Dear Casey,

As I understand your argument, it runs roughly like this:

> 1. "speech sounds within a category appear more similar to each other than to those outside the category. … Categorical perception results in a characteristic pattern or structure of graded similarity and difference among appearances." (p. 21)
>
> 2. "Experiences are phenomenally similar or different in some phenomenal respect." (p. 21)
>
> 3. The fact that (2) is best explained by the hypothesis that experiences of speech involve phonemes appearing as such (p. 22).
>
> 4. Thus "perceiving phonemes counts as a way of being aware of phonemes." And "we perceptually experience speech and phonemes in much the same way that we perceptually experience shapes." (p. 22)

I take the inference from 3 to 4 to be uncontroversial. I want to offer two lines of reply to this argument. The first is to suggest that (3) is probably false. The second is to deny (1). I think both lines of reply are correct but as they are independent it seemed worth presenting them separately.

First response: (3) is probably false. In brief, this response involves three distinctions. First, there's a distinction between the claim that pairs isolated segments of speech are experienced as similar or dissimilar to each other and the claim that such segments are so experienced in the much richer context of ordinary speech. Second, there's a distinction between saying that adults' experience of speech involves experiencing segments of speech as similar or dissimilar and saying that these are experiences as similar or dissimilar in respect of their *phonetic* properties. (I agree with your point that that "[e]xperiences are phenomenally similar or different in some phenomenal respect" (p. 21); the issue is *which* respect, and it may be that what we are aware of in speech is approximately syllabic rather than phonetic.) Third, there's a distinction between saying that adults confronted with speech have experiences involving certain phenomenal contents and saying that experiences with these phenomenal contents play some role in explaining their abilities to process speech.

In slightly more detail, take Ute Jekosch's (2005) experiment on quality spaces which you cite. This experimenter presented subjects with pairs of isolated syllables such as /ta/ acoustically. Subjects were then asked to rate how similar they perceived these to be on a five point scale (Jekosch 2005:149). Do the results support the claim that speech sounds "form complex appearance structures" in roughly the way that colours do? What Jekosch's findings immediately show is that adult listeners can rate isolated syllables as more or less similar. It isn't obvious to me that these ratings reflect patterns of similarity that are part of how subjects experience the stimuli rather than part of how they interpret their experiences in the light of background knowledge. As this point raises some hard questions about the boundary between perception and cognition which neither of us think has been sharply drawn (p. 35), I'll assume for the sake of argument that speech sounds do "form complex appearance structures". This assumption leads to another question. How accurately do the categories induced by judgements of similarity map onto phonemes actually distinguished in ordinary speech perception? It is notable that the two main subjects' results are markedly different in some respects (Jekosch 2005:155ff.). Unless subjects with relevantly similar abilities to perceive speech have the roughly the same similarity spaces, it seems unlikely that the similarity space identifies phonemes. Also, in tests of phonological awareness such as sorting words according to their initial phonemes, adults make mistakes derived from their knowledge of spelling. Similarly, there may well be cases of systematic mismatch between similarity judgements and actual speech perception. That is, adults will sometimes judge one pair of isolated syllables to be similar and a second pair of isolated syllables to be dissimilar even though in processing ordinary speech the first pair of syllables is not harder to distinguish than the second. It's one thing to say experience of speech sometimes involve some things sounding similar in some respect; it's another to say that the respects in which they sound similar are phonetic. My guess is that phenomenological contents involve syllables rather than phonemes and that phenomenological contents will reflect subjects' explicit knowledge the structure of words, including spellings, an so not correspond directly to the phonetic structure uncovered by the perceptual processes involved in interpreting speech. (By the way, I'm not sure how to test claims about the phenomenology of specific experiences, but it seems to me not at all obvious that syllables appear as composed of phonemes in experience.)

Let's assume for the sake of argument that the relevant similarities are similarities with respect to phonetic categories. What does the experience of such similarities have to do with speech perception? It is important in my view to distinguish between speech perception and perception of speech (I realise that we might disagree about whether there is any such distinction; more on this below). Roughly speaking, speech perception is that perceptual process (if any) which is primarily responsible for rapid oral communication. Perception of speech refers to any kind of perceptual processing which may occur when encountering speech. So far I've allowed that adult listeners experience pairs of syllables presented in isolation as similar or dissimilar with respect to phonetic properties. It's not obvious that the same phenomenology attaches to experiences of these syllables in ordinary speech. But assume for the sake of argument that it is. Assume, that is, that speech perception involves a sequence of experiences with phenomenal contents corresponding to the phenomenal contents which occur when subjects in Jekosch's trial encounter individual syllables.[1] How, if at all, are such phenomenal contents involved in speech perception? I claim that speech perception does not depend on any such phenomenal contents. The main argument for this claim is developmental:

i. If speech perception involved phenomenal contents corresponding to phonetic gestures, then it should be easy to perform tasks such as phoneme deletion

ii. Children enjoy adult-like speech perception from around six months old. (Note that this claim concerns speech perception and not perception of speech more generally; it may be that adults' explicit knowledge of syllables alters aspects of their auditory perception of speech in ways that do not directly relate to their ordinary ability to process speech.)

iii. Children cannot perform tasks such as phoneme deletion

iv. The best explanation for (3) is that children's speech perception does not involve phenomenal contents corresponding to phonetic gestures.

v. Therefore adult-like speech perception does not depend on phenomenal contents corresponding to phonetic gestures.

---

[1] Your partial endorsement of the quote from Matthen on p. 30 suggests that you would accept this assumption suitably modified to take account of the distinction you make between experiencing individuals and experiencing properties.

This line of response clearly does not show that your conclusion, (4), is wrong. In developing this response I raised some questions about the positive argument for (4), but these are merely open questions as far as I know (and I guess we could go quite a few rounds further on these). And I certainly take your argument to show that my position was inadequately stated and needs careful refinement. The claim I think I can support positively concerns speech perception rather than perception of speech more generally. I stand by the claim that speech perception—that is, the perceptual processes by which we are able to recognise spoken sentences—involves distinguishing phonetic gestures but it does not involve phonetic gestures appearing to the perceiver as such. It's consistent with this to suppose that in perceiving speech—in bringing all our perceptual abilities to bear on speech stimuli, including auditory abilities not directly involve in speech perception—we do experience phonetic gestures as such.

Here's a crude statement of the position I want to defend in response to your argument (1)–(4). Phonetic structure in speech perception is generally processed without the perceiver becoming aware of that structure, much as some researchers suppose that patterns of illumination characteristic of edges are processed in vision without perceivers becoming aware of those patterns of illumination. These patterns are not part of the phenomenal character of the experience.[2] In both cases, speech perception and vision, it is possible for subjects to reflect on the stimuli and to become aware of the these patterns. And subjects who have a long history of such reflection can become so skilled that they can recognise the patterns immediately. Depending on what phenomenal character, this could justify concluding that the patterns are part of the phenomenal character of these skilled individuals' experiences.

Given the aims I have, I'm happy to admit that adults overall experience of speech may involve awareness of phonetic structure. I'm even persuaded by your paper that the difference between experiencing speech in languages you do and don't know may be, at

[2] We clearly agree that there is an issue here. You write, "It is one thing to say speech perception bears information about intended articulatory gestures; it is another to say that intended articulatory gestures are among the apparent objects of perceptual experience, or that they figure as described in the phenomenal contents of perceptual experience" (ms. p. 32)

least in large part, best explained by perception of language-specific units smaller than words and patterns in the ways those units are arranged. What matters to me is rather that speech perception begins without any such awareness, and that mere speech perception does not provide knowledge of the phonemes or patterns of illumination. Instead, these have to be rediscovered by reflection on the stimuli.

There is still a point of disagreement between us. You base part of your argument on the claim that infant research shows that components of speech are experienced as similar with respect to phonetic properties (pp. 14–7). Because of (i)–(v) above, I doubt that this is correct. You write "learning a language alters one's perceptual experience of similarity and difference among its speech sounds" (p. 17). I would disagree because by "learning a language" you are referring to the development of speech perception. But if "learning a language" referred to the kind of reflective enterprise involved in acquiring literacy, I would agree. Behind all of this is a disagreement about the nature of categorical perception …

Second response: reject (1). You write, "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" (p. 16). Other philosophers also define categorical perception in terms of how things appear, as have some researchers. I don't want a barely terminological dispute, it seems to me that the term 'categorical perception' gets used in different ways. Here's a second way of thinking about categorical perception:

> Categorical perception is supposed to explain facts about phenomenological character, heightened discrimination, pop-out effects and verbal communication. It is these facts-to-be-explained that give us a fix on what categorical perception is. Confidence that there is any such thing as categorical perception hinges on the conjecture that explanations of these various facts in one or more of the several putative cases (speech, colour and the rest) share some unifying features.

On this second way of thinking about categorical perception, categorical perception is an explanatory notion: it is whatever explains heightened discrimination, pop-out and the rest. How things appear is just one of several facts to be explained, and

distinctive phenomenology is not obviously a feature of categorical perception in these senses: not all cases of categorical perception need involve phenomenology (e.g. some researchers claim to have identified categorical perception in frogs and crickets), and where categorical perception does involve patterns of experienced similarity, this may be a fairly far-removed effect of categorical perception plus cognition. So we should not assume that the two ways of thinking about "categorical perception" are ways of thinking about the same thing. Also, on the second, explanatory way of thinking about categorical perception, it is plausible that what is categorical is not a subjects' overall perception of a stimulus but rather only one of several concurrent perceptual processes. So, for example, it is probably a mistake to think that when a subject experiences a blue stimulus, we can ask whether her total experience is categorical. We can only ask whether any of the perceptual processes are categorical. One reason why categorical perception is hard to study empirically is the difficulty of isolating perceptual processes from each other and from memory and cognition.

So far I've suggested that there are two ways of thinking about categorical perception that are not obviously ways of thinking about one and the same thing. For arguments based on empirical research on categorical perception, it seems to me that the second way of thinking about categorical perception is the right one. On this way of thinking about categorical perception, is it true that categorical perception "yields a distinctive pattern of experienced similarity and difference relations"? My own view is that categorical perception involves similarity of experiences rather than experiences of similarity. Put crudely: colour stimuli A, B and C give rise to experiences A', B' and C'. Subjects report that A and B appear similar to each other and that each appears dissimilar to C. The standard view is that these experiences involve awareness of colour categories. That is, A', B' and C' are experiences which involve awareness of the respects in which A, B and C are similar and dissimilar. This would be experience of similarity. My counter suggestion is that the experiences are similar to each other in much the way that sensations can be similar (for example, two stinging sensations can more or less similar to each other). What gives rise to the judgement that A appears similar to B is just that A' is similar to B', not that A' is an experience of some respect in which A is similar to anything else. I don't deny that skilled subjects' overall experiences of the stimuli might involve appearances of similarity; I

just think that these appearances involve more than categorical perception of the stimuli. Setting this out and defending it takes a while; I could send you a draft if you could bear to read more.

***

I also have a general comment on your paper. I take the primary question to be what explains why "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" (p. 6). The answer is, roughly, that knowing a language involves knowing its phonemes and how they are characteristically organised in that language, and experiencing these when hearing speakers of that language. It should be possible to distinguish this answer from one which invokes semantic knowledge in a direct way with generating sentences of novel nonsense words that the right sort of prosodic and phonetic structures to be sentences of a particular language. For your position implies that the experience of listening to fake sentences compatible with your own language differs from the experience of listening to fake sentences not compatible with your own language in the same way that experience of speech in known and unknown languages differ, whereas the semantic theorists' position implies the converse. Of course this is just what you aim to establish with aphasias. But using artificial stimuli allows for finer-grained studies. It might allow one to test exactly which features of speech are responsible for the phenomenological difference.

Small points:
p. 8: "What is the auditory perceptual experience of meaningfulness or of a meaning like? What are the recognizably audible characteristics of meanings?" It's not obvious to me that your opponent would have to regard perceptual experiences of meaningfulness as auditory. After all, she might think that some phenomenological differences between experiences of known and unknown languages are not auditory.

p. 10: "temporal characteristics". What is the evidence that we actually perceive pauses etc? In principle it could be that we think we experience pauses but we don't. (Curiously, on a relationalist view of perception it would seem that we can't experience pauses that aren't actually in the stimuli.)

p. 13: “So, if the experience of listening to spoken language with PWD differs from the experience of listening to spoken language with TSA, then …”. The next paragraph assumes the antecedent of this conditional is true (“This shows the cognitive account …”) but you haven’t supported it I think.

Right, sorry for the long letter and many thanks again for the opportunity to read your paper. I have learnt a lot from it (and your other paper on speech) and very much hope we might be in further contact.

best wishes,
Steve