Cutting Through the Noise: Noise-Induced Cochlear Synaptopathy and Individual Differences in Speech Understanding Among Listeners With Normal Audiograms

Mishaela DiNino, Lori L. Holt, and Barbara G. Shinn-Cunningham

Abstract: Following a conversation in a crowded restaurant or at a lively party poses immense perceptual challenges for some individuals with normal hearing thresholds. A number of studies have investigated whether noise-induced cochlear synaptopathy (CS; damage to the synapses between cochlear hair cells and the auditory nerve following noise exposure that does not permanently elevate hearing thresholds) contributes to this difficulty. A few studies have observed correlations between proxies of noise-induced CS and speech perception in difficult listening conditions, but many have found no evidence of a relationship. To understand these mixed results, we reviewed previous studies that have examined noise-induced CS and performance on speech perception tasks in adverse listening conditions in adults with normal or near-normal hearing thresholds. Our review suggests that superficially similar speech perception paradigms used in previous investigations actually placed very different demands on sensory, perceptual, and cognitive processing. Speech perception tests that use low signal-to-noise ratios and maximize the importance of fine sensory details— specifically by using test stimuli for which lexical, syntactic, and semantic cues do not contribute to performance—are more likely to show a relationship to estimated CS levels. Thus, the current controversy as to whether or not noise-induced CS contributes to individual differences in speech perception under challenging listening conditions may be due in part to the fact that many of the speech perception tasks used in past studies are relatively insensitive to CS-induced deficits.

Key Words: Cochlear synaptopathy, Hidden hearing loss, Obscure auditory dysfunction, Speech-in-noise, Speech perception.

Abbreviations: ABR = auditory brainstem response; AN = auditory nerve; AP = action potential; CS = cochlear synaptopathy; EFR = envelope following response; ITD = interaural timing difference; NHT = normal hearing threshold; MEMR = middle ear muscle reflex; SP = summating potential; SR = spontaneous rate.

INTRODUCTION

A number of animal models demonstrate cochlear synaptopathy (CS), a loss of the synapses between inner hair cells and the auditory nerve, following exposure to high-intensity noise, even if the damage does not result in a permanent increase in hearing thresholds (Kujawa & Liberman 2009; Furman, Kujawa, & Liberman 2013; Valero et al. 2017). Less clear is the extent to which noise-induced CS occurs in humans and, if it does, whether it precipitates any perceptually relevant deficits. A large number of carefully controlled studies in humans with normal hearing thresholds (NHTs) have failed to find relationships between performance on perceptual tasks and proxies of noise-induced CS, such as noise exposure history or auditory nerve (AN) integrity metrics. These negative results have called into question the link between CS and clinically relevant perceptual impairments, and even the very existence of noise-induced CS in the human population (e.g., Prendergast et al. 2017; Le Prell et al. 2018; Johannesen et al. 2019).

Yet, interest in noise-induced CS persists because evidence in animal models suggests that it may contribute to a particularly distressing auditory perceptual deficit: impaired speech-in-noise perception in adults with NHTs. This symptomology, which has been labeled auditory processing disorder (e.g., Palfery & Duff 2007), King-Kopetzky syndrome (Hinchcliffe 1992), or obscure auditory dysfunction (Saunders & Haggard 1989) has been linked to a broad range of deficits, from peripheral (Shaw et al. 1996; Zhao & Stephens 2000; Zhao & Stephens 2006; Badri et al. 2011) to central (Jerger et al. 1991; Saunders & Haggard 1992; Zhao & Stephens 2000). CS may be an additional candidate to explain this constellation of symptoms; synaptic loss reduces the number of available AN fibers (particularly fibers with relatively greater importance for encoding loud sound; Furman et al. 2013) and is thus thought to impair perception of speech in the presence of competing auditory signals much more than it affects speech perception in quiet (e.g., Lopez-Poveda 2014). Evidence for this theory is supported by animal research: animals with reduced AN response magnitudes but normal detection thresholds following noise exposure (consistent with noise-induced CS) show behavioral deficits in broadband noise (Lobarinas et al. 2017), and animals with cochlear neural degeneration exhibit poor detection of tones in noise, but not in quiet (Resnik & Polley 2021). Noise-induced CS thus may impair human speech perception in background noise, even when hearing thresholds are normal.

Animals with CS-related AN degeneration demonstrate decreased phase-locking of neural firing to auditory signals (Shaheen et al. 2015; Parthasarathy & Kujawa 2018), suggesting that perceptual consequences of CS arise due to reduced temporal precision of auditory information. Presenting sound with simultaneous, competing sounds introduces greater demands on temporal processing than presenting a single sound—such as speech in quiet. For instance, the main effects of a competing steady state or fluctuating noise is to degrade the representation of a target sound's auditory features, an effect often known as energetic masking (Durlach et al. 2003b). Specifically, noise renders portions of the target signal inaudible, but also reduces the prominence of amplitude modulations important for conveying sound content. Because it is not spectrotemporally sparse, even multitalker speech babble causes a fair amount of energetic masking; its effects are more similar to that of competing noise than to competing speech with the same total energy (Lu & Cooke 2008). In contrast, when listening to speech in the presence of another spectrotemporally sparse signal (such

---

Department of Psychology, Neuroscience Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

0196/0202/XXX/XXXX-00/0 • Ear & Hearing • Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

<zdoi; 10.1097/AUD.0000000000001147>

as one other speech stream), audibility of the target content is not an issue. Instead, the problem of understanding the target speech depends upon perceptually segregating the target and masker (Bregman 1990). Temporal sound cues are critical for source segregation. CS, then, is likely to have larger effects when speech is presented in competing speech rather than in noise or babble.

In animal models, CS also occurs with aging (e.g., Sergeyenko et al. 2013) and noise-induced CS accelerates age-related CS (Fernandez et al. 2015; Liberman & Kujawa 2017). Human age-related CS may at least partly explain why older adults, even those with NHTs, consistently demonstrate temporal processing deficits relative to young adults [see Anderson & Karawani (2020) for review] that correlate with speech perception difficulties in noisy listening environments (Gelfand et al. 1988; Fullgrabe et al. 2014; Babkoff & Fostick 2017). Whether CS is caused by noise exposure, the aging process, or both, its perceptual consequences should be similar: the common denominator is damage to the synapse. As temporal bone studies suggest that age-related CS does occur in humans (Makary et al. 2011; Viana et al. 2015; Wu et al. 2019), it is important to explore not only whether noise-induced CS exists in humans, but how it may exacerbate effects of age-related CS.

Here, we review results from 25 studies in adults with NHTs or near-NHTs that asked whether individual differences in the ability to understand speech are related to any proxy of noise-induced CS. To find articles for this review, we conducted four separate searches on PubMed using the keywords "normal hearing speech communication challenges," "cochlear synaptopathy," "cochlear synaptopathy speech perception," and "noise exposure speech perception." The investigations that we included in this review met the following criteria: the manuscript was peer-reviewed, the participants were human adults with NHTs or near-NHTs, the total participant group was expected to have a range of CS levels (based on variability in occupational or recreational noise exposure history, a self-report noise exposure history metric, and/or age), and the authors performed a statistical comparison between at least one proxy of noise-induced CS (noise exposure history or an electrophysiological metric) and speech perception performance in challenging listening conditions. We excluded studies in which adults with tinnitus were the participant group of interest, as tinnitus may influence speech perception abilities beyond the physiological effects of CS. Three studies of tinnitus were excluded from the review.

Together, the 25 reviewed studies included 47 separate experiments, each of which compared performance on one speech perception task to measures of one or more proxies of noise-induced CS. Table 1 summarizes methods and the factors that affect performance for each test of speech perception. Table 2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A935 includes additional methodological details and lists the proxy(ies) of CS used in each experiment. Across the studies, these proxies included noise exposure history metrics as well as electrophysiological measures of peripheral auditory function [the auditory brainstem response (ABR) wave I amplitude, ABR wave I/wave V ratio, ABR summating potential/action potential ratio, ABR Wave I growth function in response to increasing sound intensity, ABR wave latencies, the envelope following response (EFR), and the middle ear muscle reflex (MEMR)]. Table 1 and Table 2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A935 show that the reviewed studies differed greatly in their experimental parameters; we did not perform a formal meta-analysis as it would have been underpowered given the large number of covariates among the investigations.

Of the 47 experiments reviewed, 22 (46.8%) observed a significant relationship between speech perception performance and one or more proxies of noise-induced CS (highlighted in gray, Table 1). With less than half of the literature finding a significant relationship between speech understanding performance and estimated CS levels, one might justifiably question whether noise-induced CS affects human perception. The lack of direct assessments of CS in living humans complicates attempts to link synaptic damage to auditory perceptual impairments. Some proxies of CS may also be more sensitive than others—heterogenous methods to predict CS levels likely contribute to the inconsistent results of prior studies, as has been discussed in several recent reviews (Bharadwaj et al. 2019; Bramhall et al. 2019; Le Prell et al. 2019). Further, individual differences in synapse counts from genetic and/or developmental factors may be a source of variability that obscures correlations between noise exposure history and speech perception in challenging listening conditions. Still, as we describe below, the speech perception tasks used in the studies reviewed here placed very different demands on the listener. Some emphasized sensory processing (particularly temporal processing), whereas others used tasks for which perceptual and cognitive processes may obscure subtle perceptual deficits caused by CS.

STIMULUS AND TASK DIFFERENCES IMPACT WHETHER RELATIONSHIPS WITH CS ARE OBSERVED

Table 1 highlights how previous studies of the relationship between noise-induced CS and speech perception in challenging listening contexts have used various combinations of speech stimuli, noise types, presentation modes, and response sets (open- or closed-set) in the speech perception task. Every one of these variables on its own can influence the specific demands of the task. Given the variability of methods, it is not surprising that the human CS literature has yielded inconsistent results. Experiments with seemingly similar objectives engage very different sensory, perceptual, and cognitive processes, depending upon the kind of target speech they present, whether they present that speech in noise—and if so, what the "noise" characteristics are, and how they measure the joint interaction of speech and noise. Therefore, although each of the studies listed in Table 1 in quantifies speech understanding, the paradigms differ in substantive ways that may affect whether or not perceptual performance is observed to relate to measures of CS.

For instance, consider two hypothetical "speech-in-noise perception" experiments: one in which a participant listens diotically to a meaningful story masked by simultaneous steady-state noise (without any envelope modulation), and one in which the participant identifies an isolated, closed-set digit presented against a competing digit spoken by the same talker, but coming from a different location in space. Each task uses "speech" presented against a competing "noise." Yet, these tasks differ fundamentally in the demands placed on the system, the

<zdoi; 10.1097/AUD.0000000000001147>

TABLE 1. Summary of 25 previous studies, encompassing 47 separate experiments, investigating the relationship between human noise-induced cochlear synaptopathy and speech perception performance

Experiment No.	Study No.	Reference	Participant Ages	Task	Target Speech Stimuli	Competing Stimuli	Response Set	Presentation Mode	Contributing Factors
Studies using high context speech									
	Studies using unintelligible competing sound								
1	1	Grose et al. (2017)	18–35	Sentence identification (Modified BKB-SIN Test)	High-context sentences	Speech-shaped noise	Open	Monaural headphone	Context effects, lexical knowledge, energetic masking
2	2	Johannesen et al. (2019)	12–68	Sentence identification (HINT)	High-context sentences	Speech-shaped noise	Open	Monaural headphone	Context effects, lexical knowledge, energetic masking
3	2	Johannesen et al. (2019)	12–68	Sentence identification (HINT)	High-context sentences	Speech-like fluctuating signal (International Female Fluctuating Masker, IFFM; Holube et al. 2011)	Open	Monaural headphone	Context effects, lexical knowledge, segregation/selection
	Studies using intelligible competing speech								
4	3	Valderrama et al. (2018)	18–55	Sentence identification (LiSN-S Test)	High-context sentences (0°)	Two streams of ongoing stories from different talkers (+90° and −90°)	Open	Binaural headphone, HRTF-separated speech and noise	Context effects, lexical knowledge, non-spatial and spatial segregation/selection
5	4	Yeend et al. (2017)	30–60	Sentence identification (LiSN-S Test)	High-context sentences (0°)	Two streams of ongoing stories from different talkers (+90° and −90°)	Open	Binaural headphone, HRTF-separated target and noise	Context effects, lexical knowledge, non-spatial and spatial segregation/selection
6	4	Yeend et al. (2017)	30–60	Speech comprehension (NAL Dynamic Conversations Test)	Four min, high-context speech monologues (0°)	Conversational noise (distinct talkers, taking turns, at various locations)	Open	Multi-speaker soundfield simulation in anechoic chamber	Context effects, lexical knowledge, non-spatial and spatial segregation/selection
Studies using low-context sentences									
7	5	Bramhall et al. (2015a)	19–90	Sentence identification (QuickSIN)	Low-context sentences	Multitalker babble	Open	Monaural headphone	Context effects, lexical knowledge, energetic masking, non-spatial segregation/selection
8	6	Skoe et al. (2019)	18–24	Sentence identification (QuickSIN)	Low-context sentences	Four-talker babble	Open	Diotic headphone	Context effects, energetic masking, non-spatial segregation/selection
9	7	Smith et al. (2019)	18–30	Sentence identification (QuickSIN)	Low-context sentences	Four-talker babble	Open	Diotic headphone	Context effects, energetic masking, non-spatial segregation/selection
10	8	Grant et al. (2020)	18–63	Sentence identification (Modified QuickSIN)	Low-context sentences	Four-talker babble	Open	Monaural headphone	Context effects, lexical knowledge, energetic masking, non-spatial segregation/selection

(*Continued*)

TABLE 1. Continued.

Experiment No.	Study No.	Reference	Participant Ages	Task	Target Speech Stimuli	Competing Stimuli	Response Set	Presentation Mode	Contributing Factors
11	9	Mepani et al. (2020)	18–63	Sentence identification (Modified QuickSIN)	Low-context sentences	Four-talker babble	Open	Monaural headphone	Context effects, lexical knowledge, energetic masking, non-spatial segregation/selection
12	10	Mepani et al. (2021)	18–63	Sentence identification (Modified QuickSIN)	Low-context sentences	Four-talker babble	Open	Monaural headphone	Context effects, lexical knowledge, energetic masking, non-spatial segregation/selection
Studies using speech without semantic or syntactic context									
	Studies with no competing sound								
13	11	Liberman et al. (2016)	18–41	Word identification	NU-6 words, 45% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
14	11	Liberman et al. (2016)	18–41	Word identification	NU-6 words, 65% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
15	9	Mepani et al. (2020)	18–63	Word identification	NU-6 words, 45% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
16	9	Mepani et al. (2020)	18–63	Word identification	NU-6 words, 65% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
17	10	Mepani et al. (2021)	18–63	Word identification	NU-6 words, 45% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
18	10	Mepani et al. (2021)	18–63	Word identification	NU-6 words, 65% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
19	8	Grant et al. (2020)	18–63	Word identification	NU-6 words, 45% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
20	8	Grant et al. (2020)	18–63	Word identification	NU-6 words, 65% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
21	12	Kamerer et al. (2019)	20–86	Word identification	NU-6 words, 45% time compression	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
22	12	Kamerer et al. (2019)	20–86	Word identification	NU-6 words, 45% time compression, 0.3 s reverb	None	Open	Monaural headphone	Lexical knowledge, speech feature coding
	Studies using unintelligible competing sound								
23	11	Liberman et al. (2016)	18–41	Word identification	NU-6 words	White noise at 0 dB SNR	Open	Monaural headphone	Lexical knowledge, energetic masking

(Continued)

TABLE 1. Continued.

Experiment No.	Study No.	Reference	Participant Ages	Task	Target Speech Stimuli	Competing Stimuli	Response Set	Presentation Mode	Contributing Factors
24	11	Liberman et al. (2016)	18–41	Word identification	NU-6 words	White noise at +5 dB SNR	Open	Monaural headphone	Lexical knowledge, energetic masking
25	9	Mepani et al. (2020)	18-63	Word identification	NU-6 words	Speech-shaped noise	Open	Monaural headphone	Lexical knowledge, energetic masking
26	10	Mepani et al. (2021)	18–63	Word identification	NU-6 words	Speech-shaped noise	Open	Monaural headphone	Lexical knowledge, energetic masking
27	13	Shehorn et al. (2020)	21–54	Word identification (Modified MD CNC Test)	Words with reverb	Speech-shaped noise	Open	Diotic headphone	Energetic masking
28	8	Grant et al. (2020)	18–63	Word identification	NU-6 words	Speech-shaped noise	Open	Monaural headphone	Lexical knowledge, energetic masking
29	12	Kamerer et al. (2019)	20–86	Word identification	NU-6 words	Noise (type not reported)	Open	Monaural headphone	Lexical knowledge, energetic masking
30	2	Johannesen et al. (2019)	12–68	Word identification	Disyllabic words	Speech-shaped noise	Open	Monaural headphone	Lexical knowledge, energetic masking
31	2	Johannesen et al. (2019)	12–68	Word identification	Disyllabic words	Speech-like fluctuating signal (IFFM)	Open	Monaural headphone	Lexical knowledge, non-spatial segregation/selection
32	13	Fulbright et al. (2017)	18–30	Word identification (The WIN Test)	NU-6 words	Multitalker babble	Open	Monaural headphone	Lexical knowledge, energetic masking, non-spatial segregation/selection
33	13	Fulbright et al. (2017)	18–30	Word identification (The Words in Broadband Noise Test)	NU-6 words	Broadband noise	Open	Monaural headphone	Lexical knowledge, energetic masking
34	14	Grinn et al. (2017)	21–27	Word identification (The WIN Test)	NU-6 words	Multitalker babble	Open	Monaural headphone	Lexical knowledge, energetic masking non-spatial segregation/selection
35	15	Le Prell et al. (2018)	18–27	Word identification (The WIN Test)	Words	Multitalker babble	Open	Monaural headphone	Lexical knowledge, energetic masking, non-spatial segregation/selection
36	16	Hope et al. (2013)	24–39	Syllable identification	VCV syllables	ICRA noise	Closed	Diotic headphone	Energetic masking
37	17	Prendergast et al. (2017)	18–36	Digit stream identification (Digit Triplet Test)	Digit streams	Speech-shaped noise	Closed	Diotic headphone	Energetic masking

(*Continued*)

TABLE 1. Continued.

Experiment No.	Study No.	Reference	Participant Ages	Task	Target Speech Stimuli	Competing Stimuli	Response Set	Presentation Mode	Contributing Factors
38	18	Prendergast et al. (2019)	18–60	Digit identification (Digit Triplet Test)	Digit streams	Speech-shaped noise	Closed	Diotic headphone	Energetic masking
39	19	Carcagno & Plack (2021)	18–73	Digit identification (Digit Triplet Test)	Digit streams	Speech-shaped noise	Closed	Diotic headphone	Energetic masking
	Studies using intelligible competing speech								
40	20	Ruggles et al. (2011)	18–55	Digit stream identification	Monotonized digit streams (0°), varying reverb	Two digit streams identical to target, from −15° and +15°	Closed	Binaural headphone, HRTF-separated speech and noise	Spatial segregation/selection
41	21	Bharadwaj et al. (2015)	21–39	Digit stream identification	Monotonized digit streams (ITD 50–400 μs)	Digit stream identical to target, but with ITD of opposite sign (symmetric)	Closed	Binaural headphone, ITD-separated speech and noise	Spatial segregation/selection
42	18	Prendergast et al. (2019)	18–60	Keyword identification (CRM)	Carrier phrases with call sign, color, and number keywords	Two streams identical to target, but with different keywords and talkers	Closed	Diotic headphone	Non-spatial segregation/selection
43	17	Prendergast et al. (2017)	18–36	Keyword identification (CRM)	Carrier phrases with call sign, color, and number keywords (0°)	Two streams identical to target, but with different keywords, talkers, and locations	Closed	Binaural headphone, HRTF-separated speech and noise	Non-spatial and spatial segregation/selection
44	22	Guest et al. (2018)	18–40	Keyword identification (CRM)	Carrier phrases with call sign, color, and number keywords (0°)	Two streams identical to target, but with different keywords, talkers, and locations (−60° and +60°)	Closed	Binaural headphone, HRTF-separated speech and noise	Non-spatial and spatial segregation/selection
45	23	Couth et al. (2020)	18–27	Keyword identification (CRM)	Carrier phrases with call sign, color, and number keywords (0°)	Two streams identical to target, but with different keywords, talkers, and locations (−60° and +60°)	Closed	Binaural headphone, HRTF-separated speech and noise	Non-spatial and spatial segregation/selection
46	24	Carcagno & Plack (2021)	18–73	Keyword identification (CRM)	Carrier phrases with call sign, color, and number keywords (0°)	Two streams identical to target, but with different keywords, talkers, and locations (−65° and +65°)	Closed	Binaural headphone, HRTF-separated speech and noise	Non-spatial and spatial segregation/selection
47	25	Parthasarathy et al. (2020)	Young to middle-aged (mean age 28.3±0.9)	Digit identification	Digit streams	Two streams identical to target, but with different talkers	Closed	Binaural headphone	Non-spatial and spatial segregation/selection

Studies are grouped first by speech materials (high-context sentences, low-context sentences, open-set words, closed-set words, or syllables) and then, within these, by masker type (none, unintelligible maskers like noise and babble, intelligible speech). The summary shows the total age range of participants in each study (from both control and experimental groups, if applicable), the name of speech perception tasks used (if applicable), the target speech stimuli and noise type employed by each task, the nature of the speech stimuli (e.g., open or closed set), the presentation mode, and the factors expected to contribute to performance on each task. Most descriptions of task parameters in the table are exactly as worded by the study authors. Of the 47 experiments summarized, 22 (46.8%) found a positive relationship between speech perception performance and proxies of noise-induced CS (highlighted in gray). Results from the two experiments in light gray were influenced by the effects of traditional hearing loss, whereas those in dark gray were not. Note that some studies reported fewer task details than others; some details were obtained by contacting the authors.

BKB-SIN, Bamford-Kowal-Bench Speech-in-Noise; CRM, coordinate response measure; dB, decibel; HINT, hearing in noise test; HRTF, head-related transfer function; ICRA: International Collegium for Rehabilitative Audiology; ITD, interaural time difference; LISN-S, Listening in Spatialized Noise-Sentences; MD CNC, Maryland consonant-nucleus-consonant words; NAL, National Acoustic Laboratories; NU-6, Northwestern University Auditory Test No. 6; QuickSIN, quick speech-in-noise test; SNR, signal-to-noise ratio; WIN, Words in Noise.

information a listener can use to understand the target speech, and the response used to measure speech comprehension. Given this, behavioral performance will depend differently on various auditory pathologies—including CS.

By considering the processes that impact perception of speech and differences in experimental procedures across studies, our review of the literature identifies some factors that may help explain disparate findings across these studies. The following sections describe specific issues related to different speech perception tasks that we believe complicate interpretation of the larger literature on the impact of noise-induced CS on speech perception under adverse listening contexts:

1. CS does not affect auditory detection thresholds and thus produces much more subtle deficits than does traditional hearing loss. Clinically validated speech perception tasks used to quantify traditional hearing impairments may thus be too insensitive to quantify deficits due to CS.
2. Speech perception paradigms with high ecological validity involve cognitive processes that may obscure any relationship between CS and task performance.
3. Speech perception tasks used in past CS studies vary in the degree to which they emphasize perception of temporal features, which are particularly susceptible to CS-induced deficits (Shaheen et al. 2015; Parthasarathy & Kujawa 2018).

In total, our review reveals characteristics of speech perception tasks that are likely to reflect perceptual deficits caused by CS. Future studies directed at determining whether CS accounts for difficulties processing speech in challenging listening contexts may benefit from considering these issues.

CS LIKELY CAUSES DEFICITS TOO SUBTLE TO INFLUENCE SCORES ON MANY CLINICAL SPEECH PERCEPTION TESTS

The sensory deficits that CS may cause are likely to be rather subtle compared to those of "traditional" hearing loss (i.e., spectral loss that affects auditory detection thresholds, and thus speech perception performance in quiet). CS predominately targets synapses on low- and medium-spontaneous rate (SR) AN fibers (see Fig. 1; Furman et al. 2013; Liberman et al. 2015), which make a relatively larger contribution to auditory signal encoding as sound levels increase. A reduced low-SR fiber population response (as demonstrated in animal models of CS) likely affects the encoding fidelity of high-intensity auditory stimuli, including sounds in loud, noisy environments. Low-SR fibers also play a role in the auditory efferent pathway. Evidence from studies in animals (Kawase et al. 1993; Pang & Guinan 1997), in humans (e.g., Giraud et al. 1997; Kumar & Vanaja 2004), and from computational modeling (Brown et al. 2010) suggests that low-SR fibers enhance sound perception in the presence of competing auditory signals by adapting AN responses to ongoing noise. CS may thus reduce the effectiveness of the auditory efferent pathway, which is likely to be especially detrimental to understanding sound sources when levels are relatively high.

While CS preferentially affects low-SR fibers, loss of any type of AN fiber will have consequences on sound coding by distorting temporal representations. Decreased neural synchrony from a reduced AN population response reduces the faithfulness of auditory signal encoding, causing perceptual difficulties in listening situations with competing sounds (Lopez-Poveda 2014). Accordingly, a model of AN under-sampling (such as would occur with CS-related AN deafferentation) predicts poor sentence identification performance in noise, but not in quiet (Lopez-Poveda & Barrios 2013). Thus, instead of affecting whether a listener can *detect* a sound (like traditional hearing loss), CS-related AN degeneration likely alters the fidelity of the coding of a sound's content (Lopez-Poveda and Barrios 2013; Lopez-Poveda 2014; Plack et al. 2014; Carney 2018).

Most clinical tests of speech perception were designed to distinguish listeners with healthy cochlear function from those with traditional hearing loss. Specifically, they have been optimized to quantify damage to the cochlear amplifier, which results in inaudibility and poor frequency selectivity. Given the sensory differences between traditional hearing loss and CS, it should not be surprising that these tests are not well-suited to assessing CS-induced deficits in adults with NHTs. Such tasks utilize SNRs that may be challenging enough to quantify individual abilities for listeners with traditional hearing loss, but too easy for listeners with NHTs, even those with a sensory coding deficit from CS. For example, the Words in Broadband Noise Test [used by Fulbright et al. (2017)] and the Words in Noise Test [Wilson & Burks 2005; used by Fulbright et al. (2017);

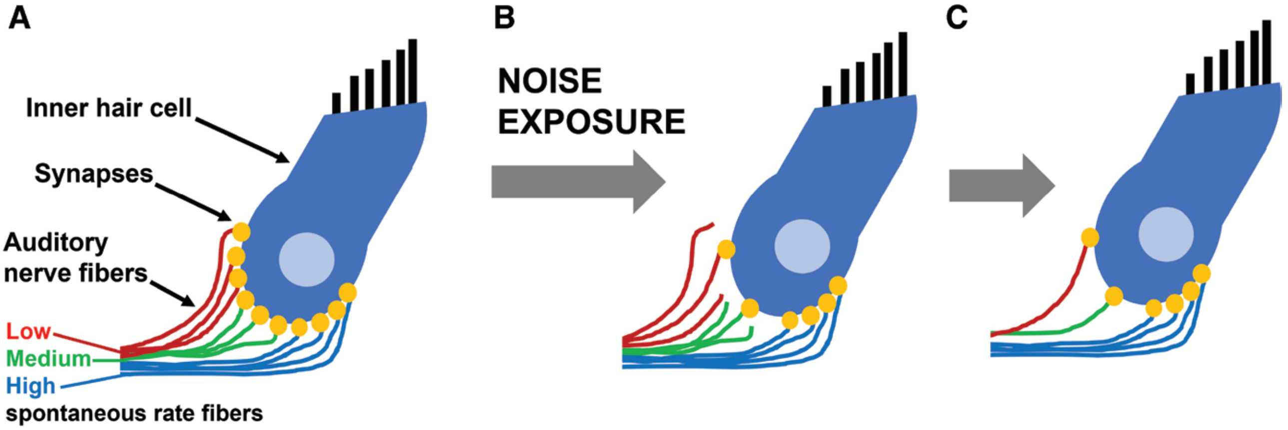

Fig. 1. Illustration of auditory nerve fiber degeneration following noise exposure. A, Prior to noise exposure, synapses between the pictured inner hair cell and the auditory nerve are intact, as are auditory nerve fibers. B, Noise exposure results in synaptic damage. C, Auditory nerve fibers degenerate following synaptic loss. Note that the low- and medium-spontaneous rate fibers, located on the modiolar side of the inner hair cell, are particularly affected.

Grinn et al. (2017); and Le Prell et al. (2018)] use SNRs ranging from +30 to +20 dB SNR and +24 dB to 0 dB SNR, respectively. In these prior studies, scores on the Words in Broadband Noise Test were not reported, but most participants with NHTs performed at ceiling on the Words in Noise Test until the SNR decreased to +8 dB SNR, leaving only 15 words (five from the three most difficult SNRs) on which participants' identification scores varied.

Similarly, the clinical version of the QuickSIN test [Killion et al. 2004; used by Bramhall et al. (2015); Skoe et al. (2019); and Smith et al. (2019)] presents sentences in noise ranging from +25 dB to 0 dB SNR. This assessment is scored clinically as "SNR loss," the total number of keywords correctly repeated (out of 30) subtracted from 25.5. A recent study of young adults confirms that most individuals with NHTs have little trouble identifying key words even at the most difficult SNR levels: participants' SNR loss fell into the limited range of −1.25 to 2.25 (with lower SNR loss representing better performance) out of the possible range of −4.50 to 25.50 (Skoe et al. 2019). Listeners with NHTs perform very well, and very similarly, on clinical speech tests that use SNRs designed to be challenging for listeners with traditional hearing loss. It is thus unsurprising that the small variation in task performance observed across listeners with NHTs does not correlate with estimates of CS severity.

Indeed, seven of the 47 experiments we reviewed (Table 1) used one of these clinical tests, but only one found any relationship to proxies of CS (Bramhall et al. 2015). The one observing a relationship included participants with traditional hearing loss—which makes it difficult to attribute any observed relationship to CS, rather than damage to the cochlear amplifier. Thus, although existing clinical speech tests and speech corpora are useful for assessing how overt hearing loss affects speech perception, those that use high SNRs are likely to be insensitive to the more subtle differences in speech perception abilities that CS may cause. As shown in Figure 2A, after excluding the study that was influenced by the effects of traditional hearing loss, no experiments using such clinical speech-in-noise perception tests demonstrated a relationship between a proxy of noise-induced CS and speech perception scores.

HIGH-CONTEXT SPEECH MATERIALS ENGAGE NONSENSORY FACTORS

Speech perception in everyday listening situations involves a host of cognitive processes, some of which may obscure observation of any potential relationship between impaired speech perception and subtle degradations in the peripheral coding of sound, such as those that CS would cause. For instance, speech perception can be guided by syntactic and semantic context that provide top-down constraints that "fill in" phonemes, syllables, or even whole words that are otherwise degraded in the input (e.g., Samuel 1981). Thus, tasks presenting sentences or narratives [e.g., the Dynamic Conversations Test; Best et al. 2016; used by Yeend et al. (2017)] provide linguistic context that individuals can leverage to fill in words they did not hear clearly [although context can hinder speech identification at very low SNRs; see Marrufo-Pérez et al. (2019)].

In addition, speech perception tasks that simulate many challenges of real-world speech perception may reveal individual differences unrelated to CS, potentially confounding discovery of a relationship between CS and speech perception under difficult listening conditions. For example, comprehension of sentences or passages requires participants to hold speech in memory before responding and captures individual differences in working memory. Thus, when a task uses meaningful sentences or stories, listeners may lean on top-down perceptual restoration to compensate for subtle sensory deficits, and/or individual differences in working memory or other cognitive processes engaged by speech perception (but unrelated to the sensory deficits of CS). Such tasks thus may conceal possibly subtle influences of CS on speech perception performance.

Six of the 47 experiments reviewed in Table 1 presented meaningful sentences embedded in different kinds of competing sound (e.g., HINT sentences, Listening in Spatialized Noise Sentences, Bamford-Kowal-Bench sentences, and sentences from the Dynamic Conversations Test). While one of these experiments reported a relationship between speech perception and estimated CS levels, that study did not rule out differences in individuals' hearing thresholds (Valderrama et al. 2018). None of the other five experiments using meaningful sentence materials found a relationship to proxies of CS [see Fig. 2B; Grose et al. 2017; two experiments in Yeend et al. (2017); two experiments in Johannesen et al. (2019)].

Some tests employ low-predictability sentences for which context provides little or no information about target words. Still, individual differences in vocabulary and access to linguistic knowledge can affect performance on even simple tasks using low-context sentences under adverse listening conditions [e.g., Banks et al. (2015); Kaandorp et al. (2016); Carroll et al. (2016)]. These confounds are a source of individual variation unrelated to sensory deficits, again reducing sensitivity to effects of CS.

Of the 47 experiments we reviewed, six presented low-context sentences. Four reported a relationship to CS proxies; however, one experiment did not rule out effects due to elevated hearing thresholds (Bramhall et al. 2015). Importantly, the other three experiments reporting a positive relationship utilized a modified, difficult version of the QuickSIN, on which participants' performance varied greatly [one experiment in Grant et al. (2020); one in Mepani et al. (2020); one in Mepani et al. (2021)]. The remaining two experiments using low-context sentence stimuli reported no relationship to CS [see Fig. 2B; Skoe et al. (2019); Smith et al. (2019)].

The other 35 prior experiments listed in Table 1 used either open-set, isolated word recognition tests, or closed-set speech identification tasks. These tasks place modest demands on working memory and remove the semantic and syntactic information that could help listeners compensate for subtle sensory deficits. Importantly, as described below and as shown in Figure 2B, the great majority of the experiments that did find significant relationships between speech perception and CS proxies used such tasks.

Twenty-three of the 47 experiments we reviewed used open-set word identification tests, in which the target word can be any possible word; participants are not limited to selecting a response from a closed-set. Although 9 of the 23 experiments reported null results [two experiments in Fulbright et al. (2017); two in Johannesen et al. (2019); three in Kamerer et al. (2019); Grinn et al. (2017); Le Prell et al. (2018)], 14 experiments did find a relationship to estimated CS levels [four experiments in Liberman et al. (2016); three in Grant et al. (2020); three in Mepani et al. (2020); three in Mepani et al. (2021); Shehorn et al. (2020)].

Relative to open-set tasks, closed-set speech identification tests provide participants with a small number of response alternatives and thus further limit the effects of individual differences in lexical knowledge and lexical access on test performance. For instance, the Digit Triplet Test [used by Prendergast et al. (2017); Prendergast et al. (2019); Carcagno & Plack (2021)] requires participants to identify three digits between one and nine presented in noise. In the Coordinate Response Measure [used by Guest et al. (2018); Prendergast et al. (2017); Prendergast et al. (2019); Carcagno & Plack (2021)], participants listen to competing streams of the form "Ready <call sign> go to <color> <number>" and are asked to report back the color (out of four options) and number (between one and four, in the investigations described in this review) of the stream that contains a target call sign, such as "Baron." Because of the structure of these stimuli and limited response options, all of these studies reduce reliance on cognitive factors that influence speech intelligibility in daily life. Such tests are clearly less natural than tests using sentences, or even open-set isolated word recognition tests, but are more likely to be sensitive to the impact of a subtle sensory deficit on speech intelligibility.

Twelve of the experiments we reviewed used such closed-set speech identification tasks. Three found that performance on the speech task in challenging listening conditions was related to proxies of CS (Ruggles et al. 2011; Hope et al. 2013; Bharadwaj et al. 2015), but the nine other experiments found no such relationship [two experiments in Prendergast et al. (2017); two in Prendergast et al. (2019); two in Carcagno & Plack (2021); Guest et al. (2018); Couth et al. (2020); and Parthasarathy et al. (2020); although see section "Speech Perception Tasks Vary in Emphasis on Temporal Acoustic Features" for potential explanations for negative findings in these specific experiments].

While each of these experiments compared perception to different CS proxies that may themselves have influenced the results (see Table 2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A935), overall, this analysis suggests that studies are more likely to reveal a relationship between estimated CS levels and speech understanding if they use speech materials and tasks that minimize context effects and other nonsensory factors (see Fig. 2B). This is a tradeoff: closed-set tasks do not have the ecological validity of more natural speech tasks, but cognitive factors may need to be minimized in order to observe the putative relationships between a subtle sensory deficit and speech perception. Ecological validity must be put aside, at least for the moment, in favor of accumulating a body of evidence regarding whether processes impacted by CS, in fact, influence human speech perception.

SPEECH PERCEPTION TASKS VARY IN EMPHASIS ON TEMPORAL ACOUSTIC FEATURES

As mentioned throughout this review, CS-related AN degeneration degrades encoding of auditory signal timing in animal models (Shaheen et al. 2015; Parthasarathy & Kujawa 2018). Thus, speech perception tasks that require a listener to rely on fine temporal features might be expected to correlate with measures of CS. Yet, as shown in Figure 2C, even the 31 experiments that used both (1) speech perception tasks with appropriate SNR levels for listeners with NHTs and (2) stimuli that limited nonsensory factors still varied in the methods they utilized to emphasize temporal processing.

Ten of the studies that we reviewed presented isolated words in which temporal features were degraded, thus stressing sensory coding (particularly of temporal representations) more than typical speech. Specifically, to degrade sensory features, these studies time-compressed the words, then added simulated reverberation. Of these ten studies, eight found a relationship to CS [see Fig. 2C; two experiments in Liberman et al. (2016); two in Grant et al. (2020); two in Mepani et al. (2020); two in Mepani et al. (2021)] and two did not [two experiments in Kamerer et al. (2019)].

The primary effects of steady-state noise, fluctuating noise, and speech babble on intelligibility of speech are to mask portions of the target speech (energetic masking), and to reduce the depth of target speech amplitude modulation, a temporal feature varying on a time scale of tens of milliseconds. This reduction in modulation can impair speech understanding; however, the modulations conveying speech information may still be slow enough that CS does not impair perception. However, the similarity of target speech and any competing sound also influences the factors that limit speech intelligibility (Durlach et al. 2003a). If target speech is presented simultaneously with other intelligible speech, the temporal precision of the auditory representation must be good enough to support both segregation of the speech from the noise and selection of the target speech from the mixture [e.g., Shinn-Cunningham & Best (2008)]. Only then can a listener successfully deploy selective attention to the target and analyze its acoustic content.

Importantly, the acoustic features that are important for source segregation and selection require temporal precision orders of magnitude more precise than those supporting speech perception in quiet or even in the presence of dissimilar noise. For instance, differences of even a few semitones in the fundamental frequencies of competing talkers are sufficient to support segregation and selection (Binns and Culling 2007; Madsen et al. 2019). Use of fundamental frequency differences between talkers requires temporal coding precision on the order of a few milliseconds. Thus, even if CS preserves timing well enough to understand shallow speech modulation in steady noise, it may impair perception of pitch cues important for segregating target speech from distracting speech.

Source location provides another important cue to support segregation and selection when listeners must focus on target speech and ignore a competing, similar sound (Hawley et al. 2004; Kidd et al. 2005). Coding of interaural timing difference, the dominant perceptual cue for sound source location (Wightman & Kistler 1992), requires even greater temporal precision than does pitch coding, on a scale of tens to hundreds of *microseconds*. Thus, tasks that require reliance on spatial cues for segregating speech streams may be especially sensitive to CS, even more than those relying on pitch cues.

It is worth noting that listeners may not rely on spatial cues to segregate target speech in *every* paradigm for which competing speech sources are spatially separated (as in the Coordinate Response Measure and the Listening in Spatialized Noise—Sentences test; see Table 1). For instance, fundamental frequency differences alone can provide sufficient differentiation of target and masker to support selective attention, rendering spatial cues irrelevant (Brungart 2001). Also, as noted above, pitch cues are likely more robust than spatial cues, so when both pitch and location cues differentiate target and masking speech, spatial cues may be irrelevant. Further, even if two otherwise identical speech streams are presented from different directions, forcing a listener to rely on spatial cues, the task may not be sensitive to subtle differences in temporal coding precision. Figure 3 illustrates this point. If competing streams are presented with a large spatial separation (for instance, as in some past studies; ~60°: Prendergast et al. 2017; Guest et al. 2018; Couth et al. 2020; Carcagno & Plack 2021; 90°: Yeend et al. 2017; Valderrama et al. 2018), even a listener with poor temporal coding nonetheless may be able to resolve the streams based on

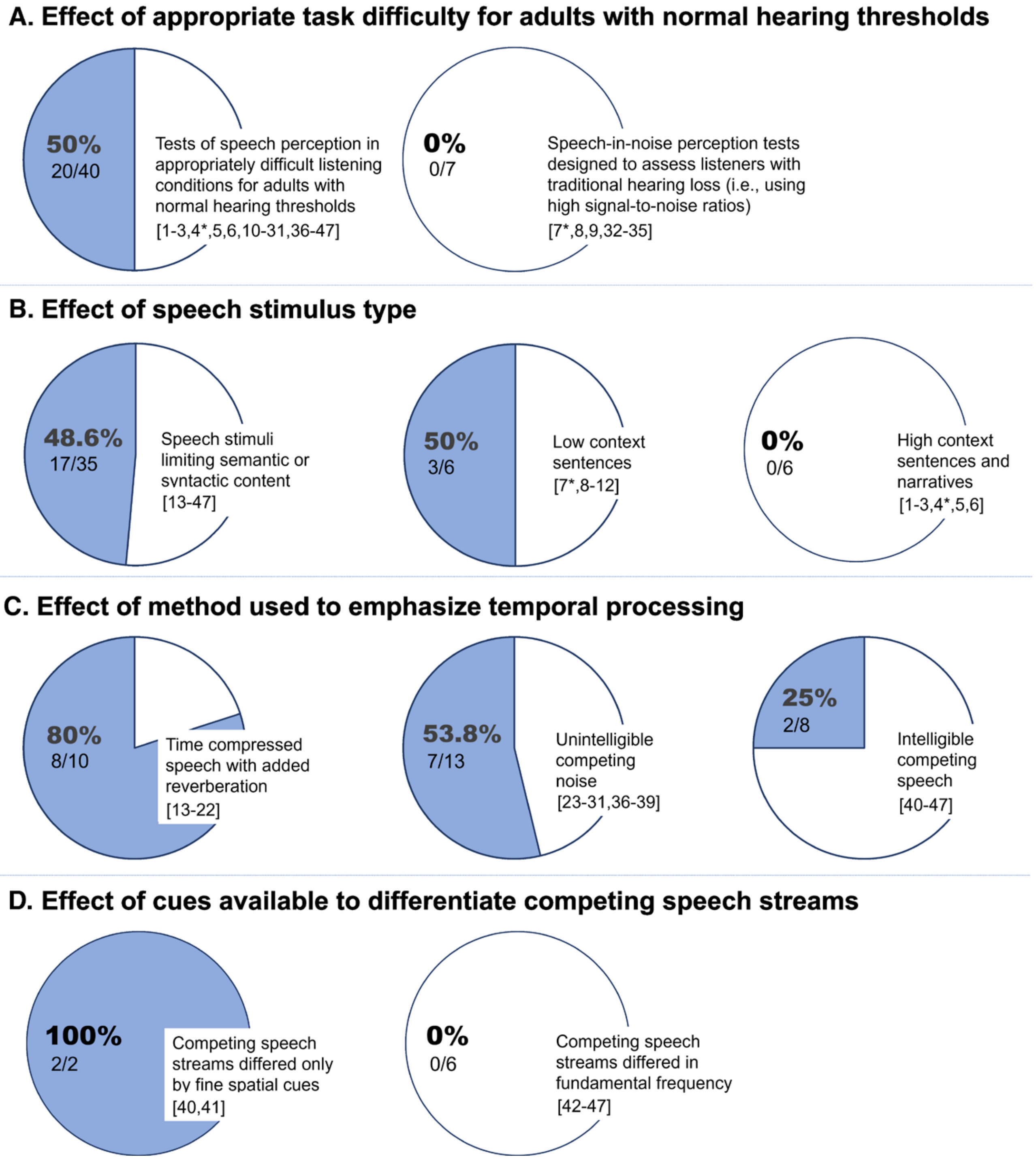

Fig. 2. Pie charts illustrating the percent of studies utilizing particular task parameters that found a relationship between a proxy of noise-induced synaptopathy and speech perception in challenging listening conditions. The experiments that contributed to data in each pie chart are listed in brackets under each—experiment numbers refer to those from Table 1. Experiments with asterisks indicate one that yielded a positive result, but was influenced by the effects of traditional hearing loss. Experiments are classified by: (A) suitability for assessing listeners with normal hearing thresholds, (B) speech stimulus, (C) method used to emphasize temporal processing, and (D) cues available for differentiating between speech streams. Note that (C) and (D) exclude experiments that used relatively high SNRs (those in panel A, left) or sentence stimuli/narratives (those in panel B, middle and right).

spatial cues. Only if the sources are close enough that listeners with "good" resolution must focus to perform the task are listeners with subtle sensory deficit like CS likely to show impaired performance.

Of the 31 experiments reviewed that used speech perception tests with appropriate SNRs for listeners with NHTs and speech stimuli that limited nonsensory contributions, eight asked listeners to report target speech played with competing, intelligible speech streams and thus emphasized acoustic cues supporting segregation and selection. Six of these found no relationship between speech intelligibility and CS proxies [one experiment in Prendergast et al. (2017); one in Prendergast et al. (2019); one in Carcagno & Plack (2021); Guest et al. (2018); Couth et al. (2020); Parthasarathy et al. (2020)]. Only two of the experiments reported a positive result (see Fig. 2C; Ruggles et al. 2011; Bharadwaj et al. 2015).

Importantly, spatial cues were critical for those two, and only those two, experiments. While some of the studies presenting target speech with competing, intelligible speech played the competing streams from different directions, the talkers also differed across streams, allowing a listener to rely on fundamental frequency cues and rendering spatial cues unnecessary (Prendergast et al. 2017, 2019; Guest et al. 2018; Couth et al. 2020; Parthasarathy et al. 2020; Carcagno & Plack 2021). In the two experiments that found a relationship, the target speech and the competing speech were from the same talker and differed only because of a small spatial separation, stressing the ability of listeners to utilize fine spatial cues to direct attention (see Fig. 3). Thus, as illustrated in Figure 2D, the influence of CS on speech perception in experiments that require listeners to segregate and select target speech may be most pronounced when the task relies upon precise spatial selective attention, which is a critical contributor to understanding speech in noisy listening environments, and which places extreme demands on temporal coding.

SUMMARY AND IMPLICATIONS

In this review, we have described several factors that may help explain the mixed results among previous studies of the relationship between noise-induced CS and speech perception in difficult listening situations. Of the 47 experiments we reviewed, 22 reported a relationship between proxies of noise-induced CS and speech perception performance (see grayed entries in Table 1). Of these, two (light gray fill in Table 1) did not rule out confounds due to traditional hearing loss (Bramhall et al. 2015; Valderrama et al. 2018). Importantly, each of the remaining 20 experiments that reported a relationship of estimated CS levels to speech perception performance, summarized below, employed speech tasks able to tap into the subtle temporal sensory deficits most associated with CS while also minimizing the higher-order perceptual and cognitive processes that can be drawn into play in speech perception.

- Two experiments observed significant relationships between EFRs and performance on a speech-against-speech task requiring fine spatial attention (Ruggles et al. 2011, Bharadwaj et al. 2015).
- One experiment found that closed-set syllable identification correlated with occupational noise-exposure history (Hope et al. 2013).
- Fourteen experiments found that isolated word recognition correlated with physiological CS proxies: twelve open-set word identification experiments presenting either words in steady-state noise or words that were sped up with reverberation added found correlations with ABR measures [four in Liberman et al. (2016); three in Grant et al. (2020)], with ABR and MEMR metrics [three in Mepani et al. (2020)], or with EFRs (three in Mepani et al. (2021)], and one using closed-set identification of words in noise with reverberation found a correlation with MEMR thresholds (Shehorn et al. 2020). Liberman et al. (2016) additionally observed significant relationships between word identification scores and noise exposure history.
- Three experiments found that performance on a modified, difficult version of the QuickSIN correlated with ABR measures [one experiment in Grant et al. (2020); one in Mepani et al. (2020)] or EFRs [one in Mepani et al. (2021)].

It is crucial to note that the last experiments, while finding a relation between one or more proxies of CS and performance on a sentence recognition task, used low-context sentences presented at SNRs that are challenging for adults with NHTs. Further, in those three investigations, the correlations between estimated CS levels and speech perception performance were weaker for sentence identification than they were for performance on the word recognition tasks—providing additional evidence that speech perception tasks with limited cognitive contributions may be most sensitive to CS.

The three studies reporting significant relationships between sentence identification performance and CS proxies also

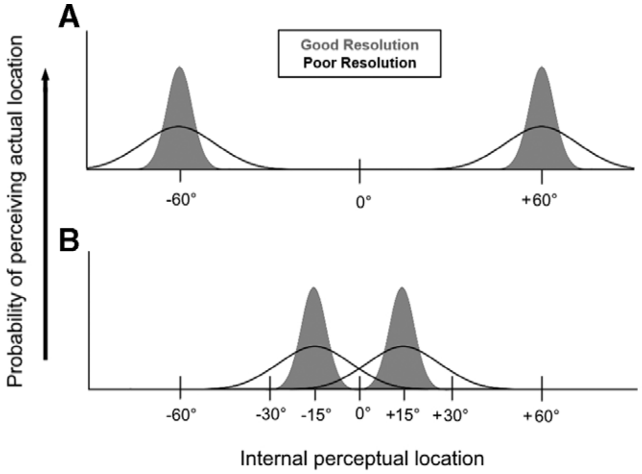

Fig. 3. Cartoon depicting the importance of using small spatial separations between speech and noise to reveal subtle temporal coding deficits. Each panel shows probability density functions representing the perceived spatial locations of two competing sources symmetrically positioned the left and right, either with a large spatial separation (A) or a spacing that is just resolvable for a listener with good temporal resolution (B). A, For large spatial separations, listeners with good temporal coding (gray, narrow distributions) and poor temporal encoding as might arise with CS (black line, board distributions), would both be able to resolve the spatial locations to perform a spatial selective listening task. Many spatial listening tasks fall into this category. B, For a small spatial separation, listeners with good temporal resolution are more likely to perform well relative to listeners with poorer temporal encoding. This design may thus be more sensitive to CS-related perceptual deficits.

employed stimuli to elicit electrophysiological responses that may be better suited to identify CS-related AN degeneration than those utilized by many other human studies [see Table 2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A935; Bharadwaj et al. 2019; Vasilkov et al. 2021; see discussions in Grant et al. (2020); Mepani et al. (2020); and Mepani et al. (2021) for further details]. These findings thus highlight the importance of developing *both* electrophysiological proxies of CS and speech perception tasks in difficult listening conditions that are sensitive to CS. Although we did not categorize studies according to the CS proxy that each used, differences in CS proxies or other experimental methods may have also influenced the results we report here. In addition, the varying sensitivity of previously used CS proxies provides one explanation for why some investigations that limited cognitive contributions and emphasized temporal processing in their speech perception task still failed to observe a relationship between predicted CS levels and speech perception performance.

Still, it is noteworthy that none of the prior investigations using speech-in-noise perception tests with relatively high SNRs and/or speech stimuli with high levels of context found a robust relationship between estimated levels of noise-induced CS and speech-in-noise understanding scores. Viewed from the perspective of the subtle sensory challenges introduced by CS, this pattern of results highlights that specific characteristics of speech tasks may be most appropriate for investigating the putative influences of CS on speech perception in challenging listening conditions. In particular, tasks that utilize appropriately difficult SNRs for listeners with NHTs (i.e., those that result in substantial variability in performance) and maximize the importance of the sensory representation of temporal acoustic features, while minimizing other perceptual and cognitive factors that could influence an individual's performance, seem most suited to quantifying the relationship of CS to speech perception performance. Such tasks can be sensitive to subtle sensory deficits while maintaining at least some ecological validity to the challenges of everyday speech perception.

Resolving the question of whether CS impacts speech perception in human listeners is essential to the future of the field, and there are important clinical implications if CS can explain otherwise puzzling perceptual deficits. A link between auditory perceptual impairments in humans and moderate- to high-intensity sound exposure that does not permanently alter hearing thresholds could motivate systemic efforts to improve hearing protection education and guidelines. Compelling evidence that CS contributes to difficulties perceiving speech under adverse listening conditions could change how clinicians diagnose and treat this type of hearing impairment. Even apart from whether CS plays a significant role in human auditory perception, this area of study has incited widespread interest that may lead to the discovery of other neural and perceptual factors that impair speech-in-noise understanding in adults with NHTs.

Perhaps, most importantly, noise-induced CS is also likely to be comorbid with the outer hair cell loss that defines traditional hearing impairment [see Hickox et al. (2017)]. The perceptual consequences of CS could potentially interact with hair cell damage to further exacerbate listeners' challenges perceiving speech-in-noise. Complicating a straightforward examination of the interactive effects of traditional hearing loss and CS, however, is that elevated audiometric thresholds decrease the AN responses that are used to predict CS levels in humans (Bramhall et al. 2015; Kamerer et al. 2019). Development of physiological and perceptual assessments specific to CS will thus not only benefit the field of human CS but could also have substantial clinical significance for listeners with traditional hearing loss. While previous reviews have focused on the need to develop precise assessments of CS levels in humans, our review highlights the importance of using speech perception tasks that tap into the specific deficits that CS may cause—such as impairments in fine temporal processing.

ACKNOWLEDGMENTS

M. D. wrote the original manuscript; L. H. H. and B. G. S.-C. provided substantial feedback on paper text and figures. All authors revised the manuscript at all stages.

This study was funded by NIH NIDCD F32 DC018460 (M. D.), NIH NIDCD T32 DC011499 (M. D.), NIH NIDCD R01 DC017734 (L. L. H.), and NIH NIDCD R01 DC013825 (B. G. S.-C.).

The authors have no conflicts of interest to disclose.

Address for correspondence: Mishaela DiNino, Carnegie Mellon University Department of Psychology, 5000 Forbes Ave, Pittsburgh, PA 15213. E-mail mdinino@andrew.cmu.edu

Received February 26, 2021; accepted September 7, 2021

REFERENCES

Anderson, S., & Karawani, H. (2020). Objective evidence of temporal processing deficits in older adults. *Hear Res, 397*, 108053.).

Babkoff, H., & Fostick, L. (2017). Age-related changes in auditory processing and speech perception: cross-sectional and longitudinal analyses. *Eur J Ageing, 14*, 269–281.

Badri, R., Siegel, J. H., Wright, B. A. (2011). Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. *J Acoust Soc Am*, *129*, 852–863.

Banks, B., Gowen, E., Munro, K. J., Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *J Acoust Soc Am, 137*, 2015–2024.

Best, V., Keidser, G., Freeston, K., Buchholz, J. M. (2016). A dynamic speech comprehension test for assessing real-world listening ability. *J Am Acad Audiol, 27*, 515–526.

Bharadwaj, H. M., Mai, A. R., Simpson, J. M., Choi, I., Heinz, M. G., Shinn-Cunningham, B. G. (2019). Non-invasive assays of cochlear synaptopathy - candidates and considerations. *Neuroscience, 407*, 53–66.

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., Shinn-Cunningham, B. G. (2015). Individual differences reveal correlates of hidden hearing deficits. *J Neurosci, 35*, 2161–2172.

Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am, 122*, 1765.

Bramhall, N., Beach, E. F., Epp, B., Le Prell, C. G., Lopez-Poveda, E. A., Plack, C. J., Schaette, R., Verhulst, S., Canlon, B. (2019). The search for noise-induced cochlear synaptopathy in humans: Mission impossible? *Hear Res, 377*, 88–103.

Bramhall, N., Ong, B., Ko, J., Parker, M. (2015). Speech perception ability in noise is correlated with auditory brainstem response wave I amplitude. *J Am Acad Audiol, 26*, 509–517.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Brown, G. J., Ferry, R. T., Meddis, R. (2010). A computer model of auditory efferent suppression: Implications for the recognition of speech in noise. *J Acoust Soc Am, 127*, 943–954.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am*, *109*, 1101–1109.

Carcagno, S., & Plack, C. J. (2021). Effects of age on psychophysical measures of auditory temporal processing and speech reception at low and high levels. *Hear Res, 400*, 108117.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. *J Assoc Res Otolaryngol, 19*, 331–352.

Carroll, R., Warzybok, A., Kollmeier, B., Ruigendijk, E. (2016). Age-related differences in lexical access relate to speech recognition in noise. *Front Psychol, 7*, 990.

Couth, S., Prendergast, G., Guest, H., Munro, K. J., Moore, D. R., Plack, C. J., Ginsborg, J., Dawes, P. (2020). Investigating the effects of noise exposure on self-report, behavioral and electrophysiological indices of hearing damage in musicians with normal audiometric thresholds. *Hear Res, 395*, 108021.

Durlach, N. I., Mason, C. R., Kidd, G. Jr, Arbogast, T. L., Colburn, H. S., Shinn-Cunningham, B. G. (2003a). Note on informational masking. *J Acoust Soc Am, 113*, 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., Kidd, G. Jr. (2003b). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J Acoust Soc Am, 114*, 368–379.

Fernandez, K. A., Jeffers, P. W., Lall, K., Liberman, M. C., Kujawa, S. G. (2015). Aging after noise exposure: Acceleration of cochlear synaptopathy in "recovered" ears. *J Neurosci, 35*, 7509–7520.

Fulbright, A. N. C., Le Prell, C. G., Griffiths, S. K., Lobarinas, E. (2017). Effects of recreational noise on threshold and suprathreshold measures of auditory function. *Semin Hear, 38*, 298–318.

Füllgrabe, C., Moore, B. C. J., Stone, M. A. (2014). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Front Aging Neurosci*, *6*, 347.

Furman, A. C., Kujawa, S. G., Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *J Neurophysiol, 110*, 577–586.

Gelfand, S. A., Ross, L., Miller, S. (1988). Sentence reception in noise from one versus two sources: Effects of aging and hearing loss. *J Acoust Soc Am, 83*, 248–256.

Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., Chéry-Croze, S. (1997). Auditory efferents involved in speech-in-noise intelligibility. *Neuroreport, 8*, 1779–1783.

Grant, K. J., Mepani, A. M., Wu, P., Hancock, K. E., de Gruttola, V., Liberman, M. C., Maison, S. F. (2020). Electrophysiological markers of cochlear function correlate with hearing-in-noise performance among audiometrically normal subjects. *J Neurophysiol, 124*, 418–431.

Grinn, S. K., Wiseman, K. B., Baker, J. A., Le Prell, C. G. (2017). Hidden hearing loss? no effect of common recreational noise exposure on cochlear nerve response amplitude in humans. *Front Neurosci, 11*, 465.

Grose, J. H., Buss, E., Hall, J. W. III. (2017). Loud music exposure and cochlear synaptopathy in young adults: Isolated auditory brainstem response effects but no perceptual consequences. *Trends Hear, 21*, 2331216517737417.

Guest, H., Munro, K. J., Prendergast, G., Millman, R. E., Plack, C. J. (2018). Impaired speech perception in noise with a normal audiogram: No evidence for cochlear synaptopathy and no relation to lifetime noise exposure. *Hear Res, 364*, 142–151.

Hawley, M. L., Litovsky, R. Y., Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J Acoust Soc Am, 115*, 833–843.

Henry, K. S., Kale, S., Scheidt, R. E., Heinz, M. G. (2011). Auditory brainstem responses predict auditory nerve fiber thresholds and frequency selectivity in hearing impaired chinchillas. *Hear Res, 280*, 236–244.

Hickox, A. E., Larsen, E., Heinz, M. G., Shinobu, L., Whitton, J. P. (2017). Translational issues in cochlear synaptopathy. *Hear Res*, *349*, 164–171.

Hinchcliffe, R. (1992). King-Kopetzky syndrome: An auditory stress disorder? *J Audiol Med*, *1*, 89–98.

Hope, A. J., Luxon, L. M., Bamiou, D. E. (2013). Effects of chronic noise exposure on speech-in-noise perception in the presence of normal audiometry. *J Laryngol Otol, 127*, 233–238.

Jerger, J., Johnson, K., Jerger, S., Coker, N., Pirozzolo, F., Gray, L. (1991). Central auditory processing disorder: A case study. *J Am Acad Audiol, 2*, 36–54.

Johannesen, P. T., Buzo, B. C., Lopez-Poveda, E. A. (2019). Evidence for age-related cochlear synaptopathy in humans unconnected to speech-in-noise intelligibility deficits. *Hear Res, 374*, 35–48.

Kaandorp, M. W., De Groot, A. M., Festen, J. M., Smits, C., Goverts, S. T. (2016). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *Int J Audiol, 55*, 157–167.

Kamerer, A. M., Kopun, J. G., Fultz, S. E., Allen, C., Neely, S. T., Rasetshwane, D. M. (2019). Examining physiological and perceptual consequences of noise exposure. *J Acoust Soc Am, 146*, 3947.

Kawase, T., Delgutte, B., Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones. *J Neurophysiol, 70*, 2533–2549.

Kidd, G. Jr, Arbogast, T. L., Mason, C. R., Gallun, F. J. (2005). The advantage of knowing where to listen. *J Acoust Soc Am, 118*, 3804–3815.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., Banerjee, S. (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 Pt 1), 2395–2405.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J Neurosci, 29*, 14077–14085.

Kumar, U. A., & Vanaja, C. S. (2004). Functioning of olivocochlear bundle and speech perception in noise. *Ear Hear, 25*, 142–146.

Le Prell, C. G. (2019). Effects of noise exposure on auditory brainstem response and speech-in-noise tasks: A review of the literature. *Int J Audiol*, *58*(Suppl1), S3–S32.

Le Prell, C. G., Siburt, H. W., Lobarinas, E., Griffiths, S. K., Spankovich, C. (2018). No reliable association between recreational noise exposure and threshold sensitivity, distortion product otoacoustic emission amplitude, or word-in-noise performance in a college student population. *Ear Hear, 39*, 1057–1074.

Liberman, L. D., Suzuki, J., Liberman, M. C. (2015). Dynamics of cochlear synaptopathy after acoustic overexposure. *J Assoc Res Otolaryngol, 16*, 205–219.

Liberman, M. C., & Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hear Res, 349*, 138–147.

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PLoS One, 11*, e0162726.

Lobarinas, E., Spankovich, C., Le Prell, C. G. (2017). Evidence of "hidden hearing loss" following noise exposures that produce robust TTS and ABR wave-I amplitude reductions. *Hear Res, 349*, 155–163.

Lopez-Poveda, E. A. (2014). Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech. *Front Neurosci, 8*, 348.

Lopez-Poveda, E. A., & Barrios, P. (2013). Perception of stochastically undersampled sound waveforms: A model of auditory deafferentation. *Front Neurosci, 7*, 124.

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *J Acoust Soc Am, 124*, 3261–3275.

Madsen, S. M. K., Dau, T., Oxenham, A. J. (2019). The effect of harmonic number and pitch salience on the ability to understand speech-on-speech based on differences in fundamental frequency. *Proc Int Symp Auditory Audiol Res*, *7*, 45–52.

Makary, C. A., Shin, J., Kujawa, S. G., Liberman, M. C., Merchant, S. N. (2011). Age-related primary cochlear neuronal degeneration in human temporal bones. *J Assoc Res Otolaryngol, 12*, 711–717.

Marrufo-Pérez, M. I., Eustaquio-Martín, A., Lopez-Poveda, E. A. (2019). Speech predictability can hinder communication in difficult listening conditions. *Cognition*, *192*, 103992.

Mepani, A. M., Kirk, S. A., Hancock, K. E., Bennett, K., de Gruttola, V., Liberman, M. C., Maison, S. F. (2020). Middle ear muscle reflex and word recognition in "normal-hearing" adults: Evidence for cochlear synaptopathy? *Ear Hear, 41*, 25–38.

Mepani, A. M., Verhulst, S., Hancock, K. E., Garrett, M., Vasilkov, V., Bennett, K., de Gruttola, V., Liberman, M. C., Maison, S. F. (2021). Envelope following responses predict speech-in-noise performance in normal-hearing listeners. *J Neurophysiol, 125*, 1213–1222.

Palfery, T. D., & Duff, D. (2007). Central auditory processing disorders: Review and case study. Axone*, 28*, 20–23.

Pang, X. D., & Guinan, J. J. Jr. (1997). Effects of stapedius-muscle contractions on the masking of auditory-nerve responses. *J Acoust Soc Am, 102*, 3576–3586.

Parthasarathy, A., & Kujawa, S. G. (2018). Synaptopathy in the aging cochlea: characterizing early-neural deficits in auditory temporal envelope processing. *J Neurosci, 38*, 7108–7119.

Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., Polley, D. B. (2020). Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing. *Elife, 9*, e51419.

Plack, C. J., Barker, D., Prendergast, G. (2014). Perceptual consequences of "hidden" hearing loss. *Trends Hear, 18*, 2331216514550621.

Prendergast, G., Couth, S., Millman, R. E., Guest, H., Kluk, K., Munro, K. J., Plack, C. J. (2019). Effects of age and noise exposure on proxy measures of cochlear synaptopathy. *Trends Hear, 23*, 2331216519877301.

Prendergast, G., Millman, R. E., Guest, H., Munro, K. J., Kluk, K., Dewey, R. S., Hall, D. A., Heinz, M. G., Plack, C. J. (2017). Effects of noise exposure on young adults with normal audiograms II: Behavioral measures. *Hear Res, 356*, 74–86.

Resnik, J., & Polley, D. B. (2021). Cochlear neural degeneration disrupts hearing in background noise by increasing auditory cortex internal noise. *Neuron, 109*, 984–996.e4.

Ruggles, D., Bharadwaj, H., Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc Natl Acad Sci U S A, 108*, 15516–15521.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *J Exp Psychol Gen, 110*, 474–494.

Saunders, G. H., & Haggard, M. P. (1989). The clinical assessment of obscure auditory dysfunction–1. Auditory and psychological factors. *Ear Hear, 10*, 200–208.

Saunders, G. H., & Haggard, M. P. (1992). The clinical assessment of "Obscure Auditory Dysfunction" (OAD) 2. Case control analysis of determining factors. *Ear Hear, 13*, 241–254.

Sergeyenko, Y., Lall, K., Liberman, M. C., Kujawa, S. G. (2013). Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline. *J Neurosci, 33*, 13686–13694.

Shaheen, L. A., Valero, M. D., Liberman, M. C. (2015). Towards a diagnosis of cochlear neuropathy with envelope following responses. *J Assoc Res Otolaryngol, 16*, 727–745.

Shaw, G. M., Jardine, C. A., Fridjhon, P. (1996). A pilot investigation of high-frequency audiometry in obscure auditory dysfunction (OAD) patients. *Br J Audiol, 30*, 233–237.

Shehorn, J., Strelcyk, O., Zahorik, P. (2020). Associations between speech recognition at high levels, the middle ear muscle reflex and noise exposure in individuals with normal audiograms. *Hear Res, 392*, 107982.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif, 12*, 283–299.

Skoe, E., Camera, S., Tufts, J. (2019). Noise exposure may diminish the musician advantage for perceiving speech in noise. *Ear Hear, 40*, 782–793.

Smith, S. B., Krizman, J., Liu, C., White-Schwoch, T., Nicol, T., Kraus, N. (2019). Investigating peripheral sources of speech-in-noise variability in listeners with normal audiograms. *Hear Res, 371*, 66–74.

Valderrama, J. T., Beach, E. F., Yeend, I., Sharma, M., Van Dun, B., Dillon, H. (2018). Effects of lifetime noise exposure on the middle-age human auditory brainstem response, tinnitus and speech-in-noise intelligibility. *Hear Res, 365*, 36–48.

Valero, M. D., Burton, J. A., Hauser, S. N., Hackett, T. A., Ramachandran, R., Liberman, M. C. (2017). Noise-induced cochlear synaptopathy in rhesus monkeys (Macaca mulatta). *Hear Res, 353*, 213–223.

Vasilkov, V., Garrett, M., Mauermann, M., Verhulst, S. (2021). Enhancing the sensitivity of the envelope-following response for cochlear synaptopathy screening in humans: The role of stimulus envelope. *Hear Res, 400*, 108132.

Viana, L. M., O'Malley, J. T., Burgess, B. J., Jones, D. D., Oliveira, C. A., Santos, F., Merchant, S. N., Liberman, L. D., Liberman, M. C. (2015). Cochlear neuropathy in human presbycusis: Confocal analysis of hidden hearing loss in post-mortem tissue. *Hear Res, 327*, 78–88.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am, 91*, 1648–1661.

Wilson, R. H., & Burks, C. A. (2005). Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol. *J Rehabil Res Dev, 42*, 839–852.

Wu, P. Z., Liberman, L. D., Bennett, K., de Gruttola, V., O'Malley, J. T., Liberman, M. C. (2019). Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. *Neuroscience, 407*, 8–20.

Yeend, I., Beach, E. F., Sharma, M., Dillon, H. (2017). The effects of noise exposure and musical training on suprathreshold auditory processing and speech perception in noise. *Hear Res, 353*, 224–236.

Zhao, F., & Stephens, D. (2000). Subcategories of patients with King-Kopetzky syndrome. *Br J Audiol, 34*, 241–256.

Zhao, F., & Stephens, D. (2006). Distortion product otoacoustic emissions in patients with King-Kopetzky syndrome. *Int J Audiol*, *45*, 34–39.