



Research

Cite this article: VanRullen R, Zoefel B, Ilhan B. 2014 On the cyclic nature of perception in vision versus audition. *Phil. Trans. R. Soc. B* **369**: 20130214.
<http://dx.doi.org/10.1098/rstb.2013.0214>

One contribution of 13 to a Theme Issue 'Understanding perceptual awareness and its neural basis'.

Subject Areas:
cognition

Keywords:

perceptual cycles, discrete perception, brain rhythms, perceptual sampling, sampling rate, speech processing

Author for correspondence:

Rufin VanRullen
e-mail: rufin.vanrullen@cerco.ups-tlse.fr

On the cyclic nature of perception in vision versus audition

Rufin VanRullen^{1,2}, Benedikt Zoefel^{1,2} and Barkin Ilhan³

¹Université de Toulouse, France

²CNRS-CerCo, UMR 5549, CHU de Purpan, Toulouse, France

³Meram Medical Faculty, Konya University, Konya, Turkey

Does our perceptual awareness consist of a continuous stream, or a discrete sequence of perceptual cycles, possibly associated with the rhythmic structure of brain activity? This has been a long-standing question in neuroscience. We review recent psychophysical and electrophysiological studies indicating that part of our visual awareness proceeds in approximately 7–13 Hz cycles rather than continuously. On the other hand, experimental attempts at applying similar tools to demonstrate the discreteness of auditory awareness have been largely unsuccessful. We argue and demonstrate experimentally that visual and auditory perception are not equally affected by temporal subsampling of their respective input streams: video sequences remain intelligible at sampling rates of two to three frames per second, whereas audio inputs lose their fine temporal structure, and thus all significance, below 20–30 samples per second. This does not mean, however, that our auditory perception must proceed continuously. Instead, we propose that audition could still involve perceptual cycles, but the periodic sampling should happen only after the stage of auditory feature extraction. In addition, although visual perceptual cycles can follow one another at a spontaneous pace largely independent of the visual input, auditory cycles may need to sample the input stream more flexibly, by adapting to the temporal structure of the auditory inputs.

1. Introduction: perceptual awareness, a discrete process?

Our conscious perception of the world appears smooth and continuous. A moving object is not seen to disappear here and reappear there, but as successively occupying all positions in between. Similarly, the sound of an approaching car seems to steadily loom closer, without being interrupted by brief recurring moments of silence. And yet, it is not at all certain that the brain mechanisms supporting our sensory perception are themselves continuous; rather, visual and auditory perception may well be intrinsically discrete or cyclic [1–5]. We are not referring here to the discreteness of individual neuronal events (action potentials, synaptic release) but to the potentially discrete nature of perceptual experience itself. In this case, the continuity of our inner experience would merely be an illusion, a temporal 'filling-in' created by our brain to hide its recurring (albeit brief) moments of blindness and deafness, perhaps in the same way as the 'blind spot' of the retina is hidden from our consciousness by spatial filling-in mechanisms [6].

The notion of perceptual 'snapshots', 'moments' or 'cycles', once popular [1–4] but later discarded without ever being firmly disproved [7] has regained momentum in recent years owing to a number of converging experimental studies. As we shall see in the following sections, however, most of this new experimental evidence concerns the periodicity of visual perception; it has been more challenging, it seems, to uncover similar signatures of auditory 'snapshots'. After reviewing the corresponding findings, we consider the major differences between the visual and auditory modalities, both in terms of cerebral organization and information processing demands, that could justify a difference in temporal perceptual organization. Finally, we speculate that both systems, under the influence of neuronal oscillations, may, indeed, represent

sensory information as a sequence of perceptual cycles, but we will argue that the properties of these cycles must be vastly different for vision versus audition.

2. Perceptual cycles in vision

The notion of discrete perception was a prevalent idea after World War II and until at least the 1960s [1–4]. Even though some authors considered this discreteness to be an intrinsic property of all sensory modalities [1], most of the available experimental evidence came from studies of visual perception [3,4]. One possible reason for such a bias is the fact that the hypothesis of discrete perception was always strongly tied to the observation of large-amplitude ‘alpha’ (8–13 Hz) oscillations in electroencephalographic (EEG) recordings [8]. Because these alpha rhythms were found to be more heavily modulated by visual [9] than by auditory inputs, scientists naturally focused on the visual modality. Most of these previous studies have been reviewed elsewhere [5]; for various reasons, they failed to convince the larger scientific community, and the notion of discrete perception was gradually proscribed. In the past 10 years, however, significant experimental advances have occurred that somewhat restored the option of a discrete perceptual organization in the visual domain. These recent advances are reviewed in the following.

(a) The continuous wagon wheel illusion

In engineering, the term ‘aliasing’ refers to a potential artefact occurring when a signal is sampled by a discrete or periodic information processing system: if the sampling rate is lower than a critical limit (the Nyquist frequency), then the system’s representation of the signal is inaccurate. A special case of aliasing occurs when the signal is a moving periodic visual pattern, and the information processing system is taking temporally discrete samples; in this case, the resulting aliasing has been termed the ‘wagon wheel illusion’, and is vividly experienced as the pattern seems to move in the wrong direction. This illusion is most commonly observed in movies or on television, owing to the periodic sampling of video cameras (generally around 24 frames or snapshots per second). But it is also possible to experience a similar effect under continuous conditions of illumination, such as in daylight [10–12]. This must imply that aliasing can also take place within the visual system itself. Thus, this continuous version of the wagon wheel illusion (or c-WWI) has been taken as evidence that the visual system samples motion information periodically [11–14].

This ‘discrete’ interpretation of the c-WWI is supported by several arguments. First, the illusory reversed motion is perceived only over a specific range of stimulus temporal frequencies, and this range is compatible with a sampling rate (the number of ‘snapshots’ per second) of approximately 13 Hz [11–13]. Second, the critical frequency range for the c-WWI was found to be largely independent of the spatial frequency of the stimulus [12,13] and of the type of motion presented (e.g. rotation versus translation, first-order versus second-order motion) [12]. Such an aliasing determined exclusively by the temporal properties of the stimulus is precisely what would be expected from a discrete sampling perceptual system. Third, during the c-WWI, there is only one frequency band of the EEG oscillatory spectrum that changes significantly, right in the same frequency

range of approximately 13 Hz [15,16]. Altogether, these experimental findings converge towards the conclusion that the motion perception system (or at least part of it) samples information periodically, at a rate of approximately 13 samples per second.

Alternative interpretations of the c-WWI have also been put forward which do not rely on temporal subsampling and aliasing. Although all authors agree that the illusion is a bistable phenomenon, coming and going with stochastic dynamics as a result of a competition between neural signals supporting the veridical and the erroneous motion directions [17], most of the disagreement is now focused on the origin of the erroneous signals. While we assume that they arise from periodic sampling and aliasing, other authors have argued that they originate instead from spurious activation of low-level motion detectors [18,19] or from motion adaptation signals that would temporarily prevail over the veridical input [20,21]. We have argued, however, that this alternative account is incompatible with the available evidence. First, the c-WWI is maximal at around the same temporal frequency for first- and second-order motion patterns, whereas motion detectors in the brain have widely different temporal frequency response properties for the two types of motion [22]. Second, focused attention was found to be necessary for the c-WWI to occur [12]; furthermore, attention modulated not only the magnitude, but also the spatial extent and even the optimal temporal frequency of the c-WWI [23,24]. Although the absolute amount of motion adaptation could be assumed to vary with attentional load [25,26], there is no evidence to date that the frequency-tuning of motion adaptation (or of low-level motion detectors) can also be modified by attention. Third, motion adaptation can be strictly dissociated from the c-WWI: by varying stimulus contrast or eccentricity, it is possible to increase the amount of motion adaptation (as measured by both the static and the dynamic motion aftereffects) while decreasing the c-WWI, and vice versa [27]. Lastly, there is converging evidence that the neural correlates of the c-WWI primarily involve the right parietal lobe [15,28,29]; if the illusion was due to adaptation of low-level motion detectors, then its correlates would probably not be expected in such a high-level hierarchical region.

In conclusion, we believe that reversed motion signals in the c-WWI originate from attention-based motion perception systems that sample inputs periodically at approximately 13 Hz, thereby producing aliasing. At the same time, other motion perception systems (e.g. the low-level or ‘first-order’ system) continue to encode the veridical motion direction; it is the ensuing competition between these opposite signals that explains the bistability of the illusion.

(b) Ongoing electroencephalographic signatures of perceptual cycles

The c-WWI implies that a certain part of the visual sensory input (namely motion signals) can be sampled discretely or ‘periodically’. One might predict, therefore, that it would be possible to record neural signatures of this sampling process in the form of a brain signal that waxes and wanes with every sample. Neuronal oscillations in various frequency bands are a natural candidate of choice for such a signature. Recently, our group and others have tested this prediction by assessing the influence of the phase of ongoing EEG oscillations (even

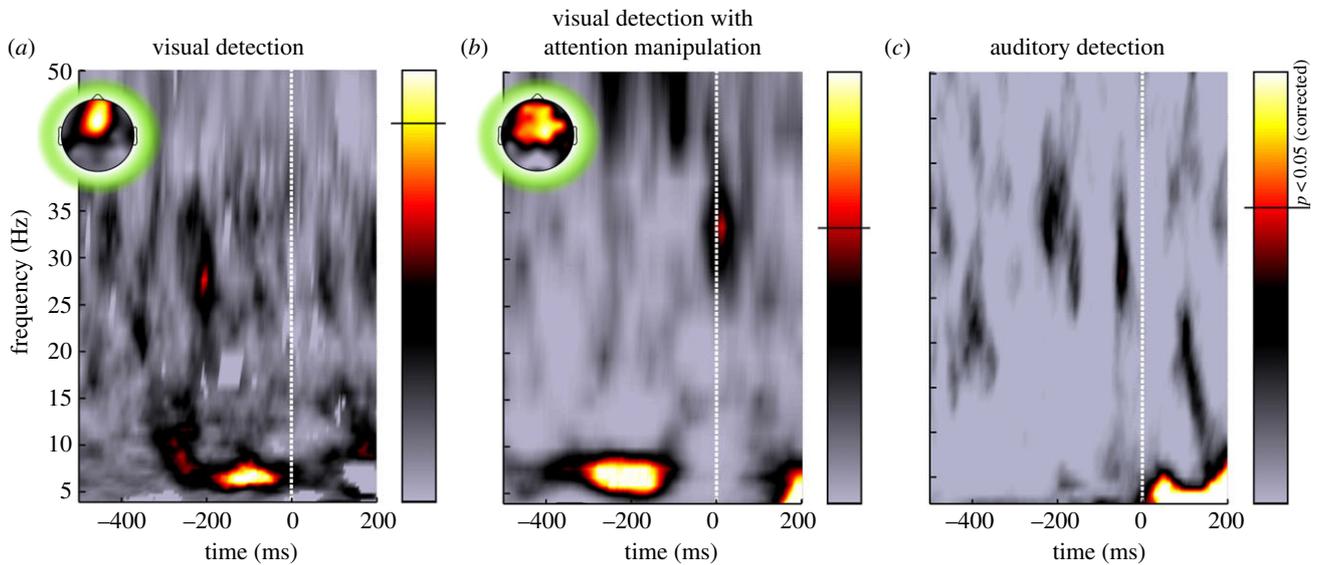


Figure 1. Pre-stimulus EEG signatures of visual, but not auditory perceptual cycles. The time–frequency maps represent the significance of ‘phase opposition’ between target-perceived and target-unperceived trials. This measure is determined by comparing the average phase-locking (or intertrial coherence, ITC) of each trial group with surrogate phase-locking values obtained over trial groups of the same size but randomly drawn among perceived and unperceived trials. A significant phase opposition at a particular time and frequency indicates that perceived and unperceived trials are associated with different phase values. (a) A pre-stimulus phase opposition was found at approximately 7 Hz in a visual experiment where subjects ($n = 12$) were free to attend to the target location (which was known in advance). The inset on the top-left represents the scalp topography of the pre-stimulus phase-opposition at this frequency, with a maximum around frontocentral electrodes. (b) The same result was replicated in a visual experiment with explicit attentional manipulation, but only for targets appearing on the attended side ($n = 13$). (c) However, no pre-stimulus phase opposition was found in an auditory experiment in which subjects ($n = 21$) were required to detect auditory clicks in a silent background. (No scalp topography is shown here as there were no significant pre-stimulus time–frequency points). (Online version in colour.)

before any stimulus is presented) on the subsequent perception of a visual stimulus (for a detailed review, see [30]).

In our first study [31], we presented dim flashes (6 ms long) in the visual periphery, with the luminance set around perceptual threshold. That is, only half of those flashes were perceived by the observers, while the other half remained unnoticed. The pre-stimulus phase locking on frontocentral electrodes was found to increase just before flash onset, for both the perceived and the unperceived trials (figure 1a). This effect occurred specifically for an EEG frequency band around 7 Hz, meaning that certain approximately 7 Hz pre-stimulus phase values facilitated the conscious perception of the flash, whereas other phase values impaired it. Indeed, when considering the phase of the 7 Hz band-pass-filtered EEG just before stimulus onset on each trial, we could predict the subsequent percept of the subject well above chance. In the same year, Mathewson *et al.* [32] also reported that the pre-stimulus phase of low-frequency oscillations (around 10 Hz) predicted the trial-by-trial perception of masked stimuli. Such a relationship between visual perception and the phase of spontaneous oscillations implies that visual inputs are not processed equally at all times, but periodically sampled by the visual system.

In our next study, we sought to determine the role of top-down attentional factors in this periodic sampling [33]. Because the target location was known in advance, we reasoned that subjects may have covertly attended to that location in order to improve their perceptual performance. Would ongoing oscillations still modulate target perception at an *unattended* location? In this new experiment, therefore, there were two possible target locations, and a central cue indicated before each trial the location at which subjects should pay attention. When the target appeared at that

attended location, everything happened exactly as in the previous experiment, and indeed, we confirmed our previous results in this condition, with a strong impact of approximately 7 Hz pre-stimulus EEG phase on target perception (figure 1b). When the target appeared on the other, unattended side, however, the phase of ongoing oscillations had no effect on perception (data not shown here). In other words, ongoing EEG phase was related to visual perception solely by the implication of attention. We thus hypothesized that attention samples visual information periodically, and that each approximately 7 Hz ongoing EEG cycle is the signature of a new attentional sample [33]. This conclusion is well in line with another body of recent experimental work that will be reviewed in §2(d).

We have also applied the same generic method, identifying pre-stimulus EEG phase opposition between the different outcomes of a given cognitive process, to perceptual tasks other than the mere detection of a peripheral flash. For example, we recently showed that the phase of ongoing EEG oscillations at approximately 10 Hz can also predict the perception of a transcranial magnetic stimulation (TMS) phosphene, i.e. an illusory visual percept that follows the administration of a TMS pulse [34]. Similarly, we showed that saccadic reaction times to a peripheral target differed for different pre-stimulus 10–15 Hz EEG oscillatory phases [35]. The likelihood of identifying a target in a difficult search array (a T among Ls) was also found to depend on pre-stimulus oscillatory phase, this time at a slower frequency of approximately 6 Hz [36].

All these studies together seem to imply that there is an ongoing succession of ‘good’ and ‘bad’ phases for visual perception and attention, i.e. that perception and attention are intrinsically periodic or cyclic phenomena. As such, these studies constitute a solid initial body of evidence for the

notion of discrete perception. It might be argued, however, that a proper demonstration of discrete perception should involve more than just a cyclic fluctuation of sensory excitability. A truly discrete system, just as in the epitomic example of the video camera, should also exhibit a periodicity in the fine-grained perception of time itself, a so-called temporal ‘framing’—meaning that two events separated by a given time interval would be perceived as occurring simultaneously or sequentially depending on whether they happened to fall within the same or distinct perceptual cycles. None of the experiments mentioned above can speak to this question, because they did not directly probe time perception. One such experiment on temporal framing was, in fact, published by Varela *et al.* [37] (see also [38]). They reported that the perception of two flashes separated by approximately 60–80 ms changed drastically as a function of the phase of the alpha rhythm (7–13 Hz) at which the first flash was presented; at one phase, they would be perceived as simultaneous, at the opposite phase as sequential. Unfortunately, this result has never been replicated, despite several attempts by our group and at least one other (D. Eagleman 2003, personal communication). Critically, however, one of our more recent experiments can also address this issue, albeit indirectly [39]. We examined the ‘flash-lag’ effect, a common illusion in which a steadily moving object is incorrectly perceived ahead of its true location at the moment of a flash [40]. The perceptual lag is generally accepted to reflect the time necessary for updating the conscious representation of the world after the ‘flash’ signal [40,41]. We showed that the trial-to-trial magnitude of this flash-lag effect systematically varied along with pre-stimulus 7–15 Hz EEG phase. That is, the oscillatory phase at (or just before) the moment of the flash determined whether an earlier or a later part of the ongoing motion sequence would be temporally grouped (or ‘framed’) with the flash. This may be the only solid evidence to date for a periodicity affecting not only sensory excitability, but also the fine-grained perception of time.

(c) Perceptual echoes

The various EEG experiments described in §2*b* indicate that ongoing brain oscillations create ‘perceptual cycles’ in which visual inputs are processed periodically. As a result, we were naturally led to ask the following questions. First, could these perceptual cycles be recorded not just before the time of stimulus presentation (i.e. in the ongoing EEG brain signals) but also afterwards, during stimulus processing itself (i.e. in the evoked EEG brain activity)? Second, for a visual event occurring at a particular instant, how many subsequent cycles would actually process the corresponding visual information? Do the perceptual cycles begin anew with each new sample, or do they also integrate the contents of past cycles, and if so, for how many successive cycles? We designed a simple experiment to answer both of these questions [42]. We presented a ‘white noise’ visual stimulus to our observers while recording their EEG activity. The stimulus was a static disc whose luminance varied randomly at each screen refresh. This random sequence of luminance intensity values had equal energy at all temporal frequencies (between 0 and 80 Hz, only limited by the 160 Hz refresh of the computer screen). We then cross-correlated the recorded EEG activity with the stimulus sequence on every trial, and averaged the results to obtain a cross-correlation function, describing the strength of

correlation between the stimulus and the brain response recorded after a certain lag, for all successive values of the lag. One might have expected this cross-correlation function to resemble a classic visual-evoked potential (VEP) [43], a sequence of positive and negative deflections lasting about 300–500 ms [44,45]. Instead, we found a much longer-lasting response in the cross-correlation functions, which took the form of an approximately 10 Hz oscillation that extended, in many subjects, for 10 or more successive cycles (figure 4*a*). This oscillatory cross-correlation response implies that visual events in the world are represented cyclically in the brain, and that this periodicity is also visible in post-stimulus EEG activity. Furthermore, it indicates that a given instant in the world is not merely represented at one instant (or in one ‘cycle’) in the brain, but in several successive cycles. Arguably, this property could provide a significant contribution to the apparent continuity of our subjective experience.

(d) Periodic attentional sampling

Many of the perceptual periodicities described in §2*a–c* are tightly linked to visual attention. For example, the temporal sampling causing the wagon wheel illusion in continuous light (c-WWI) only occurs when attention is focused on the moving pattern [12]. Similarly, the phase of ongoing EEG oscillations only modulates the probability of detection for *attended* stimuli [33]. The ongoing EEG phase can also predict the likelihood of detecting a target in a difficult search array [36], an archetypal attentional function. The 10 Hz perceptual echoes were also shown to be enhanced by focused attention [42]. In other words, ongoing perceptual cycles in the brain could be attentional by nature. Is attention a cyclic process? This question becomes particularly interesting when multiple attentional targets must be monitored: in this case, does covert attention periodically sample the targets, just like our gaze, often dubbed ‘overt’ attention, would? Or does the attentional system process all of the targets in parallel? This is a question that has been vastly debated in the past few decades [46,47]. We have recently argued that discrete versus continuous perception and sequential versus parallel attention are but two facets of the same debate [48]. The cornerstone of this theory is that attention is intrinsically periodic (figure 2): when a single attentional target is present, this periodicity is expressed as a sequence of successive discrete samples of the unique target; when multiple targets are present, this periodicity naturally provides attention with a means to scan the targets in a sequential manner.

There are many recent pieces of experimental evidence in support of this notion. The idea that rhythmic attentional sampling could occur not just in the presence of multiple potential targets (a classic form of ‘switching spotlight’ [49–51]), but also for a single attended object (a notion we called ‘blinking spotlight’) originated in a 2007 study in which we modelled the effect of set size on psychometric functions for target detection as a function of target duration [52]. To summarize, we contrasted different models of attention and found this ‘blinking spotlight’ to explain human performance better than either the ‘switching spotlight’ or the ‘parallel attention’ models. The intrinsic sampling rate of attention was estimated around 7 Hz (in agreement with several subsequent EEG experiments, such as those illustrated in figure 1).

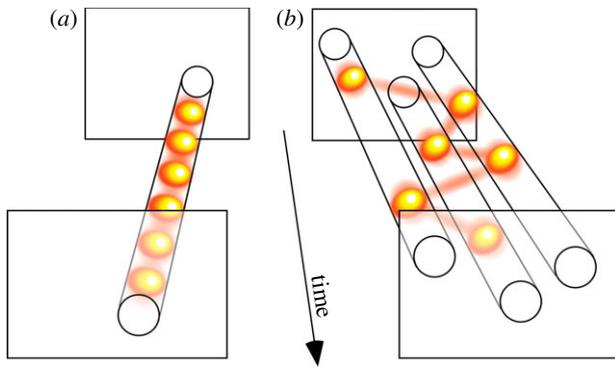


Figure 2. Discrete perception and sequential attention could reflect a unique periodic sampling mechanism. (a) A sensory process that samples a single visual input periodically illustrates the concept of discrete perception. (b) A sensory process that serially samples three simultaneously presented visual stimuli demonstrates the classic notion of a sequential or 'switching' attentional spotlight. Because many of our findings implicate attention in the periodic sampling processes displayed in panel (a), we proposed that both types of periodic operations (a,b) actually reflect a common oscillatory neuronal process. According to this view, the spotlight of attention is intrinsically rhythmic, which gives it both the ability to rapidly scan multiple objects (as in the classic 'switching' spotlight), and to discretely sample a single source. This dual behaviour is what we refer to as a 'blinking' spotlight. (The yellow balls linked by red lines illustrate successive attentional samples). (Online version in colour.)

Recently, Landau & Fries [53] used another psychophysical paradigm in which they drew attention using a salient cue at one of two possible target locations. The observers reported target detection (a contrast decrement in a continuously moving pattern) at either of the two locations. After the salient cueing event, detection performance was found to oscillate at both locations, but in counter-phase such that optimal performance at one location coincided with minimal performance at the other. In other words, it again seemed that attention periodically and sequentially sampled the two locations, with an intrinsic sampling rate of about 7–10 Hz.

We have also used the c-WWI effect to address this question [24]. We varied the set size (number of simultaneously presented moving wheels) and the wheel(s) rotation frequency while asking observers to report any occurrence of reversed motion. As previously (see §2*a*), we found that reversals were most likely to happen in a specific range of temporal frequencies. For a single target wheel, the effect was compatible with aliasing caused by attentional sampling at approximately 13 Hz, exactly as in our previous studies. But when set size increased, the effective sampling frequency systematically decreased. When four wheels were present, illusory reversals still happened, but they were now compatible with each wheel being sampled at only approximately 7 Hz. One interpretation, in line with the idea of a 'blinking spotlight', is that the successive attentional samples, instead of repeatedly sampling the same wheel, were now sequentially exploring the different wheels (or a subset of them); as a result, each wheel experienced aliasing at a lower frequency.

(e) Conclusion: discreteness in visual perception, attention and awareness

It is becoming more and more evident that, in the visual domain, neural oscillations in the 7–13 Hz range have

direct perceptual consequences that can be described as perceptual 'cycles'. This does not mean, of course, that higher-frequency oscillations, e.g. in the gamma range (30–80 Hz), do not influence perception, but these inherently more local oscillatory signals are less easily accessible to our EEG surface-recording methods. It is important to insist that it is not only sensory excitability that fluctuates cyclically at 7–13 Hz, but also higher-level perceptual representations involving visual attention, and possibly even visual awareness. There is, indeed, a tight relationship between these perceptual cycles and attentional processes, as reviewed in §2*d*. Attention is often considered as the gateway to consciousness [54,55], and it follows that if the gate opens periodically, the contents of awareness will also update periodically. Furthermore, we have described at least one instance in which the conscious perception of temporal simultaneity (i.e. which events in the world are experienced as a single 'group', a 'snapshot' or a mental 'frame') is constrained by the phase of ongoing oscillations [39]. This type of temporal framing is a hallmark of discrete perception, of the successive 'moments' of awareness [2].

3. Perceptual cycles in audition

After having reviewed the available evidence for discrete perception in vision, we now turn to the auditory system. It might be expected that the same experimental paradigms that helped uncover visual perceptual cycles could be similarly applied to audition to reveal its intrinsic discreteness. We may anticipate such auditory cycles to occur in the same frequency range as in vision (7–13 Hz), but this is not mandatory. In particular, because the frequency of visual cycles coincides roughly with the maximal range for steady-state visual-evoked responses (SSVEP) [56–58], one might predict that auditory cycles would occur instead around 40 Hz, which is the optimal frequency for auditory steady-state responses (ASSR) [59,60]. Another possibility, supported by certain theories of speech processing [61–65], could be that periodic auditory samples are taken at the same rate at which the relevant phonemic or syllabic events are expressed in normal speech, roughly between 2 and 8 Hz. Unfortunately, as we shall see, this straightforward approach of adapting our experimental paradigms to the auditory domain has not met with overwhelming success.

(a) No auditory wagon wheel illusion

In an initial attempt at directly translating the c-WWI paradigm (see §2*a*) to the auditory modality, we sought to measure the perceived motion direction of a spatially periodic sound source, such as a sound rotating around the listener through a circular array of speakers. By analogy to the illusion in the visual domain, we hoped to observe decreased perceptual performance, or even reversed motion perception, within a narrow range of temporal frequencies of the sound movement. Identifying this frequency of aliasing would then allow us to determine the intrinsic sampling frequency of the auditory system. It turned out, however, that such perceptual judgements of auditory sound motion can only be performed accurately at low temporal frequencies of sound movement, less than approximately 2–3 Hz (in agreement with previous reports [66,67]). If perceptual performance is already at chance at the hypothesized

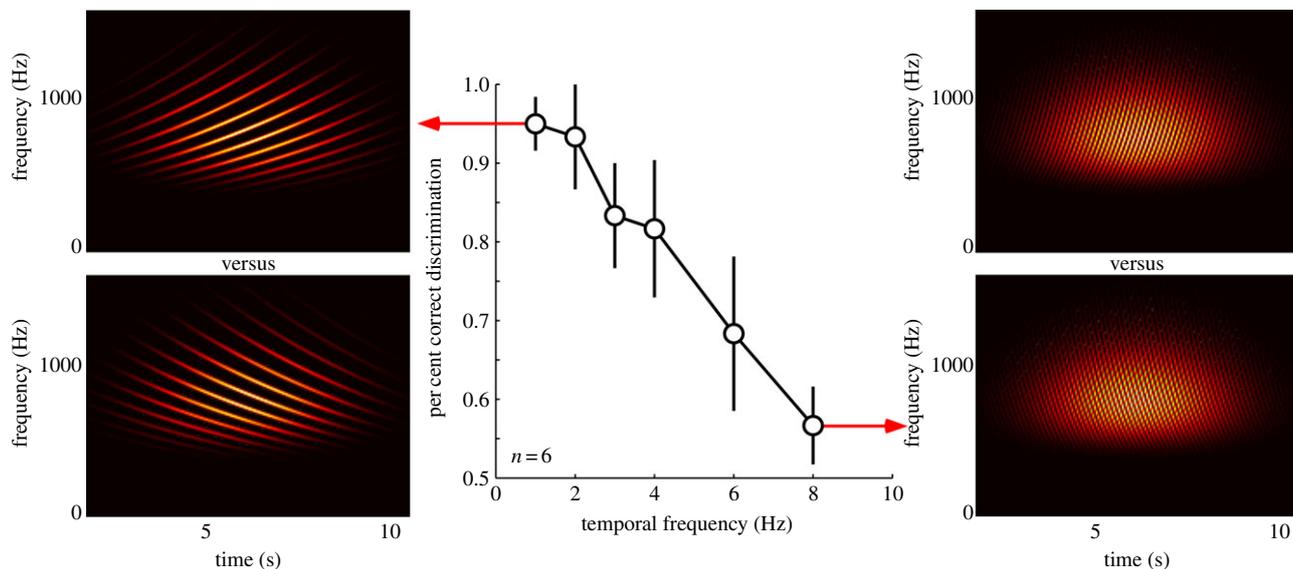


Figure 3. Auditory sensitivity for the direction of periodic sounds in the frequency domain. Shepard sequences [69] were created as a superposition of pure tone sweeps, increasing or decreasing in frequency over time. The sweep speed varied for different sequences: each tone increased or decreased logarithmically between the boundaries of audible space (set between 80 and 11025 Hz) over a fixed duration T ; a new tone was inserted into the sequence (and an old tone disappeared) every $T/40$, such that at every instant 40 sweeping tones were simultaneously present. The temporal frequency of this periodic sound motion, therefore, was defined as $TF = 40/T$. In different trials, this temporal frequency was varied between 1 and 8 Hz. Each sequence was 12 s long, with a Gaussian amplitude profile in both time and frequency space, to limit the perception of artefacts linked to sequence onset/offset and tone insertion/disappearance, respectively. Examples of upward and downward sequences for $TF = 1$ Hz and $TF = 8$ Hz are presented as spectrograms in the figure (colour map indicates stimulus energy at each time and frequency), and the corresponding sound files can be downloaded from <http://www.cerco.ups-tlse.fr/~rufin/audiovisual/>. Participants ($n = 6$) were presented with 10 sequences of each temporal frequency in randomized order, moving up or down in frequency space (randomly determined with 50% probability), and asked to report their perceived motion direction (up/down) by pressing arrow keys on the keyboard. To limit the possibility of relying on perceived pitch differences between upward and downward sequences, the frequency-domain Gaussian amplitude envelope (s.d. 0.25 log units) was centred at one of three frequencies (700, 800 or 900 Hz), randomly chosen for each trial. The direction judgements were only accurate up to 3–4 Hz (t -test against 0.5 = chance-level, $p < 0.05$), and deteriorated rapidly at higher temporal frequencies. This low-pass sensitivity function critically limits the possibility of measuring a c-WWI effect in the auditory domain. Error bars represent standard error of the mean across subjects. (Online version in colour.)

frequency of aliasing, then this aliasing will simply not be observed—whether the perceptual process relies on periodic sampling or not. In other words, the only conclusion that can be drawn from this attempt is that, if auditory perceptual cycles exist, then they must occur at a rate faster than 3 Hz—hardly a revealing conclusion.

We then reasoned that sound frequency or ‘pitch’, rather than spatial position, may be the proper equivalent to the spatial location of visual objects. Indeed, the ‘retinotopic’ neuronal organization of early visual cortex is not found in the auditory system, where neurons are instead organized in a ‘tonotopic’ manner [68]. Thus, we designed periodic stimuli that moved in particular directions in the frequency domain—so-called Shepard or Risset sequences [69]. Again, we were disappointed to find that the direction of these periodic frequency sweeps could not be reliably identified when the temporal frequency of presentation was increased beyond 3–4 Hz (figure 3; no temporal aliasing is visible, i.e. no performance below chance or local minimum in performance).

In sum, although temporal aliasing (as measured in the c-WWI) is, in principle, a choice paradigm to probe the rhythms of perception, our attempts so far at applying this technique to the auditory domain have been foiled by the strict temporal limits of auditory perception. Of course, the auditory system is widely regarded as a temporally precise one, but this precision observed for specific auditory features (discrimination of nearby pitch frequencies, interaural time delays) does not extend to periodic sound motion, either in the spatial or in the frequency domains. This limitation

precludes using the wagon wheel phenomenon to determine the sampling rate of audition or possibly, the absence of discrete auditory sampling. What we can safely conclude is that, if discrete sampling exists in audition, then it must be at a sampling rate above 3–4 Hz (if aliasing occurred at frequencies below 3 Hz, then it would have been possible for our subjects to perceive it, and they would have systematically reported reversed motion).

(b) No ongoing electroencephalographic signatures of auditory perceptual cycles

One major and undisputable piece of evidence in favour of ongoing perceptual cycles in vision is the finding that the conscious detection of a flash at luminance threshold fluctuates along with the phase of ongoing EEG oscillations (see §2*b*). Similarly, a dependence of auditory detection on ongoing EEG phase would indicate the existence of ongoing auditory perceptual cycles. We attempted to measure this relation by presenting threshold-intensity ‘clicks’ (0.5 ms square wave pulses) in a silent environment, and asking participants ($N = 21$) to report their perception via button-presses. Upon applying the same time-frequency phase opposition analysis techniques (figure 1*a,b*) as in our previous visual experiments [30,31,33], we were unable to reveal any systematic relationship between pre-stimulus phase and auditory perception in any frequency band (figure 1*c*). A similar negative report was independently published by Zoefel & Heil [70].

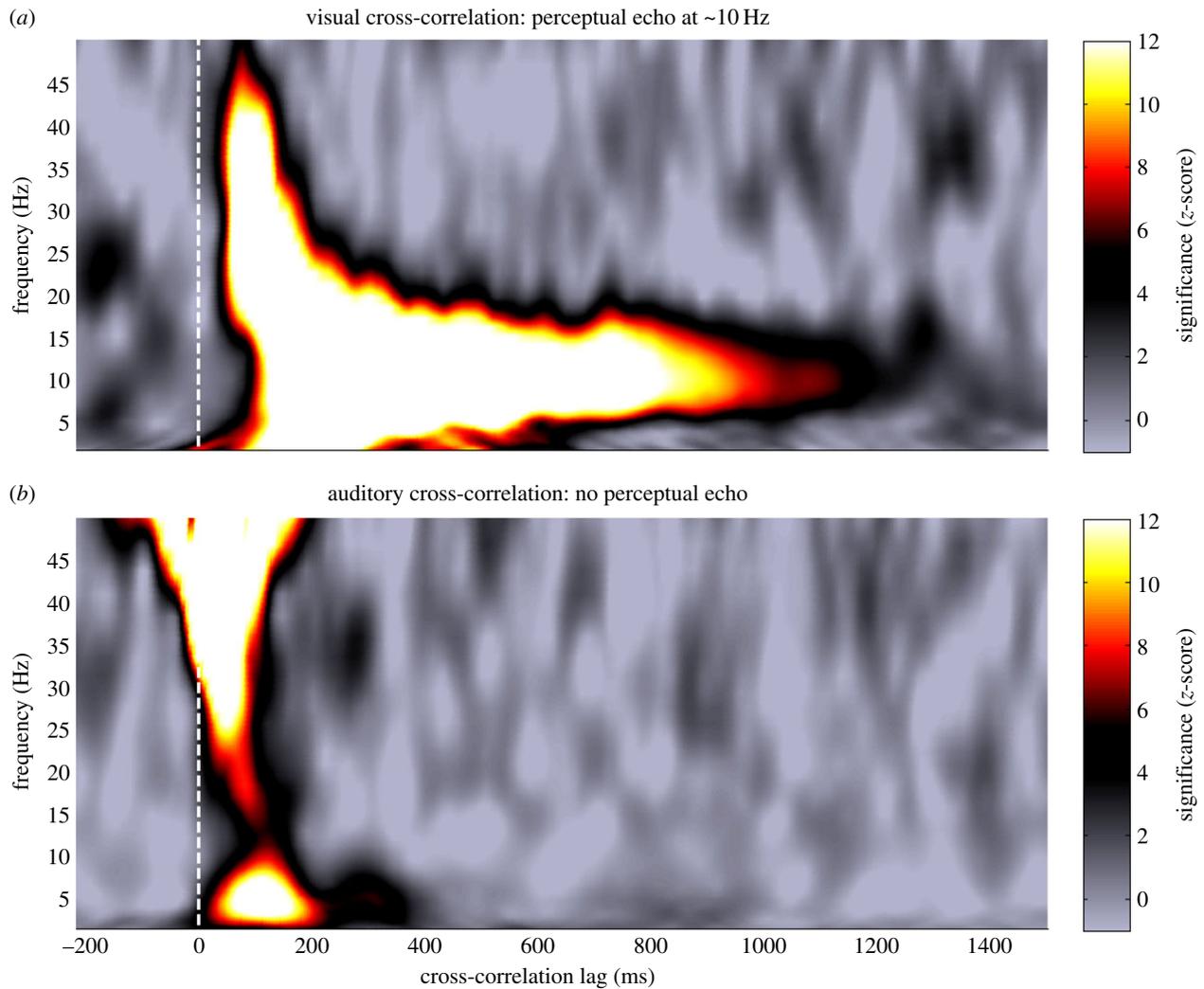


Figure 4. Perceptual echoes in the visual, but not in the auditory modality. Each panel is a time–frequency representation of the cross-correlation function between a white noise stimulus sequence and the simultaneously recorded EEG response. The cross-correlation is computed for several lags between the stimulus and EEG signals; then, a time–frequency transform is applied separately for each subject ($n = 12$); the grand-average results are expressed as a z-score (comparison against surrogate cross-correlation functions obtained by randomizing stimulus–EEG pairings). (a) When the white noise stimulus sequence reflects the changing luminance of a disc in the visual field, after a transient broadband response for time lags below 200 ms, a long-lasting reverberation (up to lags of approx. 1 s) is observed, peaking at approx. 10 Hz. (b) When the white noise stimulus sequence encodes the changing loudness of an auditory pure tone (1000 Hz carrier frequency), the transient broadband response is present, but no subsequent reverberation is observed in any frequency range. The increasing width of the transient response above 30 Hz is likely due to the auditory middle latency response (MLR) [90], a short-lived auditory potential (< 50 ms) which appears smeared in time owing to our wavelet time–frequency transform (using an eight-cycle window length at 50 Hz). The same 12 subjects participated in the visual and auditory experiments. (Online version in colour.)

It appears that auditory perceptual cycles, if they exist, cannot be detected with the very same experimental method that has successfully and repeatedly allowed us to reveal periodicities in visual perception. One critical aspect of this method was the presentation of auditory targets in a silent environment—the auditory equivalent of a flash in the dark. In fact, Ng and colleagues recently reported that auditory perception does vary with the phase of 2–6 Hz (theta-band) EEG oscillations when the target sound is embedded in an ongoing sequence made up of several superimposed naturalistic background noises [71]. However, the use of background noise in this study also implies that the relevant EEG oscillations cannot be considered as spontaneous or ongoing signals any more, but are instead driven or entrained by the background noise [72,73].

Phase entrainment to auditory streams has been demonstrated in many previous studies using rhythmic

background sounds at delta (1–4 Hz) and theta (2–8 Hz) frequencies, and auditory detection performance was found to covary with the entrained oscillatory phase [74–77]. This phase entrainment mechanism is thought to serve a critical role in speech perception [64,78–85] by aligning the optimal oscillatory phase to the peaks of the speech envelope (which also recur at a frequency roughly between 2 and 6 Hz), and thereby enhancing speech intelligibility [65,78,86,87]. However, it is difficult in such entrainment studies (even the one by Ng *et al.* [71] in which the entraining background noise contained energy in several frequency bands, including theta), to tease apart the contribution of low-level physical differences in the entraining sound to the perceptual changes recorded at different theta phases. Because the EEG is entrained by (or ‘phase-locked’ to) the background stimulus [88], different EEG phases directly correspond to different moments in the background entraining sound, with

systematic differences in auditory properties (such as loudness and pitch); in turn, these varying physical properties can conceivably affect target detection probability (e.g. through masking or contextual enhancement phenomena). In this way, a rhythmic background sound can both entrain EEG oscillations, and modulate auditory detection in a periodic fashion. Yet, the perceptual modulation in this case is not *intrinsically* periodic: should the stimulus amplitude profile resemble, say, the outline of Mount Everest or the New York City skyline, so would the listener's perceptual performance. In other words, the existing evidence so far is insufficient to decide whether the frequently observed theta-band periodic fluctuations of auditory performance reflect an intrinsic periodicity of the auditory system (i.e. true perceptual cycles) or an intrinsic periodicity of the auditory environment (or both).

To conclude, contrary to what we have observed in the visual modality, it would appear that the presence of an entraining (and ideally, rhythmic) auditory background stimulus may be a necessary condition to observe rhythmic fluctuations in auditory perception [73,76,89]. Even then, owing to the possibility of low-level confounds, it is not yet evident that such entrained rhythmic fluctuations can be considered as a signature of 'entrained' auditory perceptual cycles. It is likely, on the other hand, that purely ongoing or spontaneous oscillations (i.e. those recorded in silence) do not reflect an ongoing auditory sampling process, as they do in vision.

(c) No auditory perceptual echoes

The cross-correlation paradigm that allowed us to reveal perceptual echoes in vision (§2c) could prove a useful tool to test the hypothesis of 'entrained' (in opposition to 'ongoing') auditory perceptual cycles. Indeed, this paradigm is designed to reveal the resonance properties of a sensory system, that is, whether it presents a frequency-specific response (an 'echo', which is also a form of phase entrainment) during a white noise stimulation sequence. In the visual system, this echo was found around 10 Hz, and lasted for up to 10 cycles (figure 4a). If one assumes that auditory perceptual cycles exist, but are only active when they can be entrained by a background sound (an assumption suggested by the data reviewed in §3b), then they may be expected to show up as an auditory echo in this cross-correlation paradigm. More precisely, one might predict observing a resonance in the theta-frequency range, in accordance with the numerous theta-phase entrainment results described in §3b (and in particular the strong theta-frequency periodicity of human speech signals and human speech processing mechanisms). Another (non-exclusive) hypothesis could be that auditory echoes occur in the gamma-frequency band, around 40 Hz: indeed, while alpha (approx. 10 Hz) is the optimal visual stimulation frequency to produce an SSVEP [56–58], gamma (approx. 40 Hz) is the optimal frequency for ASSR [59,60]. A direct auditory equivalent to our approximately 10 Hz visual echoes could thus also be expected around 40 Hz.

Unfortunately, no significant auditory perceptual echo was detected in our experiments [91], either in the theta nor in the gamma range, or in any other frequency band (figure 4b). While definite, this absence still does not disprove the existence of perceptual cycles in the auditory system for at least two reasons. First, although echoes were associated with perceptual cycles in vision, this association is not mandatory:

a reverberation and integration of sensory information over several cycles is likely to be detrimental to auditory perception, so audition may instead rely on cycles that are more temporally independent (i.e. 'short-lived' echoes). In this case, figure 4b (and much of the existing literature [62,64,80]) suggests that the cycles may occur in the theta (2–8 Hz) and/or gamma (30–80 Hz) frequency ranges. Second, the absence of long-lasting auditory echoes in our experiment merely indicates that perceptual sampling and reverberation do not affect the processing of auditory loudness (the sensory feature that varied in our white noise sequences), even though they affect the visual equivalent, luminance perception. It is still possible, however, that perceptual sampling and reverberation could involve higher hierarchical levels of representation, after the extraction of basic auditory features. In accordance with this idea, oscillations have been repeatedly shown to contribute to speech perception by temporally framing the input stream according to the speech envelope [64,65,78–87]. This suggests that auditory echoes, absent with low-level stimuli such as amplitude-modulated pure tones, may still be observed with stimuli having more complex semantic content, such as speech. In that case, we predict that they should be visible around theta or gamma frequencies.

4. Different sensory inputs, different rhythmic sampling strategies

So far, all of the experimental paradigms that have succeeded in demonstrating visual perceptual cycles have also failed at revealing the auditory equivalent. Short of embracing the conclusion that perceptual cycles simply do not exist in the auditory domain, we must contemplate the possibility that these cycles could be implemented in very different ways in the two systems and thus may not be responsive to the same experimental approaches. In particular, it might prove useful to consider the different computational requirements with which each sensory modality is faced in terms of statistical properties and temporal structure of their respective sensory inputs, as well as their respective anatomical and functional architectures. This could help us explain why a processing strategy that is efficient for visual inputs may not be directly applicable to auditory inputs. In particular, we suggested above (§3c) that in audition, contrary to vision, perceptual sampling and reverberation could be restricted to higher hierarchical levels of representation, after the extraction of basic auditory features. There are two arguments to support this hypothesis: first, directly subsampling an auditory input stream (after conversion to the wavelet domain) has much more devastating consequences than the equivalent temporal subsampling of a visual input stream; second, a great deal of auditory feature extraction takes place subcortically, whereas visual processing is predominantly a cortical phenomenon. These arguments are developed in the following sections.

(a) Perceptual effects of visual versus auditory subsampling

In the visual environment, important events and changes tend to occur on a relatively slow time-scale. If one were to take two pictures of the same scene, separated by 150 ms

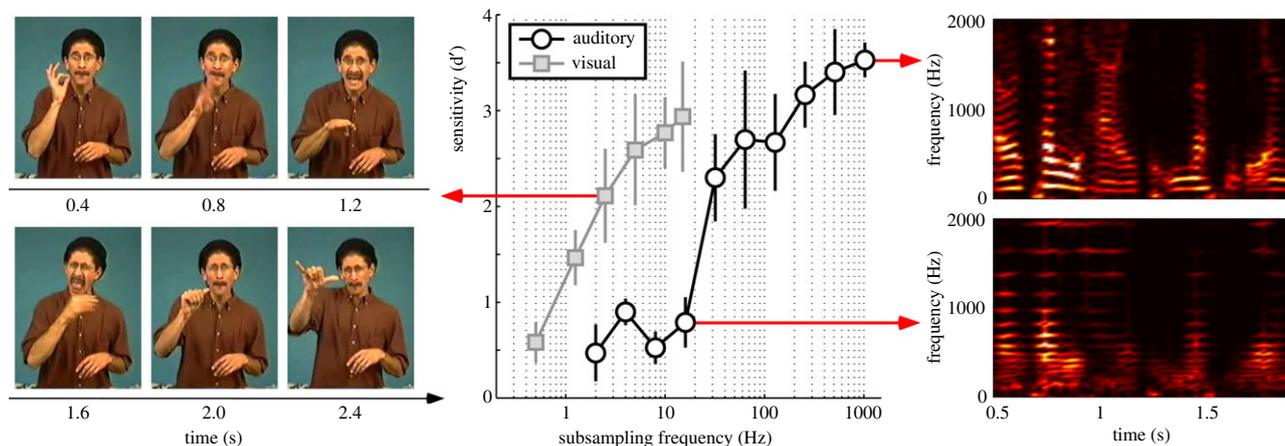


Figure 5. Auditory and visual vulnerability to input stream subsampling. The same participants ($n = 4$) watched and listened to sequences of 3 s long video and audio snippets (respectively) in different blocks while performing a two-back task (responding to a repeat of the penultimate snippet). Video snippets depicted an actor reading a children's book in American sign language, whereas audio snippets were recordings of a speaker reading an English literary classic. In each block, all snippets were temporally subsampled to the same frequency. A 2 s long excerpt from a video snippet at 2.5 frames per second is illustrated on the left, and the spectrograms of a single 2 s long excerpt from an audio stream subsampled at 1024 Hz (top) and 16 Hz (bottom) are shown on the right. The corresponding video/sound files can be downloaded from <http://www.cerco.ups-tlse.fr/~rufin/audiovisual/>. The two-back recognition task performance is expressed in terms of sensitivity (d' , corresponding to the z-scored difference between hit rates—correctly detecting a two-back repeat—and false alarm rates—incorrectly reporting a two-back repeat). It is an order of magnitude more robust to temporal subsampling for vision than for audition. Error bars represent standard error of the mean across subjects. (Online version in colour.)

(about one-seventh of a second), most if not all of the scene would likely remain unchanged between them. Movement on a biological time-scale (e.g. human actions and displays of emotions, the displacement of preys or predators) will result in only minor differences between the two pictures; furthermore, these inconsistencies can easily be recovered by temporal 'interpolation' (and indeed, the so-called apparent motion mechanisms in the brain seem to excel at this task [92]). Only the rapid movement of a spatially periodic stimulus (such as a wheel) could conceivably create a difficult 'correspondence problem' between the two images [93], but this is admittedly a rare situation (this situation describes, in fact, the temporal aliasing discussed in §2*a*). In other words, our visual system may be fairly robust to temporal subsampling of the visual environment. By contrast, auditory stimuli are defined mainly as temporal fluctuations: vocal or musical pitch, speech phoneme distinction or speech recognition all require processing fine-grained temporal information in different frequency ranges. This has moved certain authors to propose that the time dimension in audition could be equivalent to the spatial dimension in vision [94]. A periodic sampling or sensory reverberation of the auditory input stream could therefore dramatically alter signal intelligibility.

There are a number of existing studies reporting subjective judgements of video quality at different sampling rates [95,96], converging to the conclusion that frame rates above approximately 5 Hz are generally deemed acceptable. To the best of our knowledge, there is, however, no equivalent data on the perceptual effects of temporally subsampling the auditory input stream, and no direct comparison of the two modalities using the same task in the same subjects. We therefore implemented such a comparison in a new experiment (figure 5). Our comparison approach was voluntarily naive. We subsampled the visual and auditory inputs in a representation space roughly equivalent to the first sensory stage of each system: the retina (with the entire image

representing a subsampling 'frame') and the cochlea (with the instantaneous complex frequency spectrum resulting from a wavelet decomposition of the audio signal as a subsampling 'frame'), respectively. That is, for both sensory systems, we evaluated the consequences of the most severe possible temporal subsampling strategy, by subsampling the very input to the system; then we simply asked 'will the system be able to cope?' Of course, a positive answer does not imply that perceptual cycles actually occur at this frequency, but a negative answer casts serious doubt on this idea. In addition, any difference in sensitivity between the two sensory modalities can inform us about viable strategies for each system.

We hasten to mention that past studies have investigated the influence of temporal distortions on auditory perception and more particularly on speech processing [63,65,81,97–103], converging on the notion that audition can withhold temporal degrading of speech envelopes down to 16 Hz or even lower (approx. 4 Hz in [99]). But none of the distortion methods used was equivalent to a strict temporal subsampling of audio inputs. For example, the now classic 'Shannon' method [98] consists of low-pass filtering the audio signal envelope, and does it independently for several separate spectral bands.

In our experiment, one original 10 min video and one original 10 min audio sequence were used as the primary stimuli. The audio sequence was a 8000 Hz recording of a male native English speaker reading aloud an English literary classic. The video sequence was a 30 frames s^{-1} (silent) recording of a male actor reading a children's book in American sign language, shot from a static camera angle. Both audio and video recordings were cut into 3 s long 'snippets'. The snippets were contiguous excerpts that did not take into account the structure of the story. Although some snippets were certainly more informative than others, before temporal subsampling they were all intelligible or visually distinct (the participants had no prior experience with

American sign language). These 200 snippets were presented (separated by 1 s blank intervals) in a randomized order to four human observers and listeners who were instructed to perform a two-back task: indicate by a button press any snippet that matched the one presented two snippets ago. These two-back repeats occurred randomly with a probability of 33%.

Audio and video snippets were presented in separate blocks of 30 snippets, and in each block a different temporal subsampling was applied. For video subsampling at frequency TF, we selected a subset of frames (one frame every $30/TF$, rounded to the nearest frame), and simply played the videos with a frame rate set to TF. For every snippet and subsampling frequency, two subsampled versions were created by starting the frame subset selection either on the first frame, or on the nearest frame to $1 + 30/TF/2$. Whenever a two-back repeat occurred in the sequence, it was always between distinct subsampled versions (this was done to prevent the use of static information for recognition). For audio subsampling at frequency TF, we first converted the snippets into the wavelet domain to approximate cochlear transduction (continuous Morlet wavelet transform of order 6). Discrete samples were taken every $8000/TF$ point, and all points between the samples were replaced with a linear interpolation of the two surrounding samples. Both amplitude and phase of the complex wavelet coefficients were interpolated to avoid artefacts created by discrete phase transitions. As for video subsampling, we created two subsampled versions of each audio snippet by starting the samples on the first data point, or the nearest point to $1 + 8000/TF/2$. Finally, we converted the signals back to the time domain via the inverse wavelet transform.

As expected, we found a dramatic difference between the two modalities' sensitivity to temporal subsampling (figure 5). While visual performance only started to deteriorate below 2.5 frames per second, auditory performance suffered for all subsampling frequencies below 32 Hz. That is, audition was about an order of magnitude more vulnerable to this subsampling than vision.

The observed difference in temporal robustness may explain, in part, why the auditory system does not sample incoming information as the visual system does. In the visual system, we have suggested that ongoing sampling induced by brain oscillations could take place at frequencies between 7 and 13 Hz (§2). As can be appreciated from figure 5, little information is lost by directly subsampling visual inputs in this frequency range. On the other hand, directly subsampling the auditory inputs in the same frequency range has dramatic consequences: the fine temporal structure is irremediably lost, and the signals cannot be recovered (even through temporal interpolation, which was an integral part of our auditory subsampling procedure in the wavelet domain). This may be an argument for the notion that auditory sampling involves higher oscillatory frequencies, for example in the gamma range [59,62,80,104–106]. Yet our results do not imply that brain oscillations at lower frequencies have no bearing on auditory perception. As mentioned before, there are still two possible (and non-exclusive) oscillatory sampling strategies involving lower frequencies that could remain compatible with these data: first, by sampling auditory representations not in an 'ongoing' manner (a regular succession of samples, blind to the temporal structure of the inputs) but in a more flexible manner, 'entrained' by the temporal structure;

second, by sampling auditory representations not at the input level (e.g. cochlea or subcortical nuclei) but at a higher hierarchical level (e.g. auditory cortex).

In an attempt to address the former possibility, we repeated the above auditory subsampling experiment, this time comparing two modes of audio input subsampling: ongoing or 'blind' subsampling, as before, and entrained or 'flexible' subsampling. To create these 'flexible' subsampling stimuli, we first extracted the speech envelope of each snippet (weighted average of instantaneous signal energy across frequencies weighted by the average human cochlear sensitivity). Instead of selecting regular sampling points throughout the snippet ('blind' sampling), we distributed the same number of sampling points at the peaks and troughs of the 2–8 Hz band-pass-filtered speech envelope (starting with the highest peak and its immediately preceding trough; adding peak/trough pairs in decreasing order of peak amplitude; in case more sampling points were available than the number of peaks and troughs in the speech envelope, the remaining points were assigned so as to minimize the maximal sampling interval duration). In sum, this flexible subsampling kept the same average sampling rate as for blind sampling, but concentrated the samples at those moments where phonetic information was maximal. Yet we found no significant difference in the sensitivity of human listeners ($n = 7$) between the 'blind' and the 'flexible' subsampling of the input stream at frequencies between 8 and 64 Hz (two-way ANOVA with factors 'frequency' = (8,16,32,64 Hz) and 'sampling type' = [blind, flexible]; main effect of frequency $F_{3,48} = 17.45$, $p < 0.0001$, no main effect of sampling type or interaction, $p > 0.5$; data not shown). That is, audition remains an order of magnitude more vulnerable to temporal subsampling of its inputs than vision, even for a 'flexible' auditory subsampling. This finding definitely rules out the possibility that sampling at lower frequencies (less than 30 Hz) could occur early in auditory processing, since neither ongoing ('blind') nor entrained ('flexible') subsampling applied directly to the input stream would leave enough temporal information for further processing. In addition, it is worth noting that early subcortical auditory structures can display exquisite temporal resolution (greater than 100 Hz) that seems incompatible with temporal subsampling [107].

The last remaining option to rescue the notion of auditory perceptual cycles is, therefore, that they could sample auditory representations at a higher hierarchical level, after the stage of auditory feature extraction: such representations are more stable temporally, and would suffer less from a moderate loss of temporal resolution. This strategy is, in fact, the one used in modern speech compression techniques or 'vocoders' (e.g. LPC, MELP or CELP [108]) that extract phonetic features from high temporal resolution signals, but can then transmit the features in (lower resolution) temporally discrete packets or 'frames'. In future work, it may be interesting to apply temporal subsampling (either 'blind' or 'flexible' subsampling) to the output of one of these vocoders and test human auditory recognition in the same way as above: we predict that the auditory system may prove significantly more robust to this subsampling of higher-level representations.

(b) Differences in hierarchical organization

As mentioned previously, there are important architectural differences between the auditory and visual processing

hierarchies. Without going too deep into anatomical details, the most relevant discrepancy for our purposes can be summarized as follows: visual perception (even for salient low-level features such as luminance and spatial localization) depends, in great part, on cortical activity, whereas auditory stimuli reach primary auditory cortex after an already extensive processing by subcortical structures [109,110].

Consequently, applying an *architecturally* similar perceptual sampling strategy in the two systems (perceptual cycles that sample sensory representations at a similar *cortical* level, possibly under the influence of attention) could then have very different *functional* consequences, compatible with what we have observed experimentally. Apparently simple visual tasks (e.g. flash detection in the dark; figure 1) would suffer periodic fluctuations in performance, but equivalent low-level auditory tasks (e.g. click detection in silence) would appear continuous, because their outcome can be determined on the basis of subcortical representations, prior to any perceptual sampling. Perceptual cycles would only be observed with higher-level auditory stimuli such as music or speech that are not differentially processed at a subcortical level and thus require cortical activation for efficient discrimination. In agreement with this idea, a recent study demonstrated that arbitrary white noise auditory stimuli could elicit theta-band phase entrainment, but only after sufficient exposure, presumably turning the meaningless patterns into meaningful auditory objects [111]. Note finally that this reasoning remains compatible with the postulated role for attention in perceptual cycles (§2*d*), because attention is primarily a cortical function in both visual and auditory modalities [112–114]. Until direct anatomo-functional evidence is uncovered, we prefer not to speculate on whether primary sensory areas (both auditory and visual) participate or not in this ‘high-level’ periodic perceptual sampling.

(c) Conclusion: ongoing visual attentional cycles, entrained auditory attentional cycles?

To recapitulate, we can now critically evaluate the possible existence of perceptual cycles in the two modalities on the basis of experimental evidence reviewed in previous sections. We organize this evidence for perceptual cycles along two dimensions of interest, that is (i) whether they sample hierarchically ‘early’ or ‘late’ representations, and (ii) whether they

sample in an ‘ongoing’ (blind, stimulus-independent) manner or in an ‘entrained’ (flexible, stimulus-dependent) manner.

Visual perception samples sensory representations at approximately 7–13 Hz (§2). Vision could afford to sample at a hierarchically early level, because it is robust to input subsampling down to at least 5 Hz (§4*a*). But it probably does not. Indeed, we have seen that visual perceptual cycles are mainly an attentional phenomenon (§2*d*) and this is more in line with a high-level (or at least a cortical) sampling (§4*b*). Vision samples in an ongoing manner, mostly blind to the stimulus content (§2*b* and §4*a*). Yet it also has the capability to entrain to and resonate with stimuli that contain an appropriate rhythmic structure (§2*c*; [56,57]). The periodicity of visual perception, in sum, is best described as an ongoing series of attentional cycles at approximately 7–13 Hz.

Audition does not sample at a hierarchically early level in either an ongoing way (§3*b* and §4*a*) or an entrained (flexible) way (§3*c* and §4*a*). If it does sample, then it must be at a higher level, that is, a *cortical* level (§4*b*). Many reports of attentional control of phase entrainment [74,75,115] suggest that the sampling may also be attentional, as in the visual system. Can this high-level sampling be an ongoing process as in vision, or must it be entrained by the temporal structure of auditory inputs? The lack of ongoing EEG phase influence in audition (§3*b*) as well as the finding that phase entrainment strongly facilitates intelligibility [65,78,86] compels us to favour the latter alternative, just like other authors have recently argued [73,76,89].

To sum up, *if* perceptual cycles exist in audition, then they must be a relatively high-level or attentional phenomenon (as in vision), and they must proceed by stimulus entrainment (contrary to vision). Based on numerous studies of rhythmic entrainment and speech processing, we believe that the cycles are most likely to be observed in the theta-frequency range (though gamma-frequency sampling cannot be categorically ruled out). But the big ‘if’ lingers. Definite evidence for auditory perceptual cycles is still lacking.

Acknowledgements. We are grateful to Daniel Pressnitzer and Leila Reddy for providing useful comments on the manuscript.

Funding statement. This work was supported by a EURYI Award to R.V. and a Studienstiftung des deutschen Volkes scholarship to B.Z.

References

- Pitts W, McCulloch WS. 1947 How we know universals: the perception of auditory and visual forms. *Bull. Math. Biophys.* **9**, 127–147. (doi:10.1007/BF02478291)
- Stroud JM. 1956 The fine structure of psychological time. In *Information theory in psychology* (ed. H Quastler), pp. 174–205. Chicago, IL: Free Press.
- Harter MR. 1967 Excitability cycles and cortical scanning: a review of two hypotheses of central intermittency in perception. *Psychol. Bull.* **68**, 47–58. (doi:10.1037/h0024725)
- Allport DA. 1968 Phenomenal simultaneity and the perceptual moment hypothesis. *Br. J. Psychol.* **59**, 395–406. (doi:10.1111/j.2044-8295.1968.tb01154.x)
- VanRullen R, Koch C. 2003 Is perception discrete or continuous? *Trends Cogn. Sci.* **7**, 207–213. (doi:10.1016/S1364-6613(03)00095-0)
- Durgin FH, Tripathy SP, Levi DM. 1995 On the filling in of the visual blind spot: some rules of thumb. *Perception* **24**, 827–840. (doi:10.1068/p240827)
- Di Lollo V, Wilson AE. 1978 Iconic persistence and perceptual moment as determinants of temporal integration in vision. *Vision Res.* **18**, 1607–1610. (doi:10.1016/0042-6989(78)90251-1)
- Berger H. 1929 Über das Elektroenkephalogramm des Menschen. *Archiv. Psychiatr. Nervenkrankheiten* **87**, 527–570. (doi:10.1007/BF01797193)
- Chapman RM, Shelburne Jr SA, Bragdon HR. 1970 EEG alpha activity influenced by visual input and not by eye position. *Electroencephalogr. Clin. Neurophysiol.* **28**, 183–189. (doi:10.1016/0013-4694(70)90186-0)
- Schouten JF. 1967 Subjective stroboscopy and a model of visual movement detectors. In *Models for the perception of speech and visual form* (ed. I Wathen-Dunn), pp. 44–45. Cambridge, MA: MIT Press.

11. Purves D, Paydarfar JA, Andrews TJ. 1996 The wagon wheel illusion in movies and reality. *Proc. Natl Acad. Sci. USA* **93**, 3693–3697. (doi:10.1073/pnas.93.8.3693)
12. VanRullen R, Reddy L, Koch C. 2005 Attention-driven discrete sampling of motion perception. *Proc. Natl Acad. Sci. USA* **102**, 5291–5296. (doi:10.1073/pnas.0409172102)
13. Simpson WA, Shahani U, Manahilov V. 2005 Illusory percepts of moving patterns due to discrete temporal sampling. *Neurosci. Lett.* **375**, 23–27. (doi:10.1016/j.neulet.2004.10.059)
14. Andrews T, Purves D. 2005 The wagon-wheel illusion in continuous light. *Trends Cogn. Sci.* **9**, 261–263. (doi:10.1016/j.tics.2005.04.004)
15. VanRullen R, Reddy L, Koch C. 2006 The continuous wagon wheel illusion is associated with changes in electroencephalogram power at approximately 13 Hz. *J. Neurosci.* **26**, 502–507. (doi:10.1523/JNEUROSCI.4654-05.2006)
16. Piantoni G, Kline KA, Eagleman DM. 2010 Beta oscillations correlate with the probability of perceiving rivalrous visual stimuli. *J. Vis.* **10**, 18. (doi:10.1167/10.13.18)
17. Blake R, Logothetis NK. 2002 Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21. (doi:10.1038/nrn701)
18. Kline K, Holcombe AO, Eagleman DM. 2004 Illusory motion reversal is caused by rivalry, not by perceptual snapshots of the visual field. *Vision Res.* **44**, 2653–2658. (doi:10.1016/j.visres.2004.05.030)
19. Holcombe AO, Clifford CW, Eagleman DM, Pakarian P. 2005 Illusory motion reversal in tune with motion detectors. *Trends Cogn. Sci.* **9**, 559–560. (doi:10.1016/j.tics.2005.10.009)
20. Holcombe AO, Seizova-Cajic T. 2008 Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. *Vision Res.* **48**, 1743–1757. (doi:10.1016/j.visres.2008.05.019)
21. Kline K, Eagleman DM. 2008 Evidence against the temporal subsampling account of illusory motion reversal. *J. Vis.* **8**, 13 11–15.
22. Hutchinson CV, Ledgeway T. 2006 Sensitivity to spatial and temporal modulations of first-order and second-order motion. *Vision Res.* **46**, 324–335. (doi:10.1016/j.visres.2005.03.002)
23. VanRullen R. 2006 The continuous wagon wheel illusion is object-based. *Vision Res.* **46**, 4091–4095. (doi:10.1016/j.visres.2006.07.030)
24. Macdonald JSP, Cavanagh P, VanRullen R. 2013 Attentional sampling of multiple wagon wheels. *Attent. Percept. Psychophys.* **46**, 4091–4095.
25. Chaudhuri A. 1990 Modulation of the motion aftereffect by selective attention. *Nature* **344**, 60–62. (doi:10.1038/344060a0)
26. Rezec A, Krekelberg B, Dobkins KR. 2004 Attention enhances adaptability: evidence from motion adaptation experiments. *Vision Res.* **44**, 3035–3044. (doi:10.1016/j.visres.2004.07.020)
27. VanRullen R. 2007 The continuous wagon wheel illusion depends on, but is not identical to neuronal adaptation. *Vision Res.* **47**, 2143–2149. (doi:10.1016/j.visres.2007.03.019)
28. VanRullen R, Pascual-Leone A, Battelli L. 2008 The continuous wagon wheel illusion and the ‘when’ pathway of the right parietal lobe: a repetitive transcranial magnetic stimulation study. *PLoS ONE* **3**, e2911. (doi:10.1371/journal.pone.0002911)
29. Reddy L, Remy F, Vayssiere N, VanRullen R. 2011 Neural correlates of the continuous wagon wheel illusion: a functional MRI study. *Hum. Brain Mapp.* **32**, 163–170. (doi:10.1002/hbm.21007)
30. VanRullen R, Busch NA, Drewes J, Dubois J. 2011 Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Front. Percept. Sci.* **2**, 1–9.
31. Busch NA, Dubois J, VanRullen R. 2009 The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* **29**, 7869–7876. (doi:10.1523/JNEUROSCI.0113-09.2009)
32. Mathewson KE, Gratton G, Fabiani M, Beck DM, Ro T. 2009 To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci.* **29**, 2725–2732. (doi:10.1523/JNEUROSCI.3963-08.2009)
33. Busch NA, VanRullen R. 2010 Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc. Natl Acad. Sci. USA* **107**, 16 048–16 053. (doi:10.1073/pnas.1004801107)
34. Dugue L, Marque P, VanRullen R. 2011 The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. *J. Neurosci.* **31**, 11 889–11 893. (doi:10.1523/JNEUROSCI.1161-11.2011)
35. Drewes J, VanRullen R. 2011 This is the rhythm of your eyes: the phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. *J. Neurosci.* **31**, 4698–4708. (doi:10.1523/JNEUROSCI.4795-10.2011)
36. Dugue L, Marque P, VanRullen R. Submitted. Theta oscillations modulate attentional search performance periodically. *J. Neurosci.*
37. Varela FJ, Toro A, John ER, Schwartz EL. 1981 Perceptual framing and cortical alpha rhythm. *Neuropsychologia* **19**, 675–686. (doi:10.1016/0028-3932(81)90005-1)
38. Gho M, Varela FJ. 1988 A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *J. Physiol.* **83**, 95–101.
39. Chakravarthi R, VanRullen R. 2012 Conscious updating is a rhythmic process. *Proc. Natl Acad. Sci. USA* **109**, 10 599–10 604. (doi:10.1073/pnas.1121622109)
40. Nijhawan R. 1994 Motion extrapolation in catching. *Nature* **370**, 256–257. (doi:10.1038/370256b0)
41. Eagleman DM, Sejnowski TJ. 2000 Motion integration and postdiction in visual awareness. *Science* **287**, 2036–2038. (doi:10.1126/science.287.5460.2036)
42. VanRullen R, Macdonald JSP. 2012 Perceptual echoes at 10 Hz in the human brain. *Curr. Biol.* **22**, 995–999. (doi:10.1016/j.cub.2012.03.050)
43. Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ. 2006 The VESPA: a method for the rapid estimation of a visual evoked potential. *Neuroimage* **32**, 1549–1561. (doi:10.1016/j.neuroimage.2006.05.054)
44. Ciganek L. 1969 Variability of the human visual evoked potential: normative data. *Electroencephalogr. Clin. Neurophysiol.* **27**, 35–42. (doi:10.1016/0013-4694(69)90106-0)
45. Hillyard SA, Teder-Salejari WA, Munte TF. 1998 Temporal dynamics of early perceptual processing. *Curr. Opin. Neurobiol.* **8**, 202–210. (doi:10.1016/S0959-4388(98)80141-4)
46. Townsend J. 1990 Serial vs. parallel processing: sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychol. Sci.* **1**, 46–54. (doi:10.1111/j.1467-9280.1990.tb00067.x)
47. Jans B, Peters JC, De Weerd P. 2010 Visual spatial attention to multiple locations at once: the jury is still out. *Psychol. Rev.* **117**, 637–684. (doi:10.1037/a0019082)
48. VanRullen R, Dubois J. 2011 The psychophysics of brain rhythms. *Front. Psychol.* **2**, 203. (doi:10.3389/fpsyg.2011.00203)
49. Eriksen CW, Spencer T. 1969 Rate of information processing in visual perception: some results and methodological considerations. *J. Exp. Psychol.* **79**, 1–16. (doi:10.1037/h0026873)
50. Treisman A. 1969 Strategies and models of selective attention. *Psychol. Rev.* **76**, 282–299. (doi:10.1037/h0027242)
51. Kahneman D. 1973 *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
52. VanRullen R, Carlson T, Cavanagh P. 2007 The blinking spotlight of attention. *Proc. Natl Acad. Sci. USA* **104**, 19 204–19 209. (doi:10.1073/pnas.0707316104)
53. Landau AN, Fries P. 2012 Attention samples stimuli rhythmically. *Curr. Biol.* **22**, 1000–1004. (doi:10.1016/j.cub.2012.03.054)
54. Posner MI. 1994 Attention: the mechanisms of consciousness. *Proc. Natl Acad. Sci. USA* **91**, 7398–7403. (doi:10.1073/pnas.91.16.7398)
55. Mack A, Rock I. 1998 *Inattention blindness*. Cambridge, MA: MIT Press.
56. Regan D. 1977 Steady-state evoked potentials. *J. Opt. Soc. Am.* **67**, 1475–1489. (doi:10.1364/JOSA.67.001475)
57. Herrmann CS. 2001 Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* **137**, 346–353. (doi:10.1007/s002210100682)
58. Rager G, Singer W. 1998 The response of cat visual cortex to flicker stimuli of variable frequency. *Eur. J. Neurosci.* **10**, 1856–1877. (doi:10.1046/j.1460-9568.1998.00197.x)
59. Galambos R, Makeig S, Talmachoff PJ. 1981 A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl Acad. Sci. USA* **78**, 2643–2647. (doi:10.1073/pnas.78.4.2643)
60. Stapells DR, Linden D, Suffield JB, Hamel G, Picton TW. 1984 Human auditory steady state potentials. *Ear Hear.* **5**, 105–113. (doi:10.1097/00003446-198403000-00009)
61. Greenberg S. 1998 A syllable-centric framework for the evolution of spoken language. *Behav. Brain Sci.* **21**, 518. (doi:10.1017/S0140525X98301260)

62. Poeppel D. 2003 The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255. (doi:10.1016/S0167-6393(02)00107-3)
63. Ghitza O, Greenberg S. 2009 On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* **66**, 113–126. (doi:10.1159/000208934).
64. Giraud AL, Poeppel D. 2012 Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517. (doi:10.1038/nn.3063)
65. Peelle JE, Davis MH. 2012 Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* **3**, 320. (doi:10.3389/fpsyg.2012.00320)
66. Lakatos S, Shepard RN. 1997 Constraints common to apparent motion in visual, tactile, and auditory space. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 1050–1060. (doi:10.1037/0096-1523.23.4.1050)
67. Feron FX, Frissen I, Boissinot J, Guastavino C. 2010 Upper limits of auditory rotational motion perception. *J. Acoust. Soc. Am.* **128**, 3703–3714. (doi:10.1121/1.3502456)
68. Saenz M, Langers DR. 2013 Tonotopic mapping of human auditory cortex. *Hear. Res.* **307**, 42–52. (doi:10.1016/j.heares.2013.07.016)
69. Shepard RN. 1964 Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* **36**, 2346–2353. (doi:10.1121/1.1919362)
70. Zoefel B, Heil P. 2013 Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front. Psychol.* **4**, 262. (doi:10.3389/fpsyg.2013.00262)
71. Ng BS, Schroeder T, Kayser C. 2012 A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J. Neurosci.* **32**, 12 268–12 276. (doi:10.1523/JNEUROSCI.1877-12.2012)
72. Vanrullen R, McLelland D. 2013 What goes up must come down: EEG phase modulates auditory perception in both directions. *Front. Psychol.* **4**, 16. (doi:10.3389/fpsyg.2013.00016)
73. Henry MJ, Herrmann B. 2012 A precluding role of low-frequency oscillations for auditory perception in a continuous processing mode. *J. Neurosci.* **32**, 17 525–17 527. (doi:10.1523/JNEUROSCI.4456-12.2012)
74. Large EW, Jones MR. 1999 The dynamics of attending: how people track time-varying events. *Psychol. Rev.* **106**, 119–159. (doi:10.1037/0033-295X.106.1.119)
75. Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. 2008 Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320**, 110–113. (doi:10.1126/science.1154735)
76. Schroeder CE, Lakatos P. 2009 Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18. (doi:10.1016/j.tins.2008.09.012)
77. Henry MJ, Obleser J. 2012 Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl Acad. Sci. USA* **109**, 20 095–20 100. (doi:10.1073/pnas.1213390109)
78. Luo H, Poeppel D. 2007 Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**, 1001–1010. (doi:10.1016/j.neuron.2007.06.004)
79. Kerlin JR, Shahin AJ, Miller LM. 2010 Attentional gain control of ongoing cortical speech representations in a 'cocktail party'. *J. Neurosci.* **30**, 620–628. (doi:10.1523/JNEUROSCI.3631-09.2010)
80. Ghitza O. 2011 Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* **2**, 130. (doi:10.3389/fpsyg.2011.00130)
81. Ghitza O. 2012 On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Front. Psychol.* **3**, 238. (doi:10.3389/fpsyg.2012.00238)
82. Ding N, Simon JZ. 2012 Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl Acad. Sci. USA* **109**, 11 854–11 859. (doi:10.1073/pnas.1205381109)
83. Ding N, Simon JZ. 2013 Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* **33**, 5728–5735. (doi:10.1523/JNEUROSCI.5297-12.2013)
84. Zion-Golumbic EM, Poeppel D, Schroeder CE. 2012 Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* **122**, 151–161. (doi:10.1016/j.bandl.2011.12.010)
85. Zion-Golumbic EM *et al.* 2013 Mechanisms underlying selective neuronal tracking of attended speech at a 'cocktail party'. *Neuron* **77**, 980–991. (doi:10.1016/j.neuron.2012.12.037)
86. Doelling KB, Arnal LH, Ghitza O, Poeppel D. 2013 Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* **85**, 761–768. (doi:10.1016/j.neuroimage.2013.06.035)
87. Peelle JE, Gross J, Davis MH. 2013 Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* **23**, 1378–1387. (doi:10.1093/cercor/bhs118)
88. Llinas RR, Pare D. 1991 Of dreaming and wakefulness. *Neuroscience* **44**, 521–535. (doi:10.1016/0306-4522(91)90075-Y)
89. Thorne JD, Debener S. 2013 Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* **307**, 144–152. (doi:10.1016/j.heares.2013.07.002)
90. Kavanagh KT, Domico WD. 1986 High-pass digital filtration of the 40 Hz response and its relationship to the spectral content of the middle latency and 40 Hz responses. *Ear Hear.* **7**, 93–99. (doi:10.1097/00003446-198604000-00007)
91. İlhan B, VanRullen R. 2012 No counterpart of visual perceptual echoes in the auditory system. *PLoS ONE* **7**, e49287. (doi:10.1371/journal.pone.0049287)
92. Wertheimer M. 1912 Experimentelle Studien über das Sehen von Bewegung. *Zeit. Psychol.* **61**, 161–265.
93. van Santen JP, Sperling G. 1985 Elaborated Reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300–321. (doi:10.1364/JOSAA.2.000300)
94. Kubovy M. 1988 Should we resist the seductiveness of the space:time::vision:audition analogy? *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 318–320. (doi:10.1037/0096-1523.14.2.318)
95. Apteker RT, Fisher JA, Kisimov VS, Neishlos H. 1995 Video acceptability and frame rate. *IEEE Multimedia* **2**, 32–40. (doi:10.1109/93.410510)
96. McCarthy JD, Sasse MA, Miras D. 2004 Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. In *Proc. SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24–29 Apr 2004*. pp. 535–542. Aarhus, Denmark: Interaction Design Foundation.
97. Miller GA, Licklider JCR. 1950 The intelligibility of interrupted speech. *J. Acoust. Soc. Am.* **22**, 167. (doi:10.1121/1.1906584)
98. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995 Speech recognition with primarily temporal cues. *Science* **270**, 303–304. (doi:10.1126/science.270.5234.303)
99. Obleser J, Eisner F, Kotz SA. 2008 Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* **28**, 8116–8123. (doi:10.1523/JNEUROSCI.1290-08.2008)
100. Saberi K, Perrott DR. 1999 Cognitive restoration of reversed speech. *Nature* **398**, 760. (doi:10.1038/19652)
101. Stilp CE, Kieffe M, Alexander JM, Kluender KR. 2010 Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *J. Acoust. Soc. Am.* **128**, 2112–2126. (doi:10.1121/1.3483719)
102. Drullman R, Festen JM, Plomp R. 1994 Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* **95**, 2670–2680. (doi:10.1121/1.409836)
103. Drullman R, Festen JM, Plomp R. 1994 Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **95**, 1053–1064. (doi:10.1121/1.408467)
104. Joliot M, Ribary U, Llinas R. 1994 Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proc. Natl Acad. Sci. USA* **91**, 11 748–11 751. (doi:10.1073/pnas.91.24.11748)
105. Hirsh IJ, Sherrick CEJ. 1961 Perceived order in different sense modalities. *J. Exp. Psychol.* **62**, 423–432. (doi:10.1037/h0045283)
106. Poppel E. 1997 A hierarchical model of temporal perception. *Trends Cogn. Sci.* **1**, 56–61. (doi:10.1016/S1364-6613(97)01008-5)
107. Frisina RD. 2001 Subcortical neural coding mechanisms for auditory temporal processing. *Hear. Res.* **158**, 1–27. (doi:10.1016/S0378-5955(01)00296-9)
108. Kohler MA. 1997 A comparison of the new 2400 bps MELP Federal Standard with other standard coders. In *ICASSP-97: IEEE Intl. Conf. Acoustics,*

- Speech, and Signal Processing, Munich, Germany, 21–24 April 1997*, vol. 2, pp. 1587–1590. (doi:10.1109/ICASSP.1997.596256)
109. Funke K, Kisvarday F, Volgushev M, Worgotter F. 2002 The visual cortex. In *Models of neural networks IV: early vision and attention* (eds JL van Hemmen, JD Cowan, E Domany), pp. 131–160. New York: Springer.
110. Nelken I, Las L, Ulanovsky N, Farkas D. 2005 Are speech, pitch, and space early or late in the auditory system? In *The auditory cortex: a synthesis of human and animal research* (eds R Konig, P Heil, E BudingerH Scheich). Mahwah, NJ: Lawrence Erlbaum Associates.
111. Luo H, Tian X, Song K, Zhou K, Poeppel D. 2013 Neural response phase tracks how listeners learn new acoustic representations. *Curr. Biol.* **23**, 968–974. (doi:10.1016/j.cub.2013.04.031)
112. Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. 2004 Attentional modulation of human auditory cortex. *Nat. Neurosci.* **7**, 658–663. (doi:10.1038/nn1256)
113. Rinne T, Stecker GC, Kang X, Yund EW, Herron TJ, Woods DL. 2007 Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *Neuroreport* **18**, 1311–1314. (doi:10.1097/WNR.0b013e32826fb3bb)
114. Buffalo EA, Fries P, Landman R, Liang H, Desimone R. 2010 A backward progression of attentional effects in the ventral stream. *Proc. Natl Acad. Sci. USA* **107**, 361–365. (doi:10.1073/pnas.0907658106)
115. Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE. 2009 The leading sense: supramodal control of neurophysiological context by attention. *Neuron* **64**, 419–430. (doi:10.1016/j.neuron.2009.10.014)