Individual Differences Reveal the Utility of Temporal Fine-Structure Processing for Speech Perception in Noise

Short Title: TFS Coding and Speech Perception in Noise

*Agudemu Borjigin[a,c] and Hari M. Bharadwaj[a,b,d]

[a] *Weldon School of Biomedical Engineering, Purdue University, West Lafayette IN 47907, USA*

[b] *Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette IN 47907, USA*

[c] *Waisman Center, University of Wisconsin - Madison, Madison, WI 53705, USA*

[d] *Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA 15213, USA*

Funding Acknowledgements: This research was supported by funding from the National Institutes of Health [Grant R01DC015989].

Acknowledgements: We thank our research subjects for their participation in this study.

Declaration of Conflicting Interests: The Authors declare that there are no competing interests.

Data Accessibility: All data will be made publicly available upon publication of the manuscript.

*Correspondence: dagu@wisc.edu

Abstract

The auditory system is unique among sensory systems in its ability to phase lock to and precisely follow very fast cycle-by-cycle fluctuations in the phase of sound-driven cochlear vibrations. Yet, the perceptual role of this temporal fine structure (TFS) code is debated. This fundamental gap is attributable to our inability to experimentally manipulate TFS cues without altering other perceptually relevant cues. Here, we circumnavigated this limitation by leveraging individual differences across 200 participants to systematically compare variations in TFS sensitivity to performance in a range of speech perception tasks. Results suggest that robust TFS sensitivity does not confer additional masking release from pitch or spatial cues, but appears to confer resilience against the effects of reverberation. Yet, across conditions, we also found that greater TFS sensitivity is associated with faster response times, consistent with reduced listening effort. These findings highlight the perceptual significance of TFS coding for everyday hearing.

Key words: TFS, speech perception in noise, individual differences, reverberation, listening effort, online testing

1 Introduction

Human connection and communication fundamentally rely on the auditory system's capacity to encode and process complex sounds such as speech and music. Regardless of complexity, all acoustic information we receive from our environment is conveyed through the firing rate and spike timing of cochlear neurons (i.e., rate-place vs temporal coding) (Oxenham, 2018). Temporal information in sound is comprised of only two components: the rapid variations in phase—the temporal fine structure (TFS) and slower amplitude variations —the temporal envelope (Hilbert, 1906). Neurons in the auditory system can robustly track both TFS (Johnson, 1980) and envelope (Joris and Yin, 1992) through phase-locked firing. Strikingly, neural phase locking to TFS extends at least to 1000 Hz in the peripheral auditory system (Verschooten et al., 2019), a feat unmatched by other sensory modalities. In comparison, phase-locked information in the visual and somatosensory systems reaches only to about 50 Hz (Shadlen and Newsome, 1994; Baden et al., 2011). In the auditory system however, this uniquely high upper-frequency limit of phase locking only exists at the peripheral level (Verschooten et al., 2019; Wallace et al., 2000). Along the ascending pathway, the phased-locked temporal code appears to be progressively transformed into a rate-place representation (Joris et al., 2004). It seems that the auditory system initially invests heavily in this exquisite and metabolically expensive (Laughlin et al., 1998; Hasenstaub et al., 2010) phase-locked

temporal codes but then "repackages" the code into a different form for downstream processing. How this initial neural coding of TFS ultimately contributes to perception, and if and how its degradation leads to perceptual deficits is a fundamental open question not only for the neuroscience of audition, but also for clinical audiology. Yet, the significance of this peripheral TFS phase-locking in the auditory system remains controversial (Oxenham, 2013; Drullman, 1995; Swaminathan and Heinz, 2012; Oxenham and Simonson, 2009; de Cheveigné and Pressnitzer, 2006; Saddler et al., 2021; Javel and Mott, 1988; Verschooten et al., 2019).

Psychophysical experiments in *quiet* sound booths suggest that TFS may play a role in sound localization (Smith et al., 2002; Yin and Chan, 1990) and pitch perception (Moore, 1973; Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). Both spatial and pitch information can serve as primary cues for target-background segregation and selective attention in more realistic listening settings, yielding a masking release of about 5 dB each (Ihlefeld and Shinn-Cunningham, 2008; Gallun et al., 2013, 2015; Srinivasan et al., 2016; Jakien Kasey M. et al., 2017; Jakien Kasey M. and Gallun Frederick J., 2018; Brokx and Nooteboom, 1982; Bird and Darwin, 1997; Summers Van and Leek Marjorie R., 1998). Yet, whether this masking release is attributable to TFS coding is debated. This is because the other component of sound—the temporal envelope, despite eliciting weaker pitch or spatial percepts in quiet, can provide a similar degree of masking-release benefit in noise (Best et al., 2005; Oxenham and Simonson, 2009). Furthermore, TFS-based spatial cues, by virtue of being available primarily at low-frequencies, are more susceptible to corruption by reverberation than envelope-based spatial cues (Ihlefeld and Shinn-Cunningham, 2011; Richards and Wiley, 1980). Thus, despite many decades of intensive research, whether phase-locked temporal coding of TFS would introduce additional masking-release benefits in the presence of realistic listening conditions remains unclear.

A key challenge to understanding the perceptual role of TFS phase locking is that sub-band vocoding, which is the most common technique employed to investigate this question, is inherently limited (Smith et al., 2002; Hopkins et al., 2008; Hopkins and Moore, 2009; Lorenzi et al., 2006, 2009; Ardoint and Lorenzi, 2010). Vocoding has been used to acoustically dissociate TFS from envelope by creating stimuli with a constant envelope (i.e., sub-band amplitude) while manipulating the TFS (i.e., sub-band phase). Unfortunately, this clean dissociation at the acoustic level is not maintained at the output of cochlear processing, which interconverts some of the TFS cues to amplitude fluctuations (Oxenham, 2013; Swaminathan and Heinz, 2012). Recent approaches to investigate the perceptual significance of TFS coding have

leveraged deep neural networks (DNN) (Saddler et al., 2021; Saddler and McDermott, 2022). However, similar to sub-band vocoding, they are also subject to the confounding effects of cochlear processing in the absence of detailed knowledge of peripheral physiology at the individual level. Studies such as Hopkins et al. (2008) and Smith et al. (2002) kept the envelope but reduced the TFS content and completely swapped the TFS with that of a different speech, respectively, to study the role of TFS. However, these studies are subject to a broader limitation of stimulus-manipulation approaches: participants may use and weight TFS cues differently depending on the availability of other redundant cues, and thus differently in synthetic versus naturalistic stimuli. An alternative approach that can overcome these limitations is to avoid any stimulus manipulations, but directly measure individual differences in TFS processing and compare them to individual differences in speech-in-noise outcomes tested with intact, unprocessed stimuli. The individual differences approach has been successfully used to address other fundamental questions in the neuroscience of audition (McDermott et al., 2010; Bharadwaj et al., 2015; Whiteford et al., 2020). However, this approach has not been used to investigate the role of TFS given that robust measures of TFS sensitivity at the individual level were only recently established by comparing behavioral and electrophysiological assays (Borjigin et al., 2022).

Here, we leveraged these individualized TFS processing measures, and adapted them for remote testing to circumnavigate the COVID-19-related restrictions (Mok et al., 2023). We hypothesized that TFS plays an important role in everyday hearing, more specifically, in masking release from noise interference. To elucidate the role of TFS in everyday listening, we compared individual TFS sensitivity to individual participants' speech-perception outcomes under various types of noise interference. The speech-in-noise test battery included 10 different listening conditions, representing many important aspects of everyday listening where TFS phase locking has conventionally been thought to play a role. More specifically, we predicted that individuals with better TFS sensitivity would benefit more from pitch and spacial cues in noisy listening settings. Individuals with better TFS sensitivity would also be less subjected to the impact from informational masking and reverberation. Furthermore, we hypothesized that individuals with better TFS sensitivity would experience less strenuous listening in noisy listening settings. The automated and parallel nature of the online measurements allowed us to rapidly collect data from a large cohort of 200 subjects, affirming the promise and advantages of online behavioral psychoacoustical studies. Figure 1 illustrates the design of this study. The results revealed that better TFS processing, although not associated with greater masking release, provided resilience against the impact from reverberation,

and lessened listening effort. Given that reverberation is the most common source of signal corruption in everyday listening, and that listening effort is often a primary patient complaint in the audiology clinic, these findings highlight the perceptual significance of TFS coding in everyday communication.

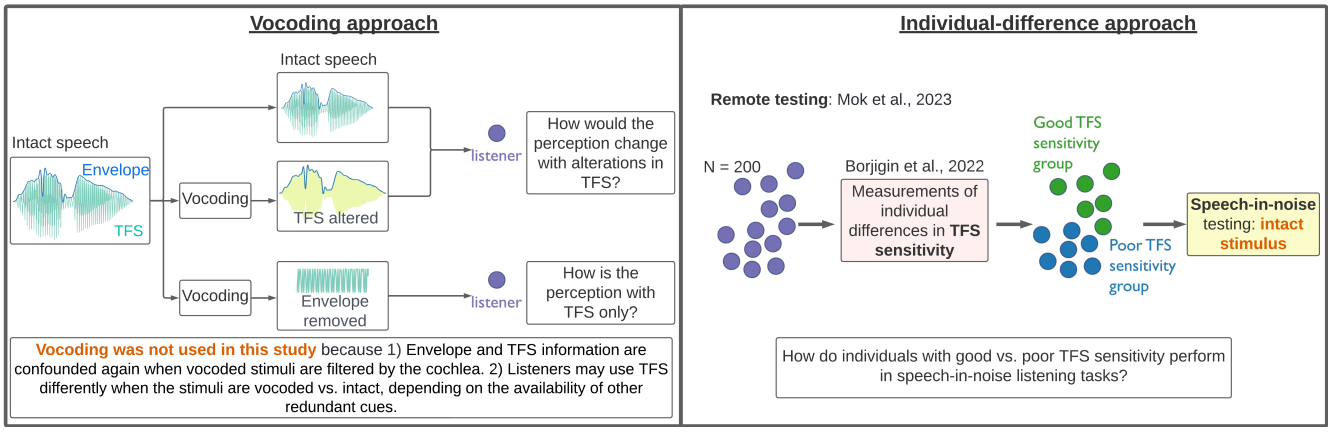


Figure 1. Contrasting the conventional vocoding approach (left) for studying TFS vs the individual-difference approach adopted in this study (right).

2 Results

2.1 Binaural temporal sensitivity measures captured individual differences in TFS processing fidelity.

Figure 2a) illustrates a scatter plot of the individual differences that we observed in our two binaural temporal sensitivity measures—ITD discrimination and binaural FM detection measurements. Metrics of individual TFS sensitivity commonly used in literature are prone to the impact of extraneous "non-sensory" variables (Borjigin et al., 2022) such as attention and motivation. Here, ILD discrimination was used as a surrogate measure to control for "non-sensory" factors as well as non-specific readout processing of binaural information. The individual differences in these "adjusted" TFS metrics are more likely driven by true individual differences in TFS processing (See Section 4.2.2 for further details). Individual ILD sensitivity data are shown in Figure 2b, which also indicates substantial individual variability.

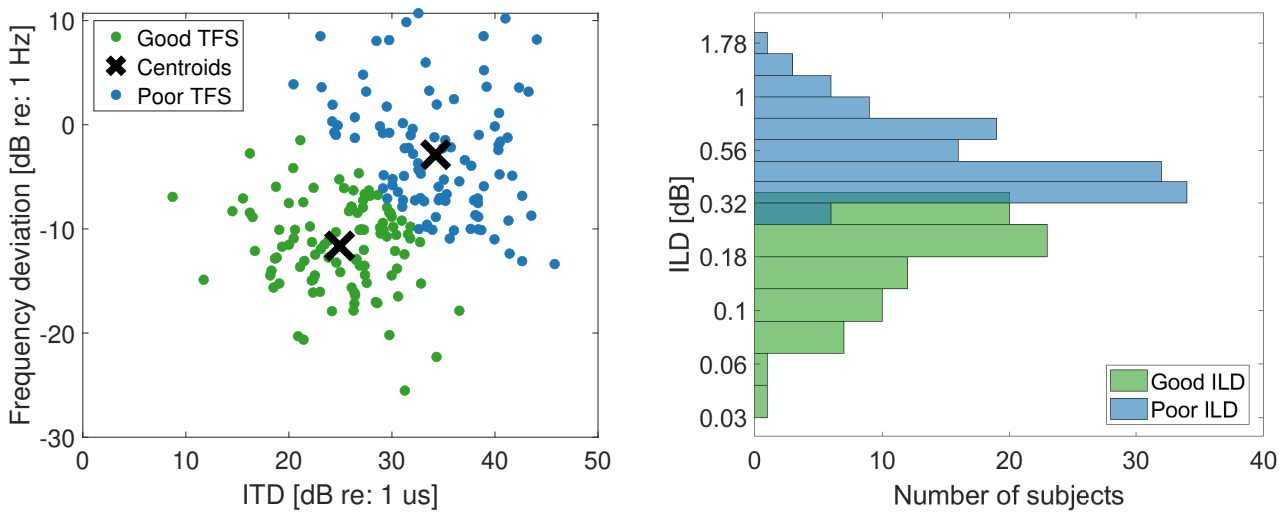
The adjusted binaural FM detection and ITD discrimination measures were significantly correlated ($r = 0.3, p < .0001$) indicating a common underlying source of variance attributable to TFS processing. Accordingly, subjects were divided into two groups by a clustering algorithm based on these two measures

into "Good-" vs "Poor-TFS" groups. Figures 3a and 3b demonstrate a clear separation of the two groups' psychometric curves in the ITD and binaural FM measurements, respectively. As expected, there's no such separation in the ILD measurement (Figure 3c). The construction of groups based on common variance across the TFS measures after eliminating differences in ILD sensitivity ensures that the grouping in Figure 2a is based on individuals' TFS sensitivity, rather on other unrelated factors. Note also that there is no significant difference in age between two groups ("Good TFS" group: mean age of 30.4 years with an std of 7.7 years; "Poor TFS" group: mean age of 32.1 years with an std of 8.4 years). To corroborate this further, individuals were also grouped based on their ILD discrimination thresholds, as shown in Figure 2b. For this alternative grouping, a clear separation is evident in the psychometric curves for ILD discrimination (Figure 3f), but not in the curves for TFS measurements (Figure 3d or 3e), consistent with the notion that the grouping in Figure 2b captures "non-TFS" variability. This alternative non-TFS regrouping of participants is also used to validate the observed group-level effects of better TFS processing on speech-in-noise outcomes, as detailed in section 2.3.

Furthermore, the web-based measurements in the present study produced data that is comparable to the data not only from our previous in-person study, but also from other labs. Figure 4 shows across-study comparisons for FM detection and ITD discrimination measurements. The left panel compares online measurement of binaural FM detection with in-person results from Grose and Mamo (2012); Whiteford and Oxenham (2015); Whiteford et al. (2017); Lelo de Larrea-Mancera et al. (2020). The right panel includes a sample of in-person studies that measured ITD discrimination (Henning, 1983; Dye, 1990; Brughera et al., 2013; Borjigin et al., 2022). Borjigin et al. (2022) is our previous in-person study. These results further validate our choice of TFS sensitivity measures.

2.2 Better TFS sensitivity does not confer additional masking-release benefit

To understand the functional role of TFS in everyday hearing, we measured participants' speech intelligibility under various types of noise interference, in addition to evaluating TFS sensitivity. Rather than absolute speech reception threshold (SRT, the lowest/noisiest level at which a person can understand speech in noise), Figure 5a depicts the masking release. Masking release refers to improved noise tolerance associated with the following cues: pitch difference between the target and background speakers, spatial separation between the target and maskers, combination of both pitch and spatial cues, and fi-



(a) Grouping based on TFS sensitivity **(b)** Grouping based on ILD sensitivity

Figure 2. (a) An illustration of cluster assignment based on the combination of ITD and binaural FM thresholds. (b) Group assignment based on ILD sensitivity. Note that the ILD sensitivity values displayed are the residuals after regressing out TFS scores from the raw ILD measurements.

nally when the background noise is non-speech stationary noise instead of speech babble. The masking release effects observed in this study are consistent with those reported in previous research: 1) With pitch separation, the participants could more easily identify the target compared to when the target and background had similar pitch (i.e., the reference condition, not shown). This pitch-based masking release is about 5 dB (Good TFS group: mean = 4.78, std = 0.47; Poor TFS group: mean = 5.3, std = 0.41), which matches previous reports from Brokx and Nooteboom (1982); Bird and Darwin (1997); Summers Van and Leek Marjorie R. (1998). 2) Similarly, participants obtained a masking release of around 3 dB when the target and background are spatially separated from each other (Good TFS group: mean = 2.94, std = 0.39; Poor TFS group: mean = 3.73, std = 2.31), which again aligns with earlier reports (Ihlefeld and Shinn-Cunningham, 2008; Gallun et al., 2013, 2015; Srinivasan et al., 2016; Jakien Kasey M. et al., 2017; Jakien Kasey M. and Gallun Frederick J., 2018). 3) In addition, when both pitch and spatial cues were presented to listeners, the masking-release benefits appeared to be cumulative, totaling about 10 dB as demonstrated in the "pitch + space" condition (Good TFS group: mean = 10.25, std = 0.81; Poor TFS group: mean = 9.45, std = 0.64). Indeed, it has previously been shown that pitch differences aid listeners in spatially separating competing sounds (Darwin, 2005). The voices from various spatial locations being distinct is perhaps also more ecologically relevant. 4) A masking release of about 19 dB was observed

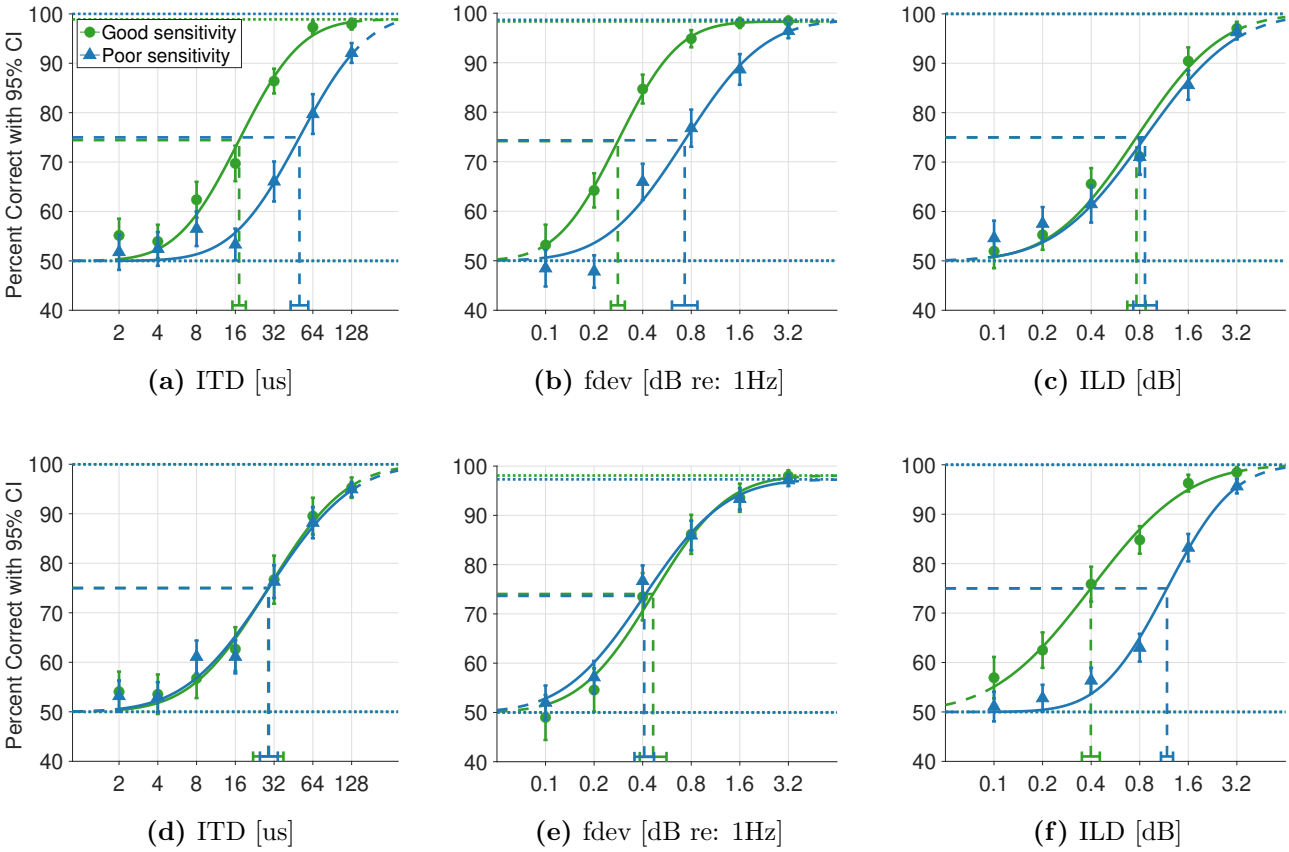


Figure 3. (a)-(c): Psychometric curves for Good vs. Poor TFS-sensitivity groups. Left: ITD; mid: binaural FM (fdev: frequency deviation); right: ILD. (d)-(f): Psychometric curves for Good vs. Poor ILD-sensitivity groups. Error bars represent within-group standard error of the mean.

when the background noise was switched from 4-talker babble speech to non-speech stationary noise, as shown in the "steady noise" condition (Good TFS group: mean = 19.22, std = 0.71; Poor TFS group: mean = 18.3, std = 0.7). This suggests that a substantial component of the masking associated with 4-talker babble derives from acoustic-linguistic similarities between the target, which is often referred to as "informational masking" (Viswanathan et al., 2021b; Arbogast et al., 2002; Kidd and Colburn, 2017). Just as with the TFS sensitivity measures (Figure 4), the consistency of these results with prior literature further confirms the viability of the online testing platform in reproducing in-person measurements.

Figure 5b illustrates the masking release in the same four conditions as in Figure 5a, except for the addition of reverberation to all conditions. Note that the reference condition (i.e., babble-speech condition, with no pitch or spatial cues, not shown) also contained reverberation. Reverberation generally reduced the masking-release benefit, except for the pitch-only condition. This is consistent with previous studies showing that reverberation has a smaller impact on monaural cues (Ruggles et al., 2011; Shinn-

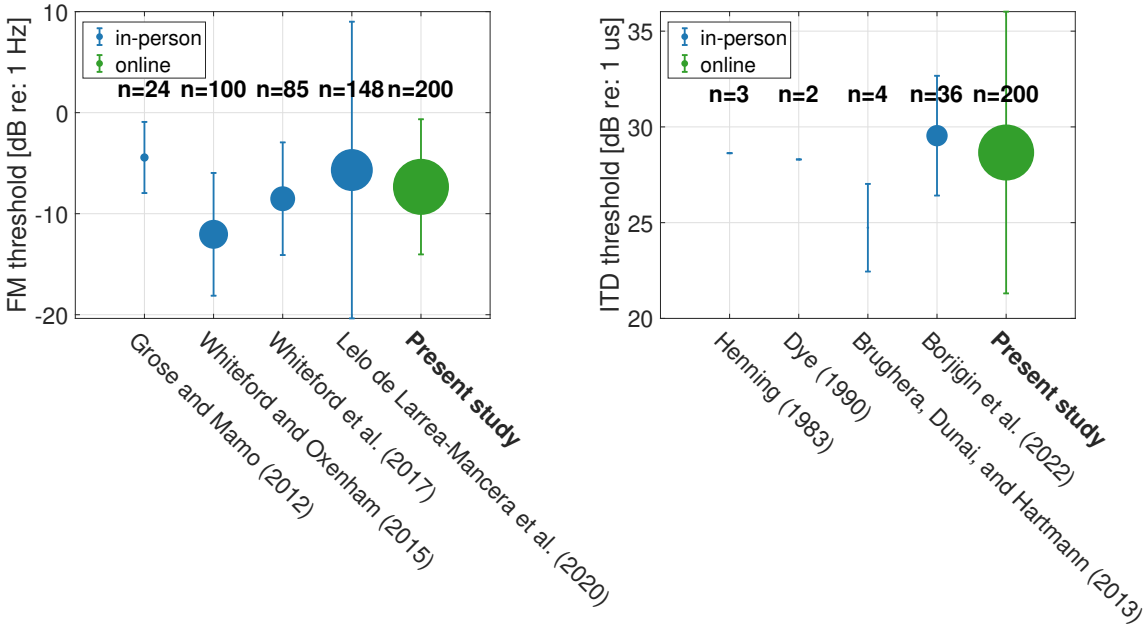
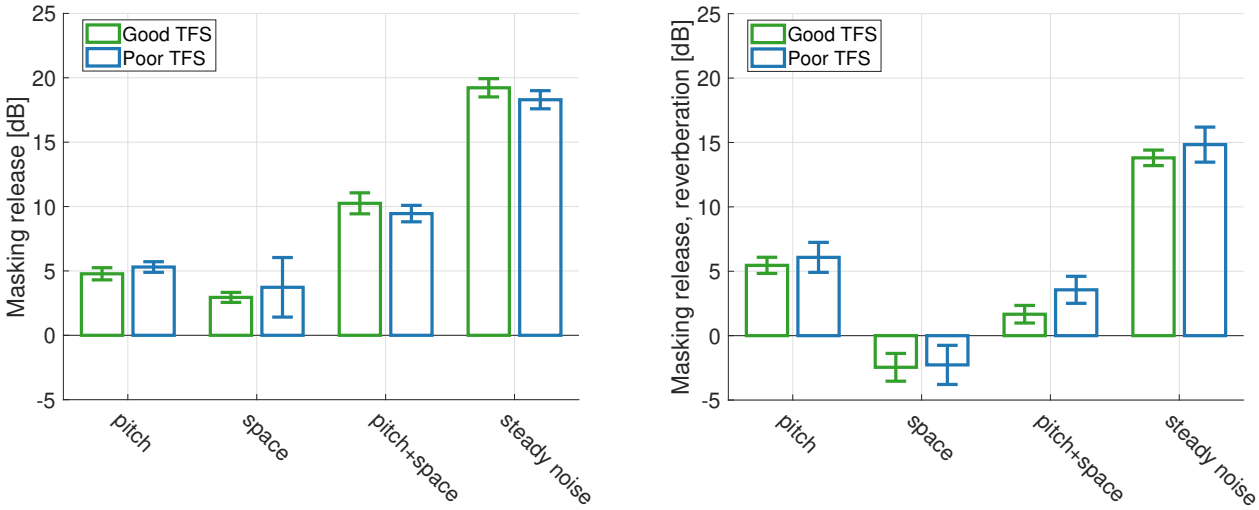


Figure 4. A sample of published reports of binaural FM detection thresholds (left) and ITD discrimination threshold (right) for comparison against the present study. Error bar represents standard deviation (std). The std of ITD detection thresholds could not be determined for Henning (1983) and Dye (1990). The size of the dot represents the number of subjects.

Cunningham et al., 2017; Culling et al., 1994), while spatial hearing is subject to substantial degradation (Ihlefeld and Shinn-Cunningham, 2011; Ruggles et al., 2011; Palomäki et al., 2004).

More importantly, Figure 5a and 5b demonstrate similar masking release for participants divided across the two groups based on their TFS sensitivity (see the grouping result in 2a). This comparison between groups directly addresses the primary research objective of this study: the perceptual significance of TFS for everyday hearing. In both non-reverberant (Figure 5a) and reverberant conditions (Figure 5b), the Good-TFS group did not benefit more from the cues in terms of masking release in any of the conditions tested. This is consistent with other studies suggesting that better TFS processing might not necessarily benefit a listener by conferring *more* masking release when other envelope-based cues are also available (Best et al., 2005; Oxenham and Simonson, 2009; Freyman et al., 2012).



(a) Masking release, without reverberation

(b) Masking release, with reverberation

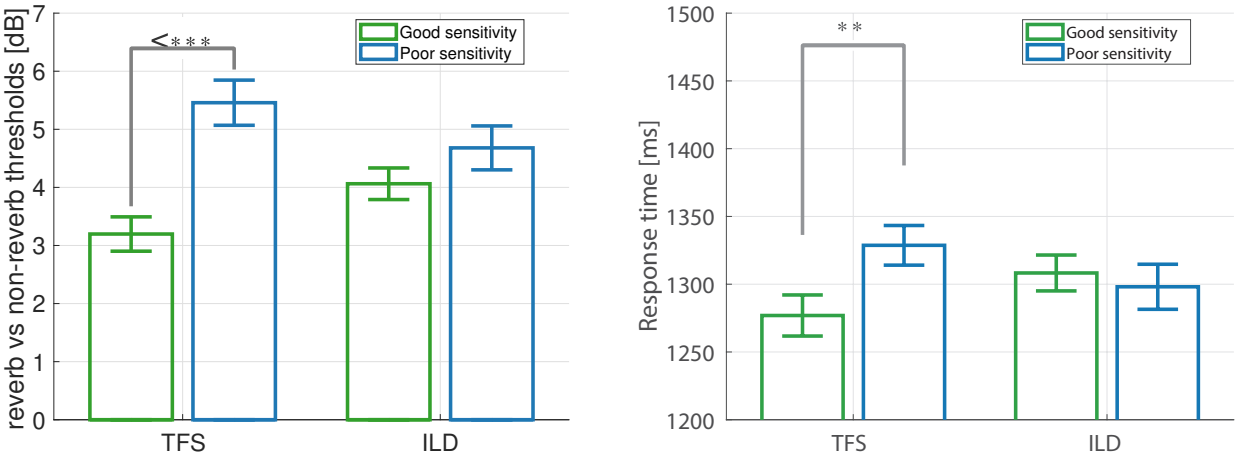
Figure 5. Masking release across conditions. The height of the bars represents the mean, error bars represent 1 standard deviation. Masking release was calculated by subtracting the SRT in each condition from that of the reference condition (not shown). Note that the reference condition in (a) does not have reverberation, whereas the reference condition in (b) contains reverberation. A positive masking release means that the SRT was lower/better than that of the reference condition.

2.3 Better TFS processing reduces the impact of reverberation and lessens cognitive load associated with speech perception in noise.

Despite the lack of group differences in masking release, the absolute SRTs for the individual conditions, including the reference condition, worsened substantially due to reverberation. Note that the masking release is the difference between the SRT in a specific condition and that of the corresponding reference condition. To investigate the role of TFS in mitigating the effects of reverberation, the threshold increase from non-reverberant to reverberant conditions was also calculated, shown by the height of the bars in Figure 6a. When the participants were grouped based on their TFS sensitivity (Figure 6a, left), the two groups exhibited a significant difference in the increase/worsening of SRTs due to reverberation: the group with poor TFS sensitivity experienced a much greater threshold increase in reverberant settings than their good-TFS counterparts. When the participants were divided based on their non-TFS sensitivity (i.e., ILD sensitivity), there was no longer a significant group difference, indicating an important role for TFS-processing fidelity. This result suggests that improved TFS sensitivity can mitigate the negative impact of reverberation, which is perhaps the most common source of signal degradation in everyday

listening scenarios.

It is well known that behavioral measures of performance can obscure important differences in the cognitive effort expended by different listeners in achieving a given level of performance (Peelle, 2018; Alhanbali et al., 2019). To investigate whether robust TFS sensitivity is associated with less effortful listening, we examined response times as a proxy for the degree to which a listener invokes different cognitive resources to understand speech in noise. The response times are indicated by the height of the bars in Figure 6b. The abslute values of the response times are consistent with prior literature (Apoux et al., 2001). When the participants were divided into two groups based on their TFS sensitivity (Figure 6b, left), the Good-TFS group exhibited significantly shorter reaction times compared to the Poor-TFS group, consistent with reduced listening effort in performing the task. Similar to the previous analysis, when the listeners were regrouped based on non-TFS characteristics (i.e., ILD sensitivity), there was no longer a significant difference between the two groups (Figure 6b, right). Taken together, these results show that robust TFS sensitivity, rather than non-sensory factors, is associated with shorter reaction times, suggesting a possible benefit of reduced listening effort conferred by better TFS processing.



(a) Threshold increase in reverberation

(b) Response time

Figure 6. (a) Increase in SRT due to reverberation for each group of participants. (b) Average response times. Data were pooled across all conditions shown in Figure 5. All error bars represent estimated standard error of the mean.

3 Discussion

Despite decades of intensive research, the contribution of TFS to everyday hearing is still debated. Our inability to precisely pinpoint the utility of TFS cues is likely due to limitations inherent to the vocoding approach historically used to address this question (Smith et al., 2002; Hopkins et al., 2008; Hopkins and Moore, 2009; Lorenzi et al., 2009; Ardoint and Lorenzi, 2010). To circumnavigate these limitations, here, we used intact stimuli and leveraged an individual difference approach. Specifically, we measured individual differences in TFS sensitivity and speech-in-noise perception in the same large cohort of participants (N=200) and compared them directly. Overall, we found that improved TFS processing confers resilience against the deleterious effects of reverberation on speech intelligibility, and reduces response times in noisy environments consistent with a lowering of listening effort.

Surprisingly, no *additional* spatial release from masking benefit was observed in the "good-TFS" group despite the strong theoretical connection between TFS phase locking and binaural temporal processing (Yin and Chan, 1990; Smith et al., 2002) (Figure 5a). Brainstem binaural circuits compare temporal information encoded by TFS phase locking from each ear and can encode microsecond ITDs that form one of two main cues supporting spatial hearing along the horizontal plane. Accordingly, we hypothesized that individuals with better TFS sensitivity would benefit more from the spatial cues in speech-in-noise tasks. One of the reasons why we did not see a group difference here may be that the participants were all typically hearing; individual differences in TFS sensitivity may not have been sufficiently large. A group difference may be observable if a broader range of TFS sensitivity is represented in the cohort by including individuals with hearing loss. Furthermore, the use of ILD detection as a reference for non-TFS factors could have also contributed to the lack of group difference in spatial release from masking. ILD also activates binaural circuits, although ILD-based binaural processing does not rely on TFS phase locking (Tollin et al., 2008). Regressing out ILD scores from binaural TFS measurements could have removed any individual variability in aspects of spatial hearing that go beyond sensitivity to TFS cues, such as the efficacy with which downstream "readout" processes use binaural information. Thus, rather than contradicting the prevailing view that TFS processing is critical to spatial hearing (Churchill et al., 2014; Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002; Brughera et al., 2013; Yin and Chan, 1990; Smith et al., 2002), our result simply suggests that the range of individual differences observed in ITD thresholds did not translate to measurable differences in the degree of spatial release from masking.

Similarly, no significant group difference was observed for pitch-based masking release. Although TFS processing is widely acknowledged as important for low-frequency spatial hearing, its role in pitch perception has been debated for over 150 years (Seebeck, 1841; Ohm, 1843). Humans perceive low-frequency periodic sounds as having a stronger pitch than high-frequency sounds (Moore, 1973; Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). Pitch discrimination declines above 4-5 kHz (Attneave and Olson, 1971; Oxenham et al., 2011), which aligns with the low-pass characteristic of TFS phase locking in the auditory nerve (Rose et al., 1967; Palmer and Russell, 1986). Furthermore, the deficits in TFS coding have been invoked to explain speech perception deficits in fluctuating noise designed to emphasize the role of pitch cues (Lorenzi et al., 2006; Qin and Oxenham, 2003; Stickney et al., 2007). While these findings appear to suggest that TFS may play a role in pitch perception, the same observations also permit alternative interpretations based on place coding, which also worsens at higher frequencies and in individuals with hearing loss (Oxenham, 2012, 2013). The result from the present study, i.e., the similar pitch-based masking release across good- and poor-TFS groups is consistent with "place-coding" based explanations of pitch phenomena.

The "steady noise" condition used in the present study (Figure 5a) was designed to maximize energetic masking (Stone and Moore, 2014) (see Methods), while the 4-talker babble masker in the reference condition contained many sources of informational masking (acoustic modulations, phonetic/lexical/semantic content) in addition to energetic masking (Durlach et al., 2003). The improvement of almost 20 dB in SRTs from the reference to the "steady noise" condition [consistent with Arbogast et al. (2002)] points to the dominant role of informational masking in everyday listening (Brungart et al., 2006). Listening in the presence of informational masking is thought to involve many sensory and cognitive processes in the central auditory system including object formation and scene segregation/streaming, auditory selective attention, working memory, and linguistic processing (Darwin, 1997; Kidd and Colburn, 2017). TFS-based processing is thought to play an important role for scene segregation and attentive selection (Shinn-Cunningham, 2008; Viswanathan et al., 2021b,a). Although our results show similar degree of release from informational masking across the two TFS-sensitivity groups (Figure 5a), the group with better TFS sensitivity had a significantly shorter response time than the poorer TFS group (Figure 6b). Our results, therefore, affirm the contribution of TFS coding for enabling robust central auditory processing, possibly with lower listening effort. The fact that the group difference in reaction times did not translate into the masking-release metrics underscores the need to investigate cognitive factors beyond

performance/score metrics to fully characterize the importance of different peripheral cues (Loughrey et al., 2018; Nuesse et al., 2018; Humes, 2021; Gallun et al., 2022).

Finally, we explored the correlation between TFS processing and listening in a reverberant environment. The SRTs were considerably worsened by the presence of reverberation (Figure 6a). More importantly, the group with poor TFS sensitivity was impacted significantly more than their good TFS counterparts, indicating a possible role of TFS processing in "de-reverberation". Reverberation impairs TFS-based spatial cues (Ihlefeld and Shinn-Cunningham, 2011) and spatial selective attention (Ruggles et al., 2011). Thus, our findings suggest that stronger TFS coding may ameliorate reverberation's detrimental effects on speech perception in noise—another benefit of good TFS processing revealed in this study in addition to reducing listening effort.

Overall, this study represents a detailed examination of the role of TFS, encompassing several major aspects of speech perception in noise where TFS processing has been thought to be important. Our experiments investigated the role of pitch (fundamental frequency) cues, spatial hearing, informational masking, and listening under reverberation. Although prior literature on the topic is extensive, they have been inconclusive about the role of TFS and constrained by the limitations of the traditional vocoding approach. In contrast, our study adopted an individual-difference approach and systematically measured TFS sensitivity and speech-in-noise intelligibility in a large subject cohort (N=200), thereby addressing this foundational gap. The results not only corroborated numerous prior findings but crucially, also highlighted the importance of TFS for reducing listening effort and for listening in reverberation. These findings, together with the fact that most cochlear implants (CIs) do not convey TFS also help explain the strenuous listening experienced by CI users every day, especially in the presence of reverberation. The findings in this study also suggest that evaluating TFS processing fidelity may complement conventional assessments used in audiology clinics to help characterize speech perception deficits in background noise (Ruggles et al., 2011, 2012; Parthasarathy et al., 2020). The test battery devised in this study is individual difference-sensitive, serving as a framework for gauging TFS sensitivity in clinical settings. Finally, our results also affirm the promise of using web-based psychoacoustics to conduct larger-scale experiments (Mok et al., 2023). Automated and parallelized data collection facilitated the rapid acquisition of data from a large subject population within a few days, providing a substantial advantage over traditional in-person psychoacoustic testing in this context.

4 Materials and Methods

This study aimed to explore the perceptual role of TFS processing in everyday hearing, specifically speech perception in noise, by assessing TFS sensitivity and speech-in-noise outcomes in a large cohort of typically hearing individuals. Owing to COVID-19-related constraints, this study collected all data remotely using a platform previously developed and validated in the lab (Mok et al., 2023). The battery of measures included two TFS-sensitivity measures, a reference measurement unrelated to TFS processing to dissociate the contribution of non-sensory variables, and speech-in-noise outcome measures (snaplabonline). The study battery is summarized in Figure 7.

4.1 Participants

Two hundred participants were recruited anonymously from Prolific.co (20-55 years old (mean=31.2, SD=8.1); 93 females, 102 males, and 5 not reported). Eighty five percent of the participants self-reported English as their first language, although all participants were native speakers of North American English. In terms of race and ethnicity, 64% self reported as White, 21% Asian, 6.5% Mixed, 2.5% Black, 2.5% Other, and 3.5% not reported. Participants reported no hearing loss, neurological disorders, or persistent tinnitus, and passed headphone checks and a speech-in-noise-based hearing screening (Mok et al., 2023). The participants consented to participate in the study following Institutional Review Board (IRB) protocols established at Purdue University and were compensated for their time. All participants completed the full study battery. The median time for completion was approximately 1 hour.

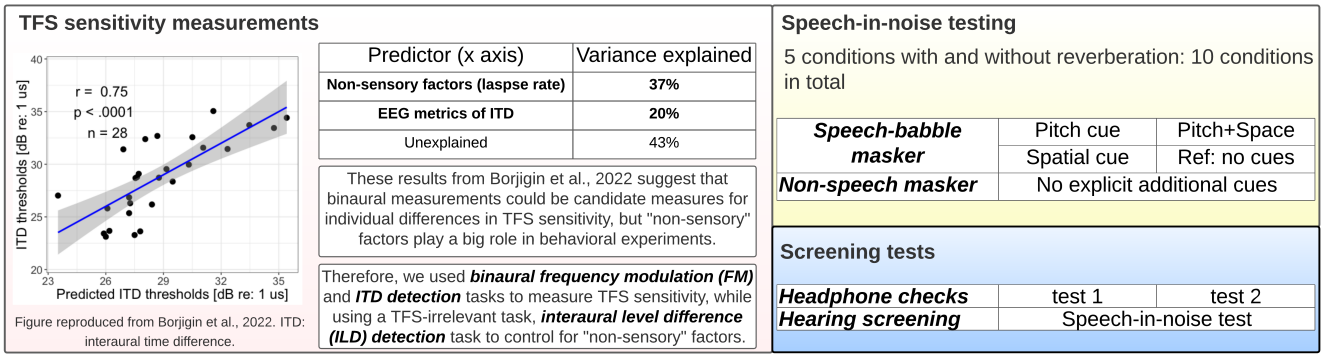


Figure 7. An illustration of all measures included in the present study.

4.2 Experimental Design and Statistical Analyses

4.2.1 Screening Measurements.

All measurements, including the screening, are listed in Figure 7. Because participants were anonymous and used their own computers and headphones, two screening procedures were administered to narrow the pool of participants to individuals with typical hearing, and to ensure stereo headphone uses critical for binaural hearing measurements.

Headphone-Check. Two tests based on previously established procedures were carried out to screen for appropriate use of headphones (Mok et al., 2023). In the first task, participants were instructed to identify the softest of a sequence of three low-frequency tones. The target tone was 6 dB softer than the two foil tones, but one of the decoy tones was presented with opposing phase between the left and right channels (Woods et al., 2017). Woods et al. (2017) reasoned that if a participant used a pair of sound field speakers instead of headphones, acoustic cancellation would result in a greater attenuation of the anti-phase decoy tone leading to an error. However, the procedure becomes ineffective if a participant uses only one speaker/channel. To catch participants who use a single-channel set up, we added a second task where participants were asked to report whether a low-frequency chirp (150–400 Hz) embedded in background low-frequency noise was rising, falling, or flat in pitch. The stimulus was designed such that chirp was at pi-phase between the left and right channels, whereas the noise was at zero phase (i.e., a so-called "N0Sπ" configuration). The signal-to-noise ratio was chosen such that the chirp would be difficult to detect with just one channel, but easily detected with binaural headphones owing to the so-called binaural masking level difference (BMLD) effect that would yield a substantial masking release (Licklider, 1948).

Hearing Screening. Participants were screened for putative normal-hearing status using a speech-in-noise task previously validated for this purpose (Mok et al., 2023). A previous meta-analysis of 15 studies suggested that speech-in-noise tasks yield a large effect size separating individuals with normal hearing and hearing loss, and can thus serve as sensitive suprathreshold tests for normal-hearing status (Mok et al., 2023). Accordingly, a speech-in-babble task was administered on a cohort of individuals with known hearing status (either audiometrically normal hearing, or known degree of hearing loss) and cutoff values were chosen based on the scores obtained on a speech-in-noise task such that the procedure yielded $> 80\%$ sensitivity to any hearing loss, and $> 95\%$ sensitivity to more-than-mild hearing loss (Mok et al.,

2023). Together with the headphone-check procedure, the speech-in-noise hearing screening helped narrow the pool of participants to the subjects who used 2-channel headphones, were of putative normal hearing, and in good compliance with the study instructions. N=200 subjects who passed all screening procedure proceeded with the main battery of the study. No training was provided except for a brief demo block for each task.

4.2.2 TFS Sensitivity Measurements

We previously established that binaural behavioral and electrophysiological (EEG) measurements of ITD sensitivity can reliably reflect individual differences in TFS processing (Borjigin et al., 2022). Therefore, in this study, we adopted the behavioral ITD detection and added a binaural version of the frequency-modulation (FM) detection. Importantly, our previous study also showed that the binaural metrics were effective in capturing individual TFS processing only if the contributions of extraneous "non-sensory" factors that are irrelevant to TFS processing, such as engagement, were measured and adjusted for (Borjigin et al., 2022). In the present study, we implemented a stand-alone measure that will also be influenced by extraneous non-sensory factors, but unaffected by individual TFS-processing fidelity. Specifically, we used an interaural level difference (ILD) discrimination task, which is also a binaural task but depends on level coding instead of TFS coding. The use of ILD discrimination as a surrogate measure not only helped mitigate non-sensory extraneous variability, but also likely enhanced the specificity of the ITD and binaural FM measures to TFS processing by removing individual variability in downstream "readout" processes that used binaural information. Further details of these measurements are provided below.

Interaural Time Difference (ITD) Discrimination. The stimulus consisted of two consecutive 400-ms-long, 500-Hz pure tones. The tones were delivered to both ears, but with a time delay in one ear (i.e., ITD). The leading ear was switched from the first to the second tone in the sequence, simulating a spatial "jump"s to the opposite side. ITDs in octave steps from 2 to 128 μs were presented in random order (8 repetitions for each step). The tone bursts were ramped on and off with a rise and fall time of 20 ms to exclude any audible transitions and to reduce reliance on onset ITDs. The gap between the two tone bursts was 200 ms. As with other tasks, participants were instructed to adjust the volume control on their devices to a comfortable loudness. The discrimination task was presented as a two-alternatives forced-choice (2AFC) task, where subjects were asked to report the direction of the "jump" between the two intervals (left-to-right or right-to-left) using a mouse click. A separate "demo" block was provided

before the experimental blocks to familiarize the subject with the task. The detection thresholds were quantified using a Bayesian approach (Kuss et al., 2005; Schütt et al., 2016). All other measurements in this study, including TFS-sensitivity, ILD sensitivity, and speech-in-noise tests (Figure 7), adopted the same method for estimating thresholds.

Binaural Frequency Modulation (FM) Detection. We employed a binaural FM detection task as a second metric for individual TFS sensitivity. Although monaural FM detection is conventionally used to probe TFS processing (Strelcyk and Dau, 2009; Parthasarathy et al., 2020), whether monaural FM detection can truly measure individual TFS processing fidelity is questionable (Whiteford et al., 2020; Borjigin et al., 2022). In contrast, binaural temporal processing has an unambiguous theoretical connection to TFS coding (Yin and Chan, 1990; Borjigin et al., 2022). The binaural FM detection measure implemented in the present study consisted of target and a reference stimuli arranged in a 2AFC task configuration. The stimuli in each interval were turned on and off with a rise and fall time of 5 ms to exclude audible transitions. The reference was a 500-ms, 500-Hz diotic pure tone. The target was a 2-Hz FM around 500 Hz with modulations going in opposite directions in two ears to introduce binaural timing cues. Low FM rates were chosen to minimize monaural envelope-based cues (Moore and Sek, 1996). FMs of magnitudes in octave steps from 0.1 to 3.2 Hz were presented in random order (8 repetitions for each step). No training was provided except for a brief "demo" block that was intended for orienting the participants before the formal testing.

Interaural Level Difference (ILD) Discrimination. ILD discrimination thresholds were measured as a reference measure to quantify the involvement of extraneous "non-sensory" factors, such as lapses in attention and variations in motivation. Stimuli consisted of two consecutive 4 kHz pure-tone bursts, a frequency where TFS phase locking is generally thought to be limited (Verschooten et al., 2019). Similar to the ITD task, the two intervals were lateralized to opposite sides through ILDs, simulating a spatial "jump" from one side to the other. ILDs in octave steps from 0.1 to 3.2 dB (8 repetitions for each step) were presented in random order. Subjects were asked to report the direction of the jump through a mouse-click response in a 2AFC setting.

4.2.3 Grouping of Participants

Participants were classified into two groups (Good vs. Poor sensitivity group), either based on TFS-sensitivity measures or the ILD measure (see Figures 2a and 2b in Section 2). A two-dimensional "k-

means" clustering algorithm was used for grouping based on the two TFS measures whereas a simple median split was used for ILD-based grouping (given that it was based on a single measure). Note that, before clustering, ILD sensitivity was "regressed-out" from the two TFS-sensitivity measures using a simple linear regression to emphasize individual differences in TFS processing and mitigate the effects of extraneous variables on the TFS measures. Similarly, the combination of the two TFS measures was regressed out from ILD measure to only retain individual differences in "non-TFS" factors to form ILD-based groups.

4.2.4 Measurements of Speech Perception in Noise

The key objective of this study was to evaluate the perceptual role of TFS coding for everyday hearing, with a focus on speech perception in noise. Accordingly, we designed speech-in-noise measures that probed several aspects of listening. The stimuli across all conditions consisted of target word with a carrier phrase and a masker. The masker was either a four-talker babble or a steady noise composed of an inharmonic complex of tones (Stone and Moore, 2014), described further below. The carrier phrase was in the same voice as the target word and said: "Please select the word ... ". The masker began after the onset of the target carrier phrase but before the target word to allow listeners to orient themselves to the target voice based on the unmasked portion of carrier phrase. A word-based test rather than a sentence-based test was chosen to minimize influence from factors such as individual differences in working memory, and ability to use linguistic context.

Participants were tested across 10 different target-masker conditions, as shown in Figure 7. Four conditions included four-talker babble as the masker and one used a non-speech, steady-noise masker. The babble masker conditions included pitch cues, spatial cues, both pitch and spatial cues, and no explicit cues at all (i.e., reference). The non-speech masker condition had a steady noise as the masker without any explicitly added cues. The remaining 5 conditions were similar but with an addition of room reverberation. Reverberation was added to investigate the role of TFS processing in reverberation because the presence of reverberation generally degrades hearing performance with individual differences often exaggerated (Ihlefeld and Shinn-Cunningham, 2011; Ruggles et al., 2011, 2012). The presentation order of these 10 different test conditions was randomized across trials. Further details about the stimulus manipulations used are provided below. For each condition, speech intelligibility was measured over a range of SNRs to estimate the speech reception threshold (SRT) defined as the SNR at which approximately

50% of the words are intelligible.

Pitch Cues. To control the available pitch cues for separating the target and masker, the audio recordings for all trials were first processed to remove inherent fundamental-frequency (F0) fluctuations (i.e., monotonized to the estimated pitch median). Then, the flattened F0 contours of each target sentence and each talker in the four-talker babble were transposed to a preset value, as shown in Figure 8. Specifically, the F0 of female target voice was set to 245 Hz, and that of the male target was set to 95 Hz. Among the talkers whose sentences were mixed to create the four-talker babble background, the male talkers' F0 values were set to 85, 90, 100, and 105 Hz, respectively, and the female talkers' F0 to 235, 240, 250, and 255 Hz, respectively. Note that the target and masker of the same sex were at similar pitch values but with a small difference to ensure that the participant could still distinguish the target from the masker but could only derive minimal masking release based on pitch. The pitch was flattened for all other stimulus configurations (i.e., reference, space, pitch+space, and non-speech noise masker) to maintain consistency throughout the study. Pitch-based masking release was estimated as the threshold SNR difference between the reference condition where the target and masker stimuli were composed of recordings from same-sex talkers and the "pitch" condition where there was a large F0 separation by virtue of the target and masker stimuli being composed of recordings from opposite-sex talkers.

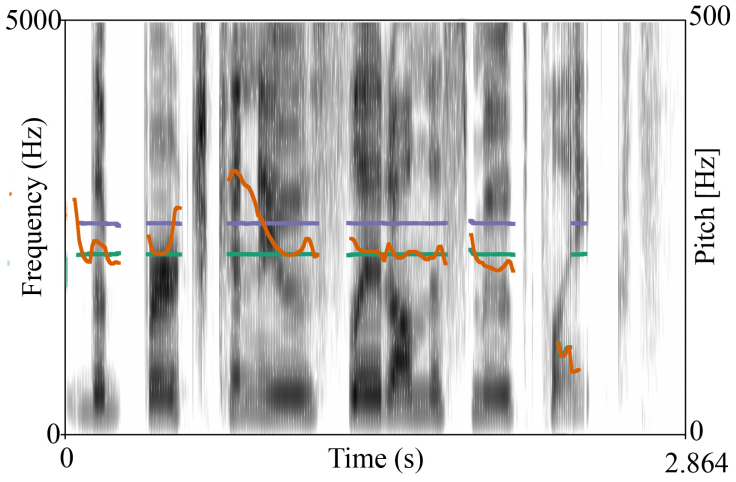


Figure 8. The spectrogram of a sentence: "The birch canoe slid on the smooth planks." The orange curve shows the estimated F0 contour with natural fluctuations; flattened F0 contour is shown in green; the flattened pitch contour that was transposed to a pre-set frequency (255 Hz in this example), is shown in purple.

Spatial Cues. Spatial cues help listeners better localize and attentively select a target of interest from background noise (Shinn-Cunningham, 2008). To simulate the perception of spatial separation using purely TFS-based cues, the polarity of the target signal in one ear was flipped while the masker signal was kept the same in both ears. This configuration is usually denoted as N_0S_π. The fully diotic condition

without this interaurral manipulation is referred to as N_0S_0. A lower SRT (i.e., better performance) is typically observed in the N_0S_π condition, with the difference in SRTs designated as the binaural masking level difference (BMLD, i.e., spatial release from masking) (Levitt and Rabiner, 1967). Studies have shown that spatial and pitch disparity between target and masker can each yield around 5 dB of masking release (Best et al., 2005; Oxenham and Simonson, 2009). Note that masking release is quantified as the threshold difference between the conditions with explicitly added beneficial cues and the corresponding reference condition without these explicit cues.

Steady Noise Masker and Reverberation. Performance in the presence of a steady noise masker was used to evaluate the role of TFS in providing release from so-called "informational masking" (Durlach et al., 2003). Accordingly, the steady noise masker was designed to have minimal intrinsic modulations and matched frequency content using the procedure described in Stone and Moore (2014). Owing to the lack of modulations (explicit and intrinsic), this masker represents a condition of maximal energetic masking while avoiding most sources of informational masking. Note that conventionally used noise maskers such as speech-spectrum stationary noise includes intrinsic modulations that can contribute to masking at more central levels of the auditory system (Culling and Stone, 2017; Hartmann and Pumplin, 1988; Kohlrausch et al., 1997; Brungart et al., 2006). Finally, to simulate listening under reverberation, the original audios that were recorded under anechoic conditions were convolved with binaural room impulse responses from a bar.

Speech Reception Threshold (SRT) Estimation To robustly estimate the mean and variance of the masking release based on different cues, SRTs for each speech-in-noise condition was estimated using a jackknife resampling procedure. Specifically, within each group (i.e., Good vs. Poor sensitivity groups), a leave-one-out procedure was used where each individual's data was removed and the SRTs was estimated by pooling the data from the remaining participants. Across individuals within a group, this procedure generated k jackknife samples for the SRT for each condition and masking release for each cue (where k is the number of individuals within the group). Following Efron and Stein (1981), group-level mean was estimated as the mean M across the jackknife samples, and the variance as the sample-variance V across the jackknife samples multipled by $(k-1)$. Note that grouping of participants were done either based on TFS sensitivity or based on ILD discrimination, as described in Section 4.2.2. The jackknife procedure avoids the need to fit psychometric curves for speech intelligibility as a function of SNR or estimate SRTs at the level of individual subject, and yet robustly estimate the variance in the SRTs (and masking release

values) across subjects within each group.

Reaction Time. SRT is one of the most common metrics for estimating an individual's ability to understand speech and is also a key component of clinical audiological assessment. However, intelligibility scores and thresholds only provide a limited view of an individual's overall listening outcome. Indeed, two listeners with comparable SRTs could nonetheless experience different levels of listening effort from having to engage cognitive resources to different degrees (Peelle, 2018; Alhanbali et al., 2019). To assess the role of cognitive effort beyond the threshold measures, the reaction time for each participant was determined by subtracting the time of the stimulus offset (or stimulus duration) from the recorded time of the mouse-click response. The same jackkniffe procedure as used for the SRT estimates was also used to estimate the mean and variance of response times. Trials with response times larger than 10 seconds were discarded, under the assumption that they were likely due to interruptions in participation rather than the engagement of cognitive processes to select a response. Response times were separately estimated for each participant grouping, and for each speech-in-noise condition.

4.2.5 Statistical Analyses

The primary analyses in the present study involved between-group comparisons of masking release or response times. Because the cohort size was large (N=200) and estimates of group mean and variance were derived using the jackknifing procedure, it was reasonable to assume that group-level estimates represented parameter estimates for normally distributed data. Accordingly, a simple two-tailed z-test was used for making inference. As described previously in section 4.2.4, among the ten speech-in-noise conditions, five simulated of speech-in-noise mixtures in anechoic environments, and the other five included room reverberation. To investigate the effects of reverberation, shown in Figure 6a, data from all five speech-in-noise configurations were combined using inverse variance pooling (Sinha et al., 2011; Hedges, 1982). For response time comparisons across groups, all ten conditions were pooled together.

References

Alhanbali, S., Dawes, P., Millman, R. E., and Munro, K. J. (2019). Measures of Listening Effort Are Multidimensional. *Ear and Hearing*, 40(5):1084–1097.

Apoux, F., Crouzet, O., and Lorenzi, C. (2001). Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: effects on identification performance and response times. *Hearing Research*, 153(1):123–131.

Arbogast, T. L., Mason, C. R., and Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, 112(5):2086–2098. Publisher: Acoustical Society of America.

Ardoint, M. and Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing Research*, 260(1):89–95.

Attneave, F. and Olson, R. K. (1971). Pitch as a Medium: A New Approach to Psychophysical Scaling. *The American Journal of Psychology*, 84(2):147–166. Publisher: University of Illinois Press.

Baden, T., Esposti, F., Nikolaev, A., and Lagnado, L. (2011). Spikes in Retinal Bipolar Cells Phase-Lock to Visual Stimuli with Millisecond Precision. *Current Biology*, 21(22):1859–1869.

Bernstein, J. G. and Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *The Journal of the Acoustical Society of America*, 113(6):3323–3334. Publisher: Acoustical Society of America.

Best, V., Ozmeral, E., Gallun, F. J., Sen, K., and Shinn-Cunningham, B. G. (2005). Spatial unmasking of birdsong in human listeners: Energetic and informational factors. *The Journal of the Acoustical Society of America*, 118(6):3766–3773.

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-Cunningham, B. G. (2015). Individual Differences Reveal Correlates of Hidden Hearing Deficits. *Journal of Neuroscience*, 35(5):2161–2172.

Bird, J. and Darwin, C. J. (1997). Effects of a difference in fundamental frequency in separating two sentences.

Borjigin, A., Hustedt-Mai, A. R., and Bharadwaj, H. M. (2022). Individualized Assays of Temporal Coding in the Ascending Human Auditory System. *eNeuro*, 9(2). Publisher: Society for Neuroscience Section: Research Article: New Research.

Brokx, J. P. L. and Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10(1):23–36.

Brughera, A., Dunai, L., and Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America*, 133(5):2839–2855.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, 120(6):4007–4018.

Churchill, T. H., Kan, A., Goupell, M. J., and Litovsky, R. Y. (2014). Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners. *The Journal of the Acoustical Society of America*, 136(3):1246–1256. Publisher: Acoustical Society of America.

Culling, J. F. and Stone, M. A. (2017). Energetic Masking and Masking Release. In *The Auditory System at the Cocktail Party*.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (1994). Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels. *Speech Communication*, 14(1):71–95.

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, 1(9):327–333.

Darwin, C. J. (2005). Pitch and Auditory Grouping. In Plack, C. J., Fay, R. R., Oxenham, A. J., and Popper, A. N., editors, *Pitch: Neural Coding and Perception*, Springer Handbook of Auditory Research, pages 278–305. Springer, New York, NY.

de Cheveigné, A. and Pressnitzer, D. (2006). The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction. *The Journal of the Acoustical Society of America*, 119(6):3908–3918. Publisher: Acoustical Society of America.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America*, 97(1):585–592. Publisher: Acoustical Society of America.

Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America*, 113(6):2984–2987.

Dye, R. H. (1990). The combination of interaural information across frequencies: Lateralization on the basis of interaural delay. *The Journal of the Acoustical Society of America*, 88(5):2159–2170.

Efron, B. and Stein, C. (1981). The Jackknife Estimate of Variance. *The Annals of Statistics*, 9(3):586–596. Publisher: Institute of Mathematical Statistics.

Freyman, R. L., Griffin, A. M., and Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *The Journal of the Acoustical Society of America*, 132(4):2514–2523. Publisher: Acoustical Society of America.

Gallun, F. J., Coco, L., Koerner, T. K., Larrea-Mancera, E. S. L. d., Molis, M. R., Eddins, D. A., and Seitz, A. R. (2022). Relating Suprathreshold Auditory Processing Abilities to Speech Understanding in Competition. *Brain Sciences*, 12(6):695. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Gallun, F. J., Diedesch, A. C., Kampel, S. D., and Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Frontiers in Neuroscience*, 7. Publisher: Frontiers.

Gallun, F. J., McMillan, G. P., Kampel, S. D., Jakien, K. M., Srinivasan, N. K., Stansell, M. M., and Gordon, S. Y. (2015). Verification of an automated headphone-based test of spatial release from masking. *Proceedings of Meetings on Acoustics*, 25(1):050001. Publisher: Acoustical Society of America.

Grose, J. H. and Mamo, S. K. (2012). Frequency modulation detection as a measure of temporal processing: Age-related monaural and binaural effects. *Hearing Research*, 294(1):49–54.

Hartmann, W. M. and Pumplin, J. (1988). Noise power fluctuations and the masking of sine signals. *The Journal of the Acoustical Society of America*, 83(6):2277–2289. Publisher: Acoustical Society of America.

Hasenstaub, A., Otte, S., Callaway, E., and Sejnowski, T. J. (2010). Metabolic cost as a unifying principle governing neuronal biophysics. *Proceedings of the National Academy of Sciences*, 107(27):12329–12334. Publisher: National Academy of Sciences Section: Biological Sciences.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2):490–499. Place: US Publisher: American Psychological Association.

Henning, G. B. (1983). Lateralization of low-frequency transients. *Hearing Research*, 9(2):153–172.

Hilbert, D. (1906). *Grundz?ge einer allgemeinen theorie der linearen integralgleichungen.* . Google-Books-ID: caARAwAAQBAJ.

Hopkins, K. and Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *The Journal of the Acoustical Society of America*, 125(1):442–446. Publisher: Acoustical Society of America.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *The Journal of the Acoustical Society of America*, 123(2):1140–1153. Publisher: Acoustical Society of America.

Houtsma, A. J. M. and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *The Journal of the Acoustical Society of America*, 87(1):304–310. Publisher: Acoustical Society of America.

Humes, L. E. (2021). Factors Underlying Individual Differences in Speech-Recognition Threshold (SRT) in Noise Among Older Adults. *Frontiers in Aging Neuroscience*, 13.

Ihlefeld, A. and Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a divided speech identification task. *The Journal of the Acoustical Society of America*, 123(6):4380–4392. Publisher: Acoustical Society of America.

Ihlefeld, A. and Shinn-Cunningham, B. G. (2011). Effect of source spectrum on sound localization in an everyday reverberant room. *The Journal of the Acoustical Society of America*, 130(1):324–333.

Jakien Kasey M. and Gallun Frederick J. (2018). Normative Data for a Rapid, Automated Test of Spatial Release From Masking. *American Journal of Audiology*, 27(4):529–538.

Jakien Kasey M., Kampel Sean D., Stansell Meghan M., and Gallun Frederick J. (2017). Validating a Rapid, Automated Test of Spatial Release From Masking. *American Journal of Audiology*, 26(4):507–518. Publisher: American Speech-Language-Hearing Association.

Javel, E. and Mott, J. B. (1988). Physiological and psychophysical correlates of temporal processes in hearing. *Hearing Research*, 34(3):275–294.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *The Journal of the Acoustical Society of America*, 68(4):1115–1122. Publisher: Acoustical Society of America.

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural Processing of Amplitude-Modulated Sounds. *Physiological Reviews*, 84(2):541–577. Publisher: American Physiological Society.

Joris, P. X. and Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *The Journal of the Acoustical Society of America*, 91(1):215–232. Publisher: Acoustical Society of America.

Kidd, G. and Colburn, H. S. (2017). Informational Masking in Speech Recognition. In *The Auditory System at the Cocktail Party*.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (1997). Detection of Tones in Low-noise Noise: Further Evidence for the Role of Envelope Fluctuations. Library Catalog: www.ingentaconnect.com Publisher: S. Hirzel Verlag.

Kuss, M., Jäkel, F., and Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, 5(5):8–8. Publisher: The Association for Research in Vision and Ophthalmology.

Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, 1(1):36–41. Number: 1 Publisher: Nature Publishing Group.

Lelo de Larrea-Mancera, E. S., Stavropoulos, T., Hoover, E. C., Eddins, D. A., Gallun, F. J., and Seitz, A. R. (2020). Portable Automated Rapid Testing (PART) for auditory assessment: Validation in a young adult normal-hearing population. *The Journal of the Acoustical Society of America*, 148(4):1831–1851. Publisher: Acoustical Society of America.

Levitt, H. and Rabiner, L. R. (1967). Binaural Release From Masking for Speech and Gain in Intelligibility. *The Journal of the Acoustical Society of America*, 42(3):601–608. Publisher: Acoustical Society of America.

Licklider, J. C. R. (1948). The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise. *The Journal of the Acoustical Society of America*, 20(2):150–159. Publisher: Acoustical Society of America.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., and Moore, B. C. J. (2009). Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *The Journal of the Acoustical Society of America*, 125(1):27–30. Publisher: Acoustical Society of America.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103(49):18866–18869.

Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., and Lawlor, B. A. (2018). Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis. *JAMA Otolaryngology–Head & Neck Surgery*, 144(2):115–126.

Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111(5):2219–2236. Publisher: Acoustical Society of America.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). Individual Differences Reveal the Basis of Consonance. *Current Biology*, 20(11):1035–1041.

Mok, B. A., Viswanathan, V., Borjigin, A., Singh, R., Kafi, H., and Bharadwaj, H. M. (2023). Web-based psychoacoustics: Hearing screening, infrastructure, and validation. *Behavior Research Methods*.

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America*, 54(3):610–619. Publisher: Acoustical Society of America.

Moore, B. C. J. and Sek, A. (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *The Journal of the Acoustical Society of America*, 100(4):2320–2331.

Nuesse, T., Steenken, R., Neher, T., and Holube, I. (2018). Exploring the Link Between Cognitive Abilities and Speech Recognition in the Elderly Under Different Listening Conditions. *Frontiers in Psychology*, 9.

Ohm, G. S. (1843). Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Annalen der Physik*, 135(8):513–565. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18431350802.

Oxenham, A. J. (2012). Pitch Perception. *Journal of Neuroscience*, 32(39):13335–13338. Publisher: Society for Neuroscience Section: Mini-Reviews.

Oxenham, A. J. (2013). Revisiting place and temporal theories of pitch. *Acoustical Science and Technology*, 34(6):388–396.

Oxenham, A. J. (2018). How We Hear: The Perception and Neural Coding of Sound. *Annual Review of Psychology*, 69:27–50.

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch.

Oxenham, A. J. and Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *The Journal of the Acoustical Society of America*, 125(1):457–468.

Palmer, A. R. and Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, 24(1):1–15.

Palomäki, K. J., Brown, G. J., and Wang, D. (2004). A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation. *Speech Communication*, 43(4):361–378.

Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., and 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. Publisher: eLife Sciences Publications, Ltd.

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, 39(2):204–214.

Qin, M. K. and Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, 114(1):446–454. Publisher: Acoustical Society of America.

Richards, D. G. and Wiley, R. H. (1980). Reverberations and Amplitude Fluctuations in the Propagation of Sound in a Forest: Implications for Animal Communication. *The American Naturalist*, 115(3):381–399. Publisher: The University of Chicago Press.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, 30(4):769–793. Publisher: American Physiological Society.

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences*, 108(37):15516–15521.

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2012). Why Middle-Aged Listeners Have Trouble Hearing in Everyday Settings. *Current Biology*, 22(15):1417–1422.

Saddler, M. R., Gonzalez, R., and McDermott, J. H. (2021). Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception. *Nature Communications*, 12(1):7278. Number: 1 Publisher: Nature Publishing Group.

Saddler, M. R. and McDermott, J. H. (2022). The role of temporal coding in everyday hearing: evidence from deep neural networks.

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, 122:105–123.

Seebeck, A. (1841). Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Annalen der Physik*, 129(7):417–436. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18411290702.

Shadlen, M. N. and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579.

Shinn-Cunningham, B., Best, V., and Lee, A. K. C. (2017). Auditory Object Formation and Selection. In *The Auditory System at the Cocktail Party*.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5):182–186.

Sinha, B. K., Hartung, J., and Knapp, G. (2011). *Statistical Meta-Analysis with Applications*. John Wiley & Sons. Google-Books-ID: JEoNB_2NONQC.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876):87.

Srinivasan, N. K., Jakien, K. M., and Gallun, F. J. (2016). Release from masking for small spatial separations: Effects of age and hearing loss. *The Journal of the Acoustical Society of America*, 140(1):EL73–EL78. Publisher: Acoustical Society of America.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F.-G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *The Journal of the Acoustical Society of America*, 122(2):1069–1078. Publisher: Acoustical Society of America.

Stone, M. A. and Moore, B. C. J. (2014). On the near non-existence of "pure" energetic masking release for speech. *The Journal of the Acoustical Society of America*, 135(4):1967–1977.

Strelcyk, O. and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, 125(5):3328–3345.

Summers Van and Leek Marjorie R. (1998). F0 Processing and the Seperation of Competing Speech Signals by Listeners With Normal Hearing and With Hearing Loss. *Journal of Speech, Language, and Hearing Research*, 41(6):1294–1306.

Swaminathan, J. and Heinz, M. G. (2012). Psychophysiological Analyses Demonstrate the Importance of Neural Envelope Coding for Speech Perception in Noise. *Journal of Neuroscience*, 32(5):1747–1756.

Tollin, D. J., Koka, K., and Tsai, J. J. (2008). Interaural Level Difference Discrimination Thresholds for Single Neurons in the Lateral Superior Olive. *Journal of Neuroscience*, 28(19):4848–4860. Publisher: Society for Neuroscience Section: Articles.

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G., and Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing Research*, 377:109–121.

Viswanathan, V., Bharadwaj, H. M., Shinn-Cunningham, B. G., and Heinz, M. G. (2021a). Modulation masking and fine structure shape neural envelope coding to predict speech intelligibility across diverse listening conditions. *The Journal of the Acoustical Society of America*, 150(3):2230–2244. Publisher: Acoustical Society of America.

Viswanathan, V., Shinn-Cunningham, B. G., and Heinz, M. G. (2021b). Temporal fine structure influences voicing confusions for consonant identification in multi-talker babble. *The Journal of the Acoustical Society of America*, 150(4):2664–2676. Publisher: Acoustical Society of America.

Wallace, M. N., Rutkowski, R. G., Shackleton, T. M., and Palmer, A. R. (2000). Phase-locked responses to pure tones in guinea pig auditory cortex. *NeuroReport*, 11(18):3989–3993.

Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (2017). Assessing the Role of Place and Timing Cues in Coding Frequency and Amplitude Modulation as a Function of Age. *Journal of the Association for Research in Otolaryngology*, 18(4):619–633.

Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (2020). The role of cochlear place coding in the perception of frequency modulation. *eLife*, 9:e58468. Publisher: eLife Sciences Publications, Ltd.

Whiteford, K. L. and Oxenham, A. J. (2015). Using individual differences to test the role of temporal and place cues in coding frequency modulation. *The Journal of the Acoustical Society of America*, 138(5):3093–3104.

Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–1661. Publisher: Acoustical Society of America.

Woods, K. J. P., Siegel, M. H., Traer, J., and McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79(7):2064–2072.

Yin, T. C. and Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology*, 64(2):465–488.