Experiencing Speech

forthcoming in *Philosophical Issues*, 2010

Casey O'Callaghan

1 Listening to spoken language

Imagine traveling on a bus in a place with an unfamiliar language. Listening to the conversations around you differs dramatically from listening to conversations on a bus back home. Listening to speech in a language you know differs from listening to speech in a language you do not know.

Differs *how*?

One kind of difference is *cognitive*. You *understand* speech in your own language. Philosophers have devoted tremendous effort to explicating what understanding a language involves and under what conditions somebody understands what is said. What does an individual know in knowing a language? What is it to grasp the meaning of an expression? The answers appeal to things like possessing concepts, mastering grammar, and knowing truth conditions.

What about *perception*? Understanding spoken language also requires hearing; or seeing, as in lip reading—for simplicity, my focus is hearing spoken language. But the role of perception in understanding spoken language has received far less systematic philosophical attention.[1] Developed accounts of perception's contribution to understanding spoken language are scarce. Does listening to speech in a language you know differ *perceptually* from listening to speech in a language you do not know?[2]

The question can be approached in a number of ways. We can ask about the *objects* of perception when listening to speech. Do they differ from those of non-linguistic auditory perception?[3] Or we can ask about perceptual *processes*. Do linguistic sounds trigger some distinctively human,

[1] Noteworthy recent exceptions include Appelbaum (1998, 1999); Butterfill (2009); Matthen (2005); Mole (2009); Remez and Trout (2009); Rey (2007, 2008); Smith (2009); Trout (2001a,b).

[2] I am not denying that perception is a form of cognition. Rather, I am assuming there is a distinction between perception and extra-perceptual cognition. In this paper, when I talk about cognition, I mean *extra-perceptual* cognition.

[3] See Liberman (1996); Fowler (1986, 1996); Rey (2007, 2008); Matthen (2010).

extra-auditory perceptual capacity, module, or even modality?[4]

I want instead to focus on the *perceptual experience* of speech. In particular, I want to know whether the perceptual experience of listening to speech differs in phenomenal character—in *what it's like* from the subject's point of view—as a function of knowing the language. If it does, what is the diagnosis?

What hangs on this?

Claims about the perceptual experience of speech regularly figure in arguments about the objects of speech perception and about whether speech is perceptually special.[5] I will not pursue the connections here, but given their impact on theorizing it is important to get this piece right.

The answer also bears on debates about the *richness* of perceptual experience, and thus on what properties humans can perceive and the boundary between perception and cognition. This is my main concern. The question I want to answer is: How *expansive* a view of perceptual experience is required to account for the difference between listening to speech in a language you know and listening to speech in an unfamiliar language? My goal here is not to give a full account of the perceptual experience of speech, nor of the difference knowing a language makes. Instead, I present the case for an account of a certain shape. This is groundwork for a full account.

Tim Bayne's (2009) recent discussion about the *reach* of perceptual phenomenality provides helpful terminology. Bayne is concerned specifically with the reach of *phenomenal content*, representational content that supervenes upon phenomenal character—that is, how things are presented as being in perceptual experience. Many authors talk about the content of perceptual experience, where this is framed in terms of the veridicality or correctness conditions for perceptual experience.[6] *Conservatives* hold that perceptual experience is exhausted by awareness as of certain "low-level" sensible properties, such as colors, shapes, tastes, odors, pitches, loudness, and warmth. *Liberals*, however, argue that humans also perceptually experience "high-level" features, which might include objecthood, causal relations, faces, or kind properties such as *being a lemon*.[7] But Bayne's liberalism is most concerned with *kind* properties or other *categorical* properties, rather than objecthood or causal relations.[8] Since controversy exists over whether humans perceptually

[4]See, for instance, Fodor (1983); Pinker (1994); Liberman (1996); Trout (2001a); Hauser et al. (2002); Pinker and Jackendoff (2005).

[5]The classic debates recorded in Liberman (1996, esp. chs 13, 14, 16) and Fowler (1986, 1996) provide a good start on these issues.

[6]See Siegel (2008).

[7]Others frame the debate using the terms 'sparse' or 'thin' and 'thick'. See, e.g., Price (2009); Masrour (forthcoming).

[8]Siegel (2009) and Butterfill (2009) argue that humans perceptually experience causation. See Bayne (2009) and

experience natural kind or artifact kind properties, and since denying this does not imply conservatism, it helps to mark intermediate views. We can say that *moderates* hold that perceptual experience involves awareness as of low-level sensible features plus additional "intermediate-level" features. One might, for example, admit perceptual awareness as of objecthood, causal relations, faces, melodies, or expressions of emotion. Intermediate-level features might include structural properties or complex appearances, such as lemon-like or wrench-like looks which sculptures share with lemons and wrenches. They do not include natural kind or artifact kind properties.[9]

If hearing spoken language, known or unknown, involves auditorily experiencing just sounds and their low-level audible features, such as pitch, timbre, and loudness, this supports *conservatism* about perceptual phenomenology. If listening to speech in a known language involves perceptually experiencing intermediate-level features specific to that language, which are absent from the perceptual experience of unfamiliar language, this supports a *moderate* account. If, however, listening to a known language involves perceptually experiencing meanings or semantic properties, this favors *liberalism*.

In this paper, I argue first that the experience of listening to speech in a known language differs in phenomenal character from the experience of listening to speech in an unfamiliar language (with *experience* construed first in a broad sense not limited to perceptual experience). I consider two popular explanations for the difference, a *wholly cognitive account* and a *liberal (semantic) perceptual account*, which appeal to grasping or representing semantic properties in extra-perceptual cognition and in perception, respectively. Each raises worries that warrant considering alternatives. The liberal (semantic) perceptual account is too generous. And neither account suffices to explain the phenomenal difference, because learning a language is accompanied by changes in the perceptual experience of non-semantic features. This much is compatible with an account according to which changes in awareness as of low-level sensible features, including pitch, timbre, loudness, and duration, exhaust the phenomenal changes in perceptual experience. I argue, however, that this *conservative (low-level) perceptual account* does not suffice either. Rather, able speech perceivers perceptually experience sounds to bear additional, language-specific features.[10] In par-

Siegel (2006b) for discussion and arguments for this strong form of liberalism over conservatism.

[9]What makes something a low-level or intermediate-level feature is a critical issue that I will not address, since here I care more about whether we perceptually experience a particular class of features than about what counts as low-, mid-, or high-level. The discussion of section 6 is a start.

[10]A terminological note. I will use 'language-specific' to mean specific to a particular language, such as Quechua, rather than to mean specific to language, in general, in contrast to non-language. I will use 'speech-specific' in close to

ticular, I discuss the case of *phonemes*. Perceptually experiencing such language-specific features is necessary to explain the phenomenal difference to perceptual experience that accrues thanks to knowing a language. This *moderate (intermediate-level) perceptual account* rejects an austere conservative view about the contents of perceptual experience. But it explains why knowing a language makes a phenomenal difference without admitting semantic properties to the content of perceptual experience. Nevertheless, because I argue that the perceptual experience of phonemes is in important respects independent from the perceptual experience of low-level audible features, this account is my contribution to the growing body of work on specific features that support a *moderately liberal* view of perceptual experience.

2 An experiential difference

Consider the contrast between listening to a language you know and listening to a language you do not know. One sort of difference is an experiential difference. I have in mind a difference in what it is like, from the first-person point of view, to have each experience, or a difference in phenomenal character. If phenomenal features are aspects of what an experience is like for a subject to have, the experiences differ in phenomenal features. You don't need to look at a brain scan or do any science to know about phenomenal features. The relevant difference is one that holds between features of which a subject who has each experience in principle might be aware. I mean 'experience' in the most encompassing way, so that it includes, for example, perceptual, sensory, cognitive, and emotional experiences.[11]

Why believe in such an experiential difference across the two cases?

A number of philosophers have taken it as data that the experiences differ. For example, Gregory McCulloch says, "Listening to someone speaking clearly on a subject you know in a language you understand is a quite different kind of experience from hearing someone say the same thing in a language you do not understand" (McCulloch 1993, 46).

My bus example relies on a contrast between hearing speech in a foreign language and hearing speech in your own language. So does McCulloch's. Some might doubt whether this phenomenal difference accrues thanks to *knowing the language*. For instance, you might think the fact that spoken languages *sound* different from each other accounts for any phenomenal difference. Moti-

the latter sense, but as restricted to spoken language, in general, in contrast to non-linguistic sounds.

[11] Except where explicitly noted, as in 'perceptual experience', I will use 'experience' in the very broadest sense which also includes things like bodily sensations, imaginings, emotional feelings, and any special phenomenology of thought.

vating the claim that knowing a language affects experience means showing that the experiential difference isn't confined to hearing acoustically different sounds.

We should control for the acoustical differences among languages that make a difference to auditory experience. Accordingly, Susanna Siegel says, "Almost everyone has experienced hearing others speak in a foreign language that one doesn't understand, and that one can't parse into words and sentences. The phenomenology of hearing *the same speech* when one does understand is markedly different" (Siegel 2006b, 490, my emphasis).

So, fix the language. Thus, Jesse Prinz says, "If a monolingual English speaker hears the word *Hund* it is just a sound. If a German speaker hears it, the experience is phenomenally different" (Prinz 2006, 452).[12] And Tim Bayne says, "There is, intuitively, a difference between what it is like to hear the sentence 'il fait froid' when one does not understand French and what it is like to hear the same sentence after having learnt French, despite the fact that both involve the same auditory input" (Bayne 2009, 390).[13]

This introduces a complication. Fixing the language leaves two options. We can consider, as Prinz does, the experiences of two people, one who knows a given spoken language and one who does not, when faced with that spoken language. Or, instead, we can consider, as Bayne, Block, and Tye do, the experiences of one person before and after learning a spoken language. (Both readings are available for the Siegel quote.)

The former introduces worries about comparing phenomenology across subjects. So maybe it is better to focus upon the experience of a single individual. The reason is straightforward: the contrast can be detected from the first person. It helps not to introduce worries about comparing experiences of different people.

Even if they occupy the same mind, the experiences nevertheless are separated by the time it takes to learn a language. But this is a contrast argument for the existence of a phenomenal difference. Contrast arguments are most convincing when they point to an experiential contrast evident to a single individual at a time. It is damaging enough to have to consider experiential contrasts across time, since all sorts of factors impact experiential recall and reliability.

This means any phenomenal difference that stems from knowing the language never is directly

[12]To be clear, Prinz thinks what differs phenomenally is one's entire or overall experience. See the paragraph of the cited quotation (Prinz 2006, 452).

[13]Ned Block mentions "the difference in what it is like to hear sounds in French before and after you have learned the language" (Block 1995, 234), and Michael Tye discusses "phenomenal differences in what it's like to hear sounds in French before and after the language has been learnt" (Tye 2000, 61).

introspected. Establishing that there is a contrast thus requires imagination and more arguments.

But consider a different example that helps to make the case. An inventive paradigm that involves listening to *sinewave speech* generates a telling effect. Human speech contains a wide range of frequencies and harmonics—it is, acoustically, incredibly complex. To test what is acoustically important in speech, Remez et al. (1981) devised a kind of synthetic speech that replaces a complex human voice with a few simple sine waves whose frequencies and amplitudes vary with components (formants) of the original signal. On first hearing it, most people cannot recognize it as speech. It just sounds like random computer-generated noises and whistles. But after being told that it is speech and what it says, most people just begin to experience the signal as speech and can easily recognize what it says.[14]

A distinctive, nearly immediate shift occurs when you start to experience sinewave speech as language. This suggests that experiencing the very same sounds as speech makes a phenomenal difference. And it removes doubts about the impact of recall on judgments of experiential contrast.

But the difference might stem from experiencing the stimulus *as speech*, rather than thanks to your *knowing* the language. Sinewave speech is experienced first as mere sounds, and then as speech.

So consider a variant. Suppose you go through a sinewave speech demonstration for speech in a foreign language. You hear the synthetic sinewave sounds. Then you hear the original spoken French sentence (or pick a language you don't understand). Finally, you hear the sinewave stimulus again. Tellingly, sinewave speech in an unfamiliar language does not generate the dramatic phenomenal shift characteristic of known language sinewave speech.[15] But, it does sound different—it is recognizable as spoken language. You experience it as speech, but you cannot understand it. Contrast this phenomenal contrast with what takes place when you go through a sinewave demonstration in English. While each case involves an experiential difference, the effect on experience differs between the foreign language and known language cases. This difference is the experiential difference due to knowing or learning a language, beyond experiencing sounds as speech.

This all suggests that, abstracting from acoustic differences in the sounds of different languages, the experience of listening to speech in a language you know differs phenomenally from the experience of listening to speech in a language you do not know. Knowing a language impacts

[14]Try it here: `http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS/`

[15]Here I am endebted to exchanges with Eric Vatikiotis-Bateson and Robert Remez.

experience.

3 Two explanations

What explains the phenomenal difference between the experience of listening to speech in a language you know and the experience of listening to speech in an unfamiliar language?

According to one account, the interesting difference is entirely *cognitive* and extra-perceptual. In one case, but not in the other, you grasp the meaning of what is said. Grasping a meaning requires ascribing a meaning to what you've heard, or mapping the sounds you hear onto meanings, by way of lexical and grammatical structures. Since sounds have meanings and other linguistic properties conventionally, relative to a language, mastering the sound-meaning mapping recruits extra-perceptual cognition. On this account, perception's contribution to understanding spoken language is hearing sounds. Different languages comprise different sounds and sound patterns, so perceivers hear speech in different languages to differ. But, according to a *wholly cognitive account*, this is a relatively uninteresting difference because knowing the language does not change how you hear the sounds. Knowing a language, on this account, does not affect auditory perceptual experience. The critical contribution—grasping linguistic structure and meaning—is cognitive and extra-auditory.[16]

What explains the *experiential* difference according to the wholly cognitive account?[17] Any phenomenal difference that accrues thanks to learning or knowing a language is a matter of extra-perceptual cognitive phenomenality, or phenomenology of thought.[18] Grasping or understanding meanings changes one's overall experience in the face of speech sounds in a known language.

On an alternative account, however, the interesting difference is partly *perceptual*. In addition to hearing its sounds, listening to speech in a language you know means *hearing* or perceiving *meanings* or *semantic features*.[19] Worries about the publicity of meanings motivate John McDowell

[16]This view finds company from empirical researchers, such as Hauser et al. (2002), who believe that humans share with non-human animals the sensory-motor capacities involved in language understanding, while humans uniquely possess the (cognitive) recursive capacity that accounts for the fact that "animal communication systems lack the rich expressive and open-ended power of human language" (Hauser et al. 2002, 1570).

[17]You could hold a version of the cognitive account and think there is no phenomenal difference, and that grasping meaning is all that differs between the two cases. That is fine, but it means rejecting the case made in section 2 and explaining why those results seem compelling.

[18]See, for example, Siewert (1998, ch 8) and Pitt (2004) for arguments in favor.

[19]More carefully, we might also say that your auditory experience may represent high-level linguistic features including semantic, grammatical, syntactic, and lexical ones.

towards this *liberal (semantic) perceptual account*.[20] Others find it plausibly explains the experiential difference between hearing speech in a language you know and in one you do not know.[21] Like the wholly cognitive account, it explains the experiential difference in terms of representation or awareness as of semantic features (and perhaps other high-level linguistic features). According to the semantic perceptual account, however, the experiential difference stems partly from a *perceptual* one. When you learn a language, you begin to perceptually experience its sounds to have semantic properties.

What are the prospects of these two accounts?

Consider the *wholly cognitive account*. First, it relies on cognitive phenomenality, or the phenomenology of thought. While I don't see great reasons to resist on these grounds, the commitment is controversial, and some will object. Notice, however, that denying cognitive phenomenality leaves intact the inclination to believe the experiences differ phenomenally. Plausibly, independent from what it is like to grasp what is said, listening to speech differs phenomenally for familiar and unfamiliar languages. Imagine, for instance, the contrast between hearing a meaningless string of syllables taken from English words and hearing Japanese. What I am suggesting is that, whether or not there is something it is like cognitively to understand, the wholly cognitive account does not capture the very strong impression that your perceptual experience of speech *also* differs once you learn the language. It is easy to imagine somebody insisting, "It just sounds different." This of course should not be the last word—even Locke (1975, II.9) is particularly concerned that these kinds of hunches are unreliable because judgments influence what we take our sensory and perceptual experiences to be. But it is a prima facie indication that the perceptual

[20] "Our attention is indeed drawn to the contents of the used sentences, rather than the mere words (which are possible objects of attention even for someone who does not understand the language they are in): but not as something 'beneath' the words, to which we are to penetrate by stripping off the linguistic clothing; rather, as something present in the words—something capable of being heard or seen in the words by those who understand the language" (McDowell 1998, 99). See Smith (2009) for commentary and dissent.

[21] Siegel (2006b, 490) claims that hearing speech has a visual analog in seeing written text. When you learn to read, "You become disposed to attend to the semantic properties of the words in the text, and less disposed to attend visually to the orthographic ones" (490–1). Further, viewing highway billboards while rapidly passing indicates that you "take in" semantic properties of their text without deliberately attending to the text, and this "suggests that the 'taking in' [of semantic properties] can be merely sensory" (497). If the visual case is analogous to the auditory case, similar claims should hold for hearing spoken language. Bayne says, "It seems plausible to suppose that the [perceptual] phenomenal differences brought about by learning French involve both (low-level) changes in phonological structure *and* (high-level) semantic differences" (Bayne 2009, 393).

experiences differ phenomenally.

The *liberal (semantic) perceptual account* is more troublesome. One set of issues concerns what it takes to represent or have awareness as of meanings or semantic properties *perceptually*. Experiential or phenomenal differences have to do with what it's like from the perspective of the subject. This raises a few worries. What is the auditory perceptual experience of *meaningfulness* or of a meaning like? What are the recognizably audible characteristics of meanings? Do meanings satisfy the requirements on perceptual objectivity: are they perceptually experienced as subject-independent; as located in space?[22] Do we represent them as like abstract Fregean senses? As extensions? Do we experience meanings as *intentional* features, as *aboutness* or *being directed at* something else? Do we represent sounds as standing in semantic *relations* to things? How do you perceptually experience (hear) this intentionality or relation to be in absence of the thing represented? What about for abstract concepts or numbers?

These worries are pressing, given a further objection. Plausibly, the phenomenal character of auditory perceptual experience does not differ for *homonyms*.

One response to the homonym objection is that speech sounds are not perceptually experienced as having specific meanings, but *as meaningful*. This, however, cannot explain the difference between listening to speech in known and unknown languages. Nothing suggests you could not hear foreign language speech sounds as meaningful without hearing their specific meanings.[23]

These concerns warrant looking for an alternative to the liberal (semantic) perceptual account.

In the next section, I argue that neither the wholly cognitive account nor the semantic perceptual account as it stands suffices to explain the phenomenal difference between the experiences of listening to speech in an unfamiliar language and of listening to speech in a language you know. The wholly cognitive account explains the phenomenal difference entirely in terms of representing semantic features. The liberal (semantic) perceptual account explains the phenomenal difference in terms the perceptual experience of semantic features. However, I argue, the perceptual experience of non-semantic features of speech differs for a known language. Neither account as it stands explains this non-semantic difference. The wholly cognitive account captures any ex-

[22] See, e.g., Strawson (1959); Siegel (2006a); Burge (2009).

[23] A response to this claim called to my attention by Jeff Speaks is that perceiving speech as meaningful requires perceiving it to have a specific meaning, just as perceiving something as colored requires perceiving it to have a specific color. Foreign language speech thus can be judged, but not perceived, as meaningful. That is fine, but it renders ineffective the response to the homonym objection being considered, since it requires the perception of specific meanings, which would suffice to distinguish homonym experiences phenomenally, for the perception of meaningfulness.

periential difference due to understanding, but it fails to recognize the difference in perceptual phenomenality. The liberal (semantic) perceptual account could be supplemented to capture this difference, but that would not alleviate worries about perceptually experiencing meanings. The liberal (semantic) perceptual account's troublesome commitment to the perceptual experience of meanings is unnecessary if there is a compelling alternative account of the phenomenal difference in perceptual experience.

4 Non-semantic differences

4.1 Temporal and qualitative

Both the wholly cognitive account and the semantic perceptual account explain the phenomenal difference between experiencing speech in known and unknown languages in terms of one's grasp of semantic features. However, there is at least one kind of difference that cannot be explained this way. The experience of *non-semantic* features of speech differs phemomenally for speech in a known language. In particular, learning a spoken language impacts one's experience of the qualitative and temporal features of speech. After motivating this claim about *experience*, I will then argue it holds as a claim about the phenomenal character of auditory *perceptual* experiences. The central fact is that learning a language is accompanied by changes in discrimination for non-semantic features. The claim is that this signals a phenomenal difference in perceptual experience.

There are two forms of difference. In some cases, you effortlessly make, and must make if you are to understand what is spoken, discriminations you did not make and, in many cases, could not make before learning the language. In other cases, you cease to make and, in many cases, can no longer make, discriminations you did make before learning the language.

And, prima facie, there are at least two sorts of features for which discrimination changes: temporal and qualitative.

One kind of discrimination change concerns temporal characteristics. Learning a language requires distinguishing linguistically significant temporal features of a sound stream. This means being able to discern patterns at a finer temporal grain than you previously could resolve. And it means discerning linguistically-significant breaks, pauses, and gaps. Accordingly, the experience of temporal characteristics of speech differs when you know the language. Learning the language alters the apparent temporal characteristics of speech sounds. For example, before learning a language, its speech appears as a continuous stream of garbled sounds whose gaps and pauses have

undifferentiated significance. After learning the language, you experience fine-grained patterns of qualitative change over time, while certain gaps and pauses seem accentuated. Someone who understands spoken English hears the temporal properties, for example, the timing and duration of gaps between words, differently in an utterance of "It is incredibly hot outside" from someone who does not know English. Similarly, the English speaker hears the temporal properties of an utterance of "Soto wa sinji rare nai kurai atsui desu" differently from someone who understands spoken Japanese.

Another kind of discrimination change concerns qualitative characteristics. A spoken utterance of 's' sounds different from 'z' to a native English speaker, but not to a Spanish speaker, just as 'l' and 'r' sound different to an English speaker but not to a Japanese speaker. The 't' in 'ton' and 'stun' sounds virtually the same to speakers of English and French, but quite different to speakers of Chinese—the difference suffices to mark different words, much as 'bed' and 'bad' differ audibly to English speakers, but not to Dutch speakers. Someone who understands a given language distinguishes finer-grained qualitative differences among certain speech sounds than someone who does not understand the language.

Furthermore, someone who knows a spoken language might fail to hear, or lack the capacity to hear, linguistically insignificant distinctions among certain speech sounds. This is another respect in which knowing a language impacts experience. Some temporal features, such as gaps and discontinuities, are not apparent in experience when they lack linguistic significance. Some sound pairs are experienced to be qualitatively more similar by those who know a language than by those who do not, especially when the difference is not linguistically significant.

The claim is that learning a spoken language restructures one's experience as of non-semantic features of speech in that language; learning changes the phenomenal character of experience. In particular, it changes one's experience of qualitative and temporal characteristics of spoken language. The argument appeals to differences between the discrimination capacities of those who are competent with a language and of those who are not.

What I want to argue is that the capacities in question involve *perceptual* discrimination. So, one perceptually distinguishes different fine-grained temporal and qualitative differences in known and unknown languages. Accordingly, one's auditory perceptual experience as of the qualitative and temporal (non-semantic) features of speech sounds differs phenomenally thanks to knowing the language.

Here are some reasonable objections.

First, why think any *phenomenal* difference accompanies these different patterns of discrimination? Experience in these respects might remain the same while responses to it differ thanks to different needs and skills.

Second, if the experiences differ phenomenally, why think this cannot be explained in terms of the representation of *semantic* properties? The differences that come to matter when learning a language are just the ones that have high-level linguistic significance. For example, the experience of gaps discussed above might stem from representing discrete meanings.

Third, if the experiences differ phenomenally, why think the difference holds between phenomenal characteristics of *perceptual* experiences? Suppose auditory experience stays the same, but learning a language alters one's experience in some extra-perceptual phenomenal respect. For instance, one might cognitively seize upon different features, or the apparent cognitive salience of some aspects of speech might change, without a phenomenal change to perceptual experience.[24]

To demonstrate that neither the wholly cognitive nor the semantic perceptual account suffices, I need to defend the claim that listening to speech in a known language involves a *phenomenal* difference in *perceptual* experience that cannot be attributed to the representation of *semantic* features. First I'll defend the claim that there is a phenomenal difference in the experience of known-language speech that cannot be attributed to representing semantic features. Then I'll defend the claim that the difference holds between perceptual experiences.

4.2 Experience

How can we show that there is a phenomenal difference in the experience as of non-semantic features? I argued in section 2 that the experiences of listening to known and unknown speech differ phenomenally by appealing to the phenomenal switch that takes place in listening to sinewave speech. The problem is that when you listen to sinewave speech in a language you know, you understand meanings, so the experiential difference might stem from grasping meanings. Could sinewave speech in an unknown language address the problem? Listening to the stimulus differs phenomenally once you know it is language, even though you cannot grasp the meaning. While this establishes a difference in the experience of non-semantic features, it illustrates only a phenomenal difference between experiencing a stimulus as non-speech and as speech, rather than a phenomenal difference that stems from knowing the language. We need a different argument.

[24] For instance, Wu (forthcoming) argues that the phenomenal salience associated with conscious *attention* is cognitive rather than perceptual.

Aphasias are acquired language-related disorders. Some aphasias can be used to ground compelling contrast arguments, much as Bayne (2009) has done with agnosias. Consider three different types of aphasia. First, *Wernicke's aphasia* (WA) is caused by damage to Wernicke's area of the brain, which is associated with comprehension. Patients suffering it commonly can produce fluent, grammatical speech (a capacity associated with Broca's area), though it is error-prone and confused (paraphrasic). Though they recognize speech as such, they find spoken language unintelligible. Evidence of being able to grasp spoken words and their meanings is scarce. For instance, those with WA cannot repeat words spoken to them. Second, *transcortical sensory aphasia* (TSA) is caused by damage to an area located behind Wernicke's area. Patients suffering it are able to speak fluently, but have difficulty grasping the meanings of spoken language. However, TSA, unlike WA, leaves patients able to grasp what words have been spoken and to repeat them aloud. TSA leaves word recognition, but not comprehension, intact. Sometimes, those with TSA do not realize their impairment. Third, *pure word deafness* (PWD) is caused by disruptions in the superior temporal lobes. Patients with it produce fluent speech and text, but cannot through hearing grasp spoken language at all. They are not deaf, and hear and recognize other types of sounds, but are unable to recognize expressions in spoken language. While they can tell speech from non-speech, and even can recognize specific voices, speech just seems like a bunch of garbled sounds. PWD sufferers can speak and read, and even lip-read and use contextual cues to figure out what has been uttered, but they cannot hear what words you're saying.[25]

> The descriptions of the subjective experience of PWD are intriguing. "(Spoken language) is like a great noise all the time . . . you think you can catch it and it fades away, like foreign folks speaking in the distance. . . . When people speak loudly or quickly the words just run together" (Klein and Harper 1956). One patient experienced speech as "meaningless noise, garbled sound, or a foreign language" (Mendez and Rosenberg 1991). (Poeppel 2001, 681)

Plausibly, based on these descriptions, the experience of spoken language differs for individuals with pure word deafness. In particular, it differs between individuals with pure word deafness and individuals with transcortical sensory aphasia (though in each condition, the ability to tell speech from non-speech remains). PWD then shows that the experience of speech when one does not recognize the words differs phenomenally from the experience of speech when one does

[25] Thanks to Bob Slevc for conversations about his work with an individual with pure word deafness.

recognize the words (as in TSA and the unimpaired). (I am appealing to PWD rather than WA because WA but not PWD stems from poor overall semantic comprehension.) TSA, however, shows that one can recognize the words and sentences uttered without grasping their meanings. So, if the experience of listening to spoken language with PWD differs from the experience of listening to spoken language with TSA, then there exists a phenomenal difference between the experience of speech when one recognizes the words and when one does not recognize the words that does not depend on occurrently grasping meanings or representing semantic properties.

This shows the wholly cognitive account of section 3 does not suffice because there is an experiential difference associated with word recognition that is not explained by representing semantic properties. For the same reason, it shows the semantic perceptual account does not suffice. However, it leaves open that this phenomenal difference stems from a cognitive difference rather than a perceptual one. I still need to argue that the *perceptual* experience of listening to speech in a known language differs phenomenally from the perceptual experience of listening to speech in an unknown language in a way that cannot be attributed to the representation of semantic features.

4.3 Perceptual experience

Why think knowing a language impacts phenomenal characteristics of *perceptual* experience, rather than extra-perceptual cognition?

Consider an argument from *development*. Very young infants (from 1–6 months) distinguish very similar speech sounds, such as those of spoken instances of 'b' and 'p'.[26] That they distinguish them is determined by sucking patterns that reveal release from habituation. Further, infants distinguish speech sounds whose difference is linguistically significant for some world language; so, they discern all of the possible contrasts among speech sounds that their language might recognize as linguistically significant. Babies in New Jersey can distinguish 'k' and 'k^{h}' (a difference in Hindi but not English), and babies in Tokyo can distinguish 'l' and 'r'. However, within the first year of life (between 6–12 months), infants cease to discriminate sounds whose differences are not linguistically significant for the languages to which they are exposed. American babies stop responding to the difference between 'k' and 'k^{h}', and Japanese babies stop distinguishing 'l' from 'r'. Without exposure, babies lose the ability to detect phonetic differences that are not

[26]For now, to avoid confusion, I'm going to skip standard phonological notation. I'll explain it below. Also, for now, I'm going to start using single quotes to refer to speech sounds, so that " 'b' " should be read as "the sound of a spoken instance of 'b'."

linguistically significant to their own language.[27]

Here is the developmental argument. How babies discriminate speech sounds changes during their first year. Could this change have an entirely cognitive origin with no impact upon perceptual phenomenality? Even at one year, infants cannot distinguish as many meanings as speech sounds, so the change in discrimination behavior cannot result from representing semantic properties or grasping meanings. The change in discrimination behavior could, however, stem from changes to judgments of non-semantic similarity or difference. But it is not plausible that the change could stem from judgments *alone*. Without relying on semantic features or changes in apparent perceptual similarity and difference, the judgments would be ungrounded. So, non-semantic similarity and difference judgments that drive how infants discriminate among speech stimuli are best understood as judgments of perceptual similarity and difference. Moreover, if perceptual experience did not change during these developmental stages, we should anticipate being able to elicit discrimination behavior consonant with stable perceptual appearances, rather than or in addition to the actual acquired discrimination capacities. No evidence I know supports this. We should conclude that the actual discrimination behavior is perceptually driven. Thus, the change in discrimination patterns marks a change in how children perceive speech sounds as a result of experience and learning.

The usual story about development is that infants prune or forget how to perceive audible differences that are not linguistically significant. Humans learn perceptually to ignore differences that are irrelevant to their language. But doing so is learning to treat as perceptually equivalent sounds that for one's language are linguistically equivalent. The developmental evidence suggests that part of learning a language is learning to hear the sounds of one's language.

What are "the sounds of one's language"? Here is a way to approach the question in this context. *Phonetics* and *phonology* study speech sounds. They distinguish *phones* from *phonemes*. Consider the perceptible differences between utterances of words. Now, consider the smallest perceptible differences that are linguistically significant for a language. For instance, the difference between utterances of 'pat' and 'bat' is significant for English, but not for all languages. The *phones* are defined in terms of all of the possible minimal linguistically-significant perceptible differences that might occur in human language. Phones are usually indicated with square brackets, such as '[p]'. *Phonemes*, however, are given by the minimal linguistically-significant differences for a particular language, and are indicated with forward slashes, such as '/p/'. They are the sounds

[27] Eimas et al. (1971) is the seminal paper. Jusczyk (1997) is a good place to start in this literature.

treated as equivalent for a given spoken language.[28]

Understood this way, we can treat phones in audible speech as sound types. Phonemes then are features that can be modeled as language-specific types or categories of sounds. But the developmental evidence suggests phonemes correspond to *perceptual* categories. Listening to speech in a language you know, sounds that instantiate a common phoneme are treated in one respect as perceptually equivalent—there is an important sense in which they *sound* the same. This does not mean they are perceptually identical or indiscernible. Phoneme instances vary audibly and phonetically, and need not be indistinguishable in all respects from some canonical form. Nevertheless, instances of a given phoneme in one respect are experienced perceptually as the same. So, perceptual learning in language development is coming to perceptually discriminate sounds in a way that is consistent with their belonging to relevant language-specific categories. We learn to discern the language-specific categories of sounds that comprise our language's phonemes.

This evidence points to an argument from *categorical perception*. This argument aims to show that such learning alters perceived similarity relations among speech sounds; thus, it alters perceptual phenomenality.

Simplifying, perception is categorical when gradual differences to some physical parameter generate abrupt perceptual differences. First imagine a patch that changes gradually from black through shades of gray to white as we turn a brightness knob. Now imagine turning the knob gradually but just seeing the black patch instantly change to white. A common example is color perception. Gradually increasing wavelength from 400 nm to 495 nm causes gradually shifting experiences of shades of blue, but by around 500 nm, an abrupt shift to looking green occurs. With color, the shift is not perfectly abrupt, so color perception is not perfectly categorical. Still, colors belong to clusters whose members appear more similar to each other than to members of other clusters, even for equivalent physical differences. One mark of categorical perception is that it is much more difficult to detect differences that fall within the boundaries of a given category than to detect differences that span a category boundary.[29]

Human speech perception is *categorical*. More carefully, for certain critical kinds of speech sounds, including stop consonants, we perceive speech categorically.[30] For example, /b/ and

[28] When a language treats phonetically different sounds (phones) as equivalent, they are called *allophonic* for that language. So, for example, [k] and [k^h] are allophones of English /k/, but /k/ and /k^h/ are distinct phonemes of Hindi; [s] and [z] are allophonic for Spanish but not English.

[29] The classic resource is Harnad (1987).

[30] Vowel perception is far more continuous than categorical.

/p/ are perceived categorically by those raised around English speakers. Utterances of /ba/ and /pa/ differ only in voice onset time—/ba/ is voiced from the start, while /pa/ is voiced shortly after release. However, if we start with /ba/ and gradually increase voice onset time, you will not experience speech sounds that seem intermediate between /ba/ and /pa/. Instead, at some point, the /ba/ will just start to sound like /pa/. Likewise for other stop consonants over different physical parameters. For example, varying formant transitions reveals categorical perception for /ga/ and /da/.

Here is the argument. Humans have a strong natural propensity to perceive speech sounds categorically. Categorical perception yields a distinctive pattern of experienced similarity and difference relations among speech sounds. Those that belong to a given category are experienced to be more similar to each other than to non-members.[31] As above, it is implausible that this stems from extra-perceptual judgment alone, or from representing semantic features. However, phonetic categories are language-specific: their members and their boundaries vary according to the language. Accordingly, one's perceptual categories depend on one's language. And, experienced similarity relations among speech sounds differ depending on one's language environment. Mastering a language requires learning to hear its sounds—to perceive them categorically by discerning the distinctive patterns of relevant similarities and differences among them. So, learning a language alters one's perceptual experience of similarity and difference among its speech sounds. Perceptual experience of speech varies with language mastery.

Developmental evidence and categorical perception in adults therefore provide compelling evidence that the perceptual experience of listening to speech in a known language differs phenomenally from that of listening to speech in an unfamiliar language in respects that cannot be explained by the representation of semantic properties.

The method of phenomenal contrast plus arguments grounded in empirical considerations thus provide good reasons to conclude that the perceptual experience of listening to speech in a known language differs phenomenally from the perceptual experience of listening to speech in an unknown language, and that this difference cannot be attributed entirely to grasping or representing semantic properties.

[31] Furthermore, some categories will be more similar to each other than either is to a further category. The overall similarity space among speech sounds and categories of speech sounds may be quite complex. Articulatory *features* (voicing, place, manner, front, central, back, etc.) used to classify phonemes are a good place to start as a guide to the different dimensions of similarity and difference.

5 A conservative perceptual account

Suppose we have established that the perceptual experience of listening to speech in a known language differs from the perceptual experience of listening to speech in an unfamiliar language in a way that the representation of semantic features does not capture. Knowing a spoken language impacts the phenomenal character of auditory experience when you listen to someone talk. What is the difference?

One response is that the representation of, or awareness as of, low-level features explains the phenomenal difference. This is compatible with a *conservative* account of the phenomenal character of perceptual experience.

What are the low-level perceptible features for audition? Plausible candidates are audible qualities like pitch, timbre, loudness, and duration.[32] In addition, it is reasonable to include sensible individuals to which audible qualities are perceptually ascribed.[33] For simplicity, say the audible individuals are sounds.[34]

So a *conservative (low-level) perceptual account* of the phenomenal difference says that there are changes to what pitch, timbre, duration, and loudness one experiences sounds to have, or to what sounds one experiences, after learning a language.

This is plausible. Think of the changes in patterns of discrimination after learning a language. Some pairs of sounds seem more alike than they did, while some previously indistinguishable pairs of sounds newly seem to differ. One account of these changes in patterns of similarity and difference among sounds is that they seem to have different low-level audible features. Perhaps qualitatively more fine-grained pitches or timbres are experienced. Perhaps audible quality ascriptions are experienced at a finer temporal grain. Moreover, perhaps what seemed like a continuous sound stream when you didn't know the language comes to seem broken up into segments or discrete individual sounds, with dilated gaps and pauses.[35]

[32] Bayne says that on a conservative view "the phenomenal character of audition is exhausted by the representation of volume, pitch, timbre, and so on" (Bayne 2009, 386).

[33] I am motivated by the Many Properties argument. See Jackson (1977) and Clark (2000).

[34] Not too much hangs on this. Some will want to ascribe audible qualities such as pitch to events like collisions. This is fine, but then sounds are coinstantiated clusters of pitch, timbre, and loudness, and play a role analogous to audible individuals. Some might prefer *auditory objects*, which for now should be treated as complex sounds (see O'Callaghan 2008; Matthen 2010).

[35] On the individuation of sounds in time, see O'Callaghan (2007) and O'Callaghan (2008).

6 A moderate perceptual account

6.1 Language-specific features

The conservative (low-level) account explains the phenomenal difference in auditory experience just in terms of the representation of low-level audible features such as pitch, timbre, loudness, and duration of sounds. But now I want to argue that differences in the representation of low-level features alone do not offer the best explanation for the phenomenal difference knowing a language makes to perceptual experience. Representing *higher-level* or *intermediate-level* features, including language-specific features such as *phonemes*, I argue, is responsible for a critical phenomenal difference between the perceptual experiences of listening to known and unknown language.[36] According to this *moderate (intermediate-level) perceptual account*, the phenomenal content of perceptual experience reaches beyond low-level features at least to language-specific phonemes. While it may reach further, to syllabic, morphological, syntactic, or other higher-level structural features, my focus here is on phonemes. Despite the plausibility of the low-level treatment, phonemes ground a nice case that the perceptual experience of speech involves awareness as of intermediate-level, language-specific features.

Consider sinewave speech. Whether you're hearing the stimulus as speech or not, you perceptually experience the sounds to have very nearly the same low-level audible qualities, such as pitch, timbre, loudness, and duration. In any of those respects, the sound doesn't appear to change much. Nevertheless, upon hearing it as speech, the perceptual experience changes. A good explanation for the phenomenal change is that you begin to represent or to be aware as of additional features. Candidates include features that are specific to speech, and to the particular language, such as phonemes.

Now consider pure word deafness (see section 4.2). PWD leaves non-linguistic auditory capacities and recognition skills intact, and it leaves semantic capacities intact. But, subjects with PWD are unable to make out spoken words. PWD bears the marks of a failure to perceptually represent language-specific features, such as phonemes. Notably, those with PWD are very poor at distinguishing stop-consonant sounds, such as /p/ and /b/, which are perceived categorically. They are better with vowels, which are more continuous. One hypothesis is that pure word deafness

[36] I do not use 'higher-level' and 'high-level' interchangeably. Here the former just marks a contrast with low-level features.

stems from trouble with categorical perception for speech sounds.[37]

One might object that listening to sinewave speech before recognizing it as speech and listening with PWD both involve failing to perceptually experience the right *combinations* of low-level audible features.

Whatever plausibility this has for sinewave speech is missing for PWD. Why can't subjects hear those combinations of low-level audible qualities? They lack other major auditory deficits, and are able to hear and identify all sorts of other environmental sounds. Since PWD leaves auditory capacities otherwise intact, we would like an explanation that is not ad hoc for why it disturbs the capacity to represent just certain combinations of low-level audible qualities.

Another kind of argument suggests that no such explanation exists. Recall that the perception of speech sounds is categorical. Small acoustical changes can have a large impact on phenomenal similarity, and large acoustical changes can have relatively little impact on phenomenal similarity. A very good explanation for the patterns of phenomenal similarity and difference among speech sounds is that, when a stimulus falls within certain parameters, it is perceptually represented as being an instance of a given phoneme. Since phoneme categories and their boundaries are language-specific, a very good explanation for perceptually experiencing the patterns of similarity and difference characteristic of a given language is the representation of its phonemes.

Why is this explanation *better* than one that appeals only to low-level audible qualities? I have emphasized that speech sounds are perceived categorically. Candidates for low-level audible qualities, such as pitch, timbre, duration, and loudness, however, are not. For example, gradual changes in fundamental frequency generate gradual, rather than abrupt, changes in experienced pitch. While timbre is trickier, the same holds for duration and loudness. It is not clear how perceptually experiencing combinations of low-level audible qualities, each of which independently is perceived in a continuous rather than categorical manner, could explain the categorical pattern of perceptual experience for speech sounds. In fact, good reasons suggest it is possible to experience any given combination of low-level audible features while failing to experience speech sounds categorically. If so, low-level audible features do not guarantee the relations of similarity and difference that are distinctive to hearing speech in a known language.

My claim is that the perceptual representation of higher-level features, such as phonemes, better explains how the apparent similarity and difference relations among sounds differ when listening to speech in a known language than does the representation of low-level audible features.

[37] Thanks to Bob Slevc for discussion.

So, the best explanation we have for the phenomenal properties associated with the experienced patterns of similarity and difference characteristic of listening to speech sounds in a known language involves the perceptual representation of phonemes. Thus, the best explanation we have for the phenomenal difference between the perceptual experience of listening to speech in a known language and the perceptual experience of listening to speech in an unknown language appeals to the representation of language-specific features beyond low-level audible features.

6.2 Independence

The moderate perceptual account has another advantage over the conservative (low-level) perceptual account. Explaining the phenomenal difference in terms of the representation of higher-level, language-specific features, such as phonemes, captures the relative *independence* of phoneme experiences from low-level audible quality experiences.

First, perceptually experiencing a given phoneme is compatible with tremendous variety in low-level audible features. Consider utterances of 'diary', 'adept', 'adroit', 'bad', 'bud', and 'arid', both by a single individual and by a bunch of your friends and neighbors. Each utterance perceptually appears to include a spoken instance of the /d/ sound. However, the low-level audible qualities one experiences when experiencing the /d/ sound might differ in two relevant respects. One: The acoustical signature corresponding to the perceptually experienced /d/ sound differs depending on the surrounding phonemic context. For example, when the /d/ is followed by /u/, the corresponding formant begins at a low frequency and drops, but when it is followed by /i/, the relevant formant begins a higher frequency and rises.[38] This leads to perceptual experiences that ascribe different low-level audible qualities but a common phoneme. Two: People come in lots of forms that lead to differences in voice, and different occasions call for different voices. Thus, low-level audible qualities of speech vary dramatically with, for example, age, sex, size, mood, and social context.[39] So, the acoustic signature of someone pronouncing /d/ depends on anatomical and cultural factors. The lesson is that the different low-level audible quality experiences with which the perceptual experience as of a common phoneme can be coupled is virtually unlimited.

[38] See Liberman et al. (1967, 435, fig. 1).

[39] Physical expressions of emotion have perceptible effects on speech sounds. For instance, smiling shortens the vocal tract and increases pitch. Chemical differences due to mood have been shown to have an acoustical effect. Age impacts voice.

Second, experiencing a certain combination of low-level audible qualities at a time is compatible with perceptually experiencing different phonemes. One: Different phonemes sometimes share acoustic correlates, again depending on phonemic context. For example, for a given speaker, the acoustic signal that corresponds to /p/ is nearly identical to that of /k/ in the contexts /pi/ and /ka/.[40] The common acoustical features that lead one to perceptually experience different phonemes in these two contexts nevertheless might cause one to experience qualitatively similar low-level audible qualities. So, in a way that depends on linguistic context, different phonemes might be coupled in perceptual experience with the same combination of low-level audible qualities. Two: Different perceptible phonemes might be coupled with a given combination of audible qualities across speakers and occasions. Two sound segments indistinguishable in low-level audible qualities can play very different roles depending on who utters them and when. An audible quality combination that coincides temporally with an apparent instance of /d/ might, under different circumstances, coincide with an apparent instance of a speech sound other than /d/.[41]

These points are well illustrated by the long absence of both compelling synthetic speech and accurate speech recognition software. If simple audible qualities marked the phonemes words comprise, we wouldn't have needed the geniuses at Google.[42]

The phenomenal features of phoneme experiences therefore are not exhausted by the phenomenal features of the low-level audible quality experiences with which they are coupled. Thus, there is a strong preliminary case that, for a known spoken language, the phenomenal features of perceptual experiences as of language-specific speech sounds are in notable respects independent from the phenomenal features of coinciding perceptual experiences as of low-level audible

[40] See Cooper et al. (1952).

[41] Compare Butterfill (2009):

> Speech perception also exhibits constancy (otherwise called invariance): the location of the category boundaries changes depending on contextual factors such as the speakers dialect, or the rate at which the speaker talks; both factors dramatically affect which sounds are produced. This means that in two different contexts, different stimuli may result in the same perceptions, and the same stimulus may result in different perceptions. (Butterfill 2009, 414–5, footnotes omitted)

[42] In a compelling discussion of the prospects for reductionism about phonemes, Robert Remez and J. D. Trout ask, "Could a phoneme, a linguistic marker used to distinguish one word from another, be a sensory form?" After considering the evidence and possibilities, they answer, "The hypothesis is false" (Remez and Trout 2009, 241). Pardo and Remez (2006) contains an illuminating discussion of the kinds of variation that are relevant to the arguments of this section.

qualities.[43] If so, what it's like to hear speech sounds in a language you know isn't just what it's like to hear low-level audible qualities. The moderate perceptual account captures this, and the conservative account does not.[44]

6.3 Awareness of phonemes

I want to respond to a challenge to this approach that stems from recent remarks by Stephen Butterfill. Butterfill (2009) argues that we do not perceive phonemes in the same sense in which we perceive shapes. Unlike perceiving shapes, "perceiving phonemes is not a way of being aware of them" (426). Further, to describe someone as perceiving phonemes is not to describe an introspectible aspect of experience. He says, "I suspect introspection tells us nothing about what we perceive when we perceive speech; certainly it has no bearing on conflicts about the objects of speech perception. So speech and shape perception differ with respect to introspection" (424). Butterfill goes on to say, "From the point of view of the perceiver, categorical perception is simply a signal of sameness or difference with distinctive phenomenological effects" (427).

Butterfill clearly is concerned with some different questions than I've been considering. I am talking about phenomenal aspects of the perceptual experience of speech. Butterfill is talking about knowledge of the objects of speech perception. However, Butterfill offers a minimalist view about the phenomenal features associated with categorical perception, according to which categorical perception yields merely a phenomenal signal of *sameness* or *difference*. This might undercut my account, according to which the representation of phonemes helps explain the phenomenal difference in perceptual experience that accrues thanks to knowing a language. Representing mere sameness or difference, along with low-level audible features, might make phoneme awareness otiose.

Butterfill is right that categorical perception is not a mode of perception one experiences *as such*. No phenomenal marker revelatory of categorical perception is evident on a particular occasion of experience. Instead, categorical perception is a kind of perceptual process characterized by a certain pattern of experiential responses—it manifests as a *pattern* of appearances. For instance, speech sounds within a category appear more similar to each other than to those outside

[43] See section 7 for further discussion.

[44] What about a *pattern* of low-level audible qualities of a sound stream over time? While this raises issues beyond this discussion, it is a move in the right direction. But the considerations in the text show that the dependence cannot be just upon a determinate pattern of determinate low-level audible features—the pattern cannot be a first-order pattern. Instead, it must be a higher-order pattern of relations or structural contrasts. See also Remez and Trout (2009).

the category. And, while categorical perception does issue signals of sameness and difference, they nonetheless come in degrees. Categorical perception results in a characteristic pattern or structure of graded similarity and difference among appearances.

Consider an analogy with color. Red things look more or less similar to things with other hues. You don't introspectibly experience redness as fitting into a color solid. The patterns of appearance, evident in phenomenal similarity and difference, comprise a color solid.[45] This may, in particular when plotted against physical features, exhibit the marks characteristic of a categorical perceptual process.

Something similar holds for phonemes. Speech sounds perceptually seem to bear similarity and difference relations to each other. Utterances of /b/ sound more similar to /p/ than to /d/. Each sounds more similar to the others than any does to utterances of /a/ or /e/. So, speech sounds form complex appearance structures.[46] Some phoneme charts provide an approximation akin to the color solid. The structure of the phenomenal space, in particular its categorical structure, is not evident in the experience one has on any given occasion.

But *similarity* and *difference* are not sui generis phenomenal characters. Experiences are phenomenally similar or different in some phenomenal respect. Red things look a particular way in virtue of which red occupies a place in the color solid. Speech sounds have distinctive perceptible characters such that they occupy places in a language-specific phenomenal similarity space.

I argued above that low-level audible qualities alone do not suffice to capture the (categorical) patterns of apparent similarity and difference—the pattern of appearances—that are characteristic of the perceptual experience of listening to speech in a known language. The phenomenal similarity relations distinctive to the experience of speech sounds therefore must hold in some other phenomenal respect. It follows that awareness as of low-level audible features plus similarity and difference does not suffice to explain the phenomenal character of the perceptual experience of listening to speech in a language you know. Put briefly, phenomenal similarity is similarity in a phenomenal respect; low-level audible respects do not yield distinctive, language-specific phenomenal similarity relations among speech sounds.

So we can say the following, which is stronger than what Butterfill says about speech perception. The representation of phonemes provides a better explanation for the patterns of phenomenal similarity and difference that are revealed by perceptual discrimination tasks than does the

[45] See, e.g., Clark (1993).

[46] Jekosch (2005) is a fascinating recent discussion of qualitative similarity spaces for speech.

representation of low-level audible features alone. Thus, an important aspect of the phenomenal character of the perceptual experience of speech sounds is explained by the representation of phonemes. However, since the similarity in question holds among features things perceptually appear to have, awareness as of language-specific features such as phonemes explains the pattern of appearances that is distinctive to hearing speech sounds in a familiar language. If so, then perceiving phonemes counts as a way of being aware of phonemes. Therefore, we perceptually experience speech and phonemes in much the same way that we perceptually experience shapes. Moreover, perceptually experiencing phonemic properties of sounds helps explain the phenomenal difference between the experience of listening to speech in known and unknown languages.

In this section, I have argued that the perceptual experience of speech in a language you know is not exhausted by the representation of low-level audible features plus similarity and difference. Since the relevant further features, phonemes, are language-specific and depend for their representation upon substantial exposure to a language, representation or awareness as of higher-level features figures critically in any explanation for why the perceptual experience of listening to speech in a language you know differs phenomenally from the perceptual experience of listening to speech in an unfamiliar language.

7 Four further issues

I have argued for a moderate (intermediate-level) account of the perceptual experience of listening to speech in a familiar language. The account is framed in terms of perceptual awareness as of language-specific features, including phonemes, missing from the experience of speech in an unfamiliar language. At the outset, I said my goal here is not to present a complete account of the perceptual experience of listening to speech. Nonetheless, the account sketched raises four further issues that warrant mention and deserve full treatment elsewhere.

First, what kind of perceptible *feature* is a phoneme? In particular, are phonemes perceptible *individuals* or *properties*? A popular metaphor likens phonemes to beads assembled on a string to form words and sentences.[47] This suggests that phonemes are individuals. I favor an account on which phonemes are language-specific perceptible properties or attributes of sounds. In particular, they are perceptually experienced as properties of temporally-extended parts, portions, or segments of sounds. Speech sounds thus can be characterized by the (perhaps overlapping)

[47]The much-discussed "beads on a string" analogy traces to Bloomfield (1933). It is related to the "alphabetic conception" discussed by Appelbaum (1999).

sequences of phonemes they perceptually appear to instantiate. This implies that experiencing non-speech sounds and experiencing speech can share perceptual objects, and that listening to speech sounds is not hearing entirely novel individuals. Thus, my account is compatible with the claim that we experience speech to have audible qualities such as pitch, timbre, and loudness. This requires a full defense, since it counters a tradition, stemming primarily from Liberman's famous *Motor Theory* of speech perception, that argues forcefully that speech perception and non-linguistic audition lack shared perceptual objects.[48]

Second, I argued in section 6.2 that the moderate perceptual account captures the independence of phoneme experiences from low-level audible quality experiences. In just what sense is the perceptual experience of phonemes *independent* from the perceptual experience of low-level audible qualities? My view is that the considerations presented here show that phoneme experiences are not *derived from* or *constitutively* dependent upon low-level audible quality experiences;[49] do not *causally* depend just upon the processes responsible for low-level audible quality experiences; and do not *supervene* in any robust sense upon coincident low-level audible quality experiences. This presents an argumentative burden and raises questions beyond the scope of this discussion. However, I suspect the inclination to ground perceptual awareness as of higher-level features in awareness as of lower-level features is the persistent legacy of treating conscious sensations as the basic units of experience. The case of speech presents the strongest case I know against blanket dependence claims for higher-level phenomenality.

Third, phonemes suffice to motivate the moderate (intermediate-level) perceptual account, once we reject the liberal (semantic) perceptual account. Which features beyond phonemes figure in a complete account of the perceptual experience of listening to speech in a language you know? For instance, do syllables, words, or even grammatical features belong among the contents of perceptual experience? Or do phonemes, along with differences in low-level audible features and the individuation of sounds, including those mentioned in section 5, suffice? I suspect more than phonemes figures in a complete account, but that richer linguistic structure corresponds to increasingly cognitive phenomenality. The method, however, should be to proceed case by case.

This leads to a fourth concern. What distinguishes perceptual phenomenality from that of extra-perceptual cognition? Rather than conjecture, let me state a worry. Suppose the boundary

[48] See Liberman (1996).

[49] For instance, as in the sense of Byrne and Hilbert (2008, 385): "One idea is that sensible qualities are of two sorts: basic and derived. The basic sensible qualities are the building blocks from which the derived sensible qualities can be constructed."

between perception and cognition is not sharp, for the following reasons. Perceptual processes occur in stages leading from sensory stimulation to higher cognition and conceptual thought, and they are shaped by many kinds of influence. And perception and thought arguably lack proprietary phenomenal markers that respectively distinguish experiences. Further, no clear, uncontroversial criterion we currently grasp exhaustively sorts processes, states, representations, or experiences into perceptions and thoughts. Which mental episodes count as perceptions and which do not thus might be a matter of degree. If so, it is tough to argue in a principled way about higher-level content of *perceptual* experience. What we really need, before continuing to argue over liberalism and conservatism about perceptual experience, is either a better account of the distinction between perception and cognition, or a story about why it matters. As with so many of the good questions, straightforward cases at the ends of the spectrum belie the tangle in between.

8 Conclusion

I have evaluated four different ways to explain why the experience of listening to speech in a language you know differs phenomenally from the experience of listening to speech in an unfamiliar language: a wholly cognitive account, a liberal (semantic) perceptual account, a conservative (low-level) perceptual account, and a moderate (intermediate-level) perceptual account. I argued that the phenomenal difference stems in part from a difference in perceptual experience. In particular, the perceptual experience of speech in a known language involves representation of or awareness as of higher-level, language-specific features of speech sounds, including phonemes. The perceptual experience of listening to familiar speech therefore is not exhausted by awareness as of low-level audible features such as pitch, timbre, and loudness, but it need not involve awareness as of meanings or semantic features. The case of speech thus supports a moderately liberal view about the reach of perceptual phenomenality.[50]

[50]I owe thanks to a number of people. Audiences at Oxford and UBC asked lots of good questions and gave me plenty of feedback on earlier versions of this material, which shaped how this paper developed. Discussions with Tim Bayne, Alex Byrne, Philippe Chuard, Justin Fisher, Mohan Matthen, Georges Rey, and Bob Slevc helped me to discern how the arguments should go at critical stages. Stephen Butterfill, Maddy Kilbride, Jeff Speaks, and Cara Spencer each provided extensive and detailed written comments that helped me to avoid mistakes and to make this a better paper. Emily Koehn and Jeff Speaks deserve special thanks for conversations about speech and about details concerning presentation. Shohei Edamura helped with the Japanese translation.

References

Appelbaum, I. (1998). Fodor, modularity, and speech perception. *Philosophical Psychology*, 11(3):317–330.

Appelbaum, I. (1999). The dogma of isomorphism: A case study from speech perception. *Philosophy of Science*, 66(Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part I: Contributed Papers):S250–S259.

Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly*, 59(236):385–404.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18:227–247.

Bloomfield, L. (1933). *Language*. Holt, New York.

Burge, T. (2009). Perceptual objectivity. *Philosophical Review*, 118(3):285–324.

Butterfill, S. (2009). Seeing causes and hearing gestures. *Philosophical Quarterly*, 59(236):405–428.

Byrne, A. and Hilbert, D. (2008). Basic sensible qualities and the structure of appearance. *Philosophical Issues*, 18:385–405.

Clark, A. (1993). *Sensory Qualities*. Clarendon Press, Oxford.

Clark, A. (2000). *A Theory of Sentience*. Oxford University Press, New York.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 24:597–606.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968):303–306.

Fodor, J. (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3):1730–1741.

Harnad, S. (1987). *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge, UK.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579.

Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press, Cambridge.

Jekosch, U. (2005). *Voice and Speech Quality Perception: Assessment and Evaluation*. Springer, Berlin.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. MIT Press, Cambridge, MA.

Klein, R. and Harper, J. (1956). The problem of agnosia in the light of a case of pure word deafness. *Journal of Mental Science*, 102(426):112–120.

Liberman, A. M. (1996). *Speech: A Special Code*. MIT Press, Cambridge, MA.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6):431–461.

Locke, J. (1689/1975). *An Essay Concerning Human Understanding*. Clarendon Press, Oxford.

Masrour, F. (2010). Is perceptual phenomenology thin? *Philosophy and Phenomenological Research*.

Matthen, M. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press, Oxford.

Matthen, M. (2010). On the diversity of auditory objects. *Review of Philosophy and Psychology*, 1(1):63–89.

McCulloch, G. (1993). The very idea of the phenomenological. *Proceedings of the Aristotelian Society*, 93:39–57.

McDowell, J. (1998). *Meaning, Knowledge, and Reality*. Harvard University Press, Cambridge, MA.

Mendez, M. F. and Rosenberg, S. (1991). Word deafness mistaken for Alzheimer's disease: Differential characteristics. *Journal of the American Geriatrics Society*, 39(2):209–211.

Mole, C. (2009). The Motor Theory of speech perception. In Nudds, M. and O'Callaghan, C., editors, *Sounds and Perception: New Philosophical Essays*, chapter 10, pages 211–233. Oxford University Press, Oxford.

O'Callaghan, C. (2007). *Sounds: A Philosophical Theory*. Oxford University Press, Oxford.

O'Callaghan, C. (2008). Object perception: Vision and audition. *Philosophy Compass*, 3(4):803–829.

Pardo, J. S. and Remez, R. E. (2006). The perception of speech. In Traxler, M. and Gernsbacher, M. A., editors, *The Handbook of Psycholinguistics*, chapter 7, pages 201–248. Academic Press, New York, 2nd edition.

Pinker, S. (1994). *The Language Instinct*. William Morrow and Company, New York.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, 95:201–236.

Pitt, D. (2004). The phenomenology of cognition, or, *What is it like to think that p*? *Philosophy and Phenomenological Research*, 69(1):1–36.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25:679–693.

Price, R. (2009). Aspect-switching and visual phenomenal character. *Philosophical Quarterly*, 59(236):508–518.

Prinz, J. (2006). Beyond appearances: The content of perception and sensation. In Gendler, T. S. and Hawthorne, J., editors, *Perceptual Experience*, pages 434–459. Oxford University Press, Oxford.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carell, T. D. (1981). Speech perception without traditional speech cues. *Science*, (212):947–950.

Remez, R. E. and Trout, J. D. (2009). Philosophical messages in the medium of spoken language. In Nudds, M. and O'Callaghan, C., editors, *Sounds and Perception: New Philosophical Essays*, chapter 11, pages 234–264. Oxford University Press, Oxford.

Rey, G. (2007). Externalism and inexistence in early content. In Schantz, R., editor, *Prospects for Meaning*, volume 3 of *Current Issues in Theoretical Philosophy*. de Gruyter, New York.

Rey, G. (2008). In defense of folieism: Replies to critics. *Croatian Journal of Philosophy*, VIII(23):177–202.

Siegel, S. (2006a). Subject and object in the contents of visual experience. *Philosophical Review*, 115(3):355–388.

Siegel, S. (2006b). Which properties are represented in perception? In Gendler, T. S. and Hawthorne, J., editors, *Percepual Experience*, pages 481–503. Oxford University Press, New York.

Siegel, S. (2008). The contents of perception. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. URL = http://plato.stanford.edu/archives/win2008/entries/perception-contents/.

Siegel, S. (2009). The visual experience of causation. *Philosophical Quarterly*, 59(236):519–540.

Siewert, C. (1998). *The Significance of Consciousness*. Princeton University Press, Princeton.

Smith, B. (2009). Speech sounds and the direct meeting of minds. In Nudds, M. and O'Callaghan, C., editors, *Sounds and Perception: New Philosophical Essays*, chapter 9, pages 183–210. Oxford University Press, Oxford.

Strawson, P. F. (1959). *Individuals*. Routledge, New York.

Trout, J. D. (2001a). The biological basis of speech: What to infer from talking to the animals. *Psychological Review*, 108(3):523–549.

Trout, J. D. (2001b). Metaphysics, method, and the mouth: Philosophical lessons of speech perception. *Philosophical Psychology*, 14(3):261–291.

Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press, Cambridge, MA.

Wu, W. (201x). What is conscious attention? *Philosophy and Phenomenological Research*, forthcoming.