Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

What Do We Experience When Listening to a Familiar Language?

ANNA DROŻDŻOWICZ*
University of Oslo, Oslo, Norway

What do we systematically experience when hearing an utterance in a familiar language? A popular and intuitive answer has it that we experience understanding an utterance or what the speaker said or communicated by uttering a sentence. Understanding a meaning conveyed by the speaker is an important element of linguistic communication that might be experienced in such cases. However, in this paper I argue that two other elements that typically accompany the production of spoken linguistic utterances should be carefully considered when we address the question of what is systematically experienced when we listen to utterances in a familiar language. First, when we listen to a familiar language we register various prosodic phenomena that speakers routinely produce. Second, we typically register stable vocal characteristics of speakers, such as pitch, tempo or accent, that are often systematically related to various properties of the speaker. Thus, the answer to the question of what we typically experience when listening to a familiar language is likely to be a complex one. Dedicated attention is needed to understand the nature and scope of phenomenology that pertains to linguistic communication. This paper lays some groundwork for that project.

Keywords: Linguistic understanding, experiences, prosody, linguistic communication.

* I would like to thank Nick Allott, David Pereplyotchik and Deirdre Wilson for detailed comments on this paper, as well as Barry Smith and the participants of the Philosophy of Language and Linguistics workshop in Dubrovnik in 2019 for helpful discussions on the topic. I am grateful to Dunja Jutronić for her help with preparing the draft. I would like to specially acknowledge the Institute of Philosophy in London, where I was a visiting researcher between 2018-2020. This work was supported by the Mobility Grant Fellowship Programme (FRICON) funded by The Research Council of Norway and the Marie Skłodowska-Curie Actions (project number: 275251).

1. *Introduction*

You are at the seaside leisurely flicking through a magazine. A friend calls out your name and asks if you would like to take another swim. There is energy and enthusiasm in her voice, and you have a strong impression that she is inviting you because she wants to take a swim as well. Your answer will probably be a quick cheerful 'yes'.

What do we experience when listening to a familiar language? For example, what do we experience when we hear a friend enthusiastically inviting us to take a swim? There is a *liberal* strategy for answering that question. After all, there are so many things one might experience when listening to an utterance produced in a familiar language. An endless variety of impressions may arise when we hear others speaking and communicate with them. Upon hearing your friend's invitation to take a swim, you might have a vivid memory of last year's summer holidays. An invitation might remind you of a scene from your favourite movie. You might feel relaxed and safe. You might feel threatened, as in the case when, unbeknownst to your friend, you nearly drowned during your last swim. You might have an unpleasant impression that your friend is nudging you to embrace a certain kind of healthy lifestyle, or you might have an impression that you and your friend have the same needs.

This would then be the answer provided by those who adopt the liberal strategy: virtually anything might be experienced when listening to a familiar language. However, the liberal strategy strikes me as an evasive one. Such varied experiences and impressions as those described above can, and perhaps often do arise when we listen to others speaking in a familiar language, but they *need* not. Moreover, when they do arise, they often do so with no specific regularity or order. A more fruitful way of addressing the question of what we experience when listening to a familiar language is to focus on elements that might be *systematically* experienced. When I ask what we experience when listening to a familiar language in this paper, I am asking about, and will consider, only those elements that typically and systematically arise in linguistic communication and could thus typically be part of our overall conscious experience.

How can we decide whether something is a candidate for being experienced *systematically*? The elements that I will consider in this paper are those that: (a) result from forms of expression and information transfer that are available to typically developed speakers, and (b) that typically developed hearers register and experience thanks to specific psychological and linguistic mechanisms employed in voice perception and spoken linguistic communication. Systematicity, as understood here, will not imply that such elements would need to be present in our experience on each and every occasion. For example, if a speaker does not convey or reveal some information when producing a linguistic utterance, a hearer will not register that information. Thus, I will not

go so far as to argue that the elements discussed in this paper would be necessary and sufficient for the case described. Linguistic communication and its phenomenology are complex phenomena and providing such conditions would be an overly ambitious task, at least in one go. Still, many real-life phenomena occur sufficiently frequently in normal conditions to warrant philosophical attention, even though exceptions may arise.

So what do we *systematically* experience in typical cases of hearing an utterance in a familiar language? One strategy would be to provide a quick and intuitive answer: we experience understanding an utterance or what the speaker said or communicated by uttering a sentence. When my friend asks me whether I would like to take a swim, I have an experience of her asking whether I would like to take a swim, or, in other words, I have an experience of understanding that she asked me whether I would like to take another swim and in that way invited me to do so.[1] This intuitive answer seems both quick and simple. In recent literature on the epistemology of linguistic understanding, experiences of what was said with an utterance, experiences of meanings communicated with utterances or experiences of understanding, depending on one's preferred terminology, have been a subject of intense debate. As we shall see (section 2) neither explaining the nature of such states nor explaining their epistemic roles is an easy task. This might be one reason why most discussions that might provide insights into the question of what we experience when listening to a familiar language have so far focused on experiences of meanings communicated with an utterance (or of linguistic understanding).

This focus may suggest, that the question about the phenomenology of linguistic communication boils down to whatever we can systematically say about the phenomenology of meanings or the phenomenology of understanding. In this paper, I will argue that this is not the case. The intuitive strategy described above would be too *restrictive* to provide a satisfying answer to the question of what we systematically experience when listening to linguistic utterances in a familiar language. Understanding or grasping a meaning conveyed by the speaker might of course be an important element of what we experience there and then. But in the course of this paper I will provide evidence that apart from this, two *other* elements that typically accompany the production of spoken linguistic utterances should be considered as candidates for what is systematically experienced when we listen to linguistic utterances in a familiar language.

First, I will argue that when we listen to a familiar language we also typically register a variety of *prosodic phenomena* that speakers routinely produce, in both controlled and spontaneous manner (Whar-

[1] In this and other cases, both the content of what was said with an utterance and its force, in this case an invitation, are commonly taken to be experienced (e.g. Fricker 2003).

ton 2009). Prosodic phenomena take many forms and often make important contributions to linguistic communication, affecting the everyday interactions involved. As such, they should be considered when addressing the question of what is systematically experienced by hearers. A friendly and enthusiastic invitation to take a swim sounds different from an indifferent one, or from one that merely attempts to sound friendly and enthusiastic.

Second, I will argue that, when we listen to linguistic utterances in a familiar language, we typically also register *stable vocal characteristics* of speakers, such as pitch and tempo, which are determined by the physiology of the speaker's vocal apparatus and the circumstances of vocal production, as well as vocal characteristics that result from sociolinguistic environment, such as accent. Importantly, such stable vocal characteristics are often systematically related to various properties of the speaker, such as sex, age, ethnicity (Belin et al. 2004, Baumann and Belin 2010; Owren et al. 2007; Mulac and Giles 1996; Rakić 2019). Thus, hearing a linguistic utterance in a particular voice will normally reveal a lot of important information about the speaker that the speaker does not intend to communicate.[2]

The evidence presented in this paper will show that the answer to the question of what we typically experience when listening to a familiar language is likely to be a complex one. The phenomenology of linguistic communication is probably richer than many mainstream philosophical debates on linguistic understanding would suggest. Careful investigation of whether and to what extent the three elements presented in the paper could enter heares' consciousness and what signalling functions they have is needed. This kind of investigation will require a detailed analysis of both philosophical and empirical arguments and goes beyond the scope of this paper. The goal of this paper is to lay some groundwork for the project of addressing the question of what we systematically experience when listening to a familiar language by pointing to new areas of research. For readers who are not immediately excited and curious about the question discussed in this paper, I would like to offer a brief explanation of why I think it is both interesting and important, thus making the task worthwhile.

First, it is interesting and important to describe and understand what we could systematically experience when listening to a familiar language. While a lot of philosophical attention has been paid to describing and explaining the nature of visual experiences and (somewhat less) to the nature of auditory experience of environmental sounds, apart from debates on experiences of understanding, there has been relatively less interest in the philosophical investigation of auditory experiences of listening to a familiar language (for rare ex-

[2] In this paper I focus entirely on spoken language and its phenomenology. Parallels and differences in the typical phenomenology of reading require separate discussion.

ceptions see: Smith 2009, Di Bona 2017, Drożdżowicz 2020). And yet such experiences are without a doubt part and parcel of our auditory environment, given that linguistic communication is a pervasive form of sharing information among humans. Second, systematic impressions of speakers can and often do play a role in linguistic communication by steering, affecting and sometimes biasing social interactions that rely on it. In the example above, specific properties of your friend's voice, e.g. those relating to the length of her vocal tract, determine how she sounds, namely as a middle-aged woman. Those properties allow you to recognize the voice that utters the invitation as the one belonging to your friend. In order to understand how we interact linguistically, dedicated attention is needed to provide an inclusive picture of what is experienced in linguistic communication.

The paper is structured into three main sections, each presenting one of the three elements that are typically registered when listening to an utterance in a familiar language. In section 2 I present recent debates concerning experiences of understanding what was said or communicated with an utterance. I summarize the main results concerning this topic and focus on pointing out questions for future research. I will briefly sketch my own view on the matter but will not defend it in detail here. In section 3 I present evidence which supports the claim that when we listen to a familiar language we commonly register a whole variety of prosodic phenomena and argue that they are another candidate to consider for what we could systematically experience in linguistic communication. Section 4 presents evidence for the claim that we also systematically register stable vocal characteristics of a speaker that reveal important information about them. Impressions based on such characteristics are yet another candidate for what we could systematically experience. I conclude in section 5.

2. *Experiencing linguistic understanding*

Setting aside cases of unsuccessful communication, when you hear an utterance in a familiar language, you typically come to understand *what the speaker communicated* with that utterance on a particular occasion. Many have argued that in such cases competent language users experience states of understanding linguistic utterances (Hunter 1998; Fricker 2003; Reiland 2015; Nes 2016; Brogaard 2018). Simple though it is, this observation has been a starting point for intense debates about the nature and epistemic roles of the experience in question.

Such experiences are commonly illustrated using so-called *contrast cases*. Imagine again hearing your friend utter the invitation "Would you like to take another swim?". Imagine now that everything is the same, except that your friend is speaking in a language totally unfamiliar to you, and says the same thing, i.e. invites you to take another swim. What will strike you is that the experience you have when listening to the second utterance in a language you do not know differs dra-

matically from the experience of a language you understand and speak fluently.[3] This observation is typically taken to be important *prima facie* (although not the only) evidence for the claim that experiences of linguistic understanding exist.[4]

Here is how such experiences are often portrayed in the current debate. It is generally agreed that when a hearer has an experience of understanding an utterance, she grasps at least one proposition, which would roughly correspond to the asserted meaning of that utterance. Most participants in this debate seem also to agree that typically such experiences would involve or somehow indicate grasping the enriched, saturated meaning of an utterance and not the minimal meaning (Fricker 2003; Nes 2016; Brogaard 2018; Gasparri and Murez 2019). For example, when you have an experience of understanding your friend uttering "Would you like to take another swim?", as in the above situation, you grasp, among other things, that: "you", in this context, refers to yourself, not to anyone else who might be at the beach, that the invitation, in this context, is to take a swim now or in the near future and not at just any point in the future, etc. Apart from rare occasions, hearers do not grasp minimal meanings with unassigned referents and unresolved ambiguities (Smith 2010).

When characterizing experiences of understanding linguistic utterances, it is usual to list their *involuntary* nature, *automaticity,* and prima facie *compelling* character. Normally we have little or no control over whether upon listening to an utterance in a familiar language we experience understanding it or not (Fodor 1983: 52). Such experiences are usually taken to arise spontaneously and automatically as soon as an utterance is heard. According to many, their immediate presence serves as *prima facie* compelling evidence for beliefs about what a speaker intended to communicate with an utterance (e.g. Fricker 2003; Brogaard 2018). Those features have been argued by some to support the idea that experiences of meanings or of linguistic understanding are interestingly similar to paradigmatic perceptual experiences (Hunter 1998; Bayne 2009; Siegel 2006; Brogaard 2018), but the exact nature of this similarity has been a subject of considerable debate.

At least three sets of questions have animated recent debates about experiences of linguistic understanding. These concern: (1) their nature, (2) their epistemic roles, and (3) the methods used to investigate them. Starting with (1), according to what has been described as the *semantic perceptual view,* properties like having meaning x can be represented in the hearer's perceptual experience (Siegel 2011; Bayne

[3] Moreover, typically utterances in unfamiliar languages are not perceived as strings of words, given that hearers are not sensitive to phonemes of a particular language (O'Callaghan 2011).

[4] This is a minimal commitment that many in this debate accept. Other evidence comes from the phenomenological shift that occurs when listening to sinewave speech (Remez et al., 1981). For a detailed discussion of this case and evidence it provides, see O'Callaghan (2011; 2015).

2009; Brogaard 2018, 2019). On that approach, the nature of experiences of utterance understanding is *perceptual* (or of perceptual seemings, Brogaard 2018). The experience we have when listening to a familiar language is an experience of *hearing meanings.* Arguably, the properties that are perceived in this case are many: Tim Bayne (2009) observes that we perceive "*both* (low-level) changes in phonological structure *and* (high-level) semantic properties".[5] In order to have an experience of understanding an utterance in a familiar language, a hearer has to perceive the phonetic and phonological properties of an utterance (O'Callaghan 2015). But it is a contentious matter whether in such cases meaning properties are also perceived, and in recent years the semantic perceptual view has been the subject of some criticism.

According to Casey O'Callaghan (2011), the contrast cases that are often presented in favour of the semantic perceptual view are best explained in terms of differences in experiencing low-level phonological properties of linguistic utterances and thus do not support claims about semantic perception. The view has also been criticized on epistemological grounds. Brendan Balcerak Jackson (2019) has recently argued that the claim that experiences of hearing meaning provide immediate justification for hearers' beliefs about what a speaker communicated with an utterance (e.g. Brogaard 2018) is unfounded. Another contentious issue has been whether the semantic perceptual view is psychologically realistic and could be made compatible with our best knowledge about the psychology of linguistic comprehension (Drożdżowicz 2019), as well as whether it can accommodate the systematic role of context and background knowledge in linguistic communication (Brogaard 2018, 2019; cf Gasparri and Murez 2019). A related contender in this debate, the view that experiences of linguistic understanding are instances of *cognitive* phenomenology characteristic of cognitive states (Strawson 2010; Siewert 1998; Dodd 2014), has also been subject to criticisms on parallel grounds (Prinz 2011; see also Montague 2017).[6]

A quite different approach to the nature of such experiences maintains that meanings (or thoughts) are not the salient contents of such experiences, but rather that what hearers experience is fluency of understanding. On my own view (Drożdżowicz forthcoming), experiences of understanding are *epistemic feelings of linguistic fluency* that result from evaluative monitoring processes.[7] There is extensive evidence that such processes are typically involved in utterance comprehension (Nozarri and Novick 2017; Pickering and Garrod 2013). The perceptual appearance of understanding experiences is, on this view, explained as resulting from the deployment of early-stage auditory processes of speech perception. Thus, on my proposal, experiences of linguistic understand-

[5] There might also be other properties (morphological, syntactic) to consider.

[6] This passage draws on section 2 from Drożdżowicz 2019.

[7] For the purpose of this paper I briefly mention my view as one of the contenders in this debate. I defend it in detail in another paper (Drożdżowicz forthcoming).

ing are first and foremost metacognitive feelings that reveal the degree of the success in comprehending an utterance. For example, when you hear a friend inviting you to take another swim, you might have an experience of understanding that amounts to a quick immediate signal indicating that you have successfully comprehended your friend's invitation and can immediately proceed to produce an answer, act on that invitation, etc. For these purposes, you do not have to represent or reconstruct the communicated meaning of her utterance as part of your conscious experience, but simply take for granted the feeling that you got the message of her utterance right, and act on it.[8] Although the outcomes of the ongoing debate on the nature of experiences of understanding still remain to be seen, the most discussed semantic perceptual view is currently under a lot of pressure. This opens up space for new contenders and calls for further investigation of question (1).

Questions concerning the epistemic roles of experiences of understanding (2), or of meanings conveyed with linguistic utterances, are of immediate interest to epistemologists working on linguistic communication and testimony. In recent debates on the epistemology of language understanding, it has been argued that such experiences: (a) justify beliefs about what a speaker communicated or said with an utterance (Hunter 1998; Brogaard 2018: 2969); (b) provide justification that is necessary for acquiring knowledge about what a speaker said (Fricker 2003: 345)[9]; (c) amount to what states of language understanding are (Pettit 2002: 544); and perhaps (d) trigger the "content entertaining" states of understanding (Longworth 2018: 825).[10]

Let us look at some evidence presented in favour of the claim that experiences of linguistic understanding provide important, justification for beliefs (Brogaard 2018) and/or knowledge about what the speaker conveyed with an utterance (Fricker 2003). Assuming that we are in a *typical* communicative context, i.e. one where both speaker and hearer are using the same language in a cooperative way (Fricker 2003: 332), what could be the epistemic contribution of an experience of understanding an utterance produced by the speaker? In our toy example, your friend asks you: "Would you like to take another swim?". Accord-

[8] This is compatible with the fact that in other, albeit less prevalent, cases of obstructed communication you might need to reflectively reconstruct the meaning you have grasped.

[9] Fricker (2003) uses the notion of what is said when describing such knowledge and beliefs, but her clarification of experiences of understanding (and corresponding beliefs) suggests that their contents concern not just what is strictly speaking said with an utterance, but what a speaker intended to communicate with an utterance on a particular occasion, leaving it open whether all pragmatic meanings (e.g. implicatures, metaphors) can be experienced in a similar manner. A similar notion is used in Brogaard (2018).

[10] Guy Longworth's notion of 'perceptual encounter with an utterance' seems parallel to the notion of experience of understanding. On the other hand, his content-entertaining states seem to involve both perceptual and belief-like elements, so they might perhaps be closer to the experiences discussed here.

ing to Fricker (2003) and Brogaard (2018), in this case your experience of understanding what she communicated with that utterance provides an immediate *prima facie* justification for your belief about what she said,[11] i.e. that she has invited you to take another swim. However, having a relevant experience of understanding would rarely suffice for your belief to be justified or to afford knowledge. On their view, you would also need to have secured a kind of warrant that is captured in broadly externalist or reliabilist terms. For example, one could maintain that for such a belief to be justified, a hearer must exercise reliably functioning linguistic capacities. Lucky beliefs, based on matching experiences of understanding produced by the unreliable workings of the language system, typically would not count as justified.

How then should we understand the strength of the claim that experiences of understanding are normally required for justified beliefs (and/or knowledge) about what a speaker said? Fricker's case of Ida (2003), initially presented as an argument against the reliabilist conception of language understanding (e.g. Schiffer 1987), is often discussed in this context:

> **IDA:** Ida has an internal, module-like device implemented in her brain that provides her with correct beliefs about the meanings of utterances in Russian: "When Ida hears a sentence in Russian, it sounds like meaningless noise to her. Yet after hearing it, she finds herself with a strong inclination to believe that a certain speech act has been effected by that – to her – meaningless burst of noise. Ida instantiates the correct 'template'. Her beliefs about what is said in Russian utterances are due to a language processing, belief-generating module in her, and are reliably true." (Fricker 2003: 337)

According to Fricker, without experiences of understanding Russian utterances, Ida cannot know what Russian speakers say. The example is used by Fricker to support her claim that "the phenomenology of understanding *is essential* to how knowledge of what is said is gained, in normal language use" (345). Although Ida satisfies the reliabilist conception of understanding, she does not have any reasons available to her to support her beliefs about what utterances in Russian mean. Fricker argues that, intuitively, Ida does not know what utterances in Russian mean. On the contrary, utterances in Russian sound meaningless to her. Ida's case is different from the case of normal language understanding because her way of gaining reliable beliefs about meaning is "phenomenally lacking".[12]

I would like to suggest that the epistemic benefit that typical language users seem to have over Ida in this case could perhaps be cap-

[11] In these discussions a distinction between beliefs about what a speaker said with an utterance and beliefs in the content of their utterance is typically assumed.

[12] As presented in this scenario, Ida is of course very different from typical language users who have a normally developed language system. The dialectic role of this example is therefore constrained by the stipulations about how Ida's module works.

tured by the notion of *doxastic* justification. A belief is doxastically justified to the extent that it is epistemically supported by the reasons on which the agent bases it (Dormandy 2018: 77). Although Ida's beliefs about what Russian speakers say are reliably causally sustained by the workings of her internal module, by their very definition, the workings of her module do not fulfil the condition for doxastic justification, i.e. they cannot be *treated as a reason* for her beliefs. But an observation-based meta-belief that she is reliably forming such beliefs due to the inner workings of her module could provide such a reason. One could argue that this would be an improvement on Ida's current epistemic situation. From a believer's perspective, obtaining any good reason can increase the durability and confidence of an agent's belief (Dormandy 2018). Cases like IDA should not lead us to conclude that experiences of understanding are strictly speaking necessary for justified belief and/or knowledge about what a speaker conveyed with an utterance. After all, there may be other possibly valuable ways of forming such beliefs. Consider the following hypothetical case:

> **ADA:** Ada has just met a Portuguese friend. She does not speak Portuguese. However, there is a language pill she could take that would make her acquire reliable true beliefs about the meanings of utterances in Portuguese. Ada would be instantiating a correct template from utterances in Portuguese to beliefs about what these utterances mean. There is only one potential downside - the pill does not induce the appropriate, typical phenomenology of understanding Portuguese utterances that most speakers of Portuguese typically enjoy. Should Ada take the pill?[13]

Intuitively, Ada would be better off if she took the pill, for she could then come to correctly and reliably believe what her Portuguese friend is saying. Even without the typical corresponding phenomenology of understanding, Ada would have *more understanding* of Portuguese than before taking the pill. Ada's beliefs about what Portuguese speakers say could be epistemically supported by her meta-belief that the pill allows her to reliably form such beliefs. Notwithstanding the intuitive verdict that Ada should take the pill, her epistemic situation would still be quite different from that of typical Portuguese speakers. There seem to be some epistemic benefits which would not be available to her, namely, an experience of understanding that could also doxastically support a corresponding belief.

Arguably, a somewhat different story about the epistemic role of experiences of understanding would accompany the view that they are epistemic feelings of linguistic fluency (Drożdżowicz forthcoming), since on that view such experiences do not present meanings as their contents but merely signal the fluency of their comprehension. Many epistemic feelings can be useful for deliberate metacognitive purposes, such as guiding a subject's attention or motivating one to reconsider

[13] I thank Sandy Goldberg for suggesting this example.

one's epistemic standing (Dokic 2012; Koriat 2007). Epistemic feelings of linguistic fluency could also fulfil some such roles. In a typical case, a feeling of fluency could signal that the hearer can proceed to utilize information about an utterance in communication, belief formation, and action. On the other hand, an epistemic feeling indicating a lower level of fluency may signal a need to allocate more resources, repair, etc. Epistemic feelings of linguistic fluency could guide our cognitive functioning in several important ways and in this way fulfil some epistemic roles.[14] Unsurprisingly, at least some answers to questions about the epistemic roles of such states (2) seem to depend on our views of their nature (1).

The final set of questions (3), concerns methodological issues about how the nature of experiences of understanding should be investigated. Which considerations should bear on the above questions and shape our views? Some people investigate the phenomenology of such experiences and build arguments based on contrast cases (e.g. Siegel 2010; Dodd 2014); others emphasize their epistemic roles in acquiring information and social interactions (e.g. Brogaard 2018; Balcerak Jackson 2019); still others advocate drawing on empirical evidence from psychological research on speech and utterance comprehension (e.g. O'Callaghan 2011; Gasparri and Murez 2020). Another complexity concerns whether and to what extent the philosophers' notion of conscious experience, as used in the debates on linguistic understanding, can be made compatible with currently available research on consciousness in psychology and neuroscience and, indeed, whether we currently have robust empirical evidence that could be informative for some of these issues. The intuitive answer to the question of what we experience when listening to a familiar language is only apparently a simple one, as the complex landscape of questions and views charted in this section illustrates. Issues concerning the phenomenology of linguistic understanding require more attention, given our common reliance on information acquired through linguistic communication (e.g. Goldberg 2018). But, as I will show in the next two sections, they do not exhaust what we should consider when we investigate the question of what we systematically experience when listening to a familiar language.

3. *Registering vocal prosody*

In many typical cases of linguistic communication, speakers produce linguistic utterances in a broader behavioural context. Spoken language is accompanied by a wide variety of non-verbal phenomena including vocal, facial, and bodily gestures (Wharton 2009). Usually, such gestures indicate the speaker's internal mental state—i.e., they convey information about their emotions, feelings, and attitudes toward the meanings expressed. Although we are often aware of using

[14] This passage draws on material from (Drożdżowicz forthcoming).

such gestures while speaking, and sometimes may even intentionally exploit them to achieve certain effects in our audience, in many cases, such gestures are produced spontaneously and beyond our conscious control. Non-linguistic gestures of various kinds can influence linguistic communication and may impact our understanding of utterances (Wharton 2009).

Linguistic communication is multimodal: it commonly exploits not only words, but also non-verbal *vocal* cues, as well as a whole variety of *visual* cues from the speaker's facial expression and bodily gestures. Bearing in mind that vocal and visual cues often interact, a phenomenon that recently has been a subject of intense study (e.g. Zhang et al. 2018; Frohlich et al. 2019), I will focus here on *prosody*—i.e., *vocal* elements of speech that are not individual phonetic segments (vowels and consonants) but properties of syllables and larger units of speech that commonly accompany the production of linguistic utterances (Speer and Blodgett 2006). This abstraction from visual input to linguistic communication is both necessary and warranted. The question to be investigated is what we systematically experience when listening to a familiar language. In addressing this, my focus will therefore be primarily on the auditory modality. In this section, I will provide evidence in support of the claim that prosodic phenomena are an important element that we routinely register when listening to a familiar language. Because of that they constitute a plausible candidate to consider for what can be systematically experienced in linguistic communication. Prosody can be and often is recognized without any visual input thanks to a specific psychological system that has been studied separately in experimental psychology of language. To illustrate, when your friend enthusiastically invites you to take another swim, in order to hear enthusiasm in her voice, you do not need to lift your sunhat to see her face or other bodily gestures. Whether she is smiling and vigorously imitating a crawl stroke, her enthusiasm, when expressed in the vocal prosody that accompanies her utterance, can be independently recognized.[15]

Prosody is an umbrella term encompassing a variety of vocal phenomena occurring in speech production. Specifically, it covers suprasegmental phonetic phenomena, i.e., properties that belong to larger units than phonemes, including syllables, phones, words, various intonation phrases and utterances (Speer and Blodgett 2006). There is general agreement that prosodic contributions to linguistic communication range from the intentionally produced, properly linguistic, and often language-specific ones (e.g., lexical tone, stress or pitch accent) to spontaneous, involuntary, or 'natural' ones (e.g. an angry, agitated or enthusiastic tone of voice) (Gussenhoven 2002; Pell 2002; Wharton

[15] The facial and bodily gestures, when perceived, may of course reinforce or modify your experience. An invitation produced in an angry tone of voice, but with a smile on the face would have a different effect than the one produced in a happy tone of voice and with a smile. Due to limited space, I leave discussion of such cases for another occasion.

2009: 139).[16] It is also commonly accepted that many contributions that prosodic gestures make to linguistic communication are context-dependent (Hirschberg 2002). Context may determine the degree of their contribution (hearing enthusiasm in your friend's voice may for example depend on your expectations about how much in general she likes to swim). It may also entirely determine the nature of a specific contribution (hearing enthusiasm in your friend's voice may strike you as fake and incongruent with the invitation, given that on such occasions she almost always speaks in a fairly neutral, flat tone of voice). Because of that, the prospects for a simple mapping from many prosodic phenomena to their communicative contributions are generally agreed to be dim (Hirschberg 2002, see also Wharton 2009: ch. 6).

The category of intentionally produced *properly linguistic prosodic gestures* is wide and includes, among other things, phenomena such as: contour variation, variation in location and type of pitch accents (e.g. nuclear stress on a single word), accent on discourse markers (e.g. but, although), accent on new information as opposed to what is given, accent on focus-sensitive operators (e.g. *only*, *some*, *must*), phrasing variation to chunk information in an utterance, variation in timing and pitch range to mark speaker involvement, final lowering (see Hirschberg 2002). The spontaneous, 'natural' prosodic contributions overlap to a large degree with what has been investigated under the label of *emotional prosody*.

'Emotional prosody' is a term used to describe phenomena in which speakers communicate emotions, either unintentionally or intentionally, by modifying acoustic attributes of their voice, and how these vocal cues are perceived and recognized by listeners (Pell and Kotz 2011). It has been argued that the neurocognitive system responsible for the processing of emotional prosody in hearers is distinct from the system responsible for the processing of speech sounds (Pell 2006), as well as from systems responsible for the processing of socially-relevant information recovered from the voice, such as age or gender (Belin et al. 2004; Spreckelmeyer et al. 2009). Some evidence suggests that vocal expressions of emotions are perceived categorically (e.g. Laukka 2005), thereby corresponding to a set of basic human emotions that also have discrete forms of expression in the face (Ekman 1994, but see Barett 2017; Celeghin et al. 2017). Furthermore, vocal expressions of anger, disgust, fear, sadness, and happiness/joy can be accurately recognized when listening to a foreign language (e.g. Pell et al. 2009; Sherer et al. 2001). This suggests that at least these emotions have discrete acoustic-perceptual properties in the voice which manifest in similar ways across languages. According to empirical studies, vocally expressed emotions in speech are registered implicitly and automatically based on specific vocal cues (Kotz and Paulmann 2007). Studies also suggest

[16] Several distinct and possibly overlapping distinctions are grouped here following recent discussions on the topic.

that there are important differences in the underlying time course for typical recognition of basic emotions from vocal expressions. Anger, sadness, fear, and neutral expressions are recognized more accurately shortly after hearing vocal cues than happiness and disgust. However, as speech unfolds with time, recognition of happiness improves significantly towards the end of the utterance, while fear is recognized more accurately than other emotions (Pell and Kotz 2011).

Described in this way, emotional prosody is a particularly important channel of information about the speaker's mental state and is often a subtle but permanent aspect of what we register when listening to linguistic utterances in a familiar language (and as the above studies suggest, when we listen to foreign languages too). When your friend invites you to take another swim, a particular shade of happiness that I have labelled as 'enthusiasm' reveals how she feels about the prospects of going for a swim with you and colours her invitation in a subtle but important way. You are sensitive to that colouring, and recognize the emotional expression in her voice.

Prosodic cues of various types tend to create impressions and convey information about speakers' emotions or attitudes, rather than expressing full propositions or concepts in their own right, as words and utterances usually do (Wharton 2009: 141). But there is no doubt that prosody can and often does aid linguistic communication (Hirschberg 2002; Fodor 2002; Wharton 2009). Prosodic contributions to linguistic communication are something that hearers regularly and systematically draw on in linguistic interactions. For example, the specific emotional prosody that accompanies your friend's invitation may influence what in the end you will take her to be communicating:

(swim 1) Would you like to take another swim? (*happy, enthusiastic tone of voice*)
(swim 2) Would you like to take another swim? (*neutral tone of voice*)

When uttered in a happy, enthusiastic tone of voice (swim 1), the invitation, given certain contextual expectations that you have about your friend and her interest in swimming, will likely be reinforced by the accompanying emotional prosody. When uttered in a neutral tone of voice (swim 2), the invitation to swim may, given some contextual assumptions, indicate that your friend is not, after all, excited about the prospects of your taking another swim. Perhaps she is offering it out of politeness, knowing that you love to swim but are afraid of doing it alone, etc. Registering the emotional prosody in your friend's voice will guide your overall interpretation of what happens in this linguistic interaction.

Another common example of how prosody impacts linguistic communication is that of intentionally employing a specific type of *stress pattern*:

(swim 3) Would you like to take another **swim**?
(swim 4) Would **you** like to take **another** swim?

When uttered with a neutral stress pattern, where the nuclear pitch accent falls at the end of the utterance, and specifically, on the last word (swim 3), given certain contextual expectations, you have grounds for taking your friend's invitation at face value. When uttered with a contrastive stress pattern, with an accent on 'you' and 'another' (swim 4), your friend's utterance may strike you as indicating something different from an invitation to take another swim. With this contrastive stress pattern, your friend may be indicating that she would be surprised if you accepted the invitation or that she doubts your stamina. Whichever interpretation might be most likely in this context, there is a clear sense in which prosodic variation in the stress pattern that accompanies an utterance contributes to what you will ultimately get from this linguistic interaction. It will affect how you interpret the invitation and probably also how you respond to it.[17]

In typical linguistic interactions, when we listen to a familiar language, we might have an experience of understanding what the speaker communicated with an utterance. But as the above evidence and examples show, we also routinely register the prosody that accompanies and partly constitutes linguistic utterances. We draw on information conveyed by vocal cues that are produced by the speaker both in a spontaneous and an intentional manner. Speakers have the resources to produce prosodic phenomena and do so regularly. Hearers register and draw on prosodic phenomena thanks to specific psychological mechanisms.

The above observations have led many linguists to ask whether prosody, given its contribution to linguistic interactions, may encode some relatively stable meanings, and if so what kind of meanings those could be. As already mentioned, any claims about prosodic meanings or prosodic code are bound to be limited by the overwhelming context-dependence of the contributions that prosodic information makes (Hirschberg 2002; Wharton 2009). According to Gussenhoven and colleagues (Gussenhoven 2002; Chen and Gussenhoven 2003), our understanding of various prosodic gestures is governed by both biological and properly linguistic codes. What they call the effort code is a biological code that connects the amount of energy that speakers utilize in speech production with specific prosodic cues to a range of interpretive effects. "An increase in effort may lead to increased articulatory precision, creating an impression of 'helpfulness', or 'obligingness'; or it may result in a wider pitch range, creating an impression of 'forcefulness' or 'certainty' or conveying affective meanings such as 'agitation' or 'surprise'" (in Wharton 2009: 143). A different approach to explaining the communicative contributions of prosodic gestures can be found in Wharton (2009). In his view, both spontaneous, uncontrolled and intentionally produced, properly linguistic prosodic gestures might encode *procedural* information, i.e. information where a word (or other linguistic

[17] For other interesting examples see Wilson and Wharton 2006.

device) encodes information specifically geared to guiding the hearer during the inferential phase of comprehension (145). In this sense, he argues, many prosodic gestures can be seen as encoding procedural meaning (Blakemore 2002; Escandell et al. 2011).

For the purposes of this paper, I will not take a stance in the debate about whether and how we could model the meaningful contributions that prosody makes to linguistic communication. It suffices to say that prosody is an important element of spoken utterances that influences many linguistic interactions. Prosody is thus a plausible candidate to consider when investigating what we systematically experience when listening to a familiar language. Whether and which aspects of prosodic information could actually surface to hearers' consciousness should be carefully investigated with both philosophical and empirical tools. In recent years prosody has received a lot of attention in theoretical linguistics and in experimental psycholinguistics. Much work remains to be done, and new research avenues emerged, such as the use of prosody in artificial text-to-speech and speech-understanding systems (Hirschberg 2002). Yet, curiously, prosody is rarely noticed in mainstream philosophical discussions of linguistic communication and understanding. This is surprising, given the abundance of evidence for the claim that prosody is commonly produced by speakers, and routinely registered by hearers. Thus, a full answer to the question of what we experience when listening to a familiar language requires an empirically-informed account of the role of prosodic cues.

One might ask whether, in light of the evidence above, a distinction between possible contributions of linguistic understanding and prosody to what we experience would be in fact warranted. Prosodic cues can affect utterance understanding. The contrastive stress pattern can influence even the truth-conditions of an utterance (as in "Sue only *spoke* to Laura, vs Sue only spoke *to Laura*"), thereby affecting what meaning or proposition we grasp upon hearing it.[18] Neither allowing for prosodic contributions to utterance interpretation, nor for interactions between communicated meanings and prosody, would I think undermine the idea that we might be able to experience prosody as something different from understanding an utterance. In many cases it makes sense to distinguish between understanding an utterance and hearing the accompanying prosody.

First, in many cases there is an intuitive sense in which we seem to register understanding an utterance and its vocal prosodic material separately. This is why we can capture our understanding by paraphrasing the sentence uttered, as well as capture the prosodic characteristics of the utterance by, for example, noting the emotions conveyed by a speaker's voice or the stress pattern used. Second, we have psycholinguistic evidence that the processing of verbal speech material and prosodic speech material is performed by two largely independent

[18] I thank Deirdre Wilson for helpful comments regarding this issue.

systems (e.g. Pell 2006; Belin et al. 2004). Third, verbal and prosodic contributions can come apart: a neutral tone of voice may not aid interpretation at all, and emotional prosody may be detected even when the phonological sounds and word meanings are not, as in the case of hearing emotions in an utterance produced in a foreign language or, as some studies suggest, in an artificial meaningless speech signal (Grandjean et al. 2005). Although linguistic understanding and prosody may and often do come together in linguistic communication, their possible contributions to what we experience when listening to a familiar language can be considered (at least to some degree) separately.

4. *Registering stable vocal characteristics of a speaker*

In this section, I argue that in typical cases of listening to utterances produced in a familiar language, we also typically detect stable vocal characteristics of a speaker and that those are another candidate to consider for what could systematically experience when listening to a familiar language. Usually, such vocal characteristics are not used for the purpose of and do not aid linguistic communication, but nevertheless reveal a lot of interesting and important information about the speaker. There are certain vocal characteristics of a speaker's voice that she or he cannot easily conceal when producing linguistic utterances. Hearers are sensitive to those characteristics and register them when listening to speech from a particular speaker. Many vocal parameters that are exhibited in vocal production coming from a particular voice systematically correspond to and indicate important properties of a speaker. Voice conveys not only rich information about a speaker's emotional state and attitudes (Pell and Kotz 2011; Bänzigier et al. 2014), as explained in section 3, but also provides extra-linguistic cues that reflect more stable speaker properties, including: identity (Baumann and Belin 2010), biological sex (Owren et al. 2007), age (Mulac and Giles 1996), and even the socioeconomic status and regional background of a speaker (e.g. Rakić 2019).

How is it possible that our voices can reveal so much about us? Vocal sounds are generated by the interaction of a source (the vocal folds in the larynx) and a filter, i.e. the vocal tract above the larynx (Ghazanfar and Rendall 2008). Voiced sounds correspond to a periodic oscillation of the vocal folds with a well-defined *fundamental frequency* (f0). Although for an individual speaker the range of f0 values can vary quite a lot during normal phonation or singing, the average f0 of a particular speaker is to a large extent a function of the size of the vocal folds (Latinus and Belin 2011). Male vocal folds tend to be longer and thicker than female vocal folds, causing them to vibrate at approximately half the frequency (100–120 Hz) than average female vocal folds (200–240 Hz) (Stevens 1998). This is why female and male voices tend to differ systematically in a way that is often recognized by hearers. Human voices tend to vary extensively. Small differences in the dimensions and histology of the

individual body parts that speakers use in phonation result in great individual variability among speakers in the individual acoustic patterns they can produce. Interindividual differences in the dimensions of the vocal folds and their tension during speech production cause variation in mean fundamental frequency (f0) and voice quality. Differences in other parts of the vocal tract, such as the nasal passage, result in differences in the absolute and relative positions of the resonance frequencies of the vocal tract (for details see Schweinberger et al. 2014).

Given such a multitude of factors determining how our voices sound, it is important to understand which of the perceptible vocal characteristics are utilized by hearers to identify speakers' voices and differentiate between them. Several studies suggest that perception of the fundamental frequency of a speaker's voice is a key parameter in recognizing the voice (e.g., Bauman and Belin 2010). Other studies reveal a more complex picture, where other parameters such as jitter (local variations in period length), shimmer (local variation in period amplitude) and harmonics-to-noise ratio are also utilized in voice perception (e.g., Kreiman and Sidtis 2011). It is generally agreed that the human perceptual ability to recognize voices is realized by a particular neurocognitive system. Neuroimaging studies have identified several brain areas in the temporal cortex, located in the middle parts of the superior temporal sulcus (STS) bilaterally, which show a particular sensitivity to voices, irrespective of whether they contain speech (Belin et al. 2004; Binder et al. 2000). Moreover, there is evidence that information about the stable vocal characteristics of a speaker is processed largely independently of the prosodic vocal information described in section 3. Studies investigating the perception of affective prosody show a greater activation of the right temporal lobe and right inferior prefrontal cortex (Mitchell et al. 2003) for prosodic information.[19] Other studies (e.g. Belin and Zatorre 2003) seem to confirm the role of anterior temporal lobe regions of the right hemisphere, particularly right anterior STS regions, in processing information about speakers' voices related to their identity.[20]

For humans, voices are among the most prevalent and salient sounds in the auditory environment. Our ability to analyze the information that is contained in voices is important for many social interactions. Take our swimming example again. Even without seeing your friend approaching you at the seaside, when you hear her enthusiastically uttering the sentence with which she invites you to take another swim, you would normally immediately recognize the voice you hear as the one that belongs to your friend. When you hear her inviting you

[19] The perception of identity information in the voice has been examined in several neuroimaging studies suggesting that the anterior temporal lobes in both hemispheres are more active during speaker identification than during emotion identification (Imaizumi et al., 1997).

[20] Passages on pages 381 and 382 up till this point draw on material from section 2 of Drożdżowicz 2020.

to take another swim, you register information about the specific vocal parameters of her voice, such as its fundamental frequency, tempo, and the resonance frequencies that are determined by the anatomy of your friend's vocal folds and vocal tract. It is by registering these parameters that you can hear the voice that utters the invitation as belonging to a middle-aged female. Moreover, given your familiarity with your friend's voice, upon hearing those vocal parameters, you immediately recognize the voice as *hers*. Hearing the same invitation to take a swim uttered in the same context but in a voice that does not belong to your friend will result in a markedly different auditory experience of the stable vocal characteristics. Needless to say, an invitation from a stranger will have a very different communicative effect. Our sensitivity to human voices is an amazing perceptual advantage that allows us to effortlessly and typically accurately track a source of spoken linguistic utterances and in this way facilitates interactions based on linguistic communication.

As already mentioned, there are other properties of speakers that are systematically indicated by relatively stable vocal characteristics of a speaker's voice. Among them, regional dialect and foreign accent are properties of speakers that we are sensitive to. At this point you may no longer pay attention to her accent, but the first time you heard your friend speaking, it may have struck you that she speaks with a slight Danish accent. Having watched multiple Danish crime series, you were actually able to correctly identify her accent as Danish, though many of your friends initially struggled with that. In many cases, we recognize stable vocal characteristics of a speaker that point to an identifiable regional dialect or foreign accent. As we detect them, we can often become aware of the speaker's place of origin, ethnic background, and sometimes even their socioeconomic status. It is not only what we say, but also how we sound, that has a power to generate impressions beyond what we intend to convey and often beyond what we would like to reveal to interlocutors.

Hearers' impressions of speaker properties based on their perception of stable vocal characteristics need not always facilitate linguistic communication or the social interactions that draw on it. Take foreign accent as an example. Foreign accent can influence social interactions based on linguistic communication. Leaving aside rare exceptions in which foreign accents are perceived positively (Gibson et al 2017), there is evidence suggesting that in various sociolinguistic contexts speakers with foreign accents are judged to be less intelligent, less trustworthy, less educated and less competent than native speakers (e.g. Dewaele and McCloskey 2015; Dragojevic et al. 2016; Fraser and Kelly 2012; Fuertes et al. 2012; Gluszek and Dovidio 2010; Livingston et al. 2017). Negative bias towards foreign-accepted speech is present from early childhood (Kinzler et al. 2007). At the age of 11, children tend to trust native-accented speakers more (Kinzler et al. 2011).

The foreign accent bias may have (at least) two origins. One is linguistic: foreign accent may decrease 'processing fluency' and lead to

lower intelligibility of the speaker (Deterding and Kirkpatrick 2006; Cristia et al. 2012).[21] The other is social: foreign-accented speakers are rapidly categorised as out-group members. In this way, foreign accent may lead to negative evaluation of speaker's competence by being a function of shared negative attitudes towards the ethnicity of the accented speaker (Lippi-Green 1997; Roessel et al. 2019). The foreign accent bias has been shown to lead to discrimination in various contexts, for example in the courtroom (Solan and Tiersma 2004) and in job interviews (Huang 2013; Hansen and Dovidio 2016).

Although, your friend's Danish accent may not be an issue when she is inviting you to take a swim, information about the speaker's stable vocal characteristics can systematically influence social interactions. Registering stable vocal characteristics of a speaker and the resulting impressions may facilitate linguistic communication and social interactions, but it may also systematically impede them. Such vocal characteristics are another candidate to consider when investigating what we systematically experience in linguistic communication. Our sensitivity to stable vocal characteristics of a speaker raises interesting ethical questions about linguistic interactions that are affected by our impressions of the speaker. It also invites us to consider whether and which properties of vocal production could be perceived by hearers and which might result from inference and the underlying implicit beliefs that hearers have about speakers. This is a complex question that requires detailed treatment. Where and how exactly the border between auditory vocal perception and audition-based cognition of speaker properties is to be drawn is a difficult matter that is likely to generate an intense discussion and requires both philosophical and empirical investigation (e.g. Di Bona 2017).

5. *Concluding remarks*

I have argued that at least three elements need to be considered when we ask what we systematically experience when listening to a familiar language. (i) We perceive speech sounds and typically seem to have an experience of *understanding* what the speaker communicated with an utterance on a particular occasion. (ii) We register various forms of *prosody* and thanks to that we can learn about speaker's attitudes and mental states indicated by prosodic cues. (iii) We register the speaker's *stable vocal characteristics* and have systematic impressions about the speaker's identity. Those *three elements* should be investigated when we consider the question of what could be systematically present in our experience when we listen to a familiar language, given that they:

[21] This might involve a pragmatic component: in some contexts foreign accents might increase processing effort for native speakers when first encountered, and might therefore affect assessments of relevance, and ultimately competence etc. This effect could in some cases diminish as hearers become more familiar with the accent and it becomes easier to process. I thank Deirdre Wilson for these points.

(a) result from common forms of linguistic expression and information transfer in humans, and (b) are registered by hearers thanks to the specific psychological mechanisms that are employed in spoken linguistic communication and voice perception.

As already mentioned, the three elements need not *always* be present when we listen to linguistic utterances. An utterance may not be understood, a speaker may produce an utterance in a flat tone, or they may have a voice that makes it particularly difficult to identify any speaker properties. Typically, however, the three elements are routinely produced and registered in spoken linguistic communication. How strong is the claim about the *systematic* presence of these three elements? This type of systematicity is contingent on our biological and cultural evolution. It is not entirely impossible that there could exist forms of human linguistic communication that do not make use of prosody,[22] and we can even imagine speakers who lack typical vocal characteristics. Human speakers however, at least for now, communicate using their vocal apparatus that has evolved in a particular manner (Belin et al. 2004). Moreover, they tend to take advantage of prosodic forms of expression when producing linguistic utterances. Human hearers are sensitive to such stable vocal characteristics and prosodic phenomena.

The paper provides some preliminary work for addressing the question of what we systematically experience when listening to a familiar language. I believe that a more inclusive approach to address the question is required if we are to make progress on epistemic and moral questions concerning testimony and other forms of social interactions that draw on linguistic communication in different contexts.

References

Bayne, T. 2009. "Perception and the reach of phenomenal content." *Philosophical Quarterly* 59 (236).

Bänziger, T. et al. 2014. "The role of perceived voice and in emotion communication." *Journal of nonverbal behavior* 38 (1): 31–52.

Barrett, L. F. 2017. "The theory of constructed emotion: an active inference account of interoception and categorization." *Social cognitive and affective neuroscience* 12 (1): 1–23.

Baumann, O. and Belin, P. 2010. "Perceptual scaling of voice identity." *Psychological Research PRPF* 74 (1): 110.

Belin, P., and Zatorre, R. J. 2003. "Adaptation to the speaker's voice in the right anterior temporal lobe." *Neuroreport* 14 (16): 2105–2109.

Belin, P., Fecteau, S., and Bedard, C. 2004. "Thinking the voice: neural correlates of voice perception." *Trends in cognitive sciences* 8 (3): 129–135.

[22] This is of course the case for many artificial speakers, such as recorded and computer generated announcements where the prosody is flat or wrong and does not serve any communicative function (Hirschberg 2002). Moreover, in the literature on autism it is often claimed that emotional prosody is commonly impaired, both in production and comprehension, of speakers with ASD (Chevallier et al. 2011).

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

Blakemore, D. 2002. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99). Cambridge: Cambridge University Press.

Brogaard, B. 2018. "In defense of hearing meanings." *Synthese* 195 (7): 2967–2983.

Chen, A. and Gussenhoven, C. (2003). "Language-dependence in the signalling of attitude in speech." In *CHI 2003 conference on Human and Computer Interaction*.

Chevallier, C., Noveck, I., Happé, F. and Wilson, D. 2011. "What's in a voice? Prosody as a test case for the Theory of Mind account of autism." *Neuropsychologia* 49 (3): 507–517.

Celeghin, A., Diano, M., Bagnis, A., Viola, M. and Tamietto, M. 2017. "Basic emotions in human neuroscience: neuroimaging and beyond." *Frontiers in Psychology* 8: 1432.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A. and Floccia, C. 2012. "Linguistic processing of accented speech across the lifespan." *Frontiers in psychology* 3: 479.

Deterding, D. and Kirkpatrick, A. 2006. "Emerging south–east Asian Englishes and intelligibility." *World Englishes* 25 (3–4): 391–409.

Di Bona, E. 2017. "Towards a rich view of auditory experience." *Philosophical Studies* 174 (11): 2629–2643.

Dodd, J. 2014. "Realism and anti-realism about experiences of understanding." *Philosophical Studies* 168 (3).

Dokic, J. 2012. "Seeds of self-knowledge: noetic feelings and metacognition." In *Foundations of metacognition*, Chapter 6: 302–321.

Dormandy, K. 2018. "Epistemic Authority: Preemption or Proper Basing?" *Erkenntnis* 83 (4): 773–791.

Dragojevic, M., Giles, H., Beck, A. C. and Tatum, N. T. 2017. "The fluency principle: Why foreign accent strength negatively biases language attitudes." *Communication monographs* 84 (3): 385–405.

Drożdżowicz, A. 2019. "Do we hear meanings? – between perception and cognition." *Inquiry* 1–33.

Drożdżowicz, A. 2020. "Bringing back the voice – on the auditory objects of speech perception." *Synthese* 1–27.

Drożdżowicz, A. Forthcoming. "Experiences of understanding as epistemic feelings." *Mind & Language*.

Ekman, P. 1994. "All emotions are basic." *The nature of emotion: Fundamental questions* 15–19.

Escandell-Vidal, V., Leonetti, M. and Ahern, A. (eds.). 2011. *Procedural meaning: Problems and perspectives* (Vol. 25). Leiden: Brill.

Fodor, J. A. 1983. *The modularity of mind: An essay on faculty psychology*. Cambridge: MIT Press.

Fodor, J. D. 2002. "Psycholinguistics cannot escape prosody." In *Speech Prosody 2002, International Conference*.

Fraser, C. and Kelly, B. F. 2012. "Listening between the lines: Social assumptions around foreign accents." *Australian Review of Applied Linguistics* 35 (1): 74–93.

Fricker, E. 2003. "Understanding and knowledge of what is said." In A. Barber. *Epistemology of Language*. Oxford: Oxford University Press: 325–66

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T. and van Schaik, C. P. 2019. "Multimodal communication and language origins: integrating gestures and vocalizations." *Biological Reviews* 94 (5): 1809–1829.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C. and Giles, H. 2012. "A meta-analysis of the effects of speakers' accents on interpersonal evaluations." *European Journal of Social Psychology* 42 (1): 120–133.

Gasparri, L. and Murez, M. 2019. "Hearing meanings: the revenge of context." *Synthese* 1–24.

Ghazanfar, A. A. and Rendall, D. 2008. "Evolution of human vocal production." *Current Biology* 18 (11): R457–R460.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B. and Fedorenko, E. (2017). "Don't underestimate the benefits of being misunderstood." *Psychological science* 28 (6): 703–712.

Gluszek, A. and Dovidio, J. F. 2010. "The way they speak: A social psychological perspective on the stigma of nonnative accents in communication." *Personality and social psychology review* 14 (2): 214–237.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R. and Vuilleumier, P. 2005. "The voices of wrath: brain responses to angry prosody in meaningless speech." *Nature neuroscience* 8 (2): 145–146.

Gussenhoven, C. 2002. "Phonology of intonation." *Glot International* 6 (9/10): 271–284.

Hansen, K. and Dovidio, J. F. 2016. "Social dominance orientation, nonnative accents, and hiring recommendations." *Cultural Diversity and Ethnic Minority Psychology* 22 (4): 544.

Hirschberg, J. 2002. "Communication and prosody: Functional aspects of prosody." *Speech Communication* 36 (1–2): 31–43.

Huang, B. H. 2013. "The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech." *System* 41 (3): 770–785.

Hunter, D. 1998. "Understanding and belief." *Philosophy and Phenomenological Research* 58 (3).

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H. ... and Kojima, S. 1997. "Vocal identification of speaker and emotion activates different brain regions." *Neuroreport* 8 (12): 2809–2812.

Jackson, B. B. 2019. "Against the perceptual model of utterance comprehension." *Philosophical Studies* 176 (2): 387–405.

Kinzler, K. D., Corriveau, K. H. and Harris, P. L. 2011. "Children's selective trust in native–accented speakers." *Developmental science* 14 (1): 106–111.

Koriat, A. 2007. "Metacognition and consciousness." In P. D. Zelazo, M. Moscovitch and E. Thompson (eds.). *The Cambridge Handbook of Consciousness*. Cambridge: Cambridge University Press: 289–326.

Kotz, S. A. and Paulmann, S. 2007. "When emotional prosody and semantics dance cheek to cheek: ERP evidence." *Brain research* 1151: 107–118.

Kreiman, J. and Sidtis, D. 2011. *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Hoboken: John Wiley and Sons.

Latinus, M. and Belin, P. 2011. "Human voice perception." *Current Biology* 21 (4): R143–R145.

Laukka, P. 2005. "Categorical perception of vocal emotion expressions." *Emotion* 5 (3): 277.

Lev-Ari, S. 2015. "Comprehending non-native speakers: theory and evidence for adjustment in manner of processing." *Frontiers in psychology* 5: 1546.

Lev-Ari, S. and Keysar, B. 2010. "Why don't we believe non-native speakers? The influence of accent on credibility." *Journal of experimental social psychology* 46 (6): 1093–1096.

Lippi, R., Donati, S., Lippi-Green, R. and Donati, R. 1997. *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.

Livingston, B. A., Schilpzand, P. and Erez, A. 2017. "Not what you expected to hear: Accented messages and their effect on choice." *Journal of Management* 43 (3): 804–833.

Longworth, G. 2018. "Understanding what was said." *Synthese* 195 (2): 815–834.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A. and Woodruff, P. W. 2003. "The neural response to emotional prosody, as revealed by functional magnetic resonance imaging." *Neuropsychologia* 41 (10): 1410–1421.

Montague, M. 2017. "Perception and cognitive phenomenology." *Philosophical Studies* 174 (8): 2045–2062.

Mulac, A. and Giles, H. 1996. "'You're Only As Old As You Sound'". *Health Communication* 8 (3): 199–215.

Nes, A. 2016. "On what we experience when we speak." *Phenomenology and Mind* 10: 58–85.

Nozari, N. and Novick, J. 2017. "Monitoring and control in language production." *Current Directions in Psychological Science* 26 (5): 403–410.

O'Callaghan, C. 2011. "Against hearing meanings." *The Philosophical Quarterly* 61 (245): 783–807.

O'Callaghan, C. 2015. "Speech Perception." In *Handbook of the Philosophy of Perception*. Oxford: Oxford University Press.

Owren, M. et al. 2007. "Listeners judge talker sex more efficiently." *Perception and psychophysics* 69 (6): 930–941.

Pell, M. D. 2002. "Surveying emotional prosody in the brain." In *Speech Prosody 2002, International Conference*: 77–82.

Pell, M. D. 2006. "Cerebral mechanisms for understanding emotional prosody in speech." *Brain and language* 96 (2): 221–234.

Pell, M. D. and Kotz, S. A. 2011. "On the time course of vocal emotion recognition." *PLoS One* 6 (11): e27256.

Prinz, J. 2011. "The sensory basis of cognitive phenomenology." In *Cognitive Phenomenology*. Oxford: Oxford University Press: 174–196.

Rakić, T., Steffens, M. C. and Mummendey, A. 2011. "Blinded by the accent! The minor role of looks in ethnic categorization." *Journal of personality and social psychology* 100 (1): 16.

Rakić, T. 2019. "How accent and gender influence perceptions of competence." *Journal of Language and Discrimination* 3: 218–231.

Reiland, I. 2015. "On experiencing meanings." *The Southern Journal of Philosophy* 53 (4): 481–492.

Roessel, J., Schoel, C., Zimmermann, R. and Stahlberg, D. 2019. "Shedding new light on the evaluation of accented speakers: Basic mechanisms behind nonnative listeners' evaluations of nonnative accented job candidates." *Journal of Language and Social Psychology* 38 (1): 3–32.

Scharenborg, O. 2007. "Reaching over the gap: A review of efforts to link human and automatic speech recognition research." *Speech Communication* 49 (5): 336–347.

Scherer, K. R., Banse, R. and Wallbott, H. G. 2001. "Emotion inferences from vocal expression correlate across languages and cultures." *Journal of Cross-cultural psychology* 32 (1): 76–92.

Schiffer, S. 1987. "Intentionality and the Language of Thought." In *Proceedings of the Aristotelian Society* (Vol. 87, No. 1). Oxford: Oxford University Press: 35–56.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G. and Zäske, R. 2014. "Speaker perception." *Wiley Interdisciplinary Reviews: Cognitive Science* 5 (1): 15–25.

Siegel, S. 2006. "Which properties are represented in perception." *Perceptual experience* 1: 481–503.

Siegel, S. 2011. *The contents of visual experience*. Oxford: Oxford University Press.

Smith, B. C. 2009. "Speech sounds and the direct meeting of minds." In Nudds and O'Callaghan. *Sounds and perception*. Oxford: Oxford University Press: 183–210.

Smith, B. C. 2010. "Meaning, Context, and How Language can Surprise Us." In *Explicit Communication*. London: Palgrave Macmillan: 92–108.

Speer, S. and Blodgett, A. 2006. "Prosody." In *Handbook of psycholinguistics*. Academic Press: 505–537.

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmüller, E. and Münte, T. F. 2009. "Neural processing of vocal emotion and identity." *Brain and cognition* 69 (1): 121–126.

Wharton, T. 2009. *Pragmatics and non-verbal communication*. Cambridge: Cambridge University Press.

Wilson, D. and Wharton, T. 2006. "Relevance and prosody." *Journal of pragmatics* 38 (10): 1559–1579.

Zhang, H., Chen, X., Chen, S., Li, Y., Chen, C., Long, Q. and Yuan, J. 2018. "Facial expression enhances emotion perception compared to vocal prosody: Behavioral and fMRI studies." *Neuroscience bulletin* 34 (5): 801–815.