Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms

Gover, B. N.; Bradley, J. S.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1121/1.1810300
Journal of the Acoustical Society of America, 116, December 6, pp. 3480-3490, 2004-12-01

Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms

Gover, B.N.; Bradley, J.S.

NRCC-47039

A version of this document is published in / Une version de ce document se trouve dans :
Journal of the Acoustical Society of America, v. 116 no. 6, Dec. 2004, pp. 3480-3490
Doi: 10.1121/1.1810300

Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms

Bradford N. Gover[a] and John S. Bradley[b]

Institute for Research in Construction, National Research Council, 1200 Montreal Road,

Ottawa, Ontario K1A 0R6, Canada

Date uploaded to JASA manuscript system: 07 September 2004

Suggested Running Title: Measures for assessing architectural speech security

[a] Electronic mail: brad.gover@nrc-cnrc.gc.ca

[b] Electronic mail: john.bradley@nrc-cnrc.gc.ca

ABSTRACT

Objective measures were investigated as predictors of the speech security of closed offices and rooms. A new signal-to-noise type measure is shown to be a superior indicator for security than existing measures such as the Articulation Index, the Speech Intelligibility Index, the ratio of the loudness of speech to that of noise, and the A-weighted level difference of speech and noise. This new measure is a weighted sum of clipped one-third-octave-band signal-to-noise ratios; various weightings and clipping levels are explored. Listening tests had 19 subjects rate the audibility and intelligibility of 500 English sentences, filtered to simulate transmission through various wall constructions, and presented along with background noise. The results of the tests indicate that the new measure is highly correlated with sentence intelligibility scores and also with three security thresholds: the threshold of intelligibility (below which speech is unintelligible), the threshold of cadence (below which the cadence of speech is inaudible), and the threshold of audibility (below which speech is inaudible). The ratio of the loudness of speech to that of noise, and simple A-weighted level differences are both shown to be well correlated with these latter two thresholds (cadence and audibility), but not well correlated with intelligibility.

PACS numbers: 43.55.Hy, 43.71.Gv

I. INTRODUCTION

Offices and meeting rooms are often intended for confidential discussions where eavesdroppers and others outside the room should not be able to listen in to these discussions. Speech originating inside such a room being difficult to hear or understand in the adjoining spaces implies that the room provides good speech privacy. In cases where the degree of privacy is sufficiently high, one can speak of architectural speech "security". Improved security would be provided, for instance, by a room constructed with boundaries having higher sound transmission loss. To assess the degree of security, it is necessary to measure whether the "transmitted" speech is audible or intelligible at positions outside the room. This paper reports the results of new subjective studies to evaluate measures of architectural speech security.

Defining the problem of architectural speech security in terms of the fraction of speech that can be understood makes it one of speech intelligibility. Investigations into speech privacy for open-plan office situations have used this same approach.[1] The distinction made by security designers is that "privacy" is a less stringent description of sound isolation than "security". A condition where overheard speech is audible but only slightly intelligible corresponds to excellent privacy, but could be described as imperfect security. A high degree of speech security implies not only very low (or zero) speech intelligibility, but also minimal audibility of the transmitted speech sounds.

Speech intelligibility, speech privacy, and speech security are related to the level of the speech signal relative to the level of the noise at the listener position. The Articulation

Index (AI)[2] and the Speech Intelligibility Index (SII)[3] are both frequency-weighted signal-to-noise ratio type measures, and can be calculated from the speech and noise spectra at the position of a listener. AI and SII are good indicators of the degree to which the speech is intelligible, but they do not necessarily represent zero intelligibility at their minimum values of zero. Furthermore, for conditions of zero intelligibility, where the rhythm of speech or some speech sounds may be audible, AI and SII provide no information. They are therefore seemingly ill-suited to the problem of assessing architectural speech security. A useful measure of security will indicate the degree to which the transmitted speech is intelligible or, in cases of zero intelligibility, audible.

The most significant published study of architectural speech "privacy" in buildings was by Cavanaugh *et al*. in 1962.[4] They presented a report on occupants' impressions of privacy in buildings, and how they related to AI. The main focus of the work was on privacy (i.e., freedom from distraction), however there was a subcomponent of the study assessing so-called "confidential privacy", which subjects were instructed corresponded to a situation where there was an "assurance of not being overheard." One of their main results states that, "the most critical 10% of the subjects began to feel a lack of [confidential] privacy when the articulation index reached 0.05." An important point is that the assessments were based on how private the subjects *felt* a situation was. The actual fraction of the speech they *could* understand was not measured. Nevertheless, this paper lays the basis for relating privacy in offices to an objective measure.

In 1965, Young[5] published a revised computational procedure based on the data in Cavanaugh *et al*. He proposed a measure derivable from A-weighted levels of speech and noise, and single number indicators of transmission loss. This is a practical approach, easy to calculate, but at best no more accurate than the original Cavanaugh *et al*. method. This simplified method was not supported with additional subjective tests. The so-called "Speech-Privacy Calculation" has become accepted practice,[6] yet being based on the Articulation Index, is likely not appropriate for rating speech security, where zero intelligibility is expected.

This paper presents the derivation of a suitable objective measure, providing information regarding the degree of security both above and below the threshold of intelligibility.

II. EXPERIMENTAL DESIGN

Subjective listening tests were conducted, presenting subjects with sound fields simulating listening outside of a room in which a person was speaking. The spectra of the speech and of the noise at the listening position were separately measured, and were used to compute various objective measures. Relations were sought between these measures and the subjects' responses.

A. Listening tests

All listening tests were conducted in a sound-isolated test room located at the National Research Council in Ottawa. Anechoically-recorded speech was filtered to simulate transmission through a wall and, along with background noise, was played into the room. Subjects in the room heard these sounds and responded to a test operator seated outside the room, who scored their response. Details of the tests and of the subjects follow. All tests were approved by NRC's Ethics Review Board.

1. Facility and hardware

The test room measured 9.2 m long by 4.7 m wide by 3.6 m high, and was constructed from concrete. The room is not connected to the building, and is resting on springs for vibration isolation. Sounds existing outside the room, therefore, are largely isolated from penetrating within. For the present study, the interior walls of the room were lined with 10 cm-thick absorbing foam, covered by curtains. There was a conventional T-bar ceiling with 25 mm-thick glass fibre ceiling tiles installed, and the floor was covered with carpet. This interior treatment yielded a quite "dead" space. The measured background noise level in the room was 13.7 dBA.

Test speech was played over loudspeakers positioned at the front of the room. The background noise was played over another set of loudspeakers positioned above the ceiling, directly above the subject. Figure 1 shows a diagram of the setup. Having the noise and speech originate from different spatial locations is important for a realistic test. Listeners are better able to recognize speech in noise when the speech and noise arrive from different directions.[7]

A block diagram of the electroacoustic system used to produce the test sounds is shown in Fig. 2. The two blocks labeled “DME32” are Yamaha Digital Mixing Engines, which are highly flexible signal processing boxes, able to perform the functions of many interconnected devices such as equalizers, filters, oscillators, etc. The outputs of the DME32s run through the power amplifiers into high-quality loudspeaker systems (Paradigm Compact Monitors, Paradigm PW sub-woofers). One component in each DME32 was initially configured under computer control (via the RS232 interface) to equalize the playback path through the power amplifiers and loudspeakers to be flat at position of the listener’s head (± 1 dB from 60 to 12000 Hz).

The background noises for the test sound fields were generated by the internal noise generator of one of the DME32 units. This same unit shaped the spectrum and adjusted the level as desired, responding to control commands sent by the computer over the MIDI interface. One channel of the noise output was delayed by 300 ms relative to the other so as to avoid any unnatural perceptual effects caused by movement of the listener’s head (as can be observed when listening to monaural material over a pair of loudspeakers).

The speech sounds were generated from playback of anechoically-recorded source material stored on the computer in 16-bit, 44.1 kHz wave-file format. The output of the sound card ran into the second DME32, which performed the necessary equalization and level adjustment, again with commands to switch settings received over the MIDI interface. Equalizer components in this DME32 simulated the various wall transmission loss curves.

2. Speech material

The test material used was phonetically-balanced English sentences.[8] These sentences are of low predictability, and are more representative of conversation than isolated words. The response and scoring procedure had the subjects state out loud the words they thought they heard, and the test operator noted on a score sheet those that were correctly identified. In this way the fraction of words actually understood was scored. In intelligibility tests using rhyming words, the scoring can be done similarly, but subjects are able to guess correctly as often as 20% of the time,[9] which means they are particularly unsuitable in assessing security situations, where scores as low as 0% are expected. Furthermore, sentences can be more accurately scored than running speech, or so-called "connected discourse", whereby the procedure usually involves allowing the subject to estimate the fraction of the words they are capable of understanding.[10] When the issue is security, rather than a sense of privacy, this distinction is important.

A total of nine recorded versions of the test sentences were acquired, spoken with different efforts by different talkers of both genders. The choice of recording used for the tests was the best quality recording available (16-bit, 44.1 kHz "CD-quality" digital), and was of a male talker speaking clearly. This speaking voice was most easily identified by subjects in pilot testing. That is, from a security point of view, this recording was the "worst-case" talker. The average of the spectral magnitude of four test sentences is shown in Fig. 3. The spectrum varies slightly from sentence to sentence; this curve is more representative of a "long-term" average. Also shown in Fig. 3 are "typical" speech spectra for a male talker speaking with two different speaking efforts, taken from Ref. 11.

At this playback level (68 dBA), it can be seen that the speech corresponds to an effort somewhere between “Raised” and “Loud”. All speech spectral levels were measured over a 60 second period, looping the sentences continuously.

3. Walls

Walls to be simulated in the tests were selected from past measurements of actual wall samples in the wall testing facility at NRC. The transmission loss (TL) curves for the four walls used in the test are shown in Fig. 4, their descriptions are given in Table I. These four curves were selected since they are of different shapes, and representative of wall constructions typical of office environments.

The speech transmitted through any of these walls will be attenuated and spectrally-distorted. Figure 5 shows typical speech spectra measured in the test room after filtering by each TL curve.

4. Noise

All background noises used in the testing were spectrally-shaped random noise. As discussed above, one of the Yamaha DME32s was used to generate the noise internally, and equalizer components were used to shape the spectrum. Five different spectra were used, shown in Fig. 6.

The base case noise is the "Neutral" spectrum, so-named under the RC naming convention.[12] It has a –5-dB/octave roll off. The other spectra are derived from this by boosting "low" (50–200 Hz), "mid" (250–1600 Hz), or "high" (2000–10000 Hz) frequency sections by 10 dB. These other noise types are used to systematically vary the spectral qualities of the interfering noise, representing cases that are more or less "rumbly" or "hissy".

B. Objective Measures

From measurements of the third-octave spectral levels of the "transmitted" speech and noise, various numerical indicators can be calculated. The issue is to find a measure that is well-correlated with the subjective responses from the listening tests.

1. AI and SII

The Articulation Index (AI)[2] and its more recent replacement the Speech Intelligibility Index (SII)[3] are measures calculated from the ratio of speech to noise in various frequency bands. The basic idea is that the contribution to the overall intelligibility from a particular frequency band is dependent on the "effective" ratio of signal to noise in that band, and on the importance of the band. The index itself is a weighted sum of the band contributions. Properties of the hearing system are built into the measure through the method in which the effective signal-to-noise ratio is determined from the signal and noise levels, and through the specification of the frequency band importance weights.

For the calculation of AI, the effective signal-to-noise level difference in a band is obtained by clipping the actual level difference, and by adjusting for the difference in level between the peak speech level and the rms speech level. The clipping is performed by setting all values below a specified minimum value L = –12 dB to be equal to –12 dB, and all values above a specified maximum value U = +18 dB equal to +18 dB. This clipped signal-to-noise is then shifted by adding 12 dB, to adjust for the difference between the peak and rms levels. The AI is given mathematically by

$$\mathrm{AI} = \frac{1}{30}\sum_b w_b \cdot \left[12 + \min\left(\max\left(S_b - N_b, -12\right), 18\right)\right], \tag{1}$$

where S_b is the speech level in decibels in frequency band b, N_b is the noise level in decibels in band b, and w_b are the frequency weightings. $\max(x, L)$ is the larger of x or L; that is, L is the minimum clipped value of x. Values of x below L are set to L. Similarly, $\min(x, U)$ clips the upper value to U. The resulting value of AI is between 0 and 1. The calculation of SII is similar, but the determination of effective band signal-to-noise ratio is a little more sophisticated, taking into account masking. With regard to the practical minimum and maximum values for band signal-to-noise level differences, the SII calculation uses L = –15 dB, U = +15 dB.

2. Weighted signal-to-noise

A simplified approach that is similar to some of the steps involved in computing the AI or SII is to compute the actual (as opposed to effective) signal-to-noise level difference in each frequency band, and then simply perform a weighted sum across all bands. It does seem prudent, however, to specify a lower limit for the signal-to-noise level difference, clipping to this minimum value. In speech security situations, very high signal-to-noise level differences are unlikely to occur, so clipping will not be specified on the upper end. The expression for the resulting weighted signal-to-noise ratio $X_w(L)$ is

$$X_w(L) = \sum_b w_b \cdot \max(S_b - N_b, L), \tag{2}$$

where as above, S_b is the speech level in decibels in frequency band b, N_b is the noise level in decibels in band b, and w_b are the frequency weightings. $\max(S_b - N_b, L)$ is the larger of L or the signal-to-noise level difference; that is, L is the minimum clipped value of the actual signal-to-noise level difference. Several weighting strategies for one-third-octave bands were considered, those resulting in the best-correlated indices are presented here: uniform weighting, AI-band importance weighting, and SII-band importance weighting. The weights, normalized so that they sum to unity, are given in Table II.

3. Loudness

Loudness (in sones) is a quantity that can be calculated from the spectrum of a signal, and has been shown to be related to the "perceived loudness" of sounds. The calculation of loudness is explained in Ref. 13. Measuring the speech loudness Λ_S and the noise loudness Λ_N separately, the loudness ratio r_Λ given by

$$r_\Lambda = \frac{\Lambda_S}{\Lambda_N} \tag{3}$$

can be computed. This is a linear measure, so for instance if the speech is half as loud as the noise, the value of r_Λ will be 0.5.

4. A-weighted level difference

Weighted speech and noise levels can be computed from their spectra separately, and then the difference in these levels computed. The weighting scheme considered here is the familiar A-weighting, the result being the A-weighted level difference L_A, given by

$$L_A = 10\log\left(\sum_b w_b 10^{(S_b/10)}\right) - 10\log\left(\sum_b w_b 10^{(N_b/10)}\right), \tag{4}$$

where S_b, N_b, b, and w_b are as above. The terms on the right-hand side are the A-weighted signal and noise levels, respectively. The weights w_b are merely the non-logarithmic A-weighting factors; $10\log(w_b)$ is the decibel correction factor for band b.

III. LISTENING TEST DESCRIPTIONS

A. Test 1: Intelligibility

The test conducted first consisted of a large number of sentences processed to simulate wide ranges of the conditions thought to be representative of actual offices and meeting rooms. Each of the four wall types and five background noise types were included. For each wall/noise combination, three signal-to-noise ratios were designed, corresponding to varying "difficulty": easy, moderately difficult, and difficult. In the "easy" case it was judged that most listeners should be able to identify all words in a sentence. The "difficult" case was intended to be just above the threshold of intelligibility—some listeners could identify some (but not necessarily all) of the words. (It is not possible to design in advance the conditions that define the threshold we are seeking to find. Some of the difficult cases still had to correspond to non-zero intelligibility in order to find the threshold point.) In addition to these 4 x 5 x 3 = 60 cases, 8 additional cases were constructed corresponding to 2 additional difficulties ("very difficult" and "very very difficult") for 1 noise type (Neutral), for all 4 walls. These cases were constructed by reducing the speech level 3 dB and 9 dB below that used for the "difficult" case. In total, this yielded 68 physical conditions. Five different sentences were included for each condition, resulting in a test 340 sentences long. The range of levels for the "source" speech (notionally behind the simulated wall) resulted in measured "transmitted" speech levels of 28–49 dBA in the test room. The range of levels for the background noise was from 27–51 dBA.

The test was conducted in the following manner: the subject sat in the room with no test sounds playing. The noise was turned on, and a second or so later, the speech began. After the end of the sentence, the noise was switched off. At this point, the subject said out loud (into a talkback microphone) the words they thought they had heard. The subjects were encouraged to guess. The operator was outside the test room (monitoring the talkback microphone) scoring the responses, and would ask the subject to repeat an answer deemed ambiguous or incomprehensible. When the operator was satisfied, the computer was cued to play the next test sentence. The score for each sentence was computed as the percentage of words correctly identified; all words were counted, and no part-scores were given.

Subjects completed the test over three different testing sessions, usually on different days. Each session consisted of two runs of about 57 sentences, separated by a brief break to avoid fatigue. The subjects listened to several practice sentences before each run. In total, one run of 57 sentences took about 20 minutes; the session for the day taking just under an hour, including breaks.

B. Test 2: Thresholds

A second test was conducted as a follow-up to the previous one, consisting of an additional 160 test sentences. These resulted from the 32 combinations of 2 wall types (G13 and G16) x 2 noise types (Neutral and Bass Boost) x 8 signal-to-noise ratios, with 5 different sentences for each combination. These cases were distributed over the range of difficulty from all subjects able to understand at least one word, through to all subjects unable to detect the presence of speech at all. The range of notional "source" speech levels resulted in measured "transmitted" levels ranging from 20–45 dBA. The background noise levels ranged from 43–46 dBA. The "best" listeners from the previous test were used as subjects. ("Best" means these subjects correctly identified the most words, that is, they were the "worst-case" listeners from a security point of view.)

The test was conducted in a manner exactly like the previous one, except in how the subjects responded. They still spoke aloud, but were asked to respond to the following questions, written on an instruction sheet: *1.) Did you hear any speech sounds? 2.) If yes, did you hear the rhythm or cadence of the speech? 3.) If yes, did you understand any of the words? (Tell the experimenter the words you were able to understand.)*

C. Subjects

Subjects participating in the tests were volunteers; all were fluent English speakers, and none was compensated for participating. All respondents (54 in total) were given a standard audiometric hearing threshold test and a short trial of the intelligibility test, spanning all "difficulty" levels. The 36 subjects correctly identifying 62% or more of the words in the trial participated in the main test (Test 1). The mean intelligibility score (across all 340 sentences) for each of these 36 subjects was computed and was used to split the group into the "better" subjects (the 19 scoring greater than 65%) and the "worse" subjects (the 17 scoring less than 65%). These "better" 19 subjects additionally participated in Test 2. The analyses in this paper use only the scores for these 19 subjects, for both Tests 1 and 2.

From the data provided in Ref. 14, an "average" otologically normal listener was constructed by averaging the hearing level (HL) data for 30 and 40 year old males and females. The 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of these HL curves are shown in Fig. 7. Also shown in the figure is the average HL curve for the 19 "better" subjects from the tests. 25% of the population of average 30/40-year old male/female listeners can be expected to have hearing as good or better than this. We see therefore that our subjects had excellent hearing, but not "unreasonably" so.

Knowing what portion of the population was used as test subjects is important, but does not necessarily enable "extrapolation" of the test results to groups with less sensitive hearing. For certain, people with worse hearing will do more poorly on the listening tests (and therefore be less of a risk, from a security point of view). However, there are factors other than hearing loss that can cause reduced scores (for example, attention span, ability and desire to concentrate, native language).

IV. RESULTS AND ANALYSIS

A. Intelligibility Score

Figure 8 shows the intelligibility score results for all the sentences in both tests combined. Each panel of the figure shows the 19 x (340 + 160) = 9500 individual intelligibility scores for the 19 subjects for each sentence, as a percentage of the words correctly identified, plotted versus the various objective indicators computed from the speech and noise spectra (the dots). Overlaid on the plots are the least-squares best-fit Boltzmann curve to the mean of the data (the dashed curve), and fits to the 5^{th}, 50^{th}, and 95^{th} percentile curves (solid curves, from bottom to top, respectively). The N^{th} percentile curve is defined so that (100–N)% of the recorded intelligibility scores exceed the values on the line; that is, (100–N)% of the data points lie above the line, N% lie below. The square of the correlation coefficient R^2 (the "coefficient of determination") of the relationship between the scores and the best-fit curve is given. See Appendix A for the equation of the Boltzmann curve, and the fit parameters for each (Table III).

The squared correlation coefficients can be used to evaluate the goodness of the indicators as predictors of intelligibility. Values closer to unity indicate better relationships, closer to zero indicate worse ones. The R^2 for the A-weighted level difference (0.464) and the loudness ratio (0.565) are the lowest. AI and SII yield R^2 values of 0.726 and 0.745, respectively. The other six indices are weighted signal-to-noise ratios: three different weightings (uniform, AI, SII) for each of two different signal-to-noise clipping levels (L = –22 and –32 dB). The highest R^2 values are for the SII-weighting, R^2 = 0.762 for –22 dB clipping, and R^2 = 0.757 for –32 dB clipping. The AI-weighted indices yield R^2 = 0.755 and R^2 = 0.750 for –22 dB and –32 dB clipping, respectively. The uniformly-weighted indices result in R^2 = 0.750 for both –22 and –32 dB clipping. These clipping levels were selected after having assessed the relationships between the test results and the indices for clipping from –12 dB down to –32 dB, in 2 dB steps. These results are discussed in Appendix B.

All of the weighted signal-to-noise indices presented in Fig. 8 are well-correlated with the intelligibility scores. The differences among the R^2 values are statistically significant, but it is not clear that they are all practically significant. In general, the SII-weighted measures correlate slightly better than the others, the highest for –22 dB clipping of the signal-to-noise level difference.

Notice that AI and SII have the problem that at their minimum value of zero, subjects are still able to correctly identify words from the test sentences. They are therefore confirmed to be unsuitable as a security measure.

Figure 9 shows the curve fits to the 5th, 25th, 50th, 75th, and 95th percentiles of the intelligibility score data collected from both listening tests, plotted versus SII-weighted, –22 dB clipped signal-to-noise ratio. The parameters defining these Boltzmann curves are given in Appendix A in Table IV. Since the weights sum to unity, the minimum value of the measure is –22 dB, which is the endpoint of these curves. What the figure shows, for instance, is that at an index value of –15 dB, 75% of the subjects could identify about 10% of the overheard words, 50% of the subjects could identify 23% of the words, and only the best 5% of the subjects could identify 92% of the words. Dropping an additional 5 dB to an index value of –20 dB, only 25% of the subjects could identify more than 4% of the words, and only the best 5% of subjects could identify 11% of the words.

B. Thresholds

For each of the 340+160 = 500 sentences in both tests, a tally was made of the percentage of the 19 subjects able to correctly identify at least one word. These data are plotted in Fig. 10 versus the objective indices (the dots). Also shown in each panel is the least-squares Boltzmann function fit, and associated R^2 value (function parameters are given in Appendix A in Table V). These are the Threshold of Intelligibility results. Appendix B discusses R^2 values for correlations with indices computed with other clipping levels.

Notice that as above for the word intelligibility scores, A-weighted level difference and loudness ratio were the poorest indicators (R^2 = 0.586 and 0.737, respectively). The R^2 values for AI and for SII were 0.889 and 0.904, respectively. The –22 dB clipped signal-to-noise ratios had the highest R^2 values (0.919, 0.910, and 0.908 for the SII, AI, and uniform weights, respectively). The R^2 for the –32 dB clipped schemes were: 0.905 for SII weighting, 0.896 for AI weighting, and 0.900 for uniform-weighted.

For the 160 sentences of Test 2, a tally was made of the fraction of the 19 subjects able to: 1.) identify the cadence or rhythm of the speech (including those identifying words), and 2.) hear the presence of speech in the background noise (including those identifying cadence or words). These results are shown in Fig. 11, for the Threshold of Cadence, and in Fig. 12 for the Threshold of Audibility. Each panel shows the percentage of subjects (the dots), the least-squares Boltzmann function fit, and the R^2 value for the fit (see Table V in Appendix A for the fit parameters). Appendix B discusses R^2 values for correlations with indices computed with other clipping levels.

For the threshold of cadence analysis, R^2 was 0.918 for the A-weighted level difference, 0.956 for the loudness ratio. The R^2 values for AI and SII were 0.672 and 0.770, respectively. The –22 dB clipped measures had R^2 values of 0.912, 0.798, and 0.815 for uniform, AI, and SII weighting, respectively. The –32 dB clipped measures had R^2 values of 0.858, 0.686, and 0.691 for uniform, AI, and SII weighting, respectively.

For the threshold of audibility analysis, R^2 was 0.835 for the A-weighted level difference, 0.899 for the loudness ratio. The R^2 values for AI and SII were 0.389 and 0.566, respectively. The –22 dB clipped measures had R^2 values of 0.816, 0.681, and 0.693 for uniform, AI, and SII weighting, respectively. The –32 dB clipped measures had R^2 values of 0.741, 0.581, and 0.583 for uniform, AI, and SII weighting, respectively.

In general, for both cadence and audibility thresholds, the A-weighted level difference and loudness ratio are superior indicators than the signal-to-noise ratio type measures, including AI and SII. The thresholds of cadence and of audibility involve the detection of sounds, not necessarily understanding of speech. They are more related to audibility and loudness than to intelligibility; the strong correlations with measures of such (loudness, A-weighted levels) indicate this. The –22 dB clipped measures were better-correlated than the –32 dB clipped, but inspection of the figure indicates that the best-fit curves do not reach zero. This is analogous to the problem with AI and SII for intelligibility scores. For both thresholds, the uniformly-weighted measures are better-correlated than the AI and SII weighted ones. This possibly indicates the relative importance of the lower frequencies (below 800 Hz) for audibility, which are reduced by the intelligibility-derived weighting schemes, but are relatively important for transmitted speech sounds (see Fig. 5).

Figure 13 shows the curves for the threshold of cadence and the threshold of audibility: the percent of subjects able to detect the cadence or any speech sounds, versus A-weighted level difference. At –10 dB (i.e., speech level 10 dB *lower* than background noise level), 95% of the subjects could identify that there was speech, and 77% of the subjects could identify its cadence. 5 dB lower than this, at a level difference of –15 dB, 67% of subjects could hear the speech sounds, 27% identifying the cadence. The threshold of intelligibility curve was not included in Fig. 13 since the correlation with A-weighted level difference is so poor (R^2 = 0.586). A-weighted level difference is not a good measure for assessment of threshold of intelligibility.

A measure that is well-correlated with all three thresholds is the uniformly-weighted –32 dB clipped signal-to-noise. This measure can be used to assess all thresholds, and indicate relative relations among them. The usefulness of uniformly-weighted band levels has been explored by Tachibana *et al*. as an indicator of loudness.[15] This ties in well with the above-noted observation that detection of the thresholds of audibility and of cadence has to do with the loudness of the speech sounds. Figure 14 shows the threshold curves for the percentage of subjects able to correctly identify at least one word (Intelligibility), able to identify at least the cadence (Cadence), and able to at least hear some speech sounds (Audibility) versus uniformly-weighted, –32 dB clipped signal-to-noise index. For a measure value of –15 dB, 98% of the subjects could hear some speech sounds, and 60% could identify at least one word. 5 dB lower, at a value of –20 dB, only 8% of subjects could identify at least one word. At an index value of –25 dB, less than 1% of subjects could identify a word, and only 20% could hear speech sounds. If one considers 50% of the subjects as a threshold point (as is the norm), then the threshold of intelligibility could be said to be at about –15.5 dB, the threshold of audibility about 7 dB lower, at –22.5 dB.

V. CONCLUSIONS

Listening tests simulating speech transmission through a range of typical office wall constructions have been used to find objective measures of intelligibility and audibility suitable for architectural speech security situations. A speech-signal-to-noise ratio, restricting the 1/3-octave-band level differences to –22 dB and weighted using the band importance frequency weights from the SII calculation, was found to be a good measure of speech intelligibility. It has also been found that, in cases of zero intelligibility, both loudness ratio and simple A-weighted level difference are able to accurately predict audibility of the speech or its cadence. These measures (loudness ratio and A-weighted level difference) should not, however, be used to assess intelligibility. A uniformly-weighted 1/3-octave-band signal-to-noise ratio clipped to –32 dB is a good indicator for all three thresholds.

The existing measures AI and SII are not suitable for evaluating speech security. While they are highly-correlated with the listening test intelligibility scores, both fail to indicate zero intelligibility at their minimum values of zero. Furthermore, they provide no information regarding the thresholds of audibility and of cadence.

These results indicate that the Speech-Privacy Calculation,[5] which uses estimates of A-weighted signal-to-noise ratio to assess privacy relative to AI=0.05, is not ideal for assessing the threshold of intelligibility.

The relationships derived between intelligibility scores and thresholds and the objective measures are "worst-case" from a security point of view. They are therefore broadly applicable. The listeners were acute-hearing and cued to expect to overhear speech, the speaking voice was strong and clear. Predicting a level of security from this work should err on the conservative side relative to listeners with less sensitive hearing and/or talkers with less clear voices.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of Ms. Kimberlee Cuthbert, who was responsible for scheduling subjects, running subject experiments, and assembling the measurement data, and Mr. Ketana Teav, who set up much of the equipment and wrote control software for the computer and hardware. We also would like to thank the subjects who each volunteered several hours of their time to help us do this work.

APPENDIX A: BOLTZMANN FUNCTION FITS

The sigmoidal curves fitted to the data from the listening tests are Boltzmann functions.[16] The functional form of this curve is given by

$$F(x) = 1 - \frac{1}{1 + \exp\left(\frac{x - A}{B}\right)} \tag{A1}$$

where A and B are parameters defining the midpoint of the rise and the slope, respectively. The parameter A is the abscissa value corresponding to the 50% point of the curve: $F(A) = 0.5$. The curve has asymptotes of +1 at $x = \infty$ and 0 at $x = -\infty$. The values of A and B for the intelligibility score curves shown in Fig. 8 are given in Table III, those for the intelligibility score percentile curves shown in Fig. 9 are given in Table IV, and those for the threshold curves shown in Figs. 10–12 are given in Table V.

APPENDIX B: SIGNAL-TO-NOISE CLIPPING LEVEL

The parameter L in Eq. (2) defines the minimum value of the signal-to-noise level difference that is taken as contributing to the measure $X_w(L)$. The results of the listening tests were correlated with $X_w(L)$ for L = –12 dB to L = –32 dB, in 2 dB steps for the three weighting strategies (uniform, AI, and SII). Figure 15 shows the resulting R^2 values for the relationships of the measures with each of: (a) intelligibility scores, (b) threshold of intelligibility, (c) threshold of cadence, and (d) threshold of audibility. Clipping below –32 dB does not appreciably change the relationships; little of the collected data lies in that range.

For the intelligibility scores and the threshold of intelligibility, the correlations are much the same for all three weightings; the SII only slightly higher (notice the scale on the ordinate axis). There is a peak at around –22 dB, which is also about the point where the fit curve trends to zero intelligibility (see Figs. 8 and 10). Clipping below this, the R^2 trends to a constant value. Above this, not only does the correlation drop, but also the data starts to clip before the intelligibility reaches zero, as happens with AI and SII.

The thresholds of cadence and audibility show that uniform weighting is better correlated than the AI and SII for all clipping levels. The R^2 actually drops with decreasing clipping level, but from inspection of Figs. 11 and 12, a level of –32 dB is necessary to ensure the data points are not clipped before audibility reaches zero.

REFERENCES

[1] ASTM E1130-02, “Standard Test Method for Objective Measurement of Speech Privacy in Open Offices Using Articulation Index,” ASTM International, West Conshohocken, PA, U.S.A.

[2] ANSI S3.5-1969, “American National Standard Methods for the Calculation of the Articulation Index,” ANSI, New York, NY, U.S.A.

[3] ANSI S3.5-1997, “American National Standard Methods for Calculation of the Speech Intelligibility Index,” ANSI, New York, NY, U.S.A.

[4] W.J. Cavanaugh, W.R. Farrell, P.W. Hirtle, and B.G. Watters, “Speech Privacy in Buildings,” J. Acoust. Soc. Am., 34, 475–492 (1962).

[5] R.W. Young, “Re-Vision of the Speech-Privacy Calculation,” J. Acoust. Soc. Am., 38, 524–530 (1965).

[6] M.D. Egan, *Architectural Acoustics* (McGraw-Hill, New York, 1988.)

[7] F. Santon, "Intelligibilité de la parole masquée et rôle de la direction de la source masquante," Acustica, 61, 67–74 (1986).

[8] "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio and Electroacoustics, 227–246 (1969).

[9] A.C.C. Warnock, "Acoustical privacy in the landscaped office," J. Acoust. Soc. Am., 53, 1535-1543 (1973).

[10] C. Speaks, B. Parker, C. Harris, and P. Kuhl, "Intelligibility of connected discourse," J. Speech and Hearing Research, 15, 590-602 (1972).

[11] W.O. Olsen, "Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson [sic], Bennett, & Fidell (1977) report," Am. J. of Audiology, 7, 1–5 (1998).

[12] ANSI S12.2-1995 (R1999), "American National Standard Criteria for Evaluating Room Noise," ANSI, New York, NY, U.S.A.

[13] ISO 532-1975 "Acoustics -- Method for calculating loudness level," ISO, Geneva, Switzerland.

[14] ISO 7029-1984 (E), "Acoustics — Threshold of hearing by air conduction as a function of age and sex for otologically normal persons," ISO, Geneva, Switzerland.

[15] H. Tachibana, Y. Sonoda, K. Iwamoto, S. Kuwano, and S. Namba, "Validity of arithmetic average of sound pressure levels in octave bands as a loudness index," J. Acoust. Soc. Jpn. (E), 14, 197–204 (1993).

[16] "Origin User's Manual Version 5," Microcal Software Inc., 1997.

TABLES

Table I Wall descriptions and sound transmission class (STC) ratings for the walls simulated in the listening tests.

Wall Descriptor	Wall Description	STC Rating
Door	Solid core wood door, no seals.	20
Plenum	5/8” mineral fibre ceiling.	32
G13	89 mm wood stud wall with 13 mm gypsum board on both sides; cavity filled with glass fibre batts.	34
G16	90 mm steel stud wall with 16 mm gypsum board on both sides; cavity filled with glass fibre batts.	46

Table II 1/3-octave-band signal-to-noise weighting strategies considered in computing a weighted signal-to-noise ratio measure.

Weighting	Frequency (Hz)															
	160	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
Uniform	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
AI	0.0000	0.0120	0.0300	0.0300	0.0420	0.0420	0.0601	0.0601	0.0721	0.0901	0.1111	0.1141	0.1021	0.1021	0.0721	0.0601
SII	0.0088	0.0101	0.0159	0.0306	0.0466	0.0612	0.0691	0.0752	0.0866	0.0893	0.0933	0.0950	0.0918	0.0893	0.0816	0.0558

Table III Values of the Boltzmann function parameters *A* and *B* from Eq. (A1) for the intelligibility score curves plotted in Fig. 8.

	Mean		5th Percentile		50th Percentile		95th Percentile	
	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*
A-weighted level difference	–5.1418	5.9696	9.3204	3.3068	–5.2999	3.6593	–18.6277	2.3182
Loudness ratio	0.4893	0.1693	1.6817	0.6347	0.4689	0.1120	0.1073	0.1047
AI	0.0767	0.0352	0.2240	0.0683	0.0756	0.0297	0.0138	0.0168
SII	0.1182	0.0459	0.2863	0.0776	0.1110	0.0428	0.0376	0.0132
Uniformly-wtd, –22 dB clipped	–11.9266	2.3234	–5.2475	2.8743	–11.9736	1.7240	–16.4794	1.1815
AI-weighted, –22 dB clipped	–12.9251	2.2389	–6.5831	2.8026	–13.0123	1.7106	–17.7037	0.9459
SII-weighted, –22 dB clipped	–12.9106	2.1987	–6.8542	2.9735	–12.8981	1.6604	–17.7130	1.1150
Uniformly-wtd, –32 dB clipped	–12.1875	2.5259	–5.2621	2.7047	–12.1053	1.9800	–17.8748	1.3434
AI-weighted, –32 dB clipped	–13.2084	2.5172	–6.4154	2.7317	–13.1395	1.9902	–18.5097	1.2247
SII-weighted, –32 dB clipped	–13.2092	2.4777	–6.8974	2.8654	–13.1850	2.0601	–18.1247	1.0317

Table IV Values of the Boltzmann function parameters *A* and *B* from Eq. (A1) for the intelligibility score percentile curves shown in Fig. 9.

	Intelligibility Score Percentiles	
	A	*B*
5th percentile	–6.8542	2.9735
25th percentile	–10.9010	1.9146
50th percentile	–12.8981	1.6604
75th percentile	–14.8641	1.6355
95th percentile	–17.7130	1.1150

Table V Values of the Boltzmann function parameters *A* and *B* from Eq. (A1) for the threshold curves shown in Figs. 10–12.

	Threshold of Intelligibility		Threshold of Cadence		Threshold of Audibility	
	A	*B*	*A*	*B*	*A*	*B*
A-weighted level difference	–9.8875	4.3067	–12.7405	2.2338	–16.6796	2.3335
Loudness ratio	0.3242	0.0882	0.2099	0.0413	0.1398	0.0288
AI	0.0271	0.0117	0.0033	0.0032	–0.0001	0.0020
SII	0.0547	0.0197	0.0150	0.0087	0.0047	0.0083
Uniformly-wtd, –22 dB clipped	–15.0566	1.5448	–18.5891	0.8918	–19.9799	0.7142
AI-weighted, –22 dB clipped	–15.8685	1.3351	–18.7729	0.9573	–20.0918	1.0158
SII-weighted, –22 dB clipped	–15.8558	1.3363	–18.8710	0.8925	–20.1302	0.9513
Uniformly-wtd, –32 dB clipped	–15.6432	1.8739	–20.0467	1.4037	–22.4119	1.8053
AI-weighted, –32 dB clipped	–16.4980	1.6114	–20.1055	2.0360	–22.9373	2.5032
SII-weighted, –32 dB clipped	–16.5059	1.6267	–20.3049	2.0092	–23.1323	2.4872

FIGURE CAPTIONS

Figure 1 Schematic of cross-section through the test room, showing the location of the listener and the loudspeakers used to generate the test sound fields.

Figure 2 Block diagram of the computer-controlled electroacoustic system used to create the test sounds. The speech, played from the computer, is processed by one of the Yamaha DME32 units to simulate transmission through a wall. The other DME32 unit generates spectrally-shaped random background noise.

Figure 3 Example measured spectrum of talker used in this work. Also shown are "typical" spectra for a male talker speaking with “Raised” and “Loud” effort as given in Ref. 11.

Figure 4 Measured transmission loss curves for the walls selected to be simulated in the tests. The four wall descriptors are explained in Table I.

Figure 5 “Transmitted” speech spectra, measured in the test room after filtering the source speech (top curve) through each simulated wall.

Figure 6 Measured noise spectra, all corresponding to 45 dBA. The "Neutral" spectrum has a –5-dB/octave roll-off. The other spectra were derived from this by boosting the bass frequencies (50–200 Hz), mid-frequencies (250–1600 Hz), or high frequencies (2000–10000 Hz) by 10 dB.

Figure 7 Average HL curve for the 19 subjects shown with 10th, 25th, 50th, 75th, and 90th percentile HL curves for an "average" otologically normal listener. The percentiles give the percentage of the population with hearing loss worse than the curve. The "average" listener response is the mean of the 30-year old and 40-year old male and female responses given in Ref. 14.

Figure 8 Individual intelligibility scores for the 19 subjects from 500 sentences each (9500 points total), plotted versus the various indices. The dashed line is the least-squares Boltzmann function fit to the mean, the squared correlation coefficient (R^2) for this fit is shown. The three solid lines are the least-squares Boltzmann function fits to the 5th, 50th, and 95th percentiles (from bottom to top, respectively).

Figure 9 Intelligibility score versus SII-weighted signal-to-noise. The curves from bottom to top represent the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles of the 9500 individual responses.

Figure 10 Threshold of intelligibility: The dots are the percentage of the 19 subjects correctly identifying at least one word from each of the 340+160 sentences in both tests

(500 dots total). The solid curve is the least-squares Boltzmann function fit to the mean. The squared correlation coefficient values (R^2) for the fits are shown.

Figure 11 Threshold of cadence: The dots are the percentage of the 19 subjects identifying the cadence of speech (including those correctly identifying some words) from each of the 160 sentences in the second test (160 dots total). The solid curve is the least-squares Boltzmann function fit to the mean. The squared correlation coefficient values (R^2) for the fits are shown.

Figure 12 Threshold of audibility: The dots are the percentage of the 19 subjects identifying the presence of speech sounds (including those identifying cadence and those identifying some words) from each of the 160 sentences in the second test (160 dots total). The solid curve is the least-squares Boltzmann function fit to the mean. The squared correlation coefficient values (R^2) for the fits are shown.

Figure 13 Thresholds of cadence and audibility versus A-weighted level difference. The curves represent the percentage of subjects able to at least identify the cadence of speech (Cadence) or to at least detect the presence of speech sounds (Audibility).

Figure 14 Thresholds versus uniformly-weighted, –32 dB clipped signal-to-noise index. Each curve represents the percentage of subjects able to correctly identify: at least one word from a sentence (Intelligibility), the cadence of speech (Cadence), or the presence of speech sounds (Audibility).

Figure 15 Squared correlation coefficient (R^2) versus minimum band signal-to-noise clipping level. The correlation describes the relationship between the subjective test responses and the measure calculated using the corresponding level, for each of three weightings (uniform, AI, and SII) for (a) intelligibility scores, (b) threshold of intelligibility, (c) thresholds of cadence, and (d) threshold of audibility.

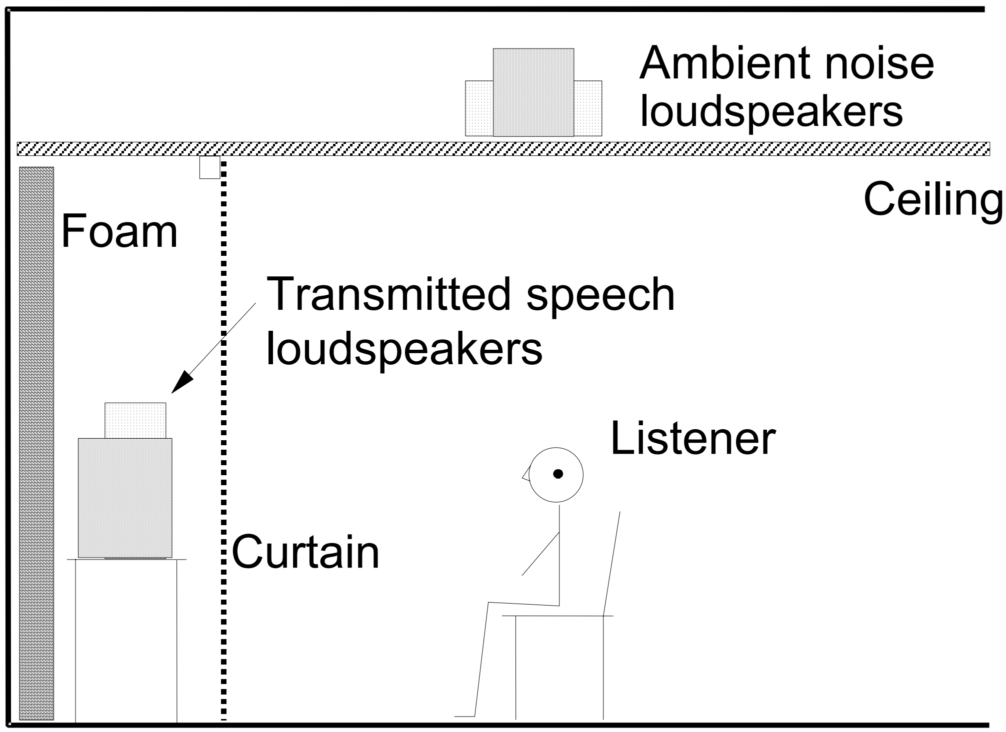

Ambient noise
loudspeakers
Ceiling
Foam
Transmitted speech
loudspeakers
Listener
Curtain

Figure 1

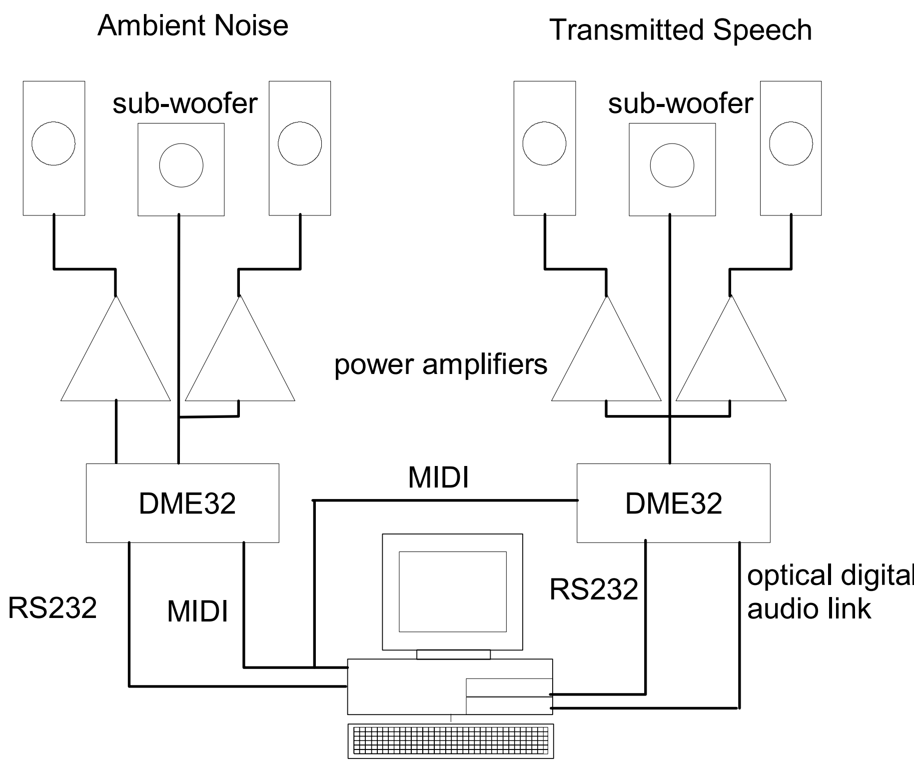

Ambient Noise
Transmitted Speech
sub-woofer
sub-woofer
power amplifiers
MIDI
DME32
DME32
RS232
MIDI
RS232
optical digital
audio link

Figure 2

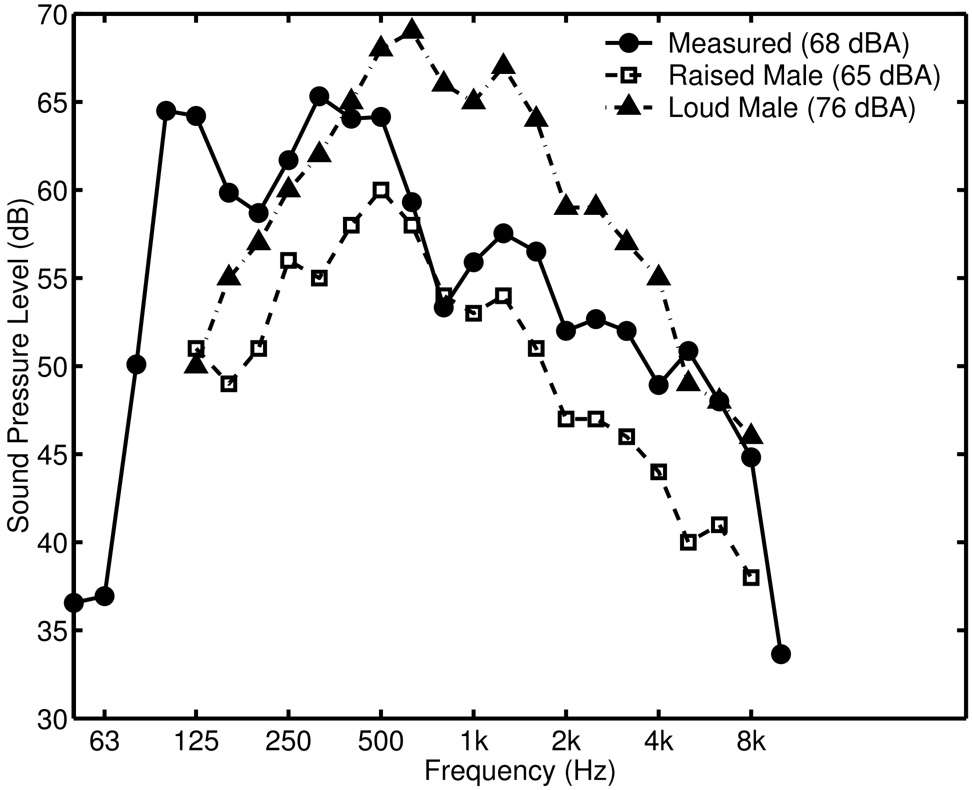

Measured (68 dBA)
Raised Male (65 dBA)
Loud Male (76 dBA)
Sound Pressure Level (dB)
70
65
60
55
50
45
40
35
30
63
125
250
500
1k
2k
4k
8k
Frequency (Hz)

Figure 3

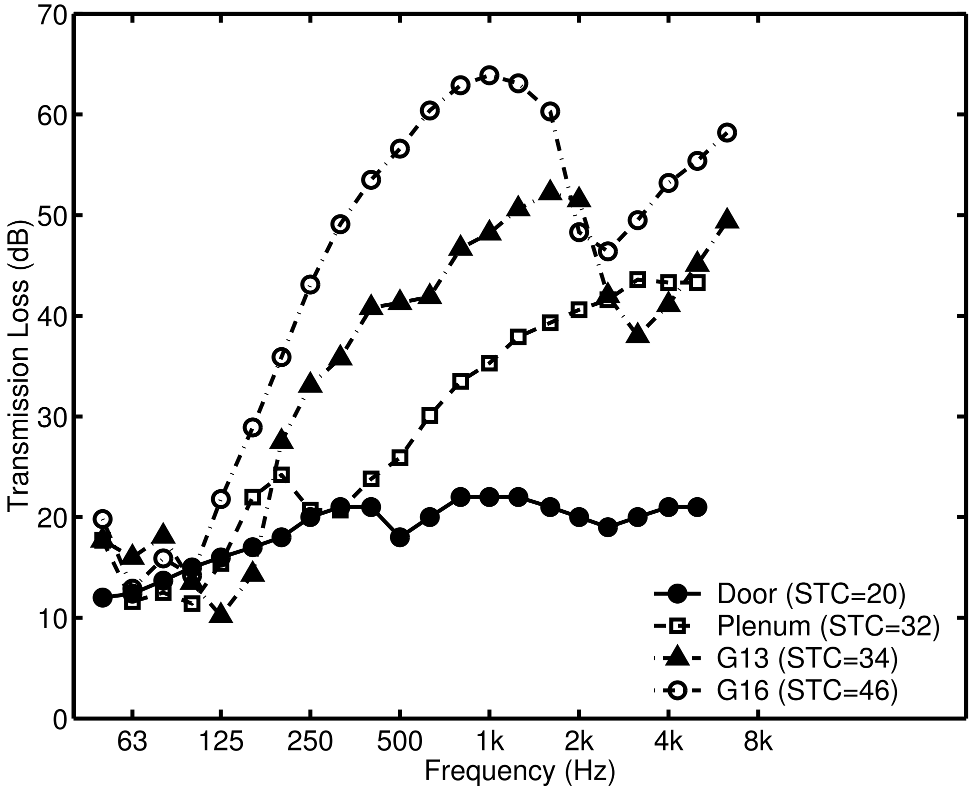
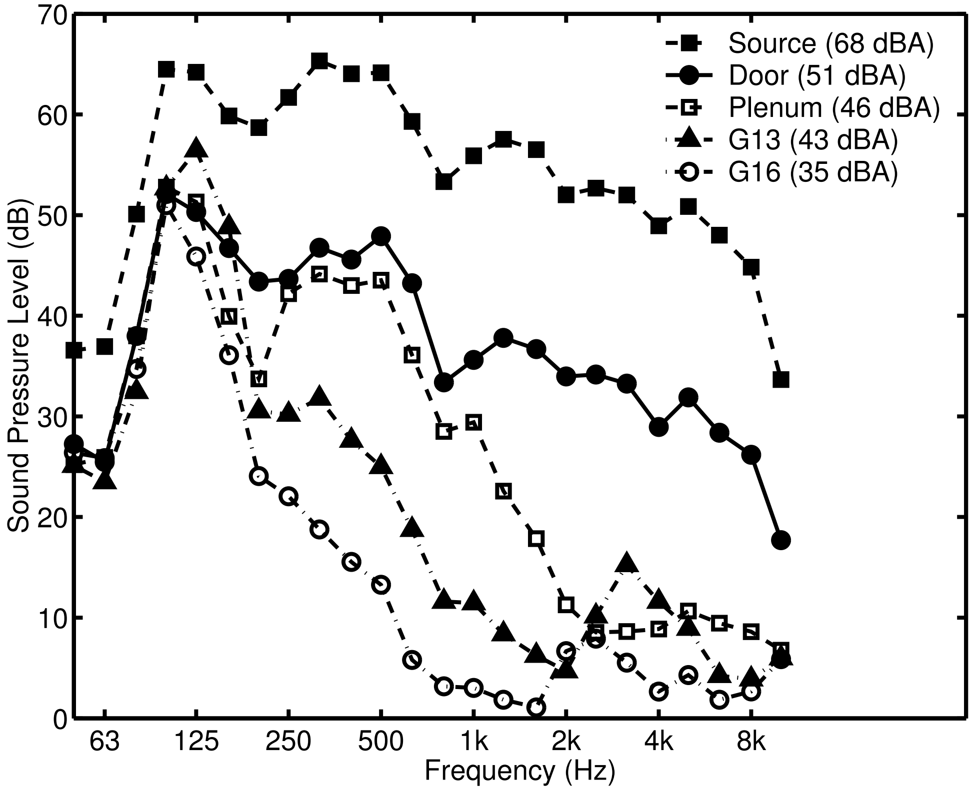

Transmission Loss (dB)
Frequency (Hz)
0
10
20
30
40
50
60
70
63
125
250
500
1k
2k
4k
8k
Door (STC=20)
Plenum (STC=32)
G13 (STC=34)
G16 (STC=46)

Figure 4



Source (68 dBA)
Door (51 dBA)
Plenum (46 dBA)
G13 (43 dBA)
G16 (35 dBA)
Sound Pressure Level (dB)
0
10
20
30
40
50
60
70
63
125
250
500
1k
2k
4k
8k
Frequency (Hz)

Figure 5

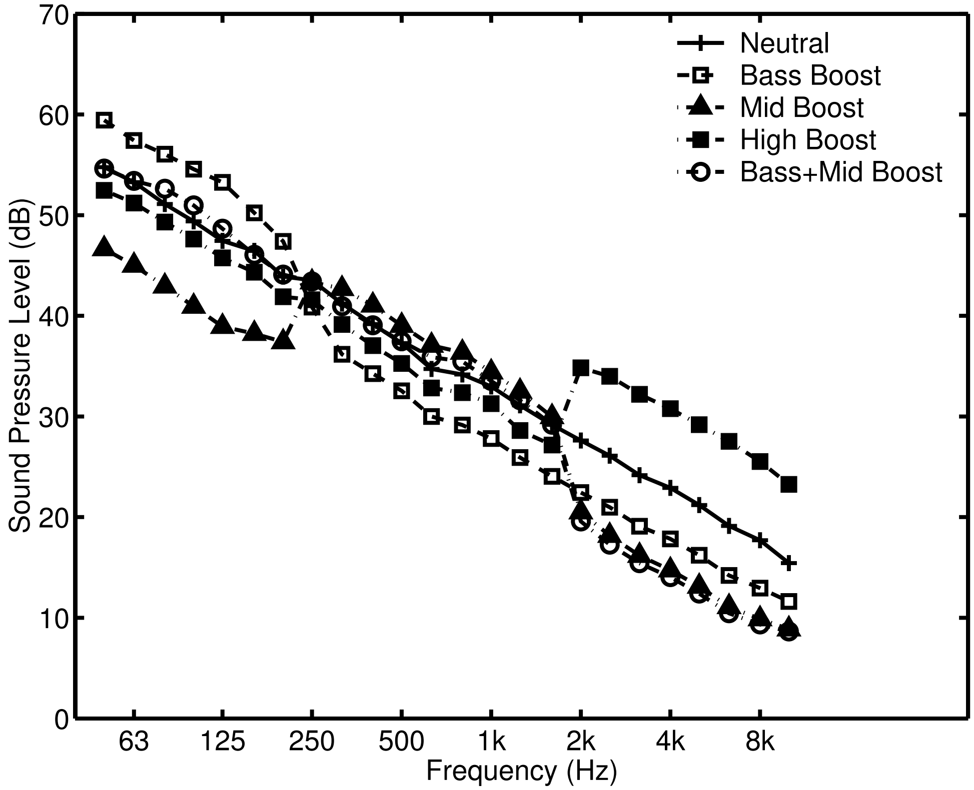
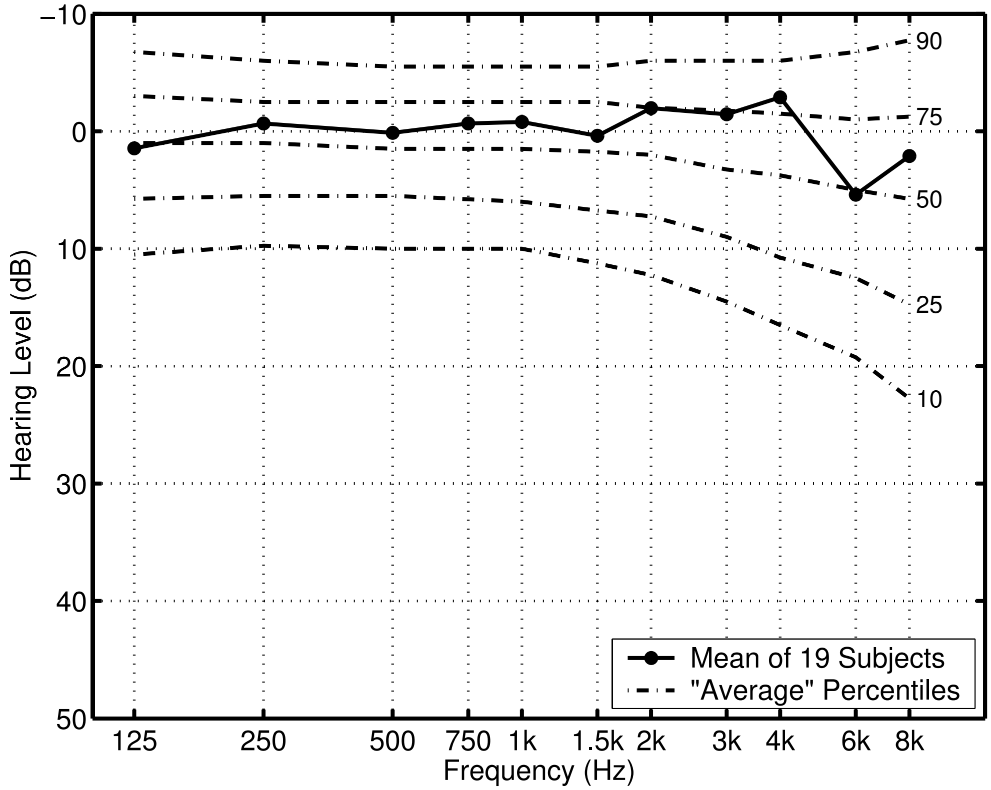

70
60
50
40
30
20
10
0
Sound Pressure Level (dB)
63
125
250
500
1k
2k
4k
8k
Frequency (Hz)
Neutral
Bass Boost
Mid Boost
High Boost
Bass+Mid Boost

Figure 6



−10
0
10
20
30
40
50
Hearing Level (dB)
125
250
500
750
1k
1.5k
2k
3k
4k
6k
8k
Frequency (Hz)
90
75
50
25
10
Mean of 19 Subjects
"Average" Percentiles

Figure 7

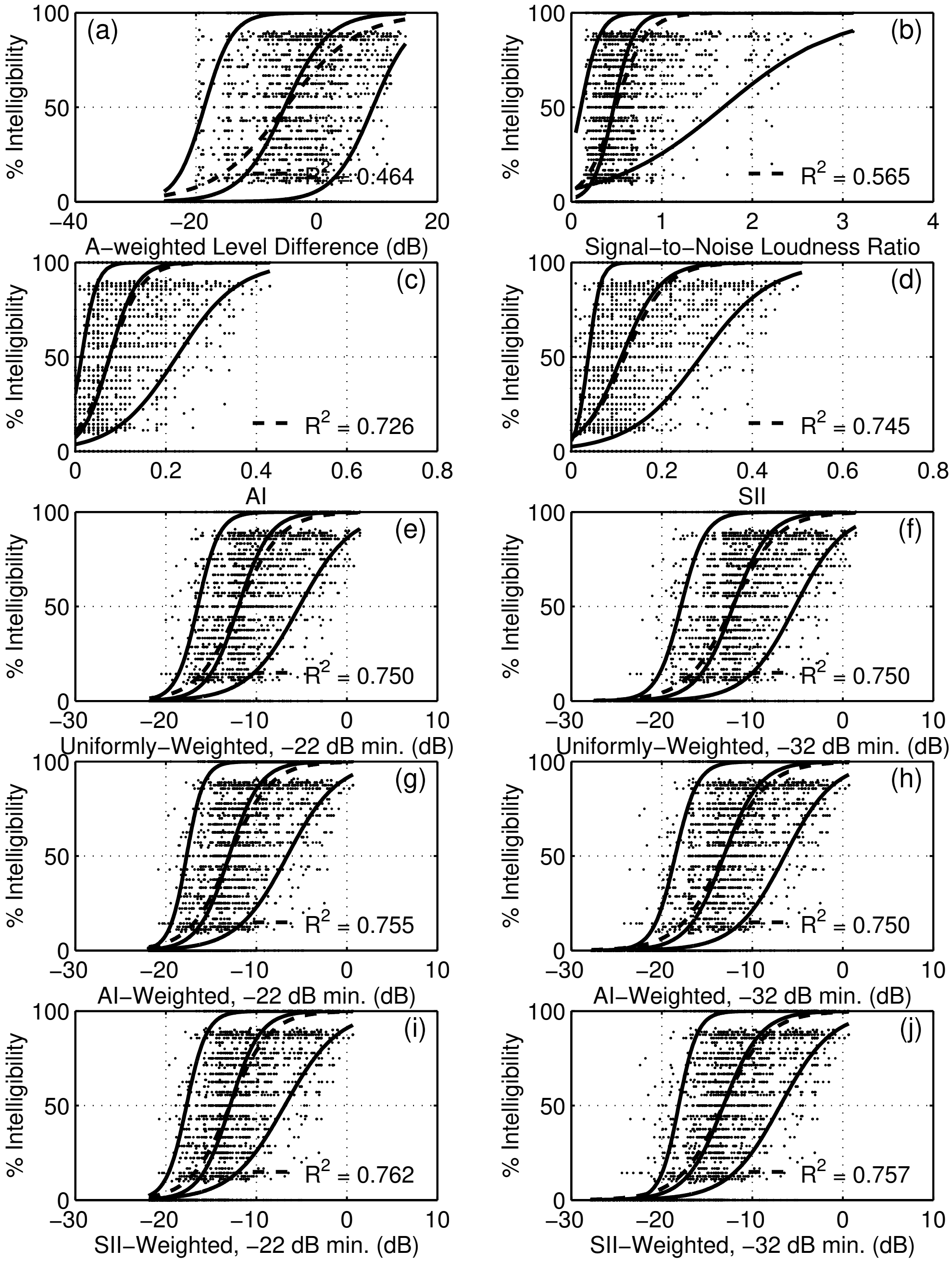

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
% Intelligibility
A−weighted Level Difference (dB)
Signal−to−Noise Loudness Ratio
AI
SII
Uniformly−Weighted, −22 dB min. (dB)
Uniformly−Weighted, −32 dB min. (dB)
AI−Weighted, −22 dB min. (dB)
AI−Weighted, −32 dB min. (dB)
SII−Weighted, −22 dB min. (dB)
SII−Weighted, −32 dB min. (dB)
R2 = 0.464
R2 = 0.565
R2 = 0.726
R2 = 0.745
R2 = 0.750
R2 = 0.750
R2 = 0.755
R2 = 0.750
R2 = 0.762
R2 = 0.757

Figure 8

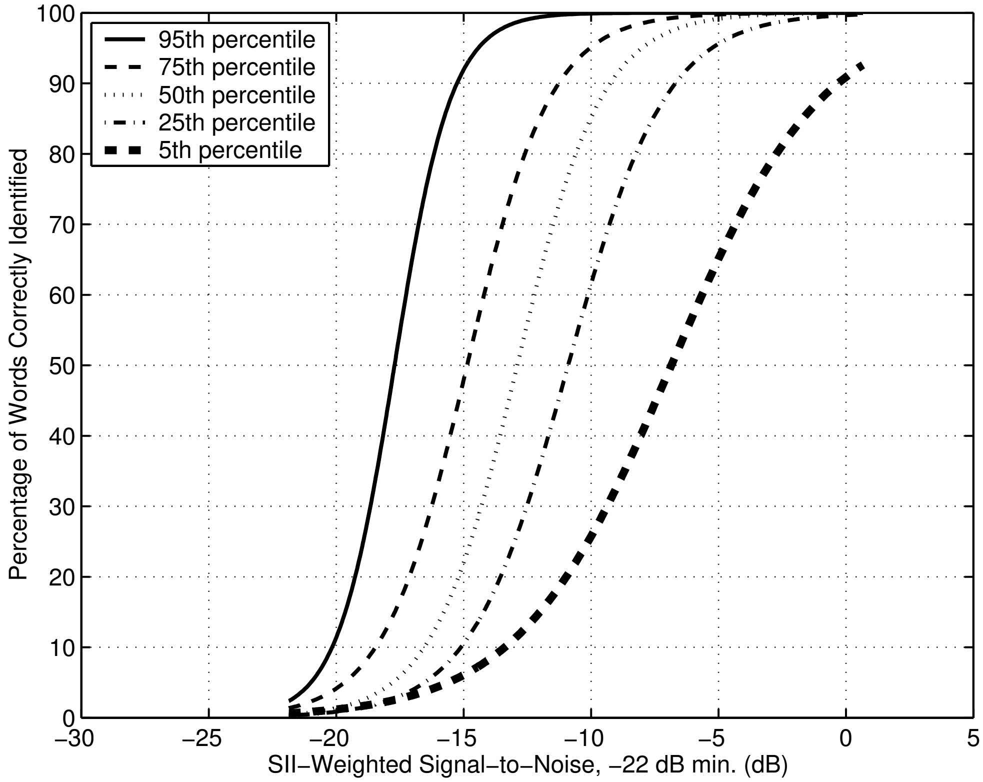

95th percentile
75th percentile
50th percentile
25th percentile
5th percentile
Percentage of Words Correctly Identified
SII−Weighted Signal−to−Noise, −22 dB min. (dB)
100
90
80
70
60
50
40
30
20
10
0
−30
−25
−20
−15
−10
−5
0
5

Figure 9

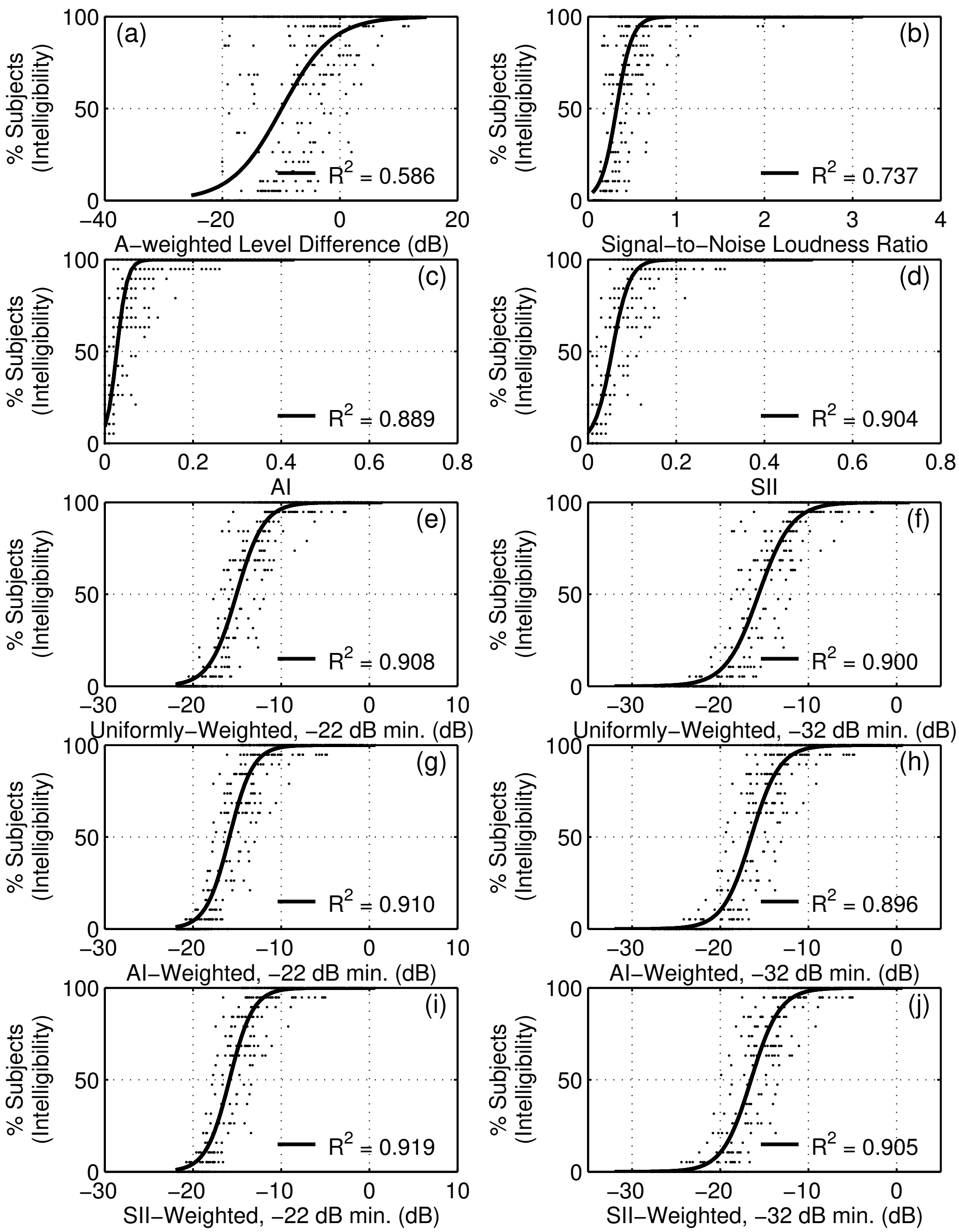

(a)
% Subjects (Intelligibility)
R^2 = 0.586
A−weighted Level Difference (dB)
(b)
R^2 = 0.737
Signal−to−Noise Loudness Ratio
(c)
R^2 = 0.889
AI
(d)
R^2 = 0.904
SII
(e)
R^2 = 0.908
Uniformly−Weighted, −22 dB min. (dB)
(f)
R^2 = 0.900
Uniformly−Weighted, −32 dB min. (dB)
(g)
R^2 = 0.910
AI−Weighted, −22 dB min. (dB)
(h)
R^2 = 0.896
AI−Weighted, −32 dB min. (dB)
(i)
R^2 = 0.919
SII−Weighted, −22 dB min. (dB)
(j)
R^2 = 0.905
SII−Weighted, −32 dB min. (dB)

Figure 10

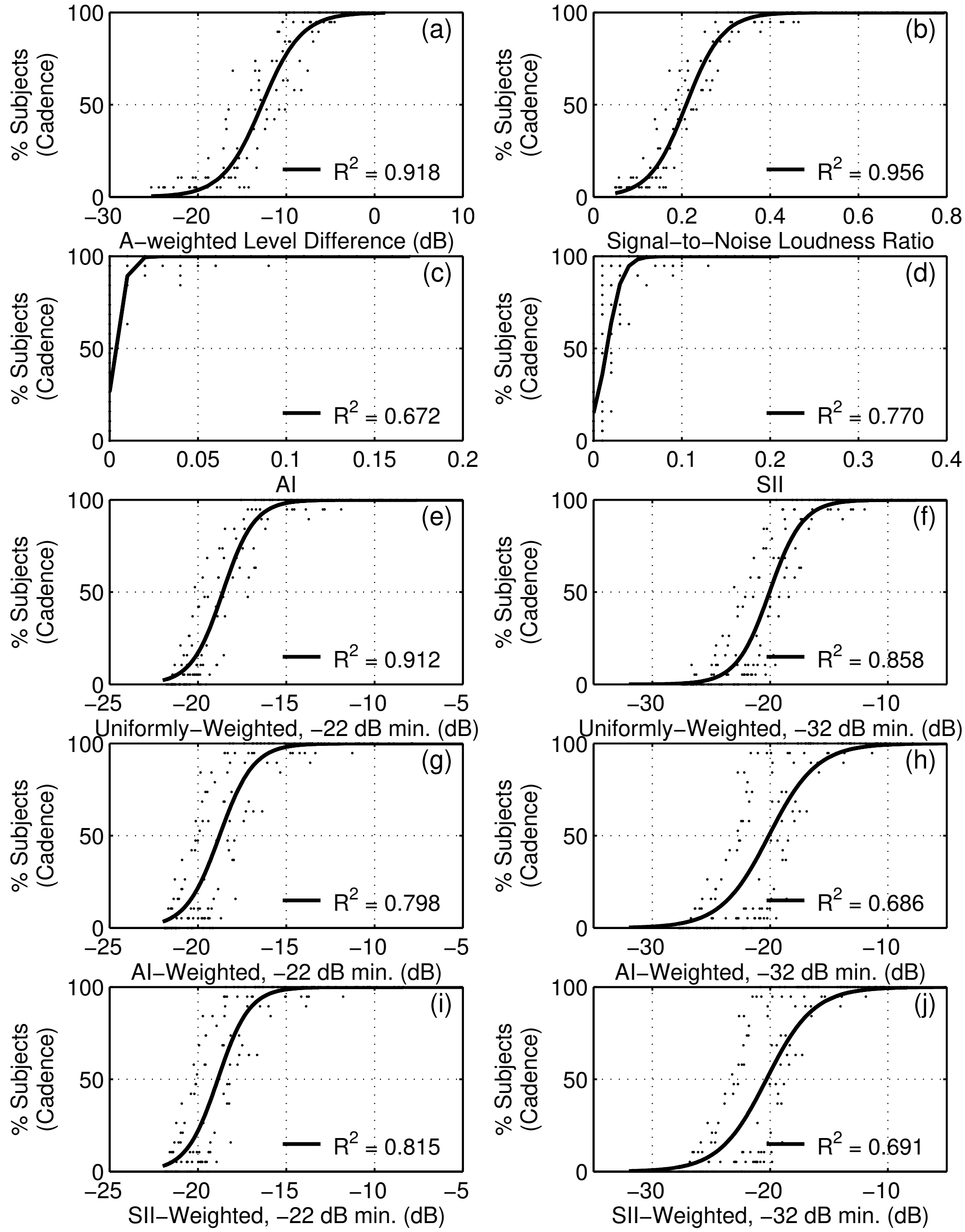
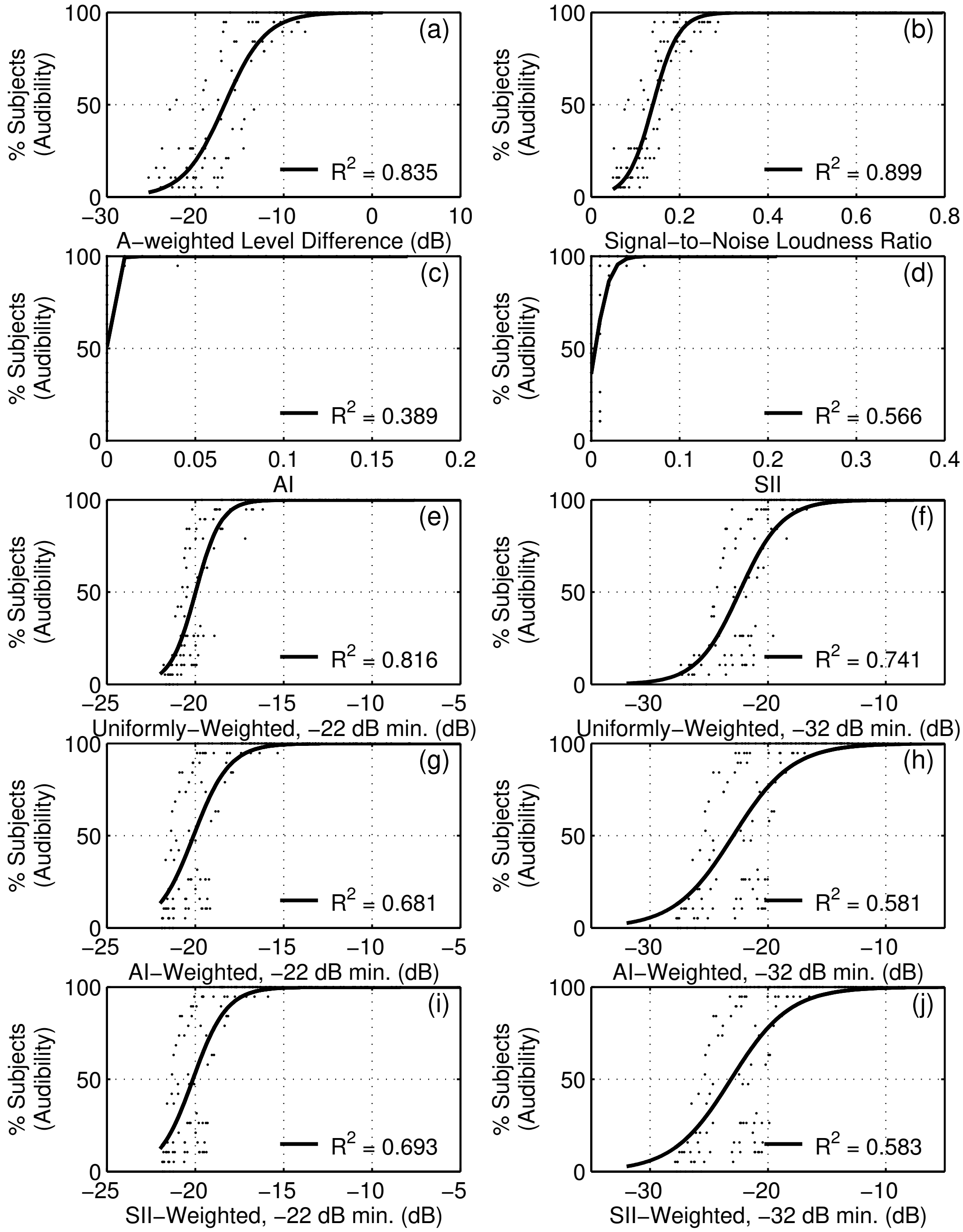
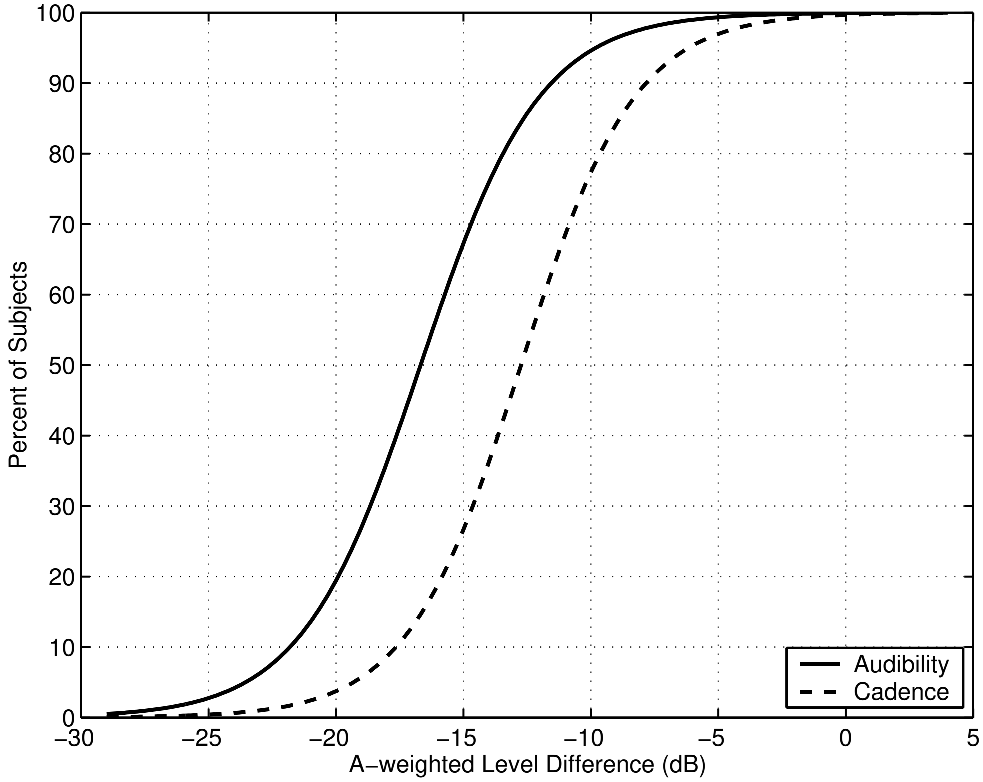

% Subjects (Cadence)
(a)
R² = 0.918
A−weighted Level Difference (dB)
(b)
R² = 0.956
Signal−to−Noise Loudness Ratio
(c)
R² = 0.672
AI
(d)
R² = 0.770
SII
(e)
R² = 0.912
Uniformly−Weighted, −22 dB min. (dB)
(f)
R² = 0.858
Uniformly−Weighted, −32 dB min. (dB)
(g)
R² = 0.798
AI−Weighted, −22 dB min. (dB)
(h)
R² = 0.686
AI−Weighted, −32 dB min. (dB)
(i)
R² = 0.815
SII−Weighted, −22 dB min. (dB)
(j)
R² = 0.691
SII−Weighted, −32 dB min. (dB)

Figure 11



(a)
% Subjects (Audibility)
R^2 = 0.835
A−weighted Level Difference (dB)
(b)
R^2 = 0.899
Signal−to−Noise Loudness Ratio
(c)
R^2 = 0.389
AI
(d)
R^2 = 0.566
SII
(e)
R^2 = 0.816
Uniformly−Weighted, −22 dB min. (dB)
(f)
R^2 = 0.741
Uniformly−Weighted, −32 dB min. (dB)
(g)
R^2 = 0.681
AI−Weighted, −22 dB min. (dB)
(h)
R^2 = 0.581
AI−Weighted, −32 dB min. (dB)
(i)
R^2 = 0.693
SII−Weighted, −22 dB min. (dB)
(j)
R^2 = 0.583
SII−Weighted, −32 dB min. (dB)

Figure 12



100
90
80
70
60
50
40
30
20
10
0
Percent of Subjects
−30
−25
−20
−15
−10
−5
0
5
A−weighted Level Difference (dB)
Audibility
Cadence

Figure 13

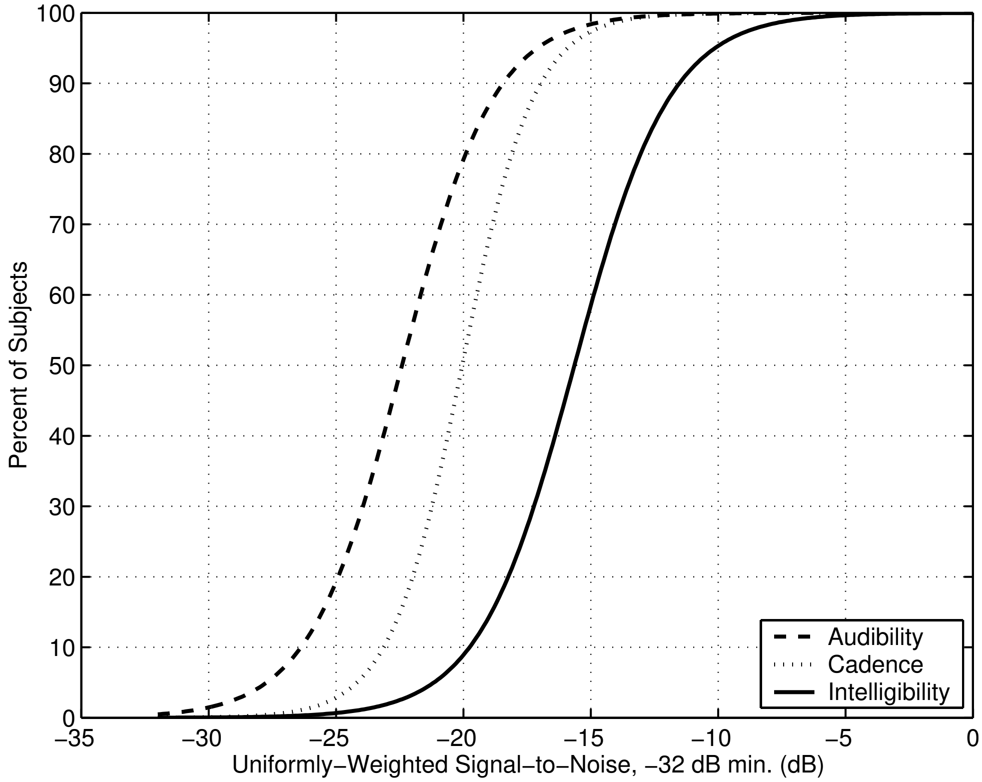

100
90
80
70
60
50
40
30
20
10
0
Percent of Subjects
−35
−30
−25
−20
−15
−10
−5
0
Uniformly−Weighted Signal−to−Noise, −32 dB min. (dB)
Audibility
Cadence
Intelligibility

Figure 14

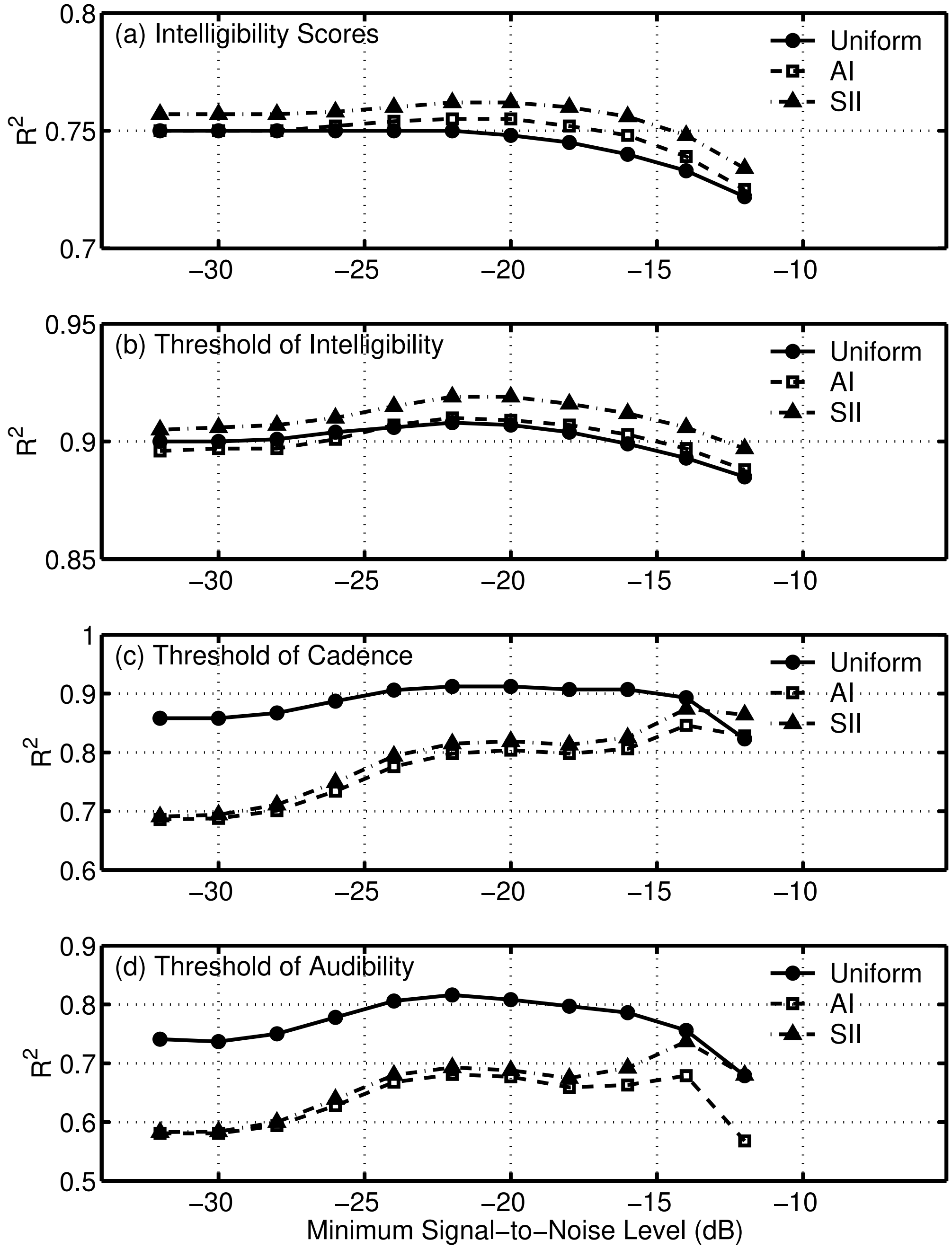

(a) Intelligibility Scores
(b) Threshold of Intelligibility
(c) Threshold of Cadence
(d) Threshold of Audibility
Uniform
AI
SII
R^2
Minimum Signal−to−Noise Level (dB)

Figure 15