

## Against hearing phonemes - A note on O'Callaghan

C. Naomi Osorio-Kupferblum, University of Vienna

### Abstract:

Casey O'Callaghan has argued that rather than hearing meanings, we hear phonemes. In this note I argue that valuable though they are in an account of speech perception – depending on how we define 'hearing' – phonemes either don't explain enough or they go too far. So, they are not the right tool for his criticism of the semantic perceptual account (SPA).

### Article:

Casey O'Callaghan has taken issue with what he calls the semantic perceptual account (SPA) in various papers ((2009), (2010), (2011), (2015); this note relies mostly on (2011).) The account considers the phenomenological difference between hearing a language we understand and one we don't, to be an indication that we actually hear meanings.<sup>1</sup> O'Callaghan agrees that there is indeed such a phenomenological difference, and, it seems to me, anyone who has ever learnt a foreign language will concur. When we first hear the language, it appears like an uninterrupted stream of sounds and what we discern in it is prosody and a couple of phonetic elements, a particular vowel or consonant perhaps, that are salient – maybe because they recur often or because they are unusual to our ears, or both. In the early stage of learning the language when we are taught individual words, we begin to pick them out from the stream of sounds, and with time, we gradually learn to discern most of the words and phrases, even if we don't always know what they mean.

The SPA attributes the phenomenological difference between the stream of sounds we heard when we first heard the language and what we hear once we know it to the fact that we understand it. Against this view, O'Callaghan argues forcefully that we don't hear meanings. His argument stresses that understanding what someone says is importantly different from perceiving our environment auditorily. First, he states what it is to form a belief based on what one perceives:

(PE) Imagining aside, I cannot have a perceptual experience in which I perceptually entertain that something is the case, or is present, without having a perceptual experience which purports that this is the case or is present. (2011:793)

---

<sup>1</sup> In his (2011:783) he specifies that this is to be understood as being "auditorily perceptually aware of [an utterance's] meaning or semantic properties".

So, the idea is that we cannot think we heard the door shut with a bang without having heard the bang, for example. His argument now runs as follows:

- (1) Understanding an utterance *u* which states that *p* means grasping the meaning of *u*, namely *p*.
- (2) However, when I hear *u*, I do not auditorily grasp, represent, experience or enjoy awareness as of *p*.<sup>2</sup>
- (3) So I don't hear meanings.

Continuing our example, this means that when we hear you say "The door has banged shut", we hear that you said that the door has banged shut, but we don't hear the bang of the door. But the meaning of your utterance just is that the door has banged shut. Therefore, if that is not what we hear, we don't hear meanings. So we don't hear meanings, and the SPA is mistaken.

O'Callaghan goes on to develop an alternative account to the effect that we hear phonemes that we will get to below. Let us first analyse the attack against the SPA.

The first thing that is noticeable is that the argument only works where we can apply a naïve verificationist truth-conditional account of meaning, i.e. for present-tense, indicative assertions about things in our surroundings. It is evidently false for speech acts, but, it seems to me, also for any utterance that expresses the speaker's mental state. It seems quite plausible to claim that hearing you say "Can I have a biscuit?" just is hearing that you want a biscuit. But O'Callaghan is aware that he is caricaturing the account<sup>3</sup> and it serves as an exposition of the line his argument takes.

O'Callaghan's *modus tollens* argument has a distinctly Pittsburghian ring to it, in that it starts from the belief.<sup>4</sup> In many contexts, this move skilfully evades the sceptic, but I propose to turn the argument around and employ sceptical doubt to tease out the direction our investigation needs to take.

Hearing what goes on around us may prompt us to form beliefs. We sometimes have reason to doubt the accuracy of what we believe ourselves to be hearing and suspend judgment until we have tested our belief. But we will test our beliefs differently when we are not sure we heard the door bang shut from when we are not sure you said that the door banged shut. In the first

---

<sup>2</sup> *ibid.*

<sup>3</sup> At the "Perceptions and Concepts" symposium (Riga, 2013), he calls it a sophism. His talk can be seen here: <https://www.youtube.com/watch?v=7OFcknu5tgo> (retrieved on 26 April 2017)

<sup>4</sup> I have argued (in a different logical form) against the conflation of hearing *u* with hearing *p* in Ruth Millikan's work (2013).

case, we may look whether the door is shut, or what else may have made the bang; in the second case, we would ask you to repeat what you have said. If we checked whether the door is shut in the second case, we would be testing the accuracy of your statement, not of our auditory perception. Of course, we may be in a situation where we aren't sure we have understood you correctly (either because we had trouble discerning your words acoustically or because we don't know your language well) and we may be too shy to ask you to repeat and therefore prefer to check the door instead. But doing this means that we assume a whole baggage of prior beliefs – that you were trying to inform us of something and therefore truthfully described an event in our surroundings (to give a rough summary) – and it does not alter the fact that we are in the first place testing our understanding of your utterance, even if the way we are doing so is by checking whether what we believe you may have said is actually the case. If, in the second case, we find the door is still open, we will have a choice between thinking you were lying or we've misunderstood you. In the first case, the choice is between thinking something else made the bang or the door did bang shut but was then opened again.

So, taken like this, it is easy to agree with O'Callaghan that the object of perception in hearing somebody say something is the utterance, not what makes the utterance true. This means that the issue is between hearing and understanding.

O'Callaghan focuses on hearing. He asks: If we don't hear meanings, what is the difference between hearing a familiar language and hearing a language we don't understand due to? He points out that every language has its own distinctive system of sounds, not only in terms of what sounds are used to make up words and how they combine, but also within what ranges variation of those sounds is possible without changing the meaning of the word they constitute. The smallest units of sounds in a language whose change makes for a change of meaning is called phoneme. O'Callaghan stresses that human infants take a special interest in language<sup>5</sup> and learn to discern phonemes of the language(s) they are born into very early, while at the same time beginning to disregard non-phonemic differences in sounds. So, learning a language means acquiring the skill to distinguish its phonetic structure at the expense of discerning sounds that are not phonemic in any language we are familiar with – we learn to hear phonemes. Backed up with examples that show that polysemy does not change the phenomenological experience, O'Callaghan considers his explanation superior to the SPA.

---

<sup>5</sup> This is surely less surprising than O'Callaghan seems to think (I infer that from how he stresses the fact) considering that foetuses hear as of the end of the second trimester, and what they hear most, apart from digestive and breathing sounds, is their mother talking (cf. Birnholz & Benacerraf 1983).

In order to assess his claim, we first need to define 'hearing' with respect to the phenomenal experience it provides and how it relates to understanding. Since the demise of the unfortunate sense-data theory, it seems to me that relative to issues like ours we have a choice between two basic options. The first is a wide-reaching option that comprises the entire mental state of the hearing subject caused by the auditory input, that I shall call the "thick option". It builds on O'Callaghan's explanation that experience is to be understood "in the broadest possible sense, so that it may encompass, for instance, sensory, perceptual, bodily, affective, emotional, imaginative and even occurrent cognitive events or states" (2011:785), and phenomenal difference as "a difference in what it is like for you as a conscious subject to have each experience" (ibid.). Its focus is on the experience and what caused it. The second option comprises merely the auditory impact our environment makes on us, the "thin option". Its focus is on the cause and its immediate effect on the perceiver.

The "thick" option therefore takes into account in a generous way what it is like for a subject to be in a state of hearing certain sounds. For instance, it would regard the shiver going down my spine at hearing a fingernail scratch a blackboard as part of what hearing that sound comprises. The "thickest" option would include the recognition that the sound is produced by a fingernail scratching a board in the phenomenological status of the subject.

In our example of a familiar language, this option would go far beyond discerning phonemes and hearing speech as a sequence of words, with pauses, coughs, uhms and ahs. It would have to comprise also the analogues to the shiver and recognising the fingernail. But what produces the analogue of the shiver? There is the sound of the voice, which conveys the speaker's emotional state and which can produce an emotional response in the hearer. But that can't be all, for compare someone screaming "My knee!" with the same person in the same tone of voice screaming "There's a fire!" The first scream is likely to produce the immediate effect of compassion, while the second scream will cause fear in the hearer. But these effects are, of course, linked to understanding what is said. In fact, it seems to me that the analogue of recognising the screech as the sound of the fingernail scratching the board is recognising the scream "My knee!" as the sound people make when they hit their knees and thereby grasping the meaning of the utterance. The "thick" mental state produced by hearing an utterance simply comprises understanding what is said.

For support of this claim, since we are more accustomed to thinking about vision than audition, let us take an example from seeing language, as it were, the Stroop test. It has names of colours printed in colours other than the one they denote; for instance "red" would be printed in yellow

ink, “yellow” in green ink, “green” in blue ink, etc. The task is to tell the colour of the ink. When a sheet of such words is shown to a person who can either not read or doesn’t understand the colour words (because they are in a foreign language), the task is very easy. But for fluent readers of the relevant language, it is really hard. When I tried it, I felt that the word I grasped at a glance interfered with my perception of the colour of the ink. I had two contradictory visual pieces of information and although I knew which one to go for, the other one was impossible to block. The visual experience was equally immediate, and, moreover, concerned the same issue. Now, if the meaning of words composed of written letters is grasped so quickly that it interferes with the perception of colours, it is likely that the meaning of words composed of phonemes would be grasped just as quickly.

Why understanding language can be as immediate as grasping anything else in perception is a topic for another day. For the thick option, where the subject’s entire mental state resulting from hearing a language utterance is taken into account, the Stroop test supports the claim that understanding what the utterance means must be part of hearing it.

The thick option would then still have to deal with O’Callaghan’s argument. It can do so by referring to the form / content dichotomy and give (PE) another twist: There is no way to perceive the content of an utterance other than through perceptual experience of the form. But the Stroop test is an indication that once we have acquired the relevant interpretive skills, we immediately perceive the content, not the form, and it takes a big effort to block perception of the content when we perceive the form.

So, on the “thick” account, SPA is right and O’Callaghan’s attack misses the point.

But O’Callaghan seems to favour some version of the “thin” option which keeps all cognitive contributions to mental states out. It must therefore focus exclusively on the acoustic effect. But as we are concerned with perceptual experience, we don’t want to speak only of the mechanical effect: soundwaves hitting the eardrum at a certain Hertz rate and being transmitted to the brain by the vestibulocochlear nerve, etc. That would be insufficient for an account of hearing; instead, such an account must also include awareness of what is thus perceived, the “sensory mode of presentation” (as Ayers puts it in a very similar account in 2004:249) – we hear music, we hear noises and we hear someone speak. The thin account leaves out the subject’s mental state beyond this bare minimum, but it packs the source of the sound into its concept of hearing (if it didn’t, it would amount to a mere sense-data theory).

But if this is the idea, O’Callaghan’s account seems to involve too much. Phonemes are the smallest phonetic units that determine meaning. They are as much semantic building blocks as letters are in written language. Just as the thick account argued that we cannot hear phonemes without hearing words and grasping their meaning, the thin account must now argue that we cannot discern phonemes without having acquired the skill to do so; but this skill was acquired jointly with knowledge of the relevant language. If the thin account wants to keep cognitive mental states out, it must stick to form. But phonemes are the first step into content and cannot be isolated from it by the hearer.

If the skill to discern phonemes is what makes hearing a familiar language phenomenally different from hearing an unfamiliar language, that skill is inseparable from understanding the language both in origin and in its employment when we hear an utterance. The thin account is not entitled to take recourse to phonemes and without them, it can account for the difference between hearing a person sing or speak, for instance, but not between hearing familiar and unfamiliar languages. An example of what the thin account can deliver is hearing sounds in the room next door and being able to tell language apart from music or the sounds an animal makes.

So, in spite of their importance in speech perception, phonemes are not the right tool to argue against SPA. For the thick account of hearing, they don’t comprise enough, and for the thin account they go too far.

### Acknowledgements

GRATIASAGOANIMADVERSIONIBUSEXCELLENTIBUSUTILLIMISQUEMICHAELO  
AYERS

### Literature

Ayers, Michael (2004) “Sense experience, concepts and content – Objections to Davidson and McDowell”, in: Schumacher, Ralph (ed.) *Perception and reality*. Paderborn: Mentis, 239-262.

Birnholz, J.C. and Benacerraf, B.R. (1983) “The development of human fetal hearing.” *Science*, 222, 516-519.

O’Callaghan, Casey (2009) “Is speech special?” *UBC Working Papers in Linguistics* 24, 57-64.

- (2010) “Experiencing speech” *Philosophical Issues*, 20(1), 305-332.
- (2011) “Against hearing meanings” *The Philosophical Quarterly*, 61(245), 783-807.
- (2015) “Speech perception” in: Mohan Matthan (ed.) *Oxford Handbook of the Philosophy of Perception*, Oxford: OUP, 475-494.

Osorio-Kupferblum, Naomi (2013) “Hearing it rain – Millikan on language learning” *Beiträge der Österreichischen Ludwig Wittgenstein Gesellschaft* 21.