

Review

Not peer-reviewed version

Pitch at the Cocktail Party: A Comparative Approach to Studying Selective Attention

Joel Ward , Veronica M Tarka , Artem Diuba , [Kerry MM Walker](#) *

Posted Date: 3 March 2026

doi: 10.20944/preprints202603.0223.v1

Keywords: pitch; attention; hearing; auditory cortex; harmonics; listening



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Pitch at the Cocktail Party: A Comparative Approach to Studying Selective Attention

Joel Ward ^{1,2,†}, Veronica M Tarka ^{1,†}, Artem Diuba ¹ and Kerry MM Walker ^{1,*}

¹ Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford, OX1 3PT, UK

² Nuffield Department of Surgical Sciences, University of Oxford, Oxford, OX3 9DU, UK

* Correspondence: kerry.walker@dpag.ox.ac.uk

[†] Joint first authors.

Simple Summary

Natural environments often contain overlapping sound waves from many different sound sources. For example, to follow a conversation in a crowded restaurant, the brain must separate these sounds so we can focus on our friend's voice. Pitch is the tonal quality that makes a sound seem low or high, and it plays a critical role in guiding our attention. However, the brain mechanisms that support this process remain poorly understood and understudied. Here, we describe how improved behavioural tasks and neural recording studies in both humans and animal models could help advance our knowledge of how pitch helps us listen in complex environments.

Abstract

Pitch is a powerful cue for segregating sound sources in complex acoustic scenes, yet the neural mechanisms through which it guides selective attention remain unclear. In this review, we synthesise behavioural and neurophysiological evidence from humans and animal models to examine how pitch supports selective listening in a two-stage process: bottom-up pitch-based feature binding, followed by top-down enhancement of an attended sound source. Behavioural studies demonstrate that even modest pitch differences substantially improve listeners' segregation of harmonic sounds, tone streams, and competing talkers. Human EEG, MEG, fMRI and ECoG studies show enhancement of target sound representations in auditory cortex during selective listening, but understanding this process at the level of individual neurons requires further study in animals that are trained in pitch-based selective listening tasks. Other key questions in this field include the relative roles of resolved and unresolved harmonic cues, the neural circuit mechanisms underlying target enhancement versus masker suppression, and how attention can target distributed cortical pitch representations. We argue that cross-species, naturalistic paradigms are essential for answering these questions and for addressing the listening difficulties associated with ageing and hearing loss.

Keywords: pitch; attention; hearing; auditory cortex; harmonics; listening

Introduction

Following a conversation in a noisy restaurant is a familiar challenge, and one that becomes especially difficult with ageing or hearing loss. Understanding how the brain solves this “cocktail party problem” is a central aim of auditory neuroscience, and it is commonly conceptualised as a two-stage process [1]. In the first *feature binding* stage, the ascending auditory pathway extracts complex acoustic features, such as harmonic structure, temporal synchrony, and spatial location, and binds these into perceptual objects [2–4]. In the second stage, *top-down attention* selects and enhances target sound representations in the auditory system [5,6]. Pitch is among the most potent cues for sound segregation [2,7], yet how it guides attention at the level of individual neurons and microcircuits remains poorly understood.

Pitch, typically defined as the tonal quality of a sound along a low-to-high scale [8], is a salient perceptual feature of speech and other harmonic sounds. Acoustically, the pitch we hear corresponds to the sound's fundamental frequency (F0). A sound that evokes a pitch is typically periodic at F0, and all its frequency components (the *harmonics*) are integer multiples of F0.

Animals, like humans, can detect, discriminate, and order pitch [9–13], and neural representations of pitch have been described in the auditory cortex of many mammalian species [14]. Attention to task-relevant sound sources modulates auditory cortical responses in both humans [6,15] and non-human animals [5,16,17]. Studies on pitch-based selective listening therefore support the translation of circuit-level measurements and neural manipulations that are only feasible in animal models to cognitive and language processes that are unique to humans. Integrating human and animal research is therefore essential to uncovering the neural mechanisms for pitch-based selective listening.

Longstanding insights into selective listening have come from experiments that used pure tones or harmonic tone complexes [1,18–20]. Determining the frequency of a pure tone is mechanistically trivial for the auditory system, as it is place mapped in the tonotopic auditory system, and does not require F0 extraction across multiple harmonics. While these simple stimuli isolate acoustical cues well, they do not capture the complexity of overlapping voices, fluctuating noise, and dynamic spectral-temporal patterns that characterise natural scenes. More ecologically relevant paradigms use speech, vocalisations, and other environmental sounds, which have highly complex cortical representations [6,21–24], and may better probe how binding and attention operate in realistic contexts [23]. Moving beyond pure-tone stimuli is critical to translate findings into clinical applications, as most people with hearing loss struggle to follow conversation in noisy environments, even if their pure tone audiogram is only mildly impaired [25,26].

This review synthesises behavioural and neurophysiological evidence in humans and animal models on how pitch perception guides selective listening. We highlight gaps in our current understanding, and propose how the open questions in this field can be addressed by methodological innovations and cross-species comparisons.

Behavioural Evidence for Pitch-Based Selective Listening

Pitch facilitates the segregation of voices in everyday listening. Listeners can more readily follow a female voice against a male voice background and vice versa, due to their pitch differences [27,28]. In fact, when talkers differ in F0 by about 100Hz, listeners showed an intelligibility improvement equivalent to a 6 dB gain in signal-to-noise ratio [27]. Even modest pitch differences of 1–4 semitones can improve identification of two simultaneously presented vowels [29–32] (Figure 1a). Thus, pitch may play a particularly important role for segregating voices that overlap in location and time.

Harmonicity refers to the extent to which the frequency components of a sound are integer multiples of F0, and this acoustical correlate of pitch may play a particularly important role in binding. Mistuning a single harmonic in a tone complex by as little as 1–3% can cause it to be heard as a separate auditory object, demonstrating the strength of harmonic fusion [33–36] (Figure 1b). Harmonicity improves detection and discrimination of speech presented in noise [37] or with a competing talker [38] (Figure 1c). At least some of these segregation effects can be explained by the equal spacing of frequency components (*spectral regularity*) in harmonic sounds, even when those components are not multiples of a common F0 [36,39]. Nevertheless, harmonicity is especially important for pitch judgements made in noisy backgrounds [40], suggesting that we bind an auditory object's harmonic components in order to segregate it from other sound sources.

Classic two-tone streaming paradigms have used alternating low (A) and high (B) frequency pure tone sequences to probe the use of frequency for auditory feature binding (Figure 1d). In these tasks, tones are often presented in a repeating ABA- sequence, where the hyphen indicates a silent gap after each tone triplet. Small frequency differences (<10%) between A and B produce a fused percept of a single auditory object with a “galloping” rhythm, and larger differences are perceived as two segregated A and B streams [1]. These effects generalise to harmonic complexes, where both F0

and the spectral envelope independently contribute to grouping [41,42]. Streaming paradigms allow the effects of pitch on sound segregation to be quantified through parametric control of F0 differences. However, their reliance on subjective reports of “one versus two streams” is a challenge for implementing these tasks in animal models.

Few studies have directly trained pitch-based selective listening tasks in animals, but there is evidence that harmonicity helps them to perceptually bind natural sounds. For example, animals can detect a mistuned harmonic in a tone complex [43–46]. Adapting two-tone streaming paradigms for animals also offers a promising comparative approach [47]. Rather than relying on subjective reports of one or two streams, these paradigms in animals instead measure objective behavioural performance on a task that requires effective stream segregation. For example, if A and B are segregated, human listeners show improved recognition of melodies in the A stream [48]. Izumi [49] took this approach, training macaques to detect frequency contours in a target tone stream while ignoring similar contours in a distractor stream. Ma et al. [50] trained a similar two-tone streaming task in ferrets. In both cases, the animals performed better when the A and B streams had larger frequency differences, mirroring the effects observed in human listeners. Another approach to training ABA- streaming tasks in non-humans is to first train animals to discriminate between galloping and isochronous rhythms in a single A stream, and then test them on a variety of ABA-streams. This approach has been successfully used to demonstrate frequency-dependent behaviour in songbirds [51] and even goldfish [52]. These tasks enable direct cross-species comparisons of frequency separation effects on selective listening, and are feasible to train in a variety of animal models. However, they are a less direct demonstration of perceptual streaming than their human counterparts. A limitation of the corpus of streaming studies in animals to date is that they have focussed on pure tone streaming, so they have not tested the role of F0 in segregating spectrally overlapping sounds.

Studying selective listening with more naturalistic tasks, particularly those involving speech, remains challenging in animals. Geissler and Ehret altered the onsets of individual harmonics in a mouse pup call relative to the fundamental frequency, and showed that nursing mothers were less likely to respond to these altered calls [53]. This behavioural effect could reflect either a difficulty in segregating the call from background noise or a more general impairment of call identification. Psychophysical designs in animals, where verbal instructions are not possible, must be careful to prevent alternative listening strategies. For example, animals may detect a brief spectrotemporal cue within a stream rather than sustaining attention to the target. These issues can be mitigated by including catch trials to measure how well animals sustain focus [54,55], or by using task-switching paradigms [56]. These approaches are well suited to capture aspects of dynamic attentional control akin to the human cocktail party scenario.

In summary, pitch plays a key role in binding and selective attention in humans, as demonstrated across a range of psychophysical paradigms. Appropriate behavioural tools already exist to more extensively examine this process in animal models, but such studies must move beyond pure tone stimuli to explore the role of complex pitch in directing attention.

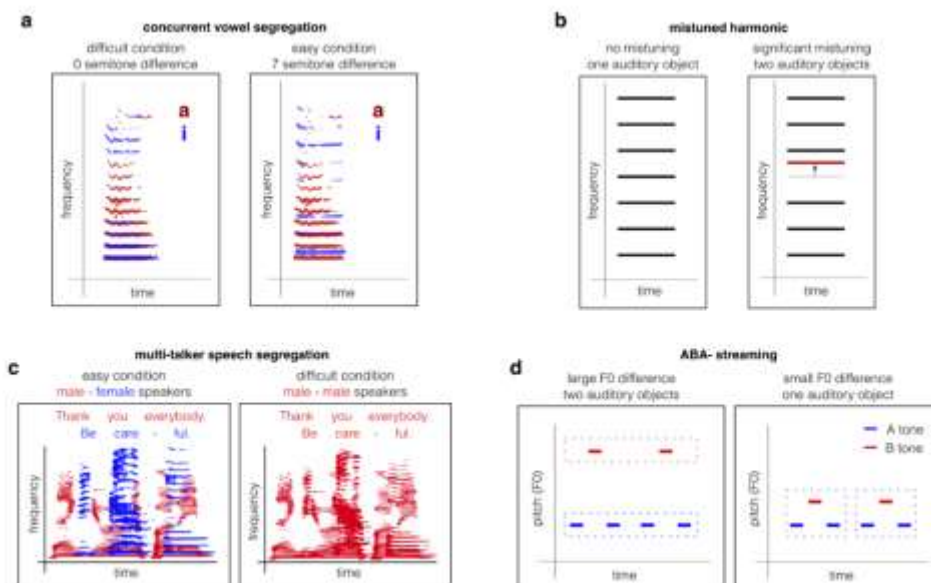


Figure 1. Common paradigms for measuring pitch-based feature binding and selective attention. **a)** Concurrent vowel segregation tasks the listener with identifying two simultaneously presented vowels that differ only in F0. **b)** When one harmonic within a tone complex is sufficiently mistuned, it is no longer a perfect multiple of the sound's F0. The mistuned harmonic 'pops out' to form a separate auditory object with a different perceived pitch from the other perceptually fused harmonics. **c)** Multi-talker speech segregation tasks present the listener with sentences spoken by two people simultaneously. The listener must identify words spoken by the target talker, who is identifiable by some feature such as gender, relative onset timing, or the first word spoken. This task becomes easier when there is a pitch separation between the talkers, typically provided by one talker being male and the other female. **d)** ABA-streaming (also known as two-tone streaming) paradigms present two interleaved sequences of tones at different rates. The sequences of A and B tones are perceived as a single stream of tone triplets with a 'galloping' rhythm when the pitch separation between the tones is sufficiently small. When the pitch separation is large, the A and B tones form two distinct perceptual streams.

The Roles of Resolved Harmonics and Temporal Pitch Cues in Selective Listening

At least two acoustic characteristics determine a sound's pitch. First, the waveform is periodic with a repetition rate of F0, enabling phase-locked temporal representations in the auditory nerve. Second, the frequency components of the sound are all harmonics of a common F0, represented as a place code across the tonotopic map (Figure 2). The place code of F0 is only *resolved* for lower harmonics (the first 5–8 harmonics in human listeners). The frequency bandwidths of neurons tuned to higher harmonics are broader on a linear frequency scale, including neighbouring harmonics, so these are *unresolved* [57]. Human psychophysical studies suggest that both resolved harmonics and temporal envelopes contribute to pitch perception, but resolved harmonics provide finer pitch acuity [20,58–60]. Other mammals also show evidence of this dual pitch extraction strategy [10–12,45,61,62], but non-humans often rely more on temporal pitch due to poorer frequency resolution in their cochlea [12,45,61].

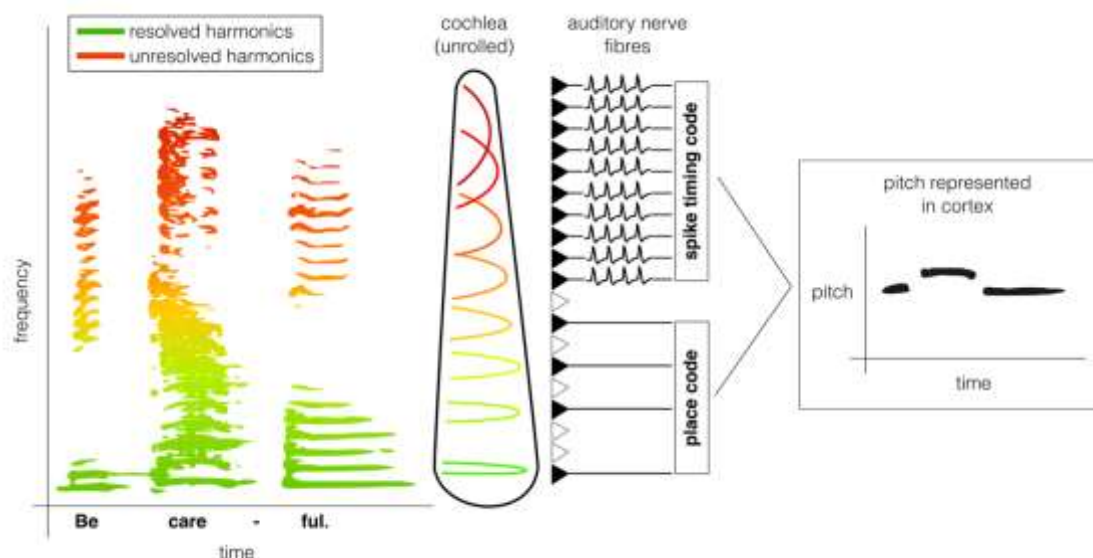


Figure 2. Resolved and unresolved harmonics can each give rise to pitch perception through distinct neural mechanisms. Most natural sounds such as speech contain both resolved (green) and unresolved (red) harmonics (left plot). On a linear scale, frequency receptive fields in the cochlea are wider for higher frequencies. As a result, low-numbered harmonics are ‘resolved’ in the cochlea, with only one harmonic falling within the receptive field of a given neuron. Auditory nerve fibres tuned to higher-numbered harmonics instead respond to multiple harmonics within their frequency receptive field, so these harmonics are termed ‘unresolved’. The resolved harmonics produce a place code representation of F0 in the auditory nerve (middle plot). The F0 of unresolved harmonics are instead encoded in an explicit spike timing code, as the auditory nerve fibres phase lock to their summed periodicity at F0. Neurons in auditory cortex may combine these two codes to provide a cue-invariant representation of pitch (right plot).

Both resolved and unresolved harmonics can support sound segregation in streaming tasks. Human listeners show similar stream build-up and oddball detection in the attended stream using either of these pitch cues [63,64]. However, pitch-based streaming effects in humans tend to be stronger when harmonic tones contain more resolved harmonics, and weaker when they rely more on unresolved harmonics [65,66]. In contrast, Madsen et al. [67] found no effect of harmonic resolvability on a multi-talker listening task. However, their process of low- and high-pass filtering speech to isolate harmonics disrupted formant cues that are critical to intelligibility and thus task performance. This can be minimised in future work by manipulating only harmonic components of speech, leaving noise-filled formant structures intact [38].

The variability in the number of resolved harmonics available across species offers a comparative framework for understanding how pitch is utilised for selective listening. For example, we might predict that ferrets and rodents might segregate sounds more effectively with unresolved harmonics, in contrast to human listeners. To date, the role of different pitch cues in selective listening performance remains unexplored in animal behaviour. Neurophysiological studies in animals, meanwhile, have provided insights into how these two F0 cues are represented in neural spiking responses. A subset of neurons in the auditory cortex of marmosets and ferrets are specialised to process resolved harmonic F0 cues, while others extract the periodicity of F0 from unresolved harmonics [11,68]. At a population level, neurons in macaque primary auditory cortex can represent resolved harmonic patterns as a place code, while low F0s are also represented by neuronal phase-locking to F0 [69].

When two-tone streaming stimuli are presented to animals, auditory cortical neurons show more distinct spiking responses to A and B sounds when their F0 difference is larger, even for unresolved harmonic complex tones that only differ in their temporal periodicity [70,71]. This provides a neural correlate of temporal-pitch-based stream segregation of the same sequences by

human listeners [72,73]. Furthermore, the differences in the neural response to A and B complexes developed over the first 5-20 seconds of sequence presentation, mirroring the time frame of streaming build-up in human perception [71]. These findings are from passively listening animals, leaving the link between neural activity and attentional effects unconfirmed. However, the fact that neural segregation occurs without active task engagement suggests that individual pitch cues, such as the temporal envelopes of unresolved harmonics, can support binding of auditory objects in these species.

Neural Activity Supporting Pitch-Based Selective Listening

Selective listening requires coordinated activity across widely distributed brain areas within and outside the auditory system. Even the neural basis for perceiving the pitch of an isolated sound source is poorly understood; auditory cortex is held to play a key role in pitch perception, but it remains unclear if pitch is processed within a specialised pitch centre or through representations spread across multiple auditory cortical regions [14,74]. Understanding the more widespread neural processes that integrate the bottom-up binding of pitch features with top-down attentional selection is an even more daunting challenge. However, non-invasive methods commonly used to study human brain function (EEG, MEG, and fMRI) and, more recently, intracranial electrocorticography (ECoG) in epilepsy patients, are ideally suited to sample such broad neural activation patterns.

EEG combines broad spatial sampling of neural activity with fast temporal fidelity, and has provided physiological evidence of selective listening being a two-stage process. Studies using concurrent vowels or mistuned harmonics as stimuli have observed an early object-related negativity, thought to reflect bottom-up feature binding, when both sounds are presented at sufficiently different pitches, even if the subject is not attending to them [75–77]. A later positive (P400) event-related potential follows only if the participant is actively listening to the sounds to perform a behavioural task, and may therefore result from top-down attention. Functional MRI, which can better localize the source of neural activity to particular brain regions, has shown that activity in left auditory thalamus and auditory cortex (including Heschl's gyrus, planum temporale and the superior temporal gyrus) is increased when listeners segregate two vowels, harmonic tones, or speech segments presented simultaneously at different pitches [78,79].

Human electrophysiological approaches have also been applied to more naturalistic multi-talker tasks, where listeners are asked to attend to one voice and ignore another. A wealth of EEG studies has shown that attention selectively enhances the neural representation of the target speech envelope [80–84]. However, these studies have usually presented the two speakers lateralised to different sides of the head, so segregation will largely result from spatial cues rather than pitch. A smaller number of investigations have instead presented a male and female voice diotically, where pitch cues will dominate segregation, and these also show enhanced phase-locked cortical tracking of the target speech envelope, with weaker neural signatures of the ignored speech [85,86]. ECoG recordings in epilepsy patients have provided further insights into the cortical distribution of selective listening effects. While neural activity in primary auditory cortex represents both voices in a multi-talker listening task, even when selectively listening to one voice, higher auditory cortical areas dynamically track the attended speaker more exclusively [6,24]. The spectrogram of the attended, but not ignored, voice can be reconstructed from these higher cortical responses [87]. Together, these studies suggest that selective attention enhances the representation of the attended talker, particularly in secondary auditory cortex. However, a limitation of the male/female talker design is that they do not carefully isolate or parameterize pitch cues; timbre may also be used to segregate voices in these experiments. Studies that present the same voice at a range of different pitches would better examine the specific role of pitch in multi-talker segregation.

Two-tone streaming sequences have also been used to identify neural markers of auditory scene analysis, where the role of pitch in segregation have been better controlled. Early EEG studies using this paradigm have shown increased positive evoked responses [88] or mismatched negativity responses [89] to pure tone sequences when they are perceptually segregated into two streams.

Functional MRI has further shown that the increased response to segregated pure tone streams is localised in auditory cortex [90] (but note that Cusack [91] reports selective activation only in higher regions of the intraparietal sulcus). At least one study has shown that the neurobiological basis of streaming the frequency of pure tones may extend to the pitch of spectrally overlapping harmonic tone complexes: Gutschalk et al. [92] employed fMRI and MEG to show increased primary and secondary auditory cortical activation when listeners perceived two streams based on F0 separation. Together, the above studies suggest that the human auditory cortex, particularly non-primary regions, may play a key role in frequency- and pitch-based selective listening.

Due to the challenges of training animals on selective listening tasks (discussed above), much of the work examining neural correlates of sound source segregation has been carried out in passively listening or anaesthetised animals. The pitch of two simultaneously presented harmonic sounds are represented with a combination of tonotopic place codes (for resolved harmonics) and envelope-locked temporal spiking patterns as early as the auditory nerve [93–96]. The F0s of spectrally overlapping sounds continue to be represented at subcortical nuclei in the auditory system, including the ventral cochlear nucleus [97], and therefore retain the information needed for later pitch-based segregation in the cortex. At the level of the primary auditory cortex, the F0s of each harmonic sound in the pair are represented as distinct rate-place codes across populations of neurons [69,98]. These neural correlates have not yet been demonstrated to drive selective attention in these species, but this remains a goal of future studies in awake, behaving animals.

Based on the human EEG and ECoG studies above, we would expect that responses in auditory cortical neurons would become dominated by representations of the target sound as animals engage attention. Many studies have demonstrated that auditory cortex provides a stable representation of dynamic sound “foregrounds” in the presence of stationary “background” noise, even in passively listening or anaesthetised animals [99–105], but these neural mechanisms may be less relevant to segregating two dynamic sounds that differ only in pitch. Noise vocoding, which degrades pitch information, reduced the decoding performance for vocalisations presented in a noise background throughout the auditory subcortical and cortical pathway [105], which may suggest that pitch cues at least partially contribute to the noise invariance observed in the above studies. However, F0 effects on segregation need to be more directly investigated while controlling for timbre and slower temporal modulation cues.

While two-tone streaming paradigms have rarely been trained in animals, the neural correlates of ABA- or AB pure tone sequences that differ in frequency have been extensively studied in passively listening animals, including microelectrode recordings in macaques [106–108], ferrets [109], bats [110], rats [111], guinea pigs [112], and songbirds [51,113]. These studies have corroborated and extended results in human neurophysiological work at the single neuron level. In these paradigms, the B tone elicits reduced spiking responses when A and B are close in frequency, mirroring the perceptual fusion observed in human streaming. Segregation is decodable in individual neurons, but population-level signals align well with the perceptual boundary between integration and segregation observed in human listeners. This suggests that population dynamics may be more relevant for perceptual organisation than the output of any one neuron. Studying these population codes is technically challenging, particularly for pitch-based selective listening, because pitch representations are distributed widely across auditory cortex in most mammals [69,114–119]. Computational models that decode auditory objects from distributed pitch representations could offer testable predictions for the neural mechanisms of pitch-based selective listening. Streaming studies that employ harmonic, spectrally overlapping stimuli and in which the animal is actively listening are also required to understand how selective attention manifests at a single neuron level.

Together, the human and animal neurophysiological studies point to a key role of auditory cortex in pitch-guided selective listening, within a hierarchical network that employs bottom-up representations of pitch for perceptual binding and top-down attention through selective modulation of a segregated target.

Does Attention Enhance the Target or Suppress the Distractor?

A key mechanistic question is whether selective listening acts primarily through *target enhancement* (increasing the gain of the harmonic target representation in auditory cortex), *masker suppression* (inhibiting representations of a competing harmonic background sound), or both. This question was addressed in concurrent vowel identification tasks in humans, which found improved target recognition when the masking vowel was harmonic rather than inharmonic [120–122]. This led to the “*harmonic cancellation*” hypothesis, which proposes that harmonic maskers are grouped and removed from the mixture to better reveal the target [123].

More recent studies have challenged the harmonic cancellation hypothesis. In a speech-in-noise task, Steinmetzger and Rosen [124] created inharmonic maskers by frequency-shifting all spectral components equally, thereby preserving spectral regularity while disrupting harmonicity. They found no difference in target speech intelligibility between harmonic and inharmonic maskers, whether the masker F0 was static or dynamic. They argued that the benefits of harmonic maskers in earlier concurrent vowel experiments may have resulted from opportunities for spectral glimpsing and reduced temporal envelope fluctuations in harmonic maskers, rather than masker harmonicity per se. Taken together, human psychophysical data suggest that any advantage of masker harmonicity is context-dependent, making it difficult to draw conclusions about the relative importance of target and maskers on attentional modulation. It is worth noting that in real world listening environments, one is often faced with segregating multiple harmonic sources, but not the frequency shifted stimuli employed by Steinmetzger and Rosen [124].

Disentangling the contributions of excitation and suppression in the human brain remains difficult even with modern neurophysiological approaches, whereas techniques employed in animal models allow cellular and circuit-level investigations of these processes. Neural responses could be compared between excitatory and inhibitory neurons in auditory cortex, and causality can be tested by manipulating activity of targeted neuronal types. The potential of this approach is illustrated by previous studies of auditory attention outside the context of pitch-based listening. For example, for zebra finches detecting a target song in a background of bird chorus, broad-spiking (putative excitatory), but not narrow-spiking (putative inhibitory), neurons in higher auditory cortex encoded target songs in a noise-robust manner [22]. This study used pharmacological approaches to further demonstrate that target selectivity arises from GABA-dependent forward suppression. In mice, inhibiting parvalbumin-positive interneurons in auditory cortex degraded cortical representations of dynamic sounds in static broadband noise [125]. These studies show how inhibitory circuits can shape foreground-background coding in auditory cortex. Future work should leverage these approaches to test whether inhibitory mechanisms preferentially mediate masker suppression, while excitatory gain and top-down feedback preferentially drive target enhancement, and if harmonicity can guide these processes.

Conclusion

The existence of a common fundamental frequency in natural harmonic sounds allows their frequency components to be easily bound together into a single auditory object. These bound feature representations can then be targeted by attentional systems to enhance neural representations of a target object (Figure 3). Therefore, our perception of pitch provides a valuable cue for selectively listening to a wide variety of auditory objects, such as an individual voice or water dripping into a puddle. While pitch underlies musical melody and speech prosody in human soundscapes, its more fundamental, cross-species role in hearing is binding and segregating harmonic sound sources.

A host of paradigms have been employed to demonstrate how humans and other animals can direct their attention to pitch cues (Table 1). While cortical correlates of pitch-based selective attention have been described, many fundamental questions remain about these neurobiological mechanisms (Box 1). For example, while attention is known to selectively enhance target representations in auditory cortex, it remains unclear how these representations are anatomically targeted by attentional

brain regions that lack acoustic details. Furthermore, does attentional modulation involve enhancement of target representations, suppression of background sounds, or a combination of both? Cross-species parallel experiments suggest that these mechanisms are at least partially conserved across species, despite species-specific differences in the weighting of resolved harmonic and temporal envelope pitch cues. Therefore, neural recordings and manipulations in animals trained on selective listening tasks offer a promising approach to better understand the neural computations through which pitch guides attention. Given the overwhelming evidence that pitch-based selective listening is impaired with ageing and hearing loss [2,74], this integrated approach will help address an important clinical priority in addition to solving a fundamental question of sensory neuroscience.

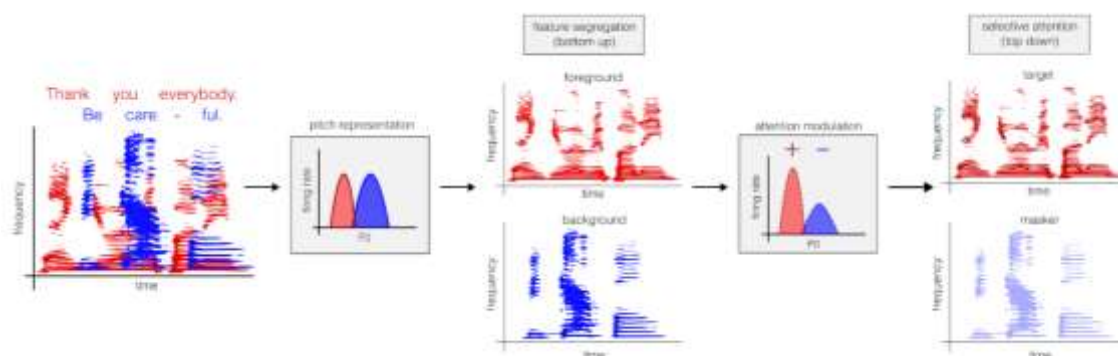


Figure 3. Selective listening conceptualized as a two-stage task requiring bottom-up feature segregation and top-down selective attention. A complex acoustic scene with multiple sound sources, such as two people talking, must first be encoded and relayed from peripheral to cortical auditory stations. Along this ascending pathway, each voice can be segregated by binding harmonic components of their F0. Higher cortical areas, potentially outside the auditory system, may enhance the auditory cortical representation of the target speaker (and/or inhibit the representation of the ignored speaker) through top-down modulation.

Table 1. Comparative paradigms used to study pitch-based segregation across humans and non-human animals, with selected citations.

<i>Paradigm / Stimulus</i>	<i>Species</i>	<i>Behavioural Findings</i>	<i>Neurophysiological Findings</i>
	Human	<i>F0 differences improve reporting both vowels [29–32,120,122].</i>	<i>Increased activation in auditory cortex when both vowels are successfully identified [75,78].</i>
Concurrent vowels	Non-human animals		<i>Neural encoding of both vowels in auditory nerve population using place and temporal codes. Population code for both vowels in place code in auditory cortex [98].</i>
	Human	<i>Detection of mistuned harmonics leads to perceptual segregation of the mistuned component as a separate object [33–36].</i>	<i>EEG/MEG show distinct responses to mistuned vs. harmonic tones; fMRI implicates auditory cortex in detecting harmonic violations [76].</i>
Mistuned harmonics	Non-human animals	<i>Animals (ferrets, gerbils, birds) detect mistuned harmonics, demonstrating perceptual grouping based on harmonicity [43–46].</i>	<i>Auditory cortical and subcortical neurons differentiate harmonic from mistuned tones, reflecting harmonicity-based segregation mechanisms [43,126].</i>

Two-tone streaming paradigms (pure tones or harmonic complexes)	Human	Complex F0 or tone frequency separation drives perceptual segregation; small differences (<10%) yield fusion (“gallop”), larger separations yield two streams [1,41,42].	EEG shows early negativity for automatic feature binding, and later positivity (P400) with active attention. fMRI/MEG shows increased auditory cortical responses for segregated vs. fused streams [88–92].
	Non-human animals	Monkeys, ferrets, birds, and even fish segregate tone streams when large enough frequency difference. Lack of studies of complex F0 streaming [47–50].	Populations of auditory cortical neurons show more distinct responses to alternating tones when presented with a larger frequency difference. Similar effects for the temporal pitch differences of harmonic complexes [51,70,71,106–113].
Multi-talker speech	Human	Female/male voice differences (high/low F0, respectively) facilitate segregation and speech intelligibility [27,28,38].	EEG and ECoG: enhanced cortical tracking of attended speech envelope MEG/fMRI: selective enhancement of target voice in secondary auditory cortex [6,24,79,85–87].

Box 1. Open research questions and suggested experimental directions.

Open research questions and future directions

Where and how is pitch represented across the auditory hierarchy? Is there a specialised “pitch centre” or distributed coding across auditory cortex? Multi-scale recordings and causal manipulations in animal models are required to help resolve this.

How does attention interact with pitch representations?

Does selective attention act primarily by enhancing target-related activity, suppressing distractors, or both?

What is the relative importance of resolved versus temporal pitch cues?

Does attention modulate these F0 cues independently, or does it act on pitch representations after these have been combined?

How does pitch-based segregation generalize from pure tones to complex sounds?

Most studies use simplified tones which can be trivially segregated via frequency place codes. Dynamic, complex harmonic stimuli with competing sources are needed to test whether tone-based mechanisms generalize to real-world listening.

How well does pitch support selective listening in multi-talker tasks?

There is a wealth of behavioural and neurophysiological studies examining selective attention to speech in humans, but almost none

Abbreviations

EEG electroencephalography

MEG magnetoencephalography

fMRI functional magnetic resonance imaging
ECoG electrocorticography
F0 fundamental frequency

Author Contributions: JW: VMT, AD and KMMW contributed to the writing the manuscript. VMT created the figures. JW wrote the first draft. JW and VMT contributed equally to the manuscript.

Funding: This work was funded by Biotechnology and Biological Sciences Research Council grants to K.M.M.W. (BB/M010929/1, BB/X013103/1), a University of Oxford Clarendon Scholarship to V.M.T., a Medical Sciences Division Studentship to J.W., an RCS England Surgical Research Fellowship (Freemasons' Royal Arch Research Fellowship with support from the Rosetrees Trust) to J.W., and a Medical Research Council Clinical research training fellowship to J.W. (UKRI449).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bregman, A.S. Auditory Scene Analysis: The Perceptual Organization of Sound; The MIT Press, 1990; ISBN 978-0-262-26920-9.
2. Micheyl, C.; Oxenham, A.J. Pitch, Harmonicity and Concurrent Sound Segregation: Psychoacoustical and Neurophysiological Findings. *Hear. Res.* **2010**, *266*, 36–51, doi:10.1016/j.heares.2009.09.012.
3. Shamma, S.; Elhilali, M.; Ma, L.; Micheyl, C.; Oxenham, A.J.; Pressnitzer, D.; Yin, P.; Xu, Y. Temporal Coherence and the Streaming of Complex Sounds. In *Basic Aspects of Hearing*; Moore, B.C.J., Patterson, R.D., Winter, I.M., Carlyon, R.P., Gockel, H.E., Eds.; Springer: New York, NY, 2013; pp. 535–543 ISBN 978-1-4614-1590-9.
4. Shinn-Cunningham, B.G. Object-Based Auditory and Visual Attention. *Trends Cogn. Sci.* **2008**, *12*, 182–186, doi:10.1016/j.tics.2008.02.003.
5. Fritz, J.B.; Elhilali, M.; Shamma, S.A. Adaptive Changes in Cortical Receptive Fields Induced by Attention to Complex Sounds. *J Neurophysiol* **2007**, *98*, 2337–2346, doi:10.1152/jn.00552.2007.
6. Mesgarani, N.; Chang, E.F. Selective Cortical Representation of Attended Speaker in Multi-Talker Speech Perception. *Nature* **2012**, *485*, 233–236, doi:10.1038/nature11020.
7. Grossberg, S. Pitch-Based Streaming in Auditory Perception. In *Musical Networks: Parallel Distributed Perception and Performance*; MIT Press: Cambridge, MA, 1996; pp. 117–140.
8. ANSI-S1.1 *Acoustical Terminology*; Acoustical Society of America, 2013;
9. Brosch, M.; Selezneva, E.; Bucks, C.; Scheich, H. Macaque Monkeys Discriminate Pitch Relationships. *Cognition* **2004**, *91*, 259–272, doi:10.1016/j.cognition.2003.09.005.
10. Klinge, A.; Klump, G.M. Frequency Difference Limens of Pure Tones and Harmonics within Complex Stimuli in Mongolian Gerbils and Humans. *J Acoust Soc Am* **2009**, *125*, 304–314, doi:10.1121/1.3021315.
11. Osmanski, M.S.; Song, X.; Wang, X. The Role of Harmonic Resolvability in Pitch Perception in a Vocal Nonhuman Primate, the Common Marmoset (*Callithrix Jacchus*). *J. Neurosci.* **2013**, *33*, 9161–9168, doi:10.1523/JNEUROSCI.0066-13.2013.
12. Walker, K.M.; Gonzalez, R.; Kang, J.Z.; McDermott, J.H.; King, A.J. Across-Species Differences in Pitch Perception Are Consistent with Differences in Cochlear Filtering. *eLife* **2019**, *8*, e41626, doi:10.7554/eLife.41626.
13. Walker, K.M.M.; Schnupp, J.W.H.; Hart-Schnupp, S.M.B.; King, A.J.; Bizley, J.K. Pitch Discrimination by Ferrets for Simple and Complex Sounds. *J. Acoust. Soc. Am.* **2009**, *126*, 1321–1335, doi:10.1121/1.3179676.
14. Wang, X.; Walker, K.M.M. Neural Mechanisms for the Abstraction and Use of Pitch Information in Auditory Cortex. *J. Neurosci.* **2012**, *32*, 13339–13342, doi:10.1523/JNEUROSCI.3814-12.2012.
15. O'Sullivan, J.; Herrero, J.; Smith, E.; Schevon, C.; McKhann, G.M.; Sheth, S.A.; Mehta, A.D.; Mesgarani, N. Hierarchical Encoding of Attended Auditory Objects in Multi-Talker Speech Perception. *Neuron* **2019**, *104*, 1195–1209.e3, doi:10.1016/j.neuron.2019.09.007.
16. Morrill, R.J.; Bigelow, J.; DeKloe, J.; Hasenstaub, A.R. Audiovisual Task Switching Rapidly Modulates Sound Encoding in Mouse Auditory Cortex. *eLife* **2022**, *11*, e75839, doi:10.7554/eLife.75839.

17. O'Connell, M.N.; Barczak, A.; Schroeder, C.E.; Lakatos, P. Layer Specific Sharpening of Frequency Tuning by Selective Attention in Primary Auditory Cortex. *J. Neurosci.* **2014**, *34*, 16496–16508, doi:10.1523/JNEUROSCI.2055-14.2014.
18. Carlyon, R.P.; Cusack, R.; Foxton, J.M.; Robertson, I.H. Effects of Attention and Unilateral Neglect on Auditory Stream Segregation. *J. Exp. Psychol. Hum. Percept. Perform.* **2001**, *27*, 115–127, doi:10.1037/0096-1523.27.1.115.
19. Okita, T. Selective Attention and Event-Related Potentials. *Jpn. J. Physiol. Psychol. Psychophysiol.* **1985**, *3*, 11–22, doi:10.5674/jjppp1983.3.11.
20. Shackleton, T.M.; Carlyon, R.P. The Role of Resolved and Unresolved Harmonics in Pitch Perception and Frequency Modulation Discrimination. *J. Acoust. Soc. Am.* **1994**, *95*, 3529–3540, doi:10.1121/1.409970.
21. Eliades, S.J.; Wang, X. Neural Substrates of Vocalization Feedback Monitoring in Primate Auditory Cortex. *Nature* **2008**, *453*, 1102–1106, doi:10.1038/nature06910.
22. Schneider, D.M.; Woolley, S.M.N. Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. *Neuron* **2013**, *79*, 141–152, doi:10.1016/j.neuron.2013.04.038.
23. Shamma, S.A.; Elhilali, M.; Micheyl, C. Temporal Coherence and Attention in Auditory Scene Analysis. *Trends Neurosci.* **2011**, *34*, 114–123, doi:10.1016/j.tins.2010.11.002.
24. Zion Golumbic, E.M.; Ding, N.; Bickel, S.; Lakatos, P.; Schevon, C.A.; McKhann, G.M.; Goodman, R.R.; Emerson, R.; Mehta, A.D.; Simon, J.Z.; et al. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party.” *Neuron* **2013**, *77*, 980–991, doi:10.1016/j.neuron.2012.12.037.
25. Gatehouse, S.; Noble, W. The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int. J. Audiol.* **2004**, *43*, 85–99, doi:10.1080/14992020400050014.
26. Parthasarathy, A.; Hancock, K.E.; Bennett, K.; DeGruttola, V.; Polley, D.B. Bottom-up and Top-down Neural Signatures of Disordered Multi-Talker Speech Perception in Adults with Normal Hearing. *eLife* **2020**, *9*, e51419, doi:10.7554/eLife.51419.
27. Brokx, J.P.L.; Nootboom, S.G. Intonation and the Perceptual Separation of Simultaneous Voices. *J. Phon.* **1982**, *10*, 23–36, doi:10.1016/S0095-4470(19)30909-X.
28. Brungart, D.S.; Simpson, B.D.; Ericson, M.A.; Scott, K.R. Informational and Energetic Masking Effects in the Perception of Multiple Simultaneous Talkers. *J. Acoust. Soc. Am.* **2001**, *110*, 2527–2538, doi:10.1121/1.1408946.
29. Assmann, P.F.; Summerfield, Q. The Contribution of Waveform Interactions to the Perception of Concurrent Vowels. *J. Acoust. Soc. Am.* **1994**, *95*, 471–484, doi:10.1121/1.408342.
30. Assmann, P.F.; Summerfield, Q. Modeling the Perception of Concurrent Vowels: Vowels with Different Fundamental Frequencies. *J. Acoust. Soc. Am.* **1990**, *88*, 680–697, doi:10.1121/1.399772.
31. Culling, J.F.; Darwin, C.J. Perceptual Separation of Simultaneous Vowels: Within and Across-formant Grouping by F0. *J. Acoust. Soc. Am.* **1993**, *93*, 3454–3467, doi:10.1121/1.405675.
32. Zwicker, U.T. Auditory Recognition of Diotic and Dichotic Vowel Pairs. *Speech Commun.* **1984**, *3*, 265–277, doi:10.1016/0167-6393(84)90023-2.
33. Hartmann, W.M.; McAdams, S.; Smith, B.K. Hearing a Mistuned Harmonic in an Otherwise Periodic Complex Tone. *J. Acoust. Soc. Am.* **1990**, *88*, 1712–1724, doi:10.1121/1.400246.
34. Moore, B.C.J.; Glasberg, B.R.; Peters, R.W. Thresholds for Hearing Mistuned Partial as Separate Tones in Harmonic Complexes. *J. Acoust. Soc. Am.* **1986**, *80*, 479–483, doi:10.1121/1.394043.
35. Moore, B.C.J.; Peters, R.W.; Glasberg, B.R. Thresholds for the Detection of Inharmonicity in Complex Tones. *J. Acoust. Soc. Am.* **1985**, *77*, 1861–1867, doi:10.1121/1.391937.
36. Roberts, B.; Brunstrom, J.M. Perceptual Segregation and Pitch Shifts of Mistuned Components in Harmonic Complexes and in Regular Inharmonic Complexes. *J. Acoust. Soc. Am.* **1998**, *104*, 2326–2338, doi:10.1121/1.423771.
37. McPherson, M.J.; Grace, R.C.; McDermott, J.H. Harmonicity Aids Hearing in Noise. *Atten. Percept. Psychophys.* **2022**, *84*, 1016–1042, doi:10.3758/s13414-021-02376-0.
38. Popham, S.; Boebinger, D.; Ellis, D.P.W.; Kawahara, H.; McDermott, J.H. Inharmonic Speech Reveals the Role of Harmonicity in the Cocktail Party Problem. *Nat. Commun.* **2018**, *9*, 2122, doi:10.1038/s41467-018-04551-8.

39. Roberts, B.; Brunstrom, J.M. Perceptual Fusion and Fragmentation of Complex Tones Made Inharmonic by Applying Different Degrees of Frequency Shift and Spectral Stretch. *J. Acoust. Soc. Am.* **2001**, *110*, 2479–2490, doi:10.1121/1.1410965.
40. McPherson, M.J.; McDermott, J.H. Diversity in Pitch Perception Revealed by Task Dependence. *Nat. Hum. Behav.* **2017**, *2*, 52–66, doi:10.1038/s41562-017-0261-8.
41. Bregman, A.S.; Liao, C.; Levitan, R. Auditory Grouping Based on Fundamental Frequency and Formant Peak Frequency. *Can. J. Psychol.* **1990**, *44*, 400–413, doi:10.1037/h0084255.
42. Singh, P.G. Perceptual Organization of Complex-tone Sequences: A Tradeoff between Pitch and Timbre? *J. Acoust. Soc. Am.* **1987**, *82*, 886–899, doi:10.1121/1.395287.
43. Fishman, Y.I.; Steinschneider, M. Neural Correlates of Auditory Scene Analysis Based on Inharmonicity in Monkey Primary Auditory Cortex. *J. Neurosci.* **2010**, *30*, 12480–12494, doi:10.1523/JNEUROSCI.1780-10.2010.
44. Homma, N.Y.; Bajo, V.M.; Happel, M.F.K.; Nodal, F.R.; King, A.J. Mistuning Detection Performance of Ferrets in a Go/No-Go Task. *J. Acoust. Soc. Am.* **2016**, *139*, EL246–EL251, doi:10.1121/1.4954378.
45. Klinge, A.; Klump, G. Mistuning Detection and Onset Asynchrony in Harmonic Complexes in Mongolian Gerbils. *J. Acoust. Soc. Am.* **2010**, *128*, 280–290, doi:10.1121/1.3436552.
46. Lohr, B.; Dooling, R.J. Detection of Changes in Timbre and Harmonicity in Complex Sounds by Zebra Finches (*Taeniopygia guttata*) and Budgerigars (*Melopsittacus undulatus*). *J. Comp. Psychol.* **1998**, *112*, 36–47, doi:10.1037/0735-7036.112.1.36.
47. Itatani, N.; Klump, G.M. Animal Models for Auditory Streaming. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160112, doi:10.1098/rstb.2016.0112.
48. Dowling, W.J. The Perception of Interleaved Melodies. *Cognit. Psychol.* **1973**, *5*, 322–337, doi:10.1016/0010-0285(73)90040-6.
49. Izumi, A. Auditory Stream Segregation in Japanese Monkeys. *Cognition* **2002**, *82*, B113–B122, doi:10.1016/S0010-0277(01)00161-5.
50. Ma, L.; Micheyl, C.; Yin, P.; Oxenham, A.J.; Shamma, S.A. Behavioral Measures of Auditory Streaming in Ferrets (*Mustela putorius*). *J. Comp. Psychol.* **2010**, *124*, 317–330, doi:10.1037/a0018273.
51. MacDougall-Shackleton, S.A.; Hulse, S.H.; Gentner, T.Q.; White, W. Auditory Scene Analysis by European Starlings (*Sturnus vulgaris*): Perceptual Segregation of Tone Sequences. *J. Acoust. Soc. Am.* **1998**, *103*, 3581–3587, doi:10.1121/1.423063.
52. Fay, R.R. Auditory Stream Segregation in Goldfish (*Carassius auratus*). *Hear. Res.* **1998**, *120*, 69–76, doi:10.1016/S0378-5955(98)00058-6.
53. Geissler, D.B.; Ehret, G. Time-Critical Integration of Formants for Perception of Communication Calls in Mice. *Proc. Natl. Acad. Sci.* **2002**, *99*, 9021–9025, doi:10.1073/pnas.122606499.
54. Caporello Bluvas, E.; Gentner, T.Q. Attention to Natural Auditory Signals. *Hear. Res.* **2013**, *305*, 10–18, doi:10.1016/j.heares.2013.08.007.
55. Schwartz, Z.P.; David, S.V. Focal Suppression of Distractor Sounds by Selective Attention in Auditory Cortex. *Cereb. Cortex* **2018**, *28*, 323–339, doi:10.1093/cercor/bhx288.
56. Rodgers, C.C.; DeWeese, M.R. Neural Correlates of Task Switching in Prefrontal Cortex and Primary Auditory Cortex in a Novel Stimulus Selection Task for Rodents. *Neuron* **2014**, *82*, 1157–1170, doi:10.1016/j.neuron.2014.04.031.
57. Moore, B.C.J.; Gockel, H.E. Resolvability of Components in Complex Tones and Implications for Theories of Pitch Perception. *Hear. Res.* **2011**, *276*, 88–97, doi:10.1016/j.heares.2011.01.003.
58. Carlyon, R.P. Masker Asynchrony Impairs the Fundamental-Frequency Discrimination of Unresolved Harmonics. *J. Acoust. Soc. Am.* **1996**, *99*, 525–533, doi:10.1121/1.414511.
59. Carlyon, R.P. Encoding the Fundamental Frequency of a Complex Tone in the Presence of a Spectrally Overlapping Masker. *J. Acoust. Soc. Am.* **1996**, *99*, 517–524, doi:10.1121/1.414510.
60. Shofner, W.P.; Campbell, J. Pitch Strength of Noise-Vocoded Harmonic Tone Complexes in Normal-Hearing Listeners. *J. Acoust. Soc. Am.* **2012**, *132*, EL398–EL404, doi:10.1121/1.4757697.
61. Shofner, W.P.; Chaney, M. Processing Pitch in a Nonhuman Mammal (*Chinchilla laniger*). *J. Comp. Psychol.* **2013**, *127*, 142–153, doi:10.1037/a0029734.

62. Song, X.; Osmanski, M.S.; Guo, Y.; Wang, X. Complex Pitch Perception Mechanisms Are Shared by Humans and a New World Monkey. *Proc Natl Acad Sci U S A* **2016**, *113*, 781–786, doi:10.1073/pnas.1516120113.
63. Vliegen, J.; Moore, B.C.J.; Oxenham, A.J. The Role of Spectral and Periodicity Cues in Auditory Stream Segregation, Measured Using a Temporal Discrimination Task. *J. Acoust. Soc. Am.* **1999**, *106*, 938–945, doi:10.1121/1.427140.
64. Vliegen, J.; Oxenham, A.J. Sequential Stream Segregation in the Absence of Spectral Cues. *J. Acoust. Soc. Am.* **1999**, *105*, 339–346, doi:10.1121/1.424503.
65. Grimault, N.; Micheyl, C.; Carlyon, R.P.; Arthaud, P.; Collet, L. Influence of Peripheral Resolvability on the Perceptual Segregation of Harmonic Complex Tones Differing in Fundamental Frequency. *J. Acoust. Soc. Am.* **2000**, *108*, 263–271, doi:10.1121/1.429462.
66. Madsen, S.M.K.; Dau, T.; Moore, B.C.J. Effect of Harmonic Rank on Sequential Sound Segregation. *Hear. Res.* **2018**, *367*, 161–168, doi:10.1016/j.heares.2018.06.002.
67. Madsen, S.M.K.; Dau, T.; Oxenham, A.J. No Interaction between Fundamental-Frequency Differences and Spectral Region When Perceiving Speech in a Speech Background. *PLOS ONE* **2021**, *16*, e0249654, doi:10.1371/journal.pone.0249654.
68. Tarka, V.M.; Gaucher, Q.; Walker, K.M.M. Pitch Selectivity in Ferret Auditory Cortex 2025.
69. Fishman, Y.I.; Micheyl, C.; Steinschneider, M. Neural Representation of Harmonic Complex Tones in Primary Auditory Cortex of the Awake Monkey. *J. Neurosci.* **2013**, *33*, 10312–10323, doi:10.1523/JNEUROSCI.0020-13.2013.
70. Itatani, N.; Klump, G.M. Neural Correlates of Auditory Streaming of Harmonic Complex Sounds With Different Phase Relations in the Songbird Forebrain. *J. Neurophysiol.* **2011**, *105*, 188–199, doi:10.1152/jn.00496.2010.
71. Knyazeva, S.; Selezneva, E.; Gorkin, A.; Aggelopoulos, N.C.; Brosch, M. Neuronal Correlates of Auditory Streaming in Monkey Auditory Cortex for Tone Sequences without Spectral Differences. *Front. Integr. Neurosci.* **2018**, *12*, doi:10.3389/fnint.2018.00004.
72. Dolležal, L.-V.; Itatani, N.; Günther, S.; Klump, G.M. Auditory Streaming by Phase Relations between Components of Harmonic Complexes: A Comparative Study of Human Subjects and Bird Forebrain Neurons. *Behav. Neurosci.* **2012**, *126*, 797–808, doi:10.1037/a0030249.
73. Roberts, B.; Glasberg, B.R.; Moore, B.C.J. Primitive Stream Segregation of Tone Sequences without Differences in Fundamental Frequency or Passband. *J. Acoust. Soc. Am.* **2002**, *112*, 2074–2085, doi:10.1121/1.1508784.
74. Oxenham, A.J. Questions and Controversies Surrounding the Perception and Neural Coding of Pitch. *Front. Neurosci.* **2023**, *16*, 1074752, doi:10.3389/fnins.2022.1074752.
75. Alain, C.; Reinke, K.; He, Y.; Wang, C.; Lobaugh, N. Hearing Two Things at Once: Neurophysiological Indices of Speech Segregation and Identification. *J. Cogn. Neurosci.* **2005**, *17*, 811–818, doi:10.1162/0898929053747621.
76. Alain, C.; Arnott, S.R.; Picton, T.W. Bottom-up and Top-down Influences on Auditory Scene Analysis: Evidence from Event-Related Brain Potentials. *J. Exp. Psychol. Hum. Percept. Perform.* **2001**, *27*, 1072–1089, doi:10.1037/0096-1523.27.5.1072.
77. Hautus, M.J.; Johnson, B.W. Object-Related Brain Potentials Associated with the Perceptual Segregation of a Dichotically Embedded Pitch. *J. Acoust. Soc. Am.* **2005**, *117*, 275–280, doi:10.1121/1.1828499.
78. Alain, C.; Reinke, K.; McDonald, K.L.; Chau, W.; Tam, F.; Pacurar, A.; Graham, S. Left Thalamo-Cortical Network Implicated in Successful Speech Separation and Identification. *NeuroImage* **2005**, *26*, 592–599, doi:10.1016/j.neuroimage.2005.02.006.
79. Hill, K.T.; Miller, L.M. Auditory Attentional Control and Selection during Cocktail Party Listening. *Cereb. Cortex* **2010**, *20*, 583–590, doi:10.1093/cercor/bhp124.
80. Horton, C.; Srinivasan, R.; D’Zmura, M. Envelope Responses in Single-Trial EEG Indicate Attended Speaker in a ‘Cocktail Party.’ *J. Neural Eng.* **2014**, *11*, 046015, doi:10.1088/1741-2560/11/4/046015.
81. Jaeger, M.; Mirkovic, B.; Bleichner, M.G.; Debener, S. Decoding the Attended Speaker From EEG Using Adaptive Evaluation Intervals Captures Fluctuations in Attentional Listening. *Front. Neurosci.* **2020**, *14*, 603, doi:10.3389/fnins.2020.00603.

82. Kerlin, J.R.; Shahin, A.J.; Miller, L.M. Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party." *J. Neurosci.* **2010**, *30*, 620–628, doi:10.1523/JNEUROSCI.3631-09.2010.
83. O'Sullivan, J.A.; Power, A.J.; Mesgarani, N.; Rajaram, S.; Foxe, J.J.; Shinn-Cunningham, B.G.; Slaney, M.; Shamma, S.A.; Lalor, E.C. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* **2015**, *25*, 1697–1706, doi:10.1093/cercor/bht355.
84. Power, A.J.; Foxe, J.J.; Forde, E.; Reilly, R.B.; Lalor, E.C. At What Time Is the Cocktail Party? A Late Locus of Selective Attention to Natural Speech. *Eur. J. Neurosci.* **2012**, *35*, 1497–1503, doi:10.1111/j.1460-9568.2012.08060.x.
85. Ding, N.; Simon, J.Z. Emergence of Neural Encoding of Auditory Objects While Listening to Competing Speakers. *Proc. Natl. Acad. Sci.* **2012**, *109*, 11854–11859, doi:10.1073/pnas.1205381109.
86. Kong, Y.-Y.; Mullangi, A.; Ding, N. Differential Modulation of Auditory Responses to Attended and Unattended Speech in Different Listening Conditions. *Hear. Res.* **2014**, *316*, 73–81, doi:10.1016/j.heares.2014.07.009.
87. Pasley, B.N.; David, S.V.; Mesgarani, N.; Flinker, A.; Shamma, S.A.; Crone, N.E.; Knight, R.T.; Chang, E.F. Reconstructing Speech from Human Auditory Cortex. *PLoS Biol.* **2012**, *10*, e1001251, doi:10.1371/journal.pbio.1001251.
88. Snyder, J.S.; Alain, C.; Picton, T.W. Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation. *J. Cogn. Neurosci.* **2006**, *18*, 1–13, doi:10.1162/089892906775250021.
89. Sussman, E.; Ritter, W.; Vaughan, H.G. An Investigation of the Auditory Streaming Effect Using Event-Related Brain Potentials. *Psychophysiology* **1999**, *36*, 22–34, doi:10.1017/s0048577299971056.
90. Wilson, E.C.; Melcher, J.R.; Micheyl, C.; Gutschalk, A.; Oxenham, A.J. Cortical fMRI Activation to Sequences of Tones Alternating in Frequency: Relationship to Perceived Rate and Streaming. *J. Neurophysiol.* **2007**, *97*, 2230–2238, doi:10.1152/jn.00788.2006.
91. Cusack, R. The Intraparietal Sulcus and Perceptual Organization. *J. Cogn. Neurosci.* **2005**, *17*, 641–651, doi:10.1162/0898929053467541.
92. Gutschalk, A.; Oxenham, A.J.; Micheyl, C.; Wilson, E.C.; Melcher, J.R. Human Cortical Activity during Streaming without Spectral Cues Suggests a General Neural Substrate for Auditory Stream Segregation. *J. Neurosci.* **2007**, *27*, 13074–13081, doi:10.1523/JNEUROSCI.2299-07.2007.
93. Larsen, E.; Cedolin, L.; Delgutte, B. Pitch Representations in the Auditory Nerve: Two Concurrent Complex Tones. *J. Neurophysiol.* **2008**, *100*, 1301–1319, doi:10.1152/jn.01361.2007.
94. Palmer, A.R. Segregation of the Responses to Paired Vowels in the Auditory Nerve of the Guinea-Pig Using Autocorrelation. In *The Auditory Processing of Speech*; Schouten, M.E., Ed.; DE GRUYTER MOUTON, 1992; pp. 115–124 ISBN 978-3-11-013589-3.
95. Sinex, D.G.; Guzik, H.; Li, H.; Henderson Sabes, J. Responses of Auditory Nerve Fibers to Harmonic and Mistuned Complex Tones. *Hear. Res.* **2003**, *182*, 130–139, doi:10.1016/S0378-5955(03)00189-8.
96. Tramo, M.J.; Cariani, P.A.; Delgutte, B.; Braida, L.D. Neurobiological Foundations for the Theory of Harmony in Western Tonal Music. *Ann. N. Y. Acad. Sci.* **2001**, *930*, 92–116, doi:10.1111/j.1749-6632.2001.tb05727.x.
97. Keilson, S.E.; Richards, V.M.; Wyman, B.T.; Young, E.D. The Representation of Concurrent Vowels in the Cat Anesthetized Ventral Cochlear Nucleus: Evidence for a Periodicity-Tagged Spectral Representation. *J. Acoust. Soc. Am.* **1997**, *102*, 1056–1071, doi:10.1121/1.419859.
98. Fishman, Y.I.; Micheyl, C.; Steinschneider, M. Neural Representation of Concurrent Vowels in Macaque Primary Auditory Cortex. *eNeuro* **2016**, *3*, ENEURO.0071-16.2016, doi:10.1523/ENEURO.0071-16.2016.
99. Bar-Yosef, O.; Nelken, I. The Effects of Background Noise on the Neural Responses to Natural Sounds in Cat Primary Auditory Cortex. *Front. Comput. Neurosci.* **2007**, *1*, doi:10.3389/neuro.10.003.2007.
100. Hamersky, G.R.; Shaheen, L.A.; Espejo, M.L.; Wingert, J.C.; David, S.V. Reduced Neural Responses to Natural Foreground versus Background Sounds in the Auditory Cortex. *J. Neurosci.* **2025**, *45*, e0121242024, doi:10.1523/JNEUROSCI.0121-24.2024.
101. Mesgarani, N.; David, S.V.; Fritz, J.B.; Shamma, S.A. Mechanisms of Noise Robust Representation of Speech in Primary Auditory Cortex. *Proc. Natl. Acad. Sci.* **2014**, *111*, 6792–6797, doi:10.1073/pnas.1318017111.

102. Moore, R.C.; Lee, T.; Theunissen, F.E. Noise-Invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise. *PLoS Comput. Biol.* **2013**, *9*, e1002942, doi:10.1371/journal.pcbi.1002942.
103. Rabinowitz, N.C.; Willmore, B.D.B.; King, A.J.; Schnupp, J.W.H. Constructing Noise-Invariant Representations of Sound in the Auditory Pathway. *PLoS Biol.* **2013**, *11*, e1001710, doi:10.1371/journal.pbio.1001710.
104. Saderi, D.; Buran, B.N.; David, S.V. Streaming of Repeated Noise in Primary and Secondary Fields of Auditory Cortex. *J. Neurosci.* **2020**, *40*, 3783–3798, doi:10.1523/JNEUROSCI.2105-19.2020.
105. Souffi, S.; Lorenzi, C.; Varnet, L.; Huetz, C.; Edeline, J.-M. Noise-Sensitive But More Precise Subcortical Representations Coexist with Robust Cortical Encoding of Natural Vocalizations. *J. Neurosci.* **2020**, *40*, 5228–5246, doi:10.1523/JNEUROSCI.2731-19.2020.
106. Fishman, Y.I.; Arezzo, J.C.; Steinschneider, M. Auditory Stream Segregation in Monkey Auditory Cortex: Effects of Frequency Separation, Presentation Rate, and Tone Duration. *J. Acoust. Soc. Am.* **2004**, *116*, 1656–1670, doi:10.1121/1.1778903.
107. Fishman, Y.I.; Reser, D.H.; Arezzo, J.C.; Steinschneider, M. Neural Correlates of Auditory Stream Segregation in Primary Auditory Cortex of the Awake Monkey. *Hear. Res.* **2001**, *151*, 167–187, doi:10.1016/S0378-5955(00)00224-0.
108. Micheyl, C.; Tian, B.; Carlyon, R.P.; Rauschecker, J.P. Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. *Neuron* **2005**, *48*, 139–148, doi:10.1016/j.neuron.2005.08.039.
109. Elhilali, M.; Ma, L.; Micheyl, C.; Oxenham, A.J.; Shamma, S.A. Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. *Neuron* **2009**, *61*, 317–329, doi:10.1016/j.neuron.2008.12.005.
110. Kanwal, J.S.; Medvedev, A.V.; Micheyl, C. Neurodynamics for Auditory Stream Segregation: Tracking Sounds in the Mustached Bat's Natural Environment. *Netw. Comput. Neural Syst.* **2003**, *14*, 413–435, doi:10.1088/0954-898X_14_3_303.
111. Noda, T.; Kanzaki, R.; Takahashi, H. Stimulus Phase Locking of Cortical Oscillation for Auditory Stream Segregation in Rats. *PLoS ONE* **2013**, *8*, e83544, doi:10.1371/journal.pone.0083544.
112. Pressnitzer, D.; Sayles, M.; Micheyl, C.; Winter, I.M. Perceptual Organization of Sound Begins in the Auditory Periphery. *Curr. Biol.* **2008**, *18*, 1124–1128, doi:10.1016/j.cub.2008.06.053.
113. Bee, M.A.; Klump, G.M. Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain. *J. Neurophysiol.* **2004**, *92*, 1088–1104, doi:10.1152/jn.00884.2003.
114. Allen, E.J.; Burton, P.C.; Olman, C.A.; Oxenham, A.J. Representations of Pitch and Timbre Variation in Human Auditory Cortex. *J. Neurosci.* **2017**, *37*, 1284–1293, doi:10.1523/JNEUROSCI.2336-16.2016.
115. Bizley, J.K.; Walker, K.M.M.; King, A.J.; Schnupp, J.W.H. Neural Ensemble Codes for Stimulus Periodicity in Auditory Cortex. *J. Neurosci.* **2010**, *30*, 5078–5091, doi:10.1523/JNEUROSCI.5475-09.2010.
116. Bizley, J.K.; Walker, K.M.M.; Silverman, B.W.; King, A.J.; Schnupp, J.W.H. Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. *J. Neurosci.* **2009**, *29*, 2064–2075, doi:10.1523/JNEUROSCI.4755-08.2009.
117. Gander, P.E.; Kumar, S.; Sedley, W.; Nourski, K.V.; Oya, H.; Kovach, C.K.; Kawasaki, H.; Kikuchi, Y.; Patterson, R.D.; Howard, M.A.; et al. Direct Electrophysiological Mapping of Human Pitch-Related Processing in Auditory Cortex. *NeuroImage* **2019**, *202*, 116076, doi:10.1016/j.neuroimage.2019.116076.
118. Griffiths, T.D.; Kumar, S.; Sedley, W.; Nourski, K.V.; Kawasaki, H.; Oya, H.; Patterson, R.D.; Brugge, J.F.; Howard, M.A. Direct Recordings of Pitch Responses from Human Auditory Cortex. *Curr. Biol.* **2010**, *20*, 1128–1132, doi:10.1016/j.cub.2010.04.044.
119. Steinschneider, M.; Reser, D.H.; Fishman, Y.I.; Schroeder, C.E.; Arezzo, J.C. Click Train Encoding in Primary Auditory Cortex of the Awake Monkey: Evidence for Two Mechanisms Subserving Pitch Perception. *J. Acoust. Soc. Am.* **1998**, *104*, 2935–2955, doi:10.1121/1.423877.
120. de Cheveigné, A. Concurrent Vowel Identification. III. A Neural Model of Harmonic Interference Cancellation. *J. Acoust. Soc. Am.* **1997**, *101*, 2857–2865, doi:10.1121/1.419480.
121. de Cheveigné, A.; McAdams, S.; Marin, C.M.H. Concurrent Vowel Identification. II. Effects of Phase, Harmonicity, and Task. *J. Acoust. Soc. Am.* **1997**, *101*, 2848–2856, doi:10.1121/1.419476.

122. de Cheveigné, A.; McAdams, S.; Laroche, J.; Rosenberg, M. Identification of Concurrent Harmonic and Inharmonic Vowels: A Test of the Theory of Harmonic Cancellation and Enhancement. *J. Acoust. Soc. Am.* **1995**, *97*, 3736–3748, doi:10.1121/1.412389.
123. de Cheveigné, A. Harmonic Cancellation—A Fundamental of Auditory Scene Analysis. *Trends Hear.* **2021**, *25*, doi:10.1177/23312165211041422.
124. Steinmetzger, K.; Rosen, S. No Evidence for a Benefit from Masker Harmonicity in the Perception of Speech in Noise. *J. Acoust. Soc. Am.* **2023**, *153*, 1064–1072, doi:10.1121/10.0017065.
125. Nocon, J.C.; Gritton, H.J.; James, N.M.; Mount, R.A.; Qu, Z.; Han, X.; Sen, K. Parvalbumin Neurons Enhance Temporal Coding and Reduce Cortical Noise in Complex Auditory Scenes. *Commun. Biol.* **2023**, *6*, 1–14, doi:10.1038/s42003-023-05126-0.
126. Homma, N.Y.; Happel, M.F.K.; Nodal, F.R.; Ohl, F.W.; King, A.J.; Bajo, V.M. A Role for Auditory Corticothalamic Feedback in the Perception of Complex Sounds. *J. Neurosci.* **2017**, *37*, 6149–6161, doi:10.1523/JNEUROSCI.0397-17.2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.