JASA ARTICLE

On the use of the TIMIT, QuickSIN, NU-6, and other widely used bandlimited speech materials for speech perception experiments[a)]

Brian B. Monson[1,b)] and Emily Buss[2]
[1]Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, Champaign, Illinois 61820, USA
[2]Department of Otolaryngology/HNS, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA

ABSTRACT:
The use of spectrally degraded speech signals deprives listeners of acoustic information that is useful for speech perception. Several popular speech corpora, recorded decades ago, have spectral degradations, including limited extended high-frequency (EHF) (>8 kHz) content. Although frequency content above 8 kHz is often assumed to play little or no role in speech perception, recent research suggests that EHF content in speech can have a significant beneficial impact on speech perception under a wide range of natural listening conditions. This paper provides an analysis of the spectral content of popular speech corpora used for speech perception research to highlight the potential shortcomings of using bandlimited speech materials. Two corpora analyzed here, the TIMIT and NU-6, have substantial low-frequency spectral degradation (<500 Hz) in addition to EHF degradation. We provide an overview of the phenomena potentially missed by using bandlimited speech signals, and the factors to consider when selecting stimuli that are sensitive to these effects. https://doi.org/10.1121/10.0013993
(Received 31 March 2022; revised 20 July 2022; accepted 20 August 2022; published online 13 September 2022)
[Editor: Matthew B. Winn] Pages: 1639–1645

I. INTRODUCTION

The selection of speech materials for speech perception research directly affects measured behavioral outcomes. Many speech materials in use today were recorded decades ago using recording procedures that spectrally degraded the speech materials. For example, the use of low sampling rates (e.g., 16 or 22 kHz) was standard practice in speech research when some of the well-known and widely used speech materials were recorded, resulting in materials that are bandlimited to 8 or 11 kHz. Additionally, transducers used for recording may have had low- and/or high-frequency roll-off or other variations in the frequency response that cut out or degraded spectral content. The result is several speech corpora that do not represent the high-fidelity speech signals that listeners encounter in their everyday lives.

Is the use of speech materials that are bandlimited to 8 kHz problematic? Several studies have demonstrated that extended high-frequency (EHF) (>8 kHz) energy in speech is audible and useful for speech perception (Moore *et al.*, 2008; Monson *et al.*, 2014a; Hunter *et al.*, 2020). EHF cues in speech have been shown to support: (1) speech recognition in noise for adults (Monson *et al.*, 2019; Motlagh-Zadeh *et al.*, 2019; Trine and Monson, 2020; Polspoel *et al.*, 2022; see also Levy *et al.*, 2015) and children (Flaherty *et al.*, 2021), (2) consonant and vowel recognition when lower frequencies are either partially removed (Lippmann, 1996) or entirely absent (Vitela *et al.*, 2015), (3) speech localization (Best *et al.*, 2005), (4) talker head orientation discrimination (Monson *et al.*, 2019), and (5) subjective speech quality (Moore and Tan, 2003). In most of these studies, the utility of EHF speech cues was demonstrated by comparing the outcome measure for speech low-pass filtered at 8 kHz to that for full-bandwidth speech. Moreover, EHF audiometric thresholds predict both EHF audibility in speech (Monson *et al.*, 2014b; Monson and Caravello, 2019) and speech-in-noise performance (Badri *et al.*, 2011; Yeend *et al.*, 2019; Motlagh-Zadeh *et al.*, 2019; Trine and Monson, 2020; Braza *et al.*, 2022; Mishra *et al.*, 2021; Mishra *et al.*, 2022).

These findings run counter to the traditional notion of a limited speech bandwidth (or "speech frequencies") with an upper edge of 7 or 8 kHz (e.g., Fletcher and Wegel, 1922; Miller, 1947; Pollack, 1948; Guinan, 2017; see Monson *et al.*, 2014a for a review on this topic). It is true that speech with a bandwidth restricted at both the low- and high-frequency ends of the spectrum is sufficient for good speech recognition under favorable signal-to-noise ratios (Fletcher and Galt, 1950), with negligible benefit of speech cues above 6 kHz (ANSI, 1997). However, early speech research demonstrating this phenomenon was often limited by transducers with poor frequency response, particularly at higher frequencies, or simply did not consider ecological conditions under which EHF cues in speech might be useful (Monson *et al.*, 2014a; Hunter *et al.*, 2020). More recent

[a)]This paper is part of a special issue on Reconsidering Classic Ideas in Speech Communication.
[b)]Also at: Department of Biomedical and Translational Sciences, Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Champaign, Illinois, 61820, USA. Electronic mail: monson@illinois.edu

TABLE I. Details of speech corpora analyzed in this study.

Acronym	Name	Talkers	Sampling rate	Source/Reference
TIMIT and PRESTO	Texas Instruments/Massachusetts Institute of Technology & Perceptually Robust English Sentence Test Open-set	Multiple male and female	16 kHz	Garofolo *et al.* (1993); Gilbert *et al.* (2013)
AzBio	Arizona Biomedical Institute Sentence Test	Multiple male and female	22 kHz	Spahr *et al.* (2012)
BKB-SIN	Bamford–Kowal–Bench Speech-In-Noise	One male		Bench *et al.* (1979); distributed by Etymotic
CNC	Consonant-Nucleus-Consonant	One male		Peterson *et al.* (1962); Auditec
NU-6	Northwestern University Test No.6	One male		Auditec
Quick-SIN	Quick Speech-in-Noise	One female		Killion *et al.* (2004); distributed by Auditec
SPIN-R	Speech-in-Noise Revised	One male		Bilger *et al.* (1984)
BEL	Basic English Lexicon	One male, two female	44.1 kHz	Rimikis *et al.* (2013)
HINT	Hearing in Noise Test	One male	20 kHz	Nilsson *et al.* (1994); distributed by Interacoustics
HIST	Hearing in Speech Test	One male	44.1 kHz	Levy *et al.* (2015)
Monson	N/A	Multiple male and female	44.1 kHz	Monson *et al.* (2012a)
Monson BKB	N/A	One female	44.1 kHz	Monson *et al.* (2019)

research demonstrated that higher frequencies contribute more to speech perception than initially thought, but these studies often used stimuli that were still bandlimited to 9, 10, or 11 kHz (e.g., Stelmachowicz *et al.*, 2001; Moore *et al.*, 2010; McCreery and Stelmachowicz, 2011, 2013). Given the traditional understanding of the bandwidth important for speech perception, it may be surprising to some that energy between 13 and 20 kHz in speech is audible to young, normal-hearing listeners. That is, the low-pass filter cutoff frequency at which listeners can detect a loss of EHF information in speech is approximately 13 kHz (Monson and Caravello, 2019).

In the spirit of this special issue on reconsidering traditional ideas, the objective of the present contribution is to provide an analysis of the spectra and recording details of several widely used speech corpora, to draw attention to the importance of evaluating frequency content of speech materials used in hearing research, and to review the effects that are missed when using bandlimited recordings. In doing so, our invitation is for researchers to reconsider the interpretation of data collected with bandlimited speech signals. We urge the auditory research community to consider potential effects of the loss of information incurred by using band limited speech recordings, and, accordingly, to select speech materials that faithfully represent the acoustics of real-world signals, as relevant to the research question.

II. METHODS

Stimuli evaluated here were publicly available recordings that are commonly used in speech perception research. Table I provides details and acronym definitions of the recordings included in the analysis. Recordings distributed on compact disks (CDs) were ripped and saved to disks. In some cases, a subset of recordings was used; results below represent a minimum of 50 samples (words or sentences) for each corpus, except the BEL corpus, which had only 20 sentences from a male talker and 40 from female talkers.

We compared the long-term average speech spectra (LTASS) of these materials to those of an anechoic full-band speech corpus collected by Monson *et al.* (2012a) and full-band BKB sentences recorded with a single female talker by Monson *et al.* (2019). Both sets of materials were recorded at a sampling rate of 44.1 kHz using Class I precision microphones with flat frequency response out to 20 kHz. These recordings were used in several of the experiments demonstrating the importance of EHFs to speech perception (e.g., Monson *et al.*, 2019; Trine and Monson, 2020; Flaherty *et al.*, 2021; Braza *et al.*, 2022).

Because there are observable gender differences in the LTASS (Monson *et al.*, 2012a), speech was analyzed separately by talker gender. Speech materials were manually edited to remove silence between speech samples to prevent the spectral content of the noise floor from affecting the LTASS. The NU-6 words were evaluated with the carrier phrase, "say the word," which was produced prior to each word; excising the carrier phrase and recomputing the LTASS did not modify the primary features noted in Sec. III, however. Speech samples by all talkers of the same gender were concatenated into a single audio file to obtain the LTASS for each corpus. The LTASS for most corpora was calculated using a 2048 point fast Fourier transform (FFT) and a hanning window with 50% overlap across frames. The LTASS for the TIMIT/PRESTO was calculated using a 1024 point FFT because of its limited bandwidth, but otherwise using identical processing procedures. To estimate expected between-talker variability, an individual LTASS was also calculated for each subject in the Monson *et al.* (2012a) recordings (10 male, 10 female); these spectra were used to characterize the range of LTASS values across

https://doi.org/10.1121/10.0013993

talkers. Each LTASS was normalized to 65 dB SPL, which is an average speech conversational level (Monson *et al.*, 2012a).

III. RESULTS

The LTASS for each corpus is shown in a zoomed view (0–3 kHz) in Fig. 1 and full-band (0–20 kHz) in Fig. 2. These traces indicate comparable distribution of energy between 500 Hz and 8 kHz. However, there are notable differences below 500 Hz and above 8 kHz. Whereas most corpora have comparable energy levels below 500 Hz, with maxima corresponding to the mean fundamental frequency of the talker gender, the TIMIT/PRESTO appears to roll off below 500 Hz. This roll-off results in losses of 10–15 dB below what would be expected at frequencies corresponding to fundamental frequencies of male and female speech (100–250 Hz). A similar low-frequency roll-off is observed in the NU-6 corpus, resulting in a highly abnormal 13 dB increase between 150 and 500 Hz.

The TIMIT/PRESTO was originally recorded with a sampling rate of 20 kHz, but then downsampled to 16 kHz prior to public distribution, so there is no energy above 8 kHz. The AzBio was recorded at 22 kHz, resulting in an upper limit of 11 kHz. Other corpora were recorded at >22 kHz and contain EHF cues, but with a wide range of magnitudes (in some cases, the exact sampling rates could not be determined; see Table I). For example, the difference between the BKB-SIN and Monson *et al.* (2012a) male recordings grows from 10 dB at 8 kHz to 20 dB at 16 kHz. The analysis of CNC words shows a similar pattern. The NU-6 and QuickSIN display a much steeper roll-off beyond 8 kHz compared to natural full-bandwidth speech. There is also evidence of high-frequency distortion just below 16 kHz for the QuickSIN. The SPIN has higher levels between 8 and 11 kHz than most of the other corpora, but the spectral slope rolls off steeply above 11 kHz. The spectral slope differences could be due to differences in the microphones' frequency responses or positions relative to the talkers' mouths, either of which can affect spectral content (Monson *et al.*, 2012b). Although some of these spectral degradations seem subtle, others likely give rise to the audible differences between recordings. Subjectively, some of the corpora exhibit audible noise or distortion, including the TIMIT/PRESTO, QuickSIN, and the NU-6.

Phonetic content of the speech materials and talker-specific differences can affect the LTASS at EHFs, but this variability is expected to be smaller than the large variations across corpora illustrated in Fig. 2. For example, a comparison between the Monson *et al.* (2012a) female corpus, Monson *et al.* (2019) single female BKB talker, and BEL female talkers demonstrates expected variability based on talker characteristics and differing phonetic content. The difference between the BEL corpus male talker, HIST male talker, and Monson *et al.* (2012a) male talkers is also within normal variation that might be expected based on these factors.

Given the low levels of EHF energy in these spectra, one might question whether EHF level differences between recordings are audible. It should be noted that the linearly scaled FFT used here does not reflect a physiological representation of energy distribution. Broadened auditory filter bands at higher frequencies give rise to increased spectral levels at EHFs, as has been previously shown (Levy *et al.*, 2015; Hunter *et al.*, 2020). Furthermore, it has been demonstrated that average difference limens for speech spectral

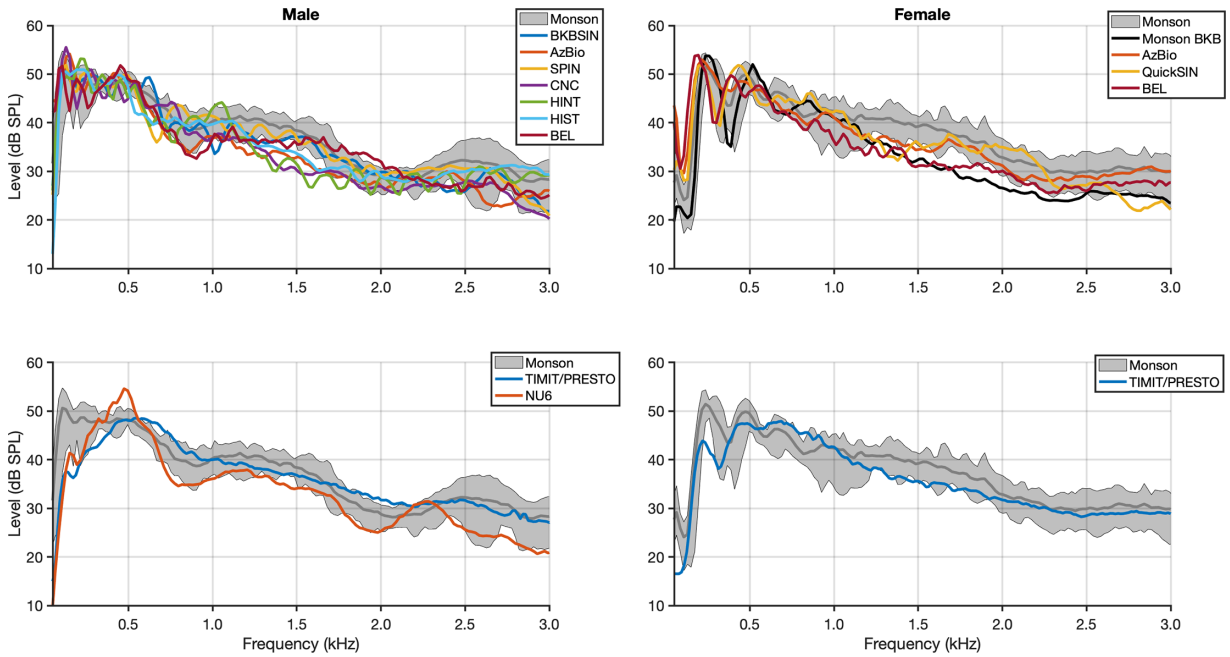

FIG. 1. (Color online) Zoomed view (0–3 kHz) of the LTASS for each corpus analyzed in this study. Left panels are male speech and right panels are female speech. For comparison, the Monson *et al.* (2012a) female and male LTASS are also shown. Shading indicates the range of the Monson recordings.

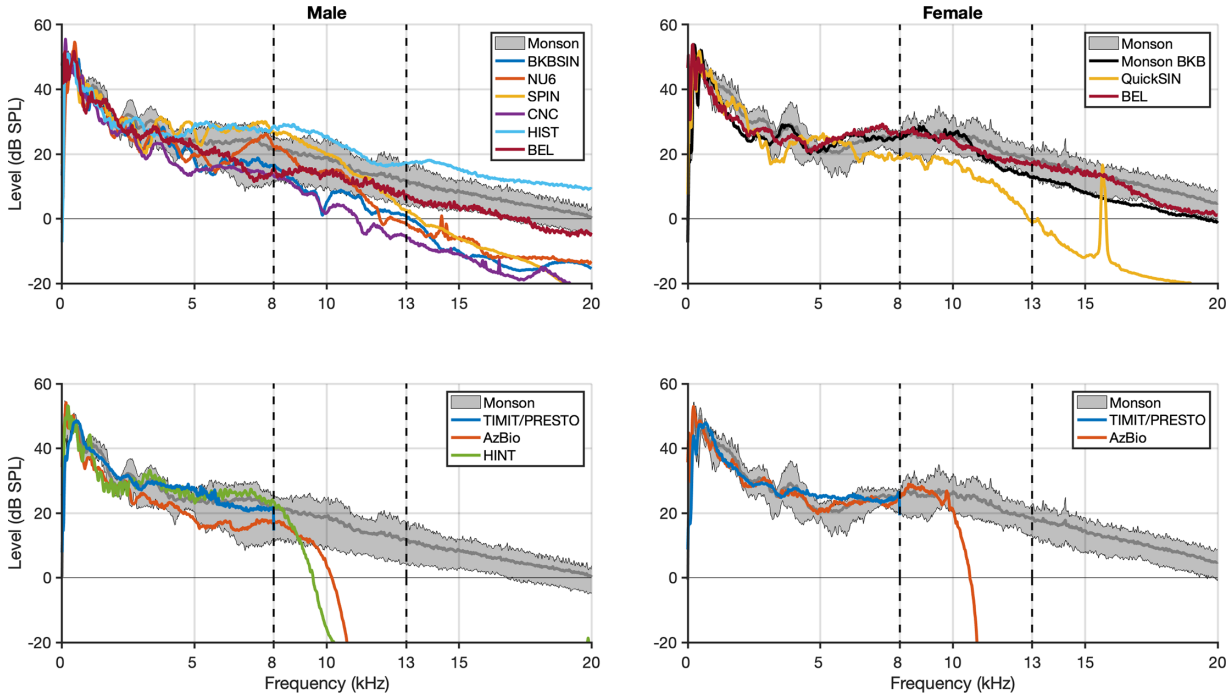

FIG. 2. (Color online) Full-band view of the LTASS for each corpus analyzed in this study. Plotting conventions follow those of Fig. 1. Dotted lines are shown at 8 kHz, representing the boundary for EHFs, and 13 kHz, representing the maximum audible low-pass filter cutoff frequency for speech for young, normal-hearing listeners (Monson and Caravello, 2019). Shading indicates the range of the Monson recordings.

levels of octave-wide spectral notches at EHFs range from 5 dB for the 8 kHz octave to approximately 10 dB for the 16 kHz octave (Monson *et al.*, 2011; Monson *et al.*, 2014b). Thus, the >10 dB differences in EHF spectral levels observed here between the Monson corpus and the other corpora are likely audible for young, normal-hearing listeners.

IV. DISCUSSION

There are circumstances in which the spectral degradations observed here may be of little practical importance. For example, EHF content is not relevant when evaluating speech perception in cochlear implant users who are mapped with input filters up to ∼8 kHz. When evaluating speech perception for cochlear implant users, the roll-off at 11 kHz for the AzBio should not affect results. Furthermore, although normal-hearing listeners derive benefit from EHF cues, providing EHF cues may not be necessary to answer questions of interest for every speech perception study, just as providing other beneficial cues (e.g., spatial cues, visual cues, etc.) is not necessary to answer every research question. Indeed, we have learned a tremendous amount about the importance of fundamental speech cues (e.g., formants, fundamental frequency) using bandlimited stimuli.

However, there are many situations in which the spectral degradations observed here could affect results. Take, for instance, the low-frequency roll-off of the TIMIT or NU-6, which appears to be a high-pass filter with cutoff frequency ∼400 Hz. Band importance functions indicate significant contributions to female speech-in-speech recognition for bands as low as 200 Hz, corresponding to the average F0 for female talkers (Buss and Bosen, 2021). The ∼10 dB reduction observed in these corpora could therefore degrade cues that listeners could otherwise use. Moreover, perceived naturalness of speech drops dramatically when speech is high-pass filtered at ∼200 Hz or higher (Moore and Tan, 2003).

One recent well-cited study used the TIMIT to demonstrate that inharmonic speech (i.e., speech with harmonics that were synthetically mistuned) was less recognizable than harmonic speech in the presence of competing speech (Popham *et al.*, 2018), suggesting harmonicity provided a modest benefit for speech-in-speech recognition. However, because this study used the spectrally degraded TIMIT, listeners had reduced access to typical cues related to harmonicity and pitch, including a ∼10 dB reduction in fundamental frequency and level reductions at the 2nd and 3rd harmonics. Thus, it is not clear whether harmonicity plays any role in speech recognition when listeners have full access to the rich cues available in full-band high-fidelity signals.

The EHF roll-offs could also affect speech perception results. Studies investigating the effects of hearing loss at standard audiometric frequencies on speech perception often include a control group of listeners with normal hearing. For example, Peters *et al.* (1998) is an oft-cited comparison of speech reception thresholds for young normal-hearing adults to those of older normal-hearing adults, young hearing-impaired adults, and older hearing-impaired adults. Because this study used the HINT speech materials, which are band limited at 8 kHz, the study design imposed an EHF

https://doi.org/10.1121/10.0013993

audibility loss on all listeners–a loss that likely disproportionately affected the young, normal-hearing group because the hearing-impaired and older listeners likely already had EHF loss. Given the value of EHF cues for young, normal-hearing listeners, the effects of hearing loss and aging on speech perception were likely underestimated based on the resulting data.

Consideration of the spectral content of speech could be particularly important for understanding hidden hearing loss. For example, Liberman *et al.* (2016) examined elevated EHF pure-tone thresholds as a marker of hidden hearing loss by testing the association between EHF thresholds and speech recognition scores but found no relationship. However, the study used the NU-6 corpus, which has substantial spectral degradations, including reduced low-frequency and EHF cues. Similarly, Smith *et al.* (2019) found no evidence for a relationship between EHF thresholds and speech-in-noise, but this study used the QuickSIN, which also has reduced EHF cues. These studies might have turned out differently if the speech corpora used to assess speech recognition had been high-quality full-bandwidth recordings (e.g., Yeend *et al.*, 2019). An association between EHF thresholds and speech recognition could reflect the utility of EHF speech cues or reflect a relationship between EHF thresholds and cochlear health at lower frequencies (Hunter *et al.*, 2020; Lough and Plack, 2022). Availability of full-bandwidth speech stimuli will be critical for differentiating these alternatives.

There are several other topics in hearing research where omission of EHF speech cues could affect outcomes. Some data indicate that the audible bandwidth requirements are broader for speech-in-speech recognition than speech-in-noise recognition (e.g., Best *et al.*, 2019), so studies of speech in a speech background may be especially sensitive to the presence of EHFs. Similarly, other data indicate talker gender differences in bandwidth requirements for speech recognition (Stelmachowicz *et al.*, 2001), suggesting that EHF cues may be more useful in female speech than male speech. Several lines of evidence suggest that young children require a wider bandwidth of audibility than adults to learn and recognize masked speech (e.g., Stelmachowicz *et al.*, 2001). Recent data from our labs suggest that speech perception is affected by differences in the EHF content when talkers are facing a listener as opposed to talkers who are not directly facing the listener. These effects of talker head orientation on EHF content are particularly pronounced when talkers are co-located on the horizontal plane (Braza *et al.*, 2022). Research to date has largely ignored these effects, so current estimates of the beneficial effects of spatial release from masking are likely to be inflated relative to those observed in natural multi-talker environments. Finally, failure to replicate the EHF content available in natural environments could undermine the naturalness of simulated auditory environments (Moore and Tan, 2003), although the consequences of reduced naturalness for other measured outcomes are not clear.

More generally, removing low-frequency or EHF spectral cues removes information that would otherwise be available to a normal-hearing listener. The practice of removing information from listeners is inherent to all lab-based auditory experiments: listeners will have lost at least some information from the loss of any real-world contextual cues, including visual cues, talker familiarity, linguistic context, natural reverberation, etc. However, bandwidth is a fundamental aspect of the acoustic signal, and it is straightforward to preserve both low-frequency and EHF energy present in natural speech using modern recording procedures and hardware.

We have focused here on the implications of using band limited vs full-band speech materials, but a final point regarding transducer frequency response is warranted. Providing EHF cues with full-band speech recordings is of little benefit if the speech is presented to listeners using transducers with poor frequency response at higher frequencies. For example, the standard transducers for audiological assessments (e.g., TDH-39 or RadioEar DD45), are only rated for stimuli ≤ 8 kHz. Just as researchers should select speech materials to faithfully represent the cues of relevance to their experiments, so too should caution be taken to select headphones and loudspeakers that will faithfully reproduce these signals with high fidelity. Many options are currently available for high-quality transducers with relatively flat frequency responses out to 20 kHz, and careful selection can help ensure full-band stimuli are reproduced as such. This consideration also applies for the selection of microphones when new speech materials are to be recorded.

V. CONCLUSIONS

The use of decades-old speech materials that are degraded and exclude what we now understand to be valuable information is a practice worthy of reconsideration. The EHF content of speech recordings could be particularly relevant for characterizing effects of hearing loss, speech-in-speech recognition, female speech recognition, spatial hearing, simulating natural listening conditions, and auditory development. For these areas of research, the choice of speech materials could have a sizable impact on the results. Multiple priorities go into selecting speech stimuli, including hypotheses of interest, consistency with other research, and convenience. Growing recognition of the importance of EHFs should also motivate greater consideration of the spectral fidelity of the speech materials used in future research, and how exclusion of EHF cues could affect experimental results. The use of corpora that exhibit both EHF and low-frequency degradations (i.e., TIMIT, QuickSIN, and NU-6) may be especially problematic, depending on the research question or questions of interest. We recommend consistent reporting of bandwidths and recording details of speech materials when publishing speech perception research. Finally, we recommend discontinuing the use of the terms "speech bandwidth" or "speech frequencies" to refer to anything other than the full frequency range of

human hearing, to avoid perpetuating the idea that valuable information in speech is restricted to a bandlimited range.

ACKNOWLEDGMENTS

The authors thank Vahid Delaram for assistance editing the recordings. This work was supported by the National Institute of Deafness and Other Communication Disorders (Grant No. R01 DC019745).

ANSI (**1997**). ANSI S3.5, *Methods for Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).

Badri, R., Siegel, J. H., and Wright, B. A. (**2011**). "Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms," J. Acoust. Soc. Am. **129**(2), 852–863.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**(3), 108–112.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**(1), 353–363.

Best, V., Roverud, E., Baltzell, L., Rennies, J., and Lavandier, M. (**2019**). "The importance of a broad bandwidth for understanding 'glimpsed' speech," J. Acoust. Soc. Am. **146**(5), 3215–3221.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech. Lang. Hear. Res. **27**(1), 32–48.

Braza, M. D., Corbin, N. E., Buss, E., and Monson, B. B. (**2022**). "Effect of masker head orientation, listener age, and extended high-frequency sensitivity on speech recognition in spatially separated speech," Ear Hear. **43**(1), 90–100.

Buss, E., and Bosen, A. K. (**2021**). "Band importance for speech-in-speech recognition," JASA Express Lett. **1**(8), 084402.

Flaherty, M., Libert, K., and Monson, B. B. (**2021**). "Extended high-frequency hearing and head orientation cues benefit children during speech-in-speech recognition," Hear. Res. **406**, 108230.

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**(2), 89–151.

Fletcher, H., and Wegel, R. L. (**1922**). "The frequency—Sensitivity of normal ears," Phys. Rev. **19**(6), 553–566.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., and Pallett, D. S. (**1993**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM," NIST speech disc 1-1.1. *NASA STI/Recon technical report n, 93*, 27403.

Gilbert, J. L., Tamati, T. N., and Pisoni, D. B. (**2013**). "Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test," J. Am. Acad. Audiol. **24**(01), 026–036.

Guinan, J. J. (**2017**). "Hearing at speech frequencies is different from what we thought," J. Physiol. **595**(13), 4123.

Hunter, L. L., Monson, B. B., Moore, D. R., Dhar, S., Wright, B. A., Munro, K. J., and Siegel, J. H. (**2020**). "Extended high frequency hearing and speech perception implications in adults and children," Hear. Res. **397**, 107922.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**(4), 2395–2405.

Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C. J., and Puria, S. (**2015**). "Extended high-frequency bandwidth improves speech reception in the presence of spatially separated masking speech," Ear Hear. **36**(5), e214–224.

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., and Maison, S. F. (**2016**). "Toward a differential diagnosis of hidden hearing loss in humans," PLoS ONE **11**(9), e0162726.

Lippmann, R. P. (**1996**). "Accurate consonant perception without mid-frequency speech energy," IEEE Trans. Speech Audio Process. **4**(1), 66–69.

Lough, M., and Plack, C. J. (**2022**). "Extended high-frequency audiometry in research and clinical practice," J. Acoust. Soc. Am. **151**(3), 1944–1955.

McCreery, R. W., and Stelmachowicz, P. G. (**2011**). "Audibility-based predictions of speech recognition for children and adults with normal hearing," J. Acoust. Soc. Am. **130**(6), 4070–4081.

McCreery, R. W., and Stelmachowicz, P. G. (**2013**). "The effects of limited bandwidth and noise on verbal processing time and word recall in normal-hearing children," Ear Hear. **34**(5), 585–591.

Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**(2), 105–129.

Mishra, S. K., Saxena, U., and Rodrigo, H. (**2021**). "Extended high-frequency hearing impairment despite a normal audiogram: Relation to early aging, speech-in-noise perception, cochlear function, and routine earphone use," Ear Hear. **43**(3), 822–835.

Mishra, S. K., Saxena, U., and Rodrigo, H. (**2022**). "Extended high-frequency hearing impairment despite a normal audiogram: Relation to early aging, speech-in-noise perception, cochlear function, and routine earphone use," Ear Hear. **43**(3), 822–835.

Monson, B. B., and Caravello, J. (**2019**). "The maximum audible low-pass cutoff frequency for speech," J. Acoust. Soc. Am. **146**(6), EL496–EL501.

Monson, B. B., Hunter, E. J., Lotto, A. J., and Story, B. H. (**2014a**). "The perceptual significance of high-frequency energy in the human voice," Front. Psychol. **5**, 587.

Monson, B. B., Hunter, E. J., and Story, B. H. (**2012b**). "Horizontal directivity of low- and high-frequency energy in speech and singing," J. Acoust. Soc. Am. **132**(1), 433–441.

Monson, B. B., Lotto, A. J., and Story, B. H. (**2012a**). "Analysis of high-frequency energy in long-term average spectra of singing, speech, and voiceless fricatives," J. Acoust. Soc. Am. **132**(3), 1754–1764.

Monson, B. B., Lotto, A. J., and Story, B. H. (**2014b**). "Detection of high-frequency energy level changes in speech and singing," J. Acoust. Soc. Am. **135**(1), 400–406.

Monson, B. B., Lotto, A. J., and Ternstrom, S. (**2011**). "Detection of high-frequency energy changes in sustained vowels produced by singers," J. Acoust. Soc. Am. **129**(4), 2263–2268.

Monson, B. B., Rock, J., Schulz, A., Hoffman, E., and Buss, E. (**2019**). "Ecological cocktail party listening reveals the utility of extended high-frequency hearing," Hear. Res. **381**, 107773.

Moore, B. C. J., Fullgrabe, C., and Stone, M. A. (**2010**). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **128**(1), 360–371.

Moore, B. C. J., Stone, M. A., Füllgrabe, C., Glasberg, B. R., and Puria, S. (**2008**). "Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss," Ear Hear. **29**(6), 907–922.

Moore, B. C. J., and Tan, C. T. (**2003**). "Perceived naturalness of spectrally distorted speech and music," J. Acoust. Soc. Am. **114**(1), 408–419.

Motlagh Zadeh, L., Silbert, N. H., Sternasty, K., Swanepoel, D. W., Hunter, L. L., and Moore, D. R. (**2019**). "Extended high-frequency hearing enhances speech perception in noise," Proc. Natl. Acad. Sci. USA **116**(47), 23753–23759.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**(2), 1085–1099.

Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**(1), 577–587.

Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear. Disord. **27**(1), 62–70.

Pollack, I. (**1948**). "Effects of high pass and low pass filtering on the intelligibility of speech in noise," J. Acoust. Soc. Am. **20**(3), 259–266.

Polspoel, S., Kramer, S. E., van Dijk, B., and Smits, C. (**2022**). "The importance of extended high-frequency speech information in the recognition of digits, words, and sentences in quiet and noise," Ear Hear. **43**(3), 913–920.

Popham, S., Boebinger, D., Ellis, D. P., Kawahara, H., and McDermott, J. H. (**2018**). "Inharmonic speech reveals the role of harmonicity in the cocktail party problem," Nat. Commun. **9**(1), 1–13.

Rimikis, S., Smiljanic, R., and Calandruccio, L. (**2013**). "Nonnative English speaker performance on the Basic English Lexicon (BEL) sentences," J. Speech. Lang. Hear. Res. **56**(3), 792–804.

https://doi.org/10.1121/10.0013993

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

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes T., and Cook, S. (**2012**). "Development and validation of the AzBio sentence lists," Ear Hear. **33**(1), 112–117.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (**2001**). "Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults," J. Acoust. Soc. Am. **110**(4), 2183–2190.

Trine, A., and Monson, B. B. (**2020**). "Extended high frequencies provide both spectral and temporal information to improve speech-in-speech recognition," Trends Hear. **24**, 2331216520980299.

Vitela, A. D., Monson, B. B., and Lotto, A. J. (**2015**). "Phoneme categorization relying solely on high-frequency energy," J. Acoust. Soc. Am. **137**(1), EL65–EL70.

Yeend, I., Beach, E. F., and Sharma, M. (**2019**). "Working memory and extended high-frequency hearing in adults: Diagnostic predictors of speech-in-noise perception," Ear Hear. **40**(3), 458–467.