Individual differences elucidate the perceptual benefits associated with robust temporal fine-structure processing

Agudemu Borjigin[a,b,1] and Hari M. Bharadwaj[c]

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received October 3, 2023; accepted November 19, 2024

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. TFS sensitivity was assessed through detection of interaural time/phase differences, while speech perception was evaluated by word identification under noise interference. Results suggest that greater TFS sensitivity is not associated with greater masking release from fundamental-frequency or spatial cues but appears to contribute to resilience against the effects of reverberation. We also found that greater TFS sensitivity is associated with faster response times, indicating reduced listening effort. These findings highlight the perceptual significance of TFS coding for everyday hearing.

temporal fine structure | speech perception in noise | individual differences | reverberation | listening effort

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) (1). Temporal information in any sound is composed of two components: rapid variations in phase—the temporal fine structure (TFS), and slower amplitude variations—the temporal envelope (2). Neurons in the auditory system can robustly track both TFS (3) and envelope (4) through phase-locked firing. Strikingly, neural phase locking to TFS extends at least up to 1,400 Hz in the peripheral auditory system (5–7), a feat unmatched by other sensory modalities. In comparison, phase-locked information in the visual and somatosensory systems extends only to about 50 Hz (8, 9). However, this uniquely high upper-frequency limit of phase locking in the auditory system only exists at the peripheral level (5, 10). Along the ascending pathway, the phased-locked temporal code appears to be progressively transformed into a rate-place representation (11). It seems that the auditory system initially invests heavily in this exquisite and metabolically expensive (12, 13) phase-locked temporal code 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 (5, 14–20).

Psychophysical experiments in quiet sound booths suggest that TFS may play a role in sound localization (21, 22) and pitch perception (through fundamental-frequency or F0 cues) (23–25). Both spatial and F0 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 (26–34). 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 (17, 35). Furthermore, TFS-based spatial cues are more susceptible to corruption from reverberation than envelope-based spatial cues (36, 37) by virtue of being perceptually dominant primarily at low-frequencies up to about 1,400 Hz (7, 38), where reverberation is more pronounced (37). Thus, despite many decades of intensive research, whether phase-locked temporal coding of TFS would

Significance

Neural phase-locking to fast temporal fluctuations in sounds—temporal fine structure (TFS) in particular—is a unique mechanism by which acoustic information is encoded by the auditory system. However, despite decades of intensive research, the perceptual relevance of this metabolically expensive mechanism, especially in challenging listening settings, is debated. Here, we leveraged an individual-difference approach to circumnavigate the limitations plaguing conventional approaches and found that robust TFS sensitivity is associated with greater resilience against the effects of reverberation and is associated with reduced listening effort for speech understanding in noise.

Author affiliations: [a]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47907; [b]Waisman Center, University of Wisconsin, Madison, WI 53705; and [c]Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA 15213

Author contributions: A.B. and H.M.B. designed research; A.B. performed research; A.B. and H.M.B. contributed new reagents/analytic tools; A.B. analyzed data; and A.B. and H.M.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: dagu@wisc.edu.

Published January 3, 2025.

introduce additional masking-release benefits in reverberant listening conditions remains unclear.

A key challenge to understanding the perceptual role of TFS phase locking is that subband vocoding, which is the most common technique employed to investigate this question, is inherently limited (21, 39–43). Vocoding has been used to acoustically dissociate TFS from envelope by creating stimuli with a constant envelope (i.e., subband amplitude) while manipulating the TFS (i.e., subband 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 (16, 44, 45). Recent approaches to investigate the perceptual significance of TFS coding have leveraged deep neural networks (DNN) and evaluated how the performance of DNNs trained on a range of tasks is affected when temporal coding is manipulated in the models (19, 46). While this avoids the pitfalls of stimulus manipulation approaches, whether DNN predictions fully correspond to human perception continues to be an area of intense research. Some studies, such as Hopkins et al. (39) and Smith et al. (21) have employed stimuli that combine envelope and TFS information from distinct speech utterances 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 vs. 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, minimally processed stimuli. The individual differences approach has been successfully used to address other fundamental questions in the neuroscience of audition (47–50). At the time of this study, the individual-difference approach has not been used to explore the role of TFS for speech-in-noise perception, as robust individual-level measures were only recently established by comparing both EEG and behavioral measures of TFS coding (51). Since, however, Vinay and Moore (52) have employed the individual differences approach to examine the perceptual role of TFS, but only for the simple task of frequency discrimination around 2 kHz.

Here, we leveraged individualized TFS processing measures developed in our previous work, and adapted them for remote testing to circumnavigate the COVID-19-related restrictions (53). We hypothesized that TFS plays an important role in everyday hearing. 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 ten different listening conditions, representing many important aspects of everyday listening where TFS phase locking has conventionally been thought to play a role. We predicted that individuals with better TFS sensitivity would benefit more from F0 and spatial cues in noisy listening settings because of the hypothesized role of TFS in pitch perception and sound localization (21–25). Because reverberation impairs TFS-based spatial cues (36) and spatial selective attention (54), we predicted that individuals with better TFS sensitivity would be less affected by reverberation.

Last, we hypothesized that individuals with better TFS sensitivity would expend less listening effort and show more release from informational masking. Informational masking occurs when listeners fail to segregate or select the target sound components in the mixture despite minimal direct spectrotemporal overlap between the target and maskers. Both listening effort and listening under conditions of informational masking have been linked to a number of central auditory and cognitive processes (55–57); the availability of robust TFS cues is thought to be beneficial to these processes (58–60). There is now considerable literature suggesting that performance scores alone do not capture the widely varying degree of cognitive effort that different participants have to put in to reach the same score. Response times have thus found increasing use in the "listening effort" literature as a measure that is sensitive to differences in the cognitive burden experienced by different participants (61, 62). Accordingly, we measured response times in addition to speech-in-noise scores. The automated and parallel nature of the online measurements allowed us to rapidly collect data from a large cohort of 200 participants, affirming the promise and advantages of online behavioral psychoacoustical studies (50, 53). Fig. 1 illustrates the design of this study. The results revealed that better TFS processing, although not associated with greater masking release [confirming the results from Füllgrabe et al. (63)], provided resilience against reverberation, and lessened listening effort. Given that reverberation is a 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.

Results

Binaural Temporal Sensitivity Measures Captured Individual Differences in TFS Processing Fidelity. Fig. 2*A* is a scatter plot of the individual differences that we observed for our two binaural temporal sensitivity measures—interaural time difference

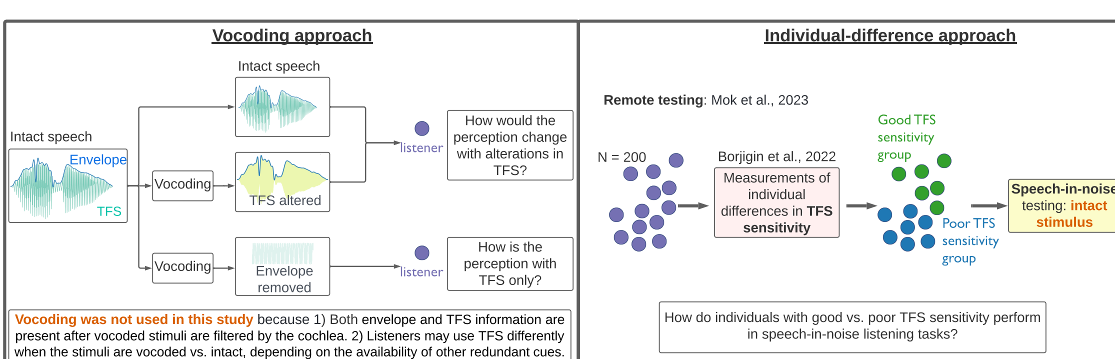

Fig. 1. Contrasting the conventional vocoding approach (*Left*) for studying TFS with the individual-difference approach adopted in this study (*Right*).

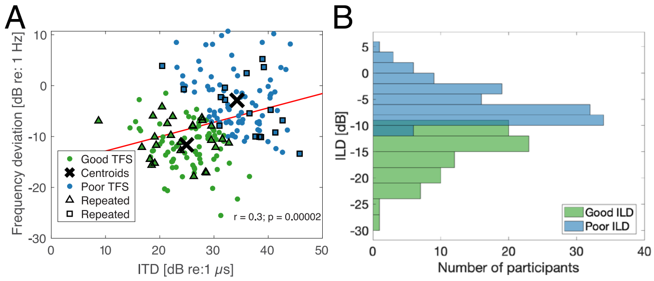

Fig. 2. (*A*) An illustration of cluster assignment based on the combination of ITD and binaural FM thresholds. Note that the marked participants (triangles and squares) returned for replication measurements (see Fig. 7 showing replication data). (*B*) Group assignment based on ILD sensitivity. Note that the ITD, ILD, binaural FM sensitivity values displayed are the residuals after regression.

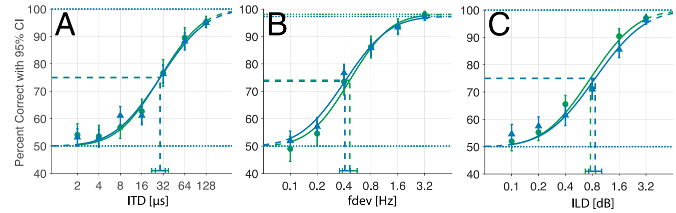

Fig. 3. (*A* and *B*) Psychometric curves for ITD and binaural FM measurement, respectively, for two groups based on ILD sensitivity; (*C*) Psychometric curves for ILD measurement for two groups based on TFS sensitivity. Error bars represent within-group SEM. Although single-point threshold estimates were used for grouping based on TFS or ILD measures, the full psychometric functions overlap in all three panels, suggesting that groupings based on TFS measures and groupings based on ILD are orthogonal with little common variance.

(ITD) discrimination and binaural frequency modulation (FM) detection (FM of opposite phase in the two ears). Metrics of individual TFS sensitivity commonly used in the literature are prone to the impact of extraneous "nonsensory" variables (51) such as attention and motivation. Here, interaural level difference (ILD) discrimination was used as a surrogate measure to control for nonsensory factors as well as aspects of binaural hearing unrelated to the basic TFS code. These TFS metrics were accordingly "adjusted" by regressing out the ILD sensitivity scores from each measure. The individual differences in these adjusted TFS metrics are more likely driven by true individual differences in TFS processing (see *Materials and Methods* for further details). Individual ILD sensitivity data are shown in Fig. 2*B*, which also indicates substantial individual variability. Note that in Fig. 2*A*, the TFS metrics are shown after regressing out the ILD measure, and vice versa.

The adjusted binaural FM detection and ITD discrimination measures were significantly correlated ($r = 0.3, P < 0.0001$) indicating a common underlying source of variance attributable to TFS processing. Accordingly, participants were divided into two groups by a clustering algorithm based on these two measures into "Good-" vs. "Poor-TFS" groups. Importantly, when the ILD data were plotted for these two TFS groups, the full psychometric curves overlapped (Fig. 3*C*), demonstrating that the elimination of common, nonsensory variance from TFS measures was successful. Note that the psychometric curves were constructed from data that were not adjusted. The construction of groups based on common variance across the TFS measures after eliminating common variance with ILD sensitivity ensures that the grouping in Fig. 2*A* is mainly based on individuals' TFS sensitivity, rather than other unrelated factors. Note also that there was no significant difference in age between two groups (Good TFS group: mean age of 30.4 y with an SD of 7.7 y; Poor TFS group: mean age of 32.1 y with an std of 8.4 y). As an additional control, individuals were also grouped based on their ILD discrimination thresholds, as shown in Fig. 2*B*. For this alternative grouping, the psychometric curves for TFS measures fully overlap (Fig. 3 *A* and *B*), consistent with the notion that the grouping in Fig. 2*B* captures "non-TFS" variability instead of TFS sensitivity. This alternative non-TFS regrouping of participants is used as a control in the experiments probing the association between TFS processing and speech-in-noise outcomes.

The web-based measurements in the present study produced data that were comparable to the data not only from our previous in-person study but also from other labs. Fig. 4 shows comparisons for FM detection and ITD discrimination measurements across studies. The *Left* panel compares online measurement of binaural FM detection with in-person results from refs. 64–67. The right panel includes a sample of in-person studies that measured ITD discrimination (7, 51, 68, 69). Ref. 51 is our previous in-person study. These results further validate our choice of TFS sensitivity measures.

Better TFS Sensitivity Is not Associated with 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), Fig. 5*A* depicts the masking release. Masking release refers to improved noise tolerance associated with the following cues: F0 difference between the target and background speakers, spatial separation between the target and maskers, combination of F0 and spatial cues, and finally when the background noise was nonspeech stationary noise instead of speech babble. The masking release effects observed in this study are consistent with those reported in previous research: 1) With F0 separation, the participants could more easily identify the target compared to when the target and background had similar F0 (i.e., the reference condition). This F0-based masking release was about 5 dB (Good TFS group: mean = 4.8, std = 0.5; Poor TFS group: mean = 5.3, std = 0.4), which matches previous reports from refs. 32–34. 2) The masking release was around 3 dB when the target and background were spatially separated (Good TFS group: mean = 2.9, std = 0.4; Poor TFS group:

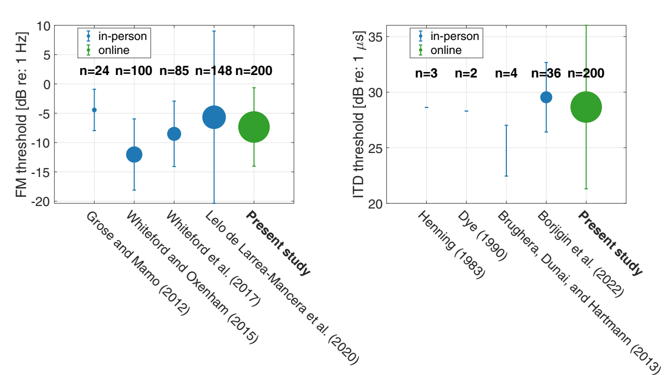

Fig. 4. A sample of published reports of binaural FM detection thresholds (*Left*) and ITD discrimination threshold (*Right*) for comparison with the present study. Error bars represent ± 1 std. The std of ITD detection thresholds could not be determined for refs. 68 and 69. The size of the dot represents the number of participants.

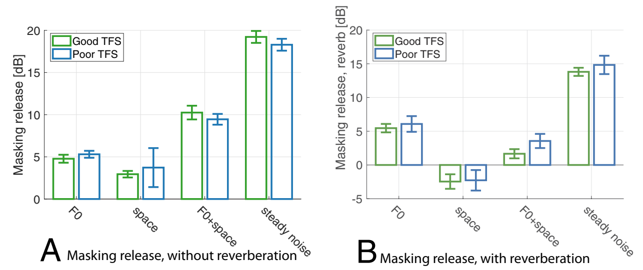

Fig. 5. Masking release across conditions. The height of the bars represents the mean, error bars represent ±1 std. Masking release was calculated by subtracting the SRT in each condition from that for the reference condition. Note that the reference condition in (*A*) does not have reverberation, whereas the reference condition in (*B*) contains reverberation. A positive masking release means that the SRT was lower/better than that for the reference condition.

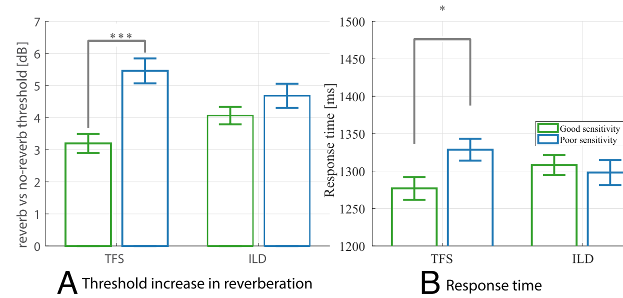

Fig. 6. (*A*) Increase in SRT due to reverberation for each group of participants. (*B*) Average response times. Data were pooled across all conditions shown in Figure 5. All error bars represent estimated SEM. Significance stars: 0.05 $>$ * $\geq$ 0.01, 0.001 $>$ *** (corrected for multiple comparisons using FDR procedures).

mean = 3.7, std = 2.3), which again aligns with earlier reports (26–31). 3) When both F0 and spatial cues were available, the masking-release benefits appeared to be cumulative, totaling about 10 dB as demonstrated in the "F0 + space" condition (Good TFS group: mean = 10.3, std = 0.8; Poor TFS group: mean = 9.5, std = 0.6). Indeed, it has previously been shown that F0 differences aid participants in spatially separating competing sounds (70). 4) A masking release of about 19 dB was observed when the background noise was switched from 4-talker babble to nonspeech stationary noise, as shown in the "steady noise" condition (Good TFS group: mean = 19.2, std = 0.7; Poor TFS group: mean = 18, 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 and background, which is often referred to as informational masking (56, 60, 71). The consistency of these results with prior literature confirms the viability of the online testing platform in reproducing in-person measurements.

Fig. 5*B* illustrates masking release for the same four conditions as in Fig. 5*A*, except for the addition of reverberation in all conditions. Note that the reference condition (i.e., babble speech with no F0 or spatial cues) also contained reverberation. Reverberation generally reduced the masking-release benefit, except for the F0-only condition. This is consistent with previous studies showing that reverberation has a smaller impact on the use of monaural cues (54, 72, 73), while spatial hearing is subject to substantial degradation (36, 54, 74).

Fig. 5 *A* and *B* demonstrate similar masking release for participants divided into two groups based on their TFS sensitivity. In both nonreverberant (Fig. 5*A*) and reverberant conditions (Fig. 5*B*), 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 envelope-based cues are also available (17, 35, 63, 75).

Better TFS Processing Is Associated with Resilience to the Effects of Reverberation and Reduced Listening Effort for Speech Perception in Noise. To illustrate the advantage associated with robust TFS processing for listening under reverberation, the threshold increase from nonreverberant to reverberant conditions is shown by the height of the bars in Fig. 6*A*. The group with poor TFS sensitivity (mean = 5.5, std = 0.4) showed a greater threshold increase in reverberant settings than their good-TFS counterparts (mean = 3.2, std = 0.3) (Fig. 6 *A*, *Left*; $z = 4.6, P = 0.2e-4$). When the participants were divided based on their ILD sensitivity, there was no significant group difference, indicating an important role for TFS processing. This result suggests that better TFS sensitivity can mitigate the negative impact of reverberation, which is a common source of signal degradation in everyday listening environments.

It is well known that behavioral measures of performance may not reveal important differences in the cognitive effort expended by participants in achieving a given level of performance (76, 77). To investigate whether robust TFS sensitivity is associated with less effortful listening, we examined response times, a measure commonly utilized in the literature for assessing listening effort (61, 62, 78, 79). The response times are indicated by the height of the bars in Fig. 6*B*. The absolute values of the response times are consistent with prior literature (80). When the participants were divided into two groups based on their TFS sensitivity, the Good-TFS group (mean = 1,277, std = 15.2) exhibited significantly shorter reaction times than the Poor-TFS group (mean = 1,328, std = 14.6) (Fig. 6 *B*, *Left*; $z = -2.5, P = 0.035$), consistent with reduced listening effort for the former. When the participants were regrouped based on non-TFS characteristics (i.e., ILD sensitivity), there was no significant difference between the two groups (Fig. 6 *B*, *Right*). Taken together, these results show that robust TFS sensitivity is associated with shorter reaction times. Both of these results, i.e., the smaller decrement in performance under reverberation and smaller overall response times in the good TFS group, remain significant after correcting for multiple comparison [10 comparisons across Figs. 5 *A* and *B* and 6 *A* and *B* using false discovery rate (FDR) procedures (81) at a 5% FDR level].

Replication Experiments in Both a Subset of the Original Participants (n = 44) and a New, Independent Sample (n = 104) Corroborate the Main Findings. In response to comments by an anonymous reviewer, we reached out to all 200 individuals who participated in 2020. Given the intermission of 3+ y, there was substantial attrition. Forty four participants responded and completed the replication measurements. Given the reduced power, the replication experiments on reinvited participants were more narrowly focused to test the main claims from the original study. Specifically, the measurements included the ITD and binaural FM threshold measures which form the basis for our grouping, ILD thresholds as a grouping control, and speech-in-noise measurements in anechoic and reverberant settings. Because the primary goal was to test whether the benefits associated with robust TFS processing in reverberation was replicable, we only included the speech-in-noise measures in the reference and F0-cue conditions. Despite a gap of more than three years, we observed statistically significant correlations between the original and repeated TFS-sensitivity measurements (ITD and binaural FM measurements, Fig. 7 *A*2 and *A*3). With the same grouping method being applied to the replication dataset for TFS-sensitivity measures (Fig. 7 *B*1), we see similar results as in Fig. 6: smaller increase in SRT due to reverberation (Fig. 7 *B*2) and shorter response times (Fig. 7 *B*3) overall for the Good-TFS group. When only the top and bottom 25% of the replication sample were chosen for grouping (Fig. 7 *C*1), to increase the group differences in TFS sensitivity and thus statistical power, the corresponding differences in the reverberation effects and response times also increased (Fig. 7 *C*2 and *C*3). Although not the focus of the replication study and not depicted in Fig. 7, note that F0-based masking release was not significantly different between groups (for groups in Fig. 7 *B*1 and *C*1), consistent with the original results from Fig. 5. We also conducted a second set of replication experiments in a newly screened, independent sample. In this new study, to maximize efficiency while retaining statistical power, we measured ITD and FM threshold from participants until we obtained a sample of $n > 100$ (specifically, $n = 104$) where each included participant fell in the same half (i.e., top or bottom half) by both the TFS measures. Participants who were in the top half by one measure and bottom half by the other were excluded. This modified screening approach was chosen a priori with the expectation that by increasing intersubject variance, we would gain greater statistical power despite using a smaller sample compared to the original experiments. All other procedures matched the original study. The results from this independent sample are shown in Fig. 7 *D*1–*D*3. As in the original experiment, in addition to the ILD control, we measured speech-in-noise scores in the reference, F0, and spatial-separations conditions in both anechoic and reverberation settings). Consistent with the original study, the good TFS group showed a smaller effect of reverberation, and

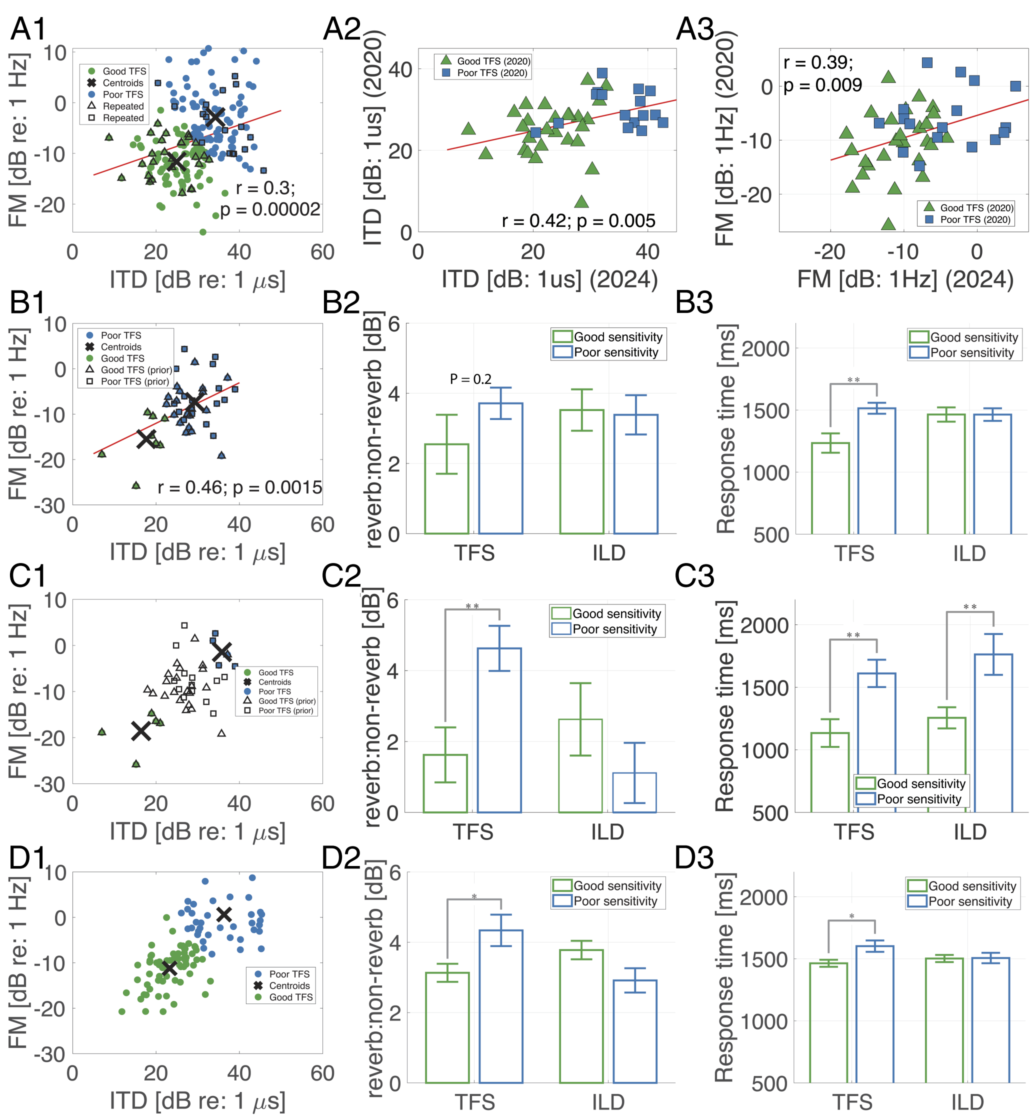

Fig. 7. Summary of the replication results obtained in 2024 from a subset of the original cohort (top three rows, $n = 44$) and an independent sample (bottom row, $n = 104$). All *P*-values shown are after adjustments for multiple comparisons using the FDR procedure (81). (*A*1–*A*3) There are statistically significant correlations between the original and replication data for our measures of TFS sensitivity; (*B*1–*B*3) the Good TFS-sensitivity group, based on replication measures, showed a smaller increase in SRTs due to reverberation and shorter response times overall. (*C*1–*C*3) Group differences in the reverberation effects and response times increased when we subselected participants to maximize group difference in TFS sensitivity. (*D*1–*D*3) We observed similar results from 104 new, independently sampled participants. Significance stars: $0.05 >^* \geq 0.01$, $0.01 > ^{**} \geq 0.001$.

shorter reaction times for speech-in-noise overall. Furthermore, although not depicted in Fig. 7, there were no differences in F0-based masking release or spatial release from masking for the groupings in Fig. 7*D*1, consistent with the original study. Taken together, the new data from both replication experiments corroborate both key findings from the original study and provided further credence to the notion that binaural measures can robustly capture individual differences in TFS processing.

Discussion

No greater spatial release from masking was observed for the Good-TFS group despite the theoretical connection between TFS phase locking and binaural temporal processing (21, 22) (Fig. 5*A*). 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 find a group difference 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. Similar to our finding, Füllgrabe et al. (63) did not observe an age effect on spatial release from masking, which might have been limited by a smaller age effect on TFS sensitivity from their typical hearing older participants. Another plausible reason could be that the spatial cue in this study was large (i.e., $S_{\pi}N_0$ vs. S_0N_0). There might have been a group difference for a small ITD between target and masker. Finally, the use of ILD discrimination as a reference for non-TFS factors could also have contributed to the lack of group difference in spatial release from masking. ILDs also activate binaural circuits, although ILD-based binaural processing does not rely on TFS phase locking (82). 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 (7, 21, 22, 38, 83, 84), 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 F0-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 y (85, 86). Humans perceive low-frequency periodic sounds as having a stronger pitch than high-frequency sounds (23–25). Frequency discrimination threshold, expressed as $\Delta F/F$, increases with increasing frequency from 2 to 8 kHz, plateauing above 8 kHz (87–89), which aligns with the low-pass characteristic of TFS phase locking in the auditory nerve (90, 91). Deficits in TFS coding have been invoked to explain speech perception deficits in fluctuating noise (41), where target-masker F0 differences are thought to play a role (92, 93). 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 (14, 94). At first glance, lack of a difference in F0-based masking release between the Good- and Poor-TFS groups appears to support place-code based explanations of pitch phenomena. However, the absence of group differences in F0-based masking release can be attributed to the same reasons discussed earlier for spatial release from masking—that the individual differences in TFS sensitivity among typical-hearing participants may have been too small to produce a significant group difference for F0-based masking release.

The "steady noise" condition used in the present study (Fig. 5*A*) was designed to minimize modulation masking (interference from modulations in the maskers) so that energetic masking would be dominant (95) (*Materials and Methods*). In contrast, the 4-talker babble masker in the reference condition contained many sources of modulations and informational masking (e.g., modulation masking, phonetic/lexical/semantic content) in addition to energetic masking (96). The improvement of almost 20 dB in SRTs from the reference to the steady noise condition [consistent with Arbogast et al. (71)] points to the dominant role of informational masking in everyday listening (97). 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 (55, 56). TFS-based processing is thought to play an important role for scene segregation and attentive selection (58–60). Although our results show similar release from informational masking across the two TFS-sensitivity groups (Fig. 5*A*), the group with better TFS sensitivity had a significantly shorter response time than the poorer TFS group (Fig. 6*B*). Our results, therefore, affirm the contribution of TFS coding to 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 (98–100).

Finally, we explored the correlation between TFS processing and listening in a reverberant environment. The SRTs were considerably worsened by the presence of reverberation (Fig. 6*A*). More importantly, the group with poor TFS sensitivity was affected significantly more than their good-TFS counterparts, indicating a possible role of TFS processing in resisting the deleterious effects of reverberation. Reverberation impairs TFS-based spatial cues (36) and spatial selective attention (54). Thus, our findings suggest that stronger TFS coding may ameliorate reverberation's detrimental effects on speech perception in noise.

These observations, together with the fact that most cochlear implants (CIs) do not convey TFS also help explain the effortful listening experience of CI users, especially in the presence of reverberation. The findings also suggest that evaluation of TFS processing may complement conventional assessments used in audiology clinics to help characterize speech perception deficits in background noise (54, 101, 102). Although the combined use of ITD, binaural FM, and ILD measures shows potential for capturing individual differences in TFS sensitivity, further validation and refinement is needed before they can be feasibly applied to clinical settings. Finally, our results also affirm the promise of using web-based psychoacoustics to conduct large-scale experiments (50, 53). Automated data collection facilitates the rapid acquisition of data from a large participant cohort over a short time frame (several days), providing a substantial advantage over traditional in-person psychoacoustic testing.

Finally, whether the perceptual benefits associated with better TFS sensitivity directly derive from the TFS code, or whether both derive from other common physiological factors, cannot be ascertained in this study. Although the contribution of nonsensory variables such as motivation and attention was mitigated by using the ILD metric as a control (51), there may be factors that preferentially affect the TFS code while also affecting speech in noise through mechanisms distinct from TFS processing. One such candidate mechanism is cochlear neural degeneration, which is hypothesized to affect temporal coding (48), and can also trigger central auditory changes which in turn can impair listening in the presence of background noise (103, 104).

Materials and Methods

Participants. Two hundred participants were recruited anonymously from Prolific.co in the original study [20 to 55 y old (mean = 31, std = 8); 93 females, 102 males, and 5 not reported]. A subset of these participants (n = 44) and an additional 104 new participants [18 to 60 y old (mean = 34, std = 9); 40 females, 64 males] were recruited in the replication measurements (*Replication Measurements*). Ninety percent of all participants self-reported English as their first language, and all participants were native speakers of North American English. In terms of race and ethnicity, 64% self-reported as White, 18% as Asian, 7% as Mixed, 4% as Black, 4% as Other, and 3% not reported. Participants reported no hearing loss, neurological disorders, or persistent tinnitus, and passed headphone checks and a speech-in-noise-based hearing screening (53). The study was approved by the Purdue University Institutional Review Board (IRB). The participants consented to participate and were compensated for their time. The median time for completion was approximately 1 h.

Experimental Design and Statistical Analyses.

Screening measurements. All measurements, including the screening, are listed in Fig. 8. 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 use.

Headphone-check. Two tests based on previously established procedures were carried out to screen for appropriate use of headphones (53). In the first, 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 opposite phase at the left and right channels (105). Woods et al. (105) reasoned that if a participant used a pair of sound field loudspeakers instead of headphones, acoustic cancellation would result in an attenuation of the anti-phase decoy tone leading to an error. However, the procedure becomes ineffective if a participant uses only one loudspeaker/channel. To catch participants who used a single-channel set up, we added a second task where participants were asked to report whether a low-frequency chirp (150 to 400 Hz) embedded in background low-frequency noise was rising, falling, or flat in F0. 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 "SπN0" 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 because of the so-called binaural masking level difference (BMLD) (106).

Hearing screening. Participants were screened for hearing status using a speech-in-noise task previously validated for this purpose (53). A previous meta-analysis of 15 studies suggested that speech-in-noise tasks yield a large effect size, separating individuals with typical hearing and hearing loss, and can thus serve as sensitive suprathreshold tests for typical-hearing status (53). A speech-in-babble task was administered to a cohort of individuals with known hearing status (either audiometrically typical hearing or known degree of hearing loss) and cutoff values were chosen based on the scores obtained such that the procedure yielded > 80% sensitivity to any hearing loss, and > 95% sensitivity to more-than-mild hearing loss (53). Together with the headphone-check procedure, the speech-in-noise hearing screening helped narrow the pool of participants to those who used two-channel headphones, had typical hearing, and were in good compliance with the study instructions. Two hundred participants who passed all screening procedures proceeded with the main battery of the study. No training was provided except for a brief demonstration block for each task.

TFS sensitivity measurements. We previously established that binaural behavioral and electrophysiological (EEG) measurements of ITD sensitivity can reliably reflect individual differences in TFS processing (51). Therefore, in this study, we adopted behavioral ITD detection and added a binaural version of frequency-modulation (FM) detection. Importantly, our previous study also showed that the binaural metrics were effective in capturing individual differences in TFS processing only if the contributions of extraneous "nonsensory" factors that are irrelevant to TFS processing, such as engagement, were measured and adjusted for (51). In the present study, we implemented a stand-alone measure that would also be influenced by extraneous nonsensory factors, but unaffected by TFS processing. Specifically, we used an 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 nonsensory 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.

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 randomly selected ear (i.e., ITD). The leading ear was switched from the first to the second tone in the sequence, simulating a spatial "jump" to the opposite side. ITDs in steps of a factor of two from 2 to 128 μs were presented in random order (eight repetitions for each step). The tone bursts were ramped on and off with a rise and fall time of 20 ms to reduce the likelihood that artifactual clicks are heard at the onset and offset of the stimuli 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. A two-alternative forced-choice (2AFC) task was used, where participants 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 participant with the task. The detection thresholds were quantified using a Bayesian approach (107, 108), using the psignifit toolkit from wichmann-lab. The same method for estimating thresholds was used for all measurements of

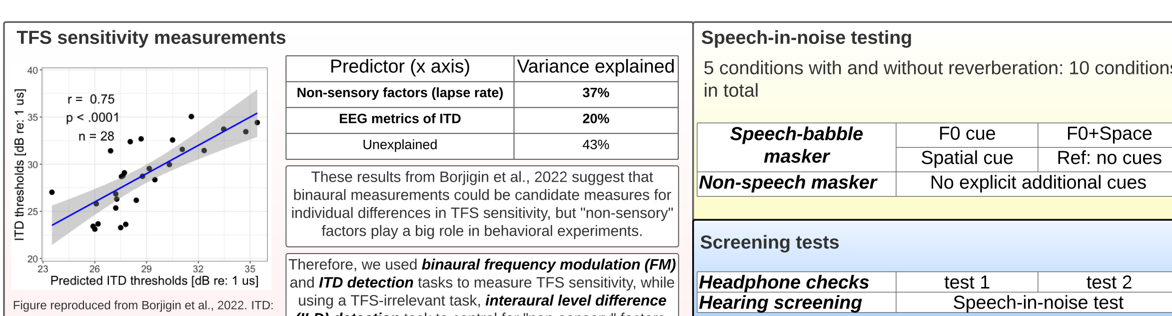

Fig. 8. An illustration of all measures included in the present study (reprinted from ref. 51).

this study, including TFS and ILD sensitivity, and speech-in-noise measurements (Fig. 8).

Binaural FM detection. We employed a binaural FM detection task as a second metric of individual TFS sensitivity. Although low-rate monaural FM detection has been used to probe TFS processing (102, 109, 110), whether monaural FM detection can truly measure individual TFS processing fidelity is questionable (49, 51). In contrast, binaural temporal processing has an unambiguous theoretical connection to TFS coding (22, 51). The binaural FM detection measure implemented in the present study consisted of target and reference stimuli in a 2AFC task. The stimuli in each interval were turned on and off with a rise and fall time of 5 ms to reduce the likelihood that artifactual clicks are heard at the onset and offset of the stimuli. The reference was a 500-ms, 500-Hz diotic pure tone. The target tone had a 2-Hz rate FM around 500 Hz with modulation out of phase in two ears to introduce binaural timing cues. A low FM rate was chosen because of the "sluggishness" of binaural system: our inability to track fast binaural modulations (111, 112). FM depths (maximum frequency deviation in one direction) in steps of a factor of two from 0.1 to 3.2 Hz were presented in random order (8 repetitions for each step). The starting phase of the stimuli was set at 0. No training was provided except for a brief demonstration block that was intended for orienting the participants before the formal testing.

ILD discrimination. ILD discrimination thresholds were measured with two consecutive 4-kHz pure-tone bursts, a frequency where TFS phase locking is generally thought to be limited (5). 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 steps of a factor of two from 0.1 to 3.2 dB (eight repetitions for each step) were presented in random order. Participants were asked to report the direction of the jump through a mouse-click response in a 2AFC task. A similar approach was used by Flanagan et al. (113), where they used intensity discrimination as a covariate in the statistical analysis to control for monaural factors since the study's focus was binaural processing. In this study, since we used binaural measurements as TFS sensitivity measures although binaural processing itself is not the focus, we used ILD discrimination to also control for the binaural factors.

Rationale. The TFS (ITD and Binaural FM) and control (ILD) measures, and sample size ($n = 200$) chosen here were guided by findings from our previous study showing robust EEG-behavior correlations in TFS measures with about 40 participants (51). However, that was an in-person study. Because the variance across participants in web-based measures is generally about 75 to 90% larger with our platform (see table 1 of ref. 53), we doubled the participant number and did so for each group (effectively quadrupling the sample size for individual difference comparisons).

Grouping of participants. Participants were classified into two groups (Good vs. Poor sensitivity), either based on TFS-sensitivity measures or the ILD measure (Fig. 2 *A* and *B*). 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. Although ILD detection is supposed to be more or less independent of TFS processing, it is subject to nonsensory contributions from variables like attention/motivation, etc., that can introduce spurious correlations between ILD detection and speech-in-noise. The mutual "regressing out" of ILD and TFS measures from each other can help reduce these nonsensory contributions.

Measurements of speech perception in noise. The stimuli consisted of a target word with a carrier phrase (Modified Rhyme Test) and a masker. The masker was either four-talker babble (IEEE speech corpus) or a steady noise composed of an inharmonic complex of tones (95), described 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 participants to orient themselves to the target voice based on the unmasked portion of the carrier phrase. A word-based test rather than a sentence-based test was chosen to minimize the influence of factors such as individual differences in working memory, and ability to use linguistic context.

Participants were tested across 10 target-masker conditions, as shown in Fig. 8. Four conditions used four-talker babble as the masker and one used a nonspeech, steady masker. The babble masker conditions included F0 cues, spatial cues, both F0 and spatial cues, and no explicit cues (i.e., reference). Note that the 4-talker babble consists of speakers of the same sex. For conditions with F0 cues, if the target was a male talker, for example, the 4-talker babble would consist of female talkers. The nonspeech masker condition had a steady masker without any explicitly added cues. The remaining five conditions were similar but with the addition of room reverberation. The presentation order of the 10 test conditions was randomized across trials. Details about the stimulus manipulations used are provided below. For each condition, speech intelligibility was measured over a range of SNRs to estimate the SRT, defined as the SNR at which approximately 50% of the words were intelligible.

F0 cues. To control the available F0 cues for separating the target and masker, the audio recordings for all trials were first processed to remove inherent F0 fluctuations (i.e., monotonized to the estimated F0 median) using Praat (version 6.4.04) and a custom Praat script (written by Matthew B. Winn). 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 Fig. 9. 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, and the female talkers' F0 values to 235, 240, 250, and 255 Hz. Note that the target and masker of the same sex had similar F0 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 F0 difference. The F0 contour was flattened for all other stimulus configurations (i.e., reference, space, F0+space, and nonspeech noise masker). F0-based masking release was estimated as the SRT difference between the reference condition where the target and masker stimuli were composed of recordings from same-sex talkers and the "F0" condition where there was a large F0 separation by virtue of the target and masker stimuli being composed of recordings from opposite-sex talkers.

Spatial cues. To simulate the perception of spatial separation using purely TFS-based cues, the polarity of the target in one ear was inverted while the masker was kept the same in the two ears. This configuration is denoted $S_\pi N_0$. The fully diotic condition without this interaural manipulation is referred to as $S_0 N_0$. A lower SRT (i.e., better performance) is typically observed in the $S_\pi N_0$ condition, with the difference in SRTs being BMLD or spatial release from masking (114).

Steady masker and reverberation. Performance in the presence of a steady masker was used to evaluate the role of TFS in providing release from so-called "informational masking" (96). Accordingly, the steady masker was designed to have minimal intrinsic modulations with tonal frequency components whose relative levels matched the relative spectral levels of different bands of the corresponding speech stimuli using the procedure described in ref. 95. The masker was dichotic, consisting of odd-numbered sinusoids delivered to one ear and even-numbered sinusoids to the opposite ear. This approach reduced the occurrence of beats generated by neighboring components in the peripheral

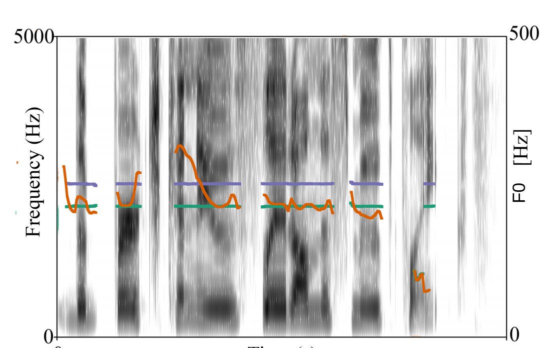

Fig. 9. The spectrogram of a sentence: "The birch canoe slid on the smooth planks." The orange curve shows the estimated F0 contour with natural fluctuations; the flattened F0 contour is shown in green; the flattened F0 contour that was transposed to a preset frequency (255 Hz in this example), is shown in purple.

auditory system, ensuring minimal amplitude fluctuations of the masker at the outputs of the auditory filters. Owing to the lack of modulations (explicit and intrinsic), this masker represents a condition where energetic masking is dominant while avoiding most sources of informational masking. Note that conventionally used noise maskers such as speech-spectrum stationary noise have intrinsic modulations that can contribute to masking at more central levels of the auditory system (97, 115–117). Finally, to simulate listening under reverberation, the stimuli that were recorded under anechoic conditions were convolved with binaural room impulse responses recorded in a bar (BarMonsieurRicard.wav from echoThief).

SRT estimation. To robustly estimate the mean and variance of the masking release based on different cues, SRTs for each speech-in-noise condition were estimated using a jackknife resampling procedure. Within each group (Good vs. Poor), a leave-one-out procedure was used: psychometric functions were fit to the percent-correct vs. SNR scores that were obtained by averaging the data across all participants except the one being left out. The SRT was then estimated as the mid-point of this psychometric curve. 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 ref. 118, the group-level mean M was estimated as the mean across the jackknife samples, and the variance as the sample-variance V across the jackknife samples multiplied by $(k-1)$. The jackknife procedure avoids the need to fit psychometric curves for speech intelligibility as a function of SNR or to estimate SRTs at the level of the individual participant, and yet robustly estimates the variance in the SRTs (and masking release values) across participants within each group.

Response time. Two participants with comparable SRTs could experience different levels of listening effort (76, 77). To assess the role of listening effort, 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 procedure as for the SRT estimates was used to estimate the mean and variance of the response times. Trials with response times larger than 10 s 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 group, and for each speech-in-noise condition.

Statistical analyses. The primary analyses 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 jackknife procedure, it was reasonable to assume that group-level estimates represented parameter estimates for normally distributed data. Accordingly, simple one-tailed z-tests were used for making inferences. As described previously, among the 10 speech-in-noise conditions, 5 simulated speech-in-noise mixtures in anechoic environments and 5 included room reverberation. To investigate the effects of reverberation, data from all 5 speech-in-noise configurations were combined using inverse variance pooling (119, 120). For response time comparisons across groups, all 10 conditions were pooled.

Replication Measurements. Two rounds of replication were conducted. The first round used a subset of the original pool of 200 participants to assess the test–retest reliability of the TFS measures, as well as to replicate the main findings of the original study. The second round of replications involved an independent, newly screened cohort of participants. For the second replication, the participant screening procedures were slightly altered a priori to gain greater statistical power. Specifically, we selected participants who fell in the top half or bottom half by both TFS sensitivity measures (ITD and binaural FM thresholds) and excluded participants who fell in the top half by one measure and bottom half by the other. All other procedures for data collection and analysis in both rounds of replication followed the same protocols as the original study.

Data, Materials, and Software Availability. Portions of the paper were developed from the doctoral dissertation submitted to Purdue University by A.B. (121). The data, scripts for setting up online experiments, data analyses, and step-by-step instructions have been uploaded on the Open Science Framework (https://osf.io/rhfw4/?view_only=a0e6731a8e4e4a1b938650a447bf855f).(122)

ACKNOWLEDGMENTS. We thank our research participants for their participation in this study. This research was supported by funding from the NIH (Grant R01DC015989). We would also like to thank Dr. Brian Moore and one anonymous reviewer for comments on previous versions of the manuscript.

1. A. J. Oxenham, How we hear: The perception and neural coding of sound. *Annu. Rev. Psychol.* **69**, 27–50 (2018).
2. D. Hilbert, "Grundz¨uge einer allgemeinen Theorie der linearen Integralgleichungen" in *Integralgleichungen und Gleichungen mit unendlich vielen Unbekannten*, D. Hilbert, E. Schmidt, A. Pietsch, Eds. (Springer, Vienna, 1989), pp. 6–169.
3. D. H. Johnson, The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.* **68**, 1115–1122 (1980).
4. P. X. Joris, T. C. T. Yin, Responses to amplitude-modulated tones in the auditory nerve of the cat. *J. Acoust. Soc. Am.* **91**, 215–232 (1992).
5. E. Verschooten *et al.*, The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hear. Res.* **377**, 109–121 (2019).
6. J. W. Hughes, The upper frequency limit for the binaural localization of a pure tone by phase difference. *Proc. R. Soc. London, Ser. B-Biol. Sci.* **128**, 293–305 (1940).
7. A. Brughera, L. Dunai, W. M. Hartmann, Human interaural time difference thresholds for sine tones: The high-frequency limit. *J. Acoust. Soc. Am.* **133**, 2839–2855 (2013).
8. M. N. Shadlen, W. T. Newsome, Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569–579 (1994).
9. T. Baden, F. Esposti, A. Nikolaev, L. Lagnado, Spikes in retinal bipolar cells phase-lock to visual stimuli with millisecond precision. *Curr. Biol.* **21**, 1859–1869 (2011).
10. M. N. Wallace, R. G. Rutkowski, T. M. Shackleton, A. R. Palmer, Phase-locked responses to pure tones in guinea pig auditory cortex. *NeuroReport* **11**, 3989–3993 (2000).
11. P. X. Joris, C. E. Schreiner, A. Rees, Neural processing of amplitude-modulated sounds. *Physiol. Rev.* **84**, 541–577 (2004).
12. S. B. Laughlin, R. R. de Ruyter, J. C. van Steveninck, Anderson,, The metabolic cost of neural information. *Nat. Neurosci.* **1**, 36–41 (1998).
13. A. Hasenstaub, S. Otte, E. Callaway, T. J. Sejnowski, Metabolic cost as a unifying principle governing neuronal biophysics. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12329–12334 (2010).
14. A. J. Oxenham, Revisiting place and temporal theories of pitch. *Acoust. Sci. Technol.* **34**, 388–396 (2013).
15. R. Drullman, Temporal envelope and fine structure cues for speech intelligibility. *J. Acoust. Soc. Am.* **97**, 585–592 (1995).
16. J. Swaminathan, M. G. Heinz, Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *J. Neurosci.* **32**, 1747–1756 (2012).
17. A. J. Oxenham, A. M. Simonson, Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *J. Acoust. Soc. Am.* **125**, 457–468 (2009).
18. A. de Cheveigné, D. Pressnitzer, The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction. *J. Acoust. Soc. Am.* **119**, 3908–3918 (2006).
19. M. R. Saddler, R. Gonzalez, J. H. McDermott, Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception. *Nat. Commun.* **12**, 7278 (2021).
20. E. Javel, J. B. Mott, Physiological and psychophysical correlates of temporal processes in hearing. *Hear. Res.* **34**, 275–294 (1988).
21. Z. M. Smith, B. Delgutte, A. J. Oxenham, Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87 (2002).
22. T. C. Yin, J. C. Chan, Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* **64**, 465–488 (1990).
23. B. C. J. Moore, Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.* **54**, 610–619 (1973).
24. A. J. M. Houtsma, J. Smurzynski, Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* **87**, 304–310 (1990).
25. J. G. Bernstein, A. J. Oxenham, Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* **113**, 3323–3334 (2003).
26. A. Ihlefeld, B. Shinn-Cunningham, Spatial release from energetic and informational masking in a divided speech identification task. *J. Acoust. Soc. Am.* **123**, 4380–4392 (2008).
27. F. J. Gallun, A. C. Diedesch, S. D. Kampel, K. M. Jakien, Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Front. Neurosci.* **7** (2013).
28. F. J. Gallun *et al.*, Verification of an automated headphone-based test of spatial release from masking. *Proc. Meet. Acoust.* **25**, 050001 (2015).
29. N. K. Srinivasan, K. M. Jakien, F. J. Gallun, Release from masking for small spatial separations: Effects of age and hearing loss. *J. Acoust. Soc. Am.* **140**, EL73–EL78 (2016).
30. K. M. Jakien, Sean D. Kampel, Meghan M. Stansell, Frederick J. Gallun, Validating a rapid, automated test of spatial release from masking. *Am. J. Audiol.* **26**, 507–518 (2017).
31. K. M. Jakien, F. J. Gallun, Normative data for a rapid, automated test of spatial release from masking. *Am. J. Audiol.* **27**, 529–538 (2018).
32. J. P. L. Brokx, S. G. Nooteboom, Intonation and the perceptual separation of simultaneous voices. *J. Phon.* **10**, 23–36 (1982).
33. J. Bird, C. J. Darwin, "Effects of a difference in fundamental frequency in separating two sentences" *in Psychophysical and Physiological Advances in Hearing*, A. R. Palmer, A. Rees, A. Q. Summerfield, R. Meddis, Eds. (Whurr Publishers, London, 1998), pp. 263–269.
34. V. Summers, M. R. Leek, F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J. Speech Lang. Hear. Res.* **41**, 1294–1306 (1998).
35. V. Best, E. Ozmeral, F. J. Gallun, K. Sen, B. G. Shinn-Cunningham, Spatial unmasking of birdsong in human listeners: Energetic and informational factors. *J. Acoust. Soc. Am.* **118**, 3766–3773 (2005).

36. A. Ihlefeld, B. G. Shinn-Cunningham, Effect of source spectrum on sound localization in an everyday reverberant room. *J. Acoust. Soc. Am.* **130**, 324–333 (2011).
37. D. G. Richards, R. H. Wiley, Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication. *Am. Nat.* **115**, 381–399 (1980).
38. F. L. Wightman, D. J. Kistler, The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **91**, 1648–1661 (1992).
39. K. Hopkins, B. C. J. Moore, M. A. Stone, Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J. Acoust. Soc. Am.* **123**, 1140–1153 (2008).
40. K. Hopkins, B. C. J. Moore, The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *J. Acoust. Soc. Am.* **125**, 442–446 (2009).
41. C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, B. C. J. Moore, Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18866–18869 (2006).
42. C. Lorenzi, L. Debruille, S. Garnier, P. Fleuriot, B. C. J. Moore, Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *J. Acoust. Soc. Am.* **125**, 27–30 (2009).
43. M. Ardoint, C. Lorenzi, Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hear. Res.* **260**, 89–95 (2010).
44. H. Voelcker, Toward a unified theory of modulation. Part I: Phase-envelope relationships. *Proc. IEEE* **54**, 340–353 (1966).
45. G. Gilbert, I. Bergeras, D. Voillery, C. Lorenzi, Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. *J. Acoust. Soc. Am.* **122**, 1336 (2007).
46. M. R. Saddler, J. H. McDermott, Models optimized for real-world tasks reveal the necessity of precise temporal coding in hearing. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.04.21.590435 (Accessed 10 July 2024).
47. J. H. McDermott, A. J. Lehr, A. J. Oxenham, Individual differences reveal the basis of consonance. *Curr. Biol.* **20**, 1035–1041 (2010).
48. H. M. Bharadwaj, S. Masud, G. Mehraei, S. Verhulst, B. G. Shinn-Cunningham, Individual differences reveal correlates of hidden hearing deficits. *J. Neurosci.* **35**, 2161–2172 (2015).
49. K. L. Whiteford, H. A. Kreft, A. J. Oxenham, The role of cochlear place coding in the perception of frequency modulation. *eLife* **9**, e58468 (2020).
50. M. J. McPherson, J. H. McDermott, Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32169–32180 (2020).
51. A. Borjigin, A. R. Hustedt-Mai, H. M. Bharadwaj, *Individualized Assays of Temporal Coding in the Ascending Human Auditory System* (Society for Neuroscience, 2022).
52. Vinay, B. C. J. Moore, Exploiting individual differences to assess the role of place and phase locking cues in auditory frequency discrimination at 2 kHz. *Sci. Rep.* **13**, 13801 (2023).
53. B. A. Mok *et al*., Web-based psychoacoustics: Hearing screening, infrastructure, and validation. *Behav. Res. Methods* **56**, 1433–1448 (2023).
54. D. Ruggles, H. Bharadwaj, B. G. Shinn-Cunningham, 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 (2011).
55. C. J. Darwin, Auditory grouping. *Trends Cogn. Sci.* **1**, 327–333 (1997).
56. G. Kidd Jr., H. S. Colburn, "Informational masking in speech recognition" in *The Auditory System at the Cocktail Party: Springer Handbook of Auditory Research 60*, J. C. Middlebrooks et al., Eds. (Springer International Publishing AG, 2017), pp. 75–109.
57. M. K. Pichora-Fuller *et al*., Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear.* **37**, 5S (2016).
58. B. G. Shinn-Cunningham, Object-based auditory and visual attention. *Trends Cogn. Sci.* **12**, 182–186 (2008).
59. V. Viswanathan, H. M. Bharadwaj, B. G. Shinn-Cunningham, M. G. Heinz, Modulation masking and fine structure shape neural envelope coding to predict speech intelligibility across diverse listening conditions. *J. Acoust. Soc. Am.* **150**, 2230–2244 (2021).
60. V. Viswanathan, B. G. Shinn-Cunningham, M. G. Heinz, Temporal fine structure influences voicing confusions for consonant identification in multi-talker babble. *J. Acoust. Soc. Am.* **150**, 2664–2676 (2021).
61. D. W. Downs, M. A. Crum, Processing demands during auditory learning under degraded listening conditions. *J. Speech Hear. Res.* **21**, 702–714 (1978).
62. A. Sarampalis, S. Kalluri, B. Edwards, E. Hafter, Objective measures of listening effort: Effects of background noise and noise reduction. *J. Speech, Lang., Hear. Res.* **52**, 1230–1240 (2009).
63. C. Füllgrabe, B. C. J. Moore, M. A. Stone, *Age-Group Differences in Speech Identification Despite Matched Audiometrically Normal Hearing: Contributions from Auditory Temporal Processing and Cognition* (Frontiers, 2015).
64. J. H. Grose, S. K. Mamo, Frequency modulation detection as a measure of temporal processing: Age-related monaural and binaural effects. *Hear. Res.* **294**, 49–54 (2012).
65. K. L. Whiteford, A. J. Oxenham, Using individual differences to test the role of temporal and place cues in coding frequency modulation. *J. Acoust. Soc. Am.* **138**, 3093–3104 (2015).
66. K. L. Whiteford, H. A. Kreft, A. J. Oxenham, Assessing the role of place and timing cues in coding frequency and amplitude modulation as a function of age. *J. Assoc. Res. Otolaryngol.* **18**, 619–633 (2017).
67. E. S. Lelo *et al*., Portable automated rapid testing (PART) for auditory assessment: Validation in a young adult normal-hearing population. *J. Acoust. Soc. Am.* **148**, 1831–1851 (2020).
68. G. B. Henning, Lateralization of low-frequency transients. *Hear. Res.* **9**, 153–172 (1983).
69. R. H. Dye, The combination of interaural information across frequencies: Lateralization on the basis of interaural delay. *J. Acoust. Soc. Am.* **88**, 2159–2170 (1990).
70. C. J. Darwin, "Pitch and auditory grouping" in *Pitch: Neural Coding and Perception. Springer Handbook of Auditory Research*, C. J. Plack, R. R. Fay, A. J. Oxenham, A. N. Popper, Eds. (Springer, New York, NY, 2005), pp. 278–305.
71. T. L. Arbogast, C. R. Mason, G. Kidd, The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am.* **112**, 2086–2098 (2002).
72. B. Shinn-Cunningham, V. Best, A. K. C. Lee, "Auditory object formation and selection" in *The Auditory System at the Cocktail Party: Springer Handbook of Auditory Research, J. C. Middlebrooks*, J. Z. Simon, A. N. Popper, R. R. Fay, Eds. (Springer International Publishing, New York, NY, 2017), pp. 7–40.
73. J. F. Culling, Q. Summerfield, D. H. Marshall, Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels. *Speech Commun.* **14**, 71–95 (1994).
74. K. J. Palomäki, G. J. Brown, D. Wang, A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation. *Speech Commun.* **43**, 361–378 (2004).
75. R. L. Freyman, A. M. Griffin, A. J. Oxenham, Intelligibility of whispered speech in stationary and modulated noise maskers. *J. Acoust. Soc. Am.* **132**, 2514–2523 (2012).
76. J. E. Peelle, Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* **39**, 204–214 (2018).
77. S. Alhanbali, P. Dawes, R. E. Millman, K. J. Munro, Measures of listening effort are multidimensional. *Ear Hear.* **40**, 1084–1097 (2019).
78. S. Gatehouse, J. Gordon, Response times to speech stimuli as measures of benefit from amplification. *Br. J. Audiol.* **24**, 63–68 (1990).
79. T. Baer, B. C. J. Moore, S. Gatehouse, Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality and response times. *J. Rehabil. Res.* **30**, 49–72 (1993).
80. F. Apoux, O. Crouzet, C. Lorenzi, Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: Effects on identification performance and response times. *Hear. Res.* **153**, 123–131 (2001).
81. Y. Benjamini, Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing* (Royal Statistical Society, 1995).
82. D. J. Tollin, K. Koka, J. J. Tsai, Interaural level difference discrimination thresholds for single neurons in the lateral superior olive. *J. Neurosci.* **28**, 4848–4860 (2008).
83. T. H. Churchill, A. Kan, M. J. Goupell, R. Y. Litovsky, Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners. *J. Acoust. Soc. Am.* **136**, 1246–1256 (2014).
84. E. A. Macpherson, J. C. Middlebrooks, Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am.* **111**, 2219–2236 (2002).
85. A. Seebeck, Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Ann. Phys.* **129**, 417–436 (1841).
86. G. S. Ohm, Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann. Phys.* **135**, 513–565 (1843).
87. F. Attneave, R. K. Olson, Pitch as a medium: A new approach to psychophysical scaling. *Am. J. Psychol.* **84**, 147–166 (1971).
88. A. J. Oxenham, C. Micheyl, M. V. Keebler, A. Loper, S. Santurette, Pitch perception beyond the traditional existence region of pitch. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7629–7634 (2011).
89. B. C. J. Moore, S. M. A. Ernst, Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *J. Acoust. Soc. Am.* **132**, 1542–1547 (2012).
90. J. E. Rose, J. F. Brugge, D. J. Anderson, J. E. Hind, Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* **30**, 769–793 (1967).
91. A. R. Palmer, I. J. Russell, Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear. Res.* **24**, 1–15 (1986).
92. M. K. Qin, A. J. Oxenham, Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J. Acoust. Soc. Am.* **114**, 446–454 (2003).
93. G. S. Stickney, P. F. Assmann, J. Chang, F. G. Zeng, Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *J. Acoust. Soc. Am.* **122**, 1069–1078 (2007).
94. A. J. Oxenham, Pitch perception. *J. Neurosci.* **32**, 13335–13338 (2012).
95. M. A. Stone, B. C. J. Moore, On the near non-existence of "pure" energetic masking release for speech. *J. Acoust. Soc. Am.* **135**, 1967–1977 (2014).
96. N. I. Durlach *et al*., Note on informational masking (L). *J. Acoust. Soc. Am.* **113**, 2984–2987 (2003).
97. D. S. Brungart, P. S. Chang, B. D. Simpson, D. Wang, Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J. Acoust. Soc. Am.* **120**, 4007–4018 (2006).
98. T. Nuesse, R. Steenken, T. Neher, I. Holube, Exploring the link between cognitive abilities and speech recognition in the elderly under different listening conditions. *Front. Psychol.* **9**, 678 (2018).
99. L. E. Humes, Factors underlying individual differences in speech-recognition threshold (SRT) in noise among older adults. *Front. Aging Neurosci.* **13**, 702739 (2021).
100. F. J. Gallun *et al*., Relating suprathreshold auditory processing abilities to speech understanding in competition. *Brain Sci.* **12**, 695 (2022).
101. D. Ruggles, H. Bharadwaj, B. G. Shinn-Cunningham, Why middle-aged listeners have trouble hearing in everyday settings. *Curr. Biol.* **22**, 1417–1422 (2012).
102. A. Parthasarathy, K. E. Hancock, K. Bennett, V. DeGruttola, D. B. Polley, Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing. *eLife* **9**, e51419 (2020).
103. J. Resnik, D. B. Polley, Cochlear neural degeneration disrupts hearing in background noise by increasing auditory cortex internal noise. *Neuron* **109**, 984–996.e4 (2021).
104. H. Bharadwaj, B. Shinn-Cunningham, What's been hidden in hidden hearing loss. *Neuron* **109**, 909–911 (2021).
105. K. J. P. Woods, M. H. Siegel, J. Traer, J. H. McDermott, Headphone screening to facilitate web-based auditory experiments. *Atten., Percept., Psychophys.* **79**, 2064–2072 (2017).
106. J. C. R. Licklider, The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise. *J. Acoust. Soc. Am.* **20**, 150–159 (1948).
107. M. Kuss, F. Jäkel, F. A. Wichmann, Bayesian inference for psychometric functions. *J. Vis.* **5**, 8 (2005).
108. H. H. Schütt, S. Harmeling, J. H. Macke, F. A. Wichmann, Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vis. Res.* **122**, 105–123 (2016).

109. O. Strelcyk, T. Dau, Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am.* **125**, 3328–3345 (2009).
110. B. C. J. Moore, A. Sek, Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *J. Acoust. Soc. Am.* **100**, 2320–2331 (1996).
111. J. Blauert, On the lag of lateralization caused by interaural time and intensity differences. *Audiology* **11**, 265–270 (1972).
112. R. Singh, H. M. Bharadwaj, Cortical temporal integration can account for limits of temporal perception: Investigations in the binaural system. *Commun. Biol.* **6**, 1–12 (2023).
113. S. A. Flanagan *et al.*, Development of binaural temporal fine structure sensitivity in children. *J. Acoust. Soc. Am.* **150**, 2967–2976 (2021).
114. H. Levitt, L. R. Rabiner, Binaural release from masking for speech and gain in intelligibility. *J. Acoust. Soc. Am.* **42**, 601–608 (1967).
115. J. F. Culling, M. A. Stone, "Energetic masking and masking release in the auditory system at the cocktail party" in *Springer Handbook of Auditory Research* (Springer Nature, 2017), vol. 60, pp. 41–73.
116. W. M. Hartmann, J. Pumplin, Noise power fluctuations and the masking of sine signals. *J. Acoust. Soc. Am.* **83**, 2277–2289 (1988).
117. A. Kohlrausch *et al.*, Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acta Acust. Acust.* **83**, 659–669 (1997).
118. B. Efron, C. Stein, The jackknife estimate of variance. *Ann. Stat.* **9**, 586–596 (1981).
119. B. K. Sinha, J. Hartung, G. Knapp, *Statistical Meta-Analysis with Applications* (John Wiley & Sons, 2011).
120. L. V. Hedges, Estimation of effect size from a series of independent experiments. *Psychol. Bull.* **92**, 490–499 (1982).
121. A. Borjigin, "The role of temporal fine structure in everyday hearing" PhD, Purdue University, Indiana (2022). https://www.proquest.com/docview/2838440365/abstract/50281C45FC954E5BPQ/1.
122. A. Borjigin, Temporal fine structure project (phase II). Open Science Framework. https://osf.io/rhfw4/. Accessed 11 December 2024.