



Cross-language comparison of long-term average speech spectrum and dynamic range for three Indian languages and British English

Vijaya Kumar Narne, Nachiketa Tiwari

Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India

Purpose: The Long-Term Average Speech Spectrum (LTASS) and Dynamic Range (DR) of speech strongly influence estimates of Speech Intelligibility Index (SII), gain and compression required for hearing aid fitting. It is also known that acoustic and linguistic characteristics of a language have a bearing on its LTASS and DR. Thus, there is a need to estimate LTASS and DR for Indian languages. The present work on three Indian languages fills this gap and contrasts LTASS and DR attributes of these languages against British English.

Methods: For this purpose, LTASS and DR were measured for 21 one-third octave bands in the frequency range of 0.1 to 10 kHz for Hindi, Kannada, Indian English and British English.

Results: Our work shows that the DR of Indian languages studied is 7-10 dB less relative to that of British English. We also report that LTASS levels for Indian languages are 7 dB lower relative to British English for frequencies above 1 kHz. Finally, we observed that LTASS and DR attributes across genders were more or less the same.

Conclusions: Given the evidence presented in this work that LTASS and DR characteristics for Indian languages analyzed are markedly different than those for BE, there is a need to determine Indian language specific SII, as well as gain and compression parameters used in hearing aids.

Keywords: Long-average speech spectrum, Dynamic range, Integration time, Dynamic range definition, Phoneme distribution



Received: July 3, 2021

Revision: August 31, 2021

Accepted: August 31, 2021

Correspondence:

Nachiketa Tiwari

Department of Mechanical Engineering,
Indian Institute of Technology, Kanpur,
India

Tel: +9559270306

Fax: +9559270306

E-mail: ntiwari@iitk.ac.in

INTRODUCTION

The Long-term Average Speech Spectrum (LTASS) and Dynamic Range (DR) of speech strongly influence estimates of Speech Intelligibility Index (SII) [1], and also prescribed gain and compression required for hearing aid [2-4]. It is also known that acoustic and linguistic characteristics of a language have a bearing on LTASS and DR. Thus, the SII [5,6] as well as optimal gain and compression required for hearing aids can be language dependent [7-9]. Unfortunately, not many studies have been conducted to estimate LTASS and DR for Indian languages, which are spoken by over 1.3 billion people. The present study addresses this need by measuring LTASS and DR for Kannada (KL) which is spoken in a part of Southern India, Hindi (HL) which is spoken by majority of North Indians, and English as spoken by Indians, termed as Indian English (IE). For purpose of comparison similar measurements were also made for British English (BE).

© 2021 The Korean Association of Speech-Language Pathologists

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Long term average speech spectrum

The LTASS is a frequency-dependent measure of time-averaged sound pressure level for speech. It provides a global representation of acoustic characteristics of continuous speech [2]. There are many factors that influence the LTASS, and language is a very important factor among these [10-12]. Byrne et al. [13] studied 13 different languages and observed that the differences in LTASS levels across languages do not exceed 3dB. They attributed such small differences to variations in the frequency of occurrence of different phonemes across languages. Based on such an understanding, they recommended the use of a universal i.e., language-independent LTASS for different applications. However, there is no hard data to show conclusively that such differences are indeed not significant, especially for languages which were not covered in Byrne's study. In fact Noh and Lee [10] showed that LTASS levels for Korean were 5-8 dB lesser relative to English at frequencies above 1.6 kHz.

We would also like to point out that there are several studies showing that Indian languages differ significantly from English in terms of phoneme distribution and acoustic properties. Differences in phoneme distribution have been reported for Hindi [14], Kannada [15] and several other Indian languages [16]. Further, acoustic properties of Indian phonemes are also markedly different vis-à-vis English, as reported by Kochetov and Sreedevi [17] and Tabain and Kochetov [18]. Such differences may be a source of a marked difference in LTASS for Indian languages relative to English.

However, there is only one study [19] which explores LTASS for some Indian languages; Malayalam, Hindi and Kannada. Among these, Malayalam was reported to have maximum energy relative to English in the 0.25-1 kHz band, in higher bands LTASS values were more or less the same. However, their LTASS estimates cannot be compared with those reported by Byrne et al. [13] for two reasons. One, Mili et al. [19] did not normalize their signal to normal conversational level and two, they measured LTASS levels in octave bands instead of one-third octave bands. Therefore, Mili's study suggests that there is indeed a strong need for a detailed study of LTASS for Indian languages.

Dynamic range

The dynamic range of speech (DR) is often defined as the difference between its maximum and minimum levels. It plays an important role in design of algorithms for predicting speech intelligibility, prescribing gain for hearing aids, design

and selection of acoustic transducers for mobile phones and head-sets, and also for compressing signal for various applications.

The DR of speech is often estimated in 1/3rd octave bands. It is calculated by breaking the speech sample over a series of small time windows (integration time), determining the speech level for each window, and finally computing the DR for the signal. Such computations are performed for different frequency bands. The DR of speech is affected by several factors; integration time [20-22], DR definition [21,22], and frequency band [3,13,20-23].

Byrne et al. [13] measured DR of speech, using a 99-1% definition, for 13 different languages and reported a frequency averaged DR of 30 dB across frequencies for all languages, sans Japanese (36 dB) and Australian English (26 dB). They also reported that DR levels across frequency bands varied by as much as 7 dB. In 2014, Jin et al. estimated DR values for three languages; Korean, Mandarin, and English. They reported that the DR for Korean is 11 dB less at lower frequencies, and 10 dB more at mid and high frequencies relative to English. They attributed such differences to variations of phoneme distributions between Korean and English. These works show that DR values can be remarkably different across languages. Thus, there is a need to characterize Indian languages in terms of DR as well.

Thus the present study had two objectives; to analyse differences in LTASS across three Indian languages (KL, HL, and IE) and BE, and to analyse differences in DR for these languages corresponding to different combinations of DR definitions, integration times, and frequency bands.

METHODS

Talkers

We recorded speech samples from 6 males and 6 females in each of the three languages studied. Further, British English samples were the same as those used by Moore et al. [3]. All the participants were native speakers of their respective languages and no one had any apparent speech abnormalities. Their age ranged between 20 and 35 years. The mean age of participants was 25.4 years across all Indian languages. BE speakers were aged between 21 and 53 years with a mean of 36 years.

Speech Material

Participants were asked to read a standardized test passage developed at All India Institute of Speech and Hearing, My-

sore in normal conversational style. A majority of speakers took about 90 seconds to read the text passage. For BE, the corresponding duration was 55 seconds.

Instrumentation and procedure for recording

For Indian languages, samples were recorded in an acoustically treated room with ambient noise level less than 30 dBA. For this a recording microphone (B & K, 4189) was placed about 20 cm away from the speaker. The microphone was at the level of speaker's mouth and was placed in front of it. The speech material was placed 30 cm away from the speaker. It was placed below the microphone level to avoid any possible reflections. Speakers read texts at normal speed and loudness levels. The sound level was also monitored through a sound level meter. Signals from the microphone were digitally recorded at a sampling frequency of 48 kHz using 24-bit A/D converter (NI-9234) in MATLAB® environment.

For BE, recordings were made in a sound-treated room with a microphone (Sennheiser, MKH 40 P 48 U3). All recordings were made at a sampling rate of 44.1 kHz. More details about recording equipment, procedure and room for BE samples are provided in Moore et al. [3].

Acoustical analysis procedure

All recordings were edited in Audacity® to remove extraneous sounds and pauses for breath. However, pauses between sentences were retained. Any low-frequency noise attributable to building vibrations and ventilation system was also filtered out through use of a fourth order high-pass elliptic filter with 0.01 dB of passband ripple and 50 dB of stopband attenuation with edge frequency of 70 Hz.

Next, LTASS and DR were calculated. For this, first the overall level of all recorded samples was scaled to 65 dB which corresponds to normal conversation level [24]. Next, the sample was filtered for 21 one-third octave bands. The filter attributes were similar to those used by Moore et al. [3]. Next, band specific LTASS was calculated. For determining DR as a function of frequency, each band specific time series data was passed through a moving average filter by convolving with a T ms long Hanning window in steps of T/2 millisecond. Here, T corresponds to integration time. For each such window, the average value of data in its first and second halves were calculated. Next, the band-specific DR values were calculated using different DR definitions. For each speech sample DR was estimated for 10 integration times (T = 1, 2, 4, 8, 16, 32, 64, 128, 256, and 512 ms) corresponding to three different definitions; 99–

20%, 99–10%, and 99–1%. Thus, for each language we computed 30 different estimates of DR corresponding to each one-third octave band. Finally, the overall dynamic range (DR_o) for each language was calculated by averaging values of DR across 21 one-third octave bands.

Statistical Analysis

Shapiro-Wilks tests for LTASS and DR indicated that these data did not exhibit normal distribution. Thus, non-parametric tests (Kruskal-Wallis test, and Dunn's test) were used to assess presence of significant differences in DR and LTASS across languages, if any. All statistical analyses were performed using R software [25].

Results and Discussion

Long-term average speech spectrum

Figure 1 is a comparison of LTASS levels for BE as measured in this study with data presented by Byrne et al. [13] and Moore et al. [3]. The figure shows that our results agree very well with earlier data. Such consistency establishes the validity of our analysis procedure.

Figure 2 shows mean LTASS levels for all languages studied in this work for male and female speakers, respectively. It can be observed from the Figure that LTASS levels for male speakers vary between 48 and 55 dB for frequencies up to 0.5 kHz regardless of languages. Beyond 0.5 kHz we note that the LTASS level for BE starts falling monotonically roughly at a rate 4.0 dB/octave and becomes 35 dB between 6.3 and 10 kHz. We also observe that among Indian languages, LTASS curve for KL exhibits the steepest decline rate (7.1 dB/octave) and the LTASS value becomes 28 dB between 6.3 and 10 kHz, while the same for HL and IE decline at a rate of 5.6 dB/octave. Thus, LTASS curves for HL and IE lie between those corresponding to KL and BE. Such trends for HL and KL are very similar to those reported by Mili et al. [19]. However, comparisons of absolute LTASS levels between these two studies cannot be made as Mili's analysis procedure was significantly different vis-à-vis ours.

We also note that LTASS for female speakers is more or less similar to that for males with two notable exceptions. One, for frequencies below 0.16 kHz LTASS levels are significantly lesser for females vis-à-vis male speakers. Two, the decay rate of LTASS levels beyond 0.5 kHz for female speakers is somewhat lesser than those for males. Both these observations are consistent with earlier studies as shown in Figure 1, and the reasons for such differences have been well documented ear-

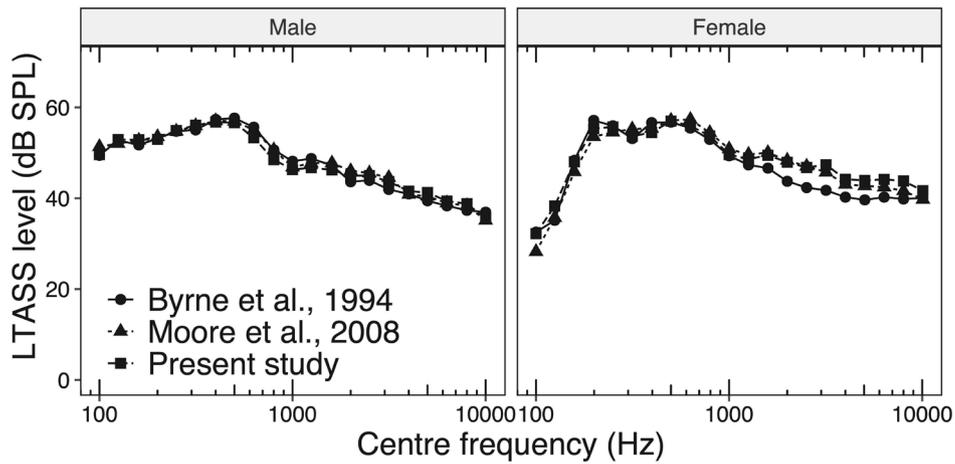


Figure 1. Comparison of LTASS as determined in present study for BE with those reported by Moore et al. (2008) and Byrne et al. (1994).

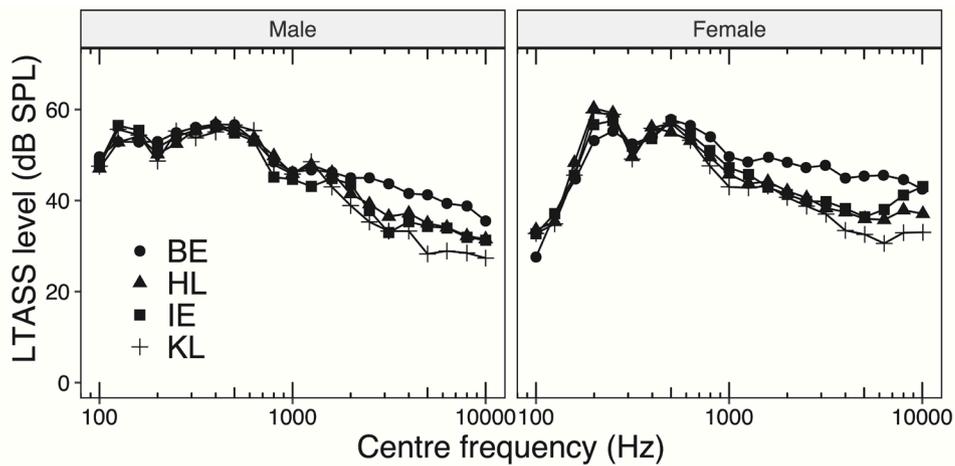


Figure 2. LTASS as a function of frequency.

lier [2,3,10,12,13,26].

Figure 2 also shows that LTASS levels of BE are 5-10 dB higher ($p < 0.05$) at frequencies above 1 kHz relative to HL and KL. Such differences in LTASS levels occur because HL and KL tend to use phonemes with high-frequency energy such as /d/, /t/, /tʃ/, /dʒ/, /f/, and /s/ far less often relative to English as shown in Table 1. We have also noted that LTASS levels for BE are appreciably higher than those for IE. Such differences could be attributable to the influence of speakers' native accents on acoustical properties of their oral renditions of passages from a non-native language [27-29].

Statistical analysis of our data shows that the standard deviation for LTASS across speakers ranges from 1.5 to 2.5 dB, and 3 to 4.7 dB for male and female speakers, respectively across languages. These values are consistent to those reported by Moore et al. [3].

Table 1. Occurrence of phonemes having dominant energy at higher frequency bands for English, Hindi and Kannada

Phonemes	English ^{1,2}	Hindi ³	Kannada ⁴
/d/	3.33%	1.51%	2.41%
/t/	5.78%	3.38%	1.61%
/f/	1.55%	-Nil-	-Nil-
/tʃ/	0.31%	0.61%	0.36%
/z/	2.75%	1.71%	0.44%
/s/	4.61%	3.76%	1.72%
/j/	0.49%	1.21%	0.61%

¹Mines et al. (1978); ²Tobias (1959); ³Ramakrishna et al. (1992); ⁴Sreedevi et al. (2012).

Dynamic Range

Figure 3 is a comparison of values of overall dynamic range (DR_o) of BE in male speakers as computed in this study, with

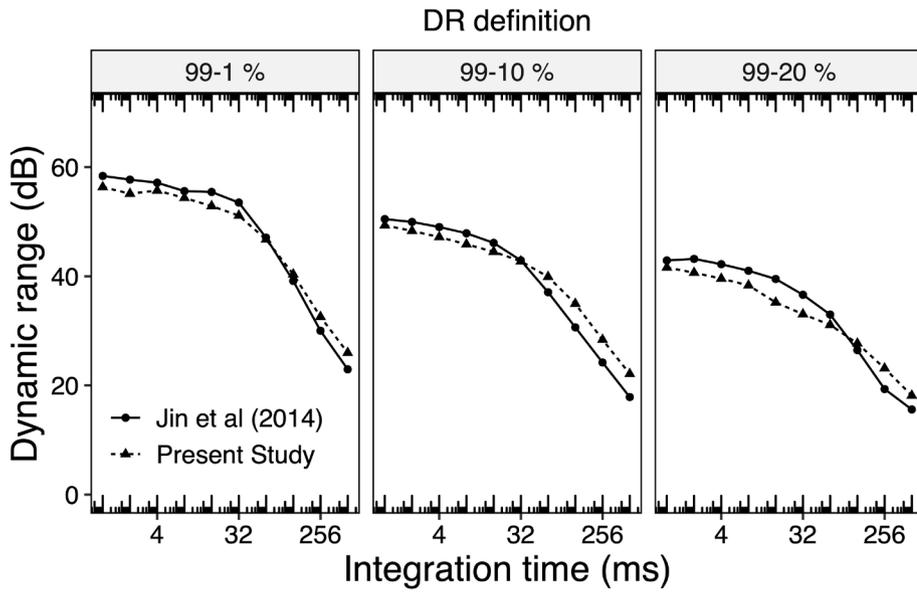


Figure 3. Comparison of overall DR of BE as detailed in this work and that by Jin et al. (2014).

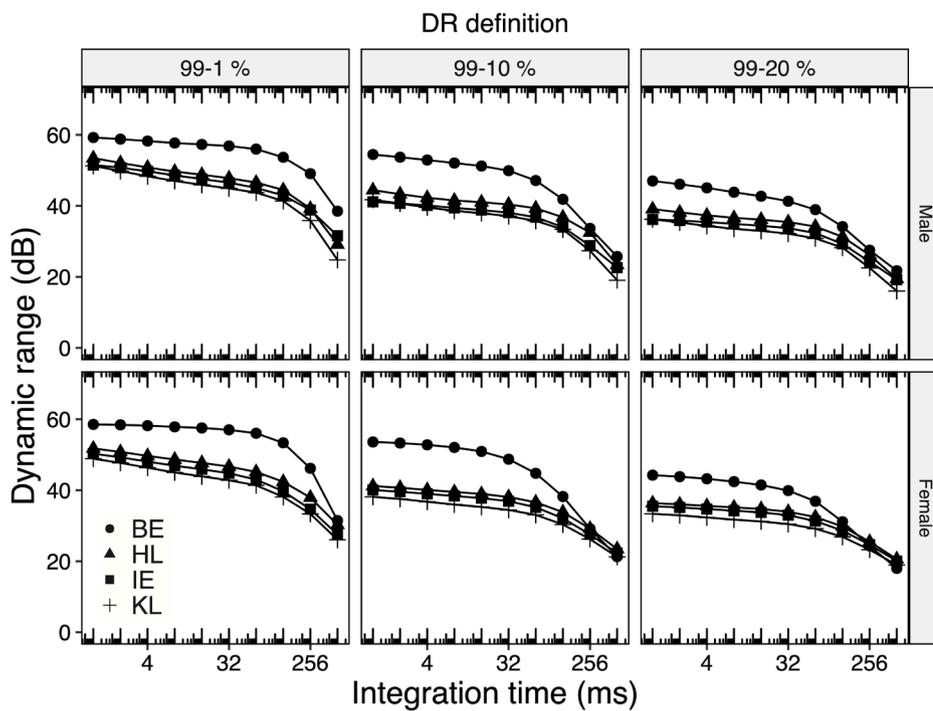


Figure 4. Overall DR as a function of integration times and DR definitions.

those reported by Jin et al. [21]. For purpose of comparison with Jin et al. [21], all the silence present between sentences in the samples of BE, were removed. The figure shows that our results are in reasonable agreement with those of Jin et al., corresponding to different combinations of integration times and DR definitions.

Figure 4 depicts values of DR_0 for all languages considered in this study as a function of different integration times and DR definitions. It is seen from the figure that DR_0 for all Indian languages is lesser than that for BE by as much as 10 dB especially when the integration time is 64 ms or less. We also note that such a difference approaches 0 dB as the integration time

rises to 512 ms. Such a difference of 10 dB between the DR₀ for Indian languages and BE was significantly larger than corresponding standard deviation.

We attribute such a marked difference in DR₀ to the fact that Indian languages tend to have more words ending with vowels and thus have lesser silence durations between words. Specifically, we found that the total duration of silence in BE, HL, and KL, were 9.1%, 5.8%, and 6.1%, respectively. This is consistent with observations made by Varnet et al. [30], who reported on modulation spectrum for different languages. They noted that languages with higher Amplitude Modulation Depth (AMD), e.g., English, tend to have shorter word durations interspersed with a larger number of silences, while languages with lesser AMD tend to have slowly fluctuating amplitude modulation.

We also observed that the DR₀ for female speakers did not exceed 2 dB relative to male speakers. Earlier studies on English [13,20] have also claimed that there is no significant differ-

ence in the DR₀ of male and female speakers.

Figure 5 shows the difference in DR for different languages with reference to BE as a function of frequency and integration time. It is seen from the figure that DR for Indian languages is significantly less at lower frequencies (0.1–0.5 kHz) as well as at higher frequencies (> 2 kHz). To explore such differences further we inspected the 99% and 10% cumulative level characteristics for all languages across different frequencies as plotted in Figure 6. The figure shows that 99% level curves for all languages are reasonably close to each other. However, 10% level curves for Indian languages are significantly above that of BE at low, as well as at high frequencies. Thus, it is the 10% level curve which is primarily responsible for lesser DR of Indian languages.

Implications

In this work we have presented strong evidence that LTASS and DR characteristics for three Indian languages are mark-

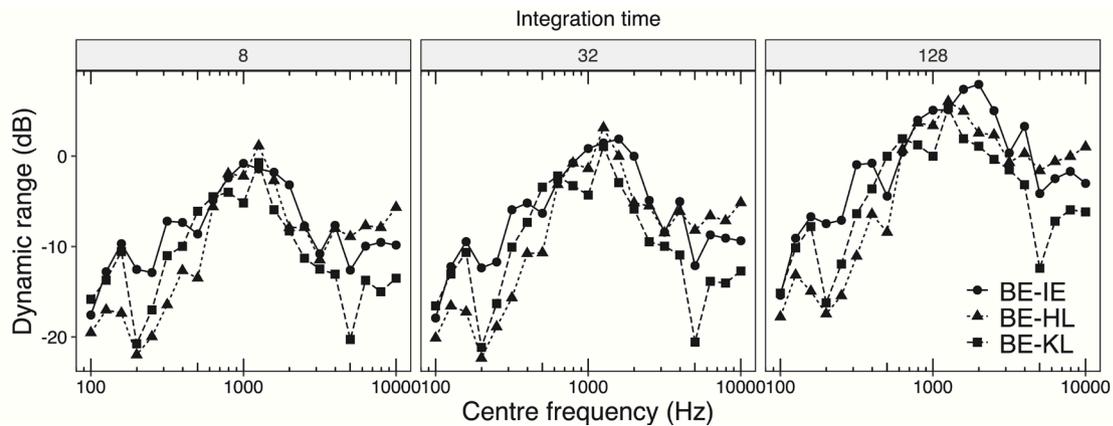


Figure 5. Difference between DRs of BE and Indian languages as function of frequency and integration time.

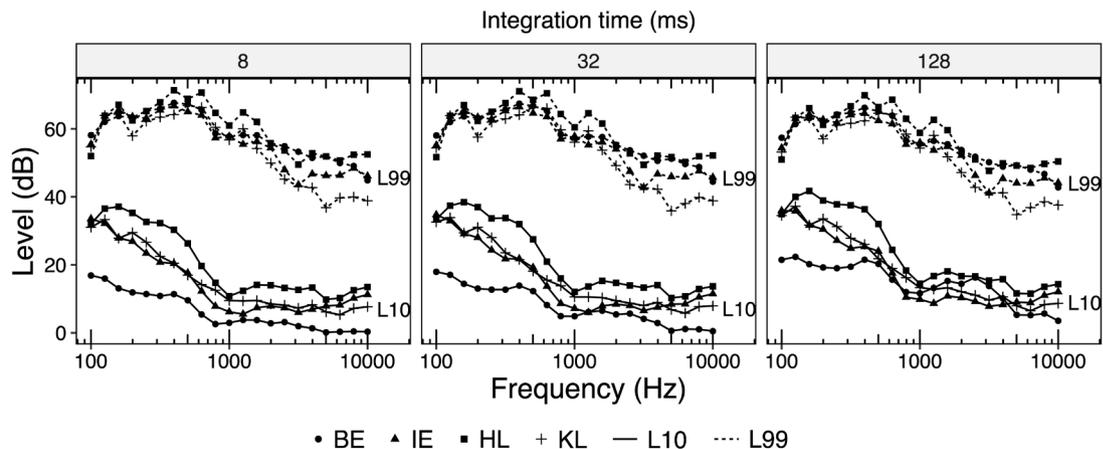


Figure 6. Cumulative histogram levels as a function of frequency and integration time.

edly different than those for BE. Such differences have significant implications with regard to calculation of Speech Intelligibility Index (SII), as well as prescription of gain and compression parameters used in hearing aids for Indian language speakers. This is discussed further.

Speech Intelligibility Index

As per ANSI S3.5 [1], SII can be calculated by using the following equation.

$$SII = \sum_{f=1}^F BIF_f \times \left[\frac{(SNR_f + K)}{DR_p} \right]$$

Here,

BIF_f : Band Importance Function for f^{th} frequency band. This parameter is language dependent and its determined empirically. As of now, its values have been determined only for English [31], Koran [32], and Madeiran [33].

SNR_f : Signal to noise ratio for f^{th} frequency band in dB. It is calculated by using language specific LTASS_f.

DR_p : Perceptual dynamic range in dB.

K : K is an offset, dependent on ratio of max speech level and its RMS level.

We note that for a given noise level band specific LTASS_f and SNR_f are positively correlated. It would also be reasonable to argue that perceptual and measured values of DR (i.e., DR_p and DR_o) are also positively correlated. Thus, the SII of a speech can be strongly language dependent, particularly if LTASS and DR_o parameters for the language under consideration are markedly different than those of English. That is indeed the case for languages analysed in this work. Traditionally, the values of DR_p have been assumed as 30 dB based on data from studies on English speech samples [34-37]. However, our study shows that there is a strong need to determine it for Indian languages. The same may be said for BIF_f as well.

Gain and compression parameter used in hearing aid

Typically gain parameters implemented in hearing aids help make speech audible over most of its dynamic range. As per ANSI S3.5 [1], the minimum gain used in hearing aids is related to output level (in dB SPL) delivered to the listener by:

$$Output = HT + G, \text{ where } G = \frac{DR_p}{2}$$

Here, HT corresponds to the hearing threshold for the listener in dB SPL and G is factor estimated from DR_p . A too large value of G may cause hearing discomfort, particularly when the HT for person is significantly high. In contrast, too little G

may lead to loss of perception of speech. Since languages studied in this work have DR_o values significantly lesser than that for BE, there is a need to reassess the current practice of setting G to at least 15 dB which is half of the DR_p of English. Further, since LTASS levels for Indian languages are at least 5 dB less than that for BE, more gain at higher frequencies may be needed for Indian speakers.

Further, hearing aids as well as audio-headsets rely on compression algorithms to reduce signal distortion and discomfort, especially when loudness level is very high. Such algorithms use DR_p to calculate the extent of compression. Once again, given that dynamic range for Indian languages is significantly different than that of BE, these algorithms might have to re-worked as well.

Declaration of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGMENTS

We thank Prof. Brain C.J. Moore and Prof. Michel Stone for providing the stimuli used in their study and also thank Prasanth Prabhu for assisting in recording the Indian speech samples. Further, we extend our thanks to the subjects for their cooperation.

REFERENCES

1. ANSI S3.5. Methods for the Calculation of the Speech Intelligibility Index. American National Standards Institute, New York. 1997.
2. Cox RM, Moore JN. Composite speech spectrum for hearing and gain prescriptions. J Speech Lang Hear Res. 1988;31:102-107.
3. Moore BCJ, Stone MA, Füllgrabe C, Glasberg BR, Puria S. 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. 2008;29:907-922.
4. Byrne D. The speech spectrum-some aspects of its significance for hearing aid selection and evaluation. Br J Audiol. 1977;11:40-46.
5. Narne VK, Prabhu P, Thuvassery P, Ramachandran R, Kumar A, Raveendran R, et al. Frequency importance function for monosyllables in Malayalam. Hear Balance Commun. 2016;14:201-206.
6. Wong LLN, Ho AHS, Chua EWW, Soli SD. Development of the Cantonese speech intelligibility index. J Acoust Soc Am. 2007;121:2350-2361.
7. Chasin M. Setting hearing aids differently for different languages. Seminars in Hearing. 2011;32:182-188.
8. Chasin M. Sentence final hearing aid gain requirements of some non-English languages. Can J Speech-Lang Pathol Audiol. 2012;36:196-202.

9. Jin IK, Lee K, Kim J, Kim D, Sohn J, Lee J. Comparison of a hearing aid fitting formula based on Korean acoustic characteristics and existing fitting formulae. *Audiol Speech Res.* 2017;13:216-221.
10. Noh H, Lee DH. Cross-language identification of long-term average speech spectra in Korean and English: Toward a better understanding of the quantitative difference between two languages. *Ear Hear.* 2012;33:441-443.
11. Tarnoczy T, Fant G. Some remarks on the average speech spectrum. *STL-QPSR.* 1964;5:13-14.
12. Yüksel M, Gündüz B. Long term average speech spectra of Turkish. *Logopedics Phoniatrics Vocology.* 2018;43:101-105.
13. Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox R, et al. An international comparison of long-term average speech spectra. *J Acoust Soc Am.* 1994;96:2108-2120.
14. Khan I, Gupta SK, Rizvi SHS. Formant frequencies of Hindi vowels in /hVd/ and C1VC2 contexts. *J Acoust Soc Am.* 1994;96:2580-2582.
15. Sreedevi N, Nair SK, Vikas M. Frequency of occurrence of phonemes in Kannada: a preliminary study. *Journal of the All India Institute of Speech & Hearing.* 2012:31.
16. Bhaskarao P. Salient phonetic features of Indian languages in speech technology. *Sadhana.* 2011;36:587-599.
17. Kochetov A, Sreedevi N. Articulation and acoustics of Kannada affricates: A case of geminate /tʃ/. *Clin Linguist Phon.* 2016;30:202-226.
18. Tabain M, Kochetov A. Acoustic Realization and Inventory Size: Kannada and Malayalam Alveolar/Retroflex Laterals and /ɹ/. *Phonetica.* 2018;75:85-109.
19. Mili M, Sairam VC, Vani R, Manjula P, Yathiraj A. Long term average speech spectrum across three Indian languages. *J Indian Speech Language Hearing Assoc.* 2004;8:1-4.
20. Cox RM, Matesich JS, Moore JN. Distribution of short-term rms levels in conversational speech. *J Acoust Soc Am.* 1988;84:1100-1104.
21. Jin IK, Kates JM, Arehart KH. Dynamic range for speech materials in Korean, English, and Mandarin: a cross-language comparison. *J Speech Lang Hear Res.* 2014;57:2024-2030.
22. Rhebergen KS, Versfeld NK, Dreschler WA. The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise. *J Acoust Soc Am.* 2009;126:3236-3245.
23. Lobdell BE, Allen JB. A model of the VU (volume-unit) meter, with speech applications. *J Acoust Soc Am.* 2007;121:279-285.
24. Pearsons K, Bennett R, Fidell S. *Speech levels in various environments*, Bolt. U.S. Environmental Protection Agency, Washington, D.C. 1977.
25. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria; 2019.
26. McCullough JA, Tu C, Lew HL. Speech-spectrum analysis of Mandarin: implications for hearing-aid fittings in a multi-ethnic society. *J Am Acad Audiol.* 1993;4:50-52.
27. Sirsa H, Redford MA. The effects of native language on Indian English sounds and timing patterns. *J Phon.* 2013;41:393-406.
28. Maxwell O, Fletcher J. Acoustic and durational properties of Indian English vowels. *World Englishes.* 2009;28:52-69.
29. Maxwell O, Fletcher J. The acoustic characteristics of diphthongs in Indian English. *World Englishes.* 2010;29:27-44.
30. Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C. A cross-linguistic study of speech modulation spectra. *J Acoust Soc Am.* 2017;142:1976-1989.
31. Studebaker GA, Pavlovic CV, Sherbecoe RL. A frequency importance function for continuous discourse. *J Acoust Soc Am.* 1987;81:1130-1138.
32. Jin IK, Lee J, Lee K, Kim J, Kim D, Sohn J, et al. The Band-Importance Function for the Korean Standard Sentence Lists for Adults. *J Audiol Otol.* 2016;20:80-84.
33. Chen J, Huang Q, Wu X. Frequency importance function of the speech intelligibility index for Mandarin Chinese. *Speech. Commun.* 2016;83:94-103.
34. Dunn HK, White SD. Statistical measurements on conversational speech. *J Acoust Soc Am.* 1940;11:278-288.
35. Beranek LL. *The design of speech communication systems*. Proc IRE. 1947;35:880-890.
36. Pavlovic CV. Articulation index predictions of speech intelligibility in hearing aid selection. *ASHA.* 1988;30:63-65.
37. Studebaker GA, Sherbecoe RL, McDaniel DM, Gwaltney CA. Monosyllabic word recognition at higher-than-normal speech and noise levels. *J Acoust Soc Am.* 1999;105:2431-2444.