Listening: Of States and Traits

Jonas Obleser

Hearing, listening, understanding: What is different, even more complicated about this process than about seeing? The sense of hearing and our ability to listen represent a special test case if one wants to make and assess hypotheses about brain functions: First of all, how does the brain manage to transform the one-dimensional fluctuations in air pressure into locatable auditory objects and then into codes, which can be interpreted, be they noise, speech or music? In the psychological and neuroscientific research presented here, the focus is on individual listeners, the biological-psychological condition in which they find themselves ("states") as well as the longer-lasting characteristics that distinguish them from other listeners ("traits"). Our listening is individual because it always serves certain behavioral purposes, and these purposes differ and shape our strategies of listening. Last, but not least, we will see that in listening, not listening, and merging our auditory and visual sensations, the psychological process of attention plays a key role.

1. Introduction

People hear. And people listen. English differentiates between "hearing" and "listening" more clearly than German does. The distinction marks a completely different, additional apparatus of psychological – and also, for me, always neurobiological – processes that follow. We quickly realize that it must be the listeners themselves who achieve the listening in the first place, while they cannot avoid hearing as such.

Listening, this "feeling at a distance", uses that curious sensorium that is our ear – especially that inner ear stimulated by the smallest ossicles, very sensitive and, therefore, easily damaged by noise, toxic substances as well as by increasing age. But it is a sense, which allows visual animals – humans – to notice dangers as well as other interesting things at a distance or behind their backs and from the side.

On the other hand, those who want to listen have only this sense of hearing available. But they must use a "process", a "module" (as one could call it in the language of the Information Age familiar to us), a whole family of processes that allow this sense to filter and evaluate the heard signal. There must be physical properties of the auditory impression in which we are interested, which allow us to distinguish it from other impressions and, thus, to declare

certain parts of this auditory impression as “signal” and others as “noise” or background. Such a “filter” (as I would like to call it without further ado), which accomplishes this, almost at will, in accordance with our intentions and goals, has, as we know, had a name in psychology for 130 years (James, 1890/1927), namely attention — even though, this alone will help us amazingly little.

So is listening nothing more than attentive hearing? The short answer is: yes. But this isn’t where it gets boring for us, because the question that has been pending for at least these 130 years is: What *is* attention? *How* do I succeed in listening? What is more important, the “bringing-into-focus” – how helpless our visually based language is here! – of an auditory *object* or rather, the successful suppression and filtering out of those auditory impressions which are not interesting to us and, above all, potentially distracting to us? But then what power do the acoustic signals themselves have to grab our attention and why do some sounds or noise sources succeed more reliably than others? (“Salience” is the technical term often used here, and “bottom-up”, i.e. from below or outside, is the direction of its effect on our listening apparatus.) And secondly, how strongly or how completely do listeners succeed in “implementing” this filter, i.e., in enforcing it – sometimes despite the high salience of a signal to be ignored (e.g., a conversation at the neighboring table)? (“Top-down”, i.e. from top to bottom or outside, listeners have to make use of their filter here.)

Most of the questions dealt with in this talk about the relationship between hearing and seeing, between listening and viewing, are psychological questions, i.e. first of all questions about human experience and behavior. As such, they can be addressed, if not exhaustively answered, with methods of experimental psychology: with elaborate experimental setups, with so-called interference and cue stimuli, with measured reaction times and error rates, with mathematical models of our perceptual and decision processes. And if attention is conceived of and measured in purely auditory terms, then this stimulus material naturally involves acoustic stimuli – music, speech, tones and noise. The famous “cocktail party” – long out of fashion after the mid-twentieth century – inspired the golden decades of attention research after World War II: How do we manage to ignore the murmuring and clatter around us so well and instead focus on the person we are conversing with? But at least as useful and even more impressive, how do we manage to listen unobtrusively to the latest gossip in the conversation behind us instead of the boring monologue of the business partner in front of us? (Cherry, 1953). Thus, it was the auditory system as a model from which the model conceptions of attention emerged that were central at the time, and in some cases still are (e.g., Broadbent’s bottleneck model; for an introduction, see Serences & Kastner, 2014).

Interestingly, in the following decades, something happened that could be called a “visual turn” in attention research (as if the sense of sight had not always dominated psychology and brain research anyway). And so today we know much better how vision and the transduction of visual information into neural signals works, and how the seeing brain organizes itself in different species, especially in humans. How our attention takes objects into the “spotlight” (as found in the terminus technicus of *attentional spotlight*) and what role the features of each object or its location in space and distracting stimuli play – all of which is now surprisingly well understood. On the other hand, in the auditory realm, the very formation of an auditory object – based on which physical features are sound stimuli grouped as belonging together? – is not without controversy. And so, the physics of the auditory stimulus and the nature of the sensory epithelium adapted to it complicate our scientific understanding from the very first conversion of sound into electrical signals.

Moreover, this research on the different sensory modalities has long been rigorously conducted separately in very reductionist research paradigms. At times they lost — perhaps necessarily — sight of (and an ear for) the fact that perception and attention occur in a naturally multi-sensory stimulating environment. And only in the last twenty years or so, also thanks to considerable progress in mathematical-statistical methods of evaluation, as well as psychological and neuroscientific studies, has been reunited what the individual can only experience in concert: Watching while hearing or listening while seeing.

II. Watching while hearing or listening while seeing

Thus, multi-sensory processing of environmental stimuli is initially the only processing conceivable for humans, even if the reductionist research paradigm often disregards this for good reasons. However, at the latest, when a rudimentary understanding of how, for example, an area of the auditory cortex encodes auditory stimuli is accomplished, we ask ourselves: What actually changes for neurons of this auditory cortex area when visual stimuli arrive concomitantly? Do they integrate the additional, simultaneous or at least temporally closely coordinated presence of visual stimuli? How early in the processing hierarchy (inner ear, brainstem, diencephalon, primary auditory cortex, etc.) does the effect of such multi-sensory signals have consequences, i.e., does it really change the neural code?

The most interesting question at the moment is how we (or more reductionistically: our brain) manage to conclude that the sound emitted and the light reflected originate from one and the same object, e.g. the second violinist there in front, whose body movements match the

sound of her instrument so well. How easy it is to disturb this process of so-called causal inference ("Is there a common cause or source of my auditory and visual sensory impressions?"). We become aware of this when we see Bruce Springsteen and the E-Street band in the Olympic Stadium in Berlin at a distance of hundreds of meters and the sound pressure reaches us only some hundreds of milliseconds later or when an audio track quickly turns out to be clearly out of sync with the video on a platform like YouTube. However, both misalignments are very quickly "repaired" by our brain. After a few minutes of unchanging conditions, we often no longer notice such deviations and again perceive a common source or cause for this audiovisual stimulation. Thus this commonality is (re-) constructed, and so a reductionist view that first describes the separation, what is initially inherent to auditory as well as visual neural processing, is not so wrong.

Is there a primacy of the visual in this construction of perception? Some findings suggest this, but first the sheer physics of seeing and hearing play their part here: light is faster than sound. Inevitably, the image of the man out front on stage with his Telecaster guitar reaches my retina and also my visual cortex faster than the sound from his guitar amplifier reaches my inner ear and my auditory cortex. Accordingly, we also perceive as "natural" what is seen milliseconds ahead of what is heard. This temporal offset is also what allows me, as a perceiver, to make predictions about the expected auditory signal based on the visual. This *predictive coding* of sensory information through a balancing integration of an internal "generative model" ("What do I actually expect of occurring sensory stimuli?") on the one hand, and what is merely deviating from it, what has real novelty value and is somewhat technically called "prediction error", on the other – this idea of the course of multisensory perceptual processes not only dominates the neuroscientific discourse of our time (Friston, 2010), but has been *de rigueur* since Hermann von Helmholtz.

As an example, consider the lip, tongue, and jaw movements of a speaker that I involuntarily look at in natural communication and that provide information about which acoustic features will immediately reach my ear – and those which have already become completely improbable because of this visual information. The McGurk effect (McGurk & MacDonald, 1976), which has become famous but cannot be reproduced quite so reliably in practice, may be mentioned here as an illustration: Seeing a mouth producing one syllable ("ga") while another syllable is sounded ("ba") makes us perceive a third syllable ("da"). The absence of the lip closure characteristic of "ba" already rules out the actual sounding syllable as a candidate, and the "da" heard instead represents, phonetically speaking, a kind of compromise between the conflicting visual and auditory evidence. A contemporary and somewhat more elaborated 2020 variant of this effect, savoring the playfulness of social

media as well as the power of written language on our perception, can be found after a brief Internet search for the words “green needle” and “brainstorm”: A young woman looks into the camera and indicates with her hand that we should read one of the two phrases shown above her, i.e. “green needle” at the top left and “brainstorm” at the top right – astonishingly, the same soundtrack, slightly reminiscent of science-fiction TV music from the 1960s, is played in an endless loop. And indeed, our brain effortlessly “hears” the respective word read quite clearly from that always identical, albeit strange, sound. What has happened? Magic, an alchemical trick? Not at all; here an “acoustic chimera” was created (Smith et al., 2002) which actually combines acoustic features of two recordings in such a way that this combination in and of itself (i.e. purely acoustically) is no longer understandable for us. Again, however, the sense of sight (here, what is read) can provide a kind of perceptual *a priori.* On the basis of this guiding signal we can then extract and use the matching auditory partial information, i.e., we “understand”.

It becomes particularly productive for our ideas of what multisensory perception can actually achieve when we return to the questions of attention and how it succeeds, which we have repeatedly dealt with here: If I can hear attentively, i.e. listen, and see attentively, i.e. watch – how do I do both at once? And, can I do this at all? If visual attention is rhythmic, i.e. the images of our retina reach our awareness only in timed windows of prioritized processing (cf. e.g. Fiebelkorn & Kastner, 2019; van Rullen & Koch, 2003), what does this mean for simultaneous listening? As long as it is about that same second violinist whom one is listening to and whose body movements one wants to follow, we can certainly assume a so-called supramodal attentional pulse, i.e., one that acts uniformly across sensory modalities. But what if our sense of hearing, on the one hand, and our sense of sight, on the other, want or ought to pay attention to different objects or sources? What at first seems to be of little use to us in everyday life (why watch the bass player closely while listening to the cellist?) is easily achieved by undergraduates in the psychological laboratory.

And in our everyday lives, too, we can find numerous situations of communication or stimulation that are both auditory and visual in nature. The child listens spellbound to a story about Arthur and the Knights of the Round Table, while creating a fairy kingdom on paper with his eyes and hands. The scientist at the hospital talks animatedly on the phone with a colleague about technical matters and is quite content to watch the coming-and-going of the delivery men, couriers and ambulances outside the office window. In situations like these, we find a fluidity and ease of switching attention between modalities that makes humans come across far less clearly as “visual animals”.

III. Is there a new aurality?

Without a doubt, hearing and listening have attributes of the moderate, the quiet, the contemplative. Both the child painting and listening to a story as described above, and the pensive person on the telephone, looking out the window appear absorbed rather than agitated. But this is of course a cliché. Humans also know how to consume, extract, utilize and exploit the spoken word. My students in the pandemic year 2020 like to listen to my statistics teaching videos at 1.5 times the speed, an invaluable advantage over the real lecture in the lecture hall, I'm sure. Conversely, many people have become fond of sending their messages to others as short voice messages – after a phase of mobile communication when it was thought that the spoken word was dying out in favor of text messages. This speaking is indeed faster than the manual translation of our thoughts into the cultural technique of written language. But this is a Pyrrhic victory for sound and spoken language, because it is now up to recipients again to listen in full to these voice messages, i.e. re-linearize them, instead of being able to "scan" them with their eyes at breakneck speed, i.e. read them, as is possible with emails or text messages.

And perhaps the most exciting, but certainly the most recent volte-face of this back-and-forth of media occurred also in the pandemic year 2020, when, in addition to the astonishingly popular medium of the so-called podcast, i.e. the exchange of spoken words as information and entertainment (not unlike good old radio), a new (and, as must be added as of fall 2021, apparently rather short-lived) social media app called Clubhouse emerged: What a person expressed was not written or displayed there, but only – spoken. A completely image-free social medium, a kind of call-in panel show or even talk radio for everyone (or at least for those with an invitation; a clubhouse, in other words). The final classification of these supposedly new or at the very least revived aural phenomena, and whether they actually signify the return or re-affirmation of the auditory in a world that is on the face of it becoming more and more visual, can gladly be left to media theorists at this point – but as an auditory researcher, I may at least find it worth noting.

IV. How do I listen? Listening as a state and a trait

In general psychological terms, we are interested in describing the regularities that determine the relationship between auditory and visual perception and attention. However, from the stance of differential psychology, we would ask: First, why does one person succeed better at listening in the morning and the other person in the car on the way home in the evening

(i.e. depending on so-called “states”)? We are investigating this in various laboratory studies by measuring and describing momentary brain activity and attempting to mathematically predict, for example, the encoding depth of what is then heard (Alavash et al., 2019; Waschke et al., 2019).

In Alavash, Tune and Obleser (2019), for example, we measured the momentary connectivity of different brain areas, e.g. classic auditory areas in the temporal lobe of the brain, with those to which we attribute more controlling or steering functions in the frontal and occipital lobes. Our subjects underwent a strenuous listening task with two competing voices while we mapped the interconnectedness of the entire brain as more or less synchronized fluctuations in the oxygen saturation of all these brain areas (this can be done with so-called functional magnetic resonance imaging, fMRI). The ability of a listener, confronted with a sudden change of plan in a large logistical network, such as an airline, to adjust and reconfigure these neural communication pathways from a state of relative calm when this difficult listening situation occurred, proved to be an interesting measurement: it predicted how successfully (how correctly, and how quickly) the individual subject managed such a “cocktail party”-type listening task.

Secondly, why is it that, in general (or over an average of situations), you may manage better than your partner to cope in difficult listening situations – in a restaurant, on the car phone, on the railway platform? This would be an indication of longer-lasting “traits” that help to distinguish two people. Again, we are interested in how we can elucidate the diversity of listening. Of course, all the traits that distinguish me from you, one person from the other, are potentially relevant here: first and foremost, our basic hearing ability in audiometric terms and, closely related to that, our biological age.

But, astonishingly, such differentiation of our listening performance can also be determined through traits of our personality. For example, emotionally unstable people tend to overestimate their own susceptibility to sonic interference; objectively, however, they tend to “perform” better than average in such situations, as we are currently showing in a large-scale study with over 1000 subjects [Wöstmann et al., *under review*]. The implications of such personality effects on our hearing experience should not be underestimated if we want to offer people individualized hearing and perceptual aids in old age and also if we want to take greater account of the relative relevance of hearing and vision for individuals in the future.

V. Addendum: The brain as metaphor or as ultimate cause?

Depending on one's own theoretical ideas (cf. Miller, 2010), we are always tempted and attracted by a biological-neuroscientific level of description alongside, underneath or above this very productive psychological one – that is, a way of looking at things that begins to describe and explain listening and, thus, auditory attention first and foremost as an achievement of the brain. This brings its own problems. First, it is hardly possible to get by without psychological and information-theoretical concepts (attention, filter, etc.) (Obleser & Kayser, 2019). Secondly, the pathologies of hearing and listening are in danger of being "biologized" particularly quickly – as if the psychological and often interpersonal phenomenon were already understood with a biological representation, no matter how crystal-clear. This applies, for example, to tinnitus or so-called verbal hallucinations, but also to the much-diagnosed "central hearing disorder" which is particularly common in children and adolescents and which could not be more unspecific (here, "central" simply means that it is not a disorder of the inner ear). Unquestionably, the first step for the individual researcher is to become aware of one's own metaphors and terminology when starting to describe and explain human hearing and listening in these different domains – e.g. am I concerned with the audiological-psychological "algorithm" of audiovisual integration or with its neurobiological "implementation"? (Cf. Marr, 1982.)

In summary, both the genuinely psychological description and explanation of listening and the neurobiological investigation of it in the past decades have provided us with a set of instruments with which we can continue to measure the relative and variable significance that our sense of hearing has in the orchestra of our sensory perception and our cognition – especially under changing conditions of communication and media mediation.

Translated from the German by Elizabeth Hormann

References

Mohsen Alavash, Sarah Tune & Jonas Obleser, "Modular reconfiguration of an auditory control brain network supports adaptive listening behavior", *Proceedings of the National Academy of Sciences of the United States of America* 116, 2 (2019, Jan 8) doi: 10.1073/pnas.1815321116

E. Colin Cherry, "Some experiments on the recognition of speech, with one and two ears", *Journal of the acoustical Society of America* 25, 5 (1953): 975-979.

Ian C. Fiebelkorn & Sabine Kastner, “A Rhythmic Theory of Attention”, *Trends in Cognitive Sciences* 23, 2 (2019 Feb): 87-101. doi: 10.1016/j.tics.2018.11.009

Karl Friston, “The free-energy principle: a unified brain theory?”, *Nature Reviews Neuroscience* 11, 2 (2010 Feb): 127-138. doi: 10.1038/nrn2787

William James, “Attention”, in *Principles of Psychology*, vol. 1 (New York: Henry Holt & Co., 1890/1927), 402-458.

David Marr, *Vision: a computational investigation into the human representation and processing of visual information* (New York: W. H. Freeman, 1982).

Harry McGurk & John MacDonald, “Hearing Lips and Seeing Voices”, *Nature* 264 (1976): 746-748. doi: 10.1038/264746a0

Gregory A. Miller, “Mistreating Psychology in the Decades of the Brain“, *Perspectives of Psychological Science* 5, 6 (2010, Nov): 716-743. doi: 10.1177/1745691610388774

Jonas Obleser & Christoph Kayser: “Neural Entrainment and Attentional Selection in the Listening Brain”, *Trends in Cognitive Sciences* 23, 11 (2019, Nov): 913-926. doi: 10.1016/j.tics.2019.08.004

John T. Serences & Sabine Kastner, “A Multi-level of Selective Attention”, in Anna C. Nobre & Sabine Kastner (eds.): *The Oxford Handbook of Attention* (New York: Oxford Univ. Press, 2014). doi: 10.1093/oxfordhb/9780199675111.013.022

Zachary M. Smith, Bertrand Delgutte & Andrew J. Oxenham, “Chimaeric sounds reveal dichotomies in auditory perception”, *Nature* 416 (2002, Mar 7): 87-90. doi: 10.1038/416087a

Rufin van Rullen & Christof Koch, “Is perception discrete or continuous?” *Trends in Cognitive Sciences* 7, 5 (2003, May): 207-213. doi: 10.1016/s1364-6613(03)00095-0

https://twitter.com/ThePrisonLawyer/status/1347626960529326080?s=20; accessed on Mar 24, 2021

Leonhard Waschke, Sarah Tune & Jonas Obleser: “Local cortical desynchronization and pupil-linked arousal differently shape brain states for optimal sensory performance”, *eLife* 8 (2019, Dec 10). doi: 10.7554/eLife.51501