The maximum audible low-pass cutoff frequency for speech

Brian B. Monson[a] **and Jacob Caravello**[a]
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, 901 South Sixth Street, Champaign, Illinois 61820, USA
monson@illinois.edu, jacobpc2@illinois.edu

Abstract: Speech energy beyond 8 kHz is often audible for listeners with normal hearing. Limits to audibility in this frequency range are not well described. This study assessed the maximum audible low-pass cutoff frequency for speech, relative to full-bandwidth speech. The mean audible cutoff frequency was approximately 13 kHz, with a small but significant effect of talker sex. Better pure tone thresholds at extended high frequencies correlated with higher audible cutoff frequency. These findings demonstrate that bandlimiting speech even at 13 kHz results in a detectable loss for the average normal-hearing listener, suggesting there is information regarding the speech signal beyond 13 kHz.

[QJF]
Date Received: August 15, 2019 **Date Accepted:** November 22, 2019

1. Introduction

For humans, the frequency range of hearing extends to approximately 20 kHz. It is widely believed that extended high frequencies (EHF) (frequencies ≥8 kHz) have little role for speech perception, being beyond the traditional "speech bandwidth." This view may be due, in part, to the mixed findings from studies that have examined whether there are benefits with audibility of high-frequency bands, some of which have included energy at EHFs (Stelmachowicz *et al.*, 2001; Stelmachowicz *et al.*, 2007; Pittman, 2008; Ricketts *et al.*, 2008; Fullgrabe *et al.*, 2010; Moore *et al.*, 2010; McCreery and Stelmachowicz, 2011, 2013; Levy *et al.*, 2015). Consequently, whereas the effects of audibility of frequency regions below 8 kHz for speech have been studied extensively (Dubno *et al.*, 1989; Ching *et al.*, 1998; McCreery and Stelmachowicz, 2011; McCreery *et al.*, 2013), the limit of audibility of the highest frequencies in the speech spectrum is not well understood.

At least some portion of EHF energy in speech is audible for normal hearing listeners. Moore and Tan (2003) demonstrated this when examining subjective judgments of speech quality. Listeners rated speech low-pass filtered at 7 kHz as less natural than speech low-pass filtered at approximately 10.9 kHz, suggesting energy between 7 and 10.9 kHz was audible. No differences were observed between speech low-pass filtered at 10.9 and 16.9 kHz, however, making it unclear whether EHF energy beyond 10.9 kHz was audible. Audibility of EHF speech energy was also demonstrated by Best *et al.* (2005) when using a speech localization task. They found that low-pass filtering speech at 8 kHz significantly increased front-back confusions, compared with low-pass filtering speech at 16 kHz, indicating EHF energy between 8 and 16 kHz was audible. Similarly, we recently found that EHFs were audible when we demonstrated that, compared to full-band speech (cutoff frequency of 22.05 kHz), low-pass filtering speech at 8 kHz substantially reduced listeners' ability to detect changes in the physical rotation (i.e., facing angle) of a talker's head (Monson *et al.*, 2019).

Speech intelligibility experiments have also demonstrated EHF audibility. Lippmann (1996) provided evidence that EHFs were audible when examining consonant recognition scores. Participants scored 44% correct for speech low-pass filtered at 800 Hz, but improved to 74% correct when 8-kHz high-pass filtered speech was added to the 800-Hz low-pass filtered speech signal. We again found EHFs to be audible by low-pass filtering speech at 8 kHz, which led to significantly poorer speech-in-speech recognition scores relative to full-band speech (Monson *et al.*, 2019).

Although these studies clearly demonstrate that some EHF energy in speech is audible, the limit of EHF audibility is unclear from the published data. Is speech energy beyond 10.9 kHz audible? At what cutoff frequency does low-pass filtering

[a]Also at: Neuroscience Program, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA.

speech materials result in a detectable loss? These questions are of interest because, as evidence regarding the ecological utility of EHF information continues to accrue, examination of EHF spectral detail will be warranted. However, study of speech spectral detail at or near 20 kHz may not be justified if it is inaudible for normal hearing listeners. For example, we previously found that more than 30% of listeners could not detect the complete attenuation of the 16-kHz octave band in speech (11.3–22.05 kHz), suggesting EHF energy beyond 11.3 kHz was inaudible for some listeners.

To better understand the limit of EHF audibility in speech, we sought to determine the maximum audible cutoff frequency for low-pass filtered speech for young, normal-hearing listeners. We hypothesized that this frequency would be higher for female speech than male speech, due to the elevated EHF levels for female speech relative to male speech (Monson *et al.*, 2012b). We also assessed whether pure tone thresholds at EHFs predicted an individual listener's maximum audible cutoff frequency.

2. Methods

Statistical analyses consisted of a one-way repeated-measures analysis of variance (ANOVA) and Pearson's correlation, as described below. All statistical analyses were conducted in R (R Core Team, 2018) using the *ezANOVA* and *correlate* functions. Signal processing and experiment presentation was implemented using MATLAB (MathWorks). All recording materials and data for this study will be made available upon request.

2.1 Participants

Twenty-one participants (seven male, age 19–27 years) participated in this experiment. Audiometric thresholds for each ear were obtained with pulsed pure tone stimuli presented over Sennheiser HDA 200 circumaural headphones, at frequencies of 0.5, 1, 2, 4, 8, 9, 10, 11.2, 12.5, 14, and 16 kHz. EHF test frequencies correspond to 1/6 octaves and were calibrated according to ISO 389-5 (ISO, 2006). An extended high-frequency pure tone threshold average for each ear was calculated as the average of thresholds for all frequencies measured from 8 to 16 kHz. Participants had normal hearing, defined as pure tone audiometric thresholds better than 20 dB hearing level (HL) in at least one ear for all frequencies between 0.5 and 16 kHz, and no history of hearing disorder. Participants for all experiments provided written informed consent. All study procedures were approved by the Institutional Review Board at the University of Illinois at Urbana-Champaign.

2.2 Stimuli

The stimuli were recordings of the phrase "amend the slower page" uttered by two male and two female talkers, taken from a database of high-fidelity (44.1-kHz sampling rate, 16-bit precision, type 1 precision microphone) anechoic recordings (Monson *et al.*, 2012a) (Fig. 1). This utterance was selected because it is phonetically representative, containing at least one voiceless fricative, voiced fricative, affricate, nasal, liquid, vowel, and plosive. Stimuli for the experimental task were generated by low-pass filtering each recording with an equiripple Parks-McClellan finite impulse response filter (order 100, 1-kHz transition band) with cutoff frequencies spanning 6 to 22 kHz in 1-kHz increments.

2.3 Procedure

Stimuli were presented to listeners seated in a sound-treated booth over a Mackie HR624mk2 loudspeaker at 70 dB sound pressure level (SPL) at 1 m. An adaptive one-up two-down three-alternative forced-choice oddity task was used to measure thresholds for detecting the difference between the reference signal consisting of a 22-kHz low-pass filtered recording, and a test signal consisting of the same recording filtered with a lower cutoff frequency. The procedure was implemented using the AFC software package (Ewert, 2013). Four tracks (one for each talker) were interleaved in a single adaptive run, with track selection randomized for each trial. A brief training block consisted of a one-up one-down two-reversal run for each talker. Following the training block, the experimental run began with an easily detectable difference (8-kHz cutoff). The low-pass filter cutoff frequency step size changed from 3 to 1 kHz after the first two reversals. The frequencies of last six reversals were averaged to obtain the maximum audible cutoff frequency. Feedback on accuracy was provided for the training block, but not for the experimental block.

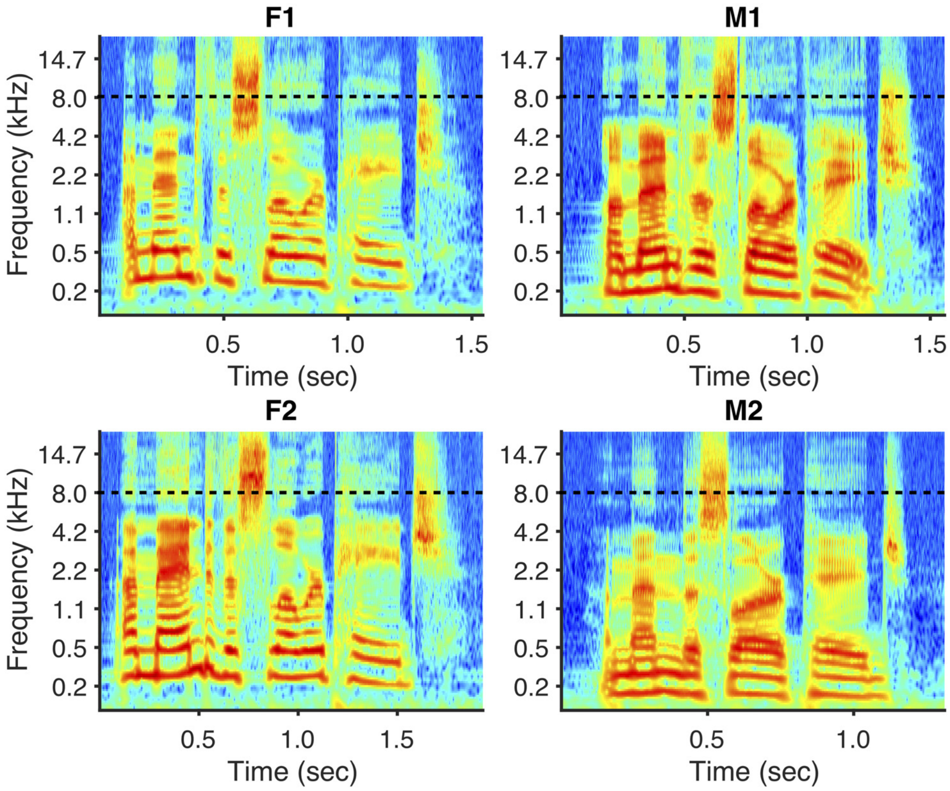

Fig. 1. (Color online) Cochleagrams of speech. The phrase "amend the slower page" was spoken by two female (F1 and F2) and two male (M1 and M2) talkers. Appreciable energy and acoustic structure are apparent beyond 8 kHz. To produce a perceptually relevant representation, data were plotted on a cochlear [equivalent rectangular bandwidth (ERB)] frequency scale, using 1-ERB-wide filter bands with 50% overlap (Glasberg and Moore, 1990; McDermott and Simoncelli, 2011). Transfer functions of the outer and middle ear were not incorporated.

3. Results

Average maximum audible cutoff frequencies were 13.1 and 12.8 kHz for female and male speech, respectively, with a small but significant effect of talker sex [$F(1,20) = 4.88$, $p = 0.04$] [Fig. 2(A)]. There was large between-subject variability in EHF pure tone thresholds for our listeners, who all had clinically normal hearing in at least one ear. Whereas standard deviations (SDs) in pure tone thresholds at individual standard audiometric frequencies (0.5–8 kHz) ranged between 4.7 and 8.5 dB, SDs for thresholds at individual EHF frequencies were almost always >9 dB, ranging up to 20 dB for 16 kHz [Figs. 2(B) and 2(C)]. There was a relationship between listener task performance (maximum audible cutoff frequency averaged across talkers) and EHF pure tone threshold average for the better ear (Pearson's $r = -0.52$, $p = 0.02$). There was a slightly stronger relationship between task performance and pure tone threshold at 16 kHz for the better ear ($r = -0.6$, $p = 0.005$) [Fig. 2(D)]. The lowest maximum audible cutoff frequency value across talkers and listeners was 10 kHz, whereas the highest value was 15 kHz. Figure 2(E) shows the proportion of listeners that could distinguish between full band speech and low-pass filtered speech as cutoff frequency increased (based on each listener's mean cutoff frequency).

4. Discussion

A particular species' and individual's auditory system develops sensitivity to the sounds that matter most for survival and success, including the spectral detail of vocalizations of predators, prey, and mates (Hoy, 1992; Hauser, 1996; Manley, 2017). Our results demonstrate that, on average, human listeners can detect the removal of a band of speech between 13 kHz and the upper limit of human hearing (approximately 20 kHz). That humans display sensitivity to this 7-kHz-wide band raises the possibility that it may provide useful information about the speech signal. The nature of this information and its potential benefit remain unclear, but multiple uses of EHF energy for speech perception have been demonstrated previously (Lippmann, 1996; Best *et al.*, 2005; Vitela *et al.*, 2015; Monson *et al.*, 2019).

We observed substantial variability in EHF pure tone thresholds for our normal hearing listeners, as has been reported by others (Yeend *et al.*, 2019). We also

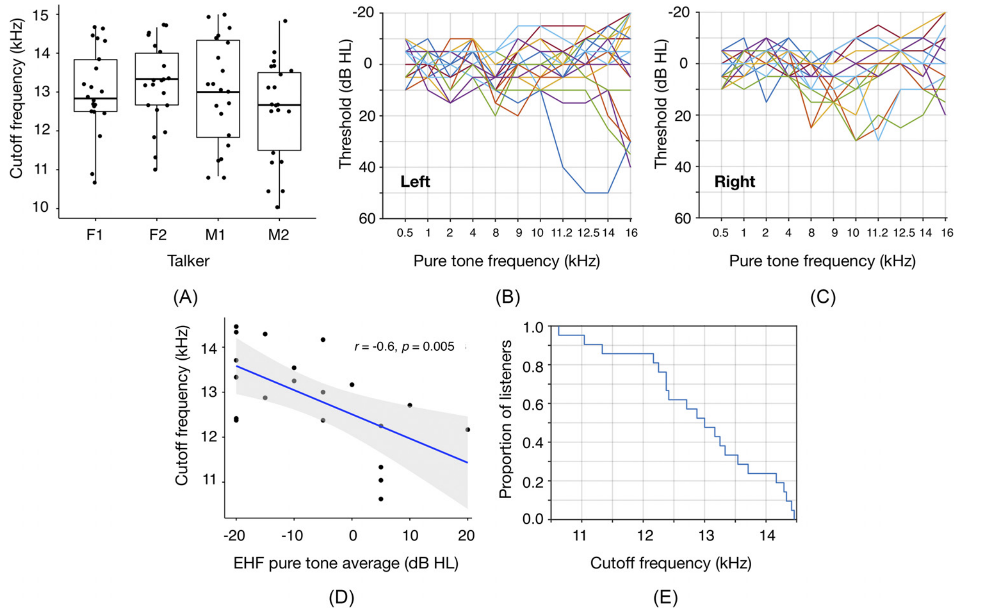

Fig. 2. (Color online) (A) Maximum audible low-pass filter cutoff frequencies for speech uttered by two female and two male talkers. The average threshold across all four talkers was approximately 13 kHz, indicating that listeners could detect the loss of the band of speech between 13 and 22 kHz. Whiskers represent the range. (B) Pure tone thresholds for the left ear for all listeners. (C) Pure tone thresholds for the right ear for all listeners. (D) Relationship between listeners' pure tone threshold at 16 kHz for the better ear and mean maximum audible cutoff frequency (averaged across talkers). (E) Proportion of listeners whose mean maximum audible cutoff frequency equaled or exceeded the cutoff frequency indicated on the abscissa.

observed a relationship between EHF pure tone detection thresholds in the better ear and the maximum audible cutoff frequency. This observation indicates that greater acuity at EHFs enables improved detection of EHF energy in speech. This is not surprising, but it is noteworthy that this relationship was observed within a group of young listeners with clinically normal hearing and relatively low variability in pure tone thresholds at standard audiometric frequencies. This result raises the possibility that individuals with better EHF pure tone thresholds (e.g., children) may have better access to information provided in the EHF range. Conversely, individuals with EHF hearing loss, which is typical for middle-age and older adults (Green *et al.*, 1987; Stelmachowicz *et al.*, 1989), lose access to the information provided by EHF energy in speech.

One reason that audibility of EHF bands in speech has not been more closely examined before now may be that seminal speech and hearing studies were focused on minimizing the bandwidth for transmission of speech over communication systems, while maintaining intelligibility (Crandall and MacKenzie, 1922; Fletcher and Steinberg, 1930; Fletcher and Galt, 1950; Monson *et al.*, 2014). These efforts gave rise to the articulation index (French and Steinberg, 1947; Fletcher and Galt, 1950), later refined and renamed the speech intelligibility index, which typically assigns little weight to EHF bands in speech. Although the earliest studies were limited by transducers and electronic equipment with poorer response at EHFs, more recent assessment using the speech intelligibility index has included speech energy up to 11 kHz (McCreery and Stelmachowicz, 2011). Our data show that speech energy beyond 13 kHz is audible, suggesting inclusion of energy beyond 13 kHz in speech intelligibility index calculations may serve to refine that method.

Our results indicate that bandlimiting speech even at 13 kHz results in a detectable loss of fidelity for the average young, normal-hearing listener. Because our speech stimulus was limited to a single utterance containing one voiceless fricative, it is possible that the audible cutoff frequency could increase using utterances with more voiceless fricatives, as voiceless fricatives have substantial energy at EHFs (Monson *et al.*, 2012b). Our findings have implications for cell phones, HD voice, hearing aids, and other communication applications which typically do not transmit sounds in this range. For example, our results suggest that speech energy beyond 13 kHz should be included to achieve true HD voice. Our data reveal that nearly 80% of listeners found speech low-pass filtered at 14 kHz to be indistinguishable from full-band speech, whereas almost all listeners failed to distinguish the two when the cutoff frequency was 14.5 kHz. These are useful benchmarks when considering design of communication systems intended to replicate speech signals with high fidelity. Finally, speech energy beyond 13 kHz may warrant further investigation in general as it is apparently audible to normal hearing listeners.

Acknowledgments

The authors thank Olivia Godnik, Melanie Flores, Elana Hunt, and Molly Cull for assistance with data collection.

References and links

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

Ching, T. Y. C., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Crandall, I. B., and MacKenzie, D. (**1922**). "Analysis of the energy distribution in speech.," Phys Rev. **19**, 221–232.

Dubno, J. R., Dirks, D. D., and Schaefer, A. B. (**1989**). "Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II: Articulation index predictions," J. Acoust. Soc. Am. **85**, 355–364.

Ewert, S. D. (**2013**). "AFC—A modular framework for running psychoacoustic experiments and computational perception models," in *Proceedings of the International Conference on Acoustics AIA-DAGA*, pp. 1326–1329.

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

Fletcher, H., and Steinberg, J. C. (**1930**). "Articulation testing methods," J. Acoust. Soc. Am. **1**, A1–A48.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Fullgrabe, C., Baer, T., Stone, M. A., and Moore, B. C. (**2010**). "Preliminary evaluation of a method for fitting hearing aids with extended bandwidth," Int. J. Audiol. **49**, 741–753.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M., Kidd, G., Jr., and Stevens, K. N. (**1987**). "High-frequency audiometric assessment of a young adult population," J. Acoust. Soc. Am. **81**, 485–494.

Hauser, M. D. (**1996**). *The Evolution of Communication* (MIT press, Cambridge, MA).

Hoy, R. R. (**1992**). "The evolution of hearing in insects as an adaptation to predation from bats," in *The Evolutionary Biology of Hearing* (Springer, New York, NY), pp. 115–129.

ISO (**2006**). ISO 398-5:2006, "Acoustics-Reference zero for the calibration of audiometric equipment-Part 5: Reference equivalent threshold sound pressure levels for pure tones in the frequency range 8 kHz to 16 kHz" (International Organization for Standardization, Geneva).

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**, e214–e224.

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

Manley, G. A. (**2017**). "Comparative auditory neuroscience: Understanding the evolution and function of ears," J. Assoc. Res. Otolaryngol.: JARO **18**, 1–24.

McCreery, R. W., Brennan, M. A., Hoover, B., Kopun, J., and Stelmachowicz, P. G. (**2013**). "Maximizing audibility and speech recognition with nonlinear frequency compression by estimating audible bandwidth," Ear Hear. **34**, e24–e27.

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**, 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**, 585–591.

McDermott, J. H., and Simoncelli, E. P. (**2011**). "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," Neuron **71**, 926–940.

Monson, B. B., Hunter, E. J., Lotto, A. J., and Story, B. H. (**2014**). "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. (**2012a**). "Horizontal directivity of low- and high-frequency energy in speech and singing," J. Acoust. Soc. Am. **132**, 433–441.

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

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**, 360–371.

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

Pittman, A. L. (**2008**). "Short-term word-learning rate in children with normal hearing and children with hearing loss in limited and extended high-frequency bandwidths," J. Speech Lang. Hear. Res. **51**, 785–797.

R Core Team (**2018**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

Ricketts, T. A., Dittberner, A. B., and Johnson, E. E. (**2008**). "High-frequency amplification and sound quality in listeners with normal through moderate hearing loss," J. Speech Lang. Hear. Res. **51**, 160–172.

Stelmachowicz, P. G., Beauchaine, K. A., Kalberer, A., Kelly, W. J., and Jesteadt, W. (**1989**). "High-frequency audiometry: Test reliability and procedural considerations," J. Acoust. Soc. Am. **85**, 879–887.

Stelmachowicz, P. G., Lewis, D. E., Choi, S., and Hoover, B. (**2007**). "Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children," Ear Hear. **28**, 483–494.

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**, 2183–2190.

Vitela, A. D., Monson, B. B., and Lotto, A. J. (**2015**). "Phoneme categorization relying solely on high-frequency energy," J. Acoust. Soc. Am. **137**, 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**, 458–467.