Hearing Research 352 (2017) 49–69

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Review article

Speech-in-noise perception in musicians: A review

Emily B.J. Coffey [a, b, c, *], Nicolette B. Mogilever [a], Robert J. Zatorre [a, b, c]

[a] Montreal Neurological Institute, McGill University, Quebec, Canada
[b] International Laboratory for Brain, Music and Sound Research (BRAMS), Quebec, Canada
[c] Centre for Research on Brain, Language and Music (CRBLM), Quebec, Canada

ARTICLE INFO

Article history:
Received 26 October 2016
Received in revised form
1 February 2017
Accepted 5 February 2017
Available online 14 February 2017

Keywords:
Speech-in-noise
Musician
Auditory system
Neuroimaging
Experience-dependent plasticity

ABSTRACT

The ability to understand speech in the presence of competing sound sources is an important neuroscience question in terms of how the nervous system solves this computational problem. It is also a critical clinical problem that disproportionally affects the elderly, children with language-related learning disorders, and those with hearing loss. Recent evidence that musicians have an advantage on this multifaceted skill has led to the suggestion that musical training might be used to improve or delay the decline of speech-in-noise (SIN) function. However, enhancements have not been universally reported, nor have the relative contributions of different bottom-up versus top-down processes, and their relation to preexisting factors been disentangled. This information that would be helpful to establish whether there is a real effect of experience, what exactly is its nature, and how future training-based interventions might target the most relevant components of cognitive processes. These questions are complicated by important differences in study design and uneven coverage of neuroimaging modality. In this review, we aim to systematize recent results from studies that have specifically looked at musician-related differences in SIN by their study design properties, to summarize the findings, and to identify knowledge gaps for future work.

Contents

* Corresponding author. 3801 rue University, Montreal Neurological Institute, McGill University, Quebec, H3A 1A1, Canada.
E-mail address: emz.coffey@mail.mcgill.ca (E.B.J. Coffey).

http://dx.doi.org/10.1016/j.heares.2017.02.006
0378-5955/

Abbreviations

ASA	auditory scene analysis
EEG	electroencephalography
FFR	frequency-following response
fMRI	functional magnetic resonance imaging
f0	fundamental frequency
H2, H3, H4, H5	2nd, 3rd, 4th and 5th harmonics
MEG	magnetoencephalography

1. Introduction

Speech-in-noise (SIN) perception, or the 'cocktail party phenomenon', may be considered a special case of auditory scene analysis (ASA) - the ability to parse complex acoustic scenes into coherent objects or sources, which involves the auditory, motor, and sometimes visual systems as they act to separate target speech from irrelevant sound (Bregman, 1994). SIN ability varies considerably even within healthy normal populations (Assmann and Summerfield, 2004). SIN deficits that impede daily function and affect quality of life are prevalent in the elderly (Parbery-Clark et al., 2011a,b) and in some paediatric populations (such as those with language-related learning disorders (Ziegler et al., 2005), making it an important topic both for improving our fundamental understanding of how the auditory system processes sound, and as a practically important matter.

SIN performance has sometimes been reported to be better among groups of musicians, but there is still controversy about this claim (Boebinger et al., 2015). It has been suggested that musical training might be used to improve the auditory system in ways that support and improve SIN perception, due to strengthening of shared resources (for a review of training studies that relate to SIN in elderly populations, see Alain et al., 2014). SIN perception appears to be supported by both the fidelity of bottom-up sound encoding (reviewed in Du et al., 2011; Anderson and Kraus, 2010) and the influence of higher-level processes such as auditory working memory (Kraus et al., 2012). Such effects may be confounded by genetic influences (Schellenberg, 2015), behavioural traits such as personality (Corrigall et al., 2013), motivation (McAuley et al., 2011), and the interactions between factors (Anderson et al., 2013). Although genetic and epigenetic factors are likely to contribute to musical and SIN-relevant cognition (Schellenberg, 2015), training studies on SIN perception (Alain et al., 2014) as well as a larger body of work on experience-dependent plasticity in the auditory system (e.g. Pantev and Herholz, 2011; de Villers-Sidani et al., 2008; Bidelman and Alain 2015) suggest that training can provide long-lasting biological benefits to auditory function, including simple perceptual enhancements, and even other functions that are necessary for higher-order cognition, like working memory and intelligence (reviewed in Moreno and Bidelman, 2014; Herholz and Zatorre, 2012).

These findings are encouraging as they demonstrate redundancy and flexibility in the neural machinery of auditory perception and might be clinically exploited. However, the mechanisms by which musical training might improve complex auditory skills like SIN perception are not yet well understood. In particular, it is not yet clear which processes and representations vary or are being modified, and which aspects of training are responsible when improvements are observed. The complexities inherent to understanding this problem are illustrated by work that shows interactions between demographic variables such as age with SIN subprocesses (Anderson et al., 2013), which suggest that different people rely on different cues and cognitive strategies to enable SIN performance. Tasks that are used clinically and in research to gauge SIN ability result in very different estimates of people's relative SIN scores, suggesting that small variations in the design of the listening tasks affect the degree to which individuals can solve SIN problems (Wilson et al., 2007; Parbery-Clark et al., 2012b). Furthermore, studies that record neurophysiological responses such as functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), in addition to behaviour in musicians are few and have been limited to a handful of specific SIN task designs, making for an incomplete understanding of the neural mechanisms supporting SIN subtasks.

Several dozen recent studies have explored the putative musician enhancement in SIN perception, but these vary in their task design. One means of disentangling the possible musical enhancements on complex SIN behaviour is to consider first what exactly is being asked of a cognitive system by the nature of the task that is presented to it (Coffey and Herholz, 2013), starting with considering what information the system is offered. Auditory stream segregation, including for natural language, offers numerous auditory cues on which the elements of a target can be separated, including spatial location, spectral and temporal regularity, and modulation (Pressnitzer et al., 2011; Moore and Gockel, 2002). It is influenced by attention (Thompson et al., 2011), and facilitated by information from vision (Suied et al., 2009) and processes of motor planning (Du et al., 2014). Predictive factors are also at work: SIN performance is known to be affected by learned knowledge including language syntax and semantics (Golestani et al., 2009; Pickering and Garrod, 2007), familiarity with the speaker's vocal timbre (Souza et al., 2013; Barker and Newman, 2004; Yonan and Sommers, 2000), and prior knowledge of the target (Bey and McAdams, 2002; Agus et al., 2010; McDermott et al., 2011), which can be used to predict, constrain, and evaluate the interpretation of incoming information (Bendixen, 2014).

The properties of both the target and the distracting stream are therefore important to how much the system will be disturbed by the distractor, and how it can reconstruct sound sources and separate out the target sound. These characteristics vary considerably among existing studies of SIN advantages in musicianship; for example, listeners may be offered whole sentences, single words or phonemes as targets, and these may be masked by noise that is made up of similar frequencies but low information content (i.e. energetic masking) as opposed to competing information-rich sound streams, like a second talker (i.e. informational masking). These differences must be taken into account if we are to better understand which aspects of SIN perception might be enhanced by experience.

In this review we first consider how SIN tasks may vary, situating each existing study along two of the most important dimensions that characterize different studies: acoustic cue richness of target versus that of noise, and we summarize the evidence in

favour of a musician advantage. We then consider coverage of the research area by different neuroimaging measures that offer complementary views of the neuophysiological basis of SIN differences between musicians and non-musicians. We suggest a means of comparing and planning future studies based on the details of their SIN task design. Finally, we highlight specific research areas in which further study would improve our understanding of SIN advantages in musicians.

2. Literature review

We restricted the scope of our literature review to studies of musician effects on SIN perception within neurologically normal populations. The literature search was executed in August 2016. Google Scholar was used to identify articles by keyword, using the following search terms: speech-in-noise, HINT, QuickSIN, auditory brainstem response, musicians/non-musicians, evoked response potentials, auditory perception, cocktail party, and background noise. Twenty studies met our criteria; the main experimental design characteristics and results for these studies are summarized in Table 1. We included nine additional studies that had not looked explicitly at SIN perception (noted in Table 1), but had studied musician enhancements in sub-skills and processes that are highly relevant to SIN processing. These are included to populate the lower range of possible listening conditions in Fig. 1; e.g. Parbery-Clark et al. studied whether musical training could offset the negative impact of aging on neural encoding of speech sounds presented in silence (2012a,b). Both the neural correlates and aging have been linked to SIN performance. However, we do not attempt to comprehensively cover studies of music-related enhancements in basic sound processing (Du et al., 2011) nor non-musical training for SIN perception improvement (Song et al., 2012). The majority of studies are cross-sectional in nature i.e. comparing musicians and non-musicians, or correlating a behavioural or neurophysiological dependent variable with a measure of musical experience such as years of practice, though several have used longitudinal training designs (Slater et al., 2015; Tierney et al., 2015) or have demonstrated a learning effect during the study (Varnet et al., 2015), which allows for some causal inference.

3. Cue richness of target and distractor sound streams

Natural SIN perception includes a large variety of potential acoustic, auditory-visual, and context-related cues that might be used to predict incoming information, but the two properties that have been most thoroughly explored in the context of musical enhancement are the acoustic properties of the target and the distractor/noise. Fig. 1 shows the studies reviewed in this paper positioned along these two dimensions. Further information about study design (and the distinctions between similar conditions that fall into the same rough categories) can be referenced in Table 1, by number and first author.

3.1. Coverage of target/distractor cue richness by existing studies

Three of the most common tasks used clinically to evaluate SIN perception (in anglophones) are the hearing-in-noise task (HINT; Nilsson, 1994), QuickSIN (Killion et al., 2004), and the words-in-noise task (WIN; Wilson, 2003; Wilson et al., 2007). These tasks, and minor variations created for reasons of experimental constraints or to address specific questions, have also been used in research; these are visible in Fig. 1 as clusters of studies. Several studies have used sentences in speech-shaped noise, which includes the HINT and its variants (Ruggles et al., 2014; Boebinger et al., 2015; Slater et al., 2015; Strait and Kraus, 2011; Coffey et al., 2016; Parbery-Clark et al., 2011a,b; Parbery-Clark et al., 2009a,b). Others have used sentences with a single-talker distractor (Boebinger et al., 2015; Parbery-Clark et al., 2011a,b; Slater and Kraus, 2016) or embedded in multi-talker babble (Slater and Kraus, 2016; Parbery-Clark et al., 2009a,b, Parbery-Clark et al., 2011a,b) which is similar to the QuickSIN task. As compared to HINT sentences, QuickSIN sentences are slightly longer (mean 8.6 words versus mean 5.3 words), they are less semantically predictable, and more complex vocabulary is used (e.g., HINT: 'A boy fell from the window' versus QuickSIN: 'A cruise in warm waters in a sleek yacht is fun'; Wilson et al., 2007), and are therefore thought to rely to a greater extent on auditory working memory (Parbery-Clark et al., 2009a,b). A smaller cluster of studies is found wherein individual words are identified in multitalker babble, which is similar to the WIN task (Parbery-Clark et al., 2011a,b; Slater and Kraus, 2016; Zendel et al., 2015).

Another group of studies looked at musician enhancements in basic sound processing in silent conditions (Zendel et al., 2015; Parbery-Clark et al., 2011a,b; Coffey et al., 2016; Bidelman and Weiss, 2014; Parbery-Clark et al., 2012; Coffey et al., 2016; Fuller et al., 2014; Lee et al., 2009); these studies offer some insight into the neural processes that operate on sound under ideal listening conditions. Several other studies have varied distractor noise types, for example, Swaminathan used backwards speech in order to manipulate the amount of informational masking while keeping the amount of energetic masking relatively constant (Swaminathan et al., 2015). A few studies used tones as maskers and/or targets, which is useful to address research questions about the involvement of pitch processing and auditory attention in SIN performance (Fuller et al., 2014; Strait and Kraus, 2011; Oxenham et al., 2003; Zendel and Alain, 2009). We have not identified any studies of musician SIN advantages that use naturalistic environmental noises, though these have been studied in relation to auditory stream analysis in other populations (e.g. restaurant noise in cochlear implant recipients, or realistic background classroom noise in children; Klatte et al., 2009; Gifford and Revit, 2010).

It is noteworthy that almost all studies reviewed here used native English speakers and English-language tests. Two studies used Dutch sentences in native Dutch speakers (Başkent and Gaudrain, 2016; Fuller et al., 2014), and two used French phonemes and monosyllabic words in French speakers (Zendel et al., 2015; Varnet et al., 2015). All three of these languages originated in north-western Europe and are closely related through common linguistic ancestry and mutual influence – and are quite unrelated to the world's two most spoken languages, Mandarin Chinese and Hindi. Languages differ in their linguistic properties, for example, tonal languages like Mandarin use changes in pitch to convey the meaning of words. Fine pitch processing might therefore be relatively more important in Mandarin than in English SIN perception. We therefore encourage further exploration of linguistic differences in SIN perception in relation to musical training, and suggest caution in generalizing conclusions based on existing work to other linguistic populations.

3.2. Additional cues

Several cues important to SIN perception in real-world environments are not represented in Fig. 1, for the sake of simplicity. Spatial separation between source and distractor is a highly useful cue for speech separation. It has been examined in only three of these studies (Clayton et al., 2016; Swaminathan et al., 2015; Parbery-Clark et al., 2009a,b). SIN perception in the absence of spatial cues, when target and distractor are co-located (or are delivered simultaneously binaurally via headphones), is generally a difficult problem for the auditory system (MacKeith and Coles,

Table 1
Main experimental design characteristics and results of reviewed studies.

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
1. Ruggles et al., 2014 (PloS ONE)	33 healthy adults, aged ~21, 1/2 musicians; native speakers of American English	Does a musician SIN advantage arise from more efficient or robust coding of periodic voiced speech, in continuous and fluctuating noise?	none	Normal (voiced) and whispered (unvoiced) nonsense sentences, in either continuous speech-spectrum noise or gated with a 16 Hz square wave (i.e. intermittent); QuickSIN and HINT	No significant group differences in any main condition, but practice hours correlated with QuickSIN and HINT in musicians
2. Boebinger et al., 2015 (The Journal of the Acoustical Society of America)	50 healthy adults, aged ~27, 1/2 musicians; native speakers of British English	Are there differences in perceptual accuracy of masked speech between trained musicians and non-musicians (using multiple maskers that vary in their energy and information content, and similarity to speech)?	none	Simple sentences in clear speech, spectrally-rotated speech, speech-amplitude modulated noise, and speech-spectrum steady-state noise	Musicians had no significant advantage in the four main conditions (despite better frequency discrimination performance); SIN performance was best predicted by non-verbal IQ scores
3. Parbery-Clark et al., 2009a (Ear and Hearing)	31 healthy adults, aged ~23, 1/2 musicians; native speakers of American English	Are there relationships between musical training and SIN performance (using common clinical measures)?	none	HINT (co-located and separated speech and noise, left and right), QuickSIN	Musicians outperformed the non-musicians on both QuickSIN and HINT (co-located but not spatially separated conditions), in addition to having more fine-grained frequency discrimination and better working memory; years of consistent musical practice correlated positively with QuickSIN, working memory, and frequency discrimination but not HINT. SIN tests were both related to working memory across groups
4. Slater et al., 2015 (Behavioural Brain Research)	38 healthy children, aged ~8, 1/2 had 2 years of musical training and 1/2 had 1 year of musical training; native speakers of American English	Does one or two years of musical training cause SIN perception improvements in children?	none	HINT	Two years of music instruction in children was associated with modest but clinically meaningful gains in SIN perception, and longer periods of training was related to greater SIN perception improvements (2 years as compared to 1 year)
5. Başkent and Gaudrain, 2016 (The Journal of the Acoustical Society of America)	38 healthy adults, aged ~22, 1/2 musicians; native speakers of Dutch	Is there is a musician advantage for speech-on-speech perception, and if so, how does the advantage depend on differences between the two voices?	none	Short, simple grammatically correct Dutch sentences masked with scrambled sentences; masker stimuli were varied in f0 and vocal tract length (9 conditions)	Musicians outperformed non-musicians significantly in all conditions, irregardless of degree of separation in f0 and VTL of target and masker
6. Clayton et al., 2016 (PLoS ONE)	34 healthy adults, aged ~22, 1/2 musicians; native speakers of American English	Are musician SIN advantages related to better cognitive processes, as measured by tests of executive function, spatial hearing, and selective attention?	none	Syntactically correct but not necessarily semantically meaningful sentences masked with similar sentences by a different speaker, presented either from the same source or separated target and noise by 15°	Musicians significantly outperformed non-musicians in the spatially separated condition, but not in the co-located condition
7. Swaminathan et al., 2015 (Scientific Reports)	24 healthy adults, aged ~23, 1/2 musicians;	Are musicians better able to understand SIN than non-musicians	none	Syntactically correct but not necessarily semantically	In the presence of backwards speech maskers (energetic

Table 1 (*continued*)

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
	native speakers of American English	under conditions of varying spatial information and information versus energetic masking?		meaningful sentences masked with similar sentences played either normally (information masking) or backwards (energetic masking) by a different speaker, presented either from the same source or separated from noise by 15°	masking), musicians had significantly better performance when noise and target were co-located; the opposite was true in the presence of competing forward-direction speech: musicians performed significantly better when the target and noise were spatially separated. Thresholds were correlated across masker types across subjects
8. Parbery-Clark et al., 2011a (PloS ONE)	37 healthy adults, aged 45–60, 1/2 musicians; native speakers of American English	Does musical experience offset age-related decline in SIN perception and associated cognitive function in an older cohort of musicians?	none	QuickSIN, HINT, WIN	Musicians performed better than non-musicians in all three SIN tests, as well as auditory working memory and auditory temporal acuity, visual working memory did not differ between groups. Auditory working memory correlated with QuickSIN and HINT but not WIN.
9. Slater and Kraus, 2016 (Cognitive Processing)	54 healthy adults, aged 18–35 (17 non-musicians, 21 vocalists and 16 percussionists); native speakers of American English	What is the role of rhythm-related expertise in SIN perception?	none	QuickSIN, WIN	Percussionists/drummers are better at perceiving sentences-in-noise than non-musicians; no group difference was found using the WIN. Better ability to discriminate rhythms (but not melodies) was associated with better sentence-in-noise (but not words-in-noise) perception across all participants. The ability to perceive words in noise (WIN) did not relate to either rhythmic or melodic competence
10. Parbery-Clark et al., 2011b (Neuropsychologia)	31 healthy adults, aged ~22, 1/2 musicians; native speakers of American English	Is musicians' enhanced speech-in-noise perception facilitated by an increased neural sensitivity to acoustic regularities (i.e. is neural encoding better in predictable conditions)?	EEG: differences in amplitudes between conditions in the frequency following response spectrum (f0, H2-H5)	Speech syllable/da/ presented in predictable and unpredictable sound streams in quiet; HINT	Musicians had a stronger representation of the f0 in the predictable condition relative to the variable condition (which correlated with years of practice), whereas nonmusicians did not. The degree of enhancement in the predictable condition for musicians was related to SIN perception scores. There were no group differences in f0 amplitude between groups. Measures of harmonic encoding were not related to condition, group, or SIN scores.

(*continued on next page*)

Table 1 (*continued*)

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
11. Coffey et al., (submitted, 2016)	20 healthy adults, aged ~22, range of musicianship; native or bilingual speakers of Canadian English	Is the strength of periodicity encoding (in silence) within different brain structures correlated with SIN scores and musicianship?	EEG/MEG: FFR-f0 amplitudes, ERP wave P2 amplitude	Speech syllable/da/ presented in silence; HINT	FFR-f0 representation, localized to auditory cortex, thalamus, and brainstem (using MEG) correlated with SIN performance. FFR-f0 in the right auditory cortex was related to measures of musical experience (onset, total hours). HINT scores correlated significantly with age of start but not total practice hours.
12. Zendel et al., 2015 (Journal of Cognitive Neuroscience)	26 healthy adults, aged ~22, 1/2 musicians; native Quebec French speakers with English competency.	Is there an interaction between musicianship and noise level on attention-dependent cognitive activity related to understanding SIN?	EEG: peak amplitude and latency of ERPs (P1, N1, P2, N400) and differences between conditions	Monosyllabic French words presented in multitalker babble; three levels of difficulty (silence, 15db and 0dB SNR); active and passive (ignore) listening conditions	Musicians heard significantly more words correctly only in the most difficult condition; age of training onset but not years of musical training was correlated with performance in the presence of noise. The pattern of ERP results suggested that more robust sound encoding in noise in musicians, and that musicians increasingly rely on acoustic information, whereas non-musicians rely more heavily on lexical information
13. Du and Zatorre, 2016 (Organization for Human Brain Mapping)	30 healthy adults, aged ~21, 1/2 musicians; native speakers of Canadian English	Do musicians show a SIN perception advantage after non-verbal IQ is controlled for? What are the contributions of auditory ventral and dorsal (motor) stream at different levels of SIN difficulty?	fMRI: univariate GLM analysis + regional multivoxel pattern analysis (MVPA)	Four English phonemes (/ba/,/ma/,/da/and/ta/) in silence and embedded in broadband noise (8dB, 4dB, 0dB, -4dB, -8dB, −12dB SNR)	Musicians outperformed non-musicians at all SNRs and showed stronger recruitment of auditory ventral and dorsal regions. Musicians showed enhanced specificity of phoneme representations in bilateral auditory ventral regions when the noise was weak, and in speech motor regions of the dorsal stream when the noise was strong
14. Musacchia et al., 2008 (Hearing Research)	26 healthy adults, aged ~26, 1/2 musicians; native speakers of American English	Does musical training shape the auditory system in a coordinated manner or in disparate ways at cortical and subcortical levels?	EEG: speech-evoked f0 peak amplitudes (FFR); onset wave delta (8–12ms post-sound onset) latency; P1-N1 and P2- N2 peak-to-peak slopes (ERP)	Speech syllable/da/ presented in silence, audio only or audiovisual (no explicit SIN task)	Musicians had larger FFR-f0 amplitudes and steeper ERP P1-N1 slopes than non-musicians in both audio and audiovisual conditions. Onset peak delta was earlier in musicians. F0 amplitude and P1-N1 slopes were correlated, and both were correlated with years of consistent musical practice.
15. Parbery-Clark et al., 2012a (Frontiers in Aging Neuroscience)	48 healthy adults, aged ~56, 1/2 musicians; native speakers of American English	How does musical experience relate to subcortical responses to speech and speech-in-noise perception in middle-aged adults?	EEG: latency and amplitude of peaks during the FFR response, correlation between waveform during quiet and noise, response consistency	Speech syllable/da/ presented in silence and in noise (unspecified, likely speech-shaped); HINT and self-report of SIN difficulty	Middle-aged musicians had better HINT scores, reported less SIN difficulty in real situations, and had greater neural fidelity of the stimulus with faster neural response

Table 1 (*continued*)

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
					timing, better envelope encoding, greater neural representation of the stimulus harmonics as well as less neural degradation with the addition of background noise. EEG measures were all associated with better speech perception in noise.
16. Parbery-Clark et al., 2009b (The Journal of Neuroscience)	31 healthy adults, aged ~23, 1/2 musicians; native speakers of American English	Do musicians have better subcortical neurophysiological responses to speech in quiet in noise as compared with non-musicians?	EEG: onset latency, consonant-vowel formant transition, and amplitude of the FFR	Speech syllable/da/ presented in silence and in noise (multitalker babble, 10dB SNR); HINT, QuickSIN (behavioural group differences were reported previously in Parbery-Clark et al., 2009a; not duplicated here)	Musicians demonstrated faster neural timing, enhanced representation of speech harmonics, and less degraded response morphology in noise. Relationships to SIN performance were observed in the noise condition or in the degree of difference between the two; not in the silent condition nor in the f0 and harmonic amplitudes.
17. Bidelman and Weiss, 2014 (European Journal of Neuroscience)	24 healthy adults, aged ~24, 1/2 musicians; native speakers of Canadian English	How do enhancements indexed by cortical and subcortical measures relate to musicianship and speech-listening behaviours?	EEG: peak amplitude and latency were measured for the prominent deflections of the cortical ERPs (Pa, P1, N1, P2, P3) in specific time windows	Measured categorical perception and neural correlates using synthetic vowels that varied only in the frequency of their first formant (phonetic continuum of/u/to/a/); no explicit SIN task was used	Musicians were faster at categorizing speech tokens and featured a more pronounced boundary between phonetic categories as compared with non-musicians, a measure that was related to years of musical training. Musicians had an advantage in neural and behavioural categorical speech processing, a higher-order linguistic operation
18. Tierney et al., 2015 (Proceedings of the National Academy of Sciences)	40 healthy children aged ~14 at start and ~18 at end, 1/2 in musical training program, 1/2 in an unrelated control training program; native speakers of American English	Does musical training in adolescence alter the course of auditory development?	EEG: FFR consistency across trials, ERP P1 and N1 amplitude	Speech syllable/da/ presented in quiet; behavioural measures of linguistic processing (phonological awareness, phonological memory, and rapid naming abilities; no explicit SIN task was used)	Both groups improved in phonological awareness relative to the general population, but the music training group improved more compared with the active controls. Training had no significant effect on phonological memory ability nor rapid naming ability. Response consistency (FFR) and P1-N1 amplitude difference were found between groups after training.
19. Parbery-Clark et al., 2012b (Neurobiology of Aging)	87 healthy adults, one younger group (~23) and older group (~50), 1/2 of each group had musical experience; native speakers of American English	Can musical training offset the negative impact of aging on neural processing?	EEG: peak latencies, specifically onset and transition responses; FFR-f0	Speech syllable/da/ presented in silence (no explicit HINT task)	Musicians showed less age-related neural delay than non-musicians in onset and transition peak latencies to syllable/da/, but no difference during the steady-state (f0) portion of the FFR.

(*continued on next page*)

Table 1 (*continued*)

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
					Results show distinct effects of aging and musicianship on the neural mechanisms responsible for encoding the different components of a stimulus
20. Strait and Kraus, 2011 (Frontiers in Psychology)	23 healthy adults, aged 18–35, 1/2 musicians; native speakers of American English	Does musical training benefit cortical mechanisms that underlie selective attention to speech?	EEG: amplitudes of ERPs	Speech syllable/da/ presented within either an attended or unattended speech stream (attended: +10dB SNR) that was separated by spatial location and sex of speaker; HINT (behavioural group differences were reported previously in Parbery-Clark et al., 2009a and Strait et al., 2010; not duplicated here), a measure of selective auditory attention	Auditory attention performance correlated with speech-in-noise perceptual ability, with better auditory attention relating to the ability to accurately perceive speech in higher levels of background noise; only musicians demonstrated decreased cortical response variability with auditory attention over the prefrontal cortex
21. Coffey et al., 2016b (Nature Communications)	20 healthy adults, aged ~22, range of musicianship; native or bilingual speakers of Canadian English	Is musical experience related to fidelity of periodic sound encoding in the auditory cortex?	EEG/MEG: FFR-f0 amplitudes, originating in right auditory cortex	Speech syllable/da/ presented in silence (no explicit HINT task)	FFR-f0 amplitude from right but not left auditory cortex correlated with age of training onset, cumulative practice hours, and fine pitch discrimination ability
22. Varnet et al., 2015 (Scientific Reports)	38 healthy adults, aged ~23, 1/2 musicians; native speakers of French (presumed)	Do musicians rely on different acoustic (spectro-temporal) cues when perceiving speech?	none	Phonemes (/da/,/ga/) presented as part of two-syllable nonsense words, embedded in white noise (SNR adapted to performance)	Musicians performed better than non-musicians and demonstrated faster learning. Musicians relied more heavily on the two main acoustic cues (selectively focusing on a small time-frequency region that is critical for correct/da/-/ga/ categorization), and they responded more consistently to stimuli
23. Musacchia et al., 2007 (Proceedings of the National Academy of Sciences)	29 healthy adults, aged 18–40, 1/2 musicians; native speakers of American English (presumed)	Do musicians have more robust (EEG) responses to speech and music, in audio and audiovisual presentations of speech and musical stimuli	EEG: onset response latencies, f0 and harmonic peak spectral amplitudes	Speech syllable/da/and cello tone 'G2′ presented in silence (no explicit HINT task)	Musicians had earlier and larger (electrophysiological) onset responses as compared with non-musician controls to both speech and music stimuli presented in auditory and audiovisual conditions, and larger FFR-f0 peak amplitudes in the speech condition. FFR-f0 amplitude was related to years of musical practice
24. Fuller et al., 2014 (Frontiers in Neuroscience)	50 healthy adults, aged ~23, 1/2 non-musicians; native speakers of Dutch	Does a musician SIN advantage persist under degraded pitch conditions of cochlear implant simulations?	none	Experiment 1: Cochlear implant simulations and normal speech (words in silence and speech-shaped noise at +10dB, +5 and 0 dB SNR; sentences in silence, speech-shaped noise, envelope-	Musicians were better able to identify words in speech-shaped noise but only in the +5db SNR condition with degraded target speech. No group differences were found using sentences as targets.

Table 1 (*continued*)

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
				modulated noise, 6-talker babble). Experiment 2: nonsense words with emotional content (unprocessed, degraded) in quiet. Experiment 3: various tonal maskers were used to assess the use of timbre and pitch cues to segregate competing melodies	Musicians showed a slight advantage identifying vocal emotions in speech stimuli. Musicians were better at the melodic contour identification task for unprocessed stimuli and most of the cochlear simulation degraded signal conditions. In general, musicians performed better than non-musicians when degraded target signals were used, though cross-domain (i.e. language stimuli) effects were weak, differences were more pronounced in tests that required pitch information.
25. Lee et al., 2009 (The Journal of Neuroscience)	26 healthy adults, aged ~25, 1/2 musicians; native speakers of American English (presumed)	Are musicians better at encoding behaviourally relevant aspects of sound?	EEG: FFR spectral amplitudes (f0, Hs)	Consonant and dissonant tone intervals presented in silence (no explicit SIN task)	Musicians had heightened responses to the harmonics of the upper tone, a feature often important in melody perception; the acoustic correlates of consonance perception (i.e., temporal envelope) were more precisely represented in the subcortical responses of musicians and correlated with musical experience
26. Strait et al., 2010 (Annals of the New York Academy of Sciences)	33 healthy adults, aged 18–40, 1/2 musicians; native speakers of American English (presumed)	Do musicians possess better-developed cognitive abilities that might account for fine-tuned auditory perception in a top-down fashion?	none	Cognitive and perceptual data on auditory working memory, auditory attention, visual attention, frequency discrimination, frequency selectivity, temporal resolution, and non-verbal IQ (no explicit SIN task)	Musicians had lower perceptual thresholds, specifically for auditory tasks that relate with cognitive abilities, such as backward masking and auditory attention
27. Oxenham et al., 2003 (The Journal of the Acoustical Society of America)	24 healthy adults, aged ~25, 1/2 musicians; native speakers of American English (presumed)	Are musicians less susceptible to information masking (which reflects central rather than peripheral limits of sound processing)?	none	Signal-to-masker ratio at thresholds measured (using repeated bursts of random-frequency sinusoids gated in precise synchrony)	Musicians had a large and statistically significant advantage over non-musicians in the high information masking condition; no difference was found between groups in the energetic masking condition
28. Zendel and Alain, 2012 (Psychology and Aging)	163 healthy adults, aged 18–91, 1/2 musicians; native speakers of Canadian English	Do musicians experience less age-related decline in central auditory processing?	none	QuickSIN, gap detection, mistuned harmonic detection	Musicians experienced less age-related decline for both gap-detection and speech-in-noise thresholds (though were not uniformly better at SIN across the lifespan). Musicians demonstrated a lifelong advantage in detecting a mistuned harmonic compared to nonmusicians.

(*continued on next page*)

Table 1 (*continued*)

	Subjects	Most relevant research question(s)	Imaging modality, main measures	Main conditions, SIN tests	Significant main results
29. Zendel and Alain 2009 (Journal of Cognitive Neuroscience)	28 healthy adults, aged ~30, 1/2 musicians; native speakers of Canadian English (presumed)	Are musicians better able to segregate concurrent sounds based on harmoncity?	EEG: amplitudes of ERPs	Complex harmonic tones in which one of the harmonics was mistuned by varying degrees (no explicit SIN task); subjects were asked whether they heard one or two separate sounds.	Performance on the mistuned harmonic and the speech-in-noise tasks were correlated with hours of music practice. Musicians were more likely to identify a mistuned harmonic as a distinct auditory object compared with nonmusicians. This was paralleled by differences in the amplitude and latencies of ERP waves.

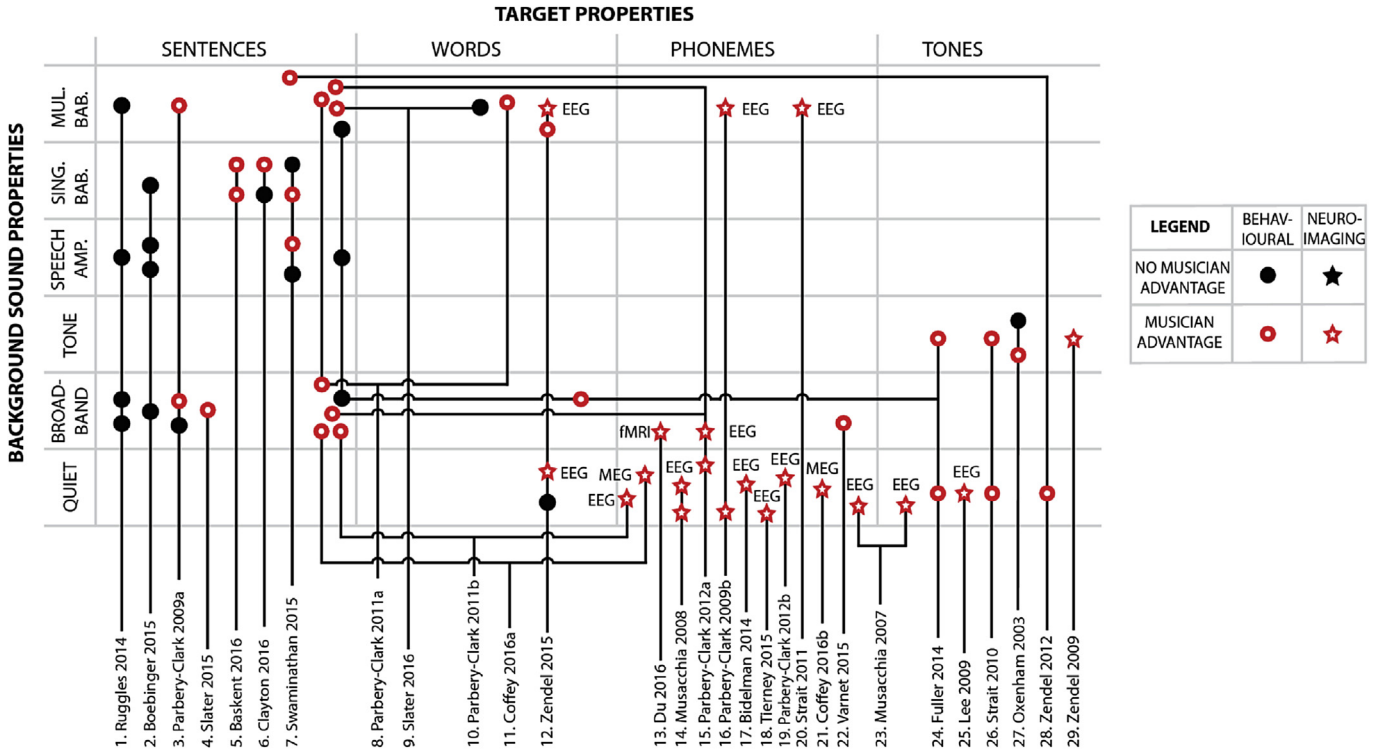

Fig. 1. Studies on speech-in-noise (SIN) perception in musicians, organized according to target and background sound properties. Studies with multiple conditions show symbols linked with lines. Conditions that included a neuroimaging component (EEG, MEG, or fMRI) are indicated by stars, whereas those that used only behavioural measures are represented by circles. Multiple conditions from a study may fit in the same area if they vary in a non-represented dimension, such as in spatial or visual cues (see Table 1 for details). Red symbols indicate conditions in which musicians were found to have performance/neural enhancements in at least one variant of the condition, relative to non-musician controls in cross-sectional designs or post-relative to pre-training in longitudinal designs. ('Broadband' = continuous broad spectrum noise including speech spectrum noise; 'Speech-amp.' = speech amplitude, refers to conditions such as backwards speech, spectrally rotated speech and speech-amplitude modulated noise that preserve some of the spectro-temporal complexity of speech yet are unintelligible; 'Sing. Bab.' = single-talker babble, may be intelligible or nonsense; 'Mul. Bab.' - multitalker babble, two or more speakers).

1971; Dirks and Wilson, 1969; Hawley et al., 2004; Avan et al., 2015). In addition to energetic masking designs, the performance benefit of spatial information is also observed when informational maskers are used (e.g. understandable sentences), suggesting that centrally-mediated processes make use of this cue, in addition to lower-level processes (e.g. Freyman et al., 2005; Hawley et al., 2004).

Another set of cues that are absent in Fig. 1 relate to multimodal interaction. The visual system can provide useful information for separating target from distractor (visual cues may co-occur with spatial cues in some natural listening situations, but can be experimentally dissociated). Watching a speaker's face considerably improves speech intelligibility under difficult listening conditions (Bernstein and Grant, 2009; Helfer and Freyman, 2005), and even the coincidence of temporally modulated yet irrelevant visual cues with a target auditory stream enhances SIN performance (Maddox et al., 2015), suggesting that there are strong interactions between the auditory and visual system during SIN tasks when visual cues are available. Recent work suggests that visual information benefits auditory processing via mechanisms of crossmodal

integration over long temporal windows (Crosse et al., 2016). Only a few of the studies reviewed here included an audio-visual condition (Musacchia et al., 2008; Musacchia et al., 2007) or measured visually-related behavioural skills such as visual attention (Strait et al., 2010).

The motor system is also likely to be involved in SIN perception. This conclusion is drawn from neuroimaging results that relate co-activation of known motor areas of the brain during SIN tasks with better performance (e.g. Du et al., 2016; Bishop and Miller, 2009), correlations between rhythmic skills with linguistic skills including reading (e.g. Woodruff Carr et al., 2014), results that show altered performance in speech discrimination after magnetic disruption to brain areas that house motor representations of the tongue and lips (D'Ausilio et al., 2012; Bartoli et al., 2015), and relationships between motor planning and rhythmic processing (reviewed in Grahn, 2012).

The role of prediction is more complicated as it can operate at multiple levels of processing. It is somewhat implicit in the categories presented in Fig. 1 themselves, although not necessarily in the same progressive order as the acoustic feature properties. Sentence targets generally offer more opportunity to use knowledge of grammar and meaning than do single words, though these cues can be examined separately, for example by using grammatically correct nonsense sentences that reduce semantic (but not syntactic) predictability (Ruggles et al., 2014). A single repeated sound is also highly predictable, though it is less likely to invoke high-level linguistic processes. The auditory system's ability to extract regularities from an acoustic signal is thought to be a key mechanism supporting SIN perception. Regularity-based predictions can be used to bias and fine-tune activity in low-level areas in a top-down fashion, via feedback loops originating in the cortex and terminating in subcortical nuclei (Parbery-Clark et al., 2011a,b; Zendel et al., 2015; Suga et al., 2000). This process may improve the representation of behaviourally-relevant sound features and may also allow for noise with regular properties to be suppressed (reviewed in Suga, 2012).

The many interactions between lower-level encoding, higher-level treatment of sound, and cross-modal integration suggests that SIN perception will not be accounted for by an exclusive focus on each in isolation. A whole-system view may be most helpful (Kraus and White-Schwoch, 2015) and will likely be facilitated by continuing to connect behaviour to brain function via neuroimaging methods (studies that have used neuroimaging techniques are marked in Fig. 1 and their coverage is discussed in a later section).

In sum, a range of target properties has been studied: sentences, to single words, phonemes, and tones, as has a range of distractor properties, from their absence, to speech-shaped noise, backwards speech, and sentences from one or more talkers. The coverage of combinations of target and distractors is less comprehensive, with little exploration of word and phoneme identification in different types of noise (see Fig. 1). When other paradigm variations such as the presence of audio-visual and spatial cues and different languages are added, the number of possible paradigm variations becomes extremely large. This motivates our summary and proposals for adopting a structured means of investigating the relative contributions of speech cue processing in SIN perception.

4. Evidence for musicianship enhancement in SIN perception

Also represented in Fig. 1, by colour, is whether or not significant differences between musicians and non-musicians were found for each study/condition. The purpose of this inclusion is to adduce evidence in favour of a musician advantage, and its possible nature, as it has previously been suggested that observed advantages in some studies may be accounted for by co-variation of higher-order cognitive factors (e.g. nonverbal IQ) with musicianship (Boebinger et al., 2015; Schellenberg, 2015).

The results summarized in Fig. 1 suggest that a musician advantage emerges over a variety of experimental conditions, although the ratio of studies finding differences to those that do not should be interpreted cautiously due to publication bias (Ioannidis et al., 2014), and some of the conditions marked as significant were so only for a subset of sound/difficulty levels: e.g. Fuller et al. tested for a musician enhancement in a words-in-noise task and found a significant difference between groups at one out of three signal-to-noise ratio (SNR) levels (see Table 1). Notwithstanding, all but two studies reported here (Ruggles et al., 2014; Boebinger et al., 2015) found a statistically significant musician enhancement in at least one condition. Among these studies are several in which higher-level cognitive factors like non-verbal IQ and working memory had been measured or controlled (Du and Zatorre, 2016; Parbery-Clark et al., 2009a,b; Strait et al., 2010; Slater and Kraus, 2016; Parbery-Clark et al., 2012a; Strait and Kraus, 2011), or that involved longitudinal training paradigms (Slater et al., 2015), including with random group assignment and a non-musical control condition (Tierney et al., 2015). Taken in the wider context that measures of predominantly bottom-up processes like periodic sound encoding are correlated with SIN scores (Coffey et al., 2016; Song et al., 2011; Hornickel et al., 2011), that longitudinal musical training designs reveal SIN gains (e.g. Song et al., 2012, see Alain et al., 2014 for a review) and the large body of research showing that neuroplasticity is induced by musical training (Herholz and Zatorre, 2012; Pantev and Herholz, 2011), it seems highly likely that musical training does influence SIN perception, and that a range of mechanisms that act on different cues (and their integration) are involved.

5. Mechanisms for cross-domain enhancement in SIN perception

The OPERA hypothesis (Overlap, Precision, Emotion, Repetition, Attention; Patel, 2014) proposes that sensory or cognitive processing mechanisms that are shared between domains might be enhanced by musical training when high task demands are combined with the emotional rewards, frequent repetition, and focused attention. These conditions are often met in musical training, making it potentially capable of inducing long-lasting changes to brain structure and function that might impact on speech processing. The degree of overlap between speech and music processing remains an active area of research; whereas there is evidence for overlap in hemodynamic response, many studies also show some degree of separation between the two domains (Leaver and Rauschecker, 2010; Angulo-Perkins et al., 2014). A recent study using a voxel decomposition method suggested that there is considerable separability of speech vs music in non-primary auditory cortex, but also that more basic acoustical features, including frequency and spectrotemporal modulation rates, recruit regions of peri-primary auditory cortex irrespective of whether the stimulus is speech or music. The degree to which processes are shared in the SIN task with musical training has been recently reviewed as regards neural overlap in co-activation studies (though neural overlap itself does not necessarily imply shared circuitry; see Peretz et al., 2015; Milovanov and Tervaniemi, 2011)., These processes include high-level cognitive processes and auditory acuity (Lee et al., 2009; Bidelman, et al., 2011; Bidelman, et al., 2013), and their interaction, for example low-level auditory signal processing enhancements might free up perceptual, attentional, and cognitive resources that could then be dedicated to flexibly adapting strategies to specific task demands or compensating for weaknesses (Zendel and Alain, 2012). However, as both musicianship and SIN

perception involve numerous sub-skills and can vary considerably, is not yet clear which overlapping aspects of cognition might be responsible. All of the SIN-relevant cues described above might be exercised to some degree during the course of musical training, and have been found to be enhanced among musicians. Of course, musical experiences can differ in ways that might affect which processes are enhanced; this is nicely illustrated in a study by Slater and Kraus (2016), in which SIN perception in percussionists, vocalists, and non-musicians was found to correlate with rhythm ability and to differ according to the nature of training. Bearing this in mind as a possible source of variability and inconsistency in heterogeneous groups of musicians, we will review evidence for musician enhancement of each cue and associated cognitive function.

6. The nature of musicianship enhancement in existing studies of SIN perception

6.1. Conditions in which musician differences are found

An examination of conditions in which group differences in SIN perception have and have not been found reveals that they are not obviously related to target and distractor cue richness (Fig. 1), nor is the pattern of results among equivalent conditions entirely consistent. For example, whereas Boebinger et al. (2015) did not find a musician advantage for conditions with high information masking (Boebinger et al., 2015); Swaminathan et al. (2015) showed a strong musician advantage using a similar design (Swaminathan et al., 2015).

Two studies showed no group differences, both using sentence-in-noise measures (Ruggles et al., 2014; Boebinger et al., 2015). Boebinger et al. studied the ability of musicians and non-musicians (who did not differ as a group in measures of working memory and non-verbal IQ) to understand simple sentences that were embedded in four different kinds of maskers that varied parametrically in their informational content: clear speech, spectrally-rotated speech (which preserves much of the spectro-temporal complexity of speech yet is unintelligible), speech-amplitude modulated noise (which includes envelope information yet lacks spectro-temporal dynamics such as formant or harmonic structure), and speech-spectrum steady-state noise (which has the same long-term average spectrum as speech, but lacks other cues; Boebinger et al., 2015). They did not find statistically significant differences between groups, nor did they find significant correlations between SIN perception and age of training onset within the musician group. Ruggles et al. found a significant difference in the full-scale IQs of musicians and non-musicians within their sample (favouring musicians), but despite that nonverbal IQ had been suggested as a cause of previously observed musician enhancements (Boebinger et al., 2015), did not find group differences on measures of speech intelligibility using normal (voiced) or whispered (unvoiced) grammatically-correct nonsense sentences embedded in continuous and intermittent speech-spectrum noise (Ruggles et al., 2014). Although no group difference was observed on clinical speech-in-noise tests in this study (QuickSIN, HINT), there were significant correlations with years of training for both tests within the musician group. Using similar tasks, Parbery-Clark et al. found that musicians did outperform the non-musicians on both QuickSIN and HINT (Parbery-Clark et al., 2009a,b). Years of consistent musical practice correlated positively with QuickSIN, working memory, and frequency discrimination but in contrast to Ruggles et al. (2014) and Parbery-Clark et al. (2012b), did not correlate with HINT scores (the discrepancy between QuickSIN and HINT results is interpreted as differences in dependence on working memory, which covaries with musicianship). The majority of conditions that do not find musician-related differences are thus those with more complex, cue-rich target stimuli, conditions that might allow for compensatory mechanisms to mask relative weaknesses in non-musicians at some levels of difficulty.

Although the results lack complete consistency and do not clearly point to a single mechanism for musician SIN enhancement, the preponderance of the evidence (18 out of 20 studies) indicates that there is indeed a musician advantage, and that this phenomenon cannot be explained on the basis of nonverbal IQ, working memory, or other confounds. Thus, the most parsimonious explanation for the few studies that failed to find effects is some combination of musician and SIN task heterogeneity, sampling error, and effect size (Boebinger et al. 2015 for example calculated that 115 subjects per group would have been required to achieve statistical power at the 0.8 level, which is much larger than most of the samples used in the reviewed studies). In the following sections we will discuss several candidate sub-processes that are known to be strengthened by musicianship: higher-level cognitive factors and executive functions, basic sound encoding, spatial cue processing, and multisensory integration.

6.2. Cognitive factors and executive functions

Enhanced SIN perception observed in musicians could be due to better higher-level functions (Boebinger et al., 2015) such as auditory attention or auditory working memory capacity (Strait et al., 2010; Carey et al., 2015; Chan et al., 1998; Ho et al., 2003; Brandler and Rammsayer, 2003), which might act to fine-tune or separate incoming auditory information and help match it to knowledge, and are known to be enhanced by musical training. The links between musical training and cognitive abilities are complex (Schellenberg and Peretz, 2008; see Moreno and Bidelman, 2014 for a recent review). Anderson et al. (2013) used structural equation modelling to evaluate the interacting contributions of cognitive ability (as measured by auditory working memory, auditory short-term memory, and auditory attention) and other factors on SIN performance, and found that cognitive ability significantly explained SIN scores, an effect that was modulated by previous musical experience. Auditory working memory has been suggested as a main mediator underpinning musicians' auditory advantages, including SIN perception as reviewed in Kraus et al. (2012). Musicians tend to have better auditory working memory ability than non-musicians (Parbery-Clark et al., 2011a,b; Kraus et al., 2012). In fMRI studies (e.g. Pallesen et al., 2010) musicians showed more brain activity during working memory performance in cortical areas involved in cortical control as compared with non-musicians (i.e. prefrontal cortex, lateral parietal cortex, insula, right putamen and anterior cingulate gyrus). Auditory working memory is in turn related to the length and timing of musical training exposure, and to rhythm processing (Bailey and Penhune, 2010) - another skill for which a causal role in language processing and acquisition has been demonstrated.

Selective auditory attention, which is a mechanism for determining which sounds will be most thoroughly processed and brought to awareness, is also enhanced in musicians (Strait and Kraus, 2011). Higher informational content in the distractor decreases how well a listener can solve SIN problems. Swaminathan et al. (2015) studied this effect in musicians and non-musicians in more depth by masking sentences with forwards and backwards speech, and found that musicians were less susceptible to informational masking, suggesting that they were better able to selectively attend to the target (as well as possibly suppress the distractor based on linguistic knowledge). Clayton et al. (2016) further confirmed the relationships between domain-general factors including both selective attention and working memory in SIN

perception among musicians by evaluating the statistical relationships between SIN scores and a variety of measures of executive function.

Computing regularities within incoming sounds is another important process that might be enhanced in musicians, possibly via enhanced processing of statistical information (Shook et al., 2013). Varnet et al. (2015) studied whether musicians and non-musicians differed in their ability to distinguish between two phonemes presented in speech-spectrum noise (/ga/, /da/) and modelled which cues each group was reliant on. Although they used similar strategies to non-musicians, musicians performed better, focused precisely on the acoustic cues that distinguished the phonemes, and learned more quickly over the course of the experiment than their non-musical counterparts (Varnet et al., 2015). These results suggest that top-down strengthening of incoming relevant sounds or suppression of irrelevant sounds may be facilitated in musicians by better mechanisms that extract spectro-temporal regularities.

Cognitive factors that support auditory function are in part determined by genetics. For example, in a twin study of musical training, Mosing et al. (2015) found that the relationship between practice and IQ could be largely accounted for by controlling genetic and shared environmental influences, despite differences in musical practice. However, evidence of musical training effects on executive functions including auditory working memory have also been demonstrated in longitudinal studies (e.g. Moreno et al., 2011, reviewed in Moreno and Bidelman, 2014). There is therefore likely to be an interplay between genetic and other predispositions with experience-dependent modulation of brain circuitry that gives rise to training-related effects (Zatorre, 2013).

These findings on cognitive factors affecting performance, taken together, imply that musical training influences SIN performance by strengthening higher-level processes related to attending to sound, repressing irrelevant sound, and holding auditory information temporarily in mind. Although domain-general cognitive factors are clearly important, the results presented in Fig. 1 also show musician group differences in a number of studies with low working-memory and attentional load, such as when phonemes or individual words are used as targets (e.g. Parbery-Clark et al., 2009a,b; Strait and Kraus, 2011; Du and Zatorre, 2016), as well as studies without two sources of sound on which to apply methods of selective attention, as when targets are presented in silent listening conditions and attention is otherwise engaged (e.g. Coffey et al., 2016; Bidelman and Weiss, 2014; Musacchia et al., 2007; Musacchia et al., 2008). These findings suggest that lower-level mechanisms are also at work.

6.3. Basic sound encoding

Poor sound encoding logically limits the ability of the auditory system to group and separate incoming sounds correctly; encoding frequency information accurately is necessary if cues like pitch are to be used to follow and separate out a speaker's voice (Moore and Gockel, 2002; Fuller et al., 2014), or to match neural representation of the incoming acoustic signal to a stored lexical representation (Zendel et al., 2015). Two reviews have looked at the relationship between measures of basic sound encoding in the brain (which are most often made using EEG) and SIN perception (Anderson and Kraus, 2010; Du et al., 2011). Here, we focus on pitch cues for their clear relationship to musical practice, but enhancements found in musicians' basic sound encoding includes faster neural response timing, higher neural response consistency, more robust encoding of harmonics, and greater neural temporal precision (Parbery-Clark et al., 2012a). Neural correlates of acoustic differences between contrastive stop consonants play an important role in SIN deficits in linguistic ability and deficit in children (e.g. Hornickel et al., 2009), and are more effectively used by musicians (Varnet et al., 2015; Parbery-Clark et al., 2012). Musicians are also better able to segregate harmonic and mistuned sound streams based on their harmonicity (Zendel and Alain, 2009), suggesting that the ability to analyze the spectral relationships within auditory scenes is improved by long-term musical training.

Pitch cues are likely candidates for music-related SIN enhancement, given that musical activities almost always extensively involve producing and attending to pitched sounds, and that musicians have superior pitch discrimination abilities, a skill that shows a clear training effect (Micheyl et al., 2006). Periodicity is used by the auditory system to promote stream segregation (Bregman, 1994), auditory object formation and speaker identification (see Parbery-Clark et al., 2011a,b). Pitch is related to encoding of the fundamental frequency (f0; Gockel et al., 2011), the slowest repeating periodic element of that sound, and can be measured using EEG via the frequency-following response (FFR; Skoe and Kraus, 2010). The recent studies covered in this paper that examine group differences in early sound measures including the FFR and their correlation with SIN perception underscore the importance of high fidelity sound encoding and its robustness in the presence of noise in supporting this skill (summarized in Table 1). For example, Coffey et al. (submitted, 2016) examined the correlation between SIN performance and the strength of the fundamental frequency's representation in the FFR in different cortical and subcortical auditory brain structures using magnetoencephalography (MEG), and found that SIN was related to FFR strength throughout the auditory system. SIN scores were also related to the age at which training started among those with musical experience, suggesting an effect of experience. Fuller et al. varied the degree to which SIN tasks were dependent on pitch cues to test their influence on SIN perception across groups (2014). Results showed that musician versus non-musician group differences were greater in conditions that relied more on pitch cues. Tierney et al. (2015) measured the FFR (presented in silence) in teenagers before and after a period of musical training, and found that as compared with a control group, musically trained individuals showed faster neural responses to sound and SIN performance was improved after training. These results support the hypothesis that superior processing of pitch cues in musicians plays an important role in SIN performance as well as a causal role of musical training on basic sound representation subserving SIN perception. The bottom row of Fig. 1 demonstrates that musicians generally represent both musical and linguistic sounds better than non-musicians even under ideal listening conditions.

6.4. Auditory spatial separation

Although spatial cues were only included in a handful of the reviewed studies (Strait et al., 2010; Parbery-Clark et al., 2009a,b; Swaminathan et al., 2015; Clayton et al., 2016; Strait and Kraus, 2011) and therefore would not account for the majority of group differences in Fig. 1, spatial information is generally available to populations with normal binaural hearing in naturalistic listening conditions, and is well-established to greatly improve stream segregation in general (Bregman, 1994), and SIN performance specifically (Pressnitzer et al., 2011). Some musical activities such as conducting an orchestra (and perhaps playing in an ensemble) appear to enhance auditory localization mechanisms (Münte et al., 2001), suggesting that enhanced processing of auditory spatial cues may represent another possible source of musician SIN enhancement.

Swaminathan et al. (2015) investigated the effect of spatial separation, and found a musician advantage only when both target

and distractor emanated from separate locations, in the condition that used normal speech (i.e. high information masking). In contrast, they found a group difference with reversed speech (i.e. low information masking) only when spatial cues were absent, suggesting that a musicianship advantage is brought out in the low information/high energetic masking condition only in the most difficult co-located listening situation. Clayton et al. (2016) used a high information masking design (forward speech) and also found a musician advantage on a measure of SIN perception when target and distractor were separated, rather than co-located; the authors speculate that there may have been an interaction between the difficulty level of the task and the degree to which each group could make use of the available cues, so this finding may not apply over all difficulty levels. Parbery-Clark et al. (2009a,b) instead found group differences in the HINT task only when speech and noise were co-located, and *not* when the distractor was delivered 90° to the left or right of the target, which was presented from straight ahead. These conflicting results suggest that a musician advantage may be partly fuelled by musicians' enhanced auditory spatial skills (Clayton et al., 2016), but only under some combinations of conditions. Spatial cues may be a useful target for improving SIN perception via training, as these cues are generally available to listeners in real-life situations.

6.5. Visual and motor system multisensory integration

Musicians must attend to visual cues to communicate timing and expressive information to other musicians, to read music, and sometimes to follow a conductor (Clayton et al., 2016). They must plan and coordinate their movements in order to produce sound; the visual and motor systems might therefore be sources of musician enhancement.

Only a couple of the studies reviewed here included an audiovisual condition (Musacchia et al., 2007; Musacchia et al., 2008). Musacchia et al. (2007) presented auditory stimuli in unimodal and audiovisual conditions (in silence). Measures of basic sound encoding (i.e. the fundamental frequency in the FFR) were found to differ between groups in the audiovisual condition, and were generally present but were smaller and less clear for the unimodal auditory condition (see Table 1 for main measures), suggesting that musicians are better able to take advantage of audiovisual integration even at very early stages of sound processing. This finding is consistent with the observation that musicians have narrower temporal integration windows for detection of misaligned auditory and visual targets (Lee and Noppeney, 2011; Lee and Noppeney, 2014). These studies did not relate audiovisual results directly to SIN perception, so the influence of this enhancement is unknown. In related work (that did not investigate musicianship), Zion Golumbic et al. (2013) studied whether congruent visual input of an attended speaker enhanced the neural representation of natural continuous speech, as measured using MEG. Their results reinforced the importance of visual input in resolving auditory perceptual ambiguity, which they speculated might act to direct attentional resources to points in time at which important acoustic input is expected (Zion Golumbic et al., 2013).

Although none of the reviewed studies explicitly examined the role of the motor system via relationship to motoric behavioural differences, Du & Zatorre showed that musicians and non-musicians differ in their recruitment of dorsal brain regions that are known to represent motoric aspects of speech when listening for words in noise (Du and Zatorre, 2016). The motor system was recruited to a greater extent in more difficult listening conditions, suggesting that it helps to compensate for impoverished sensory representations (Du et al., 2016). When full sentences are used, rhythm might increase sensitivity to timing patterns that are important for speech perception and serve as a proxy for grammatical processing, in that it may enhance the brain's ability to detect if a candidate word sequence violates a grammatically-expected rhythm (Slater and Kraus, 2016). Rhythmic processing relates to the synchronization of low-frequency cortical activity to the slow temporal modulations of speech, which could act to assist SIN perception by boosting the strength of the brain's representation of the target signal (Schön and Tillmann, 2015; Slater and Kraus, 2016). Integration of auditory processes with the visual and motor system is both an important feature of musical practice (Zatorre et al., 2007), and is relevant to naturalistic SIN perception; further investigation of these relationships in musicians is therefore warranted.

6.6. Interaction of task difficulty, cue relevance and experience

Du et al.'s results raise an interesting complication that might explain some of the observed inconsistencies in group results: listeners recruit brain regions to different degrees according to their experience (Du and Zatorre, 2016), demographics such as age (Du et al., 2016), the difficulty level of the task (i.e. SNR between target and source; Du et al., 2014; Wong et al., 2009; Zendel et al., 2015), and likely, the specific nature of the masker (Swaminathan et al., 2015). For example, in a study of SIN perception in aging Du et al. (2016) found that older adults had higher activation of frontal speech motor areas as measured by fMRI during a syllable identification task than did younger adults. This result was interpreted as a compensatory mechanism whereby older adults learned to rely on preserved phoneme specificity to achieve similar levels of SIN performance as their younger counterparts. Measures of basic encoding also show similar difficulty-dependent relationships, for example, musician-related group differences in the frequency-following response are much clearer when measured in difficult, noisy conditions than in silence (Parbery-Clark et al., 2009a,b; Strait and Kraus, 2011). These findings are in line with earlier work that used structural equation modelling to evaluate interacting contributions of peripheral hearing, central processing, cognitive ability, and life experiences to understanding SIN, and showed that older musicians rely on different cues than age-matched non-musicians (Anderson et al., 2013). Whereas Du et al.'s fMRI work suggests increased reliance on frontal motor networks in musicians as compared with non-musicians and across groups as difficulty increases (Du and Zatorre, 2016), other work investigating the relationship between difficulty level (SNR) and EEG measures (P1, N400) suggested instead that as difficulty increases, musicians might be *more* reliant on acoustic cues, which might benefit SIN perception via improved lexical access (Zendel et al., 2015). These discrepancies might be resolved by using neuroimaging methods that can bridge across spatial and temporal resolutions, like combined EEG-fMRI or MEG. The majority of studies reviewed here use adaptive or accuracy-based behavioural measures of SIN, rather than set SNR levels, which may obscure some of these effects. However, considering SNR in addition to the cognitive, acoustic, spatial, and multisensory factors described above further increases the experimental design space.

7. Application of neuroimaging to SIN perception in musicianship

Understanding musician enhancement in SIN perception may come down to determining the relative importance of multiple mechanisms that contribute to this complex task in a range of listening conditions, or for clinical purposes, those that are most problematic in everyday life. However, the challenging multifaceted nature of SIN perception means that a comprehensive

understanding of it will be difficult to achieve by gradually exploring combinations of target and distractor properties, spatial and multisensory cues, linguistic and musical variation in experience, and difficulty levels, that influence behavioural SIN results. Probing the cognitive mechanisms that support SIN perception with neuroimaing may be a more powerful approach.

Fig. 1 shows that the neural correlates of SIN ability have seen lopsided coverage to date (stars; neuroimaging modalities are labelled). About half of the studies we have reviewed have been purely behavioural, including the majority of studies that have looked at naturalistic cue-rich conditions. Application of neuroimaging is largely concentrated on neural correlates of phonemes and tones presented in silence, with a few using speech-shaped noise or broadband noise (Du and Zatorre, 2016), a tone distractor (Parbery-Clark et al., 2011a,b), or multitalker babble (e.g. Tierney et al., 2015; Parbery-Clark et al., 2009a,b). This limitation may be for practical reasons: EEG, which the majority of these studies has used, typically requires repetition of hundreds or thousands of iterations in order to obtain stable averaged measures (Skoe and Kraus, 2010). High repetition designs lend themselves to the study of basic properties of sound encoding but less so to studying sentences, although it is possible to embed repeated syllables in naturalistic sound streams (e.g. Strait and Kraus, 2011) or to study the neural tracking of a speech signal in terms of how accurately it can be reconstructed from data (e.g. the broadband speech envelope from EEG data). Stimulus reconstruction from EEG and MEG data has recently been used to study the benefit to SIN perception and mechanisms of audio-visual integration (Crosse et al., 2016) and to determine attentional selection and auditory object enhancement in naturalistic multispeaker environments and with vocoded speech (O'Sullivan et al., 2015; Ding and Simon, 2012; Rimmele et al., 2015). Techniques such as these hold promise because they allow investigators to determine the actual information content of a neural signal, rather than merely the existence of a correlation between magnitude of neural activity and a certain stimulus. As applied to the issue of musical training, it could help to determine not only where, but also how the neural signal becomes enhanced such that it leads to better behavioural performance.

In Table 1, we included the main neurophysiological results of the studies that used EEG (or combined EEG/MEG) to study SIN perception in musicians. Their measures fit into two general categories. Those that are derived from acoustic information encoded in higher-frequency band neural activity (~80–2000 Hz) include measures of the frequency-following response (FFR) and onset response, and are important in the neural representation of speech elements like vowels and consonants. The strength, timing, and variability of neural encoding can be quantified in a variety of ways, and have been associated with the brainstem and early auditory cortical activity (Skoe and Kraus, 2010; Coffey et al., 2016). Measures derived from lower-frequency band activity (~2–40 Hz) include event-related potentials (ERPs) such as 'P2′ that are generated in the auditory cortex and surrounding areas (Key et al., 2005). The relationships between FFR and SIN perception have previously been reviewed by (Du et al., 2011), as has sensory-cognitive interaction in the neural encoding of SIN (Anderson and Kraus, 2010), and functional neuroimaging of auditory scene analysis (Gutschalk and Dykstra, 2014). The studies on musicianship enhancement covered here are generally in agreement with previous work showing that spectral amplitudes in the FFR (e.g. Parbery-Clark et al., 2011a,b; Coffey et al., 2016; Lee et al., 2009), the timing of transient response peaks (e.g. Parbery-Clark et al., 2012a; Musacchia et al., 2007), as well as ERP waves like P2 and N400 (e.g. Coffey et al., 2016; Zendel et al., 2015), are all enhanced in musicians under some listening conditions and correlate with better perception. This evidence therefore points to improvements occurring at multiple levels of processing.

EEG provides the fine temporal resolution that is necessary to study rapid fluctuations in neural activity, but to localize the brain structures that are involved, methods that offer spatial information such as fMRI are also needed. Cumulative results from studies of speech perception have shown that it is supported by networks of multiple brain regions. Incoming auditory information from the brainstem and thalamus passes first to bilateral auditory cortices in the temporal lobes for spectrotemporal analysis, and from there, along two neural pathways: a bilateral ventral stream that processes speech signals for comprehension (middle temporal gyrus, inferior temporal sulcus), and a left-lateralized dorsal stream that maps acoustic speech signals to frontal lobe articulatory motor planning networks (inferior frontal gyrus, premotor cortex, parieto-temporal boundar; yHickok and Poeppel, 2007; Poeppel, 2014). fMRI-based experimental designs have been used to study SIN perception in specific populations, for example older adults, who were found to have increased activity in dorsal areas and decreased activity in the auditory cortex relative to younger controls when listening to SIN, suggesting frontal compensation for declining perceptual abilities (e.g. Wong et al., 2009; Du et al., 2014). Related work on the neural processing of different types of speech maskers suggests that the sensitivity of behavioural results to small differences in task design (Fig. 1, described above) is paralleled by differences in the dependence of these tasks on different brain structures (e.g. Scott et al., 2009).

Musical activities involve many of the same brain structures as are active in speech perception, including the auditory cortex, premotor and supplementary motor areas, and frontal areas (Zatorre et al., 2007). These areas in turn overlap with brain regions that have been implicated in SIN perception: the superior temporal gyrus, middle temporal gyrus, inferior frontal gyrus, premotor cortex, and parietal areas (Du et al., 2016; Wong et al., 2009). The mechanisms through which sharing brain structures might influence SIN perception are not yet clear. For example, regions whose fMRI activity correlated with phoneme-in-noise accuracy in Du et al. (2014) are shown in Fig. 2. These results confirm the involvement and separation of dorsal and ventral stream areas. Dorsal activation increased with poor performance and increasing noise masking. Activity in the dorsal stream might aid SIN perception by providing articulatory predictions to constrain auditory perception in noisy conditions. Only one of the musician versus SIN perception studies reviewed here used fMRI, in a phoneme-in-noise design (Du and Zatorre, 2016). The results suggest that improved SIN perception in musicians may be related to stronger phoneme specificity of both the ventral and dorsal auditory streams in both hemispheres, but for the most taxing conditions it is the dorsal, motor-related structures that play the most significant role.

Currently available neuroiming data agrees with behavioural data that both lower-level and higher-level neural processes contribute to a musician advantage in SIN perception, but have only been applied in a limited fashion to a small subset of experimental paradigms. New neuroimaging techniques such as stimulus reconstruction for EEG/MEG (e.g. Ding and Simon, 2012) and multivariate pattern recognition (e.g. Du and Zatorre, 2016) show promise as means of extending the reach of EEG and fMRI methods into more complex and fine-grained aspects of SIN processing; however, these methods offer either fine temporal resolution (EEG/MEG) or spatial resolution (fMRI), but usually not both. In order to clarify exactly which overlapping mechanisms between language and music processing might be responsible for a musician SIN advantage, it will be necessary to understand how sensory information is transformed from sensory representations of physical stimuli into perceptual representations within sound streams that

are processed at higher levels (Rauschecker and Scott, 2009; Scott et al., 2009). This effort would be greatly helped by combining methods so as to obtain both sufficient temporal and spatial resolution to bridge across studies of different levels of the auditory system.

New efforts to apply MEG distributed source modelling to the auditory system offer the possibility of studying bottom-up sound processing and higher-level processes and their interaction. Coffey et al. (2016, submitted) recorded the FFR and cortical evoked response to a repeated phoneme using MEG, and modelled neural sources on individual T1-weighted MRI anatomy. They were able to localize the sources of each signal and relate them to offline SIN performance and musical experience. These findings showed that listeners with stronger representation of the fundamental frequency in the FFR also had higher amplitude cortical event-related potentials at ~200 ms after stimulation (i.e. P2 component), and tended to be musicians with early musical experience. Both signals were localized to a right-lateralized portion of the auditory cortex, and anterior superior temporal gyrus brain regions. Based on these data the authors concluded that initial encoding of sound information, even when collected in quiet, indexes the quality of information that is available for further processing, and that musical training is related to this early encoding process. Further developing this novel method will help connect SIN sub-processes to knowledge about brain processes subserving linguistic tasks and more domain-general functions, and deepen our understanding of SIN perception.

8. Systematic consideration of task requirements

In Fig. 1, studies that systematically vary cue parameters are represented by connecting lines. For example, Fuller et al. (2014) varied the importance of pitch cues on the task performance, and Parbery-Clark et al. (2009b) used several clinical tasks that varied in noise and target properties. These studies both contributed valuable pieces of information for specific questions, but as their main conclusions about a musician advantage differed, it would be helpful to have a more precise view of how their experimental designs relate to a wider framework. While we promote neuroimaging as a means of better understanding the relative contributions of SIN task sub-processes, eventual results from such an effort could prove similarly difficult to interpret due to the vast design space. We propose a new strategy to compare study designs.

A 'task decomposition' is a depiction of how a given task may be accomplished in terms of distinct sub-tasks (Coffey and Herholz, 2013). This method borrows from a well-established process of instructional design that is used in safety-critical professions and environments (e.g. aviation, medicine, and Antarctica overwinter crew) to ensure behavioural proficiency (Dick et al., 2004); here we apply it to neuroimaging. In Fig. 3, we propose a task decomposition for sentences in speech-shaped noise. While it is not always possible to prepare a task decomposition that is free from assumptions about cognition, the goal is to produce a behavioural rather than a cognitive model. It is nonetheless useful to clearly present any cognitive assumptions such that they may be compared with existing cognitive models, challenged, and possibly tested using neuroimaging (cognitive assumptions are represented in italics in Fig. 3).

Variations of the task decomposition in Fig. 3 can readily be elaborated, for example, the words-in-noise task includes lexical and phonemic cues but not syntactic cues; or linguistic cues could be added to the irrelevant stream in a speech-on-speech task. Comparing the studies included in Fig. 1 would reveal multiple differences between tasks in the cues that are offered in addition to the target and distractor cue richness that is represented; each of the cognitive processes associated with them might be responsible for a SIN enhancement or differences in whether one is found between studies. Although it might not solve the problem of multidimensional interactions between cues, representing new study designs in this way would allow the cause of inter-group differences to be narrowed to suspected factors, and might help to pinpoint the influence of specific cues on naturalistic SIN behaviour. Task decomposition could also be combined with mathematical modelling (e.g. Varnet et al., 2015; Anderson et al., 2013) and neuroimaging tools that such as those that decode the information content of observed neurophysioloical signals (e.g. Crosse et al., 2016; O'Sullivan et al., 2015) in order to provide insight into SIN processing advantages.

9. Future directions

Results on SIN perception enhancements in relation to musicianship to date have suggested that this line of research will improve our understanding of basic and higher-level auditory processing, experience-dependent enhancements in the brain, and multimodal interaction, as well as provide guidance for the treatment of SIN perception deficits. We recommend several research approaches towards these ends:

- consider the nature of the SIN task in detail (e.g. using task decomposition) and design new studies that target individual cues and systematically vary their relevance
- study the perception of words, phonemes, and tones in a variety of background noises; when studying simpler speech units,

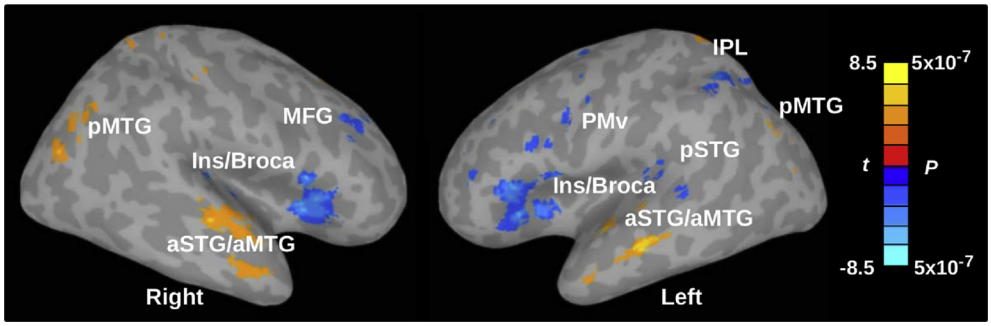

Fig. 2. fMRI BOLD activity is correlated with accuracy on a phoneme-in-noise task, with dorsal and ventral regions exhibiting opposite correlations: better performance was related to higher activity in ventral stream areas and lower activity dorsal stream areas. Reproduced from (Du et al., 2014). Maps are thresholded at FWE-corrected $p < 0.01$ with a cluster size $\geq$342 mm^3; 't' refers to the associated t-statistic (aSTG/aMTG, anterior superior temporal gyrus and anterior middle temporal gyrus; Ins/Broca, insula and Broca's area; IPL, inferior parietal lobule; MFG, middle frontal gyrus; pMTG, posterior middle temporal gyrus; PMv, ventral premotor cortex; pSTG, posterior superior temporal gyrus).

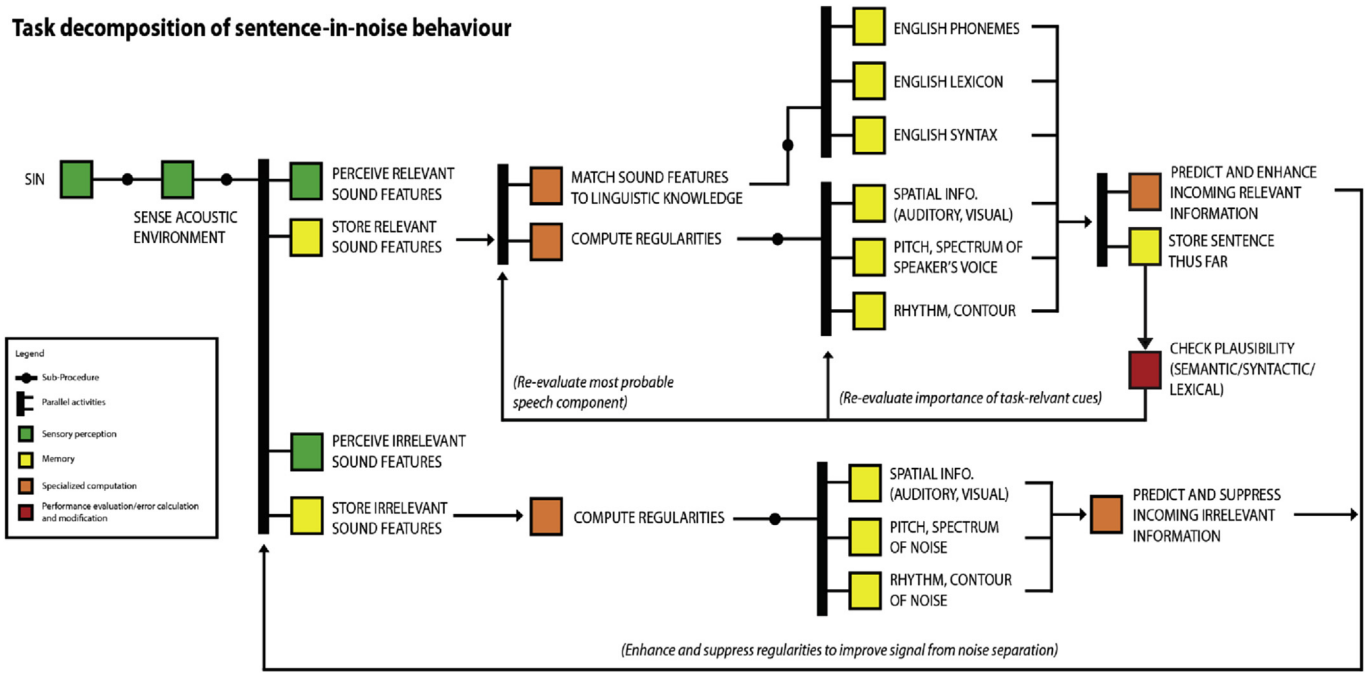

Fig. 3. Speech-in-noise (SIN) task decomposition, showing a sentence-based task. The process begins with sensing and storing in memory a combination of relevant and irrelevant auditory information (their separation is not assumed a priori, but is represented separately to illustrate feedback from top-down effects). In the relevant sound, regularities are extracted about the location of the sound's origin, the pitch of the speaker's voice, and its rhythm and contour. Incoming information is matched to knowledge of the language (to the degree that the speaker is an expert in it), by drawing on the listener's experience with phonemes, vocabulary, and syntax. Together, analysis of these cues is used to predict subsequent words, and their timing and acoustic properties, and to enhance incoming information that matches the constraints – processes that might be enhanced by better higher-level cognitive function, or more heavily relied upon when acoustic cues are degraded. As the representation of a speech stream grows with each new word, it can be evaluated for plausibility and updated with secondary guesses. The irrelevant sound in this case offers limited information, but regularities in the acoustic properties may be of some help to predict and suppress it. Symbols are defined in legend; italics indicate cognitive assumptions.

include a sentence-in-noise condition to assess relevance of findings to more naturalistic SIN listening
- apply neuroimaging tools, particularly those that offer both spatially and temporally-resolved data (EEG-fMRI, MEG), in order to clarify the underlying neural mechanisms of SIN perception throughout the auditory system
- apply stimulus reconstruction and neural decoding methods to better understand how neural representations are strengthened by training
- investigate the effect of difficulty level on listening strategy and the nature and limits of compensatory mechanisms
- explore differences in linguistic and musical experience and their relation to specific cues, particularly spatial and visual cues, which have been underrepresented; comparison with populations with SIN deficits may also be revealing
- study the relationship between SIN perception and musicianship in non-English linguistic populations, in which different cues may be most critical
- investigate causal mechanisms of SIN enhancement by using brain stimulation techniques to perturb specific circuits and evaluate their contribution to perception under varying circumstances
- use longitudinal training designs to confirm the causal effects of specific enhancing processes on naturalistic SIN perception. Although naturalistic training is likely to be more effective for their reward and motivational value (Patel, 2014), it will be necessary to establish the relevant factors in order to make sure they are represented
- develop and test interventions to support or compensate for weaker SIN sub-processes in vulnerable populations

10. Recent additions

Several additional studies were published while this manuscript was in review that bear on questions related to SIN enhancement in musicians. Non-exhaustively, Anaya et al. (2016) measured the performance of musicians and age-matched controls on SIN tests and tests of environmental sounds in noise, as well as tasks of visual perception under degraded conditions. They reported that musicians showed better performance identifying degraded speech presented in both the auditory and visual modalities, but not identifying environmental sounds in noise, suggesting that musicians possess superior processing skills that might be specific to language abilities but not limited to the auditory modality. Donai and Jennings (2016) tested the ability of musicians and non-musicians to detect the gender of speakers of spectro-temporally degraded vowel segments, as well as their gap detection thresholds. Although no differences were observed in the gender identification task, musicians demonstrated shorter gap detection thresholds, which may have implications for processing speech in degraded listening conditions. Habibi et al. (2016) conducted a longitudinal study showing improved auditory processing (i.e. enhanced ability to detect changes in tonal environment and accelerated maturation of cortical evoked potentials to musical notes) in school-aged children who participated in musical training over a 2-year period as compared with other groups of children who practised sports and visual arts. Zhao & Kuhl (2016) reported that a music intervention in 9 month-old babies appears to improve detection and prediction of auditory patterns. These two studies are not strictly studies of SIN perception, but measured how basic auditory processing skills are affected by musical training, and likely support a range of auditory skills including SIN. Generally,

these examples of ongoing work further support links between both basic processing and domain-general cognitive resources that support SIN perception. These studies also continue to demonstrate the capability for plasticity within the auditory system (which may partly explain associations in adulthood between musical training and auditory processing enhancements), and also validate ongoing efforts to develop training-based interventions for SIN deficits.

11. Conclusions

The preponderance of the evidence from the twenty or so papers that have investigated musician advantage in SIN perception supports a music training-related group difference over a wide range of conditions that vary in target and distractor characteristics, along with differences in the presence of spatial, visual, and linguistic information. Each of these cues has been shown to be both relevant to SIN perception and enhanced in musicians; however, existing data do not lead to a clear understanding of exactly how musical training might lead to SIN perception enhancement due to lack of comprehensive coverage, and inconsistency of results in cases in which group advantages are not observed. Because SIN problems can be solved using multiple cues, many paradigms lack specificity: when listeners are not forced to rely on specific cues, their performance is difficult to attribute to specific aspects of task design. It would therefore be useful to design new experiments that systematically investigate the neurophysiological correlates and performance outcomes associated with SIN task sub-components, with a view to comparing results across studies in addition to answering focused research questions. Ultimately, to influence behaviour in a specific manner, for example to design clinical interventions for SIN perceptual deficits, it will be necessary to understand the component parts and processes (and how they interact) more comprehensively. This research direction would be advanced by applying neuroimaging tools that yield spatial information to SIN research with what is known of speech processing, auditory stream segregation, and musician enhancement from cognitive neuroscience.

Funding sources

The research was supported by operating grants to R.J.Z. from the Canadian Institutes of Health Research and from the Canada Fund for Innovation. This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa Canada. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

References

Agus, Trevor R., Thorpe, Simon J., Pressnitzer, Daniel, 2010. Rapid formation of robust auditory memories: insights from noise. Neuron 66 (4), 610–618. http://dx.doi.org/10.1016/j.neuron.2010.04.014.

Alain, Claude, Rich Zendel, Benjamin, Hutka, Stefanie, Bidelman, Gavin M., 2014. Turning down the noise: the benefit of musical training on the aging auditory brain. Hear. Res. 308 (February), 162–173. http://dx.doi.org/10.1016/j.heares.2013.06.008.

Anaya, Esperanza M., Pisoni, David B., Kronenberger, William G., 2016. Long-term musical experience and auditory and visual perceptual abilities under adverse conditions. J. Acoust. Soc. Am. 140 (3) http://dx.doi.org/10.1121/1.4962628. Acoustical Society of AmericaASA: 2074–81.

Anderson, Samira, Kraus, Nina, 2010. Sensory-cognitive interaction in the neural encoding of speech in noise: a review. J. Am. Acad. Audiol. 21 (9), 575–585. http://dx.doi.org/10.3766/jaaa.21.9.3.

Anderson, Samira, White-Schwoch, Travis, Parbery-Clark, Alexandra, Kraus, Nina, 2013. A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. Hear. Res. 300 (June), 18–32. http://dx.doi.org/10.1016/j.heares.2013.03.006.

Angulo-Perkins, Arafat, Aubé, William, Peretz, Isabelle, Barrios, Fernando A., Armony, Jorge L., Concha, Luis, 2014. Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. Cortex 59, 126–137. http://dx.doi.org/10.1016/j.cortex.2014.07.013.

Assmann, Peter, Summerfield, Quentin, 2004. The Perception of Speech under Adverse Conditions. Speech Processing in the Auditory System. Springer-Verlag, New York, pp. 231–308. http://dx.doi.org/10.1007/0-387-21575-1_5.

Avan, Paul, Giraudet, Fabrice, Büki, Béla, 2015. Importance of binaural hearing. Audiol. Neuro Otol. 20 (Suppl. 1), 3–6. http://dx.doi.org/10.1159/000380741.

Bailey, Jennifer A., Penhune, Virginia B., 2010. Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. Exp. Brain Res. 204 (1), 91–101. http://dx.doi.org/10.1007/s00221-010-2299-y. Springer-Verlag.

Barker, Brittan A., Newman, Rochelle S., 2004. Listen to your mother! The role of talker familiarity in infant streaming. Cognition 94. http://dx.doi.org/10.1016/j.cognition.2004.06.001.

Bartoli, Eleonora, D'Ausilio, Alessandro, Berry, Jeffrey, Badino, Leonardo, Bever, Thomas, Fadiga, Luciano, 2015. Listener-speaker perceived distance predicts the degree of motor contribution to speech perception. Cereb. Cortex 25 (2), 281–288. http://dx.doi.org/10.1093/cercor/bht257. Oxford University Press.

Başkent, Deniz, Gaudrain, Etienne, 2016. Musician advantage for speech-on-speech perception. J. Acoust. Soc. Am. 139 (3), EL51–EL56. http://dx.doi.org/10.1121/1.4942628. Acoustical Society of America.

Bendixen, Alexandra, 2014. Predictability effects in auditory scene analysis: a review. Front. Neurosci. http://dx.doi.org/10.3389/fnins.2014.00060. Frontiers.

Bernstein, Joshua G.W., Grant, Ken W., 2009. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 125 (5), 3358. http://dx.doi.org/10.1121/1.3110132. Acoustical Society of America.

Bey, Caroline, McAdams, Stephen, 2002. Schema-based processing in auditory scene analysis. Percept. Psychophys. 64 (5), 844–854. http://dx.doi.org/10.3758/BF03194750. Springer-Verlag.

Bidelman, Gavin M., Alain, Claude, 2015. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. J. Neurosci. 35 (3), 1240–1249. http://dx.doi.org/10.1523/JNEUROSCI.3292-14.2015.

Bidelman, Gavin M., Gandour, Jackson T., Krishnan, Ananthanarayan, 2011. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cognitive Neurosci. 23 (2), 425–434. http://dx.doi.org/10.1162/jocn.2009.21362. MIT Press238 Main St., Suite 500, Cambridge, MA. 02142-1046USAjournals-info@mit.edu.

Bidelman, Gavin M., Hutka, Stefanie, Moreno, Sylvain, 2013. Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. PLoS ONE 8 (4), e60676. http://dx.doi.org/10.1371/journal.pone.0060676. Public Library of Science.

Bidelman, G.M., Weiss, M.W., 2014. "Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians." Eur. J. Neurosci. 40 (4), 2662–2672. http://onlinelibrary.wiley.com/doi/10.1111/ejn.12627/full.

Bishop, Christopher W., Miller, Lee M., 2009. A multisensory cortical network for understanding speech in noise. J. Cognitive Neurosci. 21 (9), 1790–1804. http://dx.doi.org/10.1162/jocn.2009.21118. MIT Press238 Main St., Suite 500, Cambridge, MA 02142–1046, USA. journals-info@mit.edu.

Boebinger, Dana, Evans, Samuel, Rosen, Stuart, Lima, César F., Manly, Tom, Scott, Sophie K., 2015. Musicians and non-musicians are equally adept at perceiving masked speech. J. Acoust. Soc. Am. 137 (1), 378–387. http://dx.doi.org/10.1121/1.4904537. Acoustical Society of America.

Brandler, Susanne, Rammsayer, Thomas H., 2003. Differences in mental abilities between musicians and non-musicians. Psychol. Music 31 (2), 123–138. http://dx.doi.org/10.1177/0305735603031022290. Sage PublicationsSage CA: Thousand Oaks, CA.

Bregman, Albert S., 1994. Auditory Scene Analysis: the Perceptual Organization of Sound. MIT Press, Boston, MA. https://books.google.com/books?hl=en&lr=&id=jI8muSpAC5AC&pgis=1.

Carey, Daniel, Rosen, Stuart, Krishnan, Saloni, Pearce, Marcus T., Shepherd, Alex, Aydelott, Jennifer, Dick, Frederic, 2015. Generality and specificity in the effects of musical expertise on perception and cognition. Cognition 137, 81–105. http://dx.doi.org/10.1016/j.cognition.2014.12.005.

Chan, Agnes S., Ho, Yim-Chi, Cheung, Mei-Chun, 1998. Music training improves verbal memory. Nature 396 (6707), 128. http://dx.doi.org/10.1038/24075. Nature Publishing Group.

Clayton, Kameron K., Swaminathan, Jayaganesh, Yazdanbakhsh, Arash, Zuk, Jennifer, Patel, Aniruddh D., Kidd, Gerald, 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. PLOS ONE. In: Snyder, Joel (Ed.), Public Library of Science, vol. 11(7), p. e0157638. http://dx.doi.org/10.1371/journal.pone.0157638.

Coffey, Emily B.J., Chepesiuk, A.M.P., Herholz, S.C., Baillet, S., Zatorre, Robert J., 2016. Neural correlates of early sound encoding and their relationship to speech in noise perception. bioRxiv. http://dx.doi.org/10.1101/076455 (Submitted).

Coffey, Emily B.J., Herholz, Sibylle C., 2013. Task decomposition: a framework for comparing diverse training models in human brain plasticity studies. Front. Hum. Neurosci. 7 (640), 1–6. http://dx.doi.org/10.3389/fnhum.2013.00640.

Coffey, Emily B.J., Herholz, Sibylle C., Chepesiuk, Alexander M.P., Baillet, Sylvain, Zatorre, Robert J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7 (March), 11070. http://dx.doi.org/10.1038/ncomms11070. Nature Publishing Group.

Corrigall, Kathleen A., Glenn Schellenberg, E., Misura, Nicole M., 2013. Music training, cognition, and personality. Front. Psychol. 4, 222. http://dx.doi.org/10.3389/fpsyg.2013.00222. Frontiers.

Crosse, Michael J., Di Liberto, Giovanni M., Lalor, Edmund C., 2016. Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. J. Neurosci. Official J. Soc. Neurosci. 36 (38), 9888–9895. http://dx.doi.org/10.1523/JNEUROSCI.1396-16.2016. Society for Neuroscience.

D'Ausilio, Alessandro, Bufalari, Ilaria, Salmas, Paola, Fadiga, Luciano, 2012. The role of the motor system in discriminating normal and degraded speech sounds. Cortex 48 (7), 882–887. http://dx.doi.org/10.1016/j.cortex.2011.05.017.

de Villers-Sidani, Etienne, Simpson, Kimberly L., Lu, Y.-F., Lin, Rick C.S., Merzenich, Michael M., 2008. Manipulating critical period closure across different sectors of the primary auditory cortex. Nat. Neurosci. 11 (8), 957–965. http://dx.doi.org/10.1038/nn.2144. Nature Publishing Group.

Dick, W., Carey, L., Carey, J., 2004. The Systematic Review of Instruction, sixth ed. Allyn & Bacon, Boston, MA.

Ding, Nai, Simon, Jonathan Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. U. S. A. 109 (29), 11854–11859. http://dx.doi.org/10.1073/pnas.1205381109. National Academy of Sciences.

Dirks, Donald D., Wilson, Richard H., 1969. The effect of spatially separated sound sources on speech intelligibility. J. Speech Lang. Hear. Res. 12 (1), 5. http://dx.doi.org/10.1044/jshr.1201.05. American Speech-Language-Hearing Association.

Donai, Jeremy J., Jennings, Mariah B., 2016. Gaps-in-Noise detection and gender identification from noise-vocoded vowel segments: comparing performance of active musicians to non-musicians. J. Acoust. Soc. Am. 139 (5), EL128–EL134. http://dx.doi.org/10.1121/1.4947070. Acoustical Society of AmericaASA.

Du, Yi, Buchsbaum, Bradley R., Grady, Cheryl L., Alain, Claude, 2016. Increased activity in frontal motor cortex compensates impaired speech perception in older adults. Nat. Commun. 7 (August), 12241. http://dx.doi.org/10.1038/ncomms12241. Nature Research.

Du, Yi, Buchsbaum, Bradley R., Grady, Cheryl L., Alain, Claude, 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc. Natl. Acad. Sci. U. S. A. 111 (19), 7126–7131. http://dx.doi.org/10.1073/pnas.1318738111.

Du, Yi, Kong, Lingzhi, Wang, Qian, Wu, Xihong, Li, Liang, 2011. Auditory frequency-following response: a neurophysiological measure for studying the 'cocktail-Party problem'. Neurosci. Biobehav. Rev. 35 (10), 2046–2057. http://dx.doi.org/10.1016/j.neubiorev.2011.05.008.

Du, Yi, Zatorre, Robert J., 2016. How Musicians Perceive Speech in Noise: Role of the Ventral and Dorsal Streams. In Organization for Human Brain Mapping.

Freyman, Richard L., Helfer, Karen S., Balakrishnan, Uma, 2005. Spatial and spectral factors in release from informational masking in speech recognition. Acta Acustica United Acustica 91 (3), 537–545. http://dx.doi.org/10.1121/1.1354984.

Fuller, Christina D., Galvin, John J., Maat, Bert, Free, Rolien H., Başkent, Deniz, 2014. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? Front. Neurosci. 8 (179), 1–16. http://dx.doi.org/10.3389/fnins.2014.00179.

Gifford, René H., Revit, Lawrence J., 2010. Speech perception for adult cochlear implant recipients in a realistic background noise: effectiveness of preprocessing strategies and external options for improving speech recognition in noise. J. Am. Acad. Audiol. 21 (7), 441–451. http://dx.doi.org/10.3766/jaaa.21.7.3.

Gockel, H.E., Carlyon, R.P., Mehta, A., Plack, C.J., Hristopher, C., Lack, J.P., 2011. The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. J. Assoc. Res. Otolaryngol. 782, 767–782. http://dx.doi.org/10.1007/s10162-011-0284-1.

Golestani, N., Rosen, S., Scott, S.K., 2009. Native-language benefit for understanding speech-in-noise: the contribution of semantics. Bilingualism-Language Cognition 12 (3), 385–392. http://dx.doi.org/10.1017/S1366728909990150. Cambridge University Press.

Grahn, Jessica A., 2012. Neural mechanisms of rhythm perception: current findings and future perspectives. Top. Cognitive Sci. 4 (4), 585–606. http://dx.doi.org/10.1111/j.1756-8765.2012.01213.x. Blackwell Publishing Ltd.

Gutschalk, Alexander, Dykstra, Andrew R., 2014. Functional imaging of auditory scene analysis. Hear. Res. 307, 98–110. http://dx.doi.org/10.1016/j.heares.2013.08.003.

Habibi, Assal, Cahn, B. Rael, Damasio, Antonio, Damasio, Hanna, 2016. Neural correlates of accelerated auditory processing in children engaged in music training. Dev. Cognitive Neurosci. 21, 1–14. http://dx.doi.org/10.1016/j.dcn.2016.04.003.

Hawley, Monica L., Litovsky, Ruth Y., Culling, John F., 2004. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. J. Acoust. Soc. Am. 115 (2), 833–843. http://www.ncbi.nlm.nih.gov/pubmed/15000195.

Helfer, Karen S., Freyman, Richard L., 2005. The role of visual speech cues in reducing energetic and informational masking. J. Acoust. Soc. Am. 117 (2), 842. http://dx.doi.org/10.1121/1.1836832. Acoustical Society of America.

Herholz, Sibylle C., Zatorre, Robert J., 2012. Musical training as a framework for brain plasticity: behavior, function, and structure. Neuron 76 (3), 486–502. http://dx.doi.org/10.1016/j.neuron.2012.10.011.

Hickok, Gregory, Poeppel, David, 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402. http://dx.doi.org/10.1038/nrn2113.

Ho, Yim-Chi, Cheung, Mei-Chun, Chan, Agnes S., 2003. Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. Neuropsychology 17 (3), 439–450. http://dx.doi.org/10.1037/0894-4105.17.3.439. American Psychological Association.

Hornickel, Jane, Chandrasekaran, Bharath, Zecker, Steve, Kraus, Nina, 2011. Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. Behav. Brain Res. 216 (2), 597–605. http://dx.doi.org/10.1016/j.bbr.2010.08.051.

Hornickel, Jane, Skoe, Erika, Nicol, Trent, Zecker, Steven, Kraus, Nina, 2009. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc. Natl. Acad. Sci. U. S. A. 106 (31), 13022–13027. http://dx.doi.org/10.1073/pnas.0901123106.

Ioannidis, John P.A., Munafò, Marcus R., Fusar-Poli, Paolo, Nosek, Brian A., David, Sean P., 2014. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends Cognitive Sci. 18 (5), 235–241. http://dx.doi.org/10.1016/j.tics.2014.02.010.

Key, Alexandra P Fonaryova, Dove, Guy O., Maguire, Mandy J., 2005. Linking brainwaves to the brain: an ERP primer. Dev. Neuropsychol. 27 (2), 183–215. http://dx.doi.org/10.1207/s15326942dn2702_1. Lawrence Erlbaum Associates, Inc.

Killion, Mead C., Niquette, Patricia A., Gudmundsen, Gail I., Revit, Lawrence J., Banerjee, Shilpi, 2004. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 116 (4), 2395–2405. http://dx.doi.org/10.1121/1.1784440. Acoustical Society of America.

Klatte, Maria, Lachmann, Thomas, Meis, Markus, 2009. Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. Noise Health 12 (49), 270–282. http://dx.doi.org/10.4103/1463-1741.70506.

Kraus, Nina, Strait, Dana L., Parbery-Clark, Alexandra, 2012. Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. Ann. N. Y. Acad. Sci. 1252 (April), 100–107. http://dx.doi.org/10.1111/j.1749-6632.2012.06463.x.

Kraus, Nina, White-Schwoch, Travis, 2015. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. Trends Cognitive Sci. 19 (11), 642–654. http://www.sciencedirect.com/science/article/pii/S1364661315002089.

Leaver, Amber M., Rauschecker, Josef P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. 30 (22), 7604–7612. http://dx.doi.org/10.1523/JNEUROSCI.0296-10.2010.

Lee, Hweeling, Noppeney, Uta, 2011. Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc. Natl. Acad. Sci. U. S. A. 108 (51), E1441–E1450. http://dx.doi.org/10.1073/pnas.1115267108. National Academy of Sciences.

Lee, Hweeling, Noppeney, Uta, 2014. Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music. Front. Psychol. 5 (August), 868. http://dx.doi.org/10.3389/fpsyg.2014.00868. Frontiers.

Lee, Kyung Myun, Skoe, Erika, Kraus, Nina, Ashley, Richard, 2009. Selective subcortical enhancement of musical intervals in musicians. J. Neurosci. 29 (18), 5832–5840. http://europepmc.org/abstract/MED/19420250/reload=0.

MacKeith, N.W., Coles, R.R., 1971. Binaural advantages in hearing of speech. J. Laryngol. Otol. 85 (3), 213–232. http://dx.doi.org/10.1017/S0022215100073369. Cambridge University Press.

Maddox, Ross K., Atilgan, Huriye, Bizley, Jennifer K., Lee, Adrian Kc, 2015. Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. eLife 2015 (4), 1–11. http://dx.doi.org/10.7554/eLife.04995.001 eLife Sciences Publications Limited.

McAuley, J. Devin, Henry, Molly J., Tuft, Samantha, 2011. Musician advantages in music perception: an issue of motivation, Not Just Ability. Music Percept. 28 (5), 505–518. http://dx.doi.org/10.1525/mp.2011.28.5.505.

McDermott, Josh H., Wrobleski, David, Oxenham, Andrew J., 2011. Recovering sound sources from embedded repetition. Proc. Natl. Acad. Sci. U. S. A. 108 (3), 1188–1193. http://dx.doi.org/10.1073/pnas.1004765108. National Academy of Sciences.

Micheyl, Christophe, Delhommeau, Karine, Perrot, Xavier, Oxenham, Andrew J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219 (September), 36–47. http://dx.doi.org/10.1016/j.heares.2006.05.004.

Milovanov, Riia, Tervaniemi, Mari, 2011. The interplay between musical and linguistic aptitudes: a review. Front. Psychol. http://dx.doi.org/10.3389/fpsyg.2011.00321. Frontiers.

Moore, Brian C.J., Gockel, Hedwig, 2002. Factors influencing sequential stream segregation. Acta Acustica United Acustica 88, 320–332. S. Hirzel Verlag. http://www.ingentaconnect.com/content/dav/aaua/2002/00000088/00000003/art00004.

Moreno, Sylvain, Bialystok, Ellen, Barac, Raluca, Glenn Schellenberg, E., Cepeda, Nicholas J., Chau, Tom, 2011. Short-term music training enhances verbal intelligence and executive function. Psychol. Sci. 22 (11), 1425–1433. http://dx.doi.org/10.1177/0956797611416999. SAGE Publications.

Moreno, Sylvain, Bidelman, Gavin M., 2014. Examining neural plasticity and cognitive benefit through the unique lens of musical training. Hear. Res. 308 (February), 84–97. http://dx.doi.org/10.1016/j.heares.2013.09.012.

Mosing, Miriam A., Madison, Guy, Pedersen, Nancy L., Ullén, Fredrik, 2015. Investigating cognitive transfer within the framework of music practice: genetic pleiotropy rather than causality. Dev. Sci. 3 (3), 1–9. http://dx.doi.org/10.1111/desc.12306.

Münte, Thomas F., Kohlmetz, Christine, Nager, Wido, Altenmüller, Eckart, 2001.

Neuroperception: superior auditory spatial tuning in conductors. Nature 409 (6820), 580. http://dx.doi.org/10.1038/35054668. Nature Publishing Group.
Musacchia, Gabriella, Sams, Mikko, Skoe, Erika, Kraus, Nina, 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Natl. Acad. Sci. 104 (40), 15894–15898. http://www.pnas.org/content/104/40/15894.
Musacchia, Gabriella, Strait, Dana L., Kraus, Nina, 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear. Res. 241 (1–2), 34–42. http://www.sciencedirect.com/science/article/pii/S0378595508000798.
Nilsson, Michael, 1994. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 95 (2), 1085–1099. http://dx.doi.org/10.1121/1.408469. Acoustical Society of America.
O'Sullivan, James A., Power, Alan J., Mesgarani, Nima, Rajaram, Siddharth, Foxe, John J., Shinn-Cunningham, Barbara G., Slaney, Malcolm, Shamma, Shihab A., Lalor, Edmund C., 2015. Attentional selection in a cocktail party environment can Be decoded from single-trial EEG. Cereb. Cortex 25 (7), 1697–1706. http://dx.doi.org/10.1093/cercor/bht355.
Oxenham, Andrew J., Fligor, Brian J., Mason, Christine R., Kidd, Gerald, 2003. Informational masking and musical training. J. Acoust. Soc. Am. 114 (3), 1543–1549. http://dx.doi.org/10.1121/1.1598197. Acoustical Society of America.
Pallesen, Karen Johanne, Brattico, Elvira, Bailey, Christopher J., Korvenoja, Antti, Koivisto, Juha, Gjedde, Albert, Carlson, Synnöve, 2010. Cognitive control in auditory working memory is enhanced in musicians. PLoS ONE. In: Warrant, Eric (Ed.), Public Library of Science, vol. 5(6), p. e11120. http://dx.doi.org/10.1371/journal.pone.0011120.
Pantev, Christo, Herholz, Sibylle C., 2011. Plasticity of the human auditory cortex related to musical training. Neurosci. Biobehav. Rev. 35 (10), 2140–2154. http://dx.doi.org/10.1016/j.neubiorev.2011.06.010.
Parbery-Clark, Alexandra, Anderson, Samira, Hittner, Emily, Kraus, Nina, 2012a. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. Front. Aging Neurosci. 4 (OCT), 30. http://dx.doi.org/10.3389/fnagi.2012.00030. Frontiers.
Parbery-Clark, Alexandra, Anderson, Samira, Hittner, Emily, Kraus, Nina, 2012b. Musical experience offsets age-related delays in neural timing. Neurobiol. Aging 33 (7), 1483. http://dx.doi.org/10.1016/j.neurobiolaging.2011.12.015 e1-1483.e4.
Parbery-Clark, Alexandra, Skoe, Erika, Kraus, Nina, 2009b. Musical experience limits the degradative effects of background noise on the neural processing of sound. J. Neurosci. 29 (45), 14100–14107. http://dx.doi.org/10.1523/JNEUROSCI.3256-09.2009.
Parbery-Clark, Alexandra, Skoe, Erika, Lam, Carrie, Kraus, Nina, 2009a. Musician enhancement for speech-in-noise. Ear Hear. 30 (6), 653–661. http://dx.doi.org/10.1097/AUD.0b013e3181b412e9.
Parbery-Clark, Alexandra, Strait, Dana L., Anderson, Samira, Hittner, Emily, Kraus, Nina, 2011a. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. PLoS ONE 6 (5), 1–8. http://dx.doi.org/10.1371/journal.pone.0018082.
Parbery-Clark, Alexandra, Strait, Dana L., Kraus, Nina, 2011b. Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. Neuropsychologia 49 (12), 3338–3345. http://dx.doi.org/10.1016/j.neuropsychologia.2011.08.007.
Parbery-Clark, Alexandra, Tierney, A., Strait, Dana L., Kraus, N., 2012. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience 219 (September), 111–119. http://dx.doi.org/10.1016/j.neuroscience.2012.05.042.
Patel, Aniruddh D., 2014. Can nonlinguistic musical training change the way the brain processes Speech? The expanded OPERA hypothesis. Hear. Res. 308 (February), 98–108. http://dx.doi.org/10.1016/j.heares.2013.08.011.
Peretz, Isabelle, Vuvan, Dominique, Lagrois, Marie-Élaine, Armony, Jorge L., 2015. Neural overlap in processing music and speech. Philosophical Trans. R. Soc. Lond. Ser. B Biol. Sci. 370 (1664), 20140090. http://dx.doi.org/10.1098/rstb.2014.0090.
Pickering, Martin J., Garrod, Simon, 2007. Do people use language production to make predictions during comprehension? Trends Cognitive Sci. 11 (3), 105–110. http://dx.doi.org/10.1016/j.tics.2006.12.002.
Poeppel, David, 2014. The neuroanatomic and neurophysiological infrastructure for speech and language. Curr. Opin. Neurobiol. 28C (October), 142–149. http://dx.doi.org/10.1016/j.conb.2014.07.005.
Pressnitzer, Daniel, Suied, Clara, Shamma, Shihab, 2011. Auditory scene analysis: the sweet music of ambiguity. Front. Hum. Neurosci. 5, 158. http://dx.doi.org/10.3389/fnhum.2011.00158. Frontiers.
Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12 (6), 718–724. NATURE PUBLISHING GROUP. http://discovery.ucl.ac.uk/150260/.
Rimmele, Johanna M., Zion Golumbic, Elana, Schröger, Erich, Poeppel, David, 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex 68, 144–154. http://dx.doi.org/10.1016/j.cortex.2014.12.014.
Ruggles, Dorea R., Freyman, Richard L., Oxenham, Andrew J., 2014. Influence of musical training on understanding voiced and whispered speech in noise. PloS One 9 (1), e86980. http://dx.doi.org/10.1371/journal.pone.0086980. Public Library of Science.
Schellenberg, E. Glenn, Peretz, Isabelle, 2008. Music, language and cognition: unresolved issues. Trends Cognitive Sci. 12 (2), 45–46. http://dx.doi.org/10.1016/j.tics.2007.11.005. Elsevier Science.
Schellenberg, E Glenn, 2015. Music training and speech perception: a gene-environment interaction. Ann. N. Y. Acad. Sci. 1337 (March), 170–177. http://dx.doi.org/10.1111/nyas.12627.
Schön, Daniele, Tillmann, Barbara, 2015. Short- and long-term rhythmic interventions: perspectives for language rehabilitation. Ann. N. Y. Acad. Sci. 1337 (1), 32–39. http://dx.doi.org/10.1111/nyas.12635.
Scott, S.K., Rosen, Stuart, Philip Beaman, C., Davis, Josh P., Wise, Richard J.S., 2009. The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. J. Acoust. Soc. Am. 125 (3), 1737. http://dx.doi.org/10.1121/1.3050255. Acoustical Society of America.
Shook, Anthony, Marian, Viorica, Bartolotti, James, Schroeder, Scott R., 2013. Musical experience influences statistical learning of a novel language. Am. J. Psychol. 126 (1), 95–104. NIH Public Access. http://www.ncbi.nlm.nih.gov/pubmed/23505962.
Skoe, Erika, Kraus, Nina, 2010. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 31 (3), 302–324. http://dx.doi.org/10.1097/AUD.0b013e3181cdb272.
Slater, Jessica, Kraus, Nina, 2016. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. Cogn. Process. 17 (1), 79–87. http://dx.doi.org/10.1007/s10339-015-0740-7.
Slater, Jessica, Skoe, Erika, Strait, Dana L., O'Connell, Samantha, Thompson, Elaine, Kraus, Nina, 2015. Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. Behav. Brain Res. 291 (September), 244–252. http://dx.doi.org/10.1016/j.bbr.2015.05.026.
Song, Judy H., Skoe, Erika, Banai, Karen, Kraus, Nina, 2011. Perception of speech in noise: neural correlates. J. Cognitive Neurosci. 23 (9), 2268–2279. http://dx.doi.org/10.1162/jocn.2010.21556.
Song, Judy H., Skoe, Erika, Banai, Karen, Kraus, Nina, 2012. Training to improve hearing speech in noise: biological mechanisms. Cereb. Cortex 22 (5), 1180–1190. http://dx.doi.org/10.1093/cercor/bhr196.
Souza, Pamela, Gehani, Namita, Wright, Richard, McCloy, Daniel, 2013. The advantage of knowing the talker. J. Am. Acad. Audiol. 24 (8), 689–700. http://dx.doi.org/10.3766/jaaa.24.8.6. American Academy of Audiology.
Strait, Dana L., Kraus, N., Parbery-Clark, Alexandra, Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear. Res. 261, 22–29. http://www.sciencedirect.com/science/article/pii/S0378595509003116.
Strait, Dana L., Kraus, Nina, 2011. Can you hear me Now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Front. Psychol. 2 (113), 1–10. http://dx.doi.org/10.3389/fpsyg.2011.00113.
Suga, N., Gao, E., Zhang, Y., Ma, X., Olsen, J.F., 2000. The corticofugal system for hearing: recent progress. Proc. Natl. Acad. Sci. U. S. A. 97 (22), 11807–11814. http://dx.doi.org/10.1073/pnas.97.22.11807.
Suga, Nobuo, 2012. Tuning shifts of the auditory system by corticocortical and corticofugal projections and conditioning. Neurosci. Biobehav. Rev. 36 (2), 969–988. http://dx.doi.org/10.1016/j.neubiorev.2011.11.006.
Suied, Clara, Bonneel, Nicolas, Viaud-Delmon, Isabelle, 2009. Integration of auditory and visual information in the recognition of realistic objects. Exp. Brain Res. 194 (1), 91–102. http://dx.doi.org/10.1007/s00221-008-1672-6. Springer-Verlag.
Swaminathan, Jayaganesh, Mason, Christine R., Streeter, Timothy M., Best, Virginia, Kidd, Gerald, Patel, Aniruddh D., 2015. Musical training, individual differences and the cocktail party problem. Sci. Rep. 5 (11628), 1–10. http://dx.doi.org/10.1038/srep11628. Nature Publishing Group.
Thompson, Sarah K., Carlyon, Robert P., Cusack, Rhodri, 2011. An objective measurement of the build-up of auditory streaming and of its modulation by attention. J. Exp. Psychol. Hum. Percept. Perform. 37 (4), 1253–1262. http://dx.doi.org/10.1037/a0021925. American Psychological Association.
Tierney, Adam T., Krizman, Jennifer, Kraus, Nina, 2015. Music training alters the course of adolescent auditory development. Proc. Natl. Acad. Sci. 112 (32), 1–6. http://dx.doi.org/10.1073/pnas.1505114112.
Varnet, Léo, Wang, Tianyun, Peter, Chloe, Meunier, Fanny, Hoen, Michel, 2015. How musical expertise shapes speech perception: evidence from auditory classification images. Sci. Rep. 5 (14489), 1–13. http://dx.doi.org/10.1038/srep14489.
Wilson, Richard H., 2003. Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. J. Am. Acad. Audiol. 14 (9), 453–470. American Academy of Audiology.
Wilson, Richard H., McArdle, Rachel A., Smith, Sherri L., 2007. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. J. Speech, Lang. Hear. Res. 50 (4), 844–856. http://dx.doi.org/10.1044/1092-4388(2007/059). American Speech-Language-Hearing Association.
Wong, Patrick C.M., Xumin Jin, James, Gunasekera, Geshri M., Abel, Rebekah, Lee, Edward R., Dhar, Sumitrajit, 2009. Aging and cortical mechanisms of speech perception in noise. Neuropsychologia 47 (3), 693–703. http://dx.doi.org/10.1016/j.neuropsychologia.2008.11.032.
Woodruff Carr, Kali, White-Schwoch, Travis, Tierney, Adam T., Strait, Dana L., Kraus, Nina, 2014. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. Proc. Natl. Acad. Sci. U. S. A. 111 (40), 14559–14564. http://dx.doi.org/10.1073/pnas.1406219111. National Academy of Sciences.
Yonan, Cynthia A., Sommers, Mitchell S., 2000. The effects of talker familiarity on spoken word identification in younger and older listeners. Psychol. Aging 15 (1), 88–99. http://dx.doi.org/10.1037/0882-7974.15.1.88. American Psychological Association.

Zatorre, Robert J., 2013. Predispositions and plasticity in music and speech learning: neural correlates and implications. Science 342 (6158), 585–589. http://dx.doi.org/10.1126/science.1238414. American Association for the Advancement of Science.

Zatorre, Robert J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory–motor interactions in music perception and production. Nat. Rev. Neurosci. 8, 547–558. http://www.nature.com/nrn/journal/v8/n7/abs/nrn2152.html.

Zendel, Benjamin Rich, Alain, Claude, 2009. Concurrent sound segregation is enhanced in musicians. J. Cognitive Neurosci. 21 (8), 1488–1498. http://dx.doi.org/10.1162/jocn.2009.21140. MIT Press238 Main St., Suite 500, Cambridge, MA 02142–1046, USA. journals-info@mit.edu.

Zendel, Benjamin Rich, Alain, Claude, 2012. Musicians experience less age-related decline in central auditory processing. Psychol. Aging 27 (2), 410–417. American Psychological Association. http://cat.inist.fr/?aModele=afficheN&cpsidt=25968784.

Zendel, Benjamin Rich, Tremblay, Charles-David, Belleville, Sylvie, Peretz, Isabelle, 2015. The impact of musicianship on the cortical mechanisms related to separating speech from background noise. J. Cognitive Neurosci. 27 (5), 1044–1059. http://dx.doi.org/10.1162/jocn_a_00758.

Zhao, T Christina, Kuhl, Patricia K., 2016. Musical intervention enhances infants' neural processing of temporal structure in music and speech. Proc. Natl. Acad. Sci. U. S. A. 113 (19), 5212–5217. http://dx.doi.org/10.1073/pnas.1603984113. National Academy of Sciences.

Ziegler, J.C., Pech-Georgel, C., George, F., Alario, F.-X., Lorenzi, C., 2005. Deficits in speech perception predict language learning impairment. Proc. Natl. Acad. Sci. 102 (39), 14110–14115. http://dx.doi.org/10.1073/pnas.0504446102. National Academy of Sciences.

Zion Golumbic, Elana, Cogan, Gregory B., Schroeder, Charles E., Poeppel, David, 2013. Visual input enhances selective speech envelope tracking in auditory Cortex at A 'cocktail party. J. Neurosci. 33 (4), 1417–1426. http://dx.doi.org/10.1523/JNEUROSCI.3675-12.2013.