PREDICTING SPEECH SECURITY AND SPEECH PRIVACY: CASUAL VERSUS INTENTIONAL LISTENING

CV Long Her Majesty's Government Communications Centre (HMGCC), Milton Keynes, UK
T Jackson HMGCC, Milton Keynes, UK
SB Shelley HMGCC, Milton Keynes, UK

1 INTRODUCTION

Increasingly in modern life, restrictions and demands on space inside the workplace have led to communal sharing of working space (e.g. 'open-plan' offices), and innovative design and situation of meeting rooms. By their nature, meeting rooms require a certain degree of conversational privacy, not only for the occupants of the meeting room from external noise intrusion, but also for the privacy of the meeting, and to minimise disturbance to people working outside the meeting room. In some cases, for example where there is a risk of corporate espionage, a high level of speech privacy is called for, and in cases where there is the need for minimal conversational intelligibility to those intentionally listening outside the meeting room, this may be referred to as 'speech security'.

While the requirement to rate building components in terms of their acoustic speech privacy performance has been developing since the 1950s, speech security is not yet a widely explored phenomenon in the intelligibility assessment of meeting rooms and similar environments. Caution must be exercised as the terms 'speech security' and 'speech privacy' are occasionally and, in our opinion erroneously, used interchangeably in the literature. It should be made clear that here, we refer to speech privacy as relating to casual listeners (i.e. those not intending to overhear what is being said), and speech security as relating to intentional listeners (i.e. those deliberately trying to extract intelligible words from the conversation).

Both phenomena have in common that a certain proportion of words spoken are intelligible to the casual or intentional listener. Where they begin to differ lies in the interpretation of how this information is received and processed by the listener. In the case of speech privacy, the threshold is typically based either on the level of annoyance to casual listeners (e.g. in an open plan office environment), or the sensitivity of the conversation being overheard (e.g. in a hospital waiting room environment). In the case of speech security, the threshold is based on the level of 'acceptable loss of information' to intentional listeners, or the percentage of words in the conversation that are intelligible such that the listener can gain information about the nature of that conversation (e.g. in an embassy meeting room environment). It is reasonable to assume that the acceptable threshold for speech security is lower than that for speech privacy. Ideally, the threshold of speech 'leakage' for complete speech security would be 0% intelligible words, but in reality this threshold is impossible to reliably estimate, due to the fact that prediction methods are based on average human response (i.e. it does not account for persons with better than average hearing or processing ability). The speech security threshold therefore has a degree of plasticity according to the nature of the situation and potential consequences of information 'leakage' to persons intentionally listening. Such persons may be listening 'naturally', or listening with the aid of microphones and other electronic devices[1] . This causes additional complications not inherent to speech privacy, as it allows intentional listeners the opportunity to adjust listening levels, process and enhance audio, and repeatedly listen to segments of conversation.

These differences mean that although speech privacy measures are sometimes used to rate the speech security of a structure, this may not be appropriate, or even possible in some cases. Other speech security researchers have also identified the lack of applicability of speech privacy-based measures to speech security prediction[2] . It should be recognised that small errors in the prediction

of speech privacy can be tolerated, while the risk associated with small errors in speech security prediction may be much greater.

This paper aims to highlight the unique problems posed by the assessment of speech security, as compared to those of speech privacy rating.

2 EXISTING METHODS FOR THE ASSESSMENT OF SPEECH PRIVACY AND SPEECH SECURITY

It is helpful to think of a meeting room as an enclosure with an external 'skin' that acts as a channel for speech sounds, passing from inside to outside, and vice versa. For any given communications system, it is important to be able to assess the quality of the communication channel, to determine if speech delivered at the source location (in this case inside the meeting room) by a person remains intelligible to a person at the receiver location (outside the room). To date, few methods have been developed to assess speech security as a separate entity from speech privacy.

Since both speech privacy and speech security are synonymous with the intelligibility of the transmitted conversation, the most appropriate methods for assessment are those used to predict and rate speech intelligibility[3][4][5]. The three key factors which must be accounted for with any method should include a) the vocal effort of the talkers; b) masking/background noise; and c) room containment/attenuation effects on the transmitted speech. Methods based purely on the sound attenuation characteristics of a meeting room's partitions have been found to be inappropriate for these purposes[3][6], as they do not account for the specific frequency range of speech acoustic energy, may be affected by flanking transmission *in situ*, do not account for background noise variation, assume all measurements are taken in the diffuse field, 'average' measurements across an entire surface, and are independent of conversational vocal effort levels.

2.1 Articulation Index

An early method devised to provide an estimate of speech privacy in buildings was the Articulation Index (AI) measure developed by Cavanaugh and colleagues in 1962[4]. This method scored listeners' subjective impressions of speech privacy against AI scores, ranging between 0 (unintelligible) to 1 (fully intelligible). Here, adequate privacy was defined as the listener's 'freedom from distraction', but a second category of 'confidential privacy' was used for 'assurance of not being overheard', bearing similarities to the requirements of speech security. While the threshold for 'annoyance privacy' was set at AI 0.1, it was proposed that an AI rating equal to or below 0.05 was the threshold for 'confidential speech privacy' ("some isolated words intelligible"). Listening tests by Bradley (2009)[7] showed that AI 0.05 corresponded to an intelligibility score of 39% (39 words in 100 correctly identified by the average listener). This was deemed to be too high to be an appropriate threshold for speech security ratings. The AI threshold of intelligibility was therefore re-evaluated and determined to be AI 0.027[7]. However, this is problematic for speech security assessment, since it is difficult to accurately measure such low AI ratings[5] . Nevertheless, variants of the AI measure (such as those implemented by Young (1965)[8]) are still used in modern speech privacy assessments, such as for open plan office spaces.

2.2 Privacy Factor

In 1996, the UK Department of Health published acoustic design criteria for hospitals, which included the concept of a 'Privacy Factor' (PF) for sound insulation assessment (HTM 2045[9]). This was based on both the *in situ* weighted sound reduction index of the room (R_w), as measured according to sound insulation Standard ISO 15186-2[10], and the background noise level (as defined by the Noise Rating curve according to ISO 717-1[11]). The method accounts for the vocal effort of talkers inside the assessed room by applying a correction factor, ranging from 5 dB for 'raised' vocal effort, to 20 dB for 'shout'[9] (see Figure 2). Resulting Privacy Factors are provided in Table 1.

Privacy Factor (dB)	Resulting Privacy
<70	“Clearly audible and intelligible”
70 – 75	“Audible but not intrusive”
75 – 80	“Audible but not intelligible”
>80	“Inaudible”

Table 1- Privacy Factor ratings assuming 'normal' speech (i.e. no vocal effort correction), according to HTM 2045[9].

The PF ratings are subjective in nature, with a PF > 75 dB recommended for 'confidential conversations with patients'.

It should be noted that HTM 2045 has since been withdrawn and replaced by HTM 08-01[12], which has abolished PF ratings and replaced it with an alternative method. This is based on the laboratory-measured R_w value and calculated using the *in situ* standard level difference between rooms, $D_{nT,w}$, measured according to ISO 717-1[11]. In this version, all ratings assume 'raised voice' conversational vocal efforts, and it attempts to account for background noise levels by categorising a room's “sensitivity to noise”. Room privacy requirements are selected according to four main criteria; 'Not private', 'Moderate', 'Private' (normal speech audible but not intelligible), and 'Confidential' (raised speech audible but not intelligible). An example is provided of a room requiring 'Confidential privacy', with a 'high' noise sensitivity, as needing a $D_{nT,w}$ rating of at least 52 dB. Nevertheless, PF is still currently used in some privacy assessments.

2.3 Speech Privacy Class

One method aimed to address both the speech privacy and speech security assessment of closed rooms is the Speech Privacy Class (SPC)[13], developed by Bradley, Gover and colleagues at the National Research Council of Canada. Using an incremental scale, speech is rated according to both its intelligibility and its audibility after passing through the wall of the room, thus attempting to account for frequency-dependent filtering effects (using the same method as Transmission Loss (TL) measurements[14]). Based on the results of listening tests conducted during its development, the SPC uses a Signal-to-Noise Ratio (SNR) measure to predict the resulting intelligibility of speech transmitted through a system. As this method utilises the SNR of the speech as measured *in situ,* local background noise is also included as a factor (however, Sato *et al*. (2012)[15] note that SPC-scored intelligibility may still be underestimated in some cases). The SNR is frequency-averaged over the speech spectrum (in this case taken to be 160 – 5000 Hz in third-octave bands) and uniform-weighted to give the $SNR_{uni\ 32}$ measure[6][14][16]. The SPC is then calculated as the sum of the average ambient noise level and the average TL of the structure[14]. As a result, SPC 75 is the threshold of speech privacy (“speech occasionally intelligible (1 word per 15 minutes) and audible”), and SPC 80 the threshold of speech security (“speech very rarely intelligible (4 words per 8 hours), occasionally audible”), as defined by listening tests[16] where intelligibility threshold was set at 50% of listeners able to correctly identify a single word[6][17]. SPC has been converted into an International Standard (ASTM E2638-10[18]), however Bradley stresses that the method is not designed to be applied directly to speech security rating where intentional listening is used[13]. This limitation is due to the fact that the TL of each component of a room structure *in situ* can be highly variable, while listening tests were based on average TL curve performance. In addition, Intentional listeners may seek out the weakest point of a structure, which may have significantly less ‘attenuation’ or an atypical TL response, while SPC is based on the spatially averaged performance of the closed space as a whole[17][19]. One further and important factor is that the measure is relevant only while speech within the meeting room is at a pre-estimated ‘average level’[16]. If the vocal effort level of persons within the room is higher than that average (e.g. due to the Lombard effect), this would compromise the applicability of the rating for the room. Talker location and room size/reverberation may also modify the SPC[1][17].

Gover & Bradley (2011)[19] compared the SPC measure to the Articulation Index (AI) method defined in ASTM E1130-08[20], finding both to have a high correlation with listening test intelligibility scores. However, due to the low degree of variability of the AI measure at values approaching 0, this method was not able to provide sufficient accuracy under high speech privacy/security conditions where intelligibility percentage is very low. Similar problems exist for other AI-based measures, such as Speech Transmission Index (STI) and Speech Intelligibility Index (SII)[5] (for example, the threshold of intelligibility for the SII method was found to be SII 0.055[7]). Other researchers have attempted to overcome this problem through the use of AI variants such as the Rapid Speech Transmission Index (RASTI) for speech security assessment, but there is still work to be done to tailor these methods[21].

2.4 Maximum Safe Vocal Effort

A method to rate the speech security of meeting rooms was developed by the UK National Authority for Counter-Eavesdropping (UKNACE), in collaboration with Liverpool University[22]. This Maximum Safe Vocal Effort (MSVE) method utilises the STI intelligibility rating criteria, which combines both SNR and room acoustic effects. By taking STI measurements at multiple locations outside meeting rooms, it is possible to obtain a threshold, or the point at which an 'occasional word' of transmitted speech can be understood, by relating STI score to the Phonetically Balanced (PB) word score as per Anderson & Kalb (1987)[23]. This may be taken as the point where 40% of the PB words are scored correctly, which corresponds to an STI score of 0.3 (Figure 1), the lowest value of speech 'leakage' that can be reliably measured.

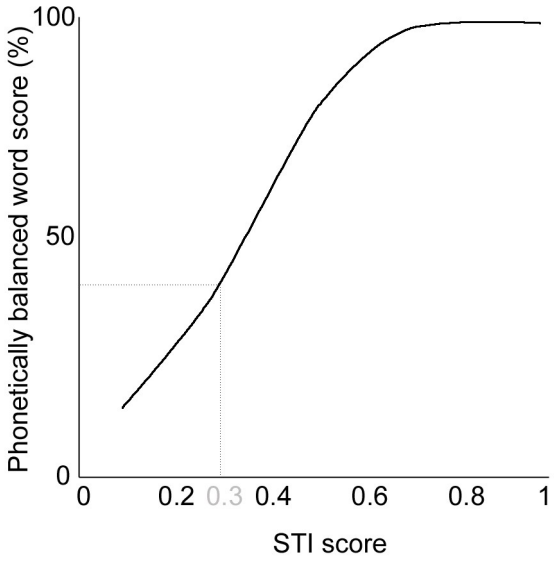

Figure 1- Demonstrating the relationship between Phonetically Balanced (PB) word score and Speech Transmission Index (STI). Dashed line indicates the 0.3 STI point, at which 40% of PB words, on average, are scored correctly by listeners.
Image adapted from Anderson & Kalb (1987)[23].

It has been proven that it is possible to offset the 0.3 STI point further still[22], down to levels where measurement becomes impossible (due to background noise or high levels of acoustic attenuation), which eliminates the problem of low-level measurements identified by Gover & Bradley (2011)[19]. Because intelligibility is highly reliant on the SNR, then speech security will be strongly dependent on background noise levels (both within and outside the meeting room), and on the vocal effort of the talker[24]. Speech security can therefore be related directly to the threshold/offset STI score, in accordance with vocal effort and background noise. As a result, the Maximum Safe Vocal Effort Level (MSVEL) is the level which talkers within the assessed meeting room should not exceed if the

conversation is to remain secure at the required STI offset (see Figure 2). The offset itself may be set according the perceived risk of information 'leakage' to an intentional listener outside the meeting room.

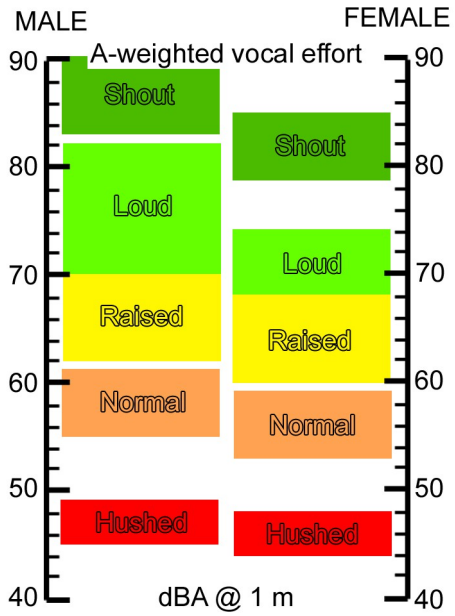

Figure 2- Vocal effort levels (as mean L_{Aeq}) for male and female typical speech spectra at a distance of 1 m on-axis (A-weighted). Vocal effort regions colour coded according to the Maximum Safe Vocal Effort rating of an assessed meeting room.
Image adapted from Cushing *et al.* (2011)[25].

This method has the benefit of being suitable for single 'weakest point' measurements of a meeting room, the ability to account for background noise outside the room, and to account for variable vocal effort levels of conversation within the meeting room. It may therefore be tailored for both male and female speech (although male speech has been found to represent the 'worst case' for speech security measurements[26]). However, unlike the SPC measure, MSVE is unable to account for the non-linear frequency transfer 'filtering effect' of speech transmitted through a meeting room wall, although recent research has identified additional STI weighting factors to reduce error associated with this phenomenon[26].

3 PROBLEMS TO BE ADDRESSED

Although there are existing methods that attempt to predict or rate speech privacy and speech security of closed rooms, there are no clear comparisons that can be made between them. Other authors have stated the need to unify thresholds for speech privacy and speech security across the various methods used, including SPC, STI and SII[7]. Additional complications arise due to the differences between the construction requirements of room walls (measured as attenuation) and the measures assessing speech privacy/security. At present, there is no way to definitively compare wall attenuation or insulation (R_w and $D_{nT,w}$) with speech rating measures, making it difficult to specify constructional requirements during the room design process.

There are also problems surrounding the distinction between speech security and speech privacy, and it is clear that methods used for the rating of one may not be appropriate for both. Much further work in the area of speech security in particular is needed, as this is often overlooked in the literature. There is a requirement for a new international standard to exclusively rate the speech security of closed rooms, particularly where a high degree of talker confidentiality is required.

A problem unique to speech security at very low thresholds exists, in that intelligibility scores are developed based on the average response of a sample to listening tests. This poses a problem of underestimating speech security in situations where an intentional listener has better than average hearing, or is very well matched to the talkers. Bradley & Gover suggest that choosing an intelligibility threshold below 50% of listeners correctly identifying a single word in listening tests would enhance speech security predictions for more sensitive or better matched listeners[17].

It should be noted that while the MSVE measure can be applied to rate speech security down to a threshold offset, it currently is not able to address situations where speech is unintelligible, but still audible to the listener. To prevent a listener from determining even speech cadence would require a substantial improvement in meeting room sound insulation than would be adequate for speech intelligibility security[27]. Whether speech security is defined as an 'absence of conversational information extraction to the intentional listener', or as 'an absence of speech sounds to the intentional listener' is subject to individual circumstantial requirements.

4 CONCLUSIONS

There are unique problems posed by the assessment of speech security that are not inherent to speech privacy rating, and therefore warrant speech security to be treated as its own entity. These problems include assessment under very low signal-to-noise ratio conditions, variation in background noise outside meeting rooms and vocal effort fluctuations inside meeting rooms, intentional listening at room weak points, and effects of room reverberation close to boundaries.

Much further work is needed in this area to explore the differences between speech privacy and speech security, in particular to define a method unique to the assessment of speech security.

REFERENCES

1. Gover, B.N. and Bradley, J.S. *Guide for Assessment of the Architectural Speech Privacy and Speech Security of Closed Rooms*. (2010) National Research Council of Canada, IRC-RR-276, p. 19.
2. Xu, J., Bradley, J.S. and Gover, B.N. (2005) 'An artificial neural network approach for predicting architectural speech security.' *J Acoust Soc Am*., 117: 1709.
3. Park, H.K., Bradley, J.S. and Gover, B.N. (2007) 'Rating sound insulation in terms of speech intelligibility.' *Proceedings of the 19th International Congress on Acoustics*, : 1-6.
4. Cavanaugh, W.J., Farrell, W.R., Hirtle, P.W. and Watters, B.G. (1962) 'Speech privacy in buildings.' *J Acoust Soc Am*., 34: 475-492.
5. Gover, B.N. and Bradley, J.S. (2004) 'Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms.' *J Acoust Soc Am*., 116: 3480-3490.
6. Bradley, J.S. and Gover, B.N. (2008) 'A new procedure for assessing the speech security of meeting rooms.' *Proceedings of the Institute of Acoustics*, 20: 1-6.
7. Bradley, J.S. (2009) 'Comparisons of speech privacy measures.' *Inter-Noise 2009*, : 1-9.
8. Young, R.W. (1965) 'Re-vision of the speech privacy calculation.' *J Acoust Soc Am*., 38: 524-533.
9. HTM 2045: *Acoustics* (1996).
10. BS EN ISO 15186-2: *Acoustics- Measurement of Sound Insulation in Buildings and of Building Elements Using Sound Intensity- Part 2: Field Measurements* (2003).
11. BS EN ISO 717-1: *Acoustics- Rating of Sound Insulation in Buildings and of Building Elements- Part 1: Airborne Sound Insulation* (2013).
12. HTM 08-01: *Acoustics* (2013).
13. Bradley, J.S., *Minimizing Speech Security Risks for Meeting Rooms*. (2009)[www document] <http://www.nrc-cnrc.gc.ca/eng/achievements/highlights/2009/meeting_security.html> (accessed 22 July, 2014).

14. Bradley, J.S. and Gover, B.N. *Selecting Walls for Speech Privacy*. (2011) National Research Council of Canada, IRC-RR-314, p. 28.
15. Sato, H., Morimoto, M., Hoshino, Y. and Odagawa, Y. (2012) 'Relationship between sound insulation performance of walls and word intelligibility scores.' *Applied Acoustics*, 73: 43-49.
16. Bradley, J.S. and Gover, B.N. *A New System of Speech Privacy Criteria in Terms of Speech Privacy Class (SPC) Values*. (2011) National Research Council of Canada, NRCC-54472, p. 8.
17. Bradley, J.S. and Gover, B.N. (2010) 'Speech levels in meeting rooms and the probability of speech privacy problems.' *J Acoust Soc Am*., 127: 815-822.
18. ASTM E2638-10: *Standard Test Method for Objective Measurement of the Speech Privacy Provided by a Closed Room* (2010).
19. Gover, B.N. and Bradley, J.S. (2011) 'ASTM metrics for rating speech privacy of closed rooms and open plan spaces.' *Canadian Acoustics*, 39: 50-51.
20. ASTM E1130-08: *Standard Test Method for Objective Measurement of Speech Privacy in Open Plan Spaces Using Articulation Index* (2008).
21. Hojbjerg, K. (1987) 'RASTI for speech isolation evaluation.' *J Acoust Soc Am*., 82: S46.
22. Robinson, M., Hopkins, C., Worrall, K. and Jackson, T. (2014) 'Thresholds of information leakage for speech security outside meeting rooms.' *J Acoust Soc Am*., [Article in press]: .
23. Anderson, B.W. and Kalb, J.T. (1987) 'English verification of the STI method for estimating speech intelligibility of a communications channel.' *J Acoust Soc Am*., 81: 1982-1985.
24. Pollack, I. (1958) 'Speech intelligibility at high noise levels: Effect of short term exposure.' J *Acoust Soc Am*., 30: 282-285.
25. Cushing, I.R., Li, F.F., Cox, T.J., Worrall, K. and Jackson, T. (2011) 'Vocal effort levels in anechoic conditions.' *Applied Acoustics*, 72: 695-701.
26. Long, C.V. (2014) 'Non-linear frequency contributions and gender-specific effects on rating the speech security of meeting rooms.' [Article in preparation], : .
27. Bradley, J.S. and Gover, B.N. (2003) 'Developing a new measure for architectural speech security.' *Canadian Acoustics*, 31: 50-51.