ICSV13 - Vienna
The Thirteenth International Congress on Sound and Vibration
Vienna, Austria, July 2-6, 2006

PREDICTING SPEECH INTELLIGIBILITY AND SECURITY USING ARTIFICIAL NEURAL NETWORK MODELS

Jingfeng Xu*[1, 2]

[1] Arup Acoustics
Level 10, 201 Kent Street, Sydney, NSW 2000, Australia
[2] Faculty of Architecture, The University of Sydney
NSW 2006, Australia
Jingfeng.xu@arup.com

Abstract

Artificial neural network models for predicting speech intelligibility scores and security thresholds have been developed in a previous work [Xu *et al.*, "An artificial neural network approach for predicting architectural speech security," Journal of the Acoustical Society of America, 117 (4), pp 1709-1712, 2005]. The present work uses an application example to show in detail how these models can be embedded into a spreadsheet application and implemented in the design stage. Using the same example, the present work also investigates how the speech intelligibility scores and security thresholds vary as a result of using different constructions for the common partition between the speech sound source room and speech sound receiving room. Results of the investigation show that, when the speech sound level is 68 dB(A) in the speech sound source room and the background noise level is 39 dB(A) in the speech sound receiving room, for a typical setup of private offices, 30% of overhead words would be intelligible when the construction of the common partition has an STC rating of around 45; only a very small percentage of the overheard words would be intelligible when the STC rating of the common partition is increased to 50; all speech sounds from the speech sound source room would be completely inaudible when the STC rating of the common partition is increased to 60.

INTRODUCTION

Speech intelligibility and privacy have been related to signal-to-noise (S/N) type measures, where the signal is the transmitted speech sound in the adjacent space and the noise is the background noise in the adjacent space. Cavanaugh *et al.* [1] presented a report on occupants' *impressions* of privacy in buildings and stated that the most critical 10% of the subjects began to *feel* a lack of privacy when the

Articulation Index (AI) [2] reached 0.05. In their work, the assessment of speech privacy was based on how private the subjects *felt* a situation was. The actual fraction of the speech, which the subjects could understand was, however, not measured. On the basis of the data in Cavanaugh *et al.*, Young [3] revised the computational procedure and proposed a measure derivable from A-weighted levels of speech and noise and single number ratings of partitions' transmission losses. Young's method is easy to calculate and has become accepted practice [4] but it is no more accurate than the original method in Cavanaugh *et al.* and not supported by additional subjective tests.

Gover and Bradley [5] carried out new subjective tests to evaluate and develop measures for objectively assessing speech intelligibility and security. In their subjective tests, four objective speech intelligibility and security results were obtained: (a) the speech intelligibility score, namely the percentage of words correctly identified by each listener; (b) the intelligibility threshold, namely the percentage of listeners able to correctly identify at least one word; (c) the cadence threshold, namely the percentage of listeners able to detect the cadence of the speech; and (d) the audibility threshold, namely the percentage of listeners able to hear the presence of the speech. Their evaluation showed that AI or its more recent replacement the Speech Intelligibility Index (SII) [6] can be related to the intelligibility score when AI≥0 or SII≥0 but cannot be used to describe conditions for high levels of speech security which would correspond to acoustical conditions below AI=0 or SII=0, where AI or SII is not defined. The difference in A-weighted levels is not limited in this way but it is much less accurately related to intelligibility scores. Consequently, they developed measures such as SII-weighted S/N ratio and S/N loudness ratio [5] to more accurately predict the speech intelligibility and security.

The current method for predicting speech intelligibility, privacy and security first requires the development of a measure and then requires the measure to be related to subjective scores using a transfer function [2, 5-6]. To avoid this two-step process, an artificial neural network (ANN) approach has been applied by Xu *et al.* [7] to directly predict the subjective speech intelligibility score and security thresholds using the S/N ratio information. The ANNs were trained and developed on the basis of the subjective test results in Gover and Bradley [6]. Compared with the work of Gover and Bradley [6] that used one-third-octave band S/N ratios, the ANN approach used only the octave-band (250 Hz to 8000 Hz) S/N ratios and performed comparably for the predictions of the intelligibility score, the intelligibility threshold and the cadence threshold, and better for the prediction of the audibility threshold.

The present work uses an application example to show in detail how the ANN models developed in Xu *et al.* [7] can be embedded into Microsoft® Excel [8] spreadsheets and implemented to predict speech intelligibility scores and security thresholds in the design stage. Using the same example, the present work also investigates how the speech intelligibility scores and security thresholds vary as a result of using different constructions for the common partition between the speech sound source room and speech sound receiving room.

AN APPLICATION EXAMPLE

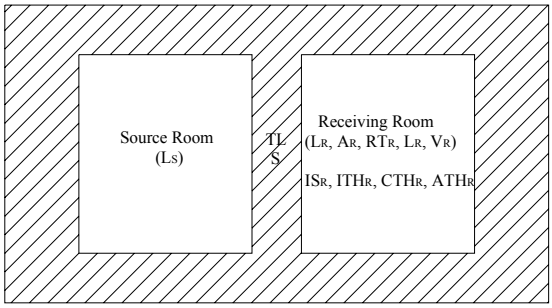

Figure 1 – Transmission of speech sound between two rooms (private offices)

As shown in Figure 1, the present application example is to consider the speech intelligibility and security conditions between two adjacent private offices. L_S is the average speech sound level in the source room; L_R is the average transmitted speech sound level in the receiving room; L_{BG} is the background noise level in the receiving room; A_R is the total sound absorption in the receiving room; RT_R is the reverberation time in the receiving room; V_R is the volume of the receiving room; TL is the transmission loss of the common partition; and S is the area of the common partition. Given the above situation, IS_R, ITH_R, CTH_R and ATH_R are respectively the resulted intelligibility score, intelligibility threshold, cadence threshold and audibility threshold in the receiving room.

Predicting the Transmitted Speech Sound Level

Assuming all the speech sound in the source room passes through the common partition, the transmitted speech sound in the receiving room can be calculated using the following equation [9]:

$$L_R = Ls - TL + 10\log S - 10\log A_R . \tag{1}$$

The above equation ignores the direct sound effect in the rooms and assumes that both rooms are diffusive. In a diffusive room, A_R can be derived from Sabine's reverberation time equation as follows:

$$A_R = \frac{0.161V_R}{RT_R} . \tag{2}$$

Substitute Equation (2) to Equation (1), Equation (1) can be transformed to

$$L_R = Ls - TL + 10\log S - 10\log V_R + 10\log RT_R + 8 . \tag{3}$$

In Gover and Bradley [6], a male speech spectrum was used. This spectrum corresponded to an effort somewhere between "raised" and "loud" [10] and was most easily identified by the subjects in their speech intelligibility tests. Figure 2 shows this speech spectrum in octave bands between 250 Hz and 8000 Hz, which was converted from the one-third-octave band spectrum provided in their work. The converted octave-band speech spectrum, corresponding to a broadband 68 dBA, is used in the present example.

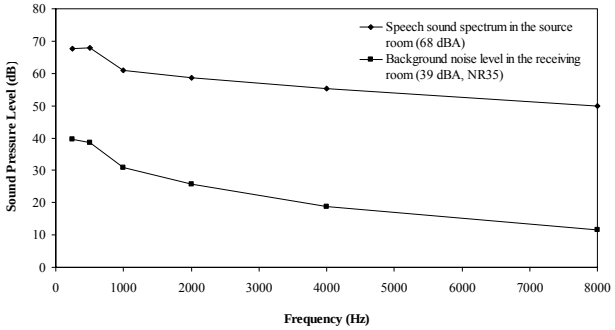

Figure 2 – Speech sound spectrum in the source room and background noise spectrum in the receiving room

Table 1 shows the different constructions of the common partition investigated in the present example. These constructions are typically seen in office buildings. Figure 3 plots their measured TLs [11-12].

Table 1 Different constructions of the common partition investigated

ID	Details of the construction	STC[a]
C1	3 mm float glass	30 [11]
C2	6.38 mm laminated glass	35 [11]
C3	64 mm steel studs, 16 mm plasterboard (12.5 kg/m^2) each side	39 [12]
C4	64 mm steel studs, 16 mm plasterboard (12.5 kg/m^2) each side, 50 mm glasswool (10.8 kg/m^3) in the stud cavity	44 [12]
C5	64 mm steel studs, 2 layers of 13 mm plasterboard (10.5 kg/m^2) each side, 50 mm glasswool (10.8 kg/m^3) in the stud cavity	50 [12]
C6	Decoupled walls: 1 layer of 13 mm plasterboard (10.5 kg/m^2), 64 mm steel studs, 20 mm air gap, 64 mm steel studs, 1 layer of 13 mm plasterboard (10.5 kg/m^2), 50mm glasswool (10.8kg/m^3) in one of the stud cavities, no connections between wall leaves	54 [12]
C7	Decoupled walls: 2 layers of 13 mm plasterboard (10.5 kg/m^2), 64 mm steel studs, 20 mm air gap, 64 mm steel studs, 2 layers of 13 mm plasterboard (10.5 kg/m^2), 50 mm glasswool (10.8kg/m^3) in one of the stud cavity, no connections between wall leaves	60 [12]

[a]STC = sound transmission class

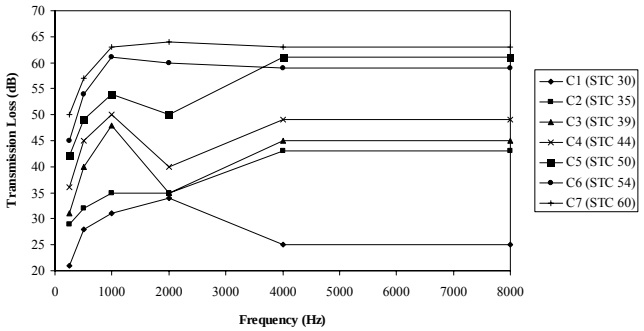

Figure 3 – Measured TLs [11-12] of constructions listed in Table 1.

In the present example, the dimensions of the receiving room are assumed to be 2.7 m (high) × 3.2 m (wide) × 4 m (long). Therefore, $S = 2.7 \times 3.2 = 8.64$ m^2 and $V_R = 2.7 \times 3.2 \times 4 = 34.56$ m^3.

Columns A to G in Table 2 show how the transmitted speech sound level can be calculated using the spreadsheet. The TL shown in Table 2 is of construction C4 and the RT_R shown in Table 2 is of a private office with primarily hard finishes. The use of IF function in Column G of Table 2 is to ensure that the values of transmitted speech sound level in the receiving room is not less than 0.

Calculating S/N Ratios in the Receiving Room

Columns G to I in Table 2 show how S/N ratio in the receiving room can be calculated using the Excel spreadsheet. The background noise level in the receiving room, L_{BG}, is from the mechanical air conditioning system serving the room. The sound spectrum of L_{BG} used in the present example is plotted in Figure 2. The use of IF function in Column I of Table 2 is to ensure that the values of S/N ratios are within the ranges of input variables of ANNs developed by Xu *et al.* [7].

Table 2 Predict the transmitted speech sound level using the Excel spreadsheet

	A	B	C	D	E	F	G	H	I
1	Frequency (Hz)	L_S (dB)	TL (dB)	S (m^2)	V_R (m^3)	RT_R (s)	L_R (dB)	L_{BG} (dB)	S/N Ratios (dB)
2	250	67.6	36	8.64	34.56	0.66	=IF(B2-C2+10*LOG(D2)-10*LOG(E2)+10*LOG(F2)+8<0,0,B2-C2+10*LOG(D2)-10*LOG(E2)+10*LOG(F2)+8)	39.5	=IF(G2-H2<-34.8,-34.8,(IF(G2-H2>6.2,6.2,G2-H2)))
3	500	67.8	45	8.64	34.56	0.60	=IF(B3-C3+10*LOG(D3)-10*LOG(E3)+10*LOG(F3)+8<0,0,B3-C3+10*LOG(D3)-10*LOG(E3)+10*LOG(F3)+8)	38.7	=IF(G3-H3<-30.2,-30.2,(IF(G3-H3>13.1,13.1,G3-H3)))
4	1000	60.9	50	8.64	34.56	0.60	=IF(B4-C4+10*LOG(D4)-10*LOG(E4)+10*LOG(F4)+8<0,0,B4-C4+10*LOG(D4)-10*LOG(E4)+10*LOG(F4)+8)	30.8	=IF(G4-H4<-32.7,-32.7,(IF(G4-H4>1.3,1.3,G4-H4)))
5	2000	58.7	40	8.64	34.56	0.54	=IF(B5-C5+10*LOG(D5)-10*LOG(E5)+10*LOG(F5)+8<0,0,B5-C5+10*LOG(D5)-10*LOG(E5)+10*LOG(F5)+8)	25.6	=IF(G5-H5<-28.2,-28.2,(IF(G5-H5>7.1,7.1,G5-H5)))
6	4000	55.4	49	8.64	34.56	0.42	=IF(B6-C6+10*LOG(D6)-10*LOG(E6)+10*LOG(F6)+8<0,0,B6-C6+10*LOG(D6)-10*LOG(E6)+10*LOG(F6)+8)	18.7	=IF(G6-H6<-27.4,-27.4,(IF(G6-H6>10,10,G6-H6)))
7	8000	49.9	49	8.64	34.56	0.30	=IF(B7-C7+10*LOG(D7)-10*LOG(E7)+10*LOG(F7)+8<0,0,B7-C7+10*LOG(D7)-10*LOG(E7)+10*LOG(F7)+8)	11.6	=IF(G7-H7<-23.2,-23.2,(IF(G7-H7>8.7,8.7,G7-H7)))

Predicting Speech Intelligibility Scores and Security Thresholds

Since the architecture of the ANN models, the range of each input variable and specifics of the ANN models developed by Xu *et al*. have been provided in their work [7], they are not repeated in the present work. Table 3 shows how the specifics of the ANN model, developed by Xu *et al.* [7], for predicting speech intelligibility scores can be embedded into the Excel spreadsheet to make predictions without the use of special ANN softwares. The procedure can be described by the following three steps.

1) Pre-process the input data, which involves the multiplication by the corresponding scale factors, followed by the addition of the corresponding shift factors.
2) Multiply the pre-processed inputs by the corresponding weights to the hidden neurons, sum the weighted inputs and subtract the corresponding bias to the hidden neurons, and pass the resulting value through a non-linear activation function. The non-linear activation function applied in Xu *et al.* [7] was the sigmoidal logistic function ($1 / (1 + e^{-x})$, where x is the resulting value).
3) Multiply the output values of Step 2 by the corresponding weights to the output neuron, sum the weighted outputs and subtract the corresponding bias to the output neuron to produce a single output value (in this case, predicted speech intelligibility scores). In Table 3, the output value of the ANN is rounded to the nearest 5%.

The same procedure can be followed to embed the specifics of the ANN models, developed by Xu *et al.* [7], for predicting speech security thresholds into spreadsheets. Other examples of embedding ANN models into spreadsheets can be found in Nannariello *et al.* [13] and Xu *et al.* [14].

Table 3 Predict speech intelligibility scores using the Excel spreadsheet

Step 1

	A	B	C
8	Input pre-processing scale factor [7]	Input pre-processing shift factor [7]	Pre-processed inputs
9	0.0244	0.8478	=I2*A9+B9
10	0.0231	0.6984	=I3*A10+B10
11	0.0295	0.9629	=I4*A11+B11
12	0.0284	0.7990	=I5*A12+B12
13	0.0268	0.7333	=I6*A13+B13
14	0.0314	0.7276	=I7*A14+B14

Step 2

	A	B	C	D	E
15	Weights to hidden neuron 1 (h1#01) [7]	Multiply weights to pre-processed inputs	Sum weighted inputs	Subtract the bias	Passing the biased summation through the activation function
16	1.1366	=C9*A16	=SUM(B16:B21)	=C16-A23	=1/(1+EXP(-D16))
17	10.5723	=C10*A17			
18	-0.2854	=C11*A18			
19	8.0602	=C12*A19			
20	2.0982	=C13*A20			
21	1.1966	=C14*A21			
22	Bias to h1#01[7]				
23	11.5184				
24	Weights to hidden neuron 2 (h1#02) [7]	Multiply weights to pre-processed inputs	Sum weighted inputs	Subtract the bias	Passing the biased summation through the activation function
25	1.4757	=C9*A25	=SUM(B25:B30)	=C25-A32	=1/(1+EXP(-D25))
26	2.7610	=C10*A26			
27	4.9302	=C11*A27			
28	1.3246	=C12*A28			
29	2.0439	=C13*A29			
30	-1.0482	=C14*A30			
31	Bias to h1#02 [7]				
32	3.4952				
33	Weights to hidden neuron 3 (h1#03) [7]	Multiply weights to pre-processed inputs	Sum weighted inputs	Subtract the bias	Passing the biased summation through the activation function
34	0.7391	=C9*A34	=SUM(B34:B39)	=C34-A41	=1/(1+EXP(-D34))
35	4.5765	=C10*A35			
36	7.2539	=C11*A36			
37	1.8921	=C12*A37			
38	-0.0099	=C13*A38			
39	-2.9519	=C14*A39			
40	Bias to h1#03 [7]				
41	2.4780				

	Step 3				
	A	B	C	D	E
42	Weights to output Neuron [7]	Multiply weights to the resulting values of Step 2	Sum weighted outputs	Subtract the bias	Output (predicted speech intelligibility scores)
43	0.5257	=E16*A43	=B43+B44+B45	=C43-A47	=MROUND(IF(D43<0,0,IF(D43>1,1,D43)),0.05)
44	1.5923	=E25*A44			
45	-1.1333	=E34*A45			
46	Bias to output neuron [7]				
47	0.0086				

Effects of Different Constructions

Table 4 shows the predicted speech intelligibility scores and security thresholds as a result of using different constructions of the common partition listed in Table 1. The predicted IS_Rs are slightly conflicting with the predicted ITH_Rs when the common partition is C5, C6 or C7 because when one of them is 0% the other should also be 0%. This conflict is mainly due to the prediction errors related to the ANN models. In any of the situations, however, both values show very low speech intelligibility will result when the common partition is C5, C6 or C7.

Table 4 Predicted speech intelligibility scores and security thresholds as a result of using different constructions of the common partition listed in Table 1

Constructions	Intelligibility score IS_R (%)	Intelligibility threshold ITH_R (%)	Cadence threshold CTH_R (%)	Audibility threshold ATH_R (%)
C1 (STC 30)	95	100	100	100
C2 (STC 35)	95	100	100	100
C3 (STC 39)	75	95	100	100
C4 (STC 44)	30	70	100	100
C5 (STC 50)	0	10	50	75
C6 (STC 54)	5	0	5	10
C7 (STC 60)	5	0	0	0

DISCUSSIONS AND CONCLUSIONS

The present work used an application example to show in detail how the ANNs for predicting speech intelligibility scores and security thresholds developed in a previous work [7] can be embedded into a spreadsheet application and implemented in the design stage. Using the same example, the effectiveness of different constructions of the common partition for the speech sound isolation between the source room and the receiving room were predicted. The prediction results indicated that 1) when the construction of the common partition had an STC rating of no more than 40, in the speech sound receiving room most of overheard words would be intelligible and consequently the speech privacy between the two offices would be very poor; 2) when the construction of the common partition had an STC rating of around 45, in the speech sound receiving room 30% of overhead words would be intelligible; 3) when the STC rating of the common partition was increased to 50, in the speech sound receiving room only a very small percentage or none of the overheard words would be intelligible; 4) when the STC rating of the common partition was between 50 and 59, in the speech sound receiving room it would still be possible to recognise the cadence or rhythm of the speech; and 5) when the STC rating of the common

partition was over 60, the speech between the two room would be inaudible and an excellent speech security could be achieved.

The present example considered a typical room setup of private offices. The speech sound in the source room had a spectrum equivalent to 68 dB(A) and the background noise level in the receiving room had a spectrum equivalent to 39 dB(A). For different room setups, source room speech sound levels and receiving room background noise levels these partitions may lead to different degrees of speech intelligibility or security.

ACKNOWLEDGMENTS

I am grateful to Dr. John S. Bradley and Dr. Bradford N. Gover for providing me with the speech intelligibility and security test results of their work [5] and the valuable comments on the present work.

REFERENCE

1. W.J. Cavanaugh, W.R. Farrell, P.W. Hirtle, and B.G. Watters, "Speech privacy in buildings," Journal of the Acoustical Society of America, **34**(4), 475-492 (1962).
2. *ANSI S3.5-1969 American National Standard Methods for the Calculation of the Articulation Index*. (American National Standards Institute, New York, NY, 1969).
3. R.W. Young, "Revision of the speech-privacy calculations," Journal of the Acoustical Society of America, **38**(4), 524-530 (1965).
4. *Australian Standard 2822-1985 Acoustics-Methods of assessing and predicting speech privacy and speech intelligibility*. (Standards Association of Australia, North Sydney, Australia, 1985).
5. B.N. Gover and J.S. Bradley, "Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms," Journal of the Acoustical Society of America, **116**(6), 3480-3490 (2004).
6. *ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index*. (American National Standards Institute, New York, NY, 1997).
7. J. Xu, J. Bradley and B.N. Gover, "A neural network approach for predicting architectural speech security," Journal of the Acoustical Society of America, **117**(4), 1709-1712 (2005).
8. *Microsoft® Excel 2002 SP3*. (Microsoft Corporation, North Ryde, Australia, 2005).
9. A.C.C. Warnock and J.D. Quirt, "Noise control in buildings," *In:* C.M. Harris, ed., *Handbook of Acoustical Measurements and Noise Control 3rd edition*. (McGraw-Hill Inc, New York, 33.1-33.41, 1991).
10. W.O. Olsen, "Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson [sic], Bennett & Fidell (1977) report," American Journal of Audiology, **7**, 1-5 (1998).
11. A.C.C. Warnock and J.D. Quirt, "Airborne sound insulation," *In:* C.M. Harris, ed., *Noise Control in Buildings*. (McGraw-Hill Inc, New York, 5.1-5.77, 1994).
12. *CSR Gyprock Fire and Acoustic Design Guide*. (CSR Gyprock, Wetherill Park, NSW, Australia, 2001).
13. J. Nannariello and F.R. Fricke, "A neural-computation method of predicting the early interaural cross-correlation coefficient (IACCE3) for auditoria," Applied Acoustics, **63**(6), 627-641, (2002).
14. J. Xu, J. Nannariello and F.R. Fricke, "Predicting and optimising the airborne sound transmission of floor-ceiling constructions using computational intelligence," Applied Acoustics, **65**(7), 693-704, (2004).