Effect of room absorption on human vocal output in multitalker situations

Lau Nijs,[a] Konca Saher,[b] and Daniël den Ouden[c]
Faculty of Architecture, Delft University of Technology, Berlageweg 1, 2628 CR Delft, The Netherlands

(Received 5 December 2006; revised 1 November 2007; accepted 6 November 2007)

People increase their vocal output in noisy environments. This is known as the Lombard effect. The aim of the present study was to measure the effect as a function of the absorption coefficient. The noise source was generated by using other talkers in the room. A-weighted sound levels were measured in a 108 m^3 test room. The number of talkers varied from one to four and the absorption coefficients from 0.12 to 0.64. A model was introduced based on the logarithmic sum of the level found in an anechoic room plus the increasing portion of noise levels up to 80 dB. Results show that the model fits the measurements when a maximum slope of 0.5 dB per 1.0 dB increase in background level is used. Hence Lombard slopes vary from 0.2 dB/dB at 50 dB background level to 0.5 dB/dB at 80 dB. In addition, both measurements and the model predict a decrease of 5.5 dB per doubling of absorbing area in a room when the number of talkers is constant. Sound pressure levels increase for a doubling of talkers from 3 dB for low densities to 6 dB for dense crowds. Finally, there was correspondence between the model estimation and previous measurements reported in the literature. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821410]

PACS number(s): 43.55.Dt, 43.55.Hy, 43.70.Bk [RYL] Pages: 803–813

I. INTRODUCTION

In 1911, Lombard published an article on the effect of people tending to raise their voices in noisy environments, now known as the Lombard effect.[1] Since then many researchers have tried to establish the increase in human vocal output as a function of noise level, type of noise, etc. Lane and Tranel conducted an overview of the literature in 1971[2] containing over 200 citations. They found different slopes but the majority of the results is close to an increase in vocal output of approximately 0.5 dB per 1.0 dB increase in noise level. This will be denoted by 0.5 dB/dB throughout this paper, where all sound levels are assumed to be A-weighted. For low noise levels, below about 50 dB, lower slopes are found and lower slopes are also found for very high levels, over 100 dB, since there is a maximum to the level of human speech.

Lane and Tranel focus heavily on the feedback loops used by a speaker. They reject the hypothesis that a speaker reacts to the sound level of her or his own voice. Speakers rather use an internal "private" loop, based on articulatory processes but also on the tension of muscles within the body, plus an external "public" loop, which is based on the response a speaker gets from the listener about the intelligibility of her or his speech. However, there is hardly any information in the paper of the building acoustics parameters that influence the vocal output as, for instance, the amount of absorption or the reverberation time. Also the distance between talker and listener, which has a strong influence on the speech level, is treated in a few sentences only, while it may be an important part of the public loop as well.

In the same year, 1971, the results of measurements by Gardner were published.[3] Gardner measured sound levels as a function of the number of people present in dining rooms and auditoria, but he also measured the absorbing areas of the rooms. He also found an increase in vocal output of approximately 0.5 dB(A) per dB(A) increase in background sound level caused by the other talkers in the room.

In 1977, Pearsons *et al.* summarized different results from measurements at different noise types. Their modal Lombard slope was somewhat higher and given as 0.6 dB/dB.[4,5] Recently, Hodgson *et al.* have published results based on measurements taken in ten eating establishments with different acoustic characteritics.[6] The values differed from 0.40 to 2.61 dB/dB but the type of background noise was very varied, since, for instance, loud café music was also taken into account.

Sound levels in a room are affected by the total absorbing area in that room. According to the principles of acoustic theory a decrease of approximately 3 to 4 dB is found if the total absorbing area is doubled when the output power of the sound source is kept constant. In a multitalker environment, however, the output powers of the human sources depend on the sound level, so a self-reinforcing effect occurs and higher decreases are found. The Lombard effect as a function of the absorbing area has not been thoroughly investigated in the literature. Examples may be found from consulting practice of "before and after measurements" where absorption is added to improve reverberant situations. Oberdörster and Tiesler, for instance, compared two similar rooms with and without a sound-absorbing ceiling.[7]

From these findings acoustic consultants were able to derive their own rule of thumb: The sound pressure level (SPL) of multitalker speech increases by 6 dB when the re-

[a]Author to whom correspondence should be addressed. Electronic mail: l.nijs@tudelft.nl.
[b]Electronic mail: k.saher@tudelft.nl.
[c]Electronic mail: d.denouden@tudelft.nl.

verberation time (RT) is doubled. Hodgson *et al.* have developed a more general formula for the vocal output of teachers by comparing university classrooms with different absorption, and giving talker sound power as a function of the logarithm of the total absorption in square meters.[8,9] The slope in their curve (see Sec. II D for more details) is given as −9.6, which means that the Lombard effect is close to 0.5 dB/dB.

We do not know of any investigations in which vocal output is measured for a series of absorption coefficients. This lack of findings from measurements in different absorption conditions is probably due to difficulties in changing the absorbing area. We were fortunate to be able to take measurements in a test facility specially designed for the new building of the Conservatory of Amsterdam. It was designed to facilitate the study of "ideal" reverberation times for music teaching and practice. The reverberation time could be varied between 1.36 and 0.21 s. This project has been and will be discussed in other articles.[10] It is the aim of this paper to present sound power levels as a function of the amount of absorption and the number of talkers in a room. From these results a simple equation will be derived that can be used in the architectural design process.

The reason we carried out the measurements is that we try to develop architectural guidelines for architects on acoustical quality in institutions for people with intellectual disabilities. It is part of research at the Faculty of Architecture at Delft University. Institutions often comprise groups of eight to ten residents and there are frequent occasions in which a "multitalker situation" occurs if two or more people talk simultaneously. The combination of room shape and sound absorption in particular is helpful in reducing sound levels,[11] but the present paper focuses on one aspect alone: The sound level of human vocal output in relation to the amount of sound absorbing material in a rectangular room in which two or more people are talking simultaneously. Both the absorption coefficient and the number of talkers in a room are taken into account. Although the number of talkers in institutions is mainly from one to four, a simple design equation will be developed and compared to earlier results from literature where the number of talkers may be as high as 100. Vocal output will be restricted to so-called "normal" conversation, which is often below the levels of speech, and sometimes screaming, found in institutions for individuals with intellectual disabilities.

II. THEORY

A. Reverberation time and sound pressure level

The RT used throughout this paper follows Sabine's equation:

$$\mathrm{RT} = \frac{55.3V}{cA} = \frac{0.16V}{A}, \tag{1}$$

where V is the total volume of the room and c the speed of sound. If c is taken as 343 m/s the value 0.16 emerges. The influence of air absorption is omitted throughout the remainder of this paper. The surface area A represents the total sound absorbing area in the room. It is found as the sum over all surfaces in the room:

$$A = \sum_i \alpha_i S_i. \tag{2}$$

For the total room the average value of the absorption coefficient can be calculated as follows:

$$S_{\mathrm{tot}} = \sum_i S_i, \tag{3}$$

$$\alpha_{\mathrm{mean}} = \frac{A}{S_{\mathrm{tot}}}. \tag{4}$$

If a room is still in its design stage, Eqs. (2)–(4) are used in that order. Once a room is finished, it is very difficult to distinguish the absorption coefficients of each surface and α_{mean} is calculated backwards from the RT, V, and S_{tot} measurements:

$$\alpha_{\mathrm{mean}} = \frac{0.16V}{\mathrm{RT}S_{\mathrm{tot}}}. \tag{5a}$$

The remainder of this paper will use a slightly different definition of S_{tot}. In an empty room, all absorption is from the ceiling, floor, and walls only. When furniture and people are added to the room, values A and S_{tot} will increase and V may decrease. In practice, the differences are slight and we will use the following *definition* of the mean absorption coefficient, denoted by α, throughout this paper:

$$\alpha \equiv \frac{0.16V}{\mathrm{RT}S}, \tag{5b}$$

where V and S are calculated for the empty room, but RT includes the influence of furniture and people if present during the measurements.

The calculation of the SPL from a source in a room is introduced as

$$\mathrm{SPL} = L_W + 10\log(H), \tag{6}$$

where L_W represents the sound power output of the sound source. The room and source characteristics are denoted by the variable H. In most acoustics textbooks (see, for instance, Pierce [12]), H is formulated as

$$H = \frac{Q}{4\pi r^2} + \frac{4(1-\alpha)}{A}. \tag{7}$$

The first term is for the direct sound between a speaker and a listener and is dependent on the distance between the two, given by r, and a factor Q, which represents the directivity of the source. Q is a number that is sometimes expressed as the directivity factor (DI) in decibels, defined as

$$\mathrm{DI} = 10\log(Q). \tag{8}$$

Hodgson *et al.* use $Q=2.0$, or DI=3.0 dB, in front of the mouth in their article.[6] In the $\mathrm{AL}_{\mathrm{cons}}$ measuring method a value of $Q=2.5$, or DI=4.0 dB, is generally used.[13]

In most practical situations the first term in Eq. (7) is greater than the second term if r is smaller than about 1 m. If the second term is much greater than the first, at larger distances from the source, the receiver is in the so-called reverberant field. In this case, H reduces to H_{dif}, defined as follows:

$$H_{\mathrm{dif}} = \frac{4(1-\alpha)}{A}. \tag{9}$$

B. Multiple talkers

The variables SPL and L_W in Eq. (6) are logarithmic values that can be written as

$$\mathrm{SPL} = 10 \log\left(\frac{p^2}{p_0^2}\right), \tag{10}$$

$$L_W = 10 \log\left(\frac{W}{W_0}\right), \tag{11}$$

where p is the sound pressure, W is the total sound power of the source, and p_0 and W_0 are the reference variables taken as 20 μPa and 1 pW, respectively.

According to common acoustic theory, the sound powers of multiple sources in a room can be summed to find the total sound power. This is not always easy since some sources (kitchen noise in dining rooms, for instance) have an impulsive nature but these problems are considered beyond the scope of the present paper. It is assumed that all sound is produced by human speech which is more or less constant.

The representation of the total sound pressure in a room from N talkers can be found by rewriting Eqs. (6) and (7) and introducing the sum over all talkers:

$$\frac{p^2}{p_0^2} = \sum_{i=1}^{N}\left(\frac{W_i}{W_0}\left(\frac{Q_i}{4\pi r_i^2} + \frac{4(1-\alpha)}{A}\right)\right). \tag{12}$$

If all talkers are in the reverberant field, Eq. (12) reduces to

$$\frac{p^2}{p_0^2} = \sum_{i=1}^{N}\left(\frac{W_i}{W_0}\frac{4(1-\alpha)}{A}\right). \tag{13}$$

After the introduction of a mean value W_{mean}, Eq. (13) can be written as follows:

$$\frac{p^2}{p_0^2} = \frac{NW_{\mathrm{mean}}}{W_0}\frac{4(1-\alpha)}{A} = \frac{NW_{\mathrm{mean}}}{W_0}H_{\mathrm{dif}}. \tag{14}$$

A new variable H_m is now introduced:

$$H_m = \sum_{i=1}^{N}\left(\frac{W_i}{NW_{\mathrm{mean}}}\frac{Q_i}{4\pi r_i^2}\right) + \frac{4(1-\alpha)}{A}, \tag{15}$$

and hence the total SPL can be written as

$$\mathrm{SPL} = L_{W,\mathrm{mean}} + 10 \log(NH_m), \tag{16}$$

where

$$L_{W,\mathrm{mean}} = 10 \log\left(\frac{W_{\mathrm{mean}}}{W_0}\right). \tag{17}$$

It is in fact the main purpose of the present paper to estimate the value of $L_{W,\mathrm{mean}}$ as a function of the number of talkers and the acoustic configuration in a room.

C. H_m vs H_{dif}

In reverberant rooms it can be expected that the second term in Eq. (15) will be greater than the first. In absorbent situations with many sound sources, however, the first term cannot be neglected. Talkers close to the listener in particular can be heard separately. It is possible to make an estimation of the near field effect with a few assumptions. The first assumption is that all talkers have the same sound power, so $W_i = W_{\mathrm{mean}}$. A second assumption is that all talkers speak in random directions and that $Q=1$. The third assumption is that all sound power is evenly distributed in circles around the receiver. An integral equation can be formulated in terms of the distance to the listener and a solution is easily found if the boundaries are circular, which means that the floor space of a rectangular room is translated into a circular floor with the same amount of square meters. So the floor space of a room S_{floor} is represented by $\pi R_{\max}^2$.

In terms of Eq. (12) we find for the first term:

$$\frac{p^2}{p_0^2}W_0 = \frac{NW_{\mathrm{mean}}}{\pi R_{\max}^2}\int_{R_{\min}}^{R_{\max}}\frac{2\pi r dr}{4\pi r^2}. \tag{18a}$$

Solving the integral yields

$$\frac{p^2}{p_0^2}W_0 = \frac{NW_{\mathrm{mean}}}{2\pi R_{\max}^2}\ln\left(\frac{R_{\max}}{R_{\min}}\right). \tag{18b}$$

The average floor space taken by N_p persons is given by

$$N_p \pi \overline{R^2} = \pi R_{\max}^2. \tag{19}$$

The mean value of R is also the best estimate for the minimum value $R_{\min}$. In most cases N_p will be the number of talkers. However, we will also produce measurement results where a target source plus a listener is surrounded by noise-talkers. If the target speaker remains silent the floor area is divided into $N+1$ sections and $N_p = N+1$.

With $N-1$ noise sources, we find for Eq. (18b):

$$\frac{p^2}{p_0^2}W_0 = \frac{(N-1)W_{\mathrm{mean}}}{2S_{\mathrm{floor}}}\ln(\sqrt{N_p}) = \frac{(N-1)W_{\mathrm{mean}}}{4S_{\mathrm{floor}}}\ln(N_p). \tag{20}$$

Numerical verification with noise sources equally spaced in a rectangular room justifies the use of this term.

The calculation of the reverberant part of the noise in a room can be expressed as follows:

$$\frac{p^2}{p_0^2}W_0 = \frac{(N-1)W_{\mathrm{mean}}}{(2S_{\mathrm{floor}} + S_{\mathrm{walls}})}\frac{4(1-\alpha)}{\alpha}. \tag{21}$$

And now H_m can be derived as

$$H_m = \frac{1}{4S_{\mathrm{floor}}}\ln(N_p) + \frac{1}{(2S_{\mathrm{floor}} + S_{\mathrm{walls}})}\frac{4(1-\alpha)}{\alpha}. \tag{22}$$

To show the influence of the contribution of the direct sound represented by the first term, an example is given in Fig. 1 for a rectangular room as a function of the absorption coefficient. For this example the sound power level L_W is taken as a constant and rather arbitrarily as 70 dB.

As Fig. 1 shows, the addition of the direct sound to the reverberant sound plays a role for high values of the absorp-

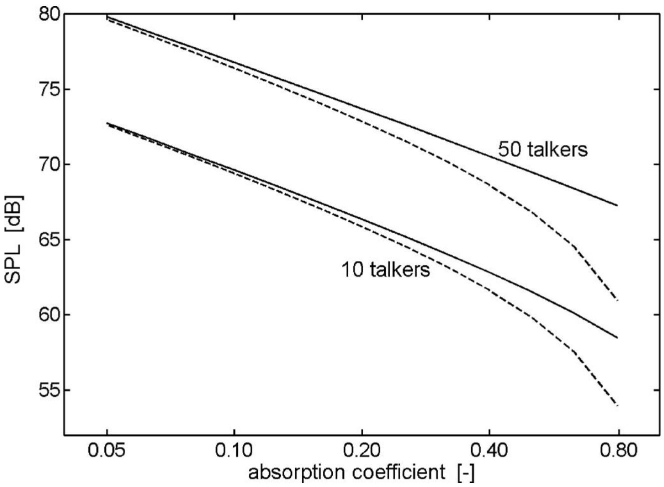

FIG. 1. Sound pressure levels for a $12 \times 10 \times 4$ m^3 room with 10 and 50 talkers. Full curves are calculated with both terms of Eq. (22); dashed lines represent the second term only. The value of L_w is taken constant as 70 dB. Absorption coefficients are taken as equal for all surfaces; they are given along a logarithmic axis.

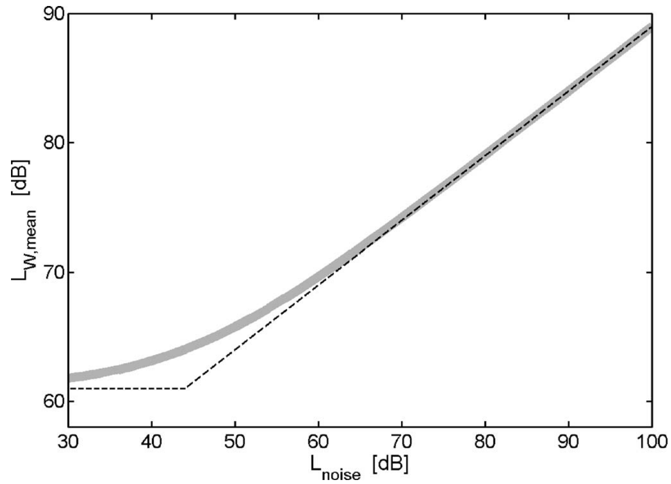

FIG. 2. An example of human vocal output as a function of background noise. The dashed curve is used in older literature. The shape of the full line represents our hypothesis for the remainder of the paper. In this example $C=61$, $D=39$, and $E=0.5$, but the actual values still have to be established.

tion coefficient or for high values of the number of talkers. In this example, 10 talkers stands for one talker per 12 m^2 floor space; 50 talkers (at least 100 people present) on 120 m^2 floor space represents a crowded cocktail party.

The absorption coefficient is given along a logarithmic horizontal axis. When only the denominator in H_{dif} is considered, a straight line will be found as a function of $\log(\alpha)$. The decrease is 3 dB per doubling of the absorption. The numerator causes a steeper descent for the higher absorption values, but the contributions of the direct sound turn the curve almost back into a straight line.

D. A model for vocal output

As a hypothesis for further investigations a model was developed at the start of the measurements. It is expressed as a simple logarithmic sum of two terms:

$$L_{W,\mathrm{mean}} = 10 \log(10^{C/10} + 10^{(D+EL_{\mathrm{noise}})/10}), \tag{23}$$

where C, D, and E are the three values under investigation. The model is shown in Fig. 2, where two asymptotic lines are also given: $L_{W,\mathrm{mean}}=C$ for low noise levels and $L_{W,\mathrm{mean}} = D + E \times L_{\mathrm{noise}}$ for higher values (dashed lines).

Equation (23) is based on previous models (see for instance Van Heusden *et al.*[14]) in which these two straight lines are used separately. For low noise levels the signal-to-noise ratio between speech and background noise is assumed sufficient and talkers do not raise their voices at all. From one specific noise level (for instance, 44 dB in Fig. 2) talkers raise their vocal output linearly. This part is the Lombard effect with slopes mainly between 0.3 and 0.6 dB/dB.

A discontinuous curve like that in Fig. 2 is based on the assumption that for noise levels below approximately 40 dB people talk at a level which is common for an anechoic chamber. It is our hypothesis that people in these situations talk somewhat louder and therefore we propose the full curve in Fig. 2 instead. A first indication of this level increase can already be found in the results by Van Heusden *et al.*[14]

When the proposed curve is used, different Lombard slopes are found. In Fig. 2 the slope is 0.6 dB/dB for noise levels above 80 dB when $E=0.6$. When noise levels are approximately 60 dB, the slope is only of the order of 0.2 to 0.3.

Recently, after we performed our measurements and curve fitting, Hodgson *et al.* have proposed a model for the total sound pressure level in an eating establishment.[6] When this model is somewhat adapted it may be used as an alternative for Eq. (23):

$$L_{W,\mathrm{mean}} = C + \frac{\mathrm{asym}}{\left\{1 + \exp\left(\frac{L_{\mathrm{mid}} - L_{\mathrm{noise}}}{\mathrm{scale}}\right)\right\}}. \tag{24}$$

The curve is shown as a dashed line in Fig. 3. The maximum slope is found when $L_{\mathrm{noise}}=L_{\mathrm{mid}}$. The slope itself equals $\mathrm{asym}/(4 \times \mathrm{scale})$ dB/dB.

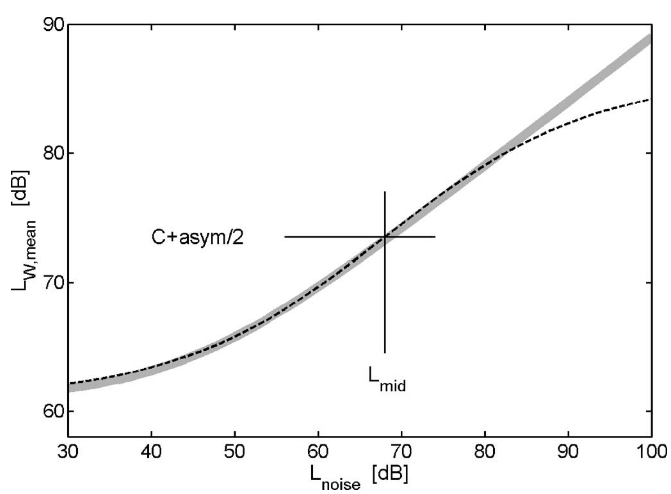

FIG. 3. Two models of human vocal output as a function of background noise. The full line is from Eq. (23); the dashed line is from Eq. (24) proposed by Hodgson *et al.* In this example $C=61$, $D=39$, and $E=0.5$ for Eq. (23); for Eq. (24) $\mathrm{asym}=25$, $L_{\mathrm{mid}}=68$, and $\mathrm{scale}=12.5$ are used. The maximum slope is 0.5 dB/dB.

The curve by Hodgson *et al.* is of course more accurate at high noise levels. There is an upper limit to the vocal output of humans which is not catered for in Eq. (23). At lower noise levels differences are slight. In the example shown in Fig. 3, differences are smaller than 0.2 dB when noise levels are below 80 dB. It is not expected that noise levels from human speech will be higher in the present investigations. In this case the present model is easier to use since it requires only three parameters.

It should be noted that Hodgson *et al.* used their equation for the sound *pressure* level, while Eq. (24) is for the sound *power* output. This has no effect on the shape of the curve. The curve is only shifted vertically by 11–DI dB.

E. Feedback from the room's SPL on the vocal output

If the noise in a room is generated by other talkers only, $L_{W,\mathrm{mean}}$ in Eqs. (16) and (17) is dependent on the noise level from $N-1$ talkers. The total sound pressure level in a room, denoted by SPL, from all N speakers is as follows:

$$\mathrm{SPL} = L_{W,\mathrm{mean}}(L_{\mathrm{noise}}) + 10\log(NH_m). \tag{25}$$

In Eq. (25), SPL and L_{noise} are mutually dependent. Hence the equation must be solved recursively. In general, recursive numerical methods do not form part of the toolbox of the average architect. But there is also a mathematical problem. A curve fitting process is a recursive process as well, and hence a second recursive method is used simultaneously to derive C, D, and E from measurements.

The recursive method to find SPL can be avoided if $E=0.5$. This method was derived after the measurements were carried out and the background will be explained at a later stage. The method itself is described in the present section.

If Eqs. (10), (11), (14), and (15) are slightly rewritten, the sound power from the noise can be written as

$$\frac{p_{\mathrm{noise}}^2}{p_0^2} = \frac{W(p_{\mathrm{noise}}^2)}{W_0}(N-1)H_m, \tag{26}$$

which, in combination with Eq. (23), can be written as

$$\frac{p_{\mathrm{noise}}^2}{p_0^2} = (10^{C/10} + 10^{(D+EL_{\mathrm{noise}})/10})(N-1)H_m, \tag{27}$$

which is

$$\frac{p_{\mathrm{noise}}^2}{p_0^2} = \left(10^{C/10} + 10^{D/10}\left(\frac{p_{\mathrm{noise}}^2}{p_0^2}\right)^E\right)(N-1)H_m. \tag{28}$$

When $E=0.5$, Eq. (28) can be solved as a quadratic equation with the solution:

$$\frac{p_{\mathrm{noise}}}{p_0} = 10^{D/10}(N-1)H_m \times\left(0.5 + 0.5\sqrt{1 + \frac{4\times 10^{(C-2D)/10}}{(N-1)H_m}}\right), \tag{29}$$

and hence:

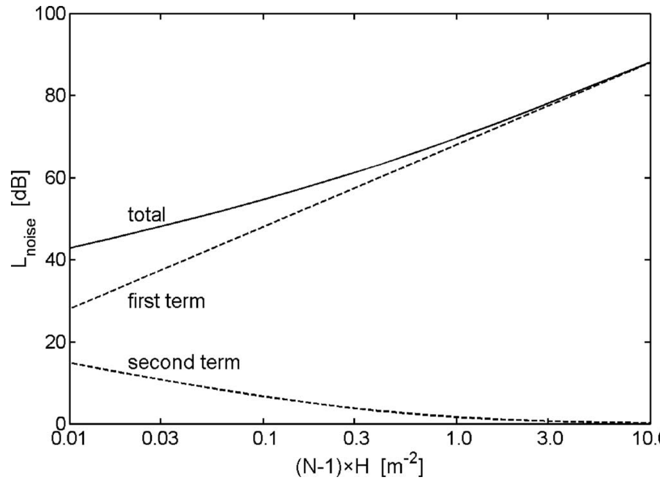

FIG. 4. L_{noise} calculated with the aid of Eq. (30) (full line) to show the difference between the first and the second term (dashed lines).

$$L_{\mathrm{noise}} = 20\log(10^{D/10}(N-1)H_m) + 20\log\left(0.5 + 0.5\sqrt{1 + \frac{4\times 10^{(C-2D)/10}}{(N-1)H_m}}\right). \tag{30}$$

Figure 4 illustrates the contributions of the first and second terms to the total noise level. If the human vocal output were not affected by noise, a straight line would be found with an increase of 3 dB per doubling of $(N-1)H_m$. Equation (30) results in a higher slope. The first term causes an increase of 6 dB per doubling of $(N-1)H_m$; the addition of the second term reduces the slope. If $(N-1)H_m$ is greater than about 2 this second term no longer has any influence and a 6 dB increase is found.

An example will serve to clarify the values of N and H_m. A rectangular room measuring $9\times 6\times 3$ m^3 has a total area of 198 m^2; the floor space is 54 m^2. So, if $\alpha=0.15$, $H_m=0.11$. When the absorption coefficient equals 0.30, H_m becomes 0.047.

If $(N-1)H_m > 2$ the contribution of the second term in Eq. (30) is less than 1 dB and the slope of the total curve varies with $20\log\{(N-1)H_m\}$. This value is found if the number of talkers is 18 or 42 for 15% and 30% absorption, respectively. The first case is possible on 54 m^2 floor space, but the second case should be considered as a very crowded cocktail party.

When the room is scaled up by a factor of 2 to $18\times 12\times 6$ m^3, the results are the same if the number of talkers is increased to 72 and 164, respectively. So the most important factor is in fact not the number of talkers, but the number of talkers per floor space.

III. MEASUREMENTS

A. Preliminary measurements

During the measurement phase participants were asked to read out excerpts from magazines at "normal conversational level as if they were talking to a listener at 1 m distance." The sound levels when reading aloud are somewhat higher than those for conversation; differences of about 2 dB have been reported.[14] To investigate this effect, preliminary

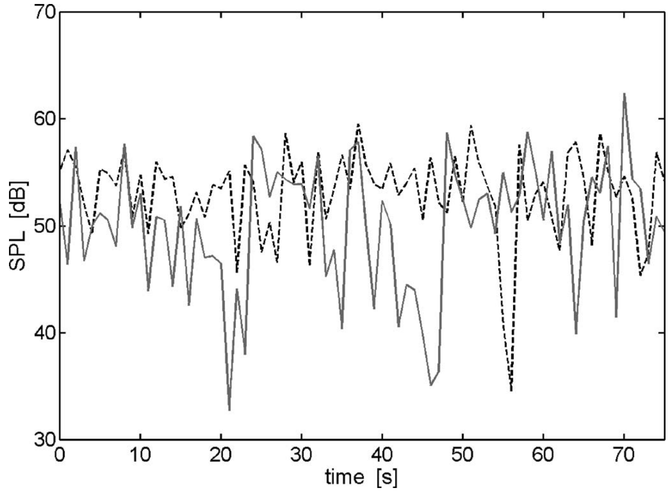

FIG. 5. SPL at 1 m in front of the head of a male test participant, recorded in an anechoic chamber, for reading (dashed curve) and normal conversation (full curve).

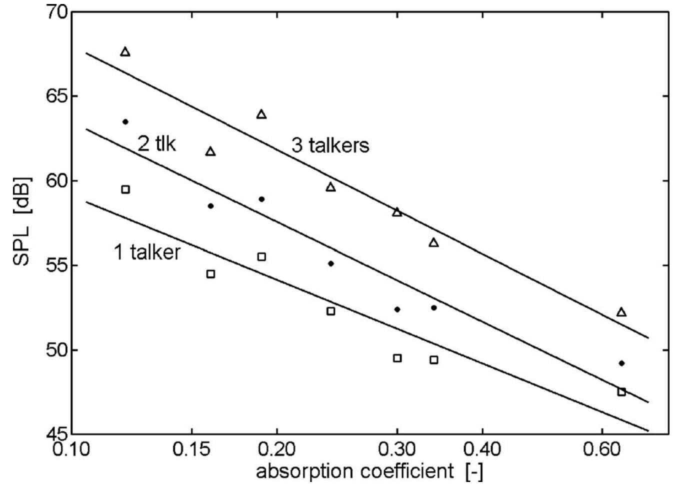

FIG. 6. SPL values from one, two, or three talkers in the reverberant field. Full lines represent best fit curves; the slopes are $-20.4\log(\alpha)$, $-19.6\log(\alpha)$, and $-16.4\log(\alpha)$ from top to bottom.

measurements were carried out in an anechoic chamber. These were also used to measure the lowest noise levels.

Six participants were asked to read excerpts from books. Equivalent sound levels were measured from 19 texts lasting approximately 2 min. The microphone was placed 1 m from participants' mouths. The measured mean A-weighted sound pressure level of the test participants was 54.6 dB, with a minimum level of 52.0 dB and a maximum level of 57.4 dB.

Figure 5 shows a typical example of the difference between reading and talking for one male test participant. The sound level of the book excerpt is fairly constant. For conversational speech the talker starts at the same level as when reading but decreases his output during the sentence, which can be observed between 15 and 22 s and between 38 and 48 s. Since the maximum levels are the same, the equivalent sound level is lower. The difference of 2 dB in equivalent sound levels agrees with that found in the literature.[14]

These findings mean that a correction is needed from reading to conversation. However, although the effect was not investigated, in the authors' opinion correction is only required at low noise levels: in noisy places talkers cannot afford to decrease their sound level during sentences, so the reading values are useful for representing conversational speech in the noisier cases without any adjustment. The difference in sound levels between participants in the anechoic chamber was up to 5.4 dB.

Measurements at 1 m in front of the mouth for 20 test persons were also done in an office room. Background noise levels varied from 25 to 80 dB. The results confirmed the Lombard slopes found in literature and are omitted here. However, the spread in the group of 20 participants is worthwhile noting. The difference between the softest and loudest voice was 12 dB when the noise level was below 30 dB. One would expect that this variation would decrease at higher noise levels as people are forced to adapt their voices to the noise. This appeared only partly the case, as differences of 9 dB are still found at a noise level of 70 dB.

B. Room with variable absorption

Speech recordings were made in a purpose-built test facility at the Conservatory of Amsterdam. The dimensions of the room were $6.4\times4.8\times3.5$ m^3, with a volume of 108 m^3 and a total surface area of 140 m^2. During the measurement phase the room contained a grand piano, a table, a few chairs, and six people.

The architect of the new Conservatory had designed special absorbers to be used in the new building, measuring 0.90×0.90 m^2 and 0.90×1.80 m^2. There were 18 large absorbers and 36 small absorbers. Special effort was invested into obtaining a flat reverberation curve as a function of frequency. Reverberation times were measured in octave bands; a mean value was calculated over the 500, 1000, and 2000 Hz bands. The mean absorption coefficients were calculated according to Eq. (5b).

One participant read a text sitting at a table. Two listeners were positioned opposite the first participant with their ears 1 m away. The noise was produced by one, two, or three real talkers at a greater distance (3 m or more) from the speaker–listener combination. They were not completely in the reverberant field of the room since the first term of Eq. (22) has some influence for the highest absorption coefficient.

Test participants were asked to read excerpts from magazines. In a test run, the main speaker at the table was asked to start reading. After approximately 30 s the noise speaker(s) started to read. After about 2 min the main speaker stopped, but the noise speaker(s) continued for a further 30 s. This method provides us not only with noise from the main speaker at 1 m, but also noise from one to three noise speakers without the main speaker.

Three runs were undertaken per absorption situation and per noise speaker. One male speaker at the table read twice; the third run was read by a female speaker.

Figure 6 shows SPLs measured at the listener's position from one, two, or three noise speakers when the target speaker remains silent as a function of the mean absorption in the room. There were seven values of the absorption co-

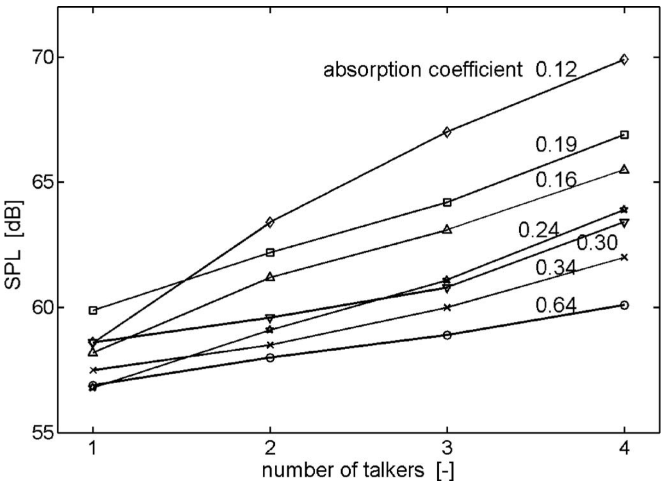

FIG. 7. SPL values at 1 m in front of the talker at the table (number of talkers=1). The total SPL is found if one, two, or three noise talkers in the reverberant field are added.

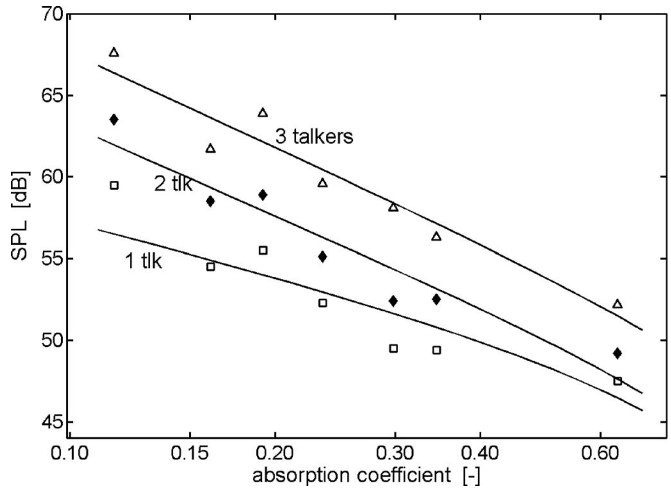

FIG. 8. Calculations of SPL for one to three talkers in the reverberant field, to be compared with Fig. 6. Triangles are for three talkers, diamonds for two talkers, and squares for one talker. See the text for C and D values.

efficient. The first situation had $\alpha=0.19$. In the second situation absorbing panels were removed and $\alpha=0.12$. Then the number of panels was increased to find $\alpha=0.16$, 0.24, 0.30, 0.34, and 0.64 in the following five situations.

The first situation, where $\alpha=0.19$, has the same amount of absorbers as the third situation, where $\alpha=0.16$. In the latter case all absorption is on the ceiling; in the first case the absorption is randomly distributed. Ray-tracing models predict the decrease of absorption as measured here.[15] However, they also predict that the SPL will stay almost constant, so the somewhat lower SPLs at $\alpha=0.16$ are not explained. More measurements are necessary to investigate this effect.

The values of NH_m are between 0.02 and 0.77, so the value of $NH_m=2$, where only the first term from Eq. (30) remains, is never reached and the second term in Eq. (30) and Fig. 1 always has an influence.

Figure 7 gives the SPL values at 1 m from the source speaker, now including the target speaker. The values are averaged over two readers in three sessions, since one reader did the test twice. Values are given as a function of the total number of speakers. So, for instance, four talkers means one target speaker plus three noise talkers.

In Fig. 7 the same increase in SPL can be observed as in Fig. 6 when the curves with $\alpha=0.16$ and 0.19 (and equal number of absorbers) are compared.

IV. CURVE FITTING

A. The curve-fitting process

In Sec. II D the background of our output curve for vocal effort was explained. Equation (23) gave the curve as a function of three variables C, D, and E. In the following, these variables will be estimated by curve fitting.

We did not find a statistical method for determining the three values simultaneously. Curve fitting methods are recursive methods, but Eq. (25) should be solved recursively as well, since SPL is on the left-hand side and the right-hand side of the equation as the noise speakers raise their voices as well. In this case a fitting process has to solve two recursive processes simultaneously.

When the noise level is kept constant, the recursive effect disappears. In this case it is possible to apply curve fitting to the data. We fitted our data using the MATLAB function lsqcurve. Babble and noise measurements were done in an office room in the early stages of our research. A comparison was also made with measurements by van Heusden *et al.*[14] Calculated E values ranged from 0.42 to 0.56, but a comparison of curves showed that if a curve was chosen with $E=0.5$, the error value produced by the fitting process increased only marginally. Differences between the best fit curves and the curves with $E=0.5$ were within 1 dB, which is well below the variation found in the measuring results. Therefore we decided to use $E=0.5$ as a given variable throughout the rest of the investigations. Equation (30) could then be applied and the recursive process for the calculation of SPL was no longer necessary, so MATLAB's lsqcurve could easily find the two remaining values for C and D.

A second problem lay in finding values for the power output (L_W) from measurements of SPL; if the microphone is close to the talker, as in Eqs. (6) and (7) and in Fig. 7, the directivity factor Q from the "target" speaker must be estimated as well.

The results from the curve fitting process will show that the value $Q=2.5$ as given in Sec. II is too low; a Q value between 4 and 5 is a better estimation, so the directivity index (DI) increases from DI=4 dB to DI=7 dB. A 2 dB increase is probably caused by the table top between the target speaker and the receiver. We measured similar differences with a loudspeaker.

B. Fitting with measurements from the Amsterdam Conservatory

The next comparison is between measurements taken at the Amsterdam Conservatory, originally given in Fig. 6, and calculations from the model, for three talkers in the reverberant field. Figure 8 shows the results of the fitting process.

Curve fitting through the measuring points for three talkers yields $C=58.3$, $D=35.2$. For the points measured with two talkers these values are almost the same: $C=57.7$, D

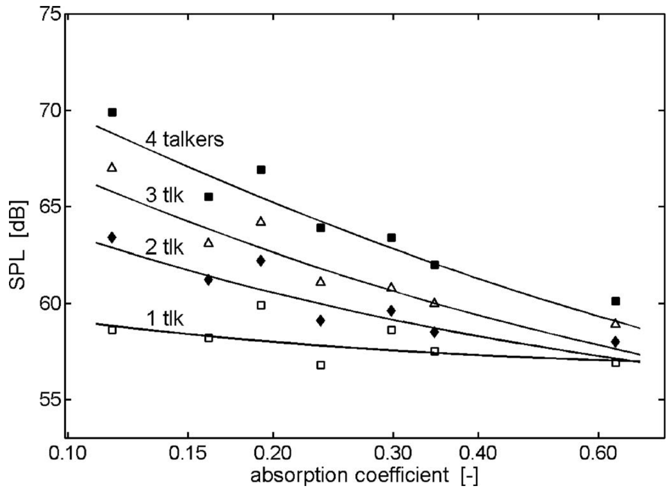

FIG. 9. Calculations of SPL at 1 m from a talker with zero, one, two, or three noise speakers in the reverberant field. The measurements from one, two, three, and four talkers are represented by open squares, closed diamonds, open triangles, and closed squares. See the text for C, D, and Q values.

=35.1. However, when there is only one talker, the noise level is very low and constant so the value of D has no meaning. The C value found in this case is 5 dB higher: C =63.0 dB. The C values for one speaker are higher than for the multitalker cases. One reason may be that our three single talkers have a voice that is louder than the mean value of three or four talkers. However, there may be another explanation. When people listen to talkers, reverberation can be regarded as noise. Common measures like speech transmission index (STI) and U_{50} for speech intelligibility are based on this principle.[16–18] Our hypothesis is that talkers react to their own sound accordingly. A simple addition to the model can be made by adding the late energy of the talker after 50 ms, as done in U_{50}. If, for instance, the late energy after 50 ms is calculated or measured as 60% of the total sound energy, a 0.6 noise speaker is added to the total number of noise speakers. In fact, it appeared possible to fit the single talker into the multitalker model. However, more measurements are required in order to check our hypothesis.

Figure 9 shows results for comparison with the readings of Fig. 7, although the results are now given with the absorption along the horizontal axis. This time the target speaker is also included: "one talker" is when he or she is the only one speaking, while two, three, and four talkers means one, two, and three noise speakers, respectively. Again, for one talker the D value has no meaning. The curve fitting process yields C=61.1 and Q=4.8. For two speakers the combination of C, D, and Q is (60.4, 32.9, 4.3), for three speakers (59.2, 33.5, 4.7), and for four speakers (59.6, 34, 4.6).

C. Gardner's results (1971)

Research undertaken at our university focuses on small numbers of talkers. It is interesting, however, to compare the model with speech from larger numbers of talkers. Gardner's results are the most appropriate, since he explicitly used multitalker backgrounds and he removed background noise from other sources.[3] More recent findings, such as those of Hodgson *et al.*, cannot be used in this instance as they include, for example, noise from radios in cafés.

Gardner investigated the Lombard effect as a function of N for some auditoria and dining rooms. In this section our intention is to estimate our C and D values from the results as given by Gardner. It is assumed that E=0.5, so all calculations are done using Eqs. (16), (22), and (30).

As an input, values are required for N, α plus the geometrical values of the rooms. N is naturally very hard to measure, so Gardner took the number of people present in the room and not the actual number of talkers. To fit our model, the value of H is needed for a specific room. This value must be derived from the measurement of the reverberation time plus the geometrical variables or at least the total area S. It is possible to calculate H_{dif} for three of Gardner's cases, since H_{dif} is very similar to the R value given by Gardner (in square feet). It is defined as

$$R = \frac{\alpha S}{(1-\alpha)} = \frac{0.05V}{\text{RT}(1-\alpha)}. \tag{31}$$

For our estimation the R value itself is not enough since we require α and the area S. Gardner gives results for a 341-seat auditorium in his Fig. 5. The R value is given as 20 000 ft^2; no dimensions are given. If α is calculated from this value by assuming the dimensions that belong to this type of auditorium, something strange occurs: either the value of the volume is extremely high, or the value of α is no less than 0.7. By coincidence, there is a 336-seat auditorium in our own faculty building. When we measured that auditorium a much smaller value R=3900 ft^2 was found, or H_{dif} =0.011 m^{-2}. It was then decided to take similar measurements to Gardner's in this auditorium.

The number of individuals present at the entrance of the auditorium was counted, but to estimate the percentage of people actually talking we installed a video camera as well. Figure 10 shows the results plotted over Gardner's findings. The open circles are Gardner's; closed triangles represent our measurements.

As can be seen from Fig. 10, Gardner's measurements and our own are very similar. The percentage of people actually talking when the auditorium was almost full was 25%. Gardner mentions 30%. If this is correct, his results are 1 dB higher. The full line shows the result of curve fitting. The total sound pressure level when 320 people are present is 71 dB. If people do not raise their voices, this level would be only 59 dB, so a 12 dB increase in vocal output is found due to the Lombard effect.

In Figs. 11 and 12 Gardner's Figs. 8 and 9 are replicated with curves from our model. In this instance Gardner gives both geometrical and acoustical values. The R value in Fig. 11 (Fig. 8 in Gardner's article) is again very high, but this time it is correct, since the room has "heavy drapes" and so the absorption coefficient is no less than 0.66. In this room it is important to use both terms in Eq. (30) because this is a typical example of a room where the direct contributions are at least as important as the sound from the reverberant field.

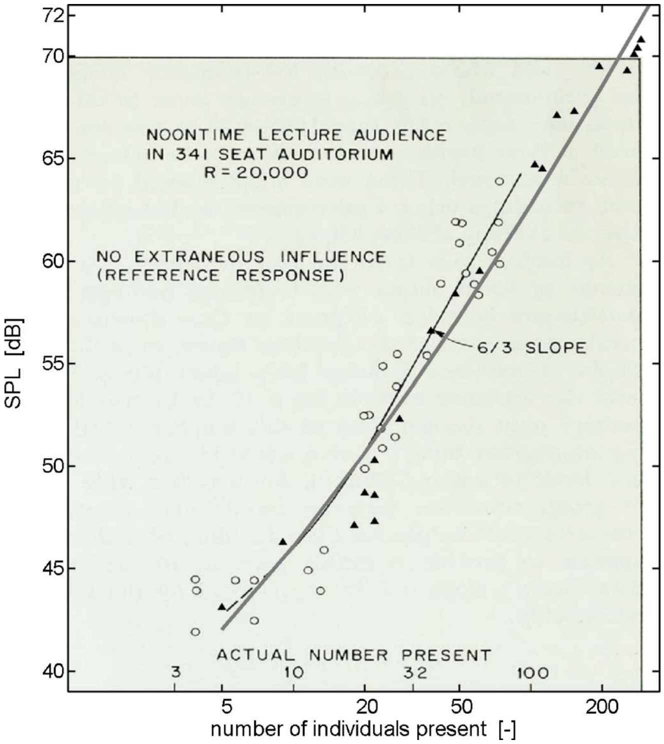

FIG. 10. (Color online) Results from Gardner's auditorium as given in his Fig. 5 (open circles). Dots are from measurements taken in a similar auditorium at our Faculty of Architecture. The full grey line gives an estimation for a rectangular room of $22 \times 15 \times 6$ m^3 where α=0.24. It is calculated using Eq. (30) when C=59 and D=35.5. The percentage of people talking is 25%.

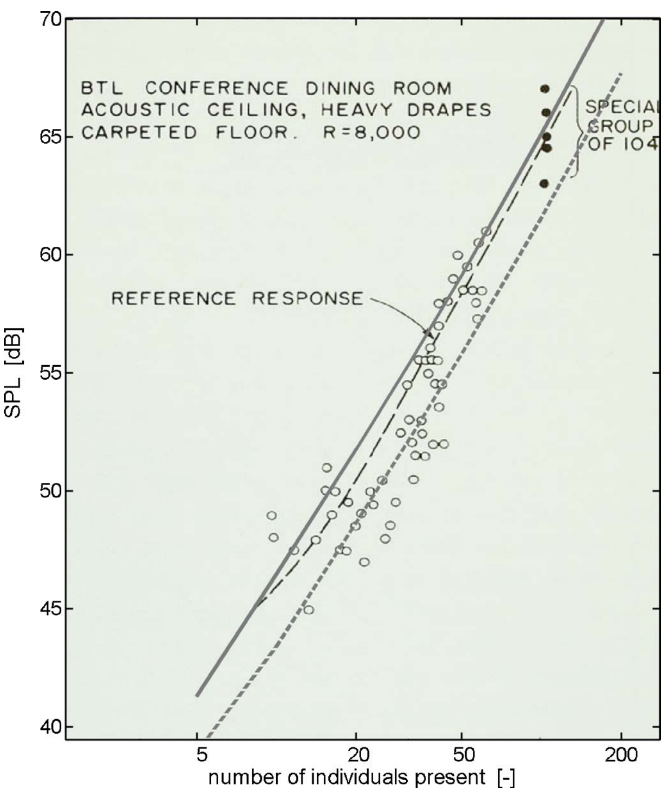

FIG. 11. (Color online) Results from Gardner's dining room as given in his Fig. 8, plus an estimation from our Eq. (30) when C=59 and D=34. The percentage of talkers is 45% for the full line and 30% for the dotted line.

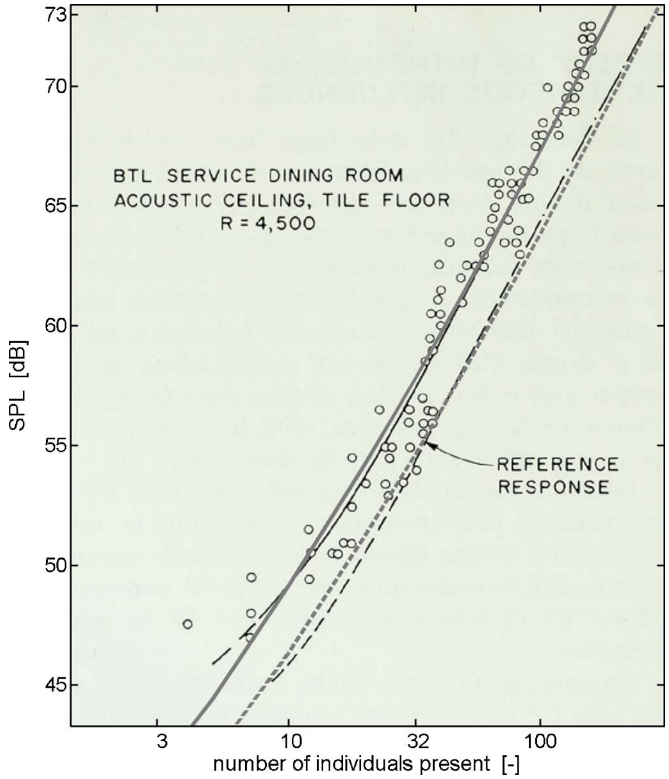

FIG. 12. (Color online) Results from Gardner's dining room as given in his Fig. 9, plus an estimation from our Eq. (30) when C=59 and D=35.5. The percentage of talkers is 45% for the full line and 30% for the dotted line.

In Fig. 12 the absorption coefficient is found as α =0.37, which is still rather a high value. Moderately reverberant rooms were not available in Gardner's article.

The model input was C=59 in both cases. To fit the results with the measurements D should be chosen as 34 in Fig. 11 and as 35.5 in Fig. 12. We will return to this subject in Sec. V.

Gardner pointed out that the best fit through his measuring points in Figs. 11 and 12 should be steeper than the 6/3 slope he derived from Fig. 10, which means a 6 dB increase per doubling of the individuals present. It could be achieved in our model by increasing the E value from 0.5 to about 0.6. This could be viewed as a flaw in our model. However, another explanation could be easily observed on the video tapes we made, but can also be observed in practical situations: The percentage of talkers often increases with the number of people present. In an auditorium people arriving early are "single" and the percentage of talkers may be as low as 20%. When more people enter they join the early attendees and when the auditorium is full the percentage of talkers may be almost doubled. A similar effect can be found at cocktail parties. People often talk in groups of four to eight at the start, but these groups break up when more people enter.[19] This time the reason is acoustic: Bigger groups create bigger talker–listener distances and groups *must* break up at higher noise levels in order to maintain the minimum signal-to-noise ratio, which is found at short talker-listener distances.

In Fig. 10 our four points well below the calculated curve are explained by a lower percentage of talkers. They are measured when 22 to 24 people are in the auditorium.

TABLE I. Overview of C, D, and Q values from five cases plus the values estimated for the calculation model.

Case		Q	C	D
A	Reading in Conservatory, talkers in reverberant field		58	35
B	Reading in Conservatory, talker at 1 m included	4.6	59.5	33.5
C	Gardner's Fig. 5, normal conversation		59	35.5
D	Gardner's Fig. 8, normal conversation		59	34
E	Gardner's Fig. 9, normal conversation		59	35.5
	Estimation for architectural purpose		59	35

The model calculates a sound level of about 50 dB when five or six people are talking (which is 25%). On the videotape it is evident that only three or four people are talking simultaneously, so the decrease in SPL is explained.

Figures 11 and 12 contain two curves from calculations when 30% and 45% of the people present in the room are actually talking; the 30% curve gives the best fit for lower noise levels and the 45% curve is for higher noise levels.

Gardner mentions that he only used normal conversation by cutting out laughter, coughing, and clashing of dishes. We followed the same procedure but this was not easy. As an auditorium fills up, outbreaks of enthusiasm and laughter increase in greater and greater proportions. In our auditorium we also suffered from noise caused by footsteps in the wooden aisles. When the auditorium was full, "normal conversation" could only be measured about one-third of the time. The nonfiltered SPL values measured are about 5 dB higher.

D. Results from Hodgson *et al.* (2007)

In Eq. (24) and Fig. 3 a comparison was made with the curve used by Hodgson *et al.* If we fit this curve to our results when $C=59$ and $D=35$ and when the noise level is kept below 80 dB, the following values are found for an exponential curve: $C=58.7$, asym$=23.4$, $L_{\mathrm{mid}}=69.8$, and scale$=12.2$. The resulting maximum slope is 0.48 dB/dB. The differences between the two curves are 0.3 dB or less.

Hodgson *et al.* give measured SPL values, while our results are for L_W, but they assume that all their sound sources are in the reverberant field of the model listener and an estimation of Q is not required. Hodgson *et al.* give room dimensions plus reverberation times and number of seats, but to calculate SPL values, the number of talkers should be estimated from the number of seats. If that factor is taken as equal to 33%, differences between the results of Hodgson *et al.* and our model range from −3 to +3 dB for their cases C, B, and R. However, comparison is difficult, since Hodgson *et al.* did not restrict the noise levels to other talkers as Gardner and we did, and their variations (Table I) in SPLs are approximately 15–20 dB. They also took loud background music into account, for instance. In the two "senior residences" from Hodgson's article, the results from our model were

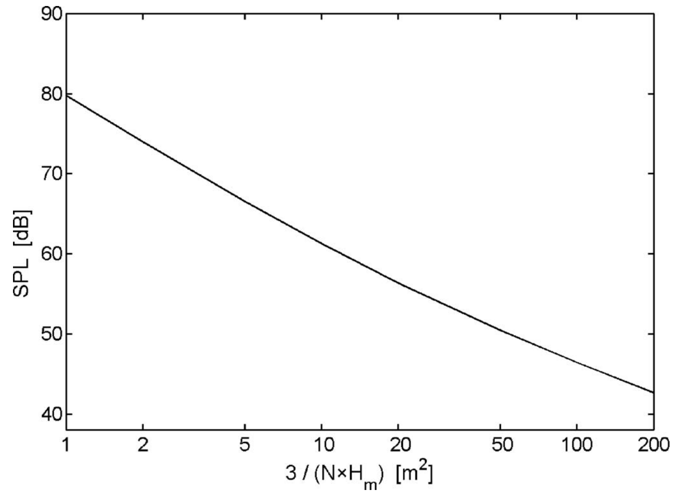

FIG. 13. The sound pressure level in a room calculated using Eq. (30). Values used in the model are $C=59$ and $D=35$. The target speaker at close distance to the listener is not included. The values plotted horizontally are almost equal to the absorbing surface per talker.

about 10 dB too high. It is possible that the number of talkers was less (the discrepancy vanishes if only one out of ten attendees is actually speaking) or the seniors have softer voices. This is an interesting question for further research.

V. SUMMARIZING THE RESULTS

A. Discussion of *C* and *D* values

In order to find the C value and D value to be used in our model, the results are summarized in Table I. The values of C and D on the lowest row of Table I are calculated from the rounded means of cases A–E.

In fact, the C value represents the sound power output of human speech in an anechoic chamber. If $C=59$ and $Q=2.5$, the resulting sound pressure level at 1 m is 52 dB. This level is 2 dB lower than the mean level we measured in the anechoic chamber, but those values were from reading a book. Lower sound pressure levels, even below 50 dB, have been reported in literature.[14,20]

Case D represents a multitalker situation in a highly damped room. Now the first term of Eq. (22), predicting the influence of direct sound from all talkers in the room, is equally important as the second term for the reverberant part. Although this term from Eq. (22) has been extensively compared with numerical models, it might overestimate the direct contributions. On the other hand, an overestimation of the contribution of the reverberant part is also very likely in damped rooms, because diffuse fields with constant levels through the room are unlikely.[11,21]

B. Curve for architectural practice

If C and D are taken as 59 and 35, respectively, Fig. 4 can be redrawn. This is shown in Fig. 13, but the variable along the horizontal axis has changed. The chosen variable $3/(NH_m)$ is equal to A/N, when $\alpha=0.25$. A/N represents the absorbing surface per talker in a room and is an easy design parameter for an architect to use.

De Ruiter[19] presented similar curves, based on measurements (including Gardner's) and on the ISO 9921 standard. The results are of the same order, but a comparison is impossible since the results are given with the number of people present in the room. Unfortunately there is no indication about the percentage of people actually talking.

According to the curve in Fig. 13, it is very unlikely to find sound pressure levels from normal conversation higher than 80 dB, since one talker–listener combination represents 1 m^2 of absorption. To verify this statement, some relatively simple readings were taken with a sound level meter. A higher level was found only once during a very crowded cocktail party. The room was very reverberant, so the absorption was almost completely provided by the people standing on the floor. In this case A/N can be below 1 when people are very close to each other.

The results agree with the results given in Table 1 by Hodgson *et al.*[6] In that table, maximum L_{eq} levels are of the order of 75–79 dB, apart from one bistro with "loud music" where levels of 82 dB have been measured.

VI. CONCLUSION

The model introduced in this paper to predict the Lombard effect is based on a simple curve, with a gradually increasing slope for the vocal sound power as a function of the actual sound pressure level in the room. The model is designed to predict human vocal output for different numbers of talkers in rooms of different sizes and with different sound absorbing properties. In crowded places, outbursts of loud talking, laughter, etc., may increase equivalent sound levels, but these effects are not incorporated in the model since it is restricted to "normal conversation" only. Therefore the model does in fact predict the minimum levels for practical situations and an increase of 5 dB can easily be found.

The model shows an increase in vocal output of 0.5 dB per 1.0 dB increase of the sound pressure level when noise levels are approximately 70–80 dB. When noise levels are around 50 dB the model predicts slopes on the order of 0.2 or 0.3 dB/dB.

The model is fitted with five different sets of measuring results. A best fit is found if the sound power level is estimated as 59 dB when noise is absent. The five cases are very well matched and show differences of only 1 dB for multitalker cases. The inaccuracy in the prediction of the sound power level is about 2 dB for high noise levels, especially in nonreverberant situations, where talkers in the vicinity of the listener can be heard separately. More measurements are needed to decrease this inaccuracy.

The main reason for developing the model was to use it as a tool in the architectural design process. What happens to the sound level if the amount of absorbing area in a room is doubled, for example? The model predicts slopes as high as −6 dB per doubling of absorption if only the reverberant field is taken into account. If the influence of direct sounds is incorporated as well, this slope is lower and a value of −5 dB is found on most occasions. This value is based on the assumption that the percentage of people actually talking remains constant. In noisy conditions this percentage is always close to 50% because groups of three or four people are simply unable to understand each other. Such poor circumstances are improved by adding absorption, so the percentage of talkers may drop as well and slopes may be steeper than −6 dB per doubling of the absorption.

[1]E. Lombard, "Le signe de l'élévation de la voix [Characteristics of the elevation of the voice]," Annales des Maladies de l'Oreille et du Larynx **37**, 101–119 (1911).

[2]H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 667–709 (1971).

[3]M. B. Gardner, "Factors affecting individual and group levels in verbal communication," J. Audio Eng. Soc. **19**, 560–569 (1971).

[4]K. Pearsons, R. L. Bennett, and S. Fidell, "Speech levels in various environments," U.S. Environmental Protection Agency, EPA-600/1–77–025, Washington D.C., 1977.

[5]W. O. Olsen, "Average speech levels and spectra in various speaking/listening conditions, a summary of the Pearson, Bennett, & Fidell (1977) Report," American Journal of Audiology **7**, 21–25 (1998).

[6]M. Hodgson, G. Steininger, and Z. Razavi, "Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments," J. Acoust. Soc. Am. **121**, 2023–2033 (2007).

[7]M. Oberdörster and G. Tiesler, "Acoustic ergonomics in schools," Federal Institute for Occupational Safety and Health, Dortmund, Germany, 2006.

[8]M. R. Hodgson, R. Rempel, and S. M. Kennedy, "Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures," J. Acoust. Soc. Am. **105**, 226–233 (1999).

[9]M. R. Hodgson, "Case-study evaluations of the acoustical designs of renovated classrooms," Appl. Acoust. **65**, 69–89 (2004).

[10]M. E. Valk, P. H. Heringa, and L. Nijs, "Het optimaliseren van de ruimteakoestiek voor de les- en oefenruimtes van het Conservatorium van Amsterdam," [Optimizing the room acoustics for the study and teaching rooms in the Conservatory of Amsterdam], Bouwfysica **17**, 11–15 (2006).

[11]L. Nijs, P. Versteeg, and M. van der Voorden, "The combination of absorbing materials and room shapes to reduce noise levels," 18th International Congress on Acoustics, Kyoto, 2004.

[12]A. D. Pierce, *Acoustics* (Acoustical Society of America, New York, 1985).

[13]J. van der Werff, *Speech Intelligibility, the ALcons Method Based on the Work of Victor Peutz* (Zoetermeer, Peutz, 2004).

[14]E. van Heusden, R. Plomp, and L. C. W. Pols, "Effect of ambient noise on the vocal output nd the preferred listening level of conversational speech," Appl. Acoust. **12**, 31–43 (1979).

[15]L. Nijs, "The distribution of absorption materials in a rectangular room," Internoise Congress on Noise Control Engineering, Rio de Janeiro, Brazil, 2005.

[16]T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077 (1985).

[17]J. S. Bradley, "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80**, 846–854 (1986).

[18]J. S. Bradley and S. R. Bistafa, "Relating speech intelligibility to useful-to-detrimental sound ratios," J. Acoust. Soc. Am. **112**, 27–29 (2002). This article refers to and corrects three earlier publications.

[19]E. Ph. J. de Ruiter, *The Great Canyon, Reclaiming Land from Urban Traffic Impact Zones* (Zoetermeer, Peutz, 2004).

[20]W. T. Chu and A. C. C. Warnock, "Detailed directivity of sound fields around human talkers," National Research Council Canada, IRC-RR-104, 2002.

[21]M. Barron, *Auditorium Acoustics and Architectural Design* (E&FN Spon, London, 1993).