inter.noise 2014
MELBOURNE AUSTRALIA
16-19 NOVEMBER

An electronic database of speech sound levels

Anthony NASH[1]

[1] Charles M. Salter Associates, USA

ABSTRACT

In early 1970's, the U.S. Environmental Protection Agency sponsored a research program involving speech sound levels in various noise environments. Bolt, Beranek and Newman (BB&N) conducted this research program under contract. Among other tasks, BB&N measured speech sound levels under controlled conditions from subjects using five categories of speech effort. In 1977, a summary report of the research program was prepared. This report was supported by a companion data supplement containing a statistical analysis of sound pressure levels measured at one meter from 97 individual subjects in an anechoic chamber. The data were obtained from a one-third-octave-band real-time analyzer, processed by a mini-computer, and then printed in the form of 482 cumulative statistical distributions. The 140,000 sound pressure levels in this printed data supplement have now been converted to an electronic database to create a convenient and efficient resource for the acoustical community. The data can be used for many purposes, including the prediction of speech privacy for a variety of speech efforts. This paper summarizes the contents of the database and discusses its possible applications.

Keywords: Speech levels, Speech intelligibility and interference, speech communication
I-INCE Classification of Subjects Number(s): 63.3

1. INTRODUCTION

The sound levels of human speech have been studied since the 1920's when Sacia, Beck and Sivian at Bell Telephone Laboratories first measured selected utterances of people speaking under controlled conditions (1, 2). Since that initial pioneering work, Sivian, Dunn, White, Fletcher and others at Bell Telephone Laboratories continued to measure the statistical levels of running speech while the subject spoke entire sentences (3, 4, 5, 6, 7, 8). Statistical amplitudes of speech levels have also been measured by Pavlovic to help quantify the assessment of intelligibility in ANSI S3.5 (9, 10). A number of other researchers refined the technique for measuring speech levels spectrally; however, their studies omitted a statistical amplitude distribution (11, 12, 13, 14).

Some researchers have also measured speech sound levels for somewhat different purposes involving the intelligibility of interrupted speech and its intelligibility under conditions of high noise levels or peak clipping (15, 16, 17, 18). Carter and Kryter were concerned with the masking of speech by tones (19).

During a study of speech privacy in landscaped ("open") offices, Warnock measured mean-square average speech amplitudes using a close-talking microphone (20).

The hearing research community has both measured and also adapted speech sound level data from others for the purpose of adjusting hearing aids for people with frequency-dependent hearing loss (21, 22). Adjusting a hearing aid to reproduce actual speech peaks is particularly difficult when the patient suffers from recruitment; hence, the integration time of the sampling spectrum analyzer has been studied as an independent parameter (23).

As part of other hearing research involving speech measured in one-third-octave bands, Brown was interested in a test subject self-identifying a "comfortable" speaking level effort when speaking a simple word (24). Holte assessed the loudness of individual one-third-octave bands of speech (25).

[1] anthony.nash@cmsalter.com

Possibly the most comprehensive study of speech sound levels was conducted in 1975-1977 by Pearsons, Bennett, and Fidell of Bolt, Beranek, and Newman (26). The Pearsons study was sponsored by the U.S. EPA and a summary report was issued in 1977. The summary report discussed speech levels measured in a variety of "real life" environments including classrooms, hospitals, department stores, offices, residences, etc. The report also extracted a few statistical charts and tables from a companion database of speech levels that was created as part of the same EPA project. In this paper, the companion database will be referred to by its original name — the "Data Supplement".

2. DATA SUPPLEMENT

The Data Supplement contains 482 controlled tests, each of which comprises a cumulative statistical distribution calculated using 1/10-second sound pressure level samples obtained from 97 subjects speaking in an anechoic chamber at one meter from a measuring microphone. The subject spoke a standard sentence at four different voice efforts while the calibrated microphone signal was recorded on a professional reel-to-reel analog tape machine.[2] The four voice categories are "normal", "raised", "loud", and "shout". The time duration of each speech test ranged from 9.5 to 24 seconds, depending on the speech cadence, the voice effort, and the number of repeats spoken.

In addition, the researcher conversed with the subject in the anechoic chamber while recording the subject's speech at one meter. In the Data Supplement, this fifth category of voice effort is called "casual conversation".

2.1 Original Data Analysis

The recordings were later played back into a Hewlett-Packard 8054A real-time analyzer that sampled the sound pressure levels in one-third-octave band from 50 Hz through 10 kHz. In addition, the analyzer sampled the A-weighted and flat-weighted levels.

The real-time analyzer was connected to a Digital Equipment Corporation PDP-8 minicomputer that received individual sample spectra and performed calculations to obtain a cumulative statistical distribution for each test. The statistical values from each test were then encoded on punched paper tape.

Neither the analog tape recording nor the punched tape output is now available — fortunately, the punched tape data were later sent to a line printer that generated 482 individual pages, each containing a single test. A bound set of pages from the line printer was retained in the Canoga Park library of the authors' acoustical consulting firm, Bolt, Beranek, and Newman.

This paper document was used to create an electronic database version of the Data Supplement.

[2] The text of the standard spoken sentence was, "Joe took father's shoe bench out; she was waiting at my lawn". This contrived compound sentence includes all the fundamental sounds found in human speech. The standard sentence was first used in 1930 by researchers at Bell Telephone Laboratories for the purpose of characterizing the time and frequency distribution of continuous speech. The spoken sentence superseded the former tedious method involving the measurement and time-domain analysis of individual phonemes. In 1930-31, Sivian and Dunn, plus White and Dunn and also Fletcher of Bell Laboratories described a custom-designed frequency analyzer used to measure the statistical distribution of short-term (1/8-second) speech amplitudes while a subject spoke continuously during a 15-second period (3, 4, 5). The frequency analysis portion of the device comprised a mix of whole-octave and one-half-octave band filters, which were employed sequentially.
The amplitude distribution portion of the device had 10 sorting bins, each with a resolution of six decibels (covering a range of 54 decibels).

2.2 Quantity of Values in Database

Each test in the Data Supplement comprises 15 percentile metrics, the minimum and maximum, the standard deviation, plus two types of mean-square averages ("Leq" and "Leq5").[3] The entire Data Supplement contains 140,493 numerical data values. Ideally, the total should have been [97 subjects x 5 voice efforts x 26 frequency bands x 20 metrics =] 252,200 values; however, many of the data cells are not populated due to the filtered signal falling below the minimum digital output capability of the analyzer.

Another reason for the disparity is that the tests of two subjects for the "Casual Conversation" voice effort and one subject for the "Shout" voice effort are either missing from the Data Supplement or were never conducted during the 1975-1977 project.

3. DEVELOPMENT OF AN ELECTRONIC DATABASE

The Data Supplement has not been available to the general public since 1977; nevertheless, its contents are deemed highly useful, especially for applications involving speech privacy and sound insulation. For this reason, we volunteered to develop an electronic relational database that could be searched using a personal computer.

3.1 Conversion From Paper To Electronic Format

In order to create an electronic version of the Data Supplement, its printed pages were scanned by an optical character reader (OCR) and converted to 482 discrete text files. Each text file was first reviewed manually for gross errors and then formatted as an individual word processing file that mimicked the appearance of the original printed page. The word processing application was then used to search each file to identify certain classes of residual typographical errors (mainly missing decimal points, extra spaces, and characters substituted for digits).

3.2 Error Correction

The scanning, formatting, and proofreading tasks consumed about 700 hours over a three-year period. Much of this effort was devoted to correcting errors and omissions from the OCR as it attempted to decode the primitive looking digits from the 1976 line printer (as the ink supply of the line printer became depleted, the increasingly lower contrast images led to greater OCR error rates).

The word processing files were printed and the new printed pages were manually proofread value-by-value to identify other classes of errors, including inappropriate digits and nonsensical artifacts from the OCR application.

Sometimes, a few of the more obscure digits from the original paper pages had to be interpolated manually by visually comparing them with readable values in adjacent columns. Less than two-dozen values in the database required such a manual intervention; in all cases, the uncertainty in the interpolated numerical values involved only the least significant digit.

3.3 Storing the Data in a Spreadsheet

The corrected data from each word processing file were copied as text and then pasted as numerical data into an electronic spreadsheet. The data in the 482 individual worksheets were also assembled into a series of five master worksheets, one for each speech category. Finally, the contents of the five master worksheets were imported into a relational database.

---

[3] The 15 percentile metrics are L_1, L_3, L_5, L_{10}, L_{20}, L_{30}, L_{40}, L_{50}, L_{60}, L_{70}, L_{80}, L_{90}, L_{95}, L_{97}, and L_{99}. Neither the "Leq" nor the "Leq5" metric is defined in the BB&N study. Using a "trial and error" technique, a few probability distributions were experimentally reconstructed from the respective cumulative distributions. Both the "Leq" and the "Leq5" metrics were then re-computed based on these experimental probability distributions and compared to their original values in the cumulative distribution. These experiments confirmed that the "Leq" is the expected mean-square average of all discrete 1/10-second sound pressure level samples measured throughout the entire measurement period. The "Leq5" metric is also a mean-square average of the samples except for the five highest values in each distribution. Evidently, these five top values were omitted from the "Leq5" average; hence, the symbol, "Leq5". The "Leq5" might have been intended to characterize the long-term average sound pressure level of speech without undue influence from peaks generated by voiced bilabial stops (as in "bench") or dental fricatives (as in "at").

4. DESCRIPTION OF DATABASE

Microsoft Access 2010 was adopted as the database application; however, any suitable relational database application could be used instead. In Table 1 below, the generic field names from the BB&N Data Supplement are compared with the explicit field names as defined in the Access database.

Table 1 – Generic and explicit field names in the database

Subject Number	Age	Sex	Voice Effort	Sample Size	Measurement Metric	Frequency Band
<SubjectNo>	<Age>	<Gender>	<VoiceLevel>	<SampleSize>	<DataType>	<freq*nnn*>

The electronic database contains five interrelated tables. A so-called *child* table holds all the numerical sound level data in 9,640 records while three so-called *parent* tables hold information about the measurement metric, the voice effort, and the subject. Another *parent* table holds the gender of the subject (male or female). This arrangement of tables helps minimize redundant information in the database records. A diagram of the table relationship is shown below.

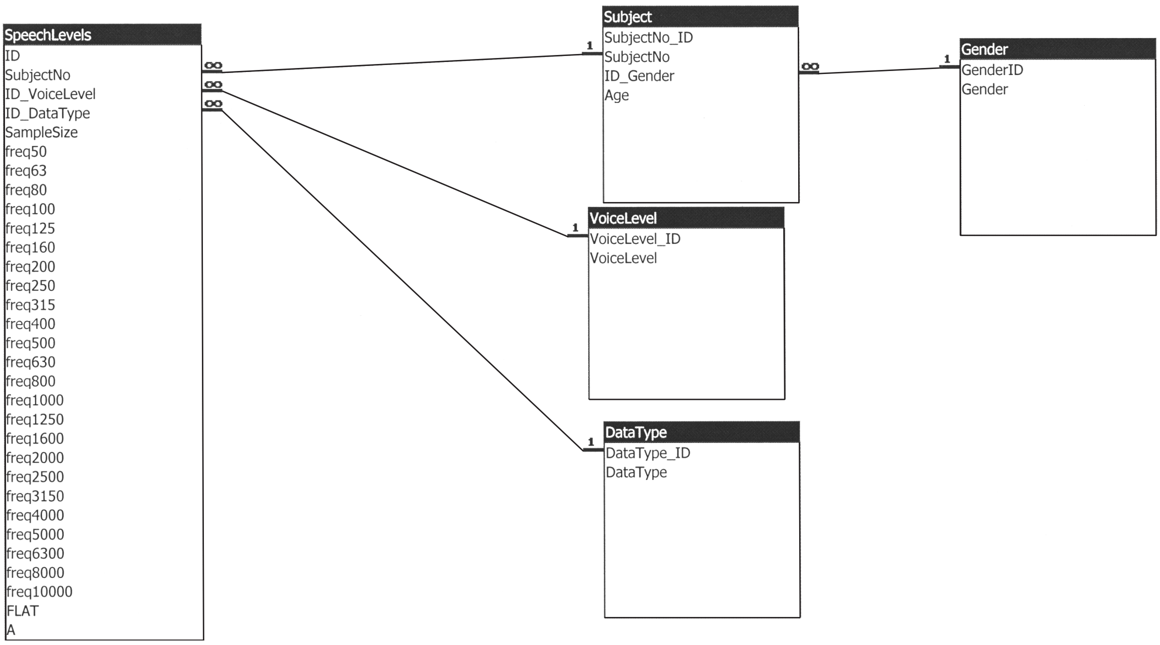

Figure 1 – Diagram illustrating the relationships among the five tables in the database file. The main *child* table is named <SpeechLevels> and contains the records of sound level data for all frequency bands. The *parent* tables that own the <SpeechLevels> table are <DataType>, <VoiceLevel>, and <Subject>. <Subject> also has a *parent* table named <Gender>. The *parent* tables hold common information in fields that would otherwise have to be repeated in many of the records held within the *child* table. Organizing the database information in this way enables it to function more efficiently.

5. QUERYING DATA

As an example, the electronic database could be queried by requesting it to display the ten-percentile levels of female talkers between the ages of 30 and 56 who are speaking with a "normal" speech effort. For the purpose of illustrating this query, the A-weighted frequency band was selected.

A possible goal of this query might be a need to compute the true mean-square average for the A-weighted, ten-percentile sound levels for all female talkers in the designated age group.

5.1 Example of Query

The first step would be to use the "Query Designer" tool in Microsoft Access to display a relationship diagram for all the relevant tables as shown in Figure 1, above. Below this relationship diagram, the Query Designer includes selectable options in a "Grid pane" view as shown below in Table 2. The search and display criteria have already been filled out for the example query.

Table 2 – Grid pane view from the Query Designer

Field:	A	DataType	Gender	Age	VoiceLevel
Table:	SpeechLevels	DataType	Gender	Subject	VoiceLevel
Sort:					
Show:	√				
Criteria:		"L10"	"F"	>=30 And <=56	"normal"
or:					

Once the query is started, Access searches through its native tables and displays only the data meeting the criteria in the Grid pane view (in this example, there are 14 records meeting the search criteria). The user could now either generate a formal report within Access or simply copy and paste the displayed data into another software application. If desired, the structure of the query could be saved with the database file so repeated future queries would be more time-efficient.

6. POSSIBLE APPLICATIONS

6.1 Calculations

The built-in functions of the electronic database could be invoked to perform mathematical calculations with the displayed data. Alternatively, the contents of the query could be pasted into an electronic spreadsheet and the calculations performed there. The data could also be imported or pasted into an advanced statistical software package for further analysis.

6.2 Speech Privacy

There has always been considerable interest in the concept of "speech privacy". The principal concern is speech being understood — i.e., whether a casual listener can gather sufficient information from the sequence of spoken words to gain intelligence or even disrupt people seated nearby as described by Warnock (20). Another aspect of speech privacy might involve the ability of an eavesdropper to detect the presence of speech — a cadence of unintelligible sounds sensed as a speech-like utterance.

Clearly, the relationship between the speech signal and the background noise level is different under these two circumstances. One could possibly use the statistical distribution of speech sounds in the database to calculate the expected "speech-to-noise" ratio for these two circumstances at various speaking efforts.

6.3 Speech Intelligibility

Another potential application for the database is to help refine the standard speech spectrum used to determine the speech intelligibility index (SII) as defined in ANSI S3.5, *Methods for Calculation of the Speech Intelligibility Index* (10). The Standard contains four columns of speech spectra in its Table 3 entitled, "One-third octave band SII procedure — frequencies, band importance function, standard speech spectra, internal noise, hearing threshold levels, and free-field to eardrum transfer function".

The spectra in Table 3 of ANSI S3.5 were developed by Pavlovic from several sources in the literature plus some speech tests that he conducted in the 1980’s using a limited population of subjects (9).[4] If one were to substitute the sound pressure levels found in Table 3 with average values from the more comprehensive BB&N Data Supplement, the uncertainty of the calculated speech intelligibility index might be improved.

6.4 Hearing Aids

The hearing research community is concerned with matching the frequency-dependent gain of a hearing aid device to encompass the dynamic range of speech. Especially for those subjects who may have recruitment, it would be desirable to avoid exceeding the threshold of discomfort whenever the aid reproduces the peak sound pressures in speech. Knowing the statistical amplitude distribution of speech in frequency bands is, therefore, of considerable importance (22, 23).

7. OBTAINING THE ELECTRONIC DATABASE

A compact disk has been prepared for public dissemination of the database. It contains the Microsoft Access database file along with 482 Microsoft Word documents, a single PDF file combining all the Word documents, five Microsoft Excel Workbooks, plus a scanned image of the original 482 printed pages formatted as single PDF file.

The scanned image of the original 482 printed pages is included so a user can visually confirm whether a typographical error has somehow found its way into the electronic version of the database. All files on the compact disk are in the public domain and are unencrypted so the information can be copied or modified without restriction.

Anyone desiring a copy of the speech database should send an inquiry to the author along with a nominal fee to cover administrative costs for reproduction, handling, and shipping.

8. SUMMARY REMARKS

The electronic database could prove quite convenient for researchers in the field of speech and hearing, practitioners in architectural acoustics, and others who are concerned with statistical sound levels and spectra generated by human talkers.

In summary, the opportunities for searching, sorting, and subsequent calculations one can perform with this speech database is limited only by the imagination of the user.

ACKNOWLEDGEMENTS

Jason Duty provided invaluable advice during the creation of the relational database. His skills and experience with Microsoft Access materially contributed to the success of the electronic database project.

[4] Personal communication with Dr. Chaslav Pavlovic.

REFERENCES

1. Sacia CF, Beck CJ. The power of fundamental speech sounds. Bell System Technical Journal. 1926;5(3):393-403.
2. Sivian LJ. Speech power and its measurement. Bell System Technical Journal. 1929;8(4):646-661.
3. Sivian LJ, Dunn HK, White SD. Absolute amplitudes and spectra of certain musical instruments and orchestras. J Acoust Soc Am. 1931;2(3):330-371.
4. Dunn HK. A new analyzer of speech and music. Bell Labs Record. 1930;10(3):118-123.
5. Fletcher H. Some physical characteristics of speech and music. Bell System Technical Journal. 1931;10(3):349-373.
6. Dunn HK, White SD. Statistical measurements on conversational speech. J Acoust Soc Am. 1940;11(3):278-288.
7. French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. J Acoust Soc Am. 1947;19(1):90-119.
8. Von Tarnóczy T. Das durchschnittliche energie-spektrum der sprache (für sechs sprachen). Acustica. 1971;24(2):57-74.
9. Pavlovic C. Derivation of primary parameters and procedures for use in speech intelligibility predictions. J Acoust Soc Am. 1987;82(2):413-422.
10. ANSI S3.5-1997. Methods for calculation of the speech intelligibility index. American National Standards Institute, New York 1997.
11. Dunn HK, Farnsworth, DW. Exploration of the pressure field around the human head during speech. J Acoust Soc Am. 1939;10(2):184-199.
12. Stevens SS, Egan JP, Miller GA. Methods of measuring speech spectra. J Acoust Soc Am. 1947;19(5):771-780.
13. Rudmose HW, Clark KC, Carlson FD, Eisenstein JC, Walker RA. Voice measurements with audio spectrometer. J Acoust Soc Am. 1948;20(4):503-512.
14. Harris CM, Waite WM. Measurements of speech spectra recorded with a close-talking microphone. J Acoust Soc Am. 1966;37(5):926-927 [Letter].
15. Miller GA, Licklider JCR. The intelligibility of of interrupted speech. J Acoust Soc Am. 1950;22(2):167-173.
16. Pickett JM. Limits of direct speech communication in noise. J Acoust Soc Am. 1958;30(4):278-281.
17. Pollack I, Pickett JM. Intelligibility of peak-clipped speech at high noise levels. J Acoust Soc Am. 1959;31(1):14-16.
18. Pickett JM, Pollack I. Prediction of speech intelligibility at high noise levels. J Acoust Soc Am. 1958;30(10):955-963.
19. Carter NL, Kryter KD. Masking of pure tones and speech. J Aud Res. 1962;(2):66-98.
20. Warnock ACC, Chu WT. Voice and background noise levels measured in open offices. Internal Report IR-837, Institute for Research in Construction, National Research Council Canada, 2002.
21. Byrne D, Dillon H, Tran K, et al. An international comparison of long-term average speech spectra. J Acoust Soc Am. 1994;96(4):2108-2120.
22. Olsen WO, Hawkins DB, Van Tasell DJ. Representations of the long-term spectra of speech. J Ear and Hearing, Am Auditory Soc. 1987;8(5):100S-108S [Supplement].
23. Cox RM, Matesich JS, Moore JN. Distribution of short-term rms levels in conversational speech. J Acoust Soc Am. 1988;84(3):1100-1104 [Letter].
24. Brown WS, Murray T, Hughes D. Comfortable effort level: an experimental variable. J Acoust Soc Am. 1976;60(3):696-699.
25. Holte L, Margolis RH. The relative loudness of third-octave bands of speech. J Acoust Soc Am. 1987;81(1):186-190 [Letter].
26. Pearsons KS, Bennett RL, Fidell S. Speech levels in various noise environments. EPA-600/1-77-025. U.S. Environmental Protection Agency, Washington, D.C., 1977 [Note: the Data Supplement appended to this report is entitled, “Conversational Speech Levels in Various Environments, Bolt, Beranek, and Newman Report No. 3281”].