Current Biology *18*, 1124–1128, August 5, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.06.053

Report

Perceptual Organization of Sound Begins in the Auditory Periphery

Daniel Pressnitzer,[1,2,*] Mark Sayles,[3] Christophe Micheyl,[4] and Ian M. Winter[3]
[1]Laboratoire Psychologie de la Perception
Centre National de la Recherche Scientifique and Université
Paris Descartes
Paris F 75006
France
[2]Département d'Etudes Cognitives
Ecole Normale Supérieure
Paris F 75005
France
[3]Centre for the Neural Basis of Hearing
The Physiological Laboratory
Cambridge CB2 3EG
United Kingdom
[4]Auditory Perception and Cognition Laboratory
Department of Psychology
University of Minnesota
Minneapolis, Minnesota 55455

Summary

Segmenting the complex acoustic mixture that makes a typical auditory scene into relevant perceptual objects is one of the main challenges of the auditory system [1], for both human and nonhuman species. Several recent studies indicate that perceptual auditory object formation, or "streaming," may be based on neural activity within the auditory cortex and beyond [2, 3]. Here, we find that scene analysis starts much earlier in the auditory pathways. Single units were recorded from a peripheral structure of the mammalian auditory brainstem, the cochlear nucleus. Peripheral responses were similar to cortical responses and displayed all of the functional properties required for streaming, including multisecond adaptation. Behavioral streaming was also measured in human listeners. Neurometric functions derived from the peripheral responses predicted accurately behavioral streaming. This reveals that subcortical structures may already contribute to the analysis of auditory scenes. This finding is consistent with the observation that species lacking a neocortex can still achieve and benefit from behavioral streaming [4]. For humans, we argue that auditory scene analysis of complex scenes is probably based on interactions between subcortical and cortical neural processes, with the relative contribution of each stage depending on the nature of the acoustic cues forming the streams.

Results and Discussion

We usually experience our acoustic environment as containing multiple "streams" of sounds, which can be selectively attended to and followed over time amid other streams (e.g., the voice of a friend in a crowded restaurant, a musical instrument within an orchestra). Analogous to the segmentation of visual scenes into objects, the parsing of acoustic sequences into streams is an essential component of the perceptual analysis of auditory scenes in humans and various other animal species [1, 5–7].

Where and how auditory streaming is implemented in the brain are as-yet-unanswered questions, but a number of physiology and brain-imaging studies have suggested that the auditory cortex plays a key role in the formation of auditory streams [6–9]. The general form of the neural correlates found in these studies can be described as "grouping by coactivation": sounds that activate the same or largely overlapping populations of neurons are perceived as forming a single stream, whereas sounds that activate different neuronal populations are perceived as separate streams. For instance, when stimulated with pure tones, most neurons of the primary auditory cortex (A1) respond selectively only to a limited range of frequencies. This is consistent with a coactivation model, given that consecutive tones with similar frequencies are grouped in a single stream, whereas tones differing widely in frequency are heard as separate streams [3, 8, 10]. Similarly, forward suppression of activity could explain the increase in sound segregation with faster rates of tone presentation [8, 10]. A more challenging feature of streaming is that it can change dynamically over time, even if the stimulus itself remains constant (similarly to bistable perception in vision [11]). Predicting the dynamics of streaming is a crucial test for any neural model of streaming. Recently, it has been proposed that multisecond adaptation of neural responses in A1 could explain the behavioral "build-up" of stream segregation when the exposure time to a sound is increased [3].

Although the relationship between neural responses in the auditory cortex and auditory streaming is being thoroughly investigated, the possible contribution of subcortical nuclei has so far remained unexplored. The auditory system contains several subcortical nuclei, which are generally believed to establish basic feature encoding before perceptual organization starts at the cortical level [12, 13]. Here, we investigated whether subcortical neural processing may in fact also take an active part in auditory perceptual organization. Single neurons were recorded from the ventral part of the cochlear nucleus (CN) of urethane-anaesthetized guinea pigs. The CN is the most peripheral brainstem structure in the ascending auditory pathways and the site of the first obligatory synapse for all auditory-nerve fibers. Its role of interface between the auditory periphery (cochlea and auditory nerve) and the higher central auditory system (inferior colliculus and auditory cortex) makes it an ideal locus to examine the origin of neural correlates of auditory streaming. The CN is made up of a variety of physiologically and histologically well-defined cell types [14]. On the one hand, bushy cells display "primary-like" response properties similar to those of the auditory-nerve fibers from which they receive their input, thus providing a window on peripheral responses. On the other hand, the CN also contains cells, such as the multipolar cells, with "chopper-sustained" or "chopper-transient" response properties far more complex than those of the auditory nerve, and that can be thought of as initial brainstem processing of sound. Like A1 neurons, most cells in the CN exhibit frequency selectivity and forward suppression

[*] Correspondence: daniel.pressnitzer@ens.fr

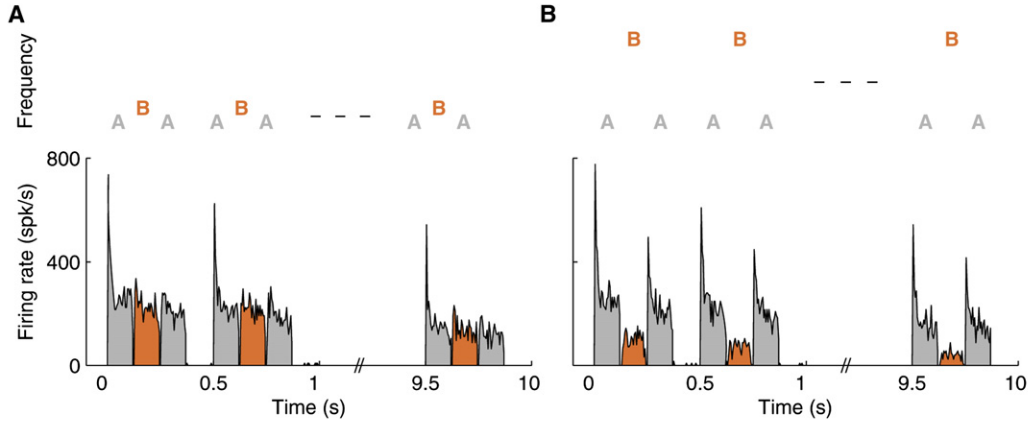

Figure 1. Illustration of the Sound Sequences and Cochlear Nucleus Single-Unit Responses

(A) The frequency difference, ΔF, between A and B tones is one semitone. Sequences of ABA tones were presented for 10 s, and the frequency of the A tone was chosen to be equal to the unit's best frequency. The neuron displayed here was classified as a multipolar cell with a transient-chopper response and a best frequency of 2.63 kHz (see Supplemental Experimental Procedures). The poststimulus time histograms (bin width: 5 ms) show that the unit responded to both A and B tones. In this case, listeners tend to hear a single stream.

(B) As (A), but ΔF is now six semitones. The responses to the B tones are reduced because of frequency selectivity and forward suppression. In this case, listeners have a probability of hearing two streams that increases over the duration of the sequence.

to a varying degree according to their response type [15]. So far, however, neural responses to long-duration sequences such as those used in psychophysical studies of auditory streaming have never been measured at the level of the CN.

To address this question, we used an experimental paradigm similar to the one used in earlier behavioral studies of auditory streaming in humans [16, 17] and in neurophysiological studies at the level of the auditory cortex [7–9, 18]. Sound stimuli were built with pure tones alternating between two frequencies, A and B, and arranged into repeating sequences of ABA triplets for a total of 10 s (ABA sequences; see Figures 1A and 1B and Supplemental Experimental Procedures available online). The percept evoked by these sound sequences depends on the frequency difference (ΔF) between the A and the B tones and on the time elapsed since the sequence is turned on. When the frequency difference is small, the sequence is perceived as a single coherent sound stream with a distinctive galloping rhythm (ABA-ABA). When frequency difference is large, the sequence is usually perceived as a single stream just after it is turned on, but after a few seconds of uninterrupted listening, it separates into two streams each with regular rhythms (stream A-A-A- and stream B-B-B-) [1, 16]. The change in percept from one stream to two streams is quite compelling and is experienced even by listeners who are aware that the physical stimulus does not change over time (online demonstrations at e.g., http://cognition.ens.fr/Audition/sup/).

An example response from a CN neuron to ABA sequences is illustrated in Figure 1; the population averages are shown in Figure 2. The frequency of the A tone was chosen equal to the neuron's best frequency (BF), and several values of ΔF were tested. Overall, responses of CN neurons closely resembled responses from single units in the primary auditory cortex [3, 10]. Importantly, they displayed all of the features of the grouping by coactivation model: At small ΔFs (e.g., 1 semitone, Figure 1A), CN neurons responded to both A and B tones, consistent with the grouped percept reported by listeners for such stimuli. As ΔF increased, neurons responded less and less to the tones that were remote from their BF (the B tones in our paradigm, Figure 1B). This result, just as in the cortex, is probably due to the combined effects of frequency selectivity and forward suppression of neural responses. The main static features of streaming are thus already apparent in the CN responses.

As mentioned above, a more challenging test for neural models of streaming relates to the dynamic percept changes that are experienced by listeners as the sequence is heard for a prolonged period of time [19]. We quantified these perceptual effects by asking normal-hearing listeners to report their percept ("one stream" or "two streams") continuously during the same 10 s stimulus sequences as the ones used for the physiology (Supplemental Experimental Procedures). The average reported percept plotted as a function of time from sequence onset shows that at all but the smallest ΔF, the proportion of two streams responses increases over time (Figures 3A and 3B). This build-up of segregation is faster and more pronounced at the largest ΔFs. It has been proposed that this build-up comes from multisecond adaptation of neural responses in the auditory cortex [3]. Here, we observed that neurons in the CN also display strong multisecond adaptation in response to the long-duration tone sequences. Both single neurons (Figure 1) and the population average (Figures 2A and 2B) showed a marked and progressive decrease in spike counts over the course of the 10 s stimulus sequence. This multisecond adaptation was present in the two main different types of cells in the ventral subdivision of the CN, including bushy cells that exhibit "primary-like" responses similar to those of auditory-nerve fibers. This shows that the multisecond adaptation observed in the auditory cortex is already present in the auditory periphery.

Adaptation over several seconds has been reported in the auditory nerve for continuous long-duration, single-frequency tones and was ascribed to neurotransmitter depletion at the synapse between hair cell and auditory-nerve fibers [20]. We simulated responses of auditory-nerve fibers to the ABA sequences using a representative model of the auditory periphery [21]. The model was chosen because it is fitted to the guinea pig's auditory periphery and it reproduces neural forward masking by means of synaptic depletion. The simulations are presented in Figure S1. The model does not exhibit multisecond adaptation. This indicates either that adaptation

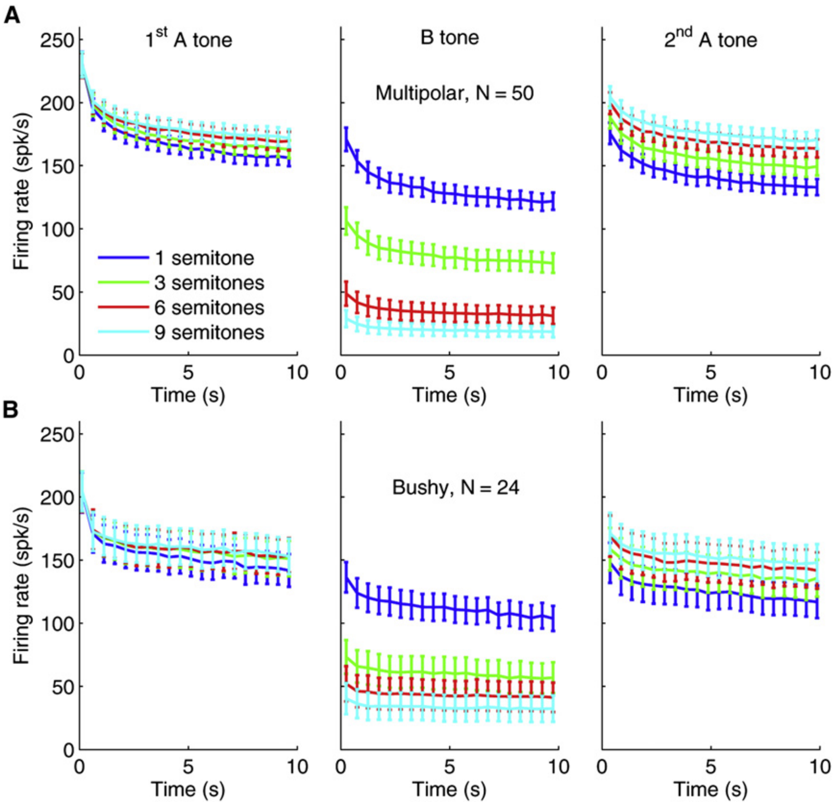

Figure 2. Multisecond Adaptation Is Present in the Cochlear Nucleus

(A) Firing rates are displayed for each tone of the triplets, as a function of the time within the sequence. Left, middle, and right panels show responses to the first A tone, B tone, and second A tone of the triplets, respectively. Single-unit firing rates were averaged for all of the multipolar cells of our sample (chopper-transient and chopper-sustained response types). Error bars represent ± 1 standard error around the mean. Each line represents a single frequency difference, ΔF, as identified in the figure legend. Multisecond adaptation is observed for all tones and all ΔFs.

(B) Same as (A), for bushy cells (primary-like response types).

to tone sequences emerges in the cochlear nucleus or that current models of the auditory nerve do not include the appropriate time constants for multisecond adaptation.

Adaptation in peripheral auditory neurons could also be influenced by descending feedback from upper processing stages, including the auditory cortex. It is highly unlikely that the multisecond adaptation we observe in all recorded neurons is a direct reflection of cortical adaptation because we recorded from the ventral part of the CN for which efferent connections are sparse [22]. It is possible, however, that the auditory cortex exerts a modulatory influence on CN activity, either via the sparse direct projections or via the more prevalent indirect projections. A possible pathway for indirect feedback is the medial olivocochlear efferent system, which can impose a form of slow gain-control on the cochlea [23] and thus on auditory-nerve and CN responses. In the VCN itself, subtle changes in adaptation are observed if feedback projections from the dorsal cochlear nucleus and medial olivary complex are removed [24, 25]. Considering the various possibilities, we suggest that multisecond adaptation to tone sequences in the VCN probably results from the interaction between long-term synaptic depression and fast recovery in peripheral neurons, with possible modulatory influences from descending projections. Whether multisecond adaptation is fully established in the periphery and simply reflected in the cortex or whether it requires an interaction between lower and higher levels in the auditory pathway remains an open question. In any case, our results show that the CN is involved in shaping this feature of auditory responses in ways not previously predicted.

The finding that neurons in the CN display frequency selectivity, forward suppression, and multisecond adaptation raises the interesting possibility that they can account quantitatively for the behavioral characteristics of auditory streaming. In order to test this possibility, we applied to the CN responses a grouping by coactivation model similar to the one proposed for A1 [3]. The model computes neurometric functions that can be compared directly with psychometric functions measured in human listeners (Supplemental Experimental Procedures). The basic idea of the model is that a one-stream percept is predicted if both A and B tones evoke an above-threshold

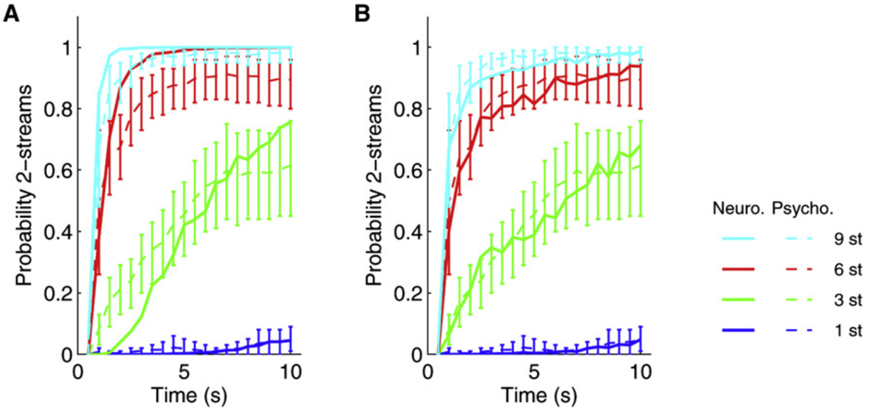

Figure 3. Responses from the Cochlear Nucleus Predict the Behavioral Build-Up of Streaming

(A) Neurometric (solid lines) functions for the multipolar cells subpopulation and psychometric functions in human listeners (dashed lines), for the ΔFs used in the experiment. Neurometric functions were estimated by a "grouping-by-coactivation" model, which predicts a one-stream percept if A and B tones recruit a same neuronal population and a two-stream percept otherwise (see Supplemental Experimental Procedures). Psychometric functions were obtained from normal-hearing human listeners. Error bars show 95% confidence intervals around the mean. There is a good correspondence between the two; neurometric functions are within the confidence intervals for the psychometric functions.

(B) Same as (A), but for the bushy cells subpopulation.

response in single neurons. In contrast, if neurons tuned to the A tones exhibit above-threshold activity during the presentation of the A tones but not during the presentation of the B tones, a two-streams percept is predicted. The average percept probability is finally obtained by tallying the model's binary decisions across a large number of simulated trials (here, 5000). Model predictions were computed for each triplet's position in the sequence, in each ΔF condition. The decision threshold in the model was adjusted to obtain the best fit between the psychometric data and the neural predictions, but it was not allowed to vary across ΔFs and triplets; therefore, variations in the predicted probability of two-streams responses as a function of these two parameters is due solely to neural-response characteristics and not to ad hoc changes in the model's threshold. The neurometric functions obtained with this procedure are presented in Figures 3A and 3B. The neurometric functions from CN neurons closely parallel the psychometric functions measured in humans. The level of agreement between neurometric and psychometric functions is just as high as that observed with cortical responses in a previous study [3]. The good fit obtained with only the bushy cells subpopulation (primary-like responses) also raises the possibility that the neural-response characteristics needed to predict the psychometric data may already be present at the level of the auditory nerve. In summary, our findings demonstrate that fundamental neural-response properties at early stages of the auditory system (frequency selectivity, forward suppression, and multisecond adaptation) can predict perceptual streaming for tone sequences. This extends to perceptual organization, the idea that adaptation is a key feature of sensory systems allowing for efficient encoding of information, as suggested by evidence in different sensory modalities [26, 27].

The present results challenge the current view that perceptual organization of sound only emerges at the level of the auditory cortex. Our findings, however, should not be interpreted as implying that the cortex plays no role in auditory scene analysis or that multisecond adaptation within frequency channels is the only mechanism of streaming. The tone sequences used here produce perceptual streaming on the basis of frequency differences, for which selectivity exists in the auditory periphery. Streaming, however, can also be observed between sounds that activate equivalently the same frequency channels but that have different temporal characteristics [28]. Under such circumstances, streaming must be based on temporal sound features that are extracted by mechanisms other than frequency selectivity, at subcortical [29] or cortical [30] levels of the auditory system. Moreover, in the general case, the sounds to be organized into streams will contain several frequency components and may overlap in time. The amount of overlap is a potent cue to auditory scene analysis, given that synchronous frequency components tend to be fused in a single-stream regardless of their frequency difference [1]. The grouping by coactivation model that we applied to the ABA sequences cannot capture these effects. It is however easy to extend the coactivation idea to the time dimension, so that a single stream is predicted if there is coactivation either in time (synchrony cue) or in frequency (neural channel cue). The neural implementation of such an extension probably requires neurons with broad receptive fields that perform frequency integration; these neurons can be found subcortically [31] and are abundant in the cortex [32]. Finally, streaming is affected by attention, context, and knowledge of the listener [16], and it is unclear whether and how such factors may influence responses at lower levels of the auditory system. Our findings must therefore be understood within the classic distinction between primitive versus schema-based processes in auditory scene analysis [1]. Neural-response properties, such as frequency selectivity, forward suppression, and multisecond adaptation, but also broadband inhibition [31], could mediate efficient primitive scene-analysis mechanisms in the auditory periphery. Other scene-analysis mechanisms, based on elaborate features or requiring plasticity, may rather involve the auditory cortex [12] and crossmodal [2] cortical regions. Humans' and other animals' remarkable ability to organize perceptually the complex mixtures of sounds encountered in natural environments is thus likely to recruit a distributed network involving interactions between subcortical and cortical neuronal processes. Such a distributed interaction might be an efficient way to achieve perceptual organization, not only for audition but also for other sensory modalities [33].

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures and one figure and can be found with this article online at http://www.current-biology.com/cgi/content/full/18/15/1124/DC1/.

Acknowledgments

This work was supported by Centre National de la Recherche Scientifique and a grant ANR-06-Neuro-022-01 (D.P.), National Institutes of Health grant RO1DC07657 (C.M.), the Biotechnology and Biological Sciences Research Council (I.M.W.), and the Frank Edward Elmore fund of the Cambridge clinical school MB PhD program (M.S.). The authors thank Josh McDermott and Ray Meddis for insightful discussions and suggestions on earlier versions of the manuscript.

Received: April 24, 2008
Revised: June 13, 2008
Accepted: June 18, 2008
Published online: July 24, 2008

References

1. Bregman, A. (1990). Auditory Scene Analysis (Cambridge, MA: MIT Press).
2. Cusack, R. (2005). The intraparietal sulcus and perceptual organization. J. Cogn. Neurosci. *17*, 641–651.
3. Micheyl, C., Tian, B., Carlyon, R.P., and Rauschecker, J.P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron *48*, 139–148.
4. Fay, R.R. (1998). Auditory stream segregation in goldfish (Carassius auratus). Hear. Res. *120*, 69–76.
5. Bee, M.A., and Klump, G.M. (2004). Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. J. Neurophysiol. *92*, 1088–1104.
6. Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., and Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. Hear. Res. *229*, 116–131.
7. Snyder, J.S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. Psychol. Bull. *133*, 780–799.
8. Fishman, Y.I., Reser, D.H., Arezzo, J.C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. *151*, 167–187.
9. Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., and Oxenham, A.J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. J. Neurosci. *25*, 5382–5388.
10. Fishman, Y.I., Arezzo, J.C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. *116*, 1656–1670.

11. Pressnitzer, D., and Hupe, J.M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr. Biol. *16*, 1351–1357.
12. Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. Curr. Opin. Neurobiol. *14*, 474–480.
13. Griffiths, T.D., and Warren, J.D. (2002). The planum temporale as a computational hub. Trends Neurosci. *25*, 348–353.
14. Young, E.D., and Oertel, D. (2003). The cochlear nucleus. In Synaptic Organization of the Brain, G.M. Shepherd, ed. (New York: Oxford University Press), pp. 125–163.
15. Bleeck, S., Sayles, M., Ingham, N.J., and Winter, I.M. (2006). The time course of recovery from suppression and facilitation from single units in the mammalian cochlear nucleus. Hear. Res. *212*, 176–184.
16. Cusack, R., Deeks, J., Aikman, G., and Carlyon, R.P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. Exp. Psychol. Hum. Percept. Perform. *30*, 643–656.
17. Moore, B.C.J., and Gockel, H. (2002). Factors influencing sequential stream segregation. Acta Acustica United with Acustica *88*, 320–332.
18. Wilson, E.C., Melcher, J., Micheyl, C., Gutschalk, A., and Oxenham, A.J. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: Relationship to perceived rate and streaming. J. Neurophysiol. *97*, 2230–2238.
19. Bregman, A.S. (1978). Auditory streaming is cumulative. J. Exp. Psychol. Hum. Percept. Perform. *4*, 380–387.
20. Javel, E. (1996). Long-term adaptation in cat auditory-nerve fiber responses. J. Acoust. Soc. Am. *99*, 1040–1052.
21. Meddis, R., and O'Mard, L.P. (2005). A computer model of the auditory-nerve response to forward-masking stimuli. J. Acoust. Soc. Am. *117*, 3787–3798.
22. Winer, J.A. (2006). Decoding the auditory corticofugal systems. Hear. Res. *212*, 1–8.
23. Sridhar, T.S., Liberman, M.C., Brown, M.C., and Sewell, W.F. (1995). A novel cholinergic "slow effect" of efferent stimulation on cochlear potentials in the guinea pig. J. Neurosci. *15*, 3667–3678.
24. Shore, S.E. (1998). Influence of centrifugal pathways on forward masking of ventral cochlear nucleus neurons. J. Acoust. Soc. Am. *104*, 378–389.
25. Mulders, W.H., Winter, I.M., and Robertson, D. (2002). Dual action of olivocochlear collaterals in the guinea pig cochlear nucleus. Hear. Res. *174*, 264–280.
26. Dean, I., Harper, N.S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. Nat. Neurosci. *8*, 1684–1689.
27. Fairhall, A.L., Lewen, G.D., Bialek, W., and de Ruyter Van Steveninck, R.R. (2001). Efficiency and ambiguity in an adaptive neural code. Nature *412*, 787–792.
28. Gutschalk, A., Oxenham, A.J., Micheyl, C., Wilson, E.C., and Melcher, J.R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. J. Neurosci. *27*, 13074–13081.
29. Winter, I.M., Wiegrebe, L., and Patterson, R.D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. J. Physiol. *537*, 553–566.
30. Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. Nature *436*, 1161–1165.
31. Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I.M. (2001). Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J. Neurosci. *21*, 6377–6386.
32. Schreiner, C.E., Read, H.L., and Sutter, M.L. (2000). Modular organization of frequency integration in primary auditory cortex. Annu. Rev. Neurosci. *23*, 501–529.
33. Leopold, D.A., and Maier, A. (2006). Neuroimaging: Perception at the brain's core. Curr. Biol. *16*, R95–R98.