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory*, 1:91–105. Reprinted in Halle (2002), pp. 105–121.

Halle, M. (2003). *From Memory to Speech and Back: Papers on Phonetics and Phonology 1954–2002.* De Gruyter Mouton, Berlin, Boston.

Harris, J. W. (1969). *Spanish phonology.* Cambridge, Mass.: Massachusetts Institute of Technology Press.

Harris, J. W. (1983). Syllable structure and stress in spanish. a nonlinear analysis. *Linguistic Inquiry Monographs Cambridge, Mass.*, (8):1–158.

Harris, J. W. (2001). Reflections on *A Phonological Grammar of Spanish*. In Herschensohn, J., Mallen, E., and Zagona, K., editors, *Features and interfaces in Romance*, pages 133–145. Benjamins, Amsterdam.

Harris, J. W. (2002). Flaps, trills, and syllable structure in Spanish. *MIT Working Papers in Linguistics*, 42:81–108.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* University of Chicago Press.

Hayes, B. (2009). *Introductory Phonology.* Wiley-Blackwell.

Hayes, B., Kirchner, R., and Steriade, D., editors (2004). *Phonetically-Based Phonology.* Cambridge University Press.

Heinz, J. and Idsardi, W. (2013). What complexity differences reveal about domains in language. *Topics in Cognitive Science*, 5(1):111–131.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2):67–99.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5):393.

Hickok, G. and Poeppel, D. (2016). Neural basis of speech perception. In *Neurobiology of Language*, pages 299–310. Elsevier.

Hockett, C. (1960). The origin of speech. *Scientific American*, 203(3):88–97.

Hockett, C. F. (1967). The Yawelmani Basic Verb. *Language*, 43(1):208–222.

Hockett, C. F. (1973). Yokuts as Testing-Ground for Linguistic Methods. *International Journal of American Linguistics*, 39(2):63–79.

Hout, K. (2017). Exceptions to hiatus resolution in Mushunguli (Somali Chizigula). In Kandybowicz, J. and Torrence, H., editors, *Africa's Endangered Languages: Documentary and Theoretical Approaches*. Oxford University Press.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., and Giraud, A.-L. (2015a). Speech encoding by coupled cortical theta and gamma oscillations. *Elife*, 4:e06213.

Hyafil, A., Giraud, A.-L., Fontolan, L., and Gutkin, B. (2015b). Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. *Trends in neurosciences*, 38(11):725–740.

Hyman, L. (1970). How concrete is phonology? *Language*, 46(1):58–76.

Idsardi, W. and Raimy, E. (2013). Three types of linearization and the temporal aspects of speech. *Challenges to linearization*, 114:31–56.

Idsardi, W. J. and Monahan, P. J. (2016). Phonology. In *Neurobiology of language*, pages 141–151. Elsevier.

Indefrey, P. and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1-2):101–144.

Jakobson, R., Gunnar, C., Fant, M., and Halle, M. (1952). *Preliminaries to Speech Analysis*. MIT Press.

Jardine, A. (2016). *Locality and non-linear representations in tonal phonology*. PhD thesis, University of Delaware.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4):485 – 499. Modelling Sociophonetic Variation.

Joshi, S. D. and Kiparsky, P. (1979). Siddha and asiddha in Paninian phonology. In Dinnsen, D. A., editor, *Current Approaches to Phonological Theory*, pages 223–250. Indiana University Press, Bloomington, IN.

Joshi, S. D. and Kiparsky, P. (2006). The extended *siddha*-principle. *Annals of the Bhandarkar Oriental Research Institute 2005*, pages 1–26.

Jueptner, M. and Krukenberg, M. (2001). Motor system: cortex, basal ganglia, and cerebellum. *Neuroimaging Clinics of North America*, 11(2):203–19.

Kaye, J. (1980). The mystery of the tenth vowel. *Journal of Linguistic research*, 1:1–14.

Kazanina, N., Phillips, C., and Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103(30):11381–11386.

Kenstowicz, M. and Kisseberth, C. (1979). *Generative Phonology: Description and Theory*. Academic Press, New York.

Kiparsky, P. (1968). How abstract is phonology? Manuscript published in O. Fujimura (ed.), *Three Dimensions of Linguistic Theory*, pp. 5–56, TEC, Tokyo, 1973. Reprinted in *Explanation in Phonology*, pp. 119–163, Foris, Dordrecht, 1982.

Kiparsky, P. (1973). Abstractness, opacity and global rules. In Fujimura, O., editor, *Three Dimensions of Linguistic Theory*, pages 57–86. Tokyo: TEC. Part 2 of "Phonological representations".

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In van der Hulst, H. and Smith, N., editors, *The Structure of Phonological Representations, Pt. 1*, pages 131–175. Foris, Dordrecht.

Kiparsky, P. (1993). Blocking in nonderived environments. In Hargus, S. and Kaisse, E. M., editors, *Studies in Lexical Phonology*, volume 4 of *Phonetics and Phonology*, pages 277–313. Academic Press, San Diego, CA.

Kiparsky, P. (2015). Stratal OT: A Synopsis and FAQs. In Hsiao, Y. E. and Wee, L.-H., editors, *Capturing Phonological Shades Within and Across Languages*, pages 2–44. Cambridge Scholars Library, Newcastle upon Tyne.

Kisseberth, C. W. (1969a). On the abstractness of phonology: The evidence from Yawelmani. *Papers in Linguistics*, 1(2):248–282.

Kisseberth, C. W. (1969b). *Theoretical Implications of Yawelmani Phonology*. Doctoral dissertation, University of Illinois, Champaign-Urbana, IL.

Klima, E. S. and Bellugi, U. (1979). *The Signs of Language*. Harvard University Press.

Kuroda, S.-Y. (1967). *Yawelmani Phonology*. MIT Press, Cambridge, MA.

Labov, W. (2020). The regularity of regular sound change. *Language*, 96(1):42–59.

Liddell, S. K. (1984). Think and believe: sequentiality in american sign language. *Language*, pages 372–399.

Liddell, S. K. and Johnson, R. E. (1989). American sign language: The phonological base. *Sign language studies*, 64(1):195–277.

Lillo-Martin, D. (1997). The modular effects of sign language acquisition. *Relations of language and thought: The view from sign language and deaf children*, pages 62–109.

Lozano, M. d. C. (1979). *Stop and Spirant Alternations: Fortition and Spirantization Processes in Spanish Phonology*. Doctoral dissertation, Ohio State University, Columbus, OH. Distributed by the Indiana University Linguistics Club, Bloomington.

Marsaja, I. G. (2008). *Desa Kolok: A deaf village and its sign language in Bali, Indonesia*. Ishara Press.

Martinet, A. (1960). Eléments de linguistique générale, paris, a. *Armand Colin, 1re éd*.

McCarthy, J. J. (2003). Sympathy, cumulativity, and the Duke-of-York gambit. In Féry, C. and van de Vijver, R., editors, *The Syllable in Optimality Theory*, pages 23–76. Cambridge University Press, Cambridge.

Meier, R. P. (2002). Why different, why the same? explaining effects and non-effects of modality upon linguistic structure in sign and speech. *Modality and structure in signed and spoken languages*, pages 1–25.

Meir, I. (2012). Word classes and word formation. *Sign language. An international handbook. Berlin: De Gruyter Mounton*, pages 77–112.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174):1006–1010.

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*, 123(2):899–909.

Monahan, P. J., Lau, E. F., and Idsardi, W. J. (2013). Computational primitives in phonology and their neural correlates. *The Cambridge Handbook of Biolinguistics*, pages 233–256.

Nespor, M. and Sandler, W. (1999). Prosody in israeli sign language. *Language and Speech*, 42(2-3):143–176.

Newkirk, D. (1981). On the temporal segmentation of movement in asl. *Unpublished manuscript, The Salk Institute for Biological Studies*.

Obleser, J., Lahiri, A., and Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, 16(1):31–39.

Odden, D. (2014). *Introducing Phonology*. Cambridge University Press, 2nd edition.

Ohala, J. (1981). The listener as a source of sound change. In Masek, C., Hendrik, R., and Miller, M., editors, *Papers from the parasession on language and behavior: Chicago Linguistics Society*, pages 178–203.

Ohala, J. (1997). The relation between phonetics and phonology. In Hardcastle, W. J. and Laver, J., editors, *Tha Handbook of Phonetic Sciences*, pages 674–694. Blackwell Publishers.

Ohala, M. (1974). The abstractness controversy: Experimental input from hindi. *Language*, pages 225–235.

Perlmutter, D. M. (1993). Sonority and syllable structure in american sign language. In *Current issues in ASL phonology*, pages 227–261. Elsevier.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: an meg mismatch study. *Journal of Cognitive Neuroscience*, 12(6):1038–1055.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology 7*, volume 4 of *Phonology and Phonetics*. Mouton de Gruyter, Berlin; New York.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2(1):33–52.

Poeppel, D. and Hackl, M. (2008). The functional architecture of speech perception. *Topics in integrative neuroscience: From cells to cognition*, pages 154–180.

Poeppel, D. and Idsardi, W. (2011). Recognizing words from speech: the perception-action-memory loop. In *Lexical representation: A multidisciplinary approach.*, pages 171–196.

Poeppel, D., Idsardi, W. J., and Van Wassenhove, V. (2007). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):1071–1086.

Poeppel, D. and Monahan, P. J. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science*, 17(2):80–85.

Pullum, G. K. (1976). The Duke of York gambit. *Journal of Linguistics*, 12:83–102.

Reiss, C. (2018). Substance free phonology. *The Routledge Handbook of Phonological Theory*, pages 425–452.

Sandler, W. (1989). *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language*, volume 32. Walter de Gruyter.

Sandler, W. (1993). Sign language and modularity. *Lingua*, 89(4):315–351.

Sandler, W. (1996). *Representing handshapes*, volume 1. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Sandler, W. and Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.

Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences*, 102(7):2661–2665.

Sapir, E. (1925). Sound patterns in language. *Language*, 1(2):37–51.

Sapir, E. (1933). La rèalitè psychologique des phonémes. *Journal de psychologie normale et pathologique*, 247-65. Reprinted in translation as 'The psychological reality of phonemes' in David G. Mandelbaum (ed.), *Edward Sapir: Selected writings in language, culture, and personality, 46–60. Berkeley: University of California Press, 1949*.

Scharinger, M., Idsardi, W. J., and Poe, S. (2011). A comprehensive three-dimensional cortical map of vowel space. *Journal of cognitive neuroscience*, 23(12):3972–3982.

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research*.

Senghas, A., Kita, S., and Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, 305(5691):1779–1782.

Stokoe, W. C. (1960). Studies in linguistics: Occasional papers 8. In *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf*. Linstock Press.

Strother-Garcia, K. (2019). *Using Model Theory in Phonology: A Novel Characterization of Syllable Structure and Syllabification*. PhD thesis, University of Delaware.

Supalla, T. and Newport, L. (1978). How many seats in a chair? the derivation of nouns and verbs in american sign language, i p. siple (ed.): Understanding language through sign language research, academic press.

Tremblay, P., Deschamps, I., and Gracco, V. L. (2016). Neurobiology of speech production: A motor control perspective. In *Neurobiology of language*, pages 741–750. Elsevier.

van der Hulst, H. (1993). Units in the analysis of signs. *Phonology*, 10(2):209–241.

van der Hulst, H. (1995). The composition of handshapes. *Trondheim Work. Papers*, 23:1–17.

van der Hulst, H. (2013). Representing rhythm. In van der Hulst, H., editor, *Word Stress: Theoretical and Typological Issues*. Cambridge University Press. To appear.

van der Hulst, H. (to appear). Sign Language Phonology. In Dresher, E. and van der Hulst, H., editors, *The handbook of the history of phonology*. Oxford University press, Oxford.

van der Hulst, H. and van der Kooij, E. (2018). Phonological structure of signs – theoretical perspectives. In Josep Quer Villanueva, R. P. and Herrmann, A., editors, *The Routledge Handbook of Theoretical and Experimental Sign Language Research*. Routledge.

Volenec, V. and Reiss, C. (2017). Cognitive Phonetics: The Transduction of Distinctive Features at the Phonology–Phonetics Interface. *Biolinguistics*, 11:251–294.

Weigel, W. (2005). *Yowlumne in the Twentieth Century*. Doctoral dissertation, UC Berkeley.

Wilbur, R. (2011). Sign syllables. *The Blackwell companion to phonology*, 1:1309–1334.

Woll, B., Sutton-Spence, R., and Elton, F. (2001). Multilingualism: The global approach to sign languages. *The sociolinguistics of sign languages*, pages 8–32.

Wootton, D. (2015). *The invention of science: a new history of the scientific revolution*. Penguin UK.

Yip, M. (2002). *Tone*. Cambridge University Press.