

This is a repository copy of *Sounds on the margins of language, at the heart of interaction*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/155325/>

Version: Published Version

Article:

Keevallik, Leelo and Ogden, Richard orcid.org/0000-0002-5315-720X (2020) Sounds on the margins of language, at the heart of interaction. Research on Language and Social Interaction. pp. 1-19. ISSN: 0835-1813

<https://doi.org/10.1080/08351813.2020.1712961>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Sounds on the Margins of Language at the Heart of Interaction

Leelo Keevallik & Richard Ogden

To cite this article: Leelo Keevallik & Richard Ogden (2020) Sounds on the Margins of Language at the Heart of Interaction, Research on Language and Social Interaction, 53:1, 1-18, DOI: [10.1080/08351813.2020.1712961](https://doi.org/10.1080/08351813.2020.1712961)

To link to this article: <https://doi.org/10.1080/08351813.2020.1712961>



© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 05 Mar 2020.



[Submit your article to this journal](#)



Article views: 121



[View related articles](#)



[View Crossmark data](#)

Sounds on the Margins of Language at the Heart of Interaction

Leelo Keevallik^a and Richard Ogden^b

^aDepartment of Culture and Society, Linköping University, Sweden; ^bDepartment of Language and Linguistic Science, Centre for Advanced Studies in Language and Communication, University of York, United Kingdom

ABSTRACT

What do people do with sniffs, lip-smacks, grunts, moans, sighs, whistles, and clicks, where these are not part of their language's phonetic inventory? They use them, we shall show, as irreplaceable elements in performing all kinds of actions—from managing the structural flow of interaction to indexing states of mind and much more besides. In this introductory essay we outline the phonetic and embodied interactional underpinnings of language and argue that greater attention should be paid to its nonlexical elements. Data are in English and Estonian.

Sounds in interaction

In spoken and embodied interaction, participants may produce sounds that are not words in the traditional sense but that nonetheless are treated by participants as meaningful. Examples from English in regular orthography include *ugh*, *mmm*, *argh!*, *pfft!*, *brrr* (used in English to mark “cold”), and *phew*. There are many others that cannot be written orthographically, such as clicks, sniffs, or sighs, or items that are made up ad hoc—singular items (Enfield, 2009) whose meaning has to be deduced from the occasion of use. Although items like these are embedded within a linguistic context, linguists do not normally think of them as linguistic objects: They do not belong to a major syntactic category, they may not conform to the phonological requirements of the language, and the relationship between their form and meaning might not be arbitrary, which to a large extent has been considered the sine qua non of linguistic forms (see, however, Dingemanse, 2018). These sounds are therefore at the boundary of language and paralanguage, or nonlanguage.

An early sociological account for the use of sounds beyond the lexicon in English was authored by Goffman (1981), who called a group of them *response cries* and provided impressionistic accounts for their use. Interaction researchers (most notably Gardner, 2002; Goodwin, Goodwin, & Olsher, 2002; Wiggins, 2013) have over the years treated a variety of single examples of sounds with great precision in their embodied context. These results are scattered, not always precise in their phonetic/prosodic detail, and mostly deal with English. Building on the pioneering work on the phonetics of talk-in-interaction (e.g., Couper-Kuhlen & Selting, 1996; Local & Kelly, 1986) as well as the well-documented embodied turn in interaction research (Neville, 2015), authors of this special issue take a closer look at a variety of human sounds in natural interaction, aiming to reveal their social organization from a participant perspective.

CONTACT Leelo Keevallik  leelo.keevallik@liu.se  Department of Culture and Society, Linköping University, Linköping 581 83, Sweden.

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the spirit of the foundations of the method used here, conversation analysis (Sacks, 1992), we should not disregard any features of behavior in advance, which is why the articles focus on sounds that at first sight often appear totally lacking lexical meaning. When studying human action from the perspective of the interacting participants, however, we may find that some sounds that are treated as relevant and communicative would not be considered as words, i.e., more or less arbitrary form-function packages, by linguists. Choosing to be agnostic about the linguistic status of the sounds produced in the human vocal tract, we hereby launch a scrutiny of their potential role in intersubjective sense-making. This analysis necessitates a serious consideration of not only the articulatory details and precise temporalities of the sound production but also careful study of the participants' treatment of these sounds within the ongoing embodied ecologies of action.

The special issue thus contributes to knowledge of how vocal tract sounds function in interaction on the one hand and to the relevance of phonetic features for social action on the other. The analytical focus of the original articles is on close qualitative and emic analysis of particular sounds, and only in the discussion will it be placed within the framework of human vocal practices in general, including language as a system of signs (Dingemanse 2020/*this issue*). In consolidating the work being carried out in the field of vocalizations, we ended up with a collection of sounds that are perhaps most remote from a commonsensical idea of language, such as sighs, moans, and sniffs, or clicks in languages where clicks are not part of the lexical phonological inventory. The languages include English, Finnish, French, and Mandarin.

An illustrative example: Clicks in a depiction

We begin with an English example of the kind of phenomenon we are focusing on. In Excerpt 1, Rachel and Laura are sitting in the sun and worrying about the possibility of getting sunburnt on their arms and shoulders. They are having trouble formulating what the resulting uneven tan will look like.

Excerpt 1: RCE Lake 17.26-17.48 hair line

01 L: Yeah, that was grim.
 02 CLICK I hOpe I get a tanned bAck,
 03 but I don't want a HAIRline?
 04 R: I hope I get a TAN:;
 05 but like I fEEl like the sUn's,
 06 like HOTter (on/than) thIs one,
 07 L: I fEEl like I'm going to get a HAIR mark
 08 and [↑thEse are really] -
 09 R: [(little) curly BIT]s just marked into your BACK.
 10 L: → CLICK CLICK
 11 R: hehe
 12 L: bOOm-
 13 r sniff+look away
 14 L: what was I saying
 15 yeah my brother

Laura and Rachel are both concerned that they might end up with an uneven tan: Laura because of her hair, Rachel because the sun is more on one side than the other. In this extract, they collaborate to find a description of the effects of the sun. Laura and Rachel recycle one another's words: "hope ... tan" (lines 2 and 4, "I feel like ... " (lines 5 and 7), and "mark" (lines 7 and 9). There is also talk where they try to formulate verbally the pattern Laura is going to have on her back: "hair line," line 3 (which normally refers to the scalp) and "hair mark," line 7 (which sounds like a mark made by the hair, not the sun); "curly bits," line 9, which describes the shape of the pattern; and the pronoun "these" (line 8), which is accompanied by looking and touching the bits that will be affected.

At line 10, Laura produces two clicks, accompanied by a tracing outline with the index finger of her right hand, which starts and ends with her gaze directed at Rachel (Figure 1, panels A and D), but in between her gaze is on the tracing outline (panels B and C). The clicks coincide with the two peaks in the tracing outline (panels B and C), and the second of the clicks (panel C) has an eyebrow flash. The clicks and tracing gesture depict visibly and audibly what Laura and Rachel have so far failed to capture in words. That they are followed by a sniff with gaze away and a change in topic (line 14) (Hoey, 2020/*this issue*) underscores the success of the nonverbal construction in demonstrating something that was not easily explained with words.



Figure 1. Extract 1, line 10. (a) Laura initiates the tracing gesture, gaze to Rachel; (b) the first click, and the first peak in the tracing outline, gaze to the outline; (c) the second click, and the second peak in the outline; Panel (d) the gesture is retracted, Laura's gaze goes back to Rachel.

Line 10 is an example of a vocalization: Clicks are sound objects (Reber, 2012) made in the vocal tract, and the clicks here contribute to the speaker's turn but do not form a recognizable conventional lexical item. This precise combination of sound and gesture is a product of an interactional contingency, but it poses no obvious problems for the participants' understanding—indeed, it solves a problem and results in a shared

understanding. This is the kind of practice that is frequent in face-to-face interaction but that linguists have commonly passed over. However, recent literature on ideophones (Akita & Dingemanse, 2019; Dingemanse, 2017) and depictions (Clark, 2016) addresses very similar issues: What do linguistically marginal contributions like these tell us about language and its connection to the physical world we live in and the bodies we inhabit?

In subsequent sections, we untangle some of the phonetic issues around several vocal tract sounds, their physical and somatic grounding, as well as adaptability to local interactional trajectories. We will thus target sense-making with inherently nonconventional (or—as we will see—semiconventional) practices.

Phonetic aspects

What do we mean by “vocalization”?

The particular sounds we are interested in in this special issue are made in the vocal tract, i.e., the part of the body used for making speech. The vocal tract consists of the larynx, pharynx (throat), the oral tract (mouth), and nasal tract (nasal passages). Sounds that are made in the vocal tract are known as vocalizations; vowels and consonants are basic vocalizations of speech.

Phonetics explores the potential of the vocal tract for making linguistic vocalizations and provides us with technologies for describing, annotating, and analyzing these sounds. These include the International Phonetic Alphabet, which encapsulates an analysis of the vocal tract and the sounds it can make. Some observable vocalizations are not known to exist in the words of any language but are nonetheless describable in phonetic terms: snores (ingressive, oro-nasal pulmonic velar trills) are one such sound. Others, like labial whistles (Reber & Couper-Kuhlen, 2020/*this issue*), sighs (Hoey, 2014), or laughter (Glenn & Holt, 2013), have no conventional phonetic description. Some vocalizations, like clicks, are considered linguistic in some languages—limited to Southern Africa, where they occur in words—but not in others, including English and Mandarin (Li, 2020/*this issue*; Ogdén, 2013, 2020/*this issue*), where they are normally considered to be paralinguistic (Gil, 2013). In English [m] has linguistic status and also paralinguistic status (Gardner, 2002; Wiggins, 2019).

The notion of “word” is central to determining linguistic distinctiveness. Linguists distinguish also between verbal and nonverbal (Laver, 1994) material. Not all linguistic contrasts are verbal, though they co-occur with words: For example, intonation contours are nonverbal, but they must co-occur with words. Their meanings are often difficult to specify, but they have been shown to be intimately connected to sequence organization and action formation (Couper-Kuhlen & Ford, 2004; Couper-Kuhlen & Selting, 1996).

The verbal/nonverbal distinction is a continuum (Wharton, 2003). For example, most English speakers can mimic the sound a duck makes. This vocalization might include nonnative sounds, like a pharyngealized open vowel (i.e., an [a]-type vowel with constriction in the throat). It can be stylized too, with the onomatopoeic form *quack* or *quack quack* (as in: “What does the duck say?”); and in turn, this stylized form can be a verb or noun in English and inflected correspondingly (“the duck quacked loudly”). Examples like this (and others like [pʁ:] vs. “to purr” or [!] vs. ‘tsk!’ vs. ‘to tut’) are evidence for a continuum between mere vocalizations and lexical items. “Words” are integrated within

the morphosyntax of a language. Vocalizations might make use of sounds that are not ordinarily found in a language, or they might not conform to the phonotactic conventions of the language, such as [pr̥:] to represent a cat purring, which contains a nonnative voiceless trill and has no vowel. In between there is a range of vocalizations, including response tokens, that are at least partly conventionalized in some dimension: They might draw on the sound inventory of the language or make use of some aspect of its meaning-making system such as certain prosodic features.

Linguistics has concentrated on the “word” end of this continuum; but more recent work on iconicity, multimodality, and—within conversation analysis—on pragmatic particles (Heritage, 1984; Sorjonen, 2001), has led many linguists to question this distinction. Articles in this special issue show examples of phenomena along this continuum, from the more somatic sniffs analyzed by Hoey to the more word-like Finnish interjection *huh-huh* (Pehkonen, 2020/this issue).

Affordances of vocalizations

Vocalizations offer different affordances from lexical items and are extremely variable in their form. This is important because it allows vocalizations to be adjusted to their local environment while still being recognizable, and it means they can be interpreted flexibly in their precise sequential and action environment. Understanding the phonetic affordances of vocalizations can serve as a way in to understanding how elements of language function more generally.

Some vocalizations are a blend of resources of different forms and types: They can be modified in ways that are conventional for the speaker’s language. Intonation, pitch register, pitch span, and voice quality are all aspects of production that are laminated on (typically) vocalic articulations and that generally carry meanings that co-occur with linguistic material. Heinemann and Koivisto (2016, p. 83) point out that *oh* in English has been shown to have “a general semantics” that can be laminated with prosodic features, or changes to vowel quality, to mark, e.g., surprise or disappointment or something about the speed or intensity of a cognitive change. Thus, the situated meaning of *oh*, and other response tokens, is compositional.

At the more word-like end, in Pehkonen’s article (2020/this issue), the Finnish *huh huh* is shown to be a stylized version of “being out of breath.” He notes that these tokens are often produced with a stylized intonation contour; perhaps this is a way of displaying “being out of breath” as something to talk about, rather than a problem to be remedied. Whatever the function of this stylization, the vocalization is a blend of consonantal, vocalic, and intonational forms that fit Finnish phonotactics (i.e., allowable combinations of sounds in the language) and are available to be interpreted compositionally.

Vocalizations that do not contain vocalic or otherwise sonorous elements can be adjusted in other ways: Repetition, loudness, laxness (such as whistles, Reber & Couper-Kuhlen, 2020/this issue), and (a-)rhythmical positioning relative to adjacent speech events are all resources that can be drawn on. Moans (Hofstetter, 2020/this issue) and clicks (Ogden, 2020/this issue) in English can be temporally integrated with surrounding talk, which expresses something about the relation of the vocalization to an adjacent utterance. This provides an orientation to the relevant vocal tract sound as something that “belongs with” the talk it is embedded in, and the participants’ current activities, and displays an

understanding of the sound as such—rather than, e.g., an extraneous noise. Rhythmicity (and arrhythmicity) is also a way to exhibit affiliative or disaffiliative relations between turns—and this is true for speech as it is for nonspeech sounds. The vocalizations thus reflect a variety of affordances for formatting action in specific local environments.

Phonetic underspecification of vocalizations

A collection of vocalizations often includes a lot of phonetic variability. Nonetheless, commonalities can be stated in general terms. We call these vocalizations “phonetically underspecified” (Keating, 1988): For the vocalization to be recognizable as a token of a particular type of vocalization, speakers have to produce some phonetic events, with freedom to vary some elements. A good example of this is moans (Hofstetter, 2020/*this issue*). These are vocalic articulations, produced with an open vocal tract, and in the open half of the vowel space, i.e., open vowels like [a ʌ ɔ] but not close vowels like [i e u]. In talking about individual tokens of moans, it is possible to describe their vowel quality with some phonetic precision, but whether this precision matters from the perspective of identifying the action of the moan is debatable: It is not trivial to show that a moan with the front open vowel [a] is substantially different from one with the back open vowel [ʌ]. Compare this with word pairs like *hat* [hat] versus *heart* [ha:t], where this front-back contrast is lexically contrastive. Probably the important element in moans is that a “moan” is constituted by an open vowel, regardless of frontness, backness, or rounding—and we happen to have no general phonetic symbol to cover vowels in such a big vowel space.

Another example of a phonetically underspecified vocalization is the kinds of sounds often made when lifting and straining, one of which we will see in Excerpt 2. These vocalizations often start with an abrupt release (such as a glottal stop), followed by a voiced, open vocalic articulation, with some friction that ranges from uvular through pharyngeal or epiglottal or laryngeal or a combination of these—i.e., the lower part of the oral tract and the pharynx—and that fade away more slowly than they start. This constitutes a phonetically underspecified depiction of current strain in the speaker’s body, and it also serves as a kind of template by which the strain can be recognized.

A more provocative example of underspecification is provided by whistles (Reber & Couper-Kuhlen, 2020/*this issue*), where the source of the sound is not at the vocal folds but at the lips. The pitch of a labial whistle is changed by moving the tongue, so it is possible to produce contours that are close to those of speech, such as a rise-fall contour or a stylized “calling contour” (Ladd, 1978; Ogden, Hakulinen, & Tainio, 2004). Whistles make it possible for a speaker to produce an intonation contour without any accompanying speech sound. The “tone-bearing unit” that is normally required by intonation contours—in English, a vowel—is replaced by something that is not part of speech. One affordance of a whistle is precisely that it contains no vocalic articulation, so it makes it possible to produce an intonation contour without having to select a syllabic such as *oh*, *ah*, *aw*, etc., which may have its own meaning. A speaker can realize an intonation contour without choosing any particular response particle. Thus, the underspecification of vocalizations is an essential asset for participants in social interaction.

The multimodal nature of vocalizations

As we have seen, phonetics—which forms the basis of phonology and in turn the basis of linguistic units like words, phrases, and sentences—is by definition about the vocal channel. Traditionally, it excludes gesture, bodily actions, and facial expression. This has not been done to deny their importance; it is more a claim that they are independent of phonetics. However, speaking is necessarily multimodal because human speech is situated in a body that is able to do other things than speaking concurrently with speaking. Some of these things are prerequisites for speaking, such as breathing (Włodarczak, Heldner, & Edlund, 2015) or preparing the articulators for speech, which can result in lip smacks or clicks.

More recently, there have been efforts to explore how speaking is connected to other physical aspects of communication. It has been shown that many facets of prosodic organization involve both phonetic and other kinds of physical expression: For example, stressed syllables have been shown to coincide with eyebrow movements (Swerts & Krahmer, 2008) and gesture peaks and eyebrow movements with intonation peaks (Borràs-Comes, Kaland, Prieto, & Swerts, 2014; Krahmer & Swerts, 2007; Leonard & Cummins, 2011; Loehr, 2007, 2012); eyebrow movements are also associated with turn taking (Guaitella, Santi, Lagrue, & Cavé, 2009) and gesture phrases with intonation phrases (Loehr, 2012). Not only are visible behaviors part of the production mechanism, they also enhance the perception of speech: Thus the traditional separation of the vocal channel from other physical modes of expression is questionable.

Some vocalizations are clearly rooted in, and contingent upon, physical conduct. For example, the phonetic form of the aforementioned strain grunts is rooted in the physical activity of doing something with effort. Holding and releasing the breath and the physical movement of lifting, bearing weight, and dropping it all impose physical demands on the body that are reflected in the sounds from the vocal tract, which in turn are rooted in the articulations of the body. The vocalizations made during such activities reflect to hearing and perhaps nonseeing participants what is going on and can be interpreted as a recruitment for assistance (Kendrick & Drew, 2016, see Excerpt 2).

One theme of the articles in this special issue is how the combination of physical events and vocal events can be deployed in interaction and for what kinds of social purposes. In our previous example (Excerpt 1), the coincidence of the two peaks in the tracing gesture and the clicks serve to produce a multimodal gestalt that can be oriented to as something that can be treated as funny (lines 11–12), as well as implicative of sequence closure (line 14). Instead of constituting a predetermined form-function package, the meaning of vocalizations emerges in their local action context between current participants in the multimodal process of sense-making.

Vocalizations and the body

The human body is both a complex organism and an intricate tool for interaction, precisely because it possesses a biological ability to move its various parts (limbs, vocal folds, torso, etc.) as well as a capacity to hear, see, smell, taste, and touch. We use all of these aspects to make sense of each other (Mondada, 2018). The accomplishment of

social action is intimately tied to people's ability to behave in a comprehensible manner and to competently interpret these very behaviors in entirety, not merely separating out a single stream of information, such as contained in the vocal sound (Goodwin, 2017). Therefore, we also need to analytically pay attention to both the multimodal aspects of action, such as the use of gaze and gesture alongside speech, and the role of multi-sensoriality, the fact that we can sense and understand others' sensing (Mondada, 2019). Crucially, the producer of a sound will be accountable for what it conveys at this very moment in this particular situation. Depending on the exact quality of the sound and its temporal placement, the bodily practices may be more or less in focus. But even in regular conversation the bodies are not trivial and passive bystanders in the intellectual exchange of abstract ideas.

Speech and vocalizations as bodily processes

Speaking is an embodied process, and linguistic practices are essentially dependent on a material substrate, including the body, the brain, and the environment (Linell, 2009, pp. 11–33). Still, a substantial amount of theorizing on language has taken it to be an abstract structure separate from the materiality of its production and perception. Apart from studies on sign language, gesture studies were perhaps the first to recognize the inherent relatedness of the movements of the hands and speaking (Schegloff, 1984) and even to argue that they originate in a single “growth point” in real time (McNeill, 2005). Already early research on interaction showed that the bodily orientation may convey relevant information in regard to the status of talk as being the main or a side concern through torquing the body in meaningful ways (Schegloff, 1998). Since then we have learned that gestures constitute a crucial aspect of sense-making in natural interaction (Streeck, 2009). We have also seen how linguistic resources are minutely coordinated with embodied ones, such as in a pointing gesture that needs to be “coupled” with a material object (Goodwin, 2007), and pointing can at the same time be used for turn taking in conversation (Mondada, 2007). Building on these studies on the fine local coordination of meaning-making resources, such as words, arms, and the body orientation, we are now moving the analytic focus to the vocal tract as one biological system among others in the human body, an aspect that has hitherto not been sufficiently highlighted, either in phonetics or in interaction research. Even though there is increasing interest in the embodied practices of interaction (Keevallik, 2018; Nevile, 2015), we tend to separate out vocal behavior from the rest of the body. One aim of our special issue is to develop a truly embodied account of sounds as they participate in accomplishing social action, such as making a joke about your imaginable tan line (Excerpt 1). We will show how some vocal tract sounds have a more obviously somatic origin, while, like laughter and crying, they can be positioned in meaningful ways in sequences of action.

Deploying the vocalizing body for social purposes

The somatic needs of the body do not define or delimit its communicative uses. Vocal tract sounds can be produced in such a way as to be audible when the person could be silent: Thus the very fact of inhaling or opening the mouth audibly take on an

interactional salience, precisely because they are hearable. Bodies feature a number of physiological needs, among them breathing, which is also necessary for speaking, breathing heavily when doing a physical exercise, and holding your breath to perform a heavy lift. Both breathing and impeding breath feature in this special issue, showing, however, that even these essential assets of a human body can be put into interactional use. Hoey (2020/this issue) demonstrates how a rapid nasal inbreath, a sniff, can organize turn taking by making evident that the person is not speaking or will not speak at this point, while Pehkonen discusses displaying affect with a conventionalized Finnish outbreath pattern *huh huh*. Furthermore, as Mondada shows (2020/this issue), in specific activities, such as tasting and smelling sessions, sniffing is organized as an accountable sequential action used for publicly demonstrating access to the source of the smell.

There are alternative more or less salient ways of, e.g., inhaling or opening one's mouth, involving various degrees of loudness (and thus hearability) as well as amplitude of movement (and thus visibility) (Mondada, 2020/this issue). These variables may be constitutive of action as being designedly public or not or even constitute segments of different kinds of social events altogether, such as either "beginning to speak" by breathing in with an open mouth or "not speaking" by sniffing through the nose (Hoey, 2020/this issue). Even sounds that seem like they would be "automatic" acquire a form and function specific to their place in ongoing activities and relative to talk; they are audible and can play a central role in the unfolding of interaction. Both Mondada and Hoey show that sniffs have implications for how further talk develops and how prior talk is treated. In addition, vocal tract sounds may be used as a socially legitimate alternative preoccupation of the body in relation to speaking (as described for humming by Stevanovic, 2013) or as a legitimization strategy for passing judgments on, e.g., a smell of a particular brand of beer. Somatically necessary sounds of the body may be deployed for social purposes and treated as meaningful by participants.

Let us consider another kind of a sound, a strain grunt, that can arguably be an involuntary "flooding out" of a body during physical exertion. Nevertheless, it is accountable for what it is doing in its local ecology, suppressable on the one hand and potentially part of action formation on the other. Excerpt 2 is taken from a recording of a clearing of a sheep stable of manure. Siim (circled in Figure 2) is digging slightly further away from the wheelbarrow where they need to place the loads. When he starts moving toward the wheelbarrow with a spadeful of manure, the person in charge of it has already begun to leave. At this very moment Siim displays urgency in trying to catch up with the rolling wheelbarrow, throwing his upper body forward, speeding up his steps and also producing a strain grunt (line 1, Figure 2). This strain grunt is quite long, with air escaping from the lungs high in his pitch register ([↑]), with rising pitch ([/]). The whole sound is represented as [↑/?(V)] in the transcript, with (V) used to capture an unspecific open vocal tract configuration and [°] to capture its mostly voiceless articulation: Together, these suggest that the vocal folds are held very tense, and his mouth is open. The wheelbarrow stops, Siim discontinues the sound by breathing out, thus audibly releasing the physical tension at the glottis (i.e., between the vocal folds). He is now able to dump the spadeful of muck into the wheelbarrow, almost falling over in the process (Figure 3).

Excerpt 2

Siim (circled) takes two steps toward the wheelbarrow with a forkful of manure. The wheelbarrow starts rolling out of the stable.

01 Siim: q#q*h [↑/?(Y) hh] #
 Fig #Fig.2 #Fig.3
 *wheelbarrow stops



Figure 2. Siim throwing himself toward the leaving wheelbarrow.



Figure 3. The wheelbarrow has stopped.

The sound here can be perceived to be part of a deliberate display of a physical effort, an action designed to be seen as well as heard, especially by the worker on the wheelbarrow. While the spadeful could not get any heavier when Siim was moving forward, the grunt instead highlights his acceleration through space, underlined by rising pitch, as well as his effort to reach the wheelbarrow before it leaves. Being intimately tied to the bodily exertion, the timing of the strain grunt can hardly be different, if only because producing it requires a considerable abdominal push. However, the sound ends at the moment when the wheelbarrow stops, which is before Siim manages to empty his spade (Figure 3), evidencing of the communicative function of the sound rather than its physiological necessity. Siim's rather spectacular throw of the manure is accomplished in silence. The strain grunt thus quite straightforwardly stops the coworker from leaving. In addition, the entire display of trying to do one's utmost to throw a last spadeful of manure into the leaving wheelbarrow achieves an overall socially preferable impression of someone really going in for the work task. A vocalization is but one aspect of the body that is collaboratively accomplishing a work assignment, belonging to the embodied social action, being indexical of it, and only vaguely meaningful on its own. By being precisely timed with the simultaneous body action, it not merely depicts it but also provides an immediate visual resource for its interpretation. As participants, we draw on our experiential knowledge of what physical action it takes to produce these sounds; we can literally sense the tension in his body.

Human bodies offer numerous opportunities for conveying meaning, some of which are intimately tied to its physiology (clearing the throat, sniffing, grunting), while others take practice

to master in a cultural context (whistling, language sounds, gesturing). Meaningful action can be a specific combination of those, such as sniffing just after putting the nose into a glass of beer when trying to become a connoisseur or grunting while showing a dance move that is too heavy when teaching students to dance. Many communicative practices involve the engagement of both the vocal tract and the rest of the body, such as an eyebrow flash together with a click that conveys appreciation of something huge (Ogden, 2020/*this issue*) or nodding while providing a type-conforming answer to a polar question (Kärkkäinen & Thompson, 2018). There is no a priori justification to prioritize the analysis of vocal action in all interaction (Mondada, 2016), especially if we want to understand the limits of language and the intersection of body and meaning.

Semantic underspecification and local understanding

All lexical items receive their precise meaning locally, in a concrete situation between specific participants, i.e., in a dialogical process (Norén & Linell, 2007). Less conventionalized vocalizations, however, seem to be especially flexible in their adaptations to immediate interactional contingencies. Partly because of the physiological body-voice connections, vocalizations tend to receive their central meaning characteristics from the current embodied trajectories of action. A vocal tract sound regularly conveys immediacy of something that the body is undergoing. An early study by Wiggins (2002) showed how a specific type of *mmm* performs currently experienced pleasure of food. It needs to be placed after visibly having put the food into the mouth. A Finnish *huh huh* is uttered precisely at transitions from strenuous activities to non-strenuous ones (Pehkonen, 2020/*this issue*). Similarly, a vocalizing of pain or discomfort needs to be immediate in respect to, e.g., doctor's pressure, in order to be deemed visceral (Weatherall, Keevallik, & Stubbe, 2019).

These kinds of various arguably internal states have been of interest to interaction researchers in regard to how they are organized “as if” externalizations of cognitive, physical or affective states. Most notably, rather than being spontaneous outbursts revealing some kind of an internal experience, emotion displays have been shown to be sequentially organized and thus designedly accomplishing social actions (Peräkylä & Sorjonen, 2012; Wilkinson & Kitzinger, 2006). Affective displays such as “surprise,” “appreciation,” or “disappointment” have furthermore been shown to feature embodied aspects in children's arguing (Goodwin & Goodwin, 1987). Hofstetter (2020/*this issue*) pursues this line of research by targeting “moans” during board games, the extended sounds made after game moves that are damaging for self or other's opportunities to win. A “moan” would not be interpretable on its own and leads participants to immediately search for what occasioned it, good or bad (Goffman, 1981). This kind of semantic underspecification, vagueness, is clearly an interactional asset, making the vocalizations adaptable to the particularities of different occasions and, in comparison to their verbal counterparts, also less vulnerable to challenge (Hofstetter, 2020/*this issue*). As the studies show, vocalizations make perfect sense for the participants within ongoing trajectories of actions. If a dad has just lifted up his kids on a high stone in the forest, the *huh huh* highlights his fatigue after a bodily accomplishment, which is exactly how one of the kids verbally interprets it (Pehkonen, 2020/*this issue*). The embeddedness of vocalizations in local temporalities of embodied action is crucial in their use for meaning-making.

Making sense of vocalizations

Working on semantically underspecified tokens forces us as analysts to push the methods of conversation analysis to their limit and at the same time exposes their true strength, as the locally relevant meaning can only be revealed in participants' next actions. One of the important findings of research on vocalizations is a better understanding of the contextualized methods and resources participants themselves have for making sense of things they have never seen or heard before, at least not in this particular combination. Across the range of articles in the collection, we see that the methods of meaning-making include attention to not only the position of the item in a turn and sequence but also its iconic or indexical aspects; exact timing in relation to current bodily action; co-occurring linguistic and prosodic features of the turn; as well as material, spatial, and other contextual matters in the local environment. These also constitute tools with which participants can make sense of singularities, i.e., one-off productions such as the click construction in our opening example, that feature vocalizations that are seemingly underspecified but cause no problem in understanding in their immediate ecology, instead featuring precise adaptation to the local communicative task, such as depicting the tan line (Excerpt 1).

Especially in activities that are more centered on the body than conversation, the exact phonetic features of a vocalization may be motivated by what the body is doing rather than the phonetic inventory of a language. Likewise, they need to be interpreted in relation to the simultaneous movements rather than as a “code” abstracted from the acting body. Excerpt 3 is taken from a dance class of Lindy Hop where two teachers in a couple (lead = TeaL and follow = TeaF) are demonstrating a choreography. They are standing side by side (see [Figure 4](#)), and the follow teacher launches the performance with a “We’re gonna go.” The ensuing lines 2–5 show beat numbers in gray above the transcription of the sounds that are rendered in orthographically approximative syllables, where [ʔ] marks a glottal stop, which gives the vowels an abrupt onset or offset.



Figure 4. Launching the demonstration.

Excerpt 3

1	TeaF:	We ('re) gonna go?# (0.3)							
	Fig	#Fig.4							
		1	2	3	4	5	6	7	8
2	TeaF:	wah	pu	ʔa		ʔa		ZA	pu
	teaL:	(back)							
		1	2	3	4	5	6	7	8
3	TeaF:	ʔa:		ʔa:		ʔa:		HEJʔ	ʔa
	teaF		x		x		x		((snaps))
		1	2	3	4	5	6	7	8
4	TeaF:	Bum		haʔ		haʔ		WAP	pum
		1	2	3	4	5	6	7	8
5	TeaF:	ZAP	pum	pum	pum	pa	pu	ʔa	ʔu ʔa

In an activity like this, the vocal production is to be heard as depicting what the bodies are doing, being in themselves partially “natural” expressions of the body and partially illustrations for the students how every current move should feel. Stops accompany sharper moves, such as steps on the floor, while lengthened vowels mark holds. Rises in pitch, such as in line 2 and 5, happen on the flashy moments in the choreography, a glottal fricative *h* (line 3) marks a kick in the air, and a voiced sibilant (line 5) accompanies a slide on the floor. Nevertheless, there is no fixed relationship between the particular steps and the sounds that accompany them; the same type of a kick can sound as a *ʔa* (line 2), a *HEJʔ* (line 3), or a *haʔ* (line 4). When presenting the same step sequence a moment later, the teachers do not use identical syllables (even though there are certainly some favored sounding patterns in the community of practice in this style of dance instruction). In any case, this phonetic variability is not in the least problematic for the participants in the activity setting of learning how to dance. The context, the respective roles of the teachers and the students, the spatial arrangements (students in a circle and the teachers in the middle), the performing bodies—all provide resources for making sense without necessarily having a conventionalized relationship between the sound and the meaning conveyed, which in this case essentially concerns the body. Human beings are apparently able to produce and understand action through a whole range of resources besides the vocal ones, and the vocal behavior may be rather coincidental in relation to what is at stake in a particular situation.

Recognizable practices for action

Decades of conversation analytic research have shown that turns-at-talk accomplish sequential actions (Schegloff, 2007), and all the vocal materials within a turn, including sound stretches and prosody, contribute to action formation (Walker, 2017). In co-present activities in particular, action formation is regularly multimodal. Depending on the activity context, it may involve artifacts, documents, computer screens, spatial positioning, others’ bodies, etc. Sequential action, such as complying with a request, can be entirely embodied (Rauniomaa & Keisanen, 2012). However, an action can also be accomplished by a vocalization, such as a “moan” that does not have an entirely conventional format but nevertheless is a recognizable practice for a bemoaning

action (Hofstetter, 2020/*this issue*). A sound accomplishing a distinct social action comes close to being similar to what linguists call *interjections* and therefore to being part of a language. Yet the phonetic underspecification of the moans suggests that they are on the margins of the more “encoded” parts of language and more indexically tied to the precise sequential and activity environments. Ultimately, conventionalization is a question of degree. An item such as *oh* can be almost equally variable as a moan. Some of the vocal(-bodily) practices get routinized and become recognizable as clear combinations of form and function for the participants across contexts, the ultimate results of this process being lexical items and grammar. In the special issue we are concerned with degrees of conventionalization of less-acknowledged sounds in respective languages. For example, Reber and Couper-Kuhlen (2020/*this issue*) show how whistling accomplishes a sequential social action in English (and Spanish), and Li (2020/*this issue*) demonstrates how midturn clicks in Mandarin are systematically used for reshaping ongoing verbal action, in close response to recipients’ bodily displays.

Yet another issue is how the vocalizations get incorporated into meaningful units in conversation. The two clicks at line 10 in Excerpt 1 are treated as a turn constructional unit (Sacks, Schegloff, & Jefferson, 1974), even though there is no conventional verbal material (see Keevallik, 2014, for arguments on vocalizations constituting turn constructional units). In Excerpt 2 the dance teacher launches a syntactic unit by “We’re gonna go” but ends it with a dance demonstration accompanied by vocalizations, together with her partner. Embodied demonstrations have been shown to systematically complete units in co-present interaction, where the temporal coordination between language and the body is crucial (Keevallik, 2013, 2015; Mori & Hayashi, 2006). Also, various embodied displays, such as smiles and frowns, have been shown to occur in preturn beginning positions and as their postcompletion, showing essentially stance (Kaukomaa, Peräkylä, & Ruusuvuori, 2013, 2014). Likewise, the positioning of various sounds with respect to emerging speech turns is crucial in regard to their function. This point is exemplified in the articles by Hoey, Ogden, and Li, who show the relevance of the temporal placement of sniffs and clicks with respect to turn constructional units and thus in their contribution to action. They display several uses where the connection between the precise phonetic form of a vocalization and its function is relatively robust in particular sequential and turn positions. Furthermore, they may display functional contrasts with other sounds in identical positions, such as centrally versus laterally released clicks (Ogden) or sniffs versus inbreaths (Hoey), thereby at least in a rudimentary way forming functional paradigms, which is considered to be one of the main features of linguistic systems.

Various sounds are shown to be able to accomplish recognizable actions in the special issue, thus displaying a degree of conventionalization. They can also be incorporated into the syntagmatic and paradigmatic structures of language. Still, competence in their production and use is not quite on a par with mastering certain more systematic aspects of grammar or the knowledge of lexicon. For example, one does not need to be able to whistle to participate in an English conversation nor perhaps to produce clicks to speak Mandarin. A sniff with a nose in a glass of beer can be done regardless of which language the smell will then be commented on. In other words, these seem to be more or less language-neutral capacities and competencies, which underlines their nonconventional character, also reflected in their design. They rely on a perceptual system that is able to associate sounds with the physical configurations behind them.

Future directions

This collection of articles celebrates the benefits of rigorous local analysis of sense-making with vocal tract sounds, both for understanding the emergent phonetic features of their real-time production but also for demonstrating their importance for coordinated action. We document the intersubjective relevance of “physiological” sounds on the one hand (such as loud inbreaths or strain grunts) and the interactional use of highly complex sounds that do not belong to the phonemic inventory of the language (e.g., whistles or clicks) on the other. Unconventional or only partially conventional vocalizations are furthermore shown to accomplish particular social tasks in the form of moans during board games or as accompaniments to dance instruction. In all the studies, the temporal organization emerges as absolutely crucial in determining the action-import of a sound, either in terms of their positioning in turns-at-talk or in relation to what the body is doing. The joint contribution of this special issue is to demonstrate how a focus on action, rather than language as a code of sounds, provides us with a new and productive angle for dissecting sense-making practices.

We have barely scratched the surface of this exciting field, leaving much work yet to be done on different languages and various sounds. More work along these lines will further our understanding of what is “natural” and what is “conventional” about human vocal behavior. Among other things, it is possible that meaning is also accomplished via certain phonetic features or their change over time, such as closeness versus openness, pitch trajectories, or loudness, as opposed to merely being conveyed through lexicon and grammar. This opens the door to a much deeper understanding of how indexicality and iconicity, as well as immediate physical contexts, contribute to participants’ sense-making practices. We are raising serious issues with the boundary drawing between language and nonlanguage, conventional and nonconventional, where language theorists may want to work further from whatever their perspectives are.

References

- Akita, K., & Dingemanse, M. (2019). Ideophones (Mimetics, Expressives). In M. Aronoff (Ed.), *Oxford research encyclopedia for linguistics*. Oxford, England: Oxford University Press.
- Borràs-Comes, J., Kaland, C., Prieto, P., & Swerts, M. (2014). Audiovisual correlates of interrogativity: A Comparative analysis of Catalan and Dutch. *Journal of Nonverbal Behavior*, 38(1), 53–66. doi:10.1007/s10919-013-0162-0
- Clark, H. (2016). Depicting as a method of communication. *Psychological Review*, 123(3), 324–347. doi:10.1037/rev0000026
- Couper-Kuhlen, E., & Ford, C. (Eds.). (2004). *Sound patterns in interaction*. Amsterdam, The Netherlands: Benjamins.
- Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in conversation: Interactional studies*. Cambridge: Cambridge University Press.
- Dingemanse, M. (2017). On the margins of language: Ideophones, interjections and dependencies in linguistic theory. In N. J. Enfield (Ed.), *Dependencies in language* (pp. 195–202). Berlin, Germany: Language Science Press.
- Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a Journal of General Linguistics*, 3(1), 4.
- Dingemanse, M. (2020). Between sound and speech: Liminal signs in interaction. *Research on Language and Social Interaction*, 53(1), 188–196.
- Enfield, N. J. (2009). *The anatomy of meaning*. Cambridge: Cambridge University Press.

- Gardner, R. (2002). When *listeners talk*: Response tokens and recipient stance with special reference to *Mm*. Amsterdam, The Netherlands: Benjamins.
- Gil, D. (2013). Para-Linguistic usages of clicks. In M. S. Dryer & M. Haspelmath Eds., *The world atlas of language structures* online. Retrieved from <http://wals.info/chapter/142>
- Glenn, P., & Holt, E. (Eds.). (2013). *Studies of laughter in interaction*. London, UK: Bloomsbury Academic.
- Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.
- Goodwin, C. (2007). Environmentally coupled gestures. In S. D. Duncan, J. Cassel, & E. T. Levy (Eds.), *Gesture and the dynamic dimension of language* (pp. 195–212). Amsterdam, The Netherlands: Benjamins.
- Goodwin, C. (2017). *Co-operative Action*. Cambridge: Cambridge University Press.
- Goodwin, C., Goodwin, M. H., & Olsher, D. (2002). Producing sense with nonsense syllables: Turn and sequence in conversations with a man with severe aphasia. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The language of turn and sequence* (pp. 56–80). Oxford: Oxford University Press.
- Goodwin, M. H., & Goodwin, C. (1987). Children's arguing. In S. Philips, S. Steele, & C. Tanz (Eds.), *Language, gender, and sex in comparative perspective* (pp. 200–248). Cambridge: Cambridge University Press.
- Guaitella, I., Santi, S., Lagrue, B., & Cavé, C. (2009). Are eyebrow movements linked to voice variations and turn-taking in dialogue? An experimental investigation. *Language and Speech*, 52 (2–3), 207–222. doi:10.1177/0023830909103167
- Heinemann, T., & Koivisto, A. (2016). Indicating a change-of-state in interaction: Cross-linguistic explorations. *Journal of Pragmatics*, 104, 83–88. doi:10.1016/j.pragma.2016.09.002
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299–345). Cambridge, U.K.: Cambridge University Press.
- Hoey, E. M. (2014). Sighing in interaction: Somatic, semiotic, and social. *Research on Language and Social Interaction*, 47(2), 175–200. doi:10.1080/08351813.2014.900229
- Hoey, E. M. (2020). Waiting to inhale: On sniffing in conversation. *Research on Language and Social Interaction*, 53(1), 118–139.
- Hofstetter, E. (2020). Non-lexical 'moans': Response cries in board game interactions. *Research on Language and Social Interaction*, 53(1), 42–65.
- Kärkkäinen, E., & Thompson, S. A. (2018). Language and bodily resources: 'Response packages' in response to polar questions in English. *Journal of Pragmatics*, 123, 220–238. doi:10.1016/j.pragma.2017.05.003
- Kaukomaa, T., Peräkylä, A., & Ruusuvuori, J. E. (2013). Turn-opening smiles: Facial expressions constructing emotional transition in conversation. *Journal of Pragmatics*, 55, 21–42. doi:10.1016/j.pragma.2013.05.006
- Kaukomaa, T., Peräkylä, A., & Ruusuvuori, J. E. (2014). Foreshadowing a problem: Turn-opening frowns in conversation. *Journal of Pragmatics*, 71, 132–147. doi:10.1016/j.pragma.2014.08.002
- Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, 5(2), 275–292. doi:10.1017/S095267570000230X
- Keevallik, L. (2013). The interdependence of bodily demonstrations and clausal syntax. *Research on Language and Social Interaction*, 46(1), 1–21. doi:10.1080/08351813.2013.753710
- Keevallik, L. (2014). Turn organization and bodily-vocal demonstrations. *Journal of Pragmatics*, 65, 103–120. doi:10.1016/j.pragma.2014.01.008
- Keevallik, L. (2015). Coordinating the temporalities of talk and dance. In A. Deppermann & S. Günthner (Eds.), *Temporality in Interaction* (pp. 309–336). Amsterdam, The Netherlands: Benjamins.
- Keevallik, L. (2018). What does embodied interaction tell us about grammar? *Research on Language and Social Interaction*, 51(1), 1–23. doi:10.1080/08351813.2018.1413887
- Kendrick, K. H., & Drew, P. (2016). Recruitment: Offers, requests, and the organization of assistance in interaction. *Research on Language and Social Interaction*, 49(1), 1–19. doi:10.1080/08351813.2016.1126436

- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396–414. doi:10.1016/j.jml.2007.06.005
- Ladd, D. R. (1978). Stylised intonation. *Language*, 54(3), 517–540. doi:10.1353/lan.1978.0056
- Laver, J. (1994). Principles of *phonetics*. Cambridge: Cambridge University Press.
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471. doi:10.1080/01690965.2010.500218
- Li, X. (2020). Click-initiated self-repair in changing the sequential trajectory of actions-in-progress. *Research on Language and Social Interaction*, 53(1), 90–117.
- Linell, P. (2009). Rethinking *language, mind and world dialogically*: Interactional and contextual theories of human sense-making. Charlotte, NC: Information Age Publishing.
- Local, J., & Kelly, J. (1986). Projection and silences: Notes on phonetic and conversational structure. *Human Studies*, 9(2–3), 185–204. doi:10.1007/BF00148126
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7(2), 179–214. doi:10.1075/gest.7.2
- Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3, 71–89. doi:10.1515/lp-2012-0006
- McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: Chicago University Press.
- Mondada, L. (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2), 195–226. doi:10.1177/1461445607075346
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3), 336–366. doi:10.1111/josl.1_12177
- Mondada, L. (2018). The multimodal interactional organization of tasting: Practices of tasting cheese in gourmet shops. *Discourse Studies*, 20(6), 743–769. doi:10.177/1461445618793439
- Mondada, L. (2019). Contemporary issues in conversation analysis. Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145, 47–62. doi:10.1016/j.pragma.2019.01.016
- Mondada, L. (2020). Audible sniffs: Smelling-in-interaction. *Research on Language and Social Interaction*, 53(1), 140–163.
- Mori, J., & Hayashi, M. (2006). The achievement of intersubjectivity through embodied completions: A study of interactions between first and second language speakers. *Applied Linguistics*, 27, 195–219. doi:10.1093/applin/aml014
- Nevile, M. (2015). The embodied turn in research on language and social interaction. *Research on Language and Social Interaction*, 48(2), 121–151. doi:10.1080/08351813.2015.1025499
- Norén, K., & Linell, P. (2007). Meaning potentials and the interaction between lexis and grammar. Some empirical substantiations. *Pragmatics*, 17, 387–416. doi:10.1075/prag.17.3.03nor
- Ogden, R. (2013). Clicks and percussives in English conversation. *Journal of the International Phonetic Association*, 43(3), 299–320. doi:10.1017/S0025100313000224
- Ogden, R. (2020). Audibly not saying something with clicks. *Research on Language and Social Interaction*, 53(1), 66–89.
- Ogden, R., Hakulinen, A., & Tainio, L. (2004). Indexing “no news” with stylization in finnish. Sound patterns in interaction. *Cross-Linguistic Studies from Conversation*, 62, 299.
- Pehkonen, S. (2020). Response cries inviting an alignment: Finnish *huh huh*. *Research on Language and Social Interaction*, 53(1), 19–41.
- Peräkylä, A., & Sorjonen, M.-L. (Eds.). (2012). *Emotion in interaction*. Oxford, England: Oxford University Press.
- Rauniomaa, M., & Keisanen, T. (2012). Two multimodal formats for responding to requests. *Journal of Pragmatics*, 44(6–7), 829–842. doi:10.1016/j.pragma.2012.03.003
- Reber, E. (2012). *Affectivity in interaction: Sound objects in English*. Amsterdam, The Netherlands: Benjamins.
- Reber, E., & Couper-Kuhlen, E. (2020). On ‘whistle’ sound objects in English everyday conversation. *Research on Language and Social Interaction*, 53(1), 164–187.
- Sacks, H. (1992). *Lectures on conversation* (Vols. 1–2). Oxford, England: Blackwell.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735. doi:10.2307/412243

- Schegloff, E. A. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 266–295). Cambridge, England: Cambridge University Press.
- Schegloff, E. A. (1998). Body torque. *Social Research*, 65(5), 536–596.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, England: Cambridge University Press.
- Sorjonen, M.-L. (2001). Responding in *conversation*. Amsterdam, The Netherlands: Benjamins.
- Stevanovic, M. (2013). Managing participation in interaction: The case of humming. *Text & Talk*, 33(1), 113–137. doi:[10.1515/text-2013-0006](https://doi.org/10.1515/text-2013-0006)
- Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. Amsterdam, The Netherlands: Benjamins.
- Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, 36(2), 219–238. doi:[10.1016/j.wocn.2007.05.001](https://doi.org/10.1016/j.wocn.2007.05.001)
- Walker, G. (2017). Pitch and the projection of more talk. *Research on Language and Social Interaction*, 50(2), 206–225. doi:[10.1080/08351813.2017.1301310](https://doi.org/10.1080/08351813.2017.1301310)
- Weatherall, A., Keevallik, L., & Stubbe, M. (2019, July 03–05). *Being incapacitated by pain during medical consultations*. Paper presented at IEMCA, Mannheim University, Mannheim.
- Wharton, T. (2003). Interjections, language and the “showing’/’saying” continuum. *Pragmatics & Cognition*, 11, 39–91. doi:[10.1075/pc.11.1.04wha](https://doi.org/10.1075/pc.11.1.04wha)
- Wiggins, S. (2002). Talking with your mouth full: Gustatory ‘mmm’s and the embodiment of pleasure. *Research on Language and Social Interaction*, 35(3), 311–336. doi:[10.1207/S15327973RLSI3503_3](https://doi.org/10.1207/S15327973RLSI3503_3)
- Wiggins, S. (2013). The social life of ‘Eugh’: Disgust as a form of evaluation in family mealtimes. *British Journal of Social Psychology*, 52(3), 489–509. doi:[10.1111/j.2044-8309.2012.02106.x](https://doi.org/10.1111/j.2044-8309.2012.02106.x)
- Wiggins, S. (2019). Moments of pleasure: A preliminary classification of gustatory mmms and the enactment of enjoyment during infant mealtimes. *Frontiers in Psychology*, 10. doi:[10.3389/fpsyg.2019.01404](https://doi.org/10.3389/fpsyg.2019.01404)
- Wilkinson, S., & Kitzinger, C. (2006). Surprise as an interactional achievement: Reaction tokens in conversation. *Social Psychology Quarterly*, 69(2), 150–182. doi:[10.1177/019027250606900203](https://doi.org/10.1177/019027250606900203)
- Wlodarczak, M., Heldner, M., & Edlund, J. (2015). Breathing in conversation: An unwritten history. In K. Jokinen & M. Vels (Eds.), *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication* (pp. 107–112). Linköping: Linköping University Press.