Speech intelligibility model including room and loudspeaker influences

L. Faiget[a)]
L.A.U.T.M, Université de Toulouse-Le Mirail, 5 allées Antonio Machado, 31058 Toulouse Cedex 1, France

R. Ruiz[b)]
L.A.R.A, Université de Toulouse-Le Mirail, 5 allées Antonio Machado, 31058 Toulouse Cedex 1, France

(Received 19 May 1998; accepted for publication 25 January 1999)

Loudspeaker directivity and frequency response are of great importance for speech intelligibility estimation. In this work their respective influence is introduced in a new predictor. Properties of the model are in good agreement with expected variations of scores when radiation and frequency response are modified. An experiment shows the accuracy of the predicted scores. Limitations of the model are discussed and future research perspectives are presented. © *1999 Acoustical Society of America.* [S0001-4966(99)03405-0]

PACS numbers: 43.55.Hy, 43.71.Gv [JDQ]

INTRODUCTION

The three main predictors of speech intelligibility in a room are the energy ratios related measures,[1,2] the Speech Transmission Index (S.T.I.),[3] and the Articulation loss of consonants (Alcons).[4] They can be directly computed from the impulse response of the loudspeaker-room-microphone system. With the two first predictors, intelligibility score prediction is very satisfactory. Correlation coefficients between predicted and measured scores are greater than 0.9 with small standard deviations.

Articulation loss of consonants depends on the distance, the volume of the room, and the reverberation time. A modified formulation introduces the directivity factor Q of the source.[5] Measured scores in different rooms with loudspeakers of high, medium, and low Q have been used to show an inaccurate prediction of Alcons method and better results for the other techniques.[6,7]

Alcons predictor seems to be the least accurate probably because the loudspeaker influence is limited to its directivity factor Q. The other techniques include *in situ* radiation and frequency response effects in the measured impulse response.

The object of this paper is to introduce loudspeaker contribution to intelligibility in a model based on an impulse response estimation.[8] Energy ratio based predictors are chosen to separate room and loudspeaker influences on intelligibility scores.

The basis of the work is a model derived from the Lochner and Burger signal-to-noise ratio[2] and from the useful-to-detrimental energy ratio of Bradley.[1] It is modified in order to introduce the loudspeaker directivity and frequency response related parameters named R_{dir} and R_{rf}. It is tested in highly reverberant conditions. Results of the prediction give a correlation coefficient of 0.96 with a standard deviation of 6%. Simulation of R_{dir} and R_{rf} variations leads to modifications of the predicted scores in good agreement with the current knowledge of the influence of directivity and frequency response on speech intelligibility.

I. INITIAL MODEL

A. Initial predictor

The concept of useful and detrimental sound energy related to speech intelligibility has been introduced by Lochner and Burger[2] and developed by Bradley.[1] The ratio U_τ of useful-to-detrimental energies can be expressed as follows:

$$U_\tau = 10\log\left[\frac{R_\tau}{(1-R_\tau)+10^{(-\mathrm{S/N})/10}}\right]. \tag{1}$$

τ is the time limit between early E_e and late E_l energy; S/N is the signal-to-noise ratio in dB(A) (i.e., the difference between speech and noise sound levels); R_τ is the ratio between early and total energy: $R_\tau = E_e/(E_e+E_l)$.

In Eq. (1) the numerator is the useful energy and the denominator the detrimental energy. The predictor U_τ leads to the following third order polynomial equation between speech intelligibility scores SI(%) (using a Fairbanks rhyme test) and U_{80} in the 1-kHz octave band (Fig. 1):[1]

$$\mathrm{SI}(\%) = 1.219.U_{80} - 0.02466U_{80}^2 + 0{,}00295U_{80}^3 + 95.65. \tag{2}$$

Intelligibility scores have been measured in rooms where reverberation time values vary from 0.8 to 3.8 s.[1]

B. Modified predictor

A third order polynomial equation is chosen to simply represent the nonlinear variation of speech intelligibility scores as a function of S.T.I.[3] or U_{80} [Eq. (2)]. But Fletcher for the Articulation Index[9] and later Lochner and Burger[2] have shown that intelligibility variations are described by a "S" form (Fig. 2) adequate with the lowest scores.

In order to follow this "S" curve, predicted scores $\hat{I}(\%)$ are computed by Eq. (3) based on Fletcher and Galt[10] and Dirks *et al.*[11] formulations:

---

a)Present address: Laboratoire d'Acoustique Appliquée, 5 rue Mchal. Joffre, 06400 Cannes; Electronic mail: Laurent.Faiget@wanadoo.fr
b)Author to whom correspondence should be addressed; Electronic mail: rruiz@univ-tlse2.fr

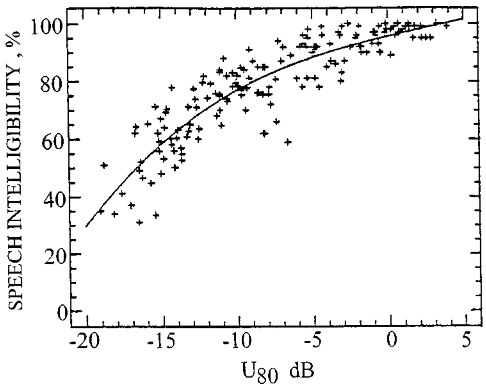

FIG. 1. Measured speech intelligibility scores versus 1-kHz U_{80} values and best-fit third order polynomial (from Ref. 1).

$$\hat{I}(\%) = 100(1 - 10^{-[(S/N)_{eq}+40]/60q})^{n}. \tag{3}$$

$(S/N)_{eq}$ is called the equivalent signal-to-noise ratio and is expressed as follows:

$$(S/N)_{eq} = 10 \log\left(\frac{R_{\tau}^{\alpha}}{(1-R_{\tau}^{\alpha}) + 10^{(-S/N)/10}} \right)$$

$$\text{with } R_{\tau}^{\alpha} = \frac{\int_0^{\tau} \alpha(t) h^2(t)\, dt}{\int_0^{T} h^2(t)\, dt}. \tag{4}$$

S/N is the signal-to-noise ratio in dB(A); α is the fraction of the energy of an individual reflection integrated in the useful energy sum (cf. Sec. I C 1),[2,1] R_{τ}^{α} is the early-to-total energy ratio; n and q are the two regression coefficients (instead of four normally used in a third order polynomial regression). $h(t)$ is the impulse response.

C. Results

1. Rooms and acoustic measurements

Measurements and speech tests have been performed in a reverberant room of 1100 m^3 and in a church of 40 000 m^3

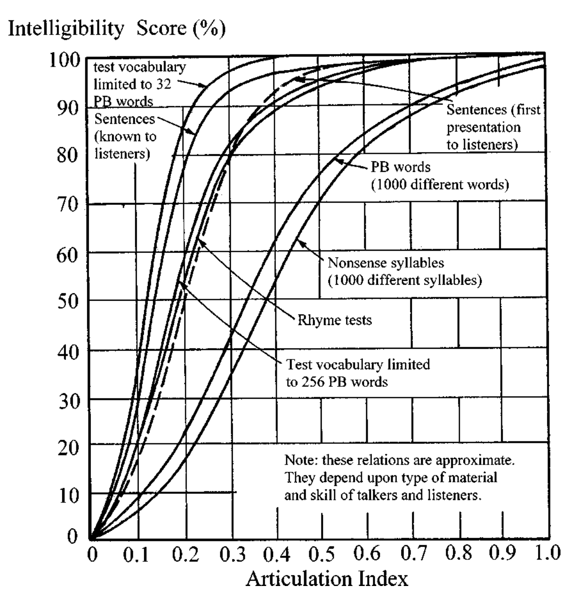

FIG. 2. ''S'' curves between intelligibility scores and Articulation Index (from Ref. 9).

TABLE I. Mean reverberation time (RT_{60}) of the reverberant room and the church A.

f (Hz)	125	250	500	1000	2000	4000	8000
RT_{60} room (s)	10,0	8,8	8,8	8,1	6,9	3,6	1,4
RT_{60} church (s)	14,5	13,4	12,0	10,0	7,3	4,3	1,8

(church A). Reverberation times are reported in Table I. They were measured from the smooth decay curve of the Schroëder integrated impulse response.[12]

Two different loudspeakers have been tested: RB33 and RB90. Their octave band directivity factors Q are in the Appendix.

Twelve on-axis source–receiver combinations are studied, six in each room. All acoustic measurements are obtained from a Maximum Length Sequence method (M.L.S.) for estimating the Impulse Response (IR). A procedure to determine the real duration of the impulse response $h(t)$ is applied.[8,13] Indeed, it has been shown that significant errors on energy ratios are obtained if the total time of acquisition is greater than the real duration T of the impulse response of the system when the measurement is corrupted by extraneous noise. The ratio R_{τ}^{α} is computed on echogram $h^2(t)$ where high sound level individual reflections are identified to apply the α weighting [Eq. (4)]. Lochner and Burger curves (Fig. 10 in Ref. 2) are approximated by the following rule: when the sound level difference between direct sound and individual reflections is greater than 2.5 dB, the 5-dB curve is applied; when the difference is between -2.5 dB and $+2.5$ dB, the 0-dB curve is used and when the difference is less than -2.5 dB, the -5-dB curve is applied. This weighting is done for all the samples of the echogram between 0 and 50 ms.

Various sound levels of white noise are emitted by another loudspeaker in order to create signal-to-noise ratios S/N varying from -10 dB(A) to $+10$ dB(A) in 5-dB(A) steps at the listeners' positions in the room. These positions are also those of the corresponding impulse response measurements. The experiment leads to 61 acoustical combinations of distances and signal-to-noise ratios.

The choice of the time limit τ is based on another set of tests in other large reverberant rooms where 99 different conditions of reverberation, noise, and loudspeakers have been considered.[8] The best correlation coefficient $r=0.99$ between measured and predicted [Eq. (3)] intelligibility scores and the smallest standard deviation $\sigma=9.3\%$ are obtained for $\tau=50$ ms. The variation of τ has been sequenced from 10 to 100 ms by steps of 5 ms. As the purpose of this study is also to predict scores in highly reverberant halls, the value of 50 ms has been selected both as time limit of the α weighting and of the useful energy. R_{50} becomes the D_{50} of Thiele;[14] corrections derived from Lochner and Burger curves are applied to individual reflections between 0 and 50 ms according to the preceding rule. Predictor $(S/N)_{eq}$ is obtained by Eq. (5):

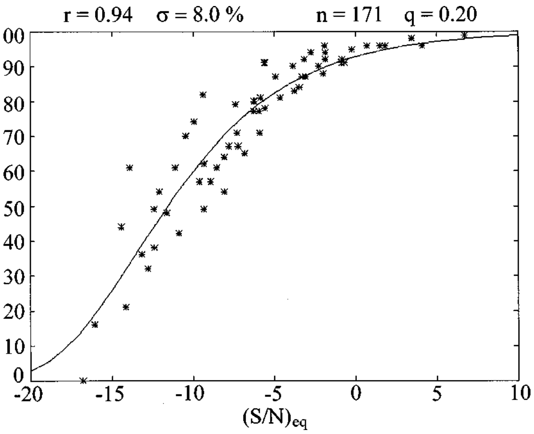

FIG. 3. Measured speech intelligibility scores versus $(S/N)_{eq}$ corresponding values [Eq. (5)] and best least-squares fit based on the model of Eq. (3).

$$(S/N)_{eq} = 10 \log\left(\frac{D_{50}^{\alpha}}{(1-D_{50}^{\alpha}) + 10^{-(S/N)/10}}\right)$$

$$\text{with } D_{50}^{\alpha} = \frac{\int_0^{50} \alpha(t) h^2(t)\,dt}{\int_0^T h^2(t)\,dt}. \tag{5}$$

2. Intelligibility test

The intelligibility test uses 10 phonetically balanced lists of 34 triphonemic French words (mono or dissyllabics). The speaking rate is about nine phonemes per second. Every word is preceded by a sentence without any semantic relation with the word to be recognized. A trial list is proposed. The subjects have to complete a form indicating the vowel(s), consonant(s), syllable(s), or word heard at the end of the sentence. They were approximately 25-year-old students without any auditory problem. The obtained score is the percentage of correctly recognized phonemes.

3. Accuracy of the model

Figure 3 plots the results of the 61 speech intelligibility scores versus $(S/N)_{eq}$ corresponding values. The curve is a regression of the form of Eq. (3). The least-squares fit is obtained for $n=171$ and $q=0.20$. The correlation coefficient r is then equal to 0.94 and the standard deviation is 8.0%. Other forms of regressions have been tested such as sigmoïdal functions. The accuracy of this model has not been improved in terms of correlation coefficient and standard deviation.

II. LIMITS OF THE MODEL

The aim of this paper is to identify and quantify separately the room and loudspeaker influence on intelligibility scores. In order to prove the necessity of a predictor depending on loudspeaker features and to evaluate the limits of the preceding global model of intelligibility, some experimental results have to be considered (Tables II and III).

First, Table II shows that measured scores I% are better predicted from the model for impulse responses obtained with the RB33 source. The higher the ratio D_{50}^{α} is, the more directive the source will be. In very noisy conditions, the RB90 leads to real scores greater than those obtained with the RB33 in the same source–receiver position. The same is true for predicted scores but with less accuracy. The increase of definition D_{50}^{α} improves both measured and estimated scores but in a different way. Therefore, directivity influence must be better accounted for by the model.

Second, measured scores differ in two source–receiver positions where loudspeaker, signal-to-noise ratio, and definition D_{50}^{α} remain the same (Table III). The difference is greater for low S/N ratios. But the predicted corresponding scores will be the same in the two rooms for the same S/N ratio. Indeed, the $(S/N)_{eq}$ predictor only depends on these parameters [Eq. (5)]. This result implies that both D_{50}^{α} and S/N ratios are not sufficient to predict intelligibility scores. Source–receiver position in the hall has an effect on measured scores, whereas parameters of the model do not act in the same way.

Finally, room and loudspeaker influence should be considered separately. A loudspeaker parameter related to its radiation in the room should be introduced to improve prediction particularly for high reverberation time values and low signal-to-noise ratios.

TABLE II. Comparison between measured I(%) and predicted $\hat{I}$(%) scores for the RB33 and the RB90 loudspeakers in the reverberant room at the distance 4 m for three signal-to-noise ratios in dB(A) (values in brackets are associated standard deviations).

Reverberant room	$(S/N)_{dB(A)}$	$(S/N)_{eq}$	I(%)	$\hat{I}$(%)
RB33	−10	−16	16,2 (11,6)	17,9
($\alpha.D_{50}$=22,7%)	−5	−11,6	48 (4,8)	47,9
	0	−8,1	63,7 (11,7)	69,7
RB90	−10	−13,9	61,0 (2,7)	31,8
($\alpha.D_{50}$=43,1%)	−5	−9,4	81,6 (4,8)	61,7
	0	−5,6	90,7 (2,6)	79,2

III. ROOM AND LOUDSPEAKER INFLUENCES

The general form of regression is the same as in Eq. (3). The objective is to introduce room and loudspeaker influence in the $(S/N)_{eq}$ predictor of Eq. (5). It is necessary to identify and separate their respective contribution in the echogram.

The prediction is based on a measurement of the impulse response by an M.L.S. technique in a given source-microphone position in the room. This impulse response $h(t)$ is defined by the following convolution equation:

TABLE III. Comparison between measured intelligibility scores I(%) for various signal-to-noise ratios but for the same value of Definition D_{50}^{α} (values into brackets are associated standard deviations).

$(S/N)_{dB(A)}$	I(%) Reverberant room 4 m	I(%) Church A 16 m
−5	81,6 (4,8)	62,0 (4,0)
0	90,7 (2,6)	78,2 (3,3)
5	91,4 (2,2)	86,8 (5,9)
10	96,1 (2,1)	91,7 (1,3)

$$h(t)=h_{hp}(t)*h_s(t)+n(t). \tag{6}$$

$h_{hp}(t)$ is the impulse response of the loudspeaker; $h_s(t)$ is the impulse response of the room; $n(t)$ is the sum of the acoustic and electrical (or computation) noises.

A procedure of deconvolution is applied to obtain $h_s(t)$. It consists in finding the inverse filter $f(t)$ which verifies $f(t)*h(t)=h_s(t)$ by inverting the module and taking the opposite of the phase of $H_{hp}(f)$, Fourier transform of $h_{hp}(t)$.[15] The deconvolution procedure uses the axial impulse response of the loudspeaker. It is measured beforehand in an anechoïc chamber or by the near-field/far-field technique.[16] The complex multiplication $F(f).H(f)$ is equal to $H_s(f)$ and an inverse Fourier transform leads to $h_s(t)$.

But the main assumption of the method is the absence of noise. It must be removed before deconvolution by averaging impulse responses after each Maximum Length Sequence emission.

When this deconvolution procedure is achieved, the room and loudspeaker influence on the speech intelligibility predictor can be studied from $h_s(t)$, $h(t)$, and $h_{hp}(t)$.

A. Room influence: D_{50}^s

The concept of useful and detrimental energies is applied to $h_s(t)$ to define room influence. The proposed ratio D_{50}^s is the same as definition D_{50} but is computed on the deconvolved impulse response $h_s(t)$ and not on the global one $h(t)$:

$$D_{50}^s=\frac{\int_0^{50} h_s^2(t)dt}{\int_0^T h_s^2(t)dt}. \tag{7}$$

T is the total time of acquisition and 0 is the time of direct sound arrival measured on $h(t)$. Integration onto the total time T does not introduce errors because impulse response is noiseless.[13] It has been demonstrated that, without noise, the values of the energy ratios are close to equal even if the total duration of acquisition changes.[8]

If E_s is the energy of speech, useful E_u and detrimental E_d ones are, respectively:

$$E_u=D_{50}^s E_s, \tag{8}$$

$$E_d=(1-D_{50}^s)E_s. \tag{9}$$

The ratio D_{50}^s represents the energetic contribution of reflections in the first 50 ms after direct sound arrival to the entire energy of reflections in the room. The concept is the same as Bradley but applied to the deconvolved impulse response.

B. Loudspeaker influence

Loudspeaker influence on speech intelligibility is studied from its on-axis impulse response measurement. The assumption is that all speech energy emitted by the loudspeaker is useful to intelligibility contrary to the room influence in which reverberant energy can act as a noise to degrade speech perception. Therefore, the corresponding features will be in the numerator of the $(S/N)_{eq}$ predictor.

TABLE IV. Values of D_{50}, D_{50}^s, and R_{dir} for RB33 and RB90 loudspeakers at the same positions in the two halls.

		Distance (m)	D_{50} (%)	D_{50}^s (%)	R_{dir}
Reverberant room	RB33	2	51	29	1,8
		4	27	19	1,4
		6	23	18	1,3
	RB90	2	59	9	6,6
		4	43	6	7,2
		6	38	5	7,6
Church A	RB33	2	79	8	9,9
		8	41	14	2,9
		16	21	17	1,2
	RB90	2	80	7	11,4
		8	64	8	8,0
		16	44	9	4,9

1. R_{dir} factor

The R_{dir} factor is defined by the following ratio.

$$R_{dir}=\frac{D_{50}}{D_{50}^s}=\frac{\int_0^{50} h^2(t)dt/\int_0^T h^2(t)dt}{\int_0^{50} h_s^2(t)dt/\int_0^T h_s^2(t)dt}. \tag{10}$$

R_{dir} is linked to the loudspeaker radiation in the room. In an ideal reflecting room with an ideal omnidirectional source, the amplitude of the first reflections is nearly the same as that of the direct signal and R_{dir} reaches 1. If a very directive source is used whose the main axis of radiation is in the direction of the microphone, D_{50}^s becomes lower than D_{50} which tends to one, and R_{dir} is greater. In an ideal anechoïc room, D_{50} is equal to one because the duration of $h(t)$ is less than 50 ms $[h(t)=h_{hp}(t)]$. D_{50}^s is also equal to 1 because $h_s(t)$ is equivalent to a Dirac impulsion; therefore $R_{dir}=1$.

The higher the directivity of a loudspeaker, the higher R_{dir} for a measurement of an impulse response at the same source–receiver location in the same room but for different loudspeakers.

By examining the results of Table IV, for the same loudspeaker but for the two halls, it can be noted that R_{dir} decreases when the source–receiver distance increases except for the reverberant room with the RB90 loudspeaker. This remark shows that R_{dir} is not a characteristic specific to the loudspeaker directivity: it is related to its radiation in the room. Both in the two halls, the values of D_{50} decrease with distance, thus indicating a higher energy in the latter part of the impulse response and/or a lower one in the earlier part.

The reverberant sound field becomes more important and the direct sound level diminishes. For D_{50}^s the variation

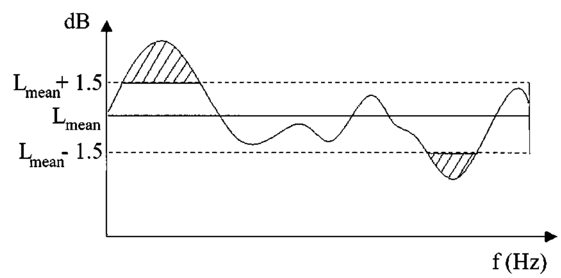

FIG. 4. Simulation of a loudspeaker frequency response with up and down limits for the computation of R_{rf} ($\pm$1,5 dB from mean sound level L_{mean}).

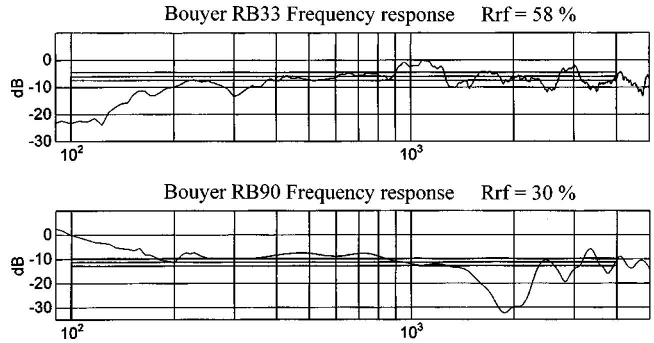

FIG. 5. Frequency responses of the two loudspeakers used in the experiments with up and down limits for the computation of R_{rf} ($\pm 1{,}5$ dB from mean sound level).

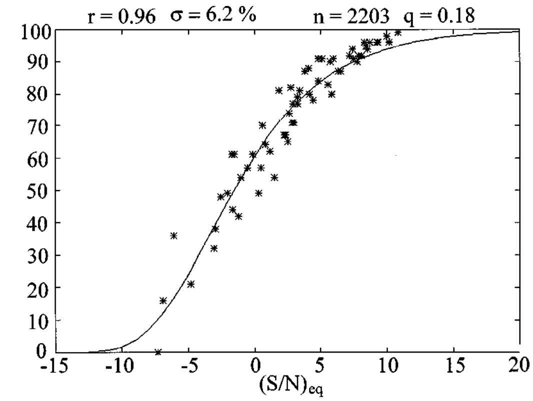

FIG. 6. Measured speech intelligibility scores versus $(\mathrm{S/N})_{\mathrm{eq}}$ predictor corresponding values [Eq. (3)] and best least-squares fit based on the model of Eq. (3).

is different in the two rooms. The deconvolution of loudspeaker impulse response shows that D_{50}^{s} increases with distance in the church for the two sources tested. This is explained by the number of reflections in the first 50 ms which grows with distance in such a long room. The result is not observed in the reverberant room because its volume is smaller. The sense of variation of D_{50} and D_{50}^{s} with distance can be different which leads also to a variable one for R_{dir}.

2. R_{rf} factor

The influence of bandwith on intelligibility has been studied in telephony applications. Experiments with high- and low-pass filtered speech and masking white noise have shown that the band 300–4000 Hz is sufficient to ensure intelligibility syllabe scores greater than 90%.[17] In other similar experiments with nonsense CVC word lists, the score was greater than 97% for the band 100–4000 Hz.[10] Considering that this band is sufficient for a good intelligibility, it is necessary to ensure a flat frequency response of the system in order to reproduce speech without any alteration. Bucklein has measured resonance and antiresonance influence on intelligibility scores.[18] The influence exists, for example, a 25-dB amplitude resonance in the band 1000–2000 Hz leads to a 4% decrease of the score. An equivalent antiresonance only leads to a 1% decrease. He concludes that resonances have a greater detrimental influence than antiresonances and that the frequency response of a system can accept irregularities but with as few wide bandwith resonances of high level as possible.

Here the purpose is to choose an amplitude tolerance in the band 100–4000 Hz and to define a feature to quantify the frequency response shape on intelligibility scores. The tolerance is fixed at ± 1.5 dB in order to be more restrictive than the usual ± 3-dB audio tolerance (but on a larger bandwith), and than the imperceptible 2-dB resonances on frequency response.[19]

The criteria R_{rf} which takes into account frequency response fluctuations in the 100–4000 Hz band is defined by the following relation:

$$R_{\mathrm{rf}}=\frac{E_{hp}-E_{n,hp}}{E_{hp}}=1-\frac{E_{n,hp}}{E_{hp}}. \tag{11}$$

Quantities E_{hp}, $E_{n,hp}$ are obtained when the tolerance $\pm 1{,}5$ dB is applied on the loudspeaker frequency response (Fig. 4). E_{hp} is the energy of the frequency response in the band

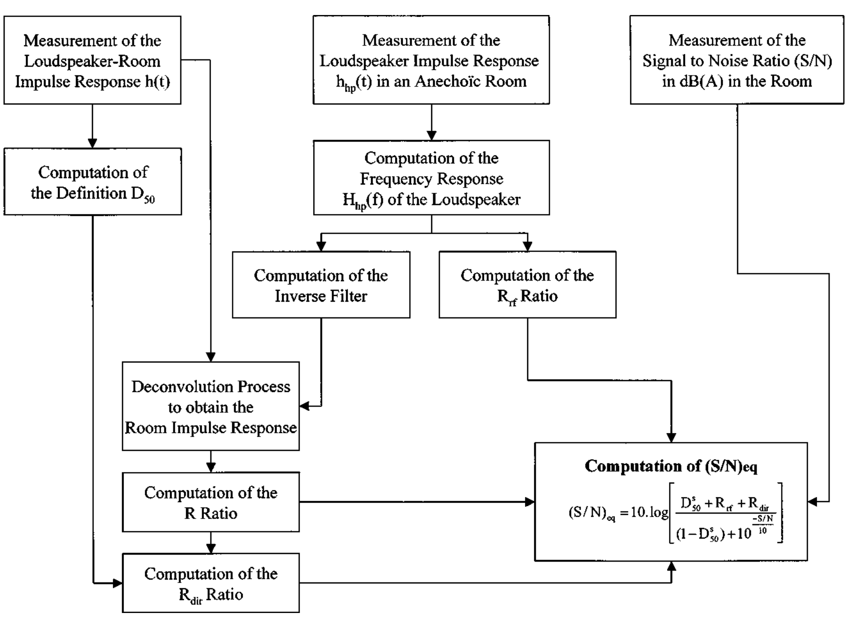

FIG. 7. Method of $(\mathrm{S/N})_{\mathrm{eq}}$ computation.

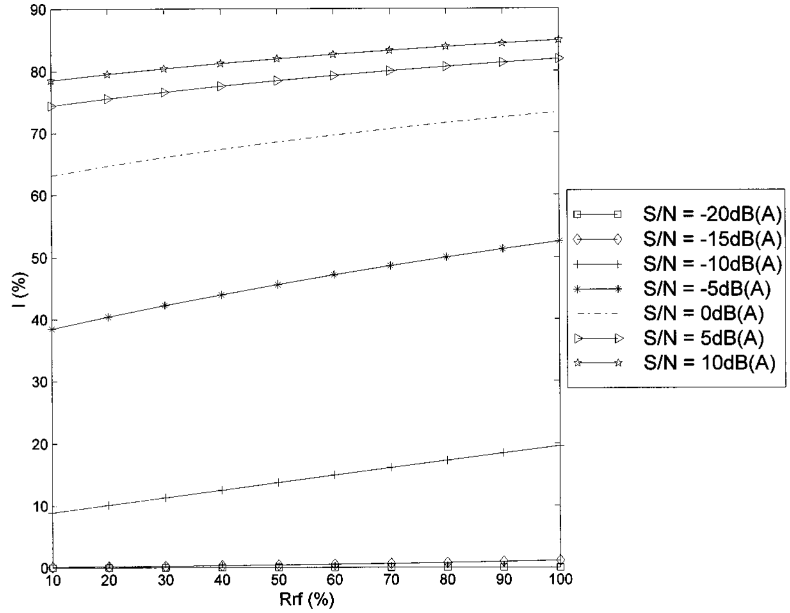

FIG. 8. Predicted intelligibility score versus R_{rf} for different signal-to-noise ratios (S/N). Values are computed from the model [Eqs. (13) and (14)] with $R_{\mathrm{dir}}=1{,}4$ and $D_{50}^{s}=20\%$.

100–4000 Hz. All of the energy above and under limits of the tolerance are summed up to give $E_{n,hp}$. In the case of antiresonances, even if energy is not present under the lower limit, this missing quantity is added to $E_{n,hp}$ like energy of the resonances above the upper limit. Therefore, if E_p is the speech energy, the useful energy transmitted by the loudspeaker is the product $R_{\mathrm{rf}}E_p$. When the frequency response is within the tolerance, $E_{n,hp}$ is null and 100% of the energy is useful. Figure 5 shows two loudspeaker frequency responses with their respective R_{rf} values.

IV. MODEL INCLUDING SEPARATED INFLUENCE OF LOUDSPEAKER AND ROOM

The new model equation is given by Eq. (3). The equivalent signal-to-noise ratio $(\mathrm{S/N})_{\mathrm{eq}}$ is:

$$(\mathrm{S/N})_{\mathrm{eq}}=10\log\left(\frac{D_{50}^{s}E_s+R_{\mathrm{rf}}E_s+R_{\mathrm{dir}}E_s}{(1-D_{50}^{s})E_s+E_n}\right). \tag{12}$$

The numerator sums the useful parts of speech energy E_s and the denominator the detrimental ones (E_n is the noise energy). After simplification, Eq. (12) becomes:

$$(\mathrm{S/N})_{\mathrm{eq}}=10\log\left(\frac{D_{50}^{s}+R_{\mathrm{rf}}+R_{\mathrm{dir}}}{(1-D_{50}^{s})+10^{-(\mathrm{S/N})/10}}\right). \tag{13}$$

S/N is the signal-to-noise ratio in dB(A).

Figure 6 plots the least-squares regression line between the 61 measured scores (cf. Sec. I C 1) and the predicted ones by the form of Eq. (3) with $(\mathrm{S/N})_{\mathrm{eq}}$ given by Eq. (13). Parameters n and q are, respectively, equal to 2203 and 0.18.

The equation of the model is the following:

$$I(\%)=100(1-10^{-[(\mathrm{S/N})_{\mathrm{eq}}+40]/(60\times0.18)})^{2203}. \tag{14}$$

The coefficient of determination r^2 is equal to 0.92, which means that 92% of the total variance is explained by the regression giving a correlation coefficient r of 0.96.[20] Standard deviation is equal to 6.2%.

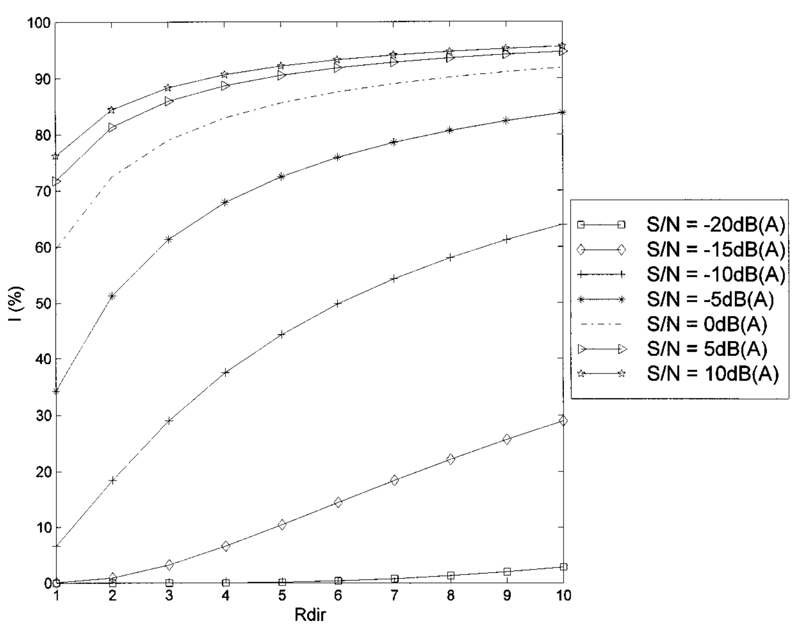

FIG. 9. Predicted intelligibility score versus R_{dir} for different signal-to-noise ratios (S/N). Values are computed from the model [Eqs. (13) and (14)] with $R_{\mathrm{rf}}=30\%$ and $D_{50}^{s}=20\%$.

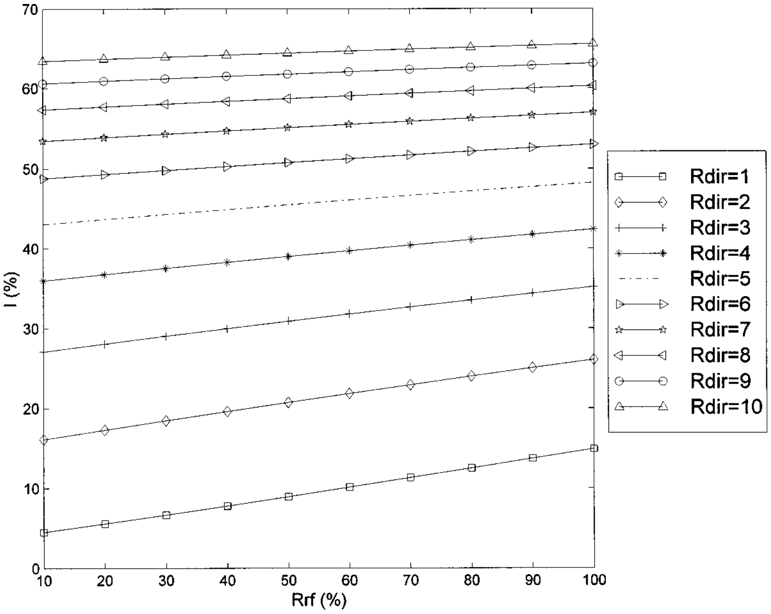

FIG. 10. Predicted intelligibility score versus R_{rf} for different R_{dir}. Values are computed from the model [Eqs. (13) and (14)] with $(S/N)=-10$ dB(A) and $D_{50}^{s}=20\%$.

It appears that the separation of room and loudspeaker influence on intelligibility scores estimation is obtained without a decrease of the quality of the estimation. The accuracy of the model is slightly better comparing with the results of the first predictor (Fig. 3).

A. Computation of the equivalent signal-to-noise ratio $(S/N)_{eq}$

Figure 7 illustrates the procedure of $(S/N)_{eq}$ computation. Three measurements are necessary. The first one is the estimation of loudspeaker impulse response on its main axis of radiation in an anechoïc room or by a near-field/far-field technique. The room-loudspeaker impulse response $h(t)$ and the signal-to-noise ratio depend on the room tested.

B. Properties of the model

Properties of the model are obtained from Eq. (14) with $(S/N)_{eq}$ given by Eq. (13). To show the respective influence of R_{rf} and R_{dir}, the variations of predicted intelligibility scores are plotted for constant values of the other parameters (D_{50}^{s}, S/N, and, respectively, R_{dir} and R_{rf}). Charts of Figs. 8–11 are then obtained. Values of S/N vary from a noisy situation [-20 dB(A)] to a comfortable one [$+10$ dB(A)] by steps of 5 dB(A). R_{rf} goes from 10% to 100%, i.e., from a very irregular loudspeaker frequency response to a perfectly flat one. Variations of R_{dir} are chosen from 1 to 10. One is representative of an omnidirectional sound source in a reverberant room and ten is representative of a more directive one in a more absorbing enclosure. Figures 8 and 9 show that the improvement of scores does not vary linearly versus S/N. For a given value of R_{rf} or R_{dir}, a S/N variation of 5 dB(A)

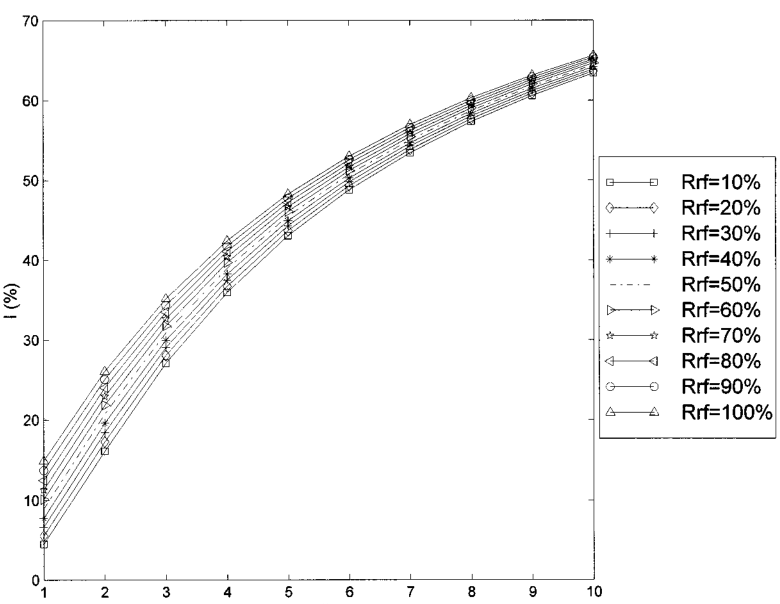

FIG. 11. Predicted intelligibility score versus R_{dir} for different R_{rf}. Values are computed from the model [Eqs. (13) and (14)] with $(S/N)=-10$ dB(A) and $D_{50}^{s}=20\%$.

TABLE V. Mean reverberation time (RT_{60}) of church B in which accuracy of the model has been tested.

f (Hz)	125	250	500	1000	2000	4000
RT_{60} (s)	5,8	6,7	7,1	6,2	5,3	3,8

between two negative values has a greater influence than between two positive ones. For example, if S/N is modified from -10 dB(A) to -5 dB(A), the gain is about 30% (Fig. 8), but if this modification occurs from 5 to 10 dB(A) the gain is less than 5%. Examination of Fig. 8 reveals that the greater improvement of score is approximatively 10% for S/N$=-10$ dB(A) when $R_{\rm rf}$ varies from 10% to 100%. The effect of a regular frequency response is important when S/N is low.

For S/N$=+10$ dB(A) improvement is about 6% between $R_{\rm rf}=10\%$ and 100%. Such a result is identical to Bucklein one. He has shown that a great degradation of frequency response can induce a 4% decrease of scores (with phoneme lists) in situations where masking noise has no influence.[20]

Influence of $R_{\rm dir}$ is shown in Fig. 9. The effect of varying $R_{\rm dir}$ is more important than the $R_{\rm rf}$ one. The gain is about 56% when $R_{\rm dir}$ varies from 1 to 10 for an S/N$=-10$ dB(A). This improvement is smaller when signal-to-noise ratio is greater [20% for S/N$=+10$ dB(A)].

When signal-to-noise ratio is too low [-15 or -20 dB(A)], influence of $R_{\rm dir}$ or $R_{\rm rf}$ is not as important as for other negative S/N. The model shows that when noise sound level becomes too high, scores cannot be improved even with directive and/or flat frequency response loudspeakers. A substantial gain of score is obtained when S/N becomes greater than or equal to -10 dB(A).

As shown in Fig. 10, for a given negative value of S/N ratio, a significant enhancement of scores can be obtained by increasing directivity whatever the value of $R_{\rm rf}$. Figure 11 makes it clear that the improvement of scores is smaller when the frequency response is flattened for the same negative S/N ratio and whatever $R_{\rm dir}$ is.

When noise is predominant, the directivity influence is greater. In a noisy ambience, the use of a highly directive source is recommended for a good intelligibility. But in a hall with a positive S/N ratio, the use of a directive loudspeaker is less necessary. Such conclusions are identical to those of Jacob in his experiments with sources of different directivity factors.[6]

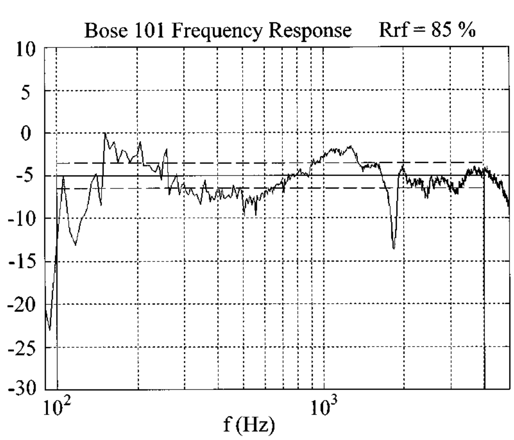

FIG. 12. Bose 101 loudspeaker frequency response with limits for the computation of $R_{\rm rf}$.

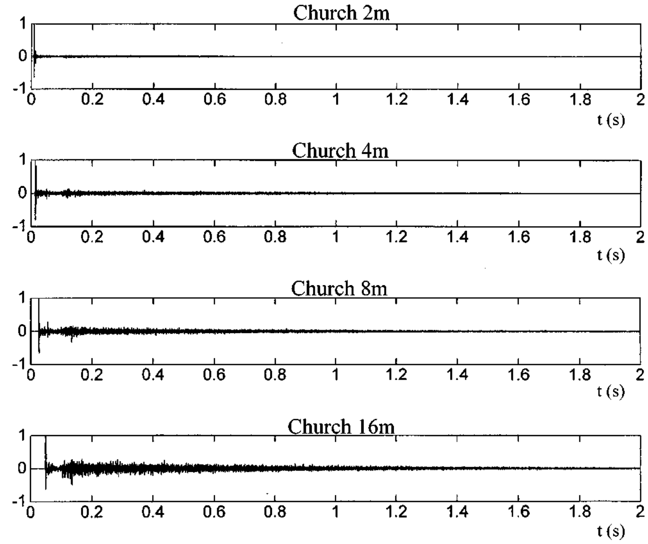

FIG. 13. Impulse responses in the church at 2, 4, 8, and 16 m from the Bose 101 loudspeaker.

All of the plots have been obtained for a $D_{50}^{s}=20\%$. When other values of D_{50}^{s} are chosen, the conclusions about $R_{\rm dir}$ and $R_{\rm rf}$ effects on scores are the same.

Properties of the model are in good agreement with known loudspeaker and room influence on intelligibility scores.

C. Model accuracy

An experiment is done to test the accuracy of the model. The hall is another empty church of 12 000 m^3 (40 m$\times$15 m$\times$20 m) (church B). A new loudspeaker (Bose 101) is placed on the altar and four points on axis are chosen at distances of 2 m, 4 m, 8 m, and 16 m from the source. Beforehand, reverberation time has been measured and averaged at this points (Table V).

Measurements of on-axis frequency response of the Bose 101 loudspeaker in an anechoïc room have given a value of 85% for $R_{\rm rf}$ (Fig. 12). Figure 13 shows impulse responses at the chosen points in church B before deconvolution procedure. Results of the computation of D_{50}^{s} and $R_{\rm dir}$ are reported in Table VI. With these values $({\rm S/N})_{\rm eq}$ is computed [Eq. (13)], and predicted scores [Eq. (14)] are indicated in Table VII for the corresponding signal-to-noise ratios.

TABLE VI. Values of D_{50}, D_{50}^{s}, $R_{\rm dir}$ in church B for the loudspeaker used ($R_{\rm rf}=85\%$).

Distance (m)	D_{50}(%)	D_{50}^{s}	$R_{\rm dir}$	$R_{\rm rf}$(%)
2	79	52	1,5	85
4	56	31	1,8	85
8	34	22	1,5	85
16	13	11	1,2	85

TABLE VII. Predicted $\hat{I}(\%)$ and measured $I(\%)$ scores for different signal-to-noise ratios (S/N) at 2, 4, 8, and 16 m from the source in church B.

Distance (m)	S/N (dB(A))	$\hat{I}(\%)$	$I(\%)$
	3	83	84
2	0	76	69
	−5	54	48
	0	74	85
4	−2	68	72
	−6	47	49
	0	70	78
8	−3	59	63
	−8	29	36
	0	64	58
16	−3	52	37
	−8	22	21

Intelligibility tests (cf. Sec. I C 2) have been performed at the same points as impulse response measurements with 12 subjects divided in 4 groups of 3, each group at one of the four distances. White masking noise is emitted by another loudspeaker in the same vertical plane as the one used for the lists. The sound level of emission is chosen in order to satisfy the required S/N ratios. The sound pressure level of the speech lists is 70 dB(A).

Scores are averaged for each group of listeners and Table VII is a comparison between the predicted $\hat{I}(\%)$ and measured $I(\%)$ scores. The mean absolute difference is 6%. Prediction is in good agreement with the measurements.

V. CONCLUSION

All of the speech tests carried out in order to build the score database have been performed with listeners in the main radiating axis of the loudspeaker. The procedure of deconvolution to obtain the room impulse response uses an on-axis measurement of the loudspeaker impulse response. Therefore, impulse response measurements in the hall are always done in this axis. Scores are predicted by the model in this particular but essential direction of propagation. It is the main actual limitation of the model. The prediction in other directions would require loudspeaker impulse response measurements in these angles of radiation. It would be necessary to build a new database of speech intelligibility tests for various directions, signal-to-noise ratios, and reverberation situations.

A second limitation of the model results in the use of a single loudspeaker for the database and the prediction. The model is not adapted to a multi-loudspeaker sound reinforcement system because in the first 50 ms of the echogram nearby loudspeakers can have a detrimental influence. Deconvolution becomes more complex. When loudspeakers are distant from each other more than 17 m, on-axis prediction acts as if close loudspeaker contributions belong to the detrimental part of the sound field. In order to examine useful or detrimental role of direct sound issued from close loudspeakers, complementary experiments are needed.

Type and spatial position of noise in the hall play an important role in speech intelligibility. All of the experiments have been performed with a wide-band white noise source (20–20 000 Hz) in the same vertical plane as the speech loudspeaker. Weighting coefficients depending on detrimental influence of position(s) and spectral (or time) properties of masking noise(s) could be introduced in the predictor.

Simulations of the R_{dir} and R_{rf} influence on score have shown that when the signal-to-noise ratio is too small, the enhancement is difficult. But in real cases, an increase of high frequency sound levels can improve intelligibility. Such a kind of modification of the source is not taken into account by the model and could also be included after complementary studies.

Another limitation concerns voice quality. In related security applications where recorded messages are not used but a ''natural voice'' must speak, quality of the voice is very important not only for message recognition but also for the emotional content transmitted. Indeed, listener reaction can depend on its perception of speaker emotion. Some acoustic modifications of vowels and consonants are measured when a speaker is under stress,[21] but they are not taken into account in the speech intelligibility models.

Finally, even if these points limit the application of the prediction for the moment, this model is a basis for future development and now includes a separation of room, loudspeaker and masking noise influence.

APPENDIX

TABLE AI. Octave band directivity factors Q for the three loudspeakers used (RB33, RB90, Bose 101).

f (Hz)	125	250	500	1000	2000	4000
Q_{RB33}	1,1	1,2	1,6	2,1	2,7	4,1
Q_{RB90}	1,7	1,7	2,5	2,5	2,8	4,9
Q_{Bose101}	1,2	1,3	2,0	3,2	3,8	6,4

[1] S. Bradley, ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845 (1986).

[2] J. P. A. Lochner and J. F. Burger, ''The influence of reflections on auditorium acoustics,'' J. Sound Vib. **1**, 426–454 (1964).

[3] H. J. M. Steeneken and T. Houtgast, ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **67**, 318–326 (1980).

[4] V. M. A. Peutz, ''Articulation loss of consonants as a criterion for speech-transmission in a room,'' J. Audio Eng. Soc. **19**, 915–919 (1971).

[5] W. Klein, ''Articulation loss of consonants as a basis for the design and judgment of sound reinforcement systems,'' J. Audio Eng. Soc. **19**, 920–922 (1971).

[6] K. D. Jacob, ''Subjective and predictive measures of speech intelligibility—the role of loudspeaker directivity,'' J. Audio Eng. Soc. **33**, 950–955 (1985).

[7] K. D. Jacob, ''Correlation of speech intelligibility tests in reverberant rooms with three predictive algorithms,'' J. Audio Eng. Soc. **37**, 1020–1029 (1989).

[8] L. Faiget, ''Séparation de l'Influence du Local et de l'Enceinte pour la Prévision de l'Intelligibilité dans des Conditions d'Écoute Difficiles,'' Thèse de Doctorat de l'Université Toulouse III, Toulouse, France (1997).

[9] ANSI S315-1969, ''American National Standard methods for the computation of the Articulation Index'' (ANSI, New York, 1969).

[10] H. Fletcher and R. H. Galt, ''The perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151 (1950).

[11] D. Dirks, T. S. Bell, R. N. Rossman, and G. E. Kincaid, ''Articulation index predictions of contextually dependent words,'' J. Acoust. Soc. Am. **80**, 82–92 (1986).

[12] M. R. Schroëder, ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412 (1965).
[13] L. Faiget, C. Legros, and R. Ruiz, ''Optimization of the impulse response length: Application to noisy and highly reverberant rooms,'' J. Audio Eng. Soc. **46**, 741–750 (1998).
[14] R. Thiele, ''Richtungsverteilung und Zeitfolge der Schallruckwurfe in Raumen,'' Acustica **3**, 291–302 (1953).
[15] J. Berkhout, D. De Vries, and M. M. Boone, ''A new method to acquire impulse responses in concert halls,'' J. Acoust. Soc. Am. **68**, 179–183 (1980).
[16] C. Struck and S. Temme, ''Simulated free field measurements,'' J. Audio Eng. Soc. **42**, 467–482 (1994).
[17] I. Pollack, ''Effects of high pass and low pass filtering on the intelligibility of speech in noise,'' J. Acoust. Soc. Am. **20**, 259–266 (1948).
[18] R. Bucklein, ''The audibility of frequency response irregularities'' J. Audio Eng. Soc. **29**, 126–131 (1981).
[19] D. E. L. Shorter, ''A survey of performance criteria and design considerations for high-quality monitoring loudspeakers,'' J. Acoust. Soc. Am. **2**, 56–71 (1958).
[20] T. H. Wonnacott and R. J. Wonnacott, *Statistique* (Wiley, New York, 1995).
[21] R. Ruiz, E. Absil, B. Harmegnies, C. Legros, and D. Poch, ''Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions,'' Speech Commun. **20**, 111–129 (1996).