HAL
open science

# Effect of masking noise on cognitive performance and annoyance in open plan offices

Laurent Brocolini, Etienne Parizet, Patrick Chevret

## To cite this version:

Laurent Brocolini, Etienne Parizet, Patrick Chevret. Effect of masking noise on cognitive performance and annoyance in open plan offices. Applied Acoustics, 2016, 114, pp.44-55. 10.1016/j.apacoust.2016.07.012 . hal-01473526

**HAL Id: hal-01473526**

**https://hal.science/hal-01473526v1**

Submitted on 19 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Effect masking noise on cognitive performance and annoyance in open plan offices

L. Brocolini [a], E. Parizet [a], P. Chevret [b]

(a) : Laboratoire Vibratios Acoustique, INSA-Lyon, Université de Lyon, F-69621 Villeurbanne Cédex

(b) : Institut National de Rercherche et Sécurité, INRS, rue du Morvan, CS 60027, 54519 Vandoeuvre les Nancy Cedex, France

## Abstract

Even for low noise levels, employees working in open-plan offices declare they are disturbed by different sound sources and especially by intelligible speech. This paper presents two experiments that aim at studying the effects of sound sources on task performance and cognitive load. In the first experience, thirty-five individuals were assigned a serial memorization task in four noise conditions: printers, intelligible speech, non-intelligible speech and phone rings. Noise annoyance was assessed by measuring task performance and cognitive workload (NASA-TLX). Although there was no significant difference between intelligible speech, unintelligible speech and phone ring tone signals in terms of annoyance, the performance in a memory task decreased in the presence of intelligible speech. In the second experiment, the relevance of a new intelligibility index was evaluated. This index was derived from SIIt as defined by Rhebergen et al. (2006) to deal with fluctuating noise as speech babble. Fifty-five subjects performed a serial memory task in four STIt (non-stationary STI) conditions (from 0.38 to 0.69) while STI values were kept constant (around 0.36). No significant difference appeared between the decreases in performance observed with multiple levels of intelligibility. However, an inter-individual variability appeared in performance measurements. The best performers seemed to be less affected by the high level of intelligibility condition than the other ones.

Keywords: open-plan office noise, task performance, cognitive workload, annoyance, sound sources, speech intelligibility

## 1 Introduction

Developed in the 1950's, the concept of the "open-plan" office is now widespread in the workplace. While offering many advantages in terms of layout and facilitating communication between colleagues, this way of organising the workspace had two major disadvantages. The feeling of privacy is lessened and noise level is increased causing discomfort for individuals working in this type of environment. Currently, it appears that

noise is a major nuisance factor in open-plan offices [1] in spite of a relatively low noise level (less than 65 dB(A)) [2]. Consequently, regulations in terms of noise levels in the workplace [3] are not representative of the actual situation. Furthermore, it seems that the noise nuisance experienced in open-plan offices affects work satisfaction [4] and employee health [5], and that this exposure to noise can reduce employee performance depending on the types of task to be carried out and the characteristics of the noise present in the workplace [6].

Numerous studies have been carried out over a number of years based on the effects of noise on performances in workspaces [7], [8]. However, very few have focussed on the nature of the various sources present in open-plan offices. The layouts of open-plan offices give rise to a multitude of sound sources: conversations or half-conversations, phone ring tones, machine noise (printers, photocopiers, computer keyboards), people moving around, ventilation systems, *etc.* Although conversations are most often thought of as being the most annoying noise source, phone ring tones can also be a significant annoyance. According to a field survey, Boyce found that out of 200 workers, 67% were upset by phone ring tones [9]. For Sundstrom *et al.*, out of more than 2000 employees, 54% said they were upset by noise, particularly from people speaking and ringing telephones [10]. In 2005, Banbury and Berry showed through a field survey that out of 88 employees, continuously-ringing telephones (which were not answered) were more annoying than other people's conversations followed by phone ring tones [11]. More recently, Pierrette *et al.* had a questionnaire filled out by 237 employees of 7 French companies [12]. In this study, it emerged that intelligible conversations caused the most significant annoyance followed by unintelligible conversations, phone ring tones and lastly the noise of machines such as printers. In view of the latest results, we wanted to check that these effects could be measured by experiments carried out in the laboratory. Indeed, although there have been numerous studies carried out in a controlled environment of the effect of speech on performance, not many have been carried out on the influence of other sound sources. In a laboratory context, Shelton *et al.* showed that in a lexical decision task, performance was reduced in the presence of phone ring tones (compared with the same task under silent conditions) [13]. This result was also demonstrated by Martellotta *et al.* [14]. In this study, the environments of individual offices characterised in particular by telephones, affect performance in a memory task and the annoyance felt. Likewise, the environments of open-plan offices, in which the presence of voices is the largest factor, also have a strong impact on performance and annoyance. In addition, these environments affect the Stroop test the most (compared with other sound environments mostly characterised by low frequency components). In order to provide more in-depth knowledge on the nature of the various sound sources present in open-plan offices, the first experiment carried out in this paper has therefore dealt with the influence of the type of sound source upsetting a cognitive task. Using a serial memory task, the performance of a number of listeners disturbed by various sound sources was measured along with the annoyance felt.

Secondly, we focussed on a single type of sound source - the intelligible speech signal - for which the level of intelligibility varied. The noise of speech is very widespread in open-plan offices and it is recognised that the intelligibility of the speech is a major nuisance factor. Widely used to measure the intelligibility of a signal, the STI (Speech Transmission Index)

[15] is a good indicator of the nuisance when the speech signal is masked by a stationary noise. A number of authors have demonstrated the influence of intelligibility on nuisance and performance [16]–[19]. However, the STI does not take account of the influence on intelligibility of a masking noise such as the one pervading an open-plan office (due to other conversations). Indeed, this noise has a fluctuating character which encourages intelligibility. One of the main reasons is a phenomenon called "listening in the gaps" which enables a listener to take in the information contained in the non-masked portions of the target signal in order to reconstruct the meaning. So for a given STI value, intelligibility will be better in the case of a fluctuating noise. Rhebergen *et al.* have proposed an indicator which takes account of the fluctuating nature of the mask. This indicator is the eSII, for extended Speech Intelligibility Index [20], [21]. The basic principle is that the speech signal and the noise are segmented into portions of short duration. The SII is calculated for each portion using the current standard (ANSI S3.5-1997) [22]. The global intelligibility index eSII is then the average of the SII values over the whole duration of the processed signal. Using the calculation method of the eSII and applying it to the STI (this being the indicator which is mostly used in all studies on open-plan offices), the STIt is an average of the sliding STIs calculated over a short time window (of the order of 12.5 ms). Thus, given an equal STI between a stationary masking noise and a fluctuating one, the STIt is larger the greater the fluctuations. Through the second experiment carried out in this paper, we wanted to know whether the STIt could be a good indicator of nuisance in the case where the noise masking the target signal is fluctuating. The second experiment therefore consisted of measuring the performance of listeners during a serial memory task while they were in the presence of a signal masked by ambient noises which fluctuated to a greater or lesser extent but which had the same STI.

# 2 Experiment 1

## 2.1 Objectives

Using a questionnaire distributed to the employees of a number of French companies, Pierrette *et al.* were able to highlight a number of troublesome sound sources in open-plan offices [12]. Intelligible conversations appear to be the sound source most often heard, followed by phone ring tones, people moving around, machines such as printers operating and lastly unintelligible conversations. However, the hierarchy in terms of annoyance is different: non-intelligible conversations appear in second position behind intelligible conversations. After this come phone ring tones, machine noise and people moving around. From these results, the objective of experiment 1 presented in this paper was to confront individuals with these sound scenarios in the laboratory in order to assess the impact of these acoustic conditions on the perceived annoyance along with performance measured by means of an intellectual activity such as a serial memory task.

## 2.2 Subjects

Thirty-five individuals took part in the experiment. Among these thirty-five individuals, twenty-eight were recruited from outside the laboratory and were paid, the others were postgraduate students of the Acoustic Vibrations Laboratory. The group of subjects was made up of seventeen women and eighteen men. Their median age was 41, the youngest was aged 23 and the oldest 59.

All participants underwent audiogram testing in the audiometric booth of the laboratory. The total hearing loss was calculated as the average of the hearing losses at frequencies 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz (weighted in the case of asymmetric losses) in accordance with the recommendations of the International Bureau for Audiophonology (BIAP) [23]. Among the 35 participants, 21 had no hearing loss, 9 had an impairment between 21 and 22 dB H.L. and 5 had an impairment between 24 et 30 dB H.L. Again according to the recommendations of the BIAP, this represents a slight hearing loss. Thus we have assumed that no subject had a hearing impairment liable to call into question the validity of the experiment.

## 2.3 Performance task

For each sound condition, the individuals have to memorise ten lists of ten words in French (presented visually) and repeat them orally (one list after the other). The words have been chosen from the lexical lists produced by D. Dubois and J. Poitou [24]. Among twenty-five lexical categories, five categories were selected: Animals, Fruits, Vegetables, Birds, Clothing. Each category had thirty words (the most frequently used [24]): "dog, cat, horse, lion, elephant, *etc.*" for the "Animals" category, "apple, pear, orange, cherry, banana, *etc.*" for the "Fruits" category, *etc.* The five categories appear twice during the same sound condition in a random order and the ten words from each list were also chosen randomly from among the thirty in each category.

Each list of words was announced on the screen by an image which was displayed for five seconds (so that the participant was prepared for the first word on the list to appear). Each word was displayed on the screen for two seconds (MS Sans Serif 48 point font). The words appeared one after the other without a pause. After the tenth word, the subject had twenty seconds to say out loud the words they had remembered. A microphone was used to record their answers. A pause of fifteen seconds followed the repeat-back phase, then a new series started: five seconds announcement, twenty seconds to memorise (10 words x 2 secs), twenty seconds of repeating back, fifteen seconds break, and so on through the ten series.

### 2.4 Sound stimuli

During this experiment, five sound conditions were used, relating to the four types of sound source and a control condition which consisted of a wide band background noise similar to a ventilation system. The sources were the sound sources identified as troublesome during the study by Pierrette *et al.*[12], namely: machine noise (printers), intelligible speech, unintelligible speech and lastly phone ring tones. All signals used were built up from sound samples lasting a few seconds found on the Internet [25]–[27]. More details regarding each of the sources are given in the following paragraphs.

#### 2.4.1 Background noise (Control)

The background noise used during this experiment came from the "Universal Soundbank" website [25]. This is the noise of a ventilation system lasting a few seconds which was repeated in order to obtain a continuous signal lasting ten minutes. The level of this background noise was set at around 35 dB(A) at the position of the listener.

#### 2.4.2 Printers (P)

The sounds of printers were also retrieved from the "Universal Soundbank" site [25]. These were three samples of respectively 12.5, 16 and 19.5 seconds. The first was the characteristic noise of an "inkjet" printer while the other two were of "laser" printers. Each of these samples was repeated twelve times to obtain thirty-six samples in total.

#### 2.4.3 Intelligible speech (I-S)

From the "Audio-lingua" website [26], the intelligible speech database had thirty-six different samples. This involved individuals speaking on various subjects in French. Each sample was cut to obtain twelve files of duration 12.5, 16 or 19.5 seconds. Half of the samples were women's voices and half men's.

#### 2.4.4 Non intelligible speech (N-I-S)

To obtain non-intelligible speech signals, nearly fifty sound extracts of intelligible speech (the shortest about twenty seconds and the longest about a minute) were retrieved from the "RFI" website [27]. These were from French courses. From these conversations, 12.5 seconds were selected randomly from eight different files (also selected randomly) and were mixed

together. Using the same procedure, eleven other samples of 12.5 seconds were built up along with twelve of 16 seconds and twelve of 19.5 seconds.

### 2.4.5 Phone rings (P-R)

In the same way as for printers, the sounds of telephones ringing were extracted from the "Universal Soundbank" website [25]. Four different ring tones were used, each being reworked into a sample lasting 12.5 seconds, one of 16 seconds and one of 19.5 seconds. Four types of ring tone were provided of three different durations making twelve samples. These signals were repeated three times to end up with a total of thirty-six samples.

For each sound sample, thirty-six samples were placed end-to-end (in a random order for each subject), with a one-second interval in between giving a total sound signal length of ten minutes. The background noise was emitted with a level $L_{Aeq,10min}$ of about 35 dB(A) while the sound sources superimposed on the background noise were played with a level $L_{Aeq,10min}$ of about 50 dB(A). All samples had been previously equalised to a level of $L_{Aeq,12.5s} = L_{Aeq,16s} = L_{Aeq,19.5s} = 50$ dB(A). The sound levels were chosen to approximate to the actual situation in an open-plan office [28].

## 2.5 Procedures

### 2.5.1 Research setting

The experiment took place in an audiometric booth in the laboratory. All stimuli were presented via a loudspeaker (Tapco S8) placed about 1 metre from the floor and from the listener behind the computer screen. The reproduction chain was frequency-corrected (31-band ULTRA-CURVE PRO DSP8024 equaliser) such that the frequency response of the loudspeaker at the listening position was flat (the standard deviation of the level was 1.4 dB over all third octave bands between 40 and 12500 Hz).

Before starting the experiment, each participant had undergone audiogram testing (Oscilla USB-300 Screening Audiometer/AudioConsole 2.3.8). It was then explained to the individual that the experiment consisted of memorising lists of ten words and repeating back as many as possible (the responses were recorded using a directional microphone placed between the listener and the computer screen). To prevent the subjects saying too many words, it was explained to them that the number of "correct" words would be counted (+1) and that a penalty (-1) would be applied for "incorrect" words. To ensure the procedure was understood, an example with only two lists of words and without any sound source was then carried out in the presence of the experimenter. Then, all participants started the experiment with a background noise to familiarise them with the experiment. Lastly, they all carried out the requested task in the presence of each sound condition (background noise, machines, intelligible speech, unintelligible speech, telephones) in a random order based on a Latin square (5x5) [29] to obviate any order effect. At the end of each sound condition, the participants had to answer a French version of the NASA-TLX questionnaire (*see* 2.5.3). Lastly, at the end of the experiment, they answered an annoyance questionnaire.

### 2.5.2 Performance measurement

The performance measured for each listener consisted of a sum of the words correctly repeated from which the number of incorrect words was subtracted. In order to assess the impact of the various sound conditions, all performances were expressed relative to that measured when the background noise only was played (the second presentation, the first was only used for learning). In other words, the measured data did not represent the raw performance but a decrease in performance calculated as the difference between the performance obtained in the control situation and the performance obtained with each sound condition:

$$Decrease\ in\ performance = \frac{Perf_{Control} - Perf_{Source}}{Perf\ max}$$

### 2.5.3 Workload measurement

Developed at the end of the 80's, the NASA-TLX (Task Load Index) [30], [31] makes it possible to assess the work load through six questions. Each of these questions relates to a component of the work load: mental demand, physical demand, temporal demand, performance, effort, frustration. The questions are assessed on a scale of 0 to 100. Secondly, a weight is allocated to each component. NASA-TLX is a weighted mean of the six scores. NASA-RTLX (Raw Task Load Index) is a simplified version, namely, non-weighted and enables similar results to be obtained in less time [32].

Each participant completed the NASA-RTLX using a graphic interface developed under Matlab after each sound condition. In the same way as for performance, the work load linked to the task carried out was expressed relative to that measured during the transmission of the background noise only. So as to work with theoretically positive data, the increase of NASA-RTLX was calculated as follows:

$$Increase\ in\ NASA_RTLX = \frac{NASA_RTLX_{Source} - NASA_RTLX_{Control}}{NASA_RTLX\ max}$$

### 2.5.4 Annoyance measurement

The annoyance felt was measured again using the Matlab graphic interface. Unlike NASA-RTLX, the annoyance was measured at the end of the experiment after hearing all the sound conditions. The question asked was as follows: "At which point did the sound sources affect you in accomplishing the task requested?". A button on which was written the name of the sound source allowed the user to listen to it again for five seconds. Beside each source, a continuous scale allowed the subjects to score the annoyance felt using a cursor. Five points on each scale were marked with the words: "Not at all", "Slightly", "Moderately", "Very", "Extremely" [33] giving the participants an idea of where to place the cursor. The position of the cursor was converted into a score between 0 (not at all annoying) and 100 (extremely annoying). To analyse the results, the score for each annoyance consisted of an increase in the annoyance level between each sound source and the background noise:

$$Increase\ in\ annoyance = \frac{Annoyance_{Source} - Annoyance_{Control}}{Annoyance\ max}$$

## 2.6 Results

### 2.6.1 Performance

The decreases in average performance according to the sound source are shown in figure 1. On average, the decrease in performance was the greatest when the sound source was intelligible speech (6.8%) followed by unintelligible speech (3.4%), printers (2.7%) and lastly phone ring tones (1.9%).

In order to check whether there was a significant difference between the various sound conditions, a repeated measures analysis of variance was carried out with one dependent variable: the decrease in performance, and one independent variable: the type of sound source. This analysis of variance showed that the sound condition had a significant effect on the decrease in performance ($F(3,102) = 5.7$, $p = 0.0012$). An Honest Significant Difference test (HSD), whose p values are given in table 1, shows that the decrease in performance is significantly greater in the presence of intelligible speech in comparison with other sound sources as can be seen from figure 1.

**Table1: HSD test p-values**

| Sound source | P | I-S | N-I-S | P-R |
|---|---|---|---|---|
| P | | **0.0104*** | 0.9496 | 0.9074 |
| I-S | **0.0104*** | | **0.0461*** | **0.0012*** |
| N-I-S | 0.9496 | **0.0461*** | | 0.6219 |
| P-R | 0.9074 | **0.0012*** | 0.6219 | |

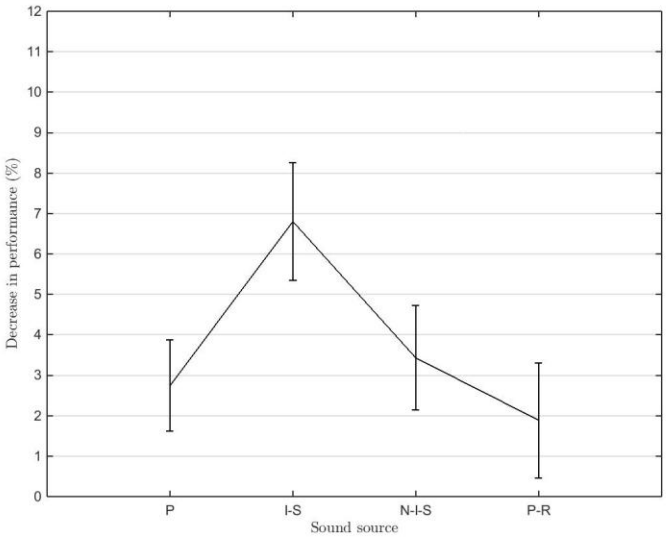


**Figure 1: Decrease in performance with sound sources. (Mean values are presented with their corresponding standard errors)**

A second repeated measures analysis of variance was carried out this time including the "age" and "sex" factors. This analysis did not reveal a link between the decrease in performance and the age or sex of the subjects.

### 2.6.2 NASA-RTLX

The most significant increase in cognitive load is obtained with intelligible speech (8.7%). However, unintelligible speech and phone ring tones can cause an extra cognitive load which is relatively close (7.8% and 6.8%). Only printers lead to a relatively low increase in NASA-RTLX (3.7%). This is confirmed by the one-way repeated measures analysis of variance carried out with the increase in NASA-RTLX as dependent variable and the type of source again as independent variable. Again, the results of this analysis are significant ($F(3,102)=2.9$, $p=0.0408$). The sound condition therefore has an influence on the increase in the cognitive load associated with a serial memory task.

The analysis of figure 2 combined with a Tukey post-hoc test shows that there is no significant difference between the "Printers", "Unintelligible speech" and "Phones" conditions. Likewise, there is no significant difference between the "Intelligible speech", "Unintelligible speech" and "Phones" conditions. On the other hand, the "Printers" and "Intelligible Speech" sound conditions are significantly different (*see* Table 2).

**Table2: HSD test p-values**

| Sound source | P | I-S | N-I-S | P-R |
|---|---|---|---|---|
| P | | **0.0356*** | 0.1166 | 0.3223 |
| I-S | **0.0356*** | | 0.9606 | 0.7316 |
| N-I-S | 0.1166 | 0.9606 | | 0.9495 |
| P-R | 0.3223 | 0.7316 | 0.9495 | |

A three-way repeated measures analysis of variance (condition, age, sex) shows that neither age nor sex has any effect on the increase in NASA-RTLX, nor any interaction whatsoever.

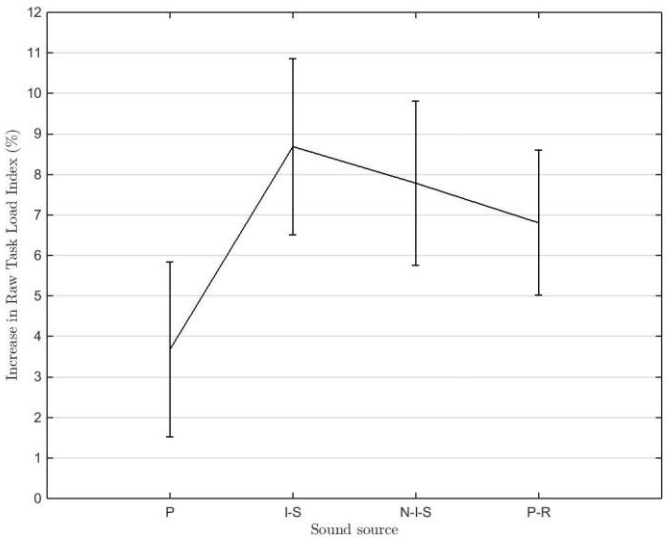


**Figure 2: Increase in RTLX with sound sources. (Mean values are presented with their corresponding standard errors)**

### 2.6.3 Annoyance

In the case of an increase in the annoyance felt, the results are similar to the increase in NASA-RTLX, even if the order of magnitude is greater. On average, the increase in annoyance felt is similar during intelligible speech, unintelligible speech and phone ring tone conditions (respectively 55.9%, 50.5% and 51.7%). The increase in annoyance felt with the printers is however mostly less (29.6%). A repeated measures analysis of variance confirms this significant effect of the sound condition ($F(3,102)=15.3$, $p=0.0000$). With regard to figure 3 and the Tukey difference test (*see* Table 3), it appears that "Intelligible speech", "Unintelligible speech" and "Phones" are not significantly different in effect. On the other hand, the increase in annoyance is significantly lower under the "Printers" condition than in the three other sound conditions.

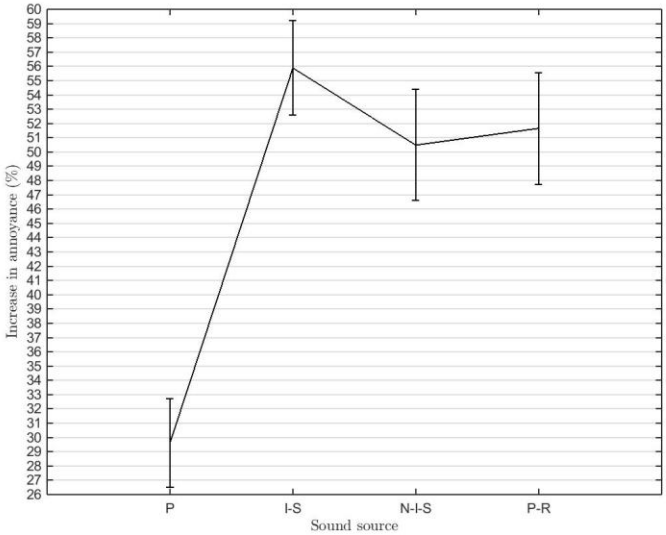


**Figure 3: Increase in annoyance with sound sources. (Mean values are presented with their corresponding standard errors)**

**Table3: HSD test p-values**

| Sound source | P | I-S | N-I-S | P-R |
|---|---|---|---|---|
| P | | **0.0001*** | **0.0002*** | **0.0001*** |
| I-S | **0.0001*** | | 0.5830 | 0.7524 |
| N-I-S | **0.0002** | 0.5830 | | 0.9927 |
| P-R | **0.0001** | 0.7524 | 0.9927 | |

We also note that with regard to the annoyance felt, the age and sex of the subjects once more had no significant effect after a three-way repeated measures analysis of variance.

## 2.7 Discussion

The aim of experiment 1 was to assess the impact of various sound sources on performance in carrying out an intellectual task as well as on the perceived annoyance. The results obtained for performance measurement were expected and in agreement with what had been shown on numerous occasions: the performance in a memory task decreases in the presence of intelligible speech [17], [34]–[36]. On the other hand, the results obtained in terms of annoyance were less expected. No significant difference appeared between intelligible speech, unintelligible speech and phone ring tone signals. However, during a field survey by Pierrette *et al.*[12], intelligible conversations were by far the most annoying sound sources. Although it is difficult to provide an explanation as to these results, they are however confirmed by the analysis of the NASA-RTLX. Indeed, although having been evaluated using a slightly different procedure (NASA-RTLX was evaluated after each sound condition while the annoyance was evaluated at the end of the experiment), the results were similar. Intelligible speech is significantly different from printer noise, but is not more annoying than unintelligible speech. Although less efficient in the presence of intelligible speech, the

individuals did not however feel more troubled. However a number of studies have shown a significant difference between these two sound sources in terms of annoyance [19], [37] and work load [36]. However, these results have been obtained with parameters associated with different stimuli which were used in our experiment. For example, in the study by Schlittmeier *et al.*[37], the stimuli were present at a level at 35 dB(A) while in our case, the sounds were broadcast at 50 dB(A). Likewise, Haka *et al.*[19] show a variation in annoyance according to intelligibility using single-voice signals transmitted at 48 dB(A). In our study, no difference was visible between an intelligible signal with a single voice and unintelligible speech with multiple voices. These different parameters may in part explain the differences in the results obtained.

Anyhow, the speech signal remains a special sound source which led us on to test the influence of the level of intelligibility in experiment 2.

# 3 Experiment 2

## 3.1 Objectives

The objective of experiment 2 was to evaluate the influence of ambient speech on performance in the case where the background noise was fluctuating. Under experimental conditions similar to experiment 1, the performance of the listeners during a serial memory task was measured in the presence of a target voice and a masking noise. A mix of different voices was used to achieve a masking noise with varying degrees of fluctuation. Five sound conditions were tested:

- 1 stationary noise only as control condition
- 1 male voice (target) + stationary noise (mask)
- 1 male voice (target) + 5 voices (mask)
- 1 male voice (target) + 3 voices (mask)
- 1 male voice (target) + 1 voices (mask)

With regard to the mask, the higher the number of voices, the more the mask tended to become stationary. For this reason, if the average spectral content of each mask used was identical, the Speech Transmission Index (STI) would remain identical while the STIt would be higher the more the masking signal fluctuated.

## 3.2 Subjects

Fifty-five individuals took part in experiment 2. The group of subjects was made up of twenty-nine women and twenty-six men. Their median age was 44, the youngest was aged 31 and the oldest 61. They were all paid for taking part. All participants underwent audiogram testing in the audiometric booth in the laboratory and the total hearing loss was calculated (*see* 2.2). Among the subjects, 41 had a total hearing loss of between 20 and 22 dB H.L, 13 had an impairment between 23 and 29 dB H.L and one participant had a hearing impairment of 34 dB H.L. Regarding these hearing losses, no subject was discounted.

### 3.3 Performance task

The performance task was strictly the same as that used during experiment 1 (*see* 2.3).

### 3.4 Sound stimuli

Five sound conditions were used, corresponding to four masking levels of a target voice and one control condition which consisted of a normalised LTASS speech noise [38] without target voice.

#### 3.4.1 Control

The control condition used during this experiment was 22 seconds of white noise generated using Matlab then LTASS normalised. The level of this background noise was set at around 47 dB(A) at the position of the listener.

#### 3.4.2 Clear speech

The target voice came from recordings available from the Collège National d'Audioprothèse. More precisely, it involved a "Man" reading phrases from Combescure [39], Fournier [40], along with lists from the Hearing in Noise Test (HINT) [41]. All the available recordings were juxtaposed and cut into 40 samples of 22 seconds. These 40 samples were all normalised using the LTASS speech spectrum.

#### 3.4.3 Masking sounds

The masking noise was made up of men's and women's voices. From the "Audio-lingua" website [26], it involved individuals speaking on various subjects in French. All the available signals were juxtaposed and cut into 90 samples of 22 seconds after having removed as many "blanks" as possible from the various speeches. Each of these 90 samples were normalised using an LTASS speech spectrum.

For each listener taking part in the experiment, the 1-voice mask was from these 90 samples, the 3-voice mask was a mix of 3 samples and the 5-voice mask a mix of 5 samples. All masking signals were equalised at the same level $L_{eq,22s}$. The $4^{th}$ masking noise was a stationary noise. This was the same signal as that used for the control condition.

To avoid any possible confusion between the target signal and the mask (particularly in the case of the "1 voice (target) + 1 voice (mask)" sound condition), the masking signals were temporally inversed. In this way, the average spectrum remained unchanged, the same as the STI, and there was no possible confusion between the target signal and the masking signal.

The signal-to-noise ratio (target/mask) was set to -4 dB, *i.e.* a STI of approximately 0.35. A STI value such as this corresponds to the area where heightened decrease in performance is at its greatest in the predictive model of decrease in performance developed by Hongisto *et al.* [17]. The STI and STIt were systematically calculated for each stimulus created (4 masks x 10 repetitions x 55 participants). Figure 4 shows the averages of 550 STI and STIt values

calculated for each masking condition. The average STI was 0.35 with the stationary masking noise and 0.36 for the other masks (5 voices, 3 voices, 1 voice). The variation in STI as expected was minimal and the standard deviation was greater the lower the number of voices present in the mask and therefore the higher the modulation. The average STIt calculated on the other hand was 0.38 with the stationary mask, 0.51 with the 5-voice signal, 0.56 with the 3-voice signal and 0.69 with the 1-voice signal. There again, the standard deviation was greater the higher the modulation of the masking signal. Only a few STIt values of the "3-voice" condition were less than those of the "5-voice" condition. Overall, the 4 masking conditions were very different from one another from the STIt point of view.

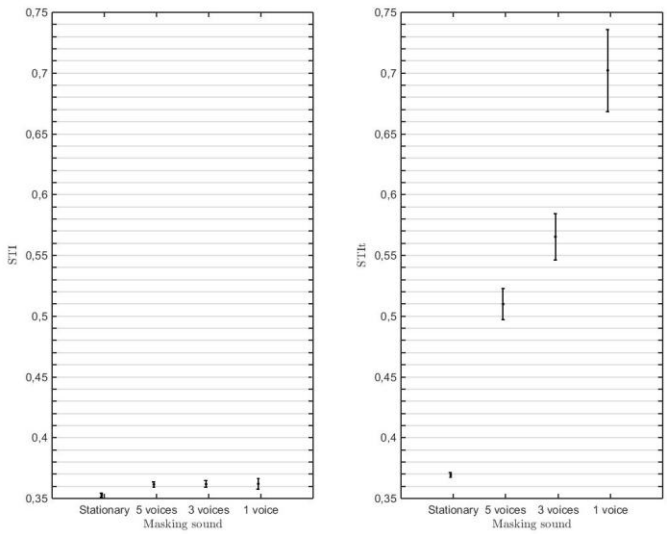


**Figure 4: Mean values of STI and STIt for each masking sound (Mean values are presented with their corresponding standard deviations).**

The difference between the average STIt calculated (0.38) and the STI calculated (0.35) for the stationary mask is explained by the method of calculating spectra in the octave band used. Indeed, for very short times (12.5 ms) the digital filters used with Matlab underestimate levels in the lowest octave bands.

During experiment 1, the stimuli were emitted throughout the period when the subjects were carrying out the requested cognitive task, namely during the word memorising and repeat-back phase (*see* 2.3 for more details of the cognitive task). In this new experiment, taking into account the signals available, the stimuli were only transmitted during the memorising phase.

### 3.5 Procedures

The sequence of experiment 2 was almost the same as for experiment 1 with a few exceptions (*see* 2.5). The main difference was in the fact that the participants in experiment 2 did not assess annoyance. Only the performance (*see* 2.5.2) and the cognitive load (*see* 2.5.3) were measured.

## 3.6 Results

### 3.6.1 Performance

The average performance decrease values are given in figure 5. On average, the decrease in performance was 3.71% when the mask was a stationary noise, 2.22% when the mask was made up of 5 voices, 3.56% with 3 voices and 3.05% with 1 voice. These results are compliant with those found by Ellermeier and Hellbrück (Exp. 2B) [16] which predicted a decrease in performance of about 3% for an STI of 0.35. This decrease in performance seemed reasonably stable whatever masking noise was used. To confirm this, a repeated measures analysis of variance was carried out with one dependent variable: the decrease in performance and one independent variable: the type of masking. This analysis of variance showed that there was no significant effect due to the fluctuation of the masking noise on the decrease in performance ($F(3,162) = 1.3$, $p = 0.2927$).

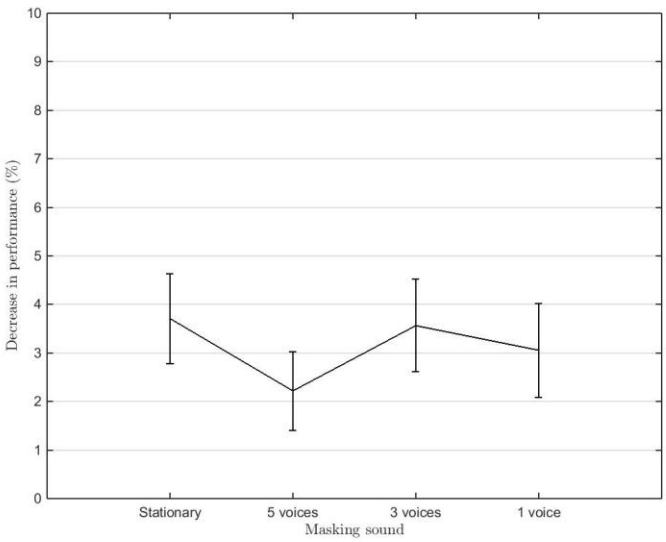


**Figure 5: Decrease in performance with masking sound (Mean values are presented with their corresponding standard errors).**

In addition to the analysis of variance, a comparison test with a standard (0) of the percentage decrease in performance was carried out. The results of this test are shown in table 4. The fact that the percentage decrease in performance is statistically non-zero implies that adding a voice worsens the performance in comparison with the control situation which does not have a target signal. Je reste perplexe sur le t-test. Pourquoi tu ne fais pas un tuckey ? Et, je ne vois pas ce que tu appelles « standard (0) ».

**Table 4: one sample t-test for percentage of decrease in performance**

| Mask | %DP | Ref | t | dl | p |
|---|---|---|---|---|---|
| Stationary | 3.71 | 0 | 3.96 | 54 | 0.0002 |
| 5 voices | 2.22 | 0 | 2.71 | 54 | 0.0089 |
| 3 voices | 3.56 | 0 | 3.70 | 54 | 0.0005 |
| 1 voice | 3.05 | 0 | 3.13 | 54 | 0.0028 |

To explore these results in greater depth, a hierarchical clustering (Ward's method) was carried out on the raw performance obtained for five sound conditions - control (control1 was used for learning and control2 was used as reference), stationary, 5 voices, 3 voices and 1 voice. This classification highlighted two groups of individuals. Group 1 comprised 21 individuals and group 2 34. The first group of 21 people had consistently better performances (whatever the condition) than those in the second group of 34 subjects as can be seen in figure 6 - no relationship between this classification and the data specific to the subjects (gender, age, hearing) could be identified. Now, the decrease in performance is similar between the two groups except under the condition with a single masking voice (*see* Figure 7). A mixed repeated measures analysis of variance (one factor within subjects, one factor between subjects) showed that the interaction between the two factors is significant ($F(3,159)=3.5$, $p=0.0177$).

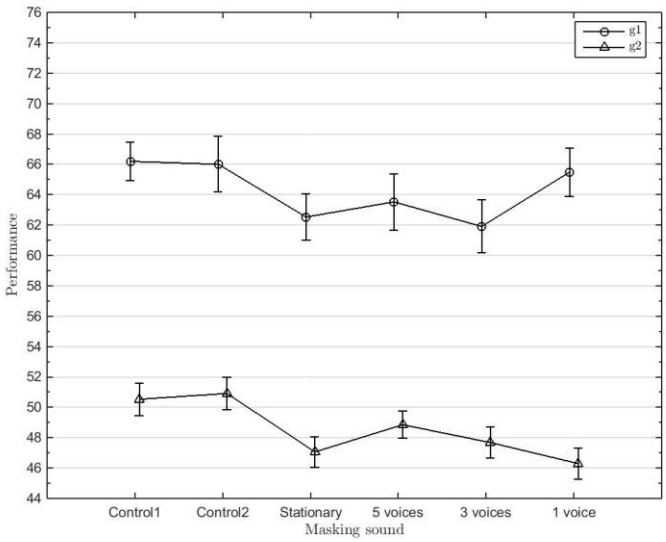


**Figure 6: Performance with masking sound for the two groups of listeners (Mean values are presented with their corresponding standard errors).**

Figure 7 shows that the effect resides in the "one masking voice" condition (corresponding to the highest STIt). The decrease in performance is therefore significantly greater for subjects with worse performances overall.

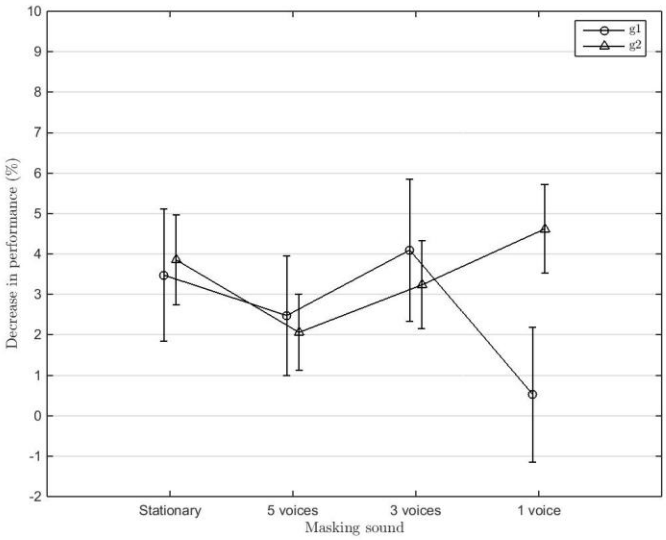


**Figure 7: Decrease in performance with acoustic condition for the two groups of listeners (Mean values are presented with their corresponding standard errors).**

### 3.6.2 NASA-RTLX

The results of the increase in work load are similar to those obtained for the decrease in performance (*see* Figure 8). As before, a one-way repeated measures analysis of variance carried out with the rate of increase in NASA-RTLX as dependent variable and the masking sound again as independent variable. Again, the results of this analysis were not significant ($F(3,102)=1.9$, $p=0.124$). The fluctuation in masking noise therefore has no influence on the increase in the cognitive load associated with a serial memory task.

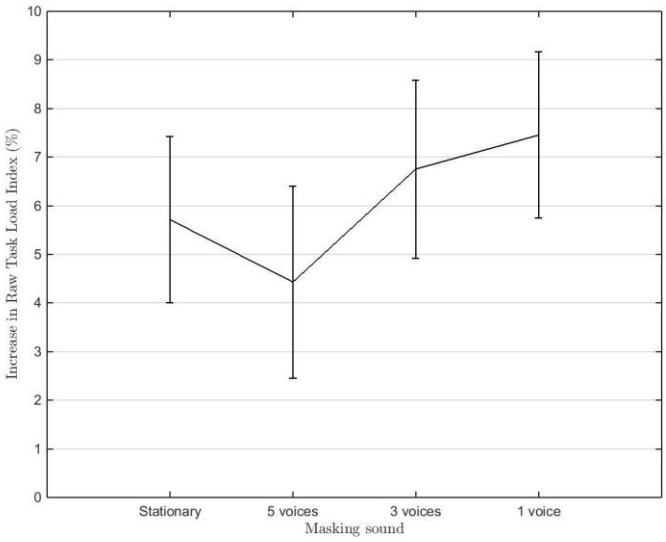


**Figure 8: Increase in Raw Task Load Index with masking sound (Mean values are presented with their corresponding standard errors).**

Following this variance analysis, a comparison test with a NASA-RTLX increase standard was also carried out. The results of this comparison test may be seen in table 5.

**Table 5: one sample t-test for percentage of increase in Raw Task Load Index**

| Mask | %IRTLX | Ref | t | dl | p |
|---|---|---|---|---|---|
| Stationary | 5.71 | 0 | 3.32 | 54 | 0.0016 |
| 5 voices | 4.43 | 0 | 2.22 | 54 | 0.0307 |
| 3 voices | 6.76 | 0 | 3.65 | 54 | 0.0006 |
| 1 voice | 7.45 | 0 | 4.32 | 54 | 0.0001 |

Once again, the significant differences relative to 0 of the increase in NASA-RTLX for each sound condition show that extra work load is required when one voice is added to the signal.

Likewise for performance, a mixed repeated measures analysis of variance with the "group" factor was carried out with the percentage increase of NASA-RTLX as dependent variable. This time, the results show no significant effect whether of the mask or the group or the interaction between the two. However, it is worthwhile noting that if the groups were formed on the basis of raw performance, the segmentation of these two groups is similar in the measurement of the cognitive load (je ne comprends toujours pas, est-ce que tu veux dire que tu obtiendrais la même segmentation pour la performance que pour la charge cognitive. Si c’est le cas, la phrase n’est pas correcte me semble-t-il). Whatever the sound condition, the memory task was less demanding for group 1 (the better performers) than for group 2 as can be seen in figure 9.

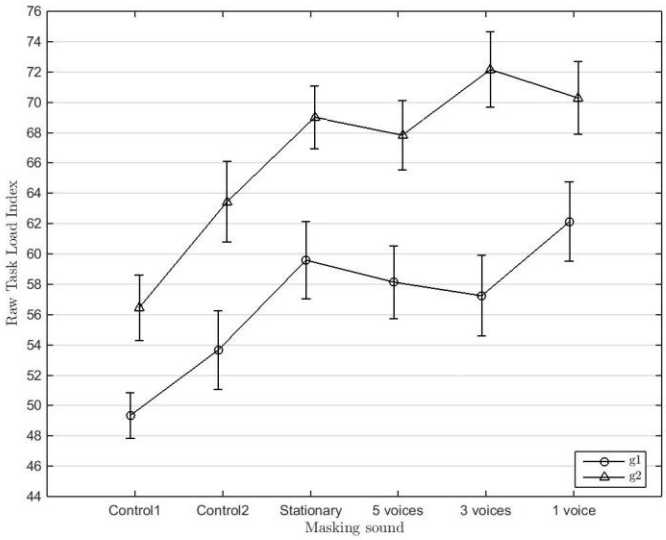


**Figure 9: NASA-RTLX with acoustic condition for the two groups of listeners (Mean values are presented with their corresponding standard errors).**

## 3.7 Discussion

In this experiment, no significant difference appeared between the decreases in performance observed with multiple levels of mask modulation. These results therefore seem to contradict the proposal to use the STIt temporal index as indicator of performance in the case where the masking noise of a speech signal fluctuates. It also goes counter to the other indicators such as the fluctuation force as proposed by Schlittmeier *et al.* in particular [42]. According to the authors, the performance in a short-term memory task is directly linked to the fluctuating force of the stimulus. Now, in our experiment, the fluctuation force is naturally greater for the "1 target voice + 1 voice" (Fs = 0.65 vacil) mix than for the "1 voice + stationary noise" (Fs = 0.1 vacil) mix and is therefore not representative of performance. On the other hand, the results obtained did not completely contradict the study carried out by Haapakangas *et al.* [43]. In this study, the performance during a serial memory task was measured in the presence of a speech masked by five different noises. Among the five masks, there was in particular a pink noise, ventilation noise and music containing words. Although not having been measured according to an objective criterion such as the fluctuation force, it appeared clear according to the authors that the ventilation noise and the pink noise were more stationary in time than the musical mask. However, the decreases in performance measured with these three masks (with equivalent STI) were not significantly different.

Between subject differences appeared for the condition "1 target voice + 1 voice". The listeners in the group with higher performance seemed less affected by the noise. Ebissou *et al.* [36] have already shown that there is a difference in performance specific to the subjects, but this difference was visible whatever the sound condition (variable STI). In our case only the "1 target voice + 1 masking voice" condition highlighted a difference between the two groups of subjects. For the other sound conditions, the two groups seemed similar.

Regarding the work load measured via the NASA-RTLX, no significant impact was found between the various sound conditions which confirms the results previously found on the subject of performance.

## 4 Conclusion

The two experiments presented in this paper have highlighted under laboratory conditions a significant effect of intelligible speech on the performance achieved during a serial memory task (in comparison with other sound sources such as unintelligible speech, phone ring tones and printers). This effect is not found, however, through a subjective measurement of the work load or an evaluation of the annoyance although this has been shown during field surveys [12].

Furthermore, no effect on performance was observed due to the fluctuation of a masking signal. However, the relationship between STIt and the intelligibility of the speech in a fluctuating noise has been confirmed in the literature. These contradictory results could mainly be explained by the short period of exposure. As the effects of the factors for the various conditions were of the second order (as the STI was equal under these conditions), the subjects were able to concentrate their cognitive resources during the ten minutes of each condition.

So, for the future, longer duration tests might be a path which would be worthwhile exploring further.

## Funding

This project was funded by ANSES (Agence nationale de sécurité sanitaire, de l'alimentation, de l'environnement et du travail), in the frame of the Environment-Health-Work Program, with the support of Governmental Offices for Environment and Work. A major portion of the work was conducted within the CeLyA (Centre Lyonnais d'Acoustique) [ANR-10-LABX-60] research network.

## References

[1] S. Amstutz, S. Kündig, and C. Monn, "Etude SBiB - Enquête suisse dans les bureaux," 2010.

[2] U. Landström, E. Akerlund, A. Kjellberg, and M. Tesarz, "Exposure levels, tonal components, and noise annoyance in working environments," *Environ. Int.*, vol. 21, no. 3, pp. 265–275, 1995.

[3] E. Directive, "10/EC of the European Parliament and of the Council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise)," *DOUE Febr.*, 2003.

[4] P. Leather, D. Beale, and L. Sullivan, "Noise, psychosocial stress and their interaction in the workplace," *J. Environ. Psychol.*, vol. 23, no. 2, pp. 213–222, 2003.

[5] J. Pejtersen, L. Allermann, T. Kristensen, and O. Poulsen, “Indoor climate, psychosocial work environment and symptoms in open-plan offices,” *Indoor Air*, vol. 16, no. 5, pp. 392–401, 2006.
[6] C. Beaman, “Auditory distraction from low-intensity noise: a review of the consequences for learning and workplace environments,” *Appl. Cogn. Psychol.*, vol. 19, no. 8, pp. 1041–1064, 2005.
[7] A. Kaarlela-Tuomaala, R. Helenius, E. Keskinen, and V. Hongisto, “Effects of acoustic environment on work in private office rooms and open-plan offices–longitudinal study during relocation,” *Ergonomics*, vol. 52, no. 11, pp. 1423–1444, 2009.
[8] J. L. Szalma and P. A. Hancock, “Noise effects on human performance: a meta-analytic synthesis.,” *Psychol. Bull.*, vol. 137, no. 4, p. 682, 2011.
[9] P. Boyce, “Users’ assessments of a landscaped office,” *J. Archit. Res.*, vol. 3, no. 3, pp. 44–62, 1974.
[10] E. Sundstrom, J. P. Town, R. W. Rice, D. P. Osborn, and M. Brill, “Office noise, satisfaction, and performance,” *Environ. Behav.*, vol. 26, no. 2, pp. 195–222, 1994.
[11] S. Banbury and D. Berry, “Office noise and employee concentration: Identifying causes of disruption and potential improvements,” *Ergonomics*, vol. 48, no. 1, pp. 25–37, 2005.
[12] M. Pierrette, E. Parizet, P. Chevret, and J. Chatillon, “Noise effect on comfort in open-space offices: development of an assessment questionnaire,” *Ergonomics*, vol. 58, no. 1, pp. 96–106, 2014.
[13] J. T. Shelton, E. M. Elliott, S. D. Eaves, and A. L. Exner, “The distracting effects of a ringing cell phone: An investigation of the laboratory and the classroom setting,” *J. Environ. Psychol.*, vol. 29, no. 4, pp. 513–521, 2009.
[14] F. Martellotta, S. Della Crociata, and A. Simone, “Laboratory study on the effects of office noise on mental performance,” in *Forum acusticum*, 2011, vol. 27.
[15] I. E. Commission and others, “Sound system equipment–Part 16: Objective rating of speech intelligibility by speech transmission index,” *Int. Stand. IEC*, pp. 60268–16, 2003.
[16] W. Ellermeier and J. Hellbrück, “Is level irrelevant in‘ irrelevant speech’? Effects of loudness, signal-to-noise ratio, and binaural unmasking.,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 24, no. 5, p. 1406, 1998.
[17] V. Hongisto, “A model predicting the effect of speech of varying intelligibility on work performance,” *Indoor Air*, vol. 15, no. 6, pp. 458–468, 2005.
[18] N. Venetjoki, A. Kaarlela-Tuomaala, E. Keskinen, and V. Hongisto, “The effect of speech and speech intelligibility on task performance,” *Ergonomics*, vol. 49, no. 11, pp. 1068–1091, 2006.
[19] M. Haka, A. Haapakangas, J. Keränen, J. Hakala, E. Keskinen, and V. Hongisto, “Performance effects and subjective disturbance of speech in acoustically different office types–a laboratory experiment,” *Indoor Air*, vol. 19, no. 6, pp. 454–467, 2009.
[20] K. S. Rhebergen and N. J. Versfeld, “A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners,” *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2181–2192, 2005.
[21] K. S. Rhebergen, N. J. Versfeld, and W. A. Dreschler, “Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise,” *J. Acoust. Soc. Am.*, vol. 120, no. 6, pp. 3988–3997, 2006.
[22] A. ANSI, “S3. 5-1997, Methods for the calculation of the speech intelligibility index,” *N. Y. Am. Natl. Stand. Inst.*, vol. 19, pp. 90–119, 1997.
[23] I. bureau for Audiophonology, “BIAP Recommendation 02/1 bis - AUDIOMETRIC CLASSIFICATION OF HEARING IMPAIRMENTS,” 1996. [Online]. Available: http://www.biap.org/index.php?option=com_content&view=article&id=5:recommandatio

n-biap-021-bis&catid=65:ct-2-classification-des-surdites&Itemid=19&lang=en. [Accessed: 07-Jul-2015].
[24] D. Dubois and J. Poitou, "'Normes catégorielles' et listes lexicales. 'Normes catégorielles' (liste de termes) pour vingt-deux catégories sémantiques," Cahiers du LCPE, vol. 5, 2002.
[25] "Universal-Soundbank," 2006. [Online]. Available: http://www.universal-soundbank.com. [Accessed: 13-Mar-2014].
[26] G. d'expérimentation pédagogique du C. R. D. P. de l'académie de Versailles, "Audiolingua," 2007. [Online]. Available: http://www.audio-lingua.eu/. [Accessed: 13-Mar-2014].
[27] RFI, "Cours de français, Comment vont les affaires ?" [Online]. Available: http://www1.rfi.fr/lfen/pages/001/liste_cva_anglais.asp. [Accessed: 13-Mar-2014].
[28] C. Pop, M. Wee, D. Cabrera, and G. Tan, "Acoustic descriptors for open plan offices - Evaluation of field measurements," in *9th European Conference on Noise Control, EURONOISE 2012*, 2012, pp. 636—641.
[29] J. V. Bradley, "Complete counterbalancing of immediate sequential effects in a Latin square design," *J. Am. Stat. Assoc.*, vol. 53, no. 282, pp. 525–528, 1958.
[30] S. Hart and L. Staveland, *NASA Task Load Index (TLX) V1. 0 Users Manual*. Hanover, MD: NASA Center for Aerospace Information (CASI), 1986.
[31] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, 2006, vol. 50, pp. 904–908.
[32] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, 1988.
[33] J. Fields, R. De Jong, T. Gjestland, I. Flindell, R. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felscher-Suhr, and R. Schumer, "Standardized General-Purpose Noise Reaction Questions for Community Noise Surveys: Research and a Recommendation," *J. Sound Vib.*, vol. 242, no. 4, pp. 641–679, 2001.
[34] A. Liebl, J. Haller, B. Jödicke, H. Baumgartner, S. Schlittmeier, and J. Hellbrück, "Combined effects of acoustic and visual distraction on cognitive performance and well-being," *Appl. Ergon.*, vol. 43, no. 2, pp. 424–434, 2012.
[35] H. Jahncke, V. Hongisto, and P. Virjonen, "Cognitive performance during irrelevant speech: effects of speech intelligibility and office-task characteristics," *Appl. Acoust.*, vol. 74, no. 3, pp. 307–316, 2013.
[36] A. Ebissou, E. Parizet, and P. Chevret, "Use of the Speech Transmission Index for the assessment of sound annoyance in open-plan offices," *Appl. Acoust.*, vol. 88, pp. 90–95, 2015.
[37] S. Schlittmeier, J. Hellbrück, R. Thaden, and M. Vorländer, "The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance," *Ergonomics*, vol. 51, no. 5, pp. 719–736, 2008.
[38] D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui, J. Kiessling, M. N. Kotby, N. H. A. Nasser, W. A. H. E. Kholy, Y. Nakanishi, H. Oyer, R. Powell, D. Stephens, R. Meredith, T. Sirimanna, G. Tavartkiladze, G. I. Frolenkov, S. Westerman, and C. Ludvigsen, "An international comparison of long–term average speech spectra," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2108–2120, 1994.
[39] Collège National d'Audioprothèse, "Vol. CD 3. Listes de phrases de Combescure, 2006," 18-Nov-2014. [Online]. Available: http://www.college-nat-audio.fr/fichiers/img92a.pdf.
[40] Collège National d'Audioprothèse, "Vol. CD 3. Listes de phrases de Fournier, 2006," 18-Nov-2014. [Online]. Available: http://www.college-nat-audio.fr/fichiers/img93a.pdf.

[41] Collège National d'Audioprothèse, "Listes de phrases HINT (Hearing In Noise Test), 2006," 18-Nov-2014. [Online]. Available: http://www.college-nat-audio.fr/fichiers/img94a.pdf.

[42] S. J. Schlittmeier, T. Weißgerber, S. Kerber, H. Fastl, and J. Hellbrück, "Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength," *Atten. Percept. Psychophys.*, vol. 74, no. 1, pp. 194–203, 2012.

[43] A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva, and E. Keskinen, "Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices," *Acta Acust. United Acust.*, vol. 97, no. 4, pp. 641–655, 2011.