



Development of a flexible, realistic hearing in noise test environment (R-HINT-E)

Laurel Trainor^{a,*}, Ranil Sonnadara^a, Karl Wiklund^b, Jeff Bondy^b, Shilpy Gupta^a,
Suzanna Becker^a, Ian C. Bruce^b, Simon Haykin^b

^aDepartment of Psychology, McMaster University, 1280 Main West Street, Hamilton, ON, Canada L8S 4K1

^bDepartment of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada

Received 15 March 2003; received in revised form 2 July 2003

Abstract

Through the use of DSP chips and multiple microphones, hearing aids now offer the possibility of performing signal-to-noise enhancement. Evaluating different algorithms before they are instantiated on a hearing aid is essential. However, commercially available tests of hearing in noise do not allow for speech perception evaluation with a variety of signals, noise types, signal and noise locations, and reverberation. Here we present a flexible realistic hearing in noise testing environment (R-HINT-E) that involves (1) measuring the impulse responses at microphones placed in the ears of a human head and torso model (KEMAR) from many different locations in real rooms of various dimensions and with various reverberation characteristics, (2) creating a corpus of sentences based on the hearing in noise test recorded in quiet from a variety of talkers, (3) creating “soundscapes” representing the input to the ears (or array of microphones in a hearing aid) by convolving specific sentences or noises with the impulse responses for specific locations in a room, and (4) using psychophysical procedures for measuring reception thresholds for speech under a variety of noise conditions. Preliminary evaluation based on the engineering signal-to-error ratio and on human perceptual tests indicates that the convolved sounds closely match real recordings from the same location in the room. R-HINT-E should be invaluable for the evaluation of hearing aid algorithms, as well as more general signal separation algorithms such as independent components analysis.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Sound; Impulse response; Hearing in noise; Speech perception

1. Introduction

The most common complaint of hearing aid users is not the degree of amplification they receive, but that this amplification does not help them to understand speech in noisy or reverberant environments [19]. The

ability of the normal mammalian auditory system to perform auditory source separation (or auditory scene analysis) is truly amazing and not yet entirely understood [5]. Objects in a real-world environment emit sound waves that radiate out in all directions and are composed of multiple frequencies varying across time. The sound wave arriving at the ear is composed of the sum of sound emissions of many such objects and their echoes as the waves bounce off walls and other

* Corresponding author. Tel.: +1-905-525-9140.
E-mail address: ljt@mcmaster.ca (L. Trainor).

objects. The normal auditory system automatically decomposes the incoming signal into groups of components representing the original sound sources. With hearing loss, however, this decomposition ability is often severely limited.

With the advent of digital hearing aids and the use of multiple microphones, it is possible, in principle, to enhance the signal-to-noise ratio (SNR) before the sound reaches a damaged ear, and a number of groups are developing algorithms to do this (e.g., [2,4,9,18,20]). However, our ability to adequately test the utility of these algorithms is falling behind our ability to develop them. In order for an algorithm to be considered successful, it must significantly enhance speech perception in real world situations. The present paper describes the development of a flexible realistic hearing-in-noise testing environment (R-HINT-E).

Historically, a number of measures of speech perception thresholds have been used to measure human performance. The speech reception threshold (SRT) is the sound pressure level at which listeners can accurately understand 50% of the words presented in quiet (ANSI S3.6-1969). The reception threshold for sentences (RTS) is the relative sound pressure level at which listeners can accurately understand 50% of sentences presented in noise [24]. On the engineering side, SNR is the most common metric for measuring how well a noise reduction algorithm is working. The SNR metric has been extended to correspond more closely with signal quality, including dealing with relative distortion due to interference, noise, or artifacts [15] and quantifying perceptual sound quality [10]. While all of these measures have some predictive power, they all fall short of predicting speech understanding across the variety of noise and reverberation conditions found in the real world.

There are a number of reasons for this. First, SNR does not take into account the inherent redundancy of speech signals in both the time and frequency domains and at levels ranging from phonotactic to syntactic to semantic. Thus, under some circumstances, human listeners can perform at a level better than that predicted by SNR. Although better engineering metrics are being developed based on models of the auditory periphery [7], there is still no substitute for testing actual performance in human subjects. Sec-

ond, the amount of reverberation in an environment also affects the RTS, although SNR metrics do not describe how well signal separation algorithms work in reverberation [28]. Third, thresholds in quiet do not entirely predict thresholds in noise, so the SRT measure is not adequate for predicting the performance of noise-reduction algorithms in the real world. Fourth, the RTS varies significantly depending on the type of noise used. For example, noise signals with spectral characteristics similar to those of speech targets can more effectively mask the target signal [22]. Thus, different results are obtained when the noise is white noise, speech-spectrum noise or competing speech. Further, recent studies indicate that with competing speech, the informational similarity of the target and interference speech affects thresholds independent of the SNR. For example, RTS is best when the target and interference talkers are different genders and worst when they are the same talker [8], and the difference between these conditions gets larger when the target and interference speakers are spatially separated [1]. In sum, there are many factors that affect the RTS and therefore a flexible testing environment is needed that can measure the RTS under various interference conditions.

Typical sensorineural hearing losses (e.g., presbycusis or losses due to noise exposure) involve hair cell damage. Outer hair cell damage results in more broadly tuned frequency channels. Deficits in temporal resolution may also be present. Thus, hearing impaired listeners are more impaired than normal listeners at perceiving speech in noise compared with perceiving speech in quiet [21]. RTSs measured in temporally unshaped long-term average speech spectrum (LTASS) noise range from about 2.5 dB worse than normal listeners for those with mild hearing losses to 7 dB worse for those with severe hearing losses [25]. This reduction in speech perception is already substantial, but it underestimates real-world performance for two reasons [21]. First, normal listeners are able to take advantage of temporal and spectral “dips” in temporally modulated noise, which are more typical of natural environments, to gain an advantage for speech perception of about 12 dB compared with performance in LTASS noise. However, those with sensorineural hearing loss are largely unable to do so (e.g., [11,16]). Second, the hearing impaired are less able than normals to take advantage of the

spatial separation of sound sources, largely because of high-frequency losses, the difference between those with mild-to-moderate hearing losses and those with normal hearing varying between 2.5 and 7 dB [6]. The sum of the inability to take advantage of temporal and spectral dips and the spatial separation of signal and source agrees well with empirical data showing that the difference between normal and impaired listeners in RTS with binaural, modulated noise is about 16 dB [12].

Two of the most popular commercially available speech in noise tests are the speech in noise (SPIN, [17]) test and the hearing in noise test (HINT, [23]), although other corpuses are available (e.g., [3]). The SPIN test requires listeners to repeat the last word of sentences spoken by a single male talker. Different levels of semantic redundancy are provided across the sentence lists. The HINT requires listeners to repeat back entire sentences, all spoken by a single male talker. The test is divided into lists of phonemically balanced sentences. For the purposes of algorithm testing, neither SPIN nor HINT adequately examine effects of different background noises (each has only one background noise: multi-speaker babble for SPIN; LTASS noise for HINT), variation in the speech signal critical for studying informational masking (each uses only one talker), reverberation and room size (each uses only a low-reverberation condition), the number of interfering signals (each has only one interference signal), or the location of target and interfering signals (each uses only one configuration with the signal in front of the listener and the noise 90° to the side). Different noise-enhancing algorithms perform differently under various noise conditions. For example, standard test benches have been developed for ICA (e.g., [26]) and for beamforming (e.g., [14]), but comparisons of ICA and beamforming are difficult because a truly fair comparison must involve a variety of noise conditions. Thus, a general, flexible hearing in noise test environment would be very useful.

Both SPIN and HINT have a further limitation for the testing of hearing aid algorithms. It is desirable to test an algorithm before it is instantiated in a DSP chip in a hearing aid, as this is a costly process. A hearing aid contains one or more microphones that record the sound input at the entrance to the ear, pass this to a DSP chip that performs the desired compress-

sion and signal-to-noise enhancement, the output of which is then played to the ear of the wearer. During algorithm development, the DSP algorithm will be implemented in software on a computer rather than on the chip in a hearing aid. Thus, in order to measure speech reception thresholds with a particular algorithm, it is necessary to first record the incoming sound wave consisting of the signal and noise and their reverberations at the location of the ear. This input then needs to be passed to the computer and processed according to the hearing loss of the individual and the signal-enhancing algorithm being tested, and the output of this processing then needs to be presented to the ear of the listener. Practicalities dictate that this is most easily done if the sound inputs are recorded ahead of time using a realistic human model, such as KEMAR, and if the signal processing is done off line, and the listeners are presented with the test sounds through headphones.

In creating our flexible R-HINT-E, we have the following goals:

1. To use realistic noise and reverberation in order to best estimate real-world performance.
2. To be able to test in a variety of environments with different reverberation characteristics.
3. To maintain experimental control, such that each listening environment is reproducible and characterized in detail. This is necessary in order to compare RTS across different signal-enhancing algorithms.
4. To use a large database of speech signals so that repeated psychophysical testing can be done. People will better understand a sentence if they have heard it before. Thus, new sentences must be used on every trial that goes into measuring a RTS, as well as across measurements of RTSs for different algorithms.
5. To have a large variety of speech sounds available including different male, female, and child voices in order to mimic challenges that will be faced in the real world.
6. To be flexible, so that the experimenter can specify a variety of sound files to be played simultaneously, each at their own intensity level and from a variety of locations in a virtual room with specified reverberation characteristics. This is necessary because the number, locations, and kinds of signal

and noise sources desired will change depending on the nature of the signal-enhancing algorithm being tested.

7. To use earphone presentation because off-line processing must be done in order to incorporate the signal-enhancing algorithm into the circuit, as discussed above.
8. To be expandable so that new databases of speech signals and new virtual rooms with particular acoustic characteristics can be readily added.

1.1. Conceptual overview of the R-HINT-E

Given that we want maximum flexibility, realistic acoustic conditions, and headphone presentation, we have developed a hybrid naturalistic-virtual system, in which impulse responses of real rooms are measured and convolved with separately recorded speech signals. There are four aspects to the system:

1. *The sound stimuli*: Speech sentences or other target sounds and noises are recorded in quiet, non-reverberant conditions. For the purposes of this paper, we have recorded the HINT sentences spoken by a variety of talkers (see Section 2). However, the system is flexible, and other quiet non-reverberant sound recordings could easily be incorporated.
2. *Characterizing room environments by impulse measures*: In order to know how sounds delivered from a particular location in a particular room will be transformed when they reach a person's ears, an impulse sound must be delivered from a speaker at that location, and the resulting sound waves reaching the ears must be recorded. The impulse responses of those speaker and recording locations can then be calculated and the speech signal or other sounds of interest can be convolved with the impulse responses in order to approximate the sound wave that would actually reach the ears had that sound actually been presented from that location. For the purposes of this paper, we recorded impulse responses with microphones located in the two ears of a human KEMAR head and torso model, which was positioned near the centre of the room. The impulse responses must be

recorded from every location where it might be desired to locate a sound source. For the purposes of the present paper, we recorded from 48 locations in each of three rooms at various angles, heights, and distances around KEMAR. Again, however, the system is flexible, and additional impulse responses measured in various environments can be added at will.

3. *Combining multiple sounds and locations*: Using MATLAB code, virtual "soundscapes" are created by the user in the following manner. First, a particular room is chosen for which impulse measures are available. Then the locations of the desired sound sources are chosen from the list where impulse responses have been measured. Then, for each location, a sound file is chosen to present from that location. Each sound file is then convolved with the impulse response for its location, and the resulting waveforms are added together to create the soundscape that will reach the person's ears. Note that this can be done separately for each of the microphone locations in order to make use of spatial information.
4. *Signal enhancing algorithm incorporation*: If the purpose is to test a particular signal-enhancing algorithm, the soundscape waveforms created in Step 3 are passed through the algorithm. Note that for subjects with a hearing loss, the particular amplification and compression algorithm of their hearing aid can also be mimicked in software at this stage.
5. *Measuring speech reception thresholds*: For a given set of locations and a given speech-enhancing algorithm, multiple trials, each with a different speech sentence, can be presented to a listener through headphones. An adaptive psychophysical procedure (e.g., [27]) can be applied at this stage. RTSs can be determined for different noise-enhancing algorithms, for different types and locations of target signals, and for different types and locations of competing signals.

2. Methods

In this section, we first describe the corpus of recorded sentences, then the details of how the impulse responses for three rooms were measured, and

finally the MATLAB program in which soundscapes are calculated and experiments presented.

2.1. Sentence recordings

The 250 sentences from the HINT are divided into 25 lists of 10 sentences each, and each of the lists are equivalently phonemically balanced. Thus they provide a good basis for a hearing in noise test. We recorded the sentences spoken by 6 males and 6 females for a total of 3000 sentences in order to better approximate the range of talkers encountered in the real world. The recordings were done with a Neumann KM131 microphone, a Dell4100 computer, and Protools Digi001 audio interface and software in a sound attenuating booth. The corpus was normalized to Fletcher and Munson's [13] iso-loudness contours at 65 dB SPL.

2.2. Impulse response measurement

Impulse responses were measured in 3 rooms. Room 1 (low reverberation) was $11' \times 11' \times 8' 6''$ high with a double row of velour drapes closed around its periphery. Room 2 was the same as Room 1, but had the drapes open for dimensions of $12' \times 12'$. Room 3 was a somewhat reverberant classroom at McMaster University, with dimensions $17' 10'' \times 32' 9'' \times 8' 8''$.

Recordings were made with Knowles FG microphones in each ear of the KEMAR human head and torso model. Microphones were connected to a custom-built preamplifier circuit board located in KEMAR's chest cavity, and from there to an Echo Layla sound card and Dell Precision M50 laptop computer running Cool Edit Pro software.

KEMAR was located in the centre of the room with the microphones 55'' above the floor. A single speaker was moved to 48 locations around KEMAR: 8 angles, starting in front of KEMAR and moving clockwise ($0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$) by 3 heights (7'', 33'' and 65'' from the floor to the center of the speaker driver) by 2 distances (3', 6') from KEMAR (Fig. 1). The impulse sound was a chirp consisting of an exponential sweep from 0 to 22,050 Hz over a period of 1486 ms (Fig. 2), and was generated with CRC-MARS software. It was presented from the Dell laptop computer connected to the Echo Layla

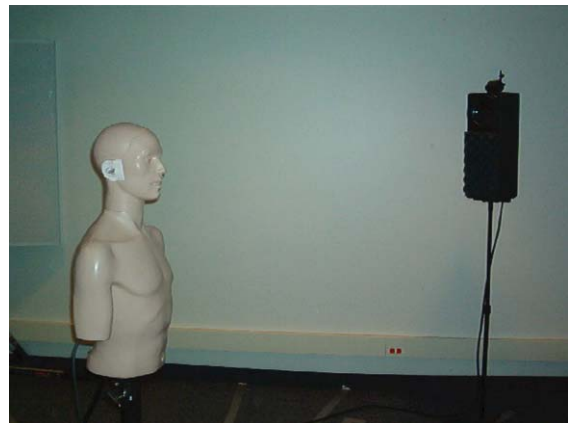
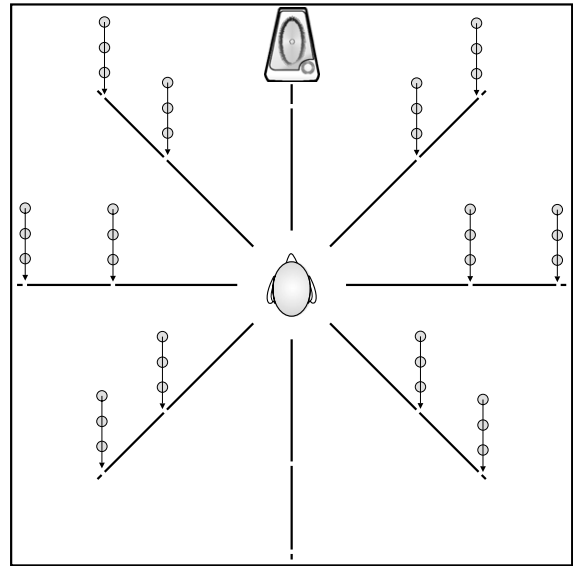


Fig. 1. The impulse measurement setup. The upper panel shows a schematic of KEMAR in the centre of the room and the locations of the speaker varying around 8 angles, 3 heights and 2 distances from KEMAR. The impulse measurements from each of the 48 locations were recorded simultaneously by microphones in the ears of KEMAR (upper right panel). The bottom panel shows KEMAR in relation to a high distant speaker location.

audio interface card (via a PCMCIA interface card) connected in turn to a Hafler P1000 amplifier and a flat-response speaker.

For each of the 48 locations in each room, the impulse response was recorded simultaneously at the microphones (see Fig. 3). For verification purposes,

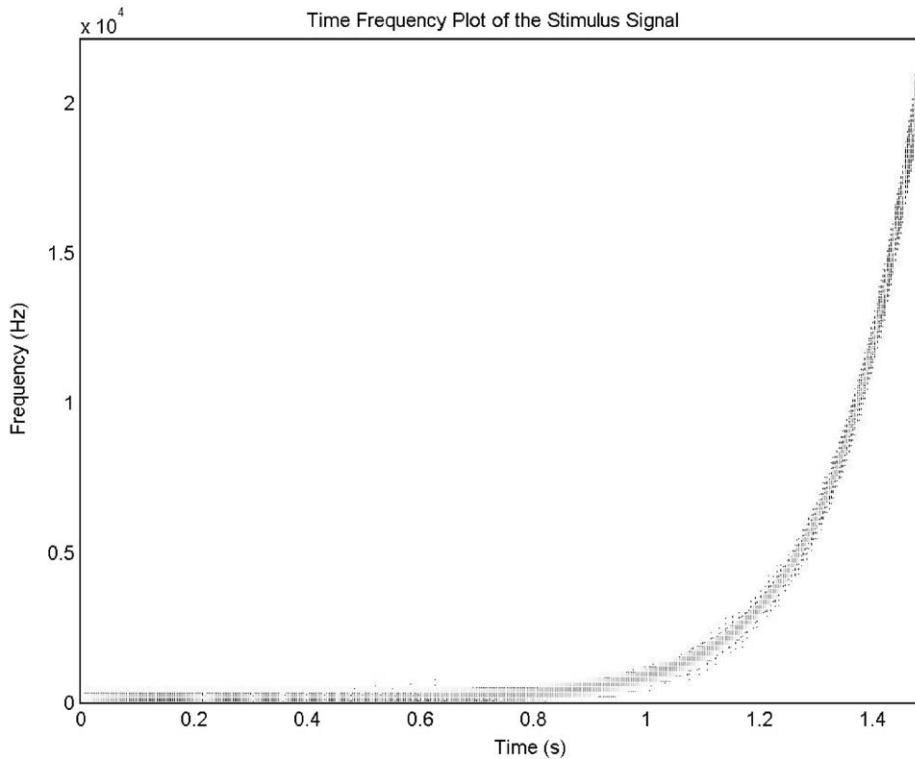


Fig. 2. A time–frequency graph of the impulse chirp stimulus.

each of 6 sentences (4 male and 2 female talkers) from the quiet non-reverberant recorded HINT sentences (see above) were also recorded simultaneously by the microphones at each of the 48 locations in each room. These were then used to check how well, at each location, the recorded sentences matched the quiet versions convolved with the impulse response at that location (see Section 3).

2.3. Experiment-presentation software

MATLAB code was written to provide an interface for experiments whereby the user chooses a room, sound locations, sound wave files for each location, and, optionally, a sound-enhancing algorithm. The program can read a text file containing an ordered list of soundscapes (trials) to play. Alternatively, a QUEST adaptive procedure [27] can be selected to present trials to a listener and calculate an RTS for this configuration.

3. Results

3.1. Engineering metric: quantifying convolved speech and straight speech differences

For each speaker location in each room we had made recordings of 6 of the HINT sentences from our database by playing the quiet recordings from the speaker and recording the results from the microphones in KEMAR (see Section 2). Using one right and one left ear channel, and a sample of 12 locations per room, we took the recordings of the sentences (measured sentences) and compared them to versions derived by convolving (MATLAB frequency domain computations) each original sentence with the impulse response for that speaker location (predicted sentences). In order to determine whether the measured impulse responses accurately described the room environments, we used a signal-to-error ratio (SER) metric to compare each real and convolved pair.

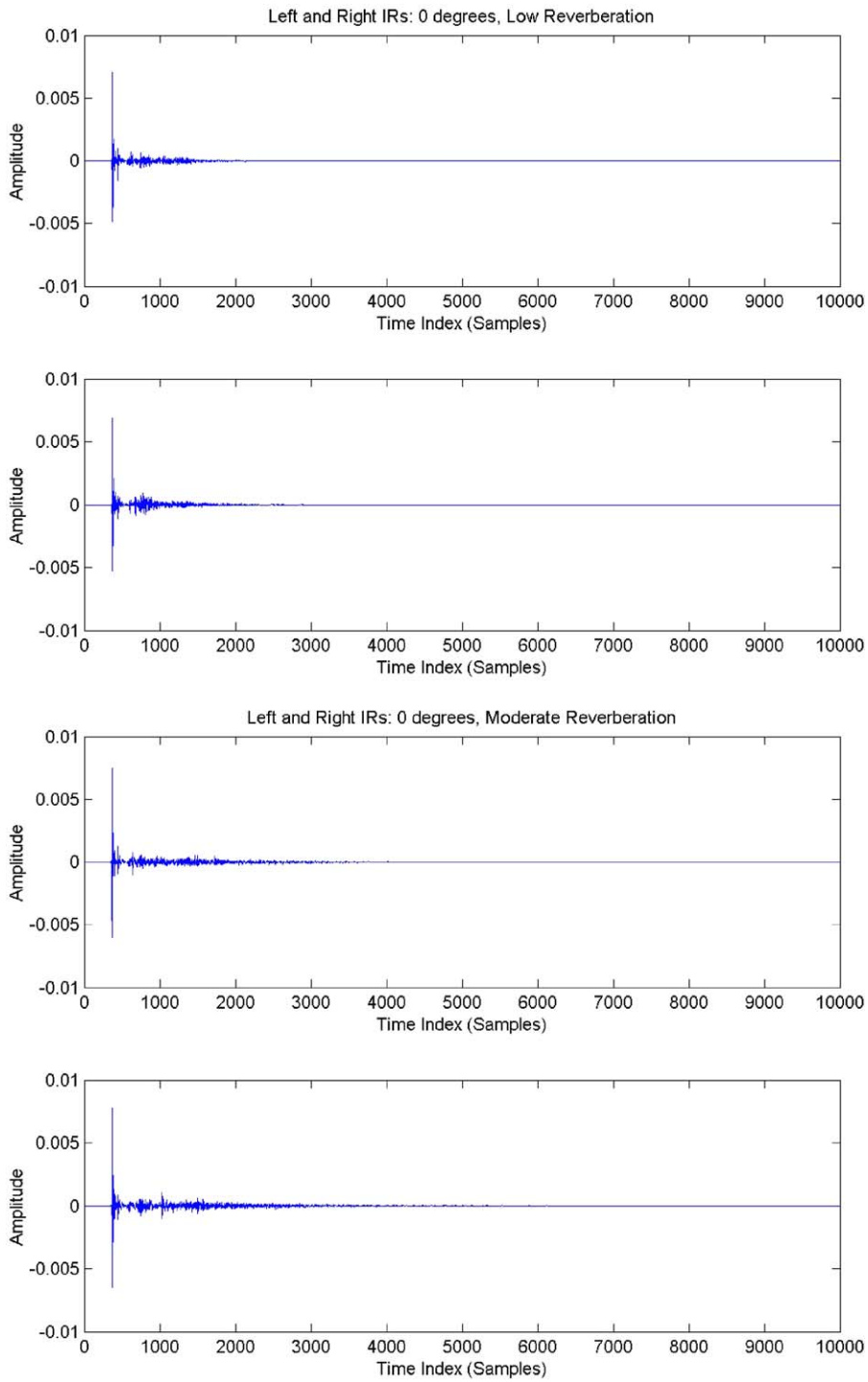


Fig. 3. Measured impulse responses for left and right channels in a room with closed velour drapes resulting in low reverberation (Room 1), and in the same room with open drapes result in moderate reverberation (Room 2).

Before applying this metric, the signals were normalized according to

$$\begin{aligned} \mathbf{x}'_{\text{left}} &= \frac{\mathbf{x}_{\text{left}}}{\sqrt{\text{var}(\mathbf{x}_{\text{left}}) + \text{var}(\mathbf{x}_{\text{right}})}}, \\ \mathbf{x}'_{\text{right}} &= \frac{\mathbf{x}_{\text{right}}}{\sqrt{\text{var}(\mathbf{x}_{\text{left}}) + \text{var}(\mathbf{x}_{\text{right}})}} \end{aligned} \quad (1)$$

where the signal (either real or simulated) is represented by \mathbf{x}_{left} and $\mathbf{x}_{\text{right}}$, and the $\text{var}()$ function simply indicates the variance. This form of normalization eliminates scale differences between the signals, while simultaneously preserving the interaural scale differences necessary for psychological testing. The SER then is simply the ratio of the measured signal over the difference between the measured and predicted signals and was calculated according to

$$\text{SER} = 10 \log_{10} \left(\frac{\text{var}(\mathbf{x}'_{\text{measured}})}{\text{var}(\mathbf{x}'_{\text{measured}} - \mathbf{x}'_{\text{predicted}})} \right). \quad (2)$$

For the case of one of the rooms (Room 3), which was a reverberant lecture room, the signals contained heating ventilation and air conditioning (HVAC) noise, which was of course greater for the straight than for the convolved sentences. Using Cool Edit Pro,¹ we eliminated as much of this noise as possible while still maintaining signal quality. As the HVAC noise is spectrally steady state, its removal should not affect the impulse response. Only after this de-noising stage, were the impulse responses calculated and the SER metric applied.

Applying the impulse responses to the speech signals used, the SERs were calculated for all of the recordings used in this study. The results are shown in Table 1.

The SERs are all high, indicating that the convolved sentences closely resemble the straight sentences. We were also interested in whether the convolved sentences matched the straight sentences across frequencies. We therefore conducted a more detailed analysis of 1.5 s speech extracts from each sentence pair.

Table 1
Mean SER for each room tested

	Mean SER (dB)	SD
Room 1 (low reverberation)	12.85	2.32
Room 2 (medium reverberation)	13.67	2.02
Room 3 (high reverberation)	13.04	1.94

A Daubechies-4 wavelet analysis was done and the SER metric was applied to each scale level. The mean SERs for each condition are shown for scale levels between 2 and 8 in Table 2.

The majority of speech energy is between 80 and 5000 Hz, and this is the range that contributes to speech intelligibility. It can be seen that the SERs are above 10 dB in this range which indicates good agreement between the measured and the convolved sentences.

3.2. Human metric

The engineering metric suggests that convolving the impulse responses with speech sentences gives an excellent approximation to presenting the sentences themselves. Here we provide human perceptual data on this issue. For each of the three rooms we chose 8 sample locations for the perceptual test. At the furthest distance and at the closest distance we chose (1) the medium height, directly in front of the listener, 0°, (2) the medium height, directly behind the listener, 180°, (3) the lower height, directly to the listener's left, 270°, and (4) the higher height, 45° to the right. For each location, we presented six binaural trials to six listeners using Sennheiser DA 200 headphones, a Dell Inspiron 8100 laptop computer, and an Echo Indigo sound card. Each trial comprised two recordings, the measured sentence recording and its predicted (convolved) sentence pair described in the previous section. Because the measured sentences had more noise than the convolved sentences, especially in Room 3 (see above), and because we did not want to distort the speech signals in any way by de-noising the straight recorded signals, we instead added a small amount of white noise to all signals, such that the SNR was approximately 15 dB. On half of the trials the measured sentence was first and on the other half the convolved sentence was first.

¹ The Cool Edit Noise Reduction parameters used were: snapshots in profile, 300; FFT size, 4096 samples; noise reduction level, high; reduce by, 40 dB; precision factor, 7; smoothing amount, 1; transition width, 0.

Table 2
Mean SER for Daubechies-4 wavelet bins for each room tested

Scale level	Room 1		Room 2		Room 3	
	Mean SER	SD	Mean	SD	Mean	SD
8 (86–172 Hz)	14.51	1.78	12.31	2.39	9.43	4.52
7 (172–344 Hz)	16.58	1.67	14.93	2.52	11.95	2.03
6 (344–689 Hz)	16.91	1.61	17.19	2.36	12.82	1.50
5 (689–1387 Hz)	16.17	1.87	16.99	2.47	12.50	1.58
4 (1387–2756 Hz)	14.88	2.52	15.55	2.94	12.12	1.85
3 (2756–5125 Hz)	12.27	3.98	12.93	4.01	11.43	2.95
2 (5125–10250 Hz)	7.91	5.89	8.34	5.82	9.06	5.57

On each trial, the listener was asked to judge which of the two sentences sounded the most “natural” by pressing a number from 1 to 7, where 1 indicated that the first recording sounded much more natural than the second, 7 indicated that the second recording sounded much more natural than the first, and 4 indicated that the two sentences sounded equally natural. The other numbers fell in between on the continuum.

The rating data was transformed such that 1 represented the measured sentence sounding more natural, and 7 the convolved sentence sounding more natural. An analysis of variance (ANOVA) with angle/height (4 levels), room (3 levels), and distance (2 levels), revealed no significant main effects or interactions. We therefore took the average response for each of the six listeners for each of the three rooms and conducted a *t*-test to determine whether the responses differed significantly from 4, the point on the scale of subjective equivalence. As can be seen in Table 3, listeners rated the measured sentences and the predicted (convolved) sentences as sounding equally natural.

In sum, these preliminary engineering and perceptual metrics indicate that the impulse responses are capturing the important acoustic features of the rooms.

4. Conclusions, limitations and future directions

A conceptual overview of R-HINT-E was presented. The R-HINT-E provides an extremely flexible hearing-in-noise testing environment, and combines the best of realistic room acoustics (by using impulse responses of real rooms rather than completely vir-

Table 3
Mean perceptual rating for each room by each subject on a scale from 1 to 7

Listener	Room 1 rating	Room 2 rating	Room 3 rating
DR	3.84	4.50	4.36
GM	4.01	3.75	3.76
KM	3.85	4.18	4.09
LD	3.52	4.28	4.56
SA	4.03	4.00	4.46
MC	4.15	4.10	4.21
Mean (SD)	3.90 (0.22)	4.13 (0.25)	4.24 (0.28)
<i>t</i> -Test probability	0.31	0.25	0.10

It can be seen listeners rate the straight and convolved sentences as equally natural (not statistically different from 4, the point of subjective equivalence)

tual physical models) with experimental control (the conditions are reproducible, SNRs can be changed from trial to trial), the ability to present a very large number of sound conditions (multiple locations and multiple sound types) without having to prerecord each one, and headphone presentation allowing the incorporation of signal-enhancing algorithms. It was shown experimentally that convolving the speech signals introduces very little error in comparison with actual presentation of the speech signals. As well, the straight and convolved signals sound very similar and listeners rate them as sounding equally natural. The next step in the development of R-HINT-E is to do further engineering and human tests of its performance with multiple sound sources. A web-based interactive version that creates soundscapes can be found at <http://trainorlab.mcmaster.ca/ahs/rhinte.htm>, or the

authors can be contacted directly for a non-interactive test copy.

The system has one major limitation as it stands. It is well known that the shape and size of the torso, head and outer ears of an individual's body impose a spectral transfer function on sounds reaching the ears. The pinna shape in particular provides important information for localizing objects in space. Because all of the recordings were done with the KEMAR model, they have the head-related transfer function (HRTF) of KEMAR, so the binaural cues will be somewhat inaccurate for each listener. This limitation could be overcome completely by measuring the HRTF of KEMAR and that of each individual listener, and doing the correction on an individual basis. Measuring individual HRTFs is very time consuming, however, so a compromise would be to have several "usual" HRTFs, such that the HRTF of most listeners can be well approximated by one of them, as is done in other stimulus presentation systems.

At McMaster University, we are developing beamforming and an adaptive neural compensator for signal-enhancing processing in hearing-aids [2,4]. The R-HINT-E will provide us with a realistic hearing in noise test that should give a good indication of the success that our new algorithms will have in the real world.

Acknowledgements

This research was supported by a CRO grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada and is associated with the Blind Source Separation and Application (BLISS) project. We thank Brian Csermak, Jim Ryan, Betty Rule, and Steve Armstrong at Gennum Corporation for help with the impulse measurements and the loan of KEMAR.

References

- [1] T.L. Arbogast, C.R. Mason, G. Y Kidd Jr., The effect of spatial separation on informational and energetic masking of speech, *J. Acoust. Soc. Amer.* 112 (5, Pt. 1) (November 2002) 2086–2098.
- [2] S. Becker, I. Bruce, Neural coding in the auditory periphery: insights from physiology and modelling lead to a novel hearing compensation algorithm, Workshop on Neural Information Coding, Les Houches France, March 2002 (abstract).
- [3] R.S. Bolia, W.T. Nelson, M.A. Ericson, A speech corpus for multitaler communications research, *J. Acoust. Soc. Amer.* 107 (2000) 1065–1066.
- [4] J. Bondy, S. Becker, I.C. Bruce, L.J. Trainor, S. Haykin, A novel signal-processing strategy for hearing aid design: neurocompensation, *Signal Processing* (2003), submitted.
- [5] A.S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambridge, MA, 1990.
- [6] A.W. Bronkhorst, R. Plomp, Binaural speech intelligibility in noise for hearing-impaired listeners, *J. Acoust. Soc. Amer.* 86 (1989) 1374–1383.
- [7] I.C. Bruce, J. Bondy, S. Haykin, S. Becker, A physiologically based predictor of speech intelligibility, Poster Presented at the International Conference of Hearing Aid Research, Lake Tahoe, USA, 2002.
- [8] D.S. Brungart, Informational and energetic masking effects in the perception of two simultaneous talkers, *J. Acoust. Soc. Amer.* 109 (2001) 1101–1109.
- [9] D.M. Chabries, D.V. Anderson, T.G. Stockham, R.W. Christiansen, Application of a human auditory model to loudness perception and hearing compensation, *Proceedings of the IEEE ASSP*, Detroit, USA, 1995, pp. 3527–3530.
- [10] C. Colomes, C. Schmidmer, T. Thiede, W.C. Truermiet, Perceptual quality assessment for digital audio (PEAQ): the proposed ITU standard for objective measurement of perceived audio quality, *Proceedings of the AES Conference*, Florence, Italy, 1995.
- [11] A.J. Duquesnoy, Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons, *J. Acoust. Soc. Amer.* 74 (1983) 739–743.
- [12] J.M. Festen, R. Plomp, Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing, *J. Acoust. Soc. Amer.* 88 (1990) 1725–1736.
- [13] H. Fletcher, W.A. Munson, Loudness, its definition, measurement, and calculation, *J. Acoust. Soc. Amer.* 5 (1933) 82–108.
- [14] J.E. Greenberg, P.M. Peterson, P.M. Zurek, Intelligibility-weighted measures of speech-to-interference ratio and speech system performance, *J. Acoust. Soc. Amer.* 94 (1993) 3009–3010.
- [15] R. Gribonval, E. Vincent, C. Fevotte, L. Benaroya, Proposals for performance measurement in source separation, *Proceedings of the Fourth Symposium on Independent Component Analysis and Blind Source Separation (ICA 2003)*, Nara, Japan, 2003, pp. 763–768.
- [16] S. Hygge, J. Ronnberg, B. Larsby, S. Arlinger, Normal-hearing and hearing impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds, *J. Speech Hearing Res.* 35 (1992) 208–215.
- [17] D.N. Kalikow, K.N. Stevens, L.L. Elliott, Development of a test of speech intelligibility in noise using sentence materials

- with controlled word predictability, *J. Acoust. Soc. Amer.* 61 (1977) 1337–1351.
- [18] J.M. Kates, M.R. Weiss, A comparison of hearing-aid array-processing techniques, *J. Acoust. Soc. Amer.* 99 (1996) 3138–3148.
- [19] S. Kochkin, Marke Trak V: Why my hearing aids are in the drawer: the consumers' perspective, *Hearing J.* 53 (2000) 34–42.
- [20] M.E. Lockwood, D.L. Jones, M.E. Elledge, R.C. Bilger, A.S. Feng, M. Goueygou, C.R. Lansing, C. Liu, W.D. O'Brien Jr., B.C. Wheeler, A minimum variance frequency-domain algorithm for binaural hearing aid processing, *J. Acoust. Soc. Amer.* 106 (1999) 2278A.
- [21] B.C.J. Moore, *Perceptual Consequences of Cochlear Damage*, Oxford University Press, Oxford, 1995.
- [22] B.C.J. Moore, R.W. Peters, M.A. Stone, Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips, *J. Acoust. Soc. Amer.* 105 (1999) 400–411.
- [23] M.J. Nilsson, S.D. Soli, J. Sullivan, Development of a hearing in noise test for the measurement of speech reception threshold, *J. Acoust. Soc. Amer.* 95 (1994) 1985–1999.
- [24] R. Plomp, A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired, *J. Speech Hearing Res.* 29 (1986) 146–154.
- [25] R. Plomp, Noise, amplification, and compression: considerations of three main issues in hearing aid design, *Ear Hearing* 15 (1994) 2–12.
- [26] D.W.E. Schobben, K. Torkkola, P. Smaragdis, Evaluation of blind signal separation methods, *Proceedings of the International Workshop Independent Component Analysis and Blind Signal Separation, France, 1999*, pp. 261–266.
- [27] A.B. Watson, D.G. Pelli, A Bayesian adaptive psychometric method, *Perception Psychophys.* 33 (1983) 113–120.
- [28] A. Westner, V. Michael Bove Jr., Applying blind source separation and deconvolution to real-world acoustic environments, *Proceedings of the 106th Audio Engineering Society (AES) Convention, Munich, Germany, 1999*.