Integrating the Acoustics of Running Speech into the Pure Tone Audiogram: A Step from Audibility to Intelligibility and Disability

Paul Corthals
Faculty of Medicine and Health Sciences, Ghent University, and Faculty of Health Care Vesalius, University College Ghent, Ghent, Belgium

Abstract

Background/Aims: The aim of the present study is to construct a simple method for visualizing and quantifying the audibility of speech on the audiogram and to predict speech intelligibility. The proposed method involves a series of indices on the audiogram form reflecting the sound pressure level distribution of running speech. The indices that coincide with a patient’s pure tone thresholds reflect speech audibility and give evidence of residual functional hearing capacity. Patients and Methods: Two validation studies were conducted among sensorineurally hearing-impaired participants \( n = 56 \) and \( n = 37 \), respectively to investigate the relation with speech recognition ability and hearing disability. Results: The potential of the new audibility indices as predictors for speech reception thresholds is comparable to the predictive potential of the ANSI 1968 articulation index and the ANSI 1997 speech intelligibility index. The sum of indices or a weighted combination can explain considerable proportions of variance in speech reception results for sentences in quiet free field conditions. The proportions of variance that can be explained in questionnaire results on hearing disability are less, presumably because the threshold indices almost exclusively reflect message audibility and much less the psychosocial consequences of hearing deficits. Conclusion: The outcomes underpin the validity of the new audibility indexing system, even though the proposed method may be better suited for predicting relative performance across a set of conditions than for predicting absolute speech recognition performance.

Introduction

Since the 1980s, following the seminal work in 1947 by French and Steinberg on the articulation index, many graphic methods have been designed to derive the amount of information available to hearing-impaired listeners from their audiogram with a minimum of calculations [1–8]. These simplified prognostic protocols rely on the relative position of auditory thresholds within the dynamic range of speech. There are several intricate aspects about prognostic extrapolations of pure tone threshold results. To begin with, the frequency content of pure tones differs drastically from the spectral composition of speech and this poses a problem when the sensation level of speech is represented on the decibel scale of an audiogram. Furthermore, one needs descriptive statistics on daily speech and these should meet stringent requirements concerning the type of speech sample, the type and number of speakers it is taken from, the intensity units it
is measured in and the conversation distance at which it is registered. Finally, prognostic extrapolations require a conversion of estimated audible speech into intelligibility measures, which implies the use of appropriate frequency importance weightings and transfer functions.

**Sensation Levels**

The sensation level of a pure tone during tonal audiometry is the difference between the hearing level of that tone and the audiometric zero level for that particular frequency. When sounds other than pure tones are projected on the decibel-Hertz plane of an audiogram, their sensation level should be determined using an appropriate zero level. For instance, the zero level of 1/3 octave band values of speech intensity should refer to specific absolute thresholds for this very kind of sounds, i.e. filtered speech. Nevertheless, the dynamic range of speech frequently appears on audiograms without clear references. Moreover, the zero level of an audiogram refers to monaural pure tone hearing thresholds with headphones, while speech intensity measurements yield free field data for a specific sample of speech at a certain conversation distance in a given room. All these restrictions bring on that the sensation level of speech on an audiogram is merely approximate and they emphasize the need for validation of any extrapolation.

**Speech Samples**

In order to yield reliable results, the dynamic range of speech should be based on a large database of measurements in which all possible conversation partners are reflected [9]. For instance, a speakers’ age and sex influence the sound intensity of speech. According to Olsen [10], average intensity levels for speech (1 sentence spoken with normal vocal effort registered at 1 m in an anechoic chamber with a time constant corresponding to the ‘fast’ response of sound level meters) are 58 dBSPL for females and 61 dBSPL for children and males. Corthals [11] investigated equivalent continuous levels (Leq) of running speech registered at a distance of 0.3 m resulting in mean Leq in decibel SPL of 69.39 for young people, 69.17 for adults and 75.06 for the elderly. Baker et al. [12] quote average values between 71.8 and 79.6 dB SPL for younger speakers and between 65.0 and 79.2 dB SPL for older individuals for the sound pressure level in syllables (200-ms mid-vowel intervals registered at 0.3 m from the mouth in a sound-treated booth). They attributed the age difference to pulmonary elasticity, chest volume, muscular strength and less efficient laryngeal adduction. This explanation is in contrast to reports stating that older people form a habit of speaking louder in an effort to compensate for decreased auditory feedback as a consequence of presbycusis [13].

The speech sample should not only be representative of different types of speakers, it must also adequately reflect everyday speech phonetically. Classes of phonemes differ in relative sonority and therefore the sound pressure level of short fragments depends on the speech material. For the purpose of representing speech on the audiogram it is better to start from longer samples of running speech, as is done in research on vocal load. Buekers et al. [14] even use an observation period of several hours. In the case of running speech, the registration technique and measurement unit have to allow some form of integration over time, i.e. a differential processing of the intensity levels depending on their duration. The Leq or equivalent continuous sound level is a good candidate. The Leq is designed for applications involving fluctuating signals. It will not reflect phenomena on a segmental level, such as short sound pressure level peaks by single speech sounds. Instead, it depends on the energy of all fragments in a larger corpus occupying a given frequency interval.

Registrations of speech intensity are only valid for a given distance in a certain space because intensity varies with distance and as a result of reverberation [15]. In phonetography, a distance of 0.3 m is preferred and measuring in sound booths or other artificial acoustic ambiences is avoided, as they might affect the speaker’s self-monitoring reflex [16]. A relatively short mouth-microphone distance and a normal, quiet acoustic ambiance yield sound pressure values almost entirely determined by direct sound waves, i.e. by speech itself and not by the acoustic properties of the room. However, these values are not a good reflection of the intensity at normal conversation distances. It is possible to extrapolate these values to other distances and to different conditions of reverberation [17].

As the literature review by Olsen et al. [9] reveals, very little research meets the above requirements on number of speakers, type of speech sample and intensity unit. Furthermore, few attempts at validation by means of correlations with speech recognition scores have been made. In an attempt to fill the void, Corthals [11] reported on the overall Leq level of continuous speech (a rainbow passage), measured at a speaking distance of 0.3 m in a representative group of speakers (n = 400) of various ages and both sexes, corresponding to the population pyramid of Belgium.
Conversions into Speech Intelligibility Measures

Once the sensation level of everyday speech for a patient is derived from his audiogram, a frequency importance function and a transfer function are applied to this estimate of audibility [18]. These steps are often reduced to the interpretation of a graphical representation, in an attempt to minimize calculation efforts. A good example is the count-the-dot audiogram [8], where the frequency importance function is incorporated in the varying density of the dots and the transfer function is emulated by the maximum number of 100 dots, resembling the maximum comprehension score in percents.

The ultimate endorsement of these methods depends on validation by means of correlations with speech perception results in the target population. In the present study a new prognostic audiogram extrapolation method is developed. It is validated by means of correlations and regressions with speech recognition data and disability questionnaire data. It is also compared with the outcome of the articulation index (ANSI S3.5-1969 R 1986) [19] and the speech intelligibility index (ANSI S3.5-1997) [20].

Participants and Methods

Indexation of Hearing Thresholds

The Corthals data [11] offer linearly weighted 1/3 octave band Leq percentiles registered with a time constant of 125 ms (‘fast’) at 0.3 m from the talkers’ mouth. These percentiles of Leq, i.e., 99 numbered sound pressure level values for each frequency band, reflect the intensity distribution of running speech in the group of 400 participants (table 1). The 125-Hz band is probably more influenced by male than by female voices. These Leq decibel values were extrapolated to a conversation distance of 1 m. Audiometric calibration values [21] were subtracted in order to express speech intensity levels relative to audiometric zero levels. There were no further adjustments in order to reflect concha or ear canal resonances. An indexation system for hearing threshold values was developed. Each decile of the distribution of speech intensity levels was given an ordinal index. Index 9 was given to audiometric thresholds coinciding with the first decile (percentile 10), index 8 to thresholds coinciding with the second decile (percentile 20) and so on. Index 0 was given to audiometric thresholds on percentile 99 and index 10 to thresholds coinciding with the first percentile of speech intensity levels. No importance function to weight the information content of different spectral areas was used and as a result the indices simply reflect the audibility of running speech.

This indexation is based on the rationale that for patients with a hearing threshold coinciding with percentile 1 of the intensity distribution, the complete dynamic range of speech is audible, which corresponds to a maximum audibility index of 10. On the other hand, when a hearing threshold falls on percentile 99, the entire dynamic range is inaudible, which is expressed by a minimum audibility index of 0. Thresholds better than percentile 99 (above the maximum index 10) get the maximum and thresholds worse than percentile 1 (under the minimum index 0) get the minimum. Decibel scales on audiogram forms are divided into 5-dB intervals, but not all deciles of the speech intensity distribution coincide with a 5-dB mark on the axis. For this reason all values were rounded to the nearest multiple of 5 dB. Consequently, there are no graphic details causing ambiguity in associating an index with a pure tone audiometry result. After being rounded, some deciles of the speech intensity level distribution correspond with the same multiple of 5 dB. In cases like this, just 1 index was attributed to a single hearing level, namely the smallest decile number involved. For instance, on 1 kHz the rounded level of the fourth, fifth and sixth deciles (i.e. percentiles 40, 50 and 60) corresponds with 35 dB HL and, consequently, thresholds falling on 35 dB HL get index 4. Projected on an audiogram form, this family of indices delineates a new representation of the dynamic range of speech (fig. 1). Anyone can reproduce it by simply writing down the indices on the 5-dB intervals of a blank audiogram.

Table 1. Percentile levels (1st and 99th) and decile levels (1st–9th) drawn from the distribution of 1/3 octave band sound pressure levels (decibel) of running speech at 0.3 m

<table>
<thead>
<tr>
<th>Percentile 1</th>
<th>125 Hz</th>
<th>250 Hz</th>
<th>500 Hz</th>
<th>1 kHz</th>
<th>2 kHz</th>
<th>4 kHz</th>
<th>8 kHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decile 1</td>
<td>35.2</td>
<td>54.6</td>
<td>59.0</td>
<td>55.1</td>
<td>51.5</td>
<td>50.4</td>
<td>48.8</td>
</tr>
<tr>
<td>Decile 2</td>
<td>38.7</td>
<td>56.7</td>
<td>53.4</td>
<td>45.3</td>
<td>42.3</td>
<td>40.7</td>
<td>38.4</td>
</tr>
<tr>
<td>Decile 3</td>
<td>41.3</td>
<td>59.0</td>
<td>55.1</td>
<td>47.5</td>
<td>43.4</td>
<td>42.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Decile 4</td>
<td>47.5</td>
<td>60.5</td>
<td>56.8</td>
<td>49.3</td>
<td>44.4</td>
<td>43.3</td>
<td>41.3</td>
</tr>
<tr>
<td>Decile 5</td>
<td>52.6</td>
<td>61.8</td>
<td>58.3</td>
<td>50.9</td>
<td>45.9</td>
<td>44.5</td>
<td>42.5</td>
</tr>
<tr>
<td>Decile 6</td>
<td>57.1</td>
<td>63.0</td>
<td>59.9</td>
<td>52.5</td>
<td>47.7</td>
<td>45.4</td>
<td>43.6</td>
</tr>
<tr>
<td>Decile 7</td>
<td>60.2</td>
<td>64.5</td>
<td>61.3</td>
<td>54.1</td>
<td>49.3</td>
<td>46.5</td>
<td>45.0</td>
</tr>
<tr>
<td>Decile 8</td>
<td>62.1</td>
<td>65.9</td>
<td>63.2</td>
<td>55.7</td>
<td>51.5</td>
<td>48.1</td>
<td>46.5</td>
</tr>
<tr>
<td>Decile 9</td>
<td>65.3</td>
<td>68.3</td>
<td>66.3</td>
<td>58.6</td>
<td>54.9</td>
<td>50.4</td>
<td>48.8</td>
</tr>
<tr>
<td>Percentile 99</td>
<td>71.3</td>
<td>72.9</td>
<td>73.0</td>
<td>67.3</td>
<td>65.9</td>
<td>57.8</td>
<td>55.3</td>
</tr>
</tbody>
</table>
Correlations and regressions were calculated in order to verify the predictive validity of the new audibility indices. Two studies were conducted. First, the new indices were applied to the data in the Corthals [22] corpus. In this corpus (n = 56) pure tone hearing thresholds as well as speech audiometry results are documented, including speech reception thresholds (SRT) pertaining to sentences. The new audibility indices were treated as predictor variables for these sentence SRTs in regression analyses. These regressions, although yielding a result in decimals, link audibility to speech reception. They are the statistical equivalent of the frequency importance and transfer functions that are applied to the raw outcome of the classic articulation-index-based methods to set the step from audibility to an estimated comprehension percentage. In this study, the articulation index (ANSI S3.5-1969 R 1986) and the speech intelligibility index (ANSI S3.5-1997) were treated as alternative SRT predictors in a regression, in order to compare the predictive powers of the proposed method and the classic articulation-index-based methods.

In a second study, a new sample (n = 37) was established in order to investigate the relation between the new indices and the outcome of a questionnaire on hearing disability. Statistical processing was done using computerized statistical packages (Winks 4.62 professional edition and SPSS version 10.1.0).

Validation Study I: Speech Intelligibility Data

Study I is based on data [22] from 56 adults (29 women and 27 men) with bilateral sensorineural hearing loss. This sample was considered to be a representative cross-section of adult hearing-impaired ENT patients. The mean hearing threshold, averaged over all octave frequencies from 250 to 4,000 Hz, was 37 dB HL for the better ear and 47 dB HL for the worse ear. SRTs for Plomp and Mimpen’s [23] sentences were determined using an adaptive paradigm. In this validation study, sentence SRT data pertaining to quiet listening conditions serve as a criterion variable. Tonal threshold data from the better ear are used to establish predictor variables. The predictors for SRT were the sum of each patients’ audibility indices, the articulation index (ANSI S3.5-1969 R 1986) and the speech intelligibility index (ANSI S3.5-1997). The ANSI 1968 articulation index was determined according to the online calculation procedure by the Hearing Aids and Audiology Information Network [24] and the ANSI 1997 speech intelligibility index was calculated using the procedure by the ASA Working Group S3-79 [25]. Calculations were performed with thresholds at octave frequency intervals only (from 250 to 4,000 Hz for ANSI S3.5-1969 R 1986 and from 250 to 8,000 Hz for ANSI S3.5-1997) using settings for quiet conditions, standard speech spectrum and band importance function.

Validation Study II: Disability Questionnaire Data

In study II, 37 participants (average age 71 years, ranging from 22 to 95) were recruited in various hearing centres. There were 18 women and 19 men. Air conduction thresholds were established according to the modified Hughson-Westlake technique for 125, 250, 500, 1,000, 2,000, 4,000 and 8,000 Hz. An index was attributed to each threshold in accordance with the above-mentioned principles. Indices for left and right ears were compared per frequency and the better values were summed for each of the 7 octave frequencies to obtain a total index, in an attempt to reflect free field listening conditions in real life. Participants were asked to fill in a questionnaire (appendix 1). This questionnaire measures difficulties in understanding speech and activity restrictions in quiet and noisy circumstances. The participants were asked to answer by marking the number 1 for ‘never’, 2 for ‘rarely’, 3 for ‘sometimes’, 4 for ‘often’ or 5 for ‘always’ on an equal-appearing interval scale. The questions were compiled after consulting several existing questionnaires. Before application, the wording of all questions was checked out by a group of speech pathology students, but there was no further investigation into the psychometric properties of the questionnaire. Half of the list pertains to quiet listening conditions and the other half to noisy situations. Questions were explained if necessary. Afterwards, the participant was given time to fill in all questions. The relation between the total index and the questionnaire sum scores for speech intelligibility and activity restrictions was investigated for quiet and noisy circumstances. Two simple linear regressions and 2 stepwise multiple regressions were calculated with the questionnaire outcome as the dependent variable and the total index as predictor variable.

Results

Validation Study I: Speech Intelligibility Data

The Pearson correlation coefficient for the relation between the sum of indices and sentence SRT is significant ($r = -0.807$ and $p < 0.001$). The Pearson correlation between the ANSI 1968 articulation index and sentence SRT
is also significant ($r = -0.780$ and $p < 0.001$), as well as the correlation between the ANSI 1968 articulation index and sentence SRT ($r = -0.745$ and $p < 0.001$). A test for the homogeneity of $r$ values points out that these 3 coefficients do not differ significantly ($p = 0.680$) and that the estimated rho, the correlation within the general population between pure tone threshold data and SRT for sentence material, lies between $-0.713$ and $-0.833$ (95% confidence interval). Figure 2 is an overlay of the scattergrams and linear regression lines for audibility indices, articulation indexes (AI) and speech intelligibility indexes versus SRT values. Compared to the regression using the sum of audibility indices as a predictor ($R^2 = 0.65$), both ANSI regressions explain a slightly smaller proportion of the variance in SRT outcome ($R^2 = 0.61$ for the articulation index and $R^2 = 0.55$ for the speech intelligibility index).

The linear regression, with the SRT as the dependent variable and the sum of indices (index sum in equation 1) as the independent variable is significant ($F = 104.79$ and $p < 0.001$). The regression equation is:

$$SRT = -0.71 \times \text{index sum} + 51.81$$

A stepwise regression with sentence SRT as the dependent variable and all threshold indices as predictors ($F = 43.85$ and $p < 0.001$) points to the audibility indices on 500, 1,000, 4,000 and 8,000 Hz (index 500 and index nk in equation 2) as the most important contributors to the prediction, explaining 77% of the variance (fig. 3). The regression equation is:

$$SRT = (-1.56 \times \text{index 500}) + (-1.35 \times \text{index 1k}) + (-0.67 \times \text{index 4k}) + (0.12 \times \text{index 8k}) + 54.67$$

**Validation Study II: Disability Questionnaire Data**

The Pearson correlation coefficient for the relation between the total index (index sum in equations 3 and 4) and the questionnaire data pertaining to quiet situations is significant ($r = 0.686$ and $p < 0.001$). The linear regression is significant ($F = 30.23$ with $p < 0.001$). The proportion of the variance it can explain is 47% ($R^2 = 0.47$). The regression equation is:

$$SRT = (\text{index sum}) + 0.84$$
Disability quiet = 9.40 + 0.266 × index sum

The Pearson correlation coefficient relating the total index to the questionnaire data pertaining to noisy situations is significant ($r = 0.677$ and $p < 0.001$). The linear regression is significant ($F = 28.79$ and $p < 0.001$). The proportion of the variance it can explain is 46% ($R^2 = 0.46$). The regression equation is:

Disability noise = 6.63 + 0.195 × index sum

Stepwise regressions with questionnaire data for quiet situations ($F = 19.49$ and $p < 0.001$) as the dependent variable or questionnaire data pertaining to difficult listening situations ($F = 22.26$ and $p < 0.001$) as the dependent variable are significant. In both cases the audibility indices at 500 and 4,000 Hz make significant contributions to the prediction. Both regressions explain about half of the variance in speech perception scores and disability estimates ($R^2 = 0.58$ and 0.54, respectively).

Discussion

When compared to most other prognostic audiogram extrapolation methods, the proposed audibility index system differs in several ways. Firstly, it starts from a large and representative range of talkers who produced running speech, which was measured in a more appropriate unit offering an adequate integration over time. The proposed indexing system scales the distribution of sound pressure levels in a non-linear fashion, not unlike Pascoe’s 10, 50 and 90th percentile curves [9]. Secondly, the proposed system solely relies on sound pressure level. This means that it reflects audibility without using an ‘importance function’ to weight the information content of frequency bands. Finally, it is validated empirically as a prediction procedure for speech intelligibility and hearing disability in a representative cross-section of adult hearing-impaired ENT patients.

The audibility indices correlate significantly with speech intelligibility and with subjective estimates of auditory disability in sensorineurally hearing-impaired patients. Compared to the classic articulation index and speech intelligibility index procedures, the proposed method requires only minimal calculation efforts and it offers equivalent predicting power. About 65% of the variance in speech reception scores can be explained by a simple sum of indices. Cumulating the indices associated with pure tone thresholds results in a global estimate of the audible part of speech in free field at a distance of 1 m.

The highest possible outcome is 66, as the indices on 125 Hz do not reach beyond –10 dB HL. Multiplying the outcome by $3/2$ transforms it linearly into a value on a scale of 0 to 100. The audibility indices on 500, 1,000, 4,000 and 8,000 Hz contributed most in the prediction of sentence comprehension by hearing-impaired subjects. A weighted combination of indices on these 4 thresholds can explain 77% of the variance. By using equations 1 or 2, it becomes possible to estimate a patient’s intelligibility losses for sentence material in quiet listening conditions. Comparing aided and unaided audibility indices promises to be an interesting hearing aid evaluation technique.

As for hearing disability scores, about half of the variance can be explained by a simple sum of indices or a weighted combination of indices. The lesser proportions of variance that can be explained in the realm of auditory disability as compared to speech intelligibility may be attributed to non-auditory factors influencing disability results. In fact, there is a psychosocial dimension to auditory disability, as it refers to limitations of daily life activities due to a hearing deficit. The audibility indices on the other hand, are based exclusively on sound pressure levels and therefore the main factor governing all predictions is simply audibility. Secondly, one cannot rule out the possibility that unknown psychometric characteristics of the questionnaire may have had an influence on the proportion of variance the threshold indices can explain.

The audibility indices on 500 and 4,000 of the best ear are significant predictors for speech reception as well as for auditory disability. This makes sense from an audiological as well as from a phonetic point of view. It is well known that the best ear contributes most to speech intelligibility in daily life. The 500-Hz frequency region corresponds to the energy peak in the long-term spectrum of speech and to the area of the first formant. This frequency zone contains crucial information about the presence of a speaker, the speaker’s voice intensity and the identity of the sonorant sounds in his message. The 4,000-Hz zone usually contains essential spectral features of obstruct sounds.

The set of indices pertaining to thresholds on 125 Hz is less representative of female voices. However, the results of the statistical analyses corroborate the usefulness of the complete set of indices in overall prognoses of abilities and disabilities in daily life. When it comes to visualizing speech on the audiogram form, which is often done during patient counselling, one should point to the 125-Hz region to signal the differences between male and female voices.
Taken together, these outcomes underpin the validity of the new audibility indexing system, even though the proposed method may be better suited for predicting relative performance across a set of conditions than for predicting absolute speech recognition performance. The same can be said about the classic articulation index and speech intelligibility index procedures [26]. An extension to the present SII model proposed by Rhebergen and Versfeld [27] to predict SRTs in noise may ameliorate predictive accuracy.

Conclusion

A new metric for general speech audibility, derived from sound pressure level statistics on running speech, was developed to predict intelligibility and hearing disability in sensorineural hearing-impaired patients. Compared to other techniques, it is based on a larger number of speakers and on a sample of running speech that was measured in an appropriate intensity unit. It also depends less on graphical details. The new metric has a predictive power comparable to the ANSI 1968 articulation index procedure and the ANSI 1997 articulation index procedure. In the realm of hearing disability only lesser proportions of the variance can be explained, presumably because it exclusively reflects message audibility and not the psychosocial consequences of hearing deficits.

References


Acknowledgements

The help from Nathalie Buys, Wendy D’Haenens, Nele Lysens, Joyce van Duyvenboden, Iana Vandenberge, Veerle Bekaert, Annika Degryse, Tine Uyttenhouwen, Sien Verschelde, Bill Kremer, Annick Ceulemans, Liesje De Wit, Sofie Vancoozer, Jasmijn Vanoutrive and Liesbet Wittouck is greatly appreciated.

Appendix

Questionnaire

1 Are you able to understand radio broadcasts without a hearing aid in a quiet room when the wireless is adjusted to its standard volume setting?
2 Are you able to hear the doorbell or the telephone without a hearing aid in a quiet room?
3 Are you able to hear a door being opened behind you without a hearing aid in a quiet room?
4 Are you able to understand someone speaking to you in a quiet room without a hearing aid?
5 Are you able to understand someone speaking in a quiet room without a hearing aid even if you cannot see his face?
6 Can you easily make a telephone call without a hearing aid when the television is on?
7 Can you easily hold a conversation without a hearing aid when other persons are talking to each other at the same time?
8 Can you easily hold a conversation without a hearing aid in a crowded place (e.g. at a party)?
9 Can you easily hold a conversation without a hearing aid when you are in a car or in a train?
10 Can you easily hold a conversation without a hearing aid in a room when there are background noises (e.g. music or the noise of a washing machine)?


