



Published in final edited form as:

Semin Hear. 2011 May 1; 32(2): 129–141. doi:10.1055/s-0031-1277234.

Neural Encoding of Speech and Music: Implications for Hearing Speech in Noise

Samira Anderson^{1,2} and Nina Kraus, Ph.D.^{1,2,3,4}

¹Auditory Neuroscience Laboratory, Northwestern University, Evanston, Illinois

²Department of Communication Sciences, Northwestern University, Evanston, Illinois

³Department of Neurobiology and Physiology, Northwestern University, Evanston, Illinois

⁴Department of Otolaryngology, Northwestern University, Evanston, Illinois

Abstract

Understanding speech in a background of competing noise is challenging, especially for individuals with hearing loss or deficits in auditory processing ability. The ability to hear in background noise cannot be predicted from the audiogram, an assessment of peripheral hearing ability; therefore, it is important to consider the impact of central and cognitive factors on speech-in-noise perception. Auditory processing in complex environments is reflected in neural encoding of pitch, timing, and timbre, the crucial elements of speech and music. Musical expertise in processing pitch, timing, and timbre may transfer to enhancements in speech-in-noise perception due to shared neural pathways for speech and music. Through cognitive-sensory interactions, musicians develop skills enabling them to selectively listen to relevant signals embedded in a network of melodies and harmonies, and this experience leads in turn to enhanced ability to focus on one voice in a background of other voices. Here we review recent work examining the biological mechanisms of speech and music perception and the potential for musical experience to ameliorate speech-in-noise listening difficulties.

Keywords

Brain stem; music; speech in noise; timing; pitch

From busy classrooms to crowded restaurants, our world is a noisy place. To participate fully in today's society, we must selectively attend and listen to one voice among competing voices and other noises. Listening in noise is particularly challenging for clinical populations, such as children with learning impairments (i.e., dyslexia, autism spectrum disorder, attention deficit disorder, auditory processing disorder, specific language impairment)^{1–8} and older adults with and without hearing loss.^{9–13} The problem with hearing in noise is often not an issue of audibility—the individual can *hear* but cannot

Copyright © 2011 by Thieme Medical Publishers, Inc.

Address for correspondence and reprint requests: Nina Kraus, Ph.D., 2240 Campus Drive, Evanston, IL 60208 (nkraus@northwestern.edu).

Proceedings of the Widex Pediatric Audiology Congress; Guest Editor, André M. Marcoux, Ph.D.

understand what is being said. Other factors including central auditory processing and cognitive abilities interact with peripheral hearing status, determining the degree of successful communication in noise. To understand the message, the individual must process the three basic components of sound: pitch, timing, and timbre. Processing these elements is important for understanding both speech and music, and the musician's ability to tune into music (along with enhanced working memory and attention) appears to transfer to the ability to tune into speech in background noise. Music is said to be *biologically powerful*,^{14,15} with effects of musical training on neural functions lasting for a lifetime. The auditory system is malleable and adaptive, making constant changes based on immediate sensory and behavioral contexts and past experiences, including musical training.¹⁶ In this review, we summarize recent work examining the neural mechanisms of speech-in-noise (SIN) perception, including encoding of stimulus regularities, timing, and pitch, and the effects of musical experience on the neural encoding of speech, particularly in noise.

SUBCORTICAL APPROACH TO EXAMINING NEURAL SOUND ENCODING

The use of auditory brain stem responses to study the neural encoding of speech and music has been fruitful for several reasons. Unlike the cortical response, the brain stem response to complex stimuli (cABR) mimics the stimulus remarkably well, both visually and auditorially.^{17,18} The periodicity of the stimulus peaks can be seen in both the response waveform as well as the fast Fourier transform of the response (Fig. 1). In addition, when the brain stem response waveform is converted to a sound file, it actually sounds like the stimulus.¹⁸ The cABR is particularly suited for testing individuals as the response is highly reliable,¹⁹ and timing differences of fractions of milliseconds can be clinically significant.^{20–24} The brain stem faithfully represents the three main features of speech and music: timing (onsets/offsets and envelope of the response), pitch (encoding of the fundamental frequency), and timbre (harmonics).^{17,25} Moreover, the response originates primarily in the rostral brain stem, and the spectral and timing features such as those that contribute to the auditory perception of emotion are represented with remarkable fidelity.²⁶ This transparency in the response then allows for the examination of musician differences in the subcortical encoding of these features.

The brain stem is traditionally referred to as the “old brain” and is considered to be evolutionarily ancient, part of the reptilian brain. Do not be misled; the brain stem is “smart,” part of a reciprocally interactive sensory and cognitive network, engaging in multiple aspects of the nervous system. Current views of the brain stem are changing, and there is evidence that the brain stem can reveal much to us about the processing of auditory signals. The brain stem's multifaceted processing is evident given the anatomic complexity of the multiple nuclei and crossed and uncrossed pathways within the brain stem. In addition to the vast system of afferent fibers carrying sensory information to the midbrain (inferior colliculus) and the auditory cortex, there is an extensive system of descending efferent fibers that synapse all along the auditory pathway extending even to the outer hair cells of the basilar membrane.²⁷ In fact, the efferent fiber count may actually exceed the number of afferent fibers, which is definitely counterintuitive if the principal purpose of the auditory system is to passively convey acoustic information from the receptor to the auditory cortex for final and more complex processing. Indeed, top-down tuning of sensory function

modulates subcortical responses by sharpening tuning, augmenting stimulus features, increasing the signal-to-noise ratio, excluding irrelevant information, increasing response efficiency, controlling contextual influences, modulating plasticity, and promoting learning.^{26,28–31} For example, top-down modulation of the pathway from the cortex to the inferior colliculus has been effectively demonstrated in ferrets, in which pharmacological blockage impeded auditory learning.³⁰

The effects of musical training on auditory brain stem processing are another demonstration of top-down modulation, and these effects are not limited to musical sound processing but generalize to speech encoding and other nonmusical neural functions. The facets of learning involved in musical training provide an explanation for these effects. Musical training requires active listening and engagement with musical sounds, including interpretation or mapping *sounds to meaning*, as reviewed in Kraus and Chandrasekaran.¹⁶ For example, the dynamics of music (tempo, volume, etc.) are used to signal mood and can differentiate otherwise identical note patterns. The ability to attach meaning to dynamics and other characteristics of music requires attention to the fine-grained properties of music (pitch, timing, and timbre) as well as cognitive skills relating to working memory and attention. Many aspects of musical training focus attention on these specific aspects of sound: the violinist must listen to pitch cues to tune his or her instrument, the instrumentalist in an orchestra monitors the timing cues of other instruments, and the conductor relies on timbre cues to distinguish between instruments.

TRANSFER EFFECTS OF MUSICAL EXPERIENCE

Musical experience generalizes to enhanced function in other areas of the auditory system, with previous studies showing transfer effects to cortical, cerebellar, and other areas.¹⁶ For example, musicians demonstrate enhanced performance on auditory-based cognitive tasks, such as auditory memory or attention.^{32,33} Interestingly, this advantage is not seen in the visual domain, suggesting that the musicians have selective improvements in the auditory domain rather than improved overall neural function. Length of musical training also relates to measures of auditory attention, indicating that this improved ability is likely not innate but rather develops with experience (Fig. 2).³² The effects of musical experience extend all the way to the cochlea,^{34–37} with stronger top-down activity demonstrated by greater olivocochlear efferent suppression of biomechanical activity in musicians compared with nonmusicians.

At the subcortical level, musicians have higher brain stem amplitudes in response to both music *and* speech than nonmusicians.³⁸ In fact, musical training generalizes to linguistic pitch encoding, with musicians showing enhancements in subcortical pitch processing to the most complex vowel contours present in Mandarin Chinese, but not to simple contours,³⁹ demonstrating that the enhancement is selective rather than an overall gain effect. Another demonstration of selective enhancement is seen in subcortical responses to a musical chord, in which musicians have larger responses to the harmonics of the upper tone in a musical chord, which are important for melody perception, but do not have larger responses to the harmonics of the lower tone or the fundamental frequency (F_0) of either upper or lower tones.³¹ Similar results are found in response to emotional nuances in sound.²⁶ Compared

with nonmusicians, musicians have greater representation of the F_0 in the response corresponding to the complex portion of a baby's cry, but actually have smaller F_0 representation for the periodic portion (easier to lock onto), suggesting that this training also results in greater processing efficiency (Fig. 3).

HEARING SPEECH IN NOISE

Understanding speech in a background of competing talkers is a highly complex task, in many ways similar to the task of the musician who must segregate and selectively attend to competing melodies from many different instruments in an orchestra or ensemble. As sound travels from the ear to the mid brain to the primary auditory cortex, the signal is processed and shaped by a myriad of interacting factors. We have already noted that musicians have enhanced auditory memory and attention, cognitive factors that also contribute to enhanced SIN perception. As stated previously, music and speech both consist of the basic elements of pitch, timing, and timbre. Here we provide examples of how the subcortical representation of these elements relates to SIN perception.

Detection of Stimulus Regularities

The brain is constantly bombarded by stimuli, which must be processed and interpreted. To provide some perspective, consider the amount of data that enters the brain in just 1 second. It is roughly estimated that in the visual system, one million retinal ganglion cells fire at a rate of 1 to 10 spikes per second, which at one bit per spike adds up to 10 million bits of data per second. In the auditory system, 30,000 auditory nerve fibers fire at a rate of 0 to 300 spikes per second, or 9 million bits of data per second at 1 bit per spike. It is impossible to process this much information in real time; therefore, our neural systems must have a mechanism for managing and interpreting these data. One of the ways this is accomplished is through detection of patterns or stimulus regularities in an ongoing stream of data, which has been demonstrated in animal^{40,41} and human studies.^{42–46}

The brain stem's response to stimulus regularities was evaluated in a group of normal-learning children ages 8 to 12.⁴³ Fast Fourier transforms were calculated for responses recorded to a speech syllable [da] presented in a repetitive, unvarying condition and to the same syllable presented in a varying condition in which the [da] syllable was one of eight different consonant-vowel syllables. The representation of the lower harmonics (H_2 and H_4) was enhanced in the repetitive compared with the variable condition. Moreover, the extent of enhancement for the repetitive condition was related to a measure of SIN perception—scores on the Hearing in Noise Test (HINT; Biologic Systems Corp., Mundelein, IL; Fig. 4). Online adaptation and enhancement were also found for music (to a repeating melody) with the greatest amplitude increase occurring for the second repeated note within the melody.⁴⁷ This ability to detect regularities is also related to a test of musical aptitude—the Rhythm Aptitude subtest of Gordon's Intermediate Measures of Audiation test (GIA Publications, Chicago, IL)—indicating that active engagement with music may improve the ability to detect and encode sound patterns.⁴⁸ The brain stem's online detection of patterns, resulting in enhancement of predictable features of the auditory stimulus, is important for providing a stable representation of the stimulus for subsequent processing by the cortex, and a more stable representation in the cortex is likely more resistant to degradation by interfering noise.

Neural Timing

Timing is a crucial element for distinguishing speech sounds; for example, consonants such as [t] and [d] are differentiated solely by voice-onset time. The brain stem's representation of timing differences is exquisitely precise,⁴⁹ and it has been suggested that neural timing differences to speech stimuli can be used to assess children with reading disabilities or impairments in language or SIN perception.^{21–23,50,51}

The brain stem nuclei are tonotopic⁵²; therefore, brain stem responses to consonants differing in their formant trajectories during the transition from the consonant to the vowel (e.g., in the syllables [ba], [da], and [ga]) reflect timing differences between these consonants, with [ga] having the shortest peak latency, followed by [da] and then [ba]. These timing differences persist for the duration of the formant transition but disappear in the steady state, the unvarying region of the response corresponding to the vowel. The extent of these timing differences is related to SIN perception in children. Brain stem responses were recorded to the syllables [ba], [ga], and [da] in children ages 8 to 12, and a differentiation score was calculated based on how closely the responses followed the expected timing patterns. Higher differentiation scores were related to better SIN perception (based on HINT; Fig. 5).

Stop consonants are particularly vulnerable to the effects of noise, which causes delays in neural timing as well as decreases in response amplitudes.^{53–55} These delays are particularly evident in children with poor SIN perception. Brain stem responses to the syllable [da] were presented in quiet and in six-talker babble conditions to children who were divided into top and bottom SIN perception groups based on the HINT. The children in the bottom group had greater delays in response timing compared with those in the top group. For both groups, the noise effect was greatest shortly after the stimulus onset, delaying the peak latencies by a full millisecond. In the top SIN group, the latency shift then reduced and leveled off by 40 milliseconds into the response, but in the bottom SIN group the latency shift did not level off until 60 milliseconds. The first 60 milliseconds of the response correspond to the formant transition of the syllable, the most perceptually vulnerable region of the response,^{21,56–58} in which the formant frequencies are dynamic. By 60 milliseconds, the formant frequencies are static, making this region of the syllable easier to process (Fig. 6). Therefore, it is not surprising that the greatest group differences were found in the transition region of the response.

Pitch Representation

Pitch is another important element for SIN perception, allowing the listener to identify the speaker and to “tag” the speaker's voice, thus facilitating the ability to follow one speaker's voice from competing voices in a noisy environment such as a crowded restaurant or cocktail party. The F_0 and lower harmonics, along with other acoustic ingredients, contribute to the perception of pitch.^{59,60} Psychophysical experiments have demonstrated that F_0 differences facilitate the ability to distinguish between competing vowels or sentences.^{61–66}

The perceptual benefit provided by F_0 separation is apparent in neural processing.⁶⁷ Both children and young adults who have greater F_0 magnitudes in their cABRs also have better

SIN perception.^{68,69} Song et al found that in young adults, the top SIN group (based on the Speech-in-Noise test [Quick-SIN]; Etymotic Research, Elmhurst, IL) had greater F_0 magnitudes than the bottom SIN group in response to the speech syllable [da] presented in background babble.¹⁹ This difference in F_0 encoding was seen in the transition but not in the steady-state region of the response. The formant transition carries phonetic information⁷⁰ and its perception can easily be disrupted by noise.^{56,58} Similarly, in children, the top SIN group (based on HINT scores) had higher F_0 magnitudes than the bottom group in the neural response corresponding to the formant transition.⁶⁹

SIN Model

Structural equation modeling was used to quantify the relationships among factors contributing to SIN perception in children.⁷¹ The brain stem construct variable consisted of three factors: degree of neural timing shift, enhancement of stimulus regularities, and magnitude of pitch cues (e.g., F_0 and lower harmonics), and the SIN construct was based on HINT scores. This model predicted 56% of the variance in SIN perception, with a p value of 0.033 (Fig. 7). Ongoing work using structural equation modeling to evaluate these relationships in older adults will include cognitive factors and peripheral hearing status.

EFFECTS OF MUSICAL TRAINING

We know that musicians have advantages for encoding both music and speech, but do these advantages extend to SIN perception? As mentioned previously, musicians are particularly adept at extracting relevant signals from a soundscape; therefore, does this skill transfer to the ability to selectively attend to one voice from among competing voices? This musician advantage has been demonstrated for SIN perception as well as auditory working memory,⁷² a skill that is highly related to SIN perception.^{73,74} In a group of young adults, musicians had higher scores on SIN assessments (HINT and Quick-SIN) as well as working memory (cluster score from the Woodcock-Johnson Test III Test of Cognitive Abilities⁷⁵) compared with nonmusicians, providing further evidence that musical training transfers to nonmusical domains. Importantly, years of musical practice correlated with SIN perception, supporting the idea that the extent of training contributes to the advantage for hearing in noise.⁷²

A neural basis for this musician benefit has been demonstrated in the subcortical encoding of sounds. Background (babble) noise had a less degradative effect on brain stem responses in musicians than in nonmusicians.⁷⁶ The brain stem responses for musicians and nonmusicians were similar in quiet, but noise had a greater (disruptive) effect on the morphology, size, timing, and frequency content of nonmusician's responses than of musicians. Parbery-Clark et al quantified this effect using stimulus-to-response correlations, which compared the timing and morphology of the response waveform to the stimulus waveform.⁷² The stimulus-to-response correlations were essentially equivalent in quiet, but in noise the musician's r value changed only slightly, and the non-musician's r value dropped significantly. This musician difference also was seen for neural timing: onset and transition peaks were more delayed in noise for nonmusicians than for musicians, and these latency delays were related to SIN perception (Fig. 8). Overall, these results demonstrate that musicians have perceptual and neural advantages for processing SIN.

CONCLUSION

Subcortical responses to complex stimuli can be used to assess auditory processing ability in individuals suspected of deficits in SIN perception. Factors such as the neural enhancement stimulus regularities, strength of neural timing (differentiation of stop consonants and decreased noise-induced delays), and representation of the F_0 are important neural signatures associated with better ability to hear in noise. Musicians have enhanced SIN abilities, both perceptually and biologically, indicating that musical training may ameliorate the deleterious effects of background noise on speech understanding. Years of musical practice are related to SIN perception, suggesting that music may be an effective vehicle for developing the auditory skills necessary for hearing well in background noise. Although the examples provided in this review represented lifelong experience, there is also evidence of subcortical malleability following short-term training.^{77–83} Future work is needed to determine whether musically based auditory training can be used as a management strategy for improving SIN perception in children and older adults who experience particular challenges in background noise.

Acknowledgments

This work was funded by the National Institutes of Health (RO1 DC01510) and the National Science Foundation SGER (0842376). We would especially like to thank the members of the Auditory Neuroscience Laboratory and our participants for their contributions to this work.

References

1. Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res.* 2003; 46(1):80–97. [PubMed: 12647890]
2. Geffner D, Lucker JR, Koch W. Evaluation of auditory discrimination in children with ADD and without ADD. *Child Psychiatry Hum Dev.* 1996; 26(3):169–179. [PubMed: 8819879]
3. Lagacé J, Jutras B, Gagné JP. Auditory processing disorder and speech perception problems in noise: finding the underlying origin. *Am J Audiol.* 2010; 19(1):17–25. [PubMed: 20308289]
4. Moore DR, Ferguson MA, Edmondson-Jones AM, Ratib S, Riley A. Nature of auditory processing disorder in children. *Pediatrics.* 2010; 126(2):e382–e390. [PubMed: 20660546]
5. Alcántara JI, Weisblatt EJ, Moore BC, Bolton PF. Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *J Child Psychol Psychiatry.* 2004; 45(6):1107–1114. [PubMed: 15257667]
6. Ziegler JC, Pech-Georgel C, George F, Lorenzi C. Speech-perception-in-noise deficits in dyslexia. *Dev Sci.* 2009; 12(5):732–745. [PubMed: 19702766]
7. Ziegler JC, Pech-Georgel C, George F, Alario F-X, Lorenzi C. Deficits in speech perception predict language learning impairment. *Proc Natl Acad Sci U S A.* 2005; 102(39):14110–14115. [PubMed: 16162673]
8. Lewis D, Hoover B, Choi S, Stelmachowicz P. Relationship between speech perception in noise and phonological awareness skills for children with normal hearing. *Ear Hear.* 2010; 31(6):761–768. [PubMed: 20562623]
9. Cruickshanks KJ, Wiley TL, Tweed TS, et al. The Epidemiology of Hearing Loss Study. Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. *Am J Epidemiol.* 1998; 148(9):879–886. [PubMed: 9801018]
10. Humes LE. Speech understanding in the elderly. *J Am Acad Audiol.* 1996; 7(3):161–167. [PubMed: 8780988]

11. Souza PE, Boike KT, Witherell K, Tremblay K. Prediction of speech recognition from audibility in older listeners with hearing loss: effects of age, amplification, and background noise. *J Am Acad Audiol*. 2007; 18(1):54–65. [PubMed: 17252958]
12. Gordon-Salant S. Hearing loss and aging: new research findings and clinical implications. *J Rehabil Res Dev*. 2005; 42(4 Suppl 2):9–24. [PubMed: 16470462]
13. Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am*. 1984; 76(1):87–96. [PubMed: 6747116]
14. Patel, A. *Music, Language, and the Brain*. New York, NY: Oxford University Press, Inc; 2010.
15. Peretz I. The nature of music from a biological perspective. *Cognition*. 2006; 100(1):1–32. [PubMed: 16487953]
16. Kraus N, Chandrasekaran B. Music training for the development of auditory skills. *Nat Rev Neurosci*. 2010; 11(8):599–605. [PubMed: 20648064]
17. Skoe E, Kraus N. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear*. 2010; 31(3):302–324. [PubMed: 20084007]
18. Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*. 1995; 6(17):2363–2367. [PubMed: 8747154]
19. Song JH, Nicol T, Kraus N. Test-retest reliability of the speech-evoked auditory brainstem response. *Clin Neurophysiol*. 2010 In press Corrected Proof.
20. Hall, J. *New Handbook of Auditory Evoked Responses*. Boston, MA: Allyn & Bacon; 2007.
21. Basu M, Krishnan A, Weber-Fox C. Brainstem correlates of temporal auditory processing in children with specific language impairment. *Dev Sci*. 2010; 13(1):77–91. [PubMed: 20121865]
22. Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Reading and subcortical auditory function. *Cereb Cortex*. 2009; 19(11):2699–2707. [PubMed: 19293398]
23. Billiet CR, Bellis TJ. The relationship between brainstem temporal processing and performance on tests of central auditory function in children with reading disorders. *J Speech Lang Hear Res*. 2011; 54(1):228–242. [PubMed: 20689038]
24. Wible B, Nicol T, Kraus N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol*. 2004; 67(3):299–317. [PubMed: 15294388]
25. Dhar S, Abel R, Hornickel J, et al. Exploring the relationship between physiological measures of cochlear and brainstem function. *Clin Neurophysiol*. 2009; 120(5):959–966. [PubMed: 19346159]
26. Strait DL, Kraus N, Skoe E, Ashley R. Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci*. 2009; 29(3):661–668. [PubMed: 19222564]
27. Gao E, Suga N. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. *Proc Natl Acad Sci U S A*. 2000; 97(14):8081–8086. [PubMed: 10884432]
28. Suga N, Xiao Z, Ma X, Ji W. Plasticity and corticofugal modulation for hearing in adult animals. *Neuron*. 2002; 36(1):9–18. [PubMed: 12367501]
29. Luo F, Wang Q, Kashani A, Yan J. Corticofugal modulation of initial sound processing in the brain. *J Neurosci*. 2008; 28(45):11615–11621. [PubMed: 18987197]
30. Bajo VM, Nodal FR, Moore DR, King AJ. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci*. 2010; 13(2):253–260. [PubMed: 20037578]
31. Lee KM, Skoe E, Kraus N, Ashley R. Selective subcortical enhancement of musical intervals in musicians. *J Neurosci*. 2009; 29(18):5832–5840. [PubMed: 19420250]
32. Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res*. 2010; 261(1–2):22–29. [PubMed: 20018234]
33. Chan AS, Ho Y-C, Cheung M-C. Music training improves verbal memory. *Nature*. 1998; 396(6707):128–128. [PubMed: 9823892]

34. Micheyl C, Carbonnel O, Collet L. Medial olivocochlear system and loudness adaptation: differences between musicians and non-musicians. *Brain Cogn.* 1995; 29(2):127–136. [PubMed: 8573328]
35. Micheyl C, Khalfa S, Perrot X, Collet L. Difference in cochlear efferent activity between musicians and non-musicians. *Neuroreport.* 1997; 8(4):1047–1050. [PubMed: 9141090]
36. Perrot X, Micheyl C, Khalfa S, Collet L. Stronger bilateral efferent influences on cochlear biomechanical activity in musicians than in non-musicians. *Neurosci Lett.* 1999; 262(3):167–170. [PubMed: 10218882]
37. Brashears SM, Morlet TG, Berlin CI, Hood LJ. Olivocochlear efferent suppression in classical musicians. *J Am Acad Audiol.* 2003; 14(6):314–324. [PubMed: 14552425]
38. Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci U S A.* 2007; 104(40):15894–15898. [PubMed: 17898180]
39. Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci.* 2007; 10(4):420–422. [PubMed: 17351633]
40. Toro JM, Trobalón JB. Statistical computations over a speech stream in a rodent. *Percept Psychophys.* 2005; 67(5):867–875. [PubMed: 16334058]
41. Hauser MD, Newport EL, Aslin RN. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition.* 2001; 78(3):B53–B64. [PubMed: 11124355]
42. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science.* 1996; 274(5294):1926–1928. [PubMed: 8943209]
43. Chandrasekaran B, Hornickel J, Skoe E, Nicol TG, Kraus N. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron.* 2009; 64(3):311–319. [PubMed: 19914180]
44. Sussman ES, Bregman AS, Wang WJ, Khan FJ. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cogn Affect Behav Neurosci.* 2005; 5(1):93–110. [PubMed: 15913011]
45. Sussman E, Steinschneider M. Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Res.* 2006; 1075(1):165–174. [PubMed: 16460703]
46. Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD. Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci.* 2001; 4(6):633–637. [PubMed: 11369945]
47. Skoe E, Kraus N. Hearing it again and again: online subcortical plasticity in humans. *PLoS ONE.* 2010; 5(10):e13645. [PubMed: 21049035]
48. Strait, DL.; Ashley, R.; Hornickel, J.; Kraus, N. Context-dependent encoding of speech in the human auditory brainstem as a marker of musical aptitude. Presented at: ARO 33rd MidWinter Meeting; 2010; Anaheim, CA.
49. Tzounopoulos T, Kraus N. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron.* 2009; 62(4):463–469. [PubMed: 19477149]
50. Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci U S A.* 2009; 106(31):13022–13027. [PubMed: 19617560]
51. Anderson S, Skoe E, Chandrasekaran B, Kraus N. Neural timing is linked to speech perception in noise. *J Neurosci.* 2010; 30(14):4922–4926. [PubMed: 20371812]
52. Kandler K, Clause A, Noh J. Tonotopic reorganization of developing auditory brainstem circuits. *Nat Neurosci.* 2009; 12(6):711–717. [PubMed: 19471270]
53. Burkard RF, Sims D. A comparison of the effects of broadband masking noise on the auditory brainstem response in young and older adults. *Am J Audiol.* 2002; 11(1):13–22. [PubMed: 12227352]
54. Russo N, Nicol T, Musacchia G, Kraus N. Brainstem responses to speech syllables. *Clin Neurophysiol.* 2004; 115(9):2021–2030. [PubMed: 15294204]
55. Cunningham J, Nicol TG, Zecker SG, Bradlow AR, Kraus N. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol.* 2001; 112(5):758–767. [PubMed: 11336890]

56. Tallal P, Stark RE. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. *J Acoust Soc Am*. 1981; 69(2):568–574. [PubMed: 7462479]
57. Tallal P, Piercy M. Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*. 1974; 12(1):83–93. [PubMed: 4821193]
58. Hedrick MS, Younger MS. Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise. *J Speech Lang Hear Res*. 2007; 50(2):254–269. [PubMed: 17463228]
59. Fellowes JM, Remez RE, Rubin PE. Perceiving the sex and identity of a talker without natural vocal timbre. *Percept Psychophys*. 1997; 59(6):839–849. [PubMed: 9270359]
60. Meddis R, O'Mard L. A unitary model of pitch perception. *J Acoust Soc Am*. 1997; 102(3):1811–1820. [PubMed: 9301058]
61. Bird, J.; Darwin, CJ. Effects of a difference in fundamental frequency in separating two sentences. In: Palmer, AR., et al., editors. *Psychophysical and Physiological Advances in Hearing*. London, United Kingdom: Whurr; 1998. p. 263-269.
62. Assmann PF, Summerfield Q. Perceptual segregation of concurrent vowels. *J Acoust Soc Am*. 1987; 82:S120.
63. Culling JF, Darwin CJ. Perceptual separation of simultaneous vowels: within and across-formant grouping by F0. *J Acoust Soc Am*. 1993; 93(6):3454–3467. [PubMed: 8326071]
64. Oxenham AJ. Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends Amplif*. 2008; 12(4):316–331. [PubMed: 18974203]
65. Brokx JP, Nootboom SG. Intonation and the perceptual separation of simultaneous voices. *J Phonetics*. 1982; 10:23–26.
66. de Cheveigne A. Concurrent vowel identification III: a neural model of harmonic interference cancellation. *J Acoust Soc Am*. 1997; 101:2857–2865.
67. 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(5):811–818. [PubMed: 15904547]
68. Song J, Skoe E, Banai K, Kraus N. Perception of speech in noise: neural correlates. *J Cogn Neurosci*. 2010; 0:1–12.
69. Anderson S, Skoe E, Chandrasekaran B, Zecker S, Kraus N. Brainstem correlates of speech-in-noise perception in children. *Hear Res*. 2010; 270(1–2):151–157. [PubMed: 20708671]
70. Redford MA, Diehl RL. The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *J Acoust Soc Am*. 1999; 106(3 Pt 1):1555–1565. [PubMed: 10489711]
71. Hornickel J, Chandrasekaran B, Zecker S, Kraus N. Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. *Behav Brain Res*. 2011; 216(2):597–605. [PubMed: 20826187]
72. Parbery-Clark A, Skoe E, Lam C, Kraus N. Musician enhancement for speech-in-noise. *Ear Hear*. 2009; 30(6):653–661. [PubMed: 19734788]
73. Pichora-Fuller MK, Souza PE. Effects of aging on auditory processing of speech. *Int J Audiol*. 2003; 42(Suppl 2):S11–S16.
74. Heinrich A, Schneider BA, Craik FI. Investigating the influence of continuous babble on auditory short-term memory performance. *Q J Exp Psychol (Colchester)*. 2008; 61(5):735–751.
75. Woodcock, RW.; McGrew, KS.; Mather, N. *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing; 2001.
76. Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci*. 2009; 29(45):14100–14107. [PubMed: 19906958]
77. Song JH, Skoe E, Wong PCM, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci*. 2008; 20(10):1892–1902. [PubMed: 18370594]
78. de Boer J, Thornton ARD. Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J Neurosci*. 2008; 28(19):4929–4937. [PubMed: 18463246]

79. Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. *Behav Brain Res.* 2005; 156(1):95–103. [PubMed: 15474654]
80. Suga N, Ma X. Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat Rev Neurosci.* 2003; 4(10):783–794. [PubMed: 14523378]
81. Kumar AU, Hegde M, Mayaleela. Perceptual learning of non-native speech contrast and functioning of the olivocochlear bundle. *Int J Audiol.* 2010; 49(7):488–496. [PubMed: 20528666]
82. Carcagno S, Plack CJ. Subcortical plasticity following perceptual learning in a pitch discrimination task. *J Assoc Res Otolaryngol.* 2011; 12(1):89–100. [PubMed: 20878201]
83. Russo NM, Hornickel J, Nicol T, Zecker S, Kraus N. Biological changes in auditory function following training in children with autism spectrum disorders. *Behav Brain Funct.* 2010; 6:60. [PubMed: 20950487]

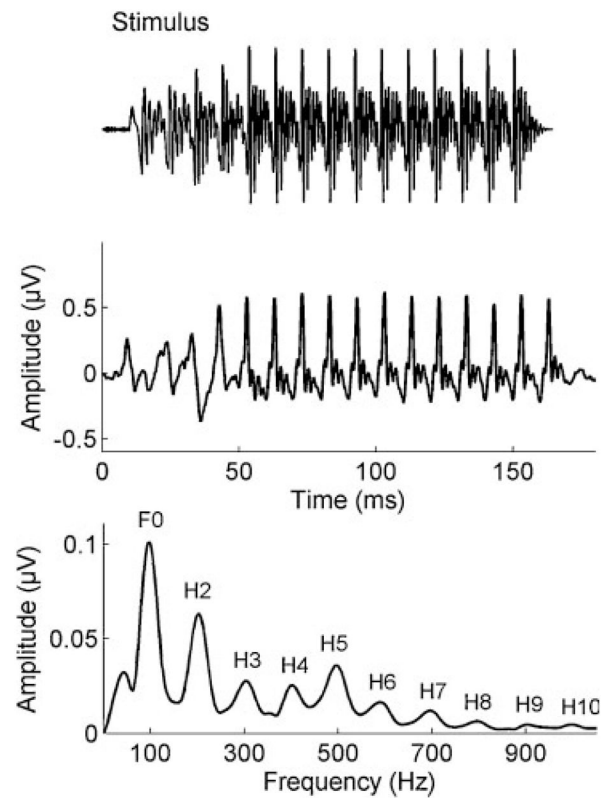


Figure 1.

The grand average brain stem response waveform obtained from 38 typically developing children (middle) is visually similar to the [da] stimulus waveform (top). The spectrum of the brain stem response contains energy at the F_0 and its integer multiples (bottom). Modified from Anderson et al, *Hear Res* 2010.⁶⁹

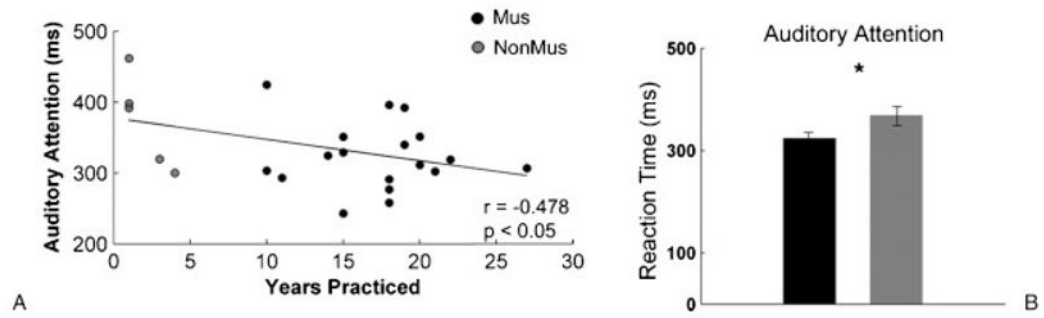


Figure 2.

(Left) Years of musical practice are related to a measure of auditory attention (subtest of the IHR Multicentre Battery for Auditory Processing; IMAP; Medical Research Council Institute of Hearing Research, Nottingham, UK). (Right) Musicians (black) have faster reaction times on the auditory attention subtest than nonmusicians (gray) ($*p < 0.05$). Modified from Strait et al, *Hear Res* 2010.³²

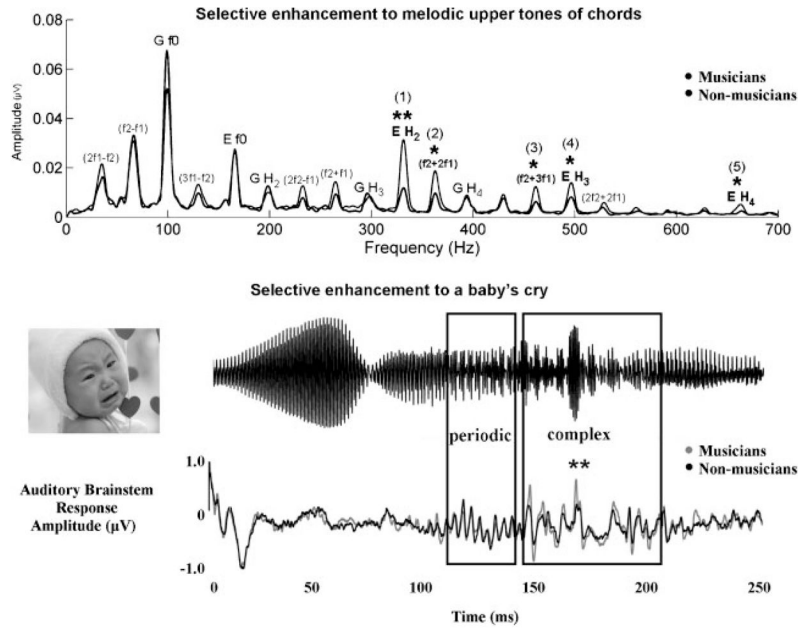


Figure 3. (Top) Compared with nonmusicians (black), musicians (grey) have higher amplitudes for the harmonics of the upper tones of a musical chord ($p < 0.05$, $**p < 0.01$). Notably, this difference was not seen for the fundamental frequency. (Bottom) Musicians again have higher amplitudes in response to the complex portion of the baby's cry compared with nonmusicians, but this enhancement was not seen in response to the periodic region of the response ($**p < 0.01$). Modified from Lee et al, J Neurosci 2009³¹; Strait et al, Eur J Neurosci 2009.²⁶

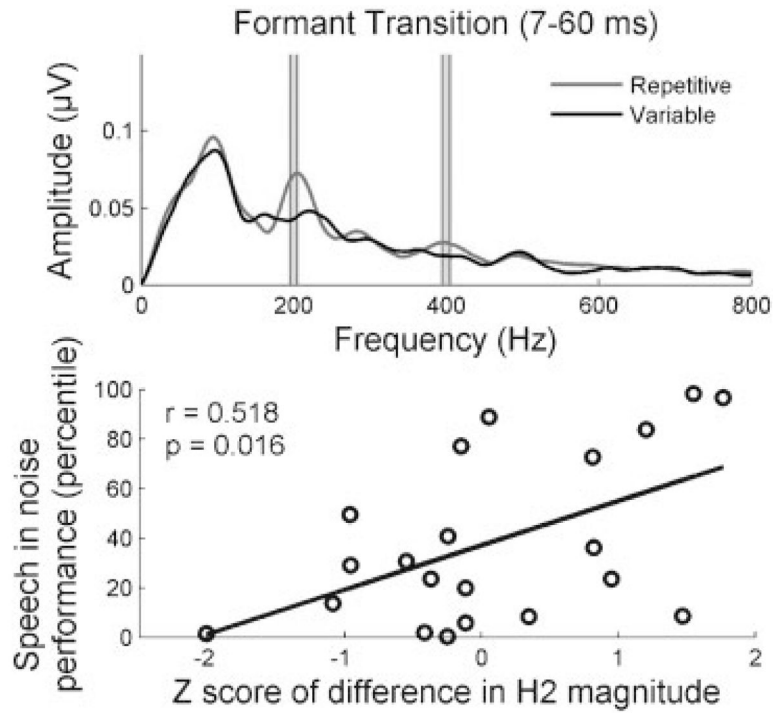


Figure 4.

The ability to take advantage of stimulus regularities is a factor in speech-in-noise (SIN) perception. (Top) Brain stem responses of typically developing children are enhanced in the repetitive, regularly occurring presentation of the [da] syllable (grey) compared with the variable presentation (black). (Bottom) The degree of enhancement in the repetitive condition is related to SIN performance (Hearing in Noise Test). Modified from Chandrasekaran et al, *Neuron* 2009.⁴³

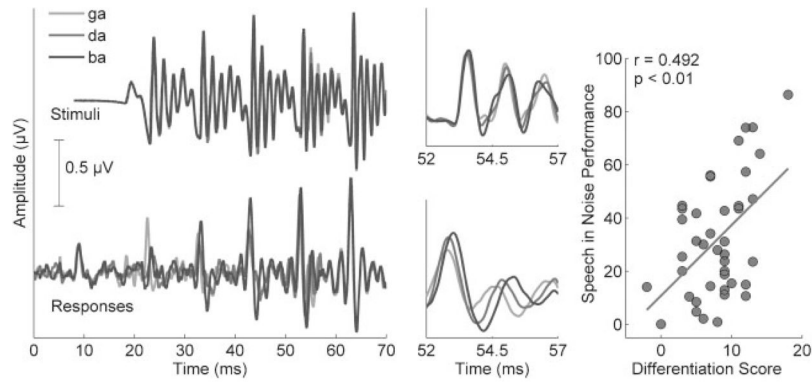


Figure 5.

Time domain grand average responses to three speech syllables ([ga], [da], and [ba]) of 20 typically developing children (bottom panels) reflect differences in stimulus timing and cochlear tonotopicity that are not seen in the stimulus waveforms (top panels). The 52- to 57-millisecond region is magnified to highlight these timing differences. The scatter plot on the right demonstrates the relationship between speech-in-noise performance (Hearing in Noise Test [HINT]) and differentiation scores (the degree to which the response latencies of the three syllables correspond to the expected pattern). Modified from Hornickel et al, *Proc Natl Acad Sci U S A* 2009.⁵⁰

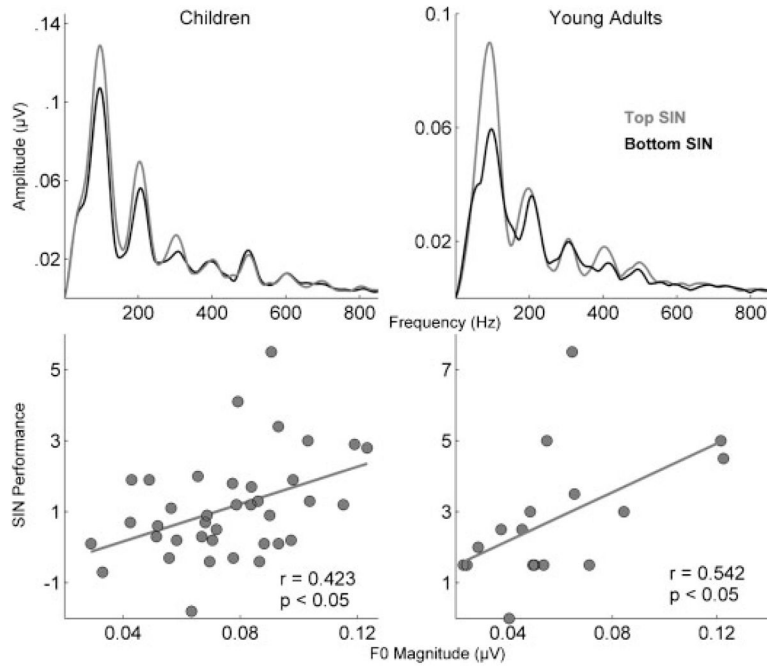


Figure 6.

The magnitude of F_0 is a factor in good speech-in-noise (SIN) perception. (Top) The top SIN perceivers in a group of 38 children (based on Hearing-in-Noise Test [HINT]) and a group of 17 young adults (based on QuickSIN) had higher F_0 magnitudes than the bottom groups ($*p < 0.05$) during the formant transition in brain stem responses to the speech syllable [da] presented in quiet (children) and in background babble (young adults). (Bottom) In both children and young adults, F_0 magnitude is related to scores on clinical measures of SIN perception (children, HINT; adults, QuickSIN). Modified from Anderson et al, *Hear Res* 2010⁶⁹ and Song et al, *J Cog Neurosci* 2010.⁶⁸

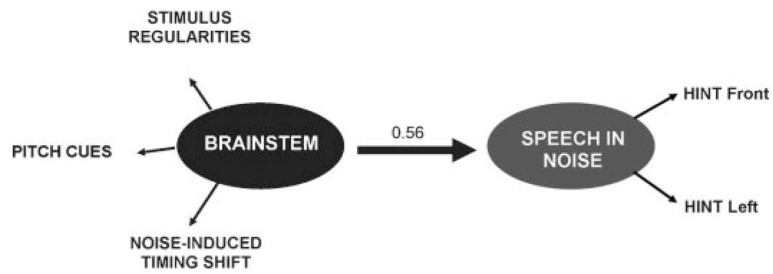


Figure 7.

Brain stem processing of speech (including enhancement of stimulus regularities, magnitude of pitch cues, and the degree of noise induced timing shifts) predicted 56% of the variance in speech-in-noise perception (based on Hearing-in-Noise Test, front and Left scores) using a structural equation model. Modified from Hornickel et al, Behav Brain Res 2011.⁷¹

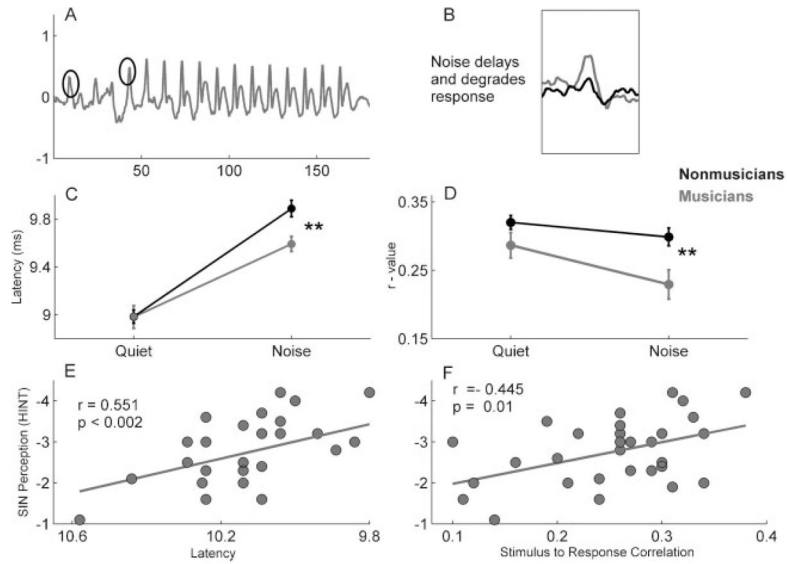


Figure 8.

In response to the syllable [da], peak timing is less delayed by noise (six-talker babble) in musicians (black) compared with nonmusicians (grey), and the overall morphology (assessed by the degree to which the response correlates with the stimulus) is less degraded by noise in musicians than in nonmusicians. (A) Grand average responses of 15 young adult nonmusicians. The circled peaks correspond to the onset and transition peaks of the response that are more delayed in noise in nonmusicians compared with musicians. (B) Noise both delays and degrades the response. (C) In quiet, the onset peaks are essentially equivalent between groups, but noise causes a greater latency delay in nonmusicians (** $p < 0.01$). (D) The stimulus-to-response correlation r values are essentially equivalent in quiet between groups, but noise causes greater degradation, as indicated by a decrease in r value in nonmusicians compared with musicians (** $p < 0.01$). When speech is presented in noise, Hearing-in-Noise Test scores are related to subcortical onset peak latencies (E) as well as stimulus-to-response correlation r -values (F). Modified from Parbery-Clark et al, J Neurosci 2009.⁷⁶