Masking Speech Noise to Reduce the Sleep Onset Latency: EEG Response Strength Predicts Disturbance

Femke Haga
s0609978
Master Thesis Artificial Intelligence
Radboud University, Nijmegen, The Netherlands

Philips Research Eindhoven supervisors:
Dzmitry Aliakseyeu, Paul Lemmens, and Mun Park
Radboud University Nijmegen supervisor:
Jason Farquhar

August 31, 2012

Abstract

A multiple sleep latency test (MSLT) was carried out to determine the impact on sleep onset latency (SOL) of disturbing speech sound and speech that was masked with rain, SOL was measured using polysomnography (PSG) and electroencephalography (EEG) responses were time-locked to designated parts of the disturbing speech. For all sound conditions subjectively perceived disturbance and speech intelligibility were measured. The analysis showed that sound condition did not have a significant effect on the SOL, but a trend was found that speech increased the SOL, and masking speech reduced the SOL, soft rain appeared to mask more effectively than loud rain. Sound condition did have significant effect on disturbance: Speech was rated as more disturbing than No Sound, Soft mask was equally disturbing as Speech, and only the Loud mask reduced the disturbance. The same effect was found for subjective intelligibility. Global prototype EEG responses showed event-related potential (ERP) components N100, N200, and P300 in the Speech and Soft mask condition, but not in the Loud mask condition. This pattern corresponded the most with subjective intelligibility, which was related to subjective disturbance, and therefore the found EEG response strengths can be a predictor for disturbance.

Keywords: Noise masking, subjective disturbance, speech intelligibility, sleep onset latency, MSLT, prototype EEG response, event-related potential, auditory evoked potential.

Contents

1 Introduction

Some people can easily fall asleep in front of the TV, some lie awake because of the ticking of the clock. Certain sounds are perceived as disturbing and can keep people from falling asleep, other sounds can be comforting and perhaps sleep promoting. In unknown environments, such as a hotel room, sounds like the airco, traffic, people walking and talking in the hallway can keep people from falling asleep. There is still not a lot known about the effects of different kinds of sound during falling asleep. Therefore, it is useful to develop a predictor for the disturbance of sound, to eliminate its effect on the sleep onset latency (SOL, the time it takes to fall asleep).

Ouis (2001) summarized some factors in assessing noise effects: acoustical and non-acoustical factors. Acoustical factors included sound pressure level (SPL), duration of exposure, and frequency spectrum. Non-acoustical factors were age, gender, sensitivity to noise, socio-economic situation, time of day, time of year, physiological and psychological state of the person, but also included more cognitive aspects, for example past experiences and top-down expectations. So, there are many factors which can influence if certain sounds are perceived as annoying.

Noise can lead to annoyance and immediate effects like physiological reactions, this can lead to behavioral modifications (Nelson, 1987). Griefahn et al. (2004) found that noise can decrease the quality of sleeping: the SOL often increases, there are more awakenings, and the REM sleep periods are shorter. People tend to wake up when noise suddenly peaks above a SPL of 55 dB, above a low background noise it will be particularly disturbing. People also tend to wake up when perceived sounds are significant, for instance whispering the sleeper's name can awake the person more easily than louder but neutral sound (Muzet, 2007).

Stanchina et al. (2004) investigated 'whether peak noise or the change in noise level from baseline is more important in inducing sleep disruption.' They used white noise (62 dB of mixed frequency sound) to reduce the magnitude of the peaks in intensive care unit (ICU) noise and tested the effect on the number of arousals. Each night the participants underwent one of the three conditions: (1) baseline, (2) exposure to recorded ICU noise and (3) exposure to ICU noise and mixed-frequency white noise. The number of arousals increased in condition (2) compared to condition (1), but not in condition (3), which proves that not peak noise but the difference between background noise and peak noise determines whether an arousal occurs.

Another study investigated what the effect of continuous white noise was on the sleep onset latency and the night awakenings in college students (Forquer & Merle-Johnson, 2007). Four students received a white noise generator and used it at home. The first few nights measurements were done without masking with white noise (baseline), then some nights were measured when using the white noise (treatment). And later some follow-up measurements were done when the white noise was discontinued (follow-up). All students showed decreased SOLs and less awakenings during the treatment, and in the follow-up the SOLs and awakenings went back to the baseline level. So masking with white noise is effective to reduce the SOL and the number of awakenings, but the effect does not persist when the masking stops.

To reduce the negative effects of disturbing sounds on the SOL, noise masking

can be used. Sounds are added to the noise to make the overall sound less unpleasant and more appropriate for sleeping. A lot of sounds can be used to mask noise, such as white noise, pink noise, and all kinds of natural sounds.

Previous studies at Philips Research showed that in a hotel environment the most disturbing noise that is keeping people awake is the sound of voices (e.g. in the hallway, in the neighboring room, outside), other noises tested were air conditioning noise, traffic noise, and noise from a refrigerator. Some other sounds are perceived as appropriate for falling asleep, the sound of rain was perceived as the most appropriate (Aliakseyeu, Bruin, & Kessels, 2009a; Aliakseyeu, Kessels, Bruin, Chen, & Loenen, 2009c).

Thus speech sound can be used as the disturbing noise, and because of the broad frequency spectrum, rain sound can be used as mask. Water-related natural sounds are one of people's most favorable sounds, because it brings them in the here and now (Andringa & Lanser, 2011).

These sounds can potentially be used to mask the disturbing sounds. However it was not clear if noise masking will have a measurable effect on the sleep onset latency, so the follow up study by Aliakseyeu et al. (2009b) was set up to measure it.

They used a multiple sleep-onset latency test (MSLT; Carskadon et al., 1986), the participants had to fall asleep under tree conditions: (1) with no sound, (2) with disturbing speech noise, and (3) with speech noise and in addition the sound of rain as a masking sound. Electroencephalography (EEG) was used to objectively detect when the participants fell asleep, which was intended to measure the SOL. After the participant fell asleep, they were woken up again. This test was performed during the day. The results of the study indicated that sound masking can have a significant effect on SOL, however due to a limited number of participants the results for conditions with masking and without masking were not statically different.

Speech is known to automatically trigger cognitive processing, therefore introducing an additional variable capable of influencing SOL. Therefore we have to take into account intelligibility of speech. Haka et al. (2009) examined in an earlier study how the intelligibility of irrelevant speech, determined with the Speech Transmission Index (STI), affects the performance on demanding cognitive tasks. STI is a measure of speech transmission quality and is correlated with subjective intelligibility (Steeneken & Houtgast, 1980). They tested three speech conditions: a private office (STI = 0.10), an acoustically excellent open office (STI = 0.35), and acoustically poor open office (STI = 0.65). In their experiment, the STI was adjusted by the relative levels of speech and masking sound. Results showed differences between the three speech conditions. Cognitive performance was disturbed in the environment with the highest STI. Thus, the less intelligible the speech is, the less distracting it is, leading to a better cognitive performance. This could mean that the speech intelligibility can also have an influence on the SOL, but factors governing sleep onset may be different than those for cognitive performance.

To obtain more insight on how noise sounds are affecting the brain when falling asleep, brain data can be analyzed. It is known that if some sounds suddenly stand out against the background noise, a reaction to it is visible in the brain. These event related potentials (ERPs) are described as "voltage fluctuations that are associated in time with some physical or mental occurrence" (Picton et al., 2000). For example, the mismatch negativity (MMN) is an ERP

with a negative peak after 150-200 ms as a reaction to an odd stimulus in a sequence of stimuli (Näätänen & Winkler, 1999).

Reducing the audibility of sound with noise masking, produces less prominent ERPs. High-pass filtered speech sounds caused increased ERP latencies and decreased amplitudes. The effect on N100 was small, larger changes were visible in the MMN (Martin, Kurtzberg, & Stapells, 1999).

A subclass of the ERPs are the Auditory Evoked Potentials (AEP). Early components originate primarily in the auditory cortex bilaterally, which include the P100-N100-P200 sequence (50-200 ms post stimulus). The N100 is a negative wave occurring approximately 100 ms after audio stimulus onset, and is visible at central scalp locations. Late cortical auditory ERPs, such as N200 and P300, are more visible in the fronto-central regions. All these ERPs are involved in auditory perception and language processing (Stapells, 2005).

It was possible to detect perceived music from the brain signal. Schaefer et al. (2011) classified seven different musical fragments using the ERP information. All musical fragments were presented several times, for each fragment a prototype EEG response was made. By a participant who showed the best result, 70% of single trials could be classified correctly from the ERP information, which improved to 100% when multiple trials were combined. Across participants the performance was 53%, this means there were some consistent responses to the musical fragments between subjects.

When analyzing EEG signals, the aim is to measure the brain response to a certain event. Since brain data are by their nature very noisy, several responses to a certain event have to be obtained to reduce the noise impact and to construct a valid average. Noise in the EEG is independent of the event itself, therefore noise be reduced by averaging (Oostenveld, Fries, Maris, & Schoffelen, 2011).

Unexpected sounds are perceived as annoying, because it arouses the auditory modality. People are constantly reminded that the sound can reoccur, the auditory system is preactivated in anticipation to the audio event (Andringa & Lanser, 2011). Unexpected sounds are known to cause more ERP peaks, and therefore stronger EEG responses (Stapells, 2005).

In the present study we investigated whether masking using the natural sound of rain could reduce the negative effects of disturbing speech on the SOL. We tried to quantify the impact of disturbing speech in three ways: (1) assessment of perceived disturbance and intelligibility, (2) SOLs, and (3) ERPs.

2   Research Questions

The goal of this project is to answer the following questions:

- Can masking speech noise with rain sound reduce the sleep onset latency (SOL)?

- Does masking speech noise with rain sound reduce the EEG response strength?

- Can you predict if the participant was disturbed by the sounds from the measured EEG?

3 Method

3.1 Participants

The participants were 16 (8M) Philips employees or interns with a mean age of 26.75±6.45, ranging from 23 to 49. All were volunteers and the interns (N = 14) received gift vouchers as compensation for participation. They were all prescreened for being normal sleepers, using the Pittsburgh Sleep Quality Index (PSQI ≤ 5) (Buysse, Reynolds, Monk, Berman, & Kupfer, 1988), and the Sleep Disorders Questionnaire (SDQ ≤ 2.5) (Violani, Devoto, Lucidi, Lombardo, & Russo, 2004). None of the participants had any hearing problems. We ensured that the five days before the test day participants had a regular sleep pattern, we checked this using a sleep diary and an Actiwatch Spectrum (Philips Respironics). To control for food which can affect sleep, the participants were instructed to conform to a restricted diet on the test day and on the night before: no alcohol, caffeine, or chocolate.

3.2 Stimuli

To answer the question if masking speech noise with rain sound can reduce the SOL a MSLT was conducted using a repeated measures design. In a MSLT study participants receive several opportunities to fall asleep during the day, each session is terminated within 30 min. So for each session at least 30 minutes of audio was needed. The period between the naps had to be at least one hour, so four naps would already fill a day. Therefore we chose to have four sound conditions: (1) One baseline condition with no sound; (2) one with speech noise; and then two kinds of noise masking, (3) one using soft rain as masking sound, and (4) one using loud rain as masking sound.

The sound conditions were chosen in such way that not only the SOL study was possible, but also the EEG analyses were possible. For the SOL study real life had to be simulated, therefore natural sounds were needed. Repeating audio samples were needed for the EEG analyses, because the more brain responses to the same audio parts are available, the better average response can be calculated.

To be able to meet these requirements, an audio fragment with disturbing speech was made from scratch. An interview-style kind of speech was chosen with questions and a lot of 'yes' and 'no' answers. The 'yes' and 'no' sounds were used as the repeating sounds, and brain responses to those sounds were used in the EEG analyses. The questions were unique and gave sufficient variation to make the speech sound naturally. The speech was recorded by three male native speakers of British English in a soundproofed room using a Roland R-26 portable recorder, sampling at 44100 Hz. Two of them were asking the questions, the third one recorded one time a 'yes' and one time a 'no'. The audio fragments of 'yes' and 'no' were 950 ms long, the actual sound in both started after 300 ms, lasted for 370 ms, and ended 670 ms after the start of the audio fragment (for visual representation see Figure 1).

These fragments were adjusted using Audacity 1.3.14 - beta (http:// audacity.sourceforge.net). With Matlab (Mathworks) the answers were concatenated to the questions in a random way, to ensure that each part of the 30 minutes of speech had about the same number of yes's and no's. In total there were 198 yes's and 197 no's. To make the speech sound as if it was heard in a hotel
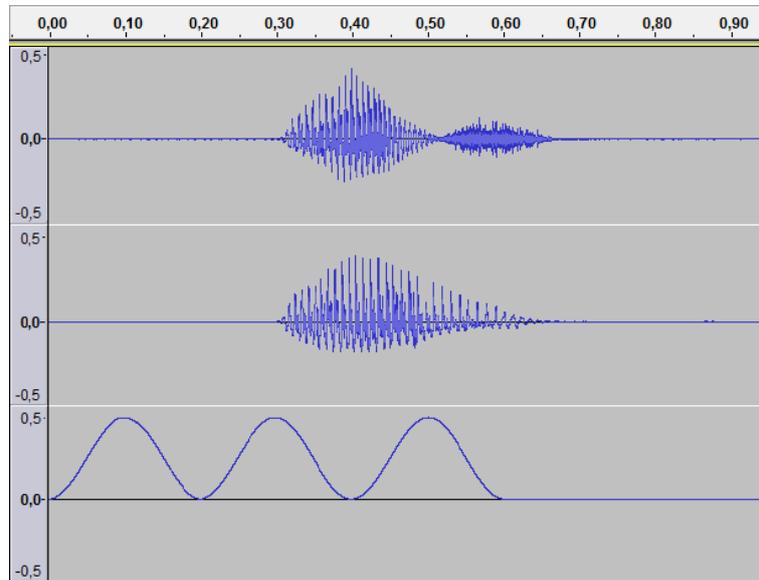
Figure 1: Visual representation of audio: 'yes' (upper), 'no' (middle), and pulses used as markers for EEG analyses.

environment, it was filtered to make it sound as if heard through a gypsum board wall (Halliwell, Nightingale, Warnock, & Birta, 1998). The STI of the speech sample was 0.87, after the filtering, when averaged over non-overlapping 30-second frames.

For the two masking conditions the sound of rain was added to the speech. The rain sample previously used by Aliakseyeu et al. (2009b) was used. The rain reduced the intelligibility of the speech. We chose to use a soft mask (STI = 0.40), and a loud mask (STI = 0.25). Both masking conditions were also gypsum board filtered.

The ambient noise level in the sleep lab was 27 dBA. The speech was calibrated to 32 dBA using a sound level meter (Extech HD600). The soft mask was 30 dBA, and the loud mask 41 dBA. For an overview of the sound conditions see Table 1.

Table 1: The four used sound conditions in the sleep study with the corresponding sound pressure levels of the speech and the rain, the Signal-to-Noise Ratio, and the Speech Transmission Index.

Sound condition	Speech	Rain	SNR	STI
A No Sound	–	–	–	–
B Speech	32 dBA	–	–	0.87 (excellent)
C Soft mask	32 dBA	30 dBA	+2 dB	0.40 (poor)
D Loud mask	32 dBA	41 dBA	−9 dB	0.25 (bad)

In Audacity a stereo audio file was made for the Speech, Soft mask, and Loud mask condition. The audible sound was put on the left channel, only this channel was audible for the participants, and was in fact mono sound. Every time a 'yes' or a 'no' occurred in the speech, three audio pulses of 5 Hz (see Figure 1) were put on the right channel. Besides the PSG channels, also this right audio channel was recorded by the Vitaport EEG recorder to be able to synchronize the audio with the EEG data. 5 Hz audio pulses were used instead of the original audio (44100 Hz), because only low frequencies (under 256 Hz) were measurable in the EEG recorder. The pulses were also used as markers in the EEG analyses.

The procedure was the same for every participant, except the order of the sound conditions. To reduce the influence of unrelated factors, the conditions were balanced using a digram-balanced Latin Squares design, see Table 2. For four conditions only four orders were required to ensure that all irrelevant influences of order effects or carry over effects were fully counterbalanced.

Table 2: The sound conditions were balanced using a digram-balanced Latin Squares design.

Order	Nap 1	Nap 2	Nap 3	Nap 4
1	A	B	D	C
2	B	C	A	D
3	C	D	B	A
4	D	A	C	B

3.3 Equipment and procedure

The study was carried out in a sleep lab where a natural environment is simulated (Philips Experience Lab, High Tech Campus, Eindhoven). The bedroom was equipped equipment that enabled following the sleeping behavior of the participants. The bedroom and the equipment were arranged in such way, that participants noticed as little as possible of the data collecting. Therefore, participants had the most natural sleep as possible in a sleep lab. For the same reason there was also a living room with kitchen and bathroom, to make the feeling of being in a home or hotel environment complete (see Figure 2).

Each participant arrived at 9:00 AM in the apartment. They were informed, signed an informed consent, and changed into night clothing. Each participant was fitted with a PSG, or sleep EEG, set-up. PSG is an abbreviation for PolySomnoGraphy, which can be translated to 'multiple sleep signal measurements'. The brain activity was recorded with 8 sensors on the head (Sleepnet, BrainNet). Six EEG electrodes were placed, two frontal (F3 and F4), two central (C3 and C4), and two occipital (O1 and O2). Two references were placed on the mastoids (A1 and A2). The EEG data were referenced to the mastoid, with the lowest impedance. The sensors were fixed in a headband, conductance gel was used on the sensors to make the contact to the skin better. The facial muscle activity and eye movements (left EOG, right EOG, and chin MEG) were also recorded using disposable gel based KittyCat electrodes (Medi-Trace). For

Figure 2: The apartment where the sleep study was carried out.

recording the PSG data, we used a Vitaport 3 PSG recorder (TEMEC Instruments). The Vitaport was connected to a computer in the control room using an Optilink fiber optic cable.

In the bedroom, a loud speaker was positioned one meter from the pillow on the bed, see Figure 3. A Philips GoGear Connect 3 in the control room was used to play the pre-recorded sounds. The player worked on batteries, therefore there was no connection between the power line and the participant.

We used a microphone in the bedroom, to record the ambient sound. This microphone was connected to the computer in the control room, where Audacity was used to record and to play the sound. The ambient sound was checked for unexpected external sounds, which could disturb the experiment. The recorded sound could be used in a later stage for analyses if something extraordinary happened.

The four naps started for all participants at 10:00 AM, 11:30 AM, 1:00 PM, and 2:30 PM, to keep the influence of circadian rhythm constant. The bedroom temperature was controlled and kept constant to 18 degrees Celsius.

Each nap the participant was put to bed, the connections and the PSG signal were checked, and the participant was asked to fall asleep. The experimenter sat in the control room, see Figure 4. The lights in the bedroom were turned off using a remote control, and the sound fragment was started. The SOL was measured from this time point until the participant fell asleep.

The Columbus software package (TEMEC Instruments) was used to monitor the sleep online. The experimenter looked at the PSG signal to score when the first sleep stage begins according to the MSLT guidelines (Carskadon et al., 1986). This was when for three 30-seconds epochs in a row the following rules were true: the alpha rhythm in the occipital channels was not visible in more than 50% of the epoch, and rolling eyes movements were visible in the EOG

Figure 3: The bedroom.

channels. When the participant fell asleep, the SOL measurement was ended and the participant was awoken. If the participant slept too long it would have had influence on the SOL in the next condition. If the participant did not fall asleep within 30 minutes, the trial was also over, and the SOL was set to 30 minutes.

Before and after the nap the participant was asked to fill in two question-naires, the Karolinska Sleepiness Scale (Akerstedt & Gillberg, 1990) and Global Vigor and Affect (Monk, 1988).

After each nap the participant was disconnected from the PSG recorder, brought to the living room, and instructed to do light office work until the next nap. The break between naps had to be long enough to reduce the influence of the previous trial.

At noon the participants were provided with a light sandwich lunch. Until the last nap the participants could not have any chocolate, alcohol or caffeine.

After the last nap, the participants received a snack, and before telling what the sound conditions were, they were asked to fill in another questionnaire. For every condition they had to assess if the sound was disturbing, and if the speech was intelligible and understandable. The scales consisted of seven items from 'totally disagree' to 'totally agree', resulting in a score between 1 and 7. Participants also had to indicate how noise sensitive they were on a visual ana-logue scale of 100 mm, rating from 'not sensitive at all' to 'extremely sensitive', resulting in a score ranging from 0 to 100.

After filling in the questionnaire one more measurement was done, an awake session. The participant was put to bed, and connected to the PSG recorder. EEG data was recorded while the participant was listening again to the speech, but now was trying to stay awake. To keep the participants from falling asleep, they had to count the yes's and no's in the conversation. Every seven minutes

Figure 4: The control room.

the speech was paused, and the participants were asked how many yes's and no's they counted. This was repeated two times, ending up with three blocks of seven minutes.

After the awake session the participants were disconnected, and the sensors were removed. All measured data were saved and the sleep lab was cleaned.

3.4 Analyses

Some participants had to be excluded from further analysis. In the EEG of participants 6 and 7 no clear alpha signal was detectable during live assessment of the SOL; they slept too long in all the naps, and they had to be excluded in the SOL analysis (N = 14, 7M). Sleeping longer could have had an influence on the perceived disturbance and intelligibility, therefore only these 14 participants were included in the analysis of questionnaires. In the EEG analysis participants 2, 5, 7, 11, and 16 were excluded (N = 11, 4M), because time alignment was not possible because of too noisy data. In the similarity analysis participant 8 was excluded, due to having too few speech samples in the nap conditions (N = 10, 3M).

3.4.1 Analysis of questionnaires

The ratings for disturbance of sound and intelligibility of speech were analyzed with SPSS (IBM, SPSS Statistics 17.0.1) using a GLM repeated measures ANOVA. Sound condition was used as within-subject factor, to address the question of whether the sound conditions had effect on these subjective measurements.

The ratings for sleepiness, global vigor, global affect, and noise sensitivity were not analyzed for this study.

3.4.2 Sleep Onset Latency analyses

A blind post-hoc verification of the SOL scores was done using Somnologica Studio 5.1.0 (Embla Systems Inc.). The verified SOL data was analyzed with SPSS. To answer the research question if noise masking reduced the SOL, a within-subjects design was used. The effect of the sound conditions was calculated using a repeated measures ANOVA.

3.4.3 EEG analyses

A microphone channel with pulses was recorded parallel to the other PSG and EEG channels, therefore the pulses could be used for time-locking when the 'yes' and 'no' fragments were audible for the participants. The EEG responses to these fragments were averaged to make prototype EEG responses.

Each participant underwent four naps and one awake session. In the No Sound condition there was no sound, so also no 'yes' and 'no' fragments were an EEG response to could follow. Therefore the No Sound condition was not further analyzed. The 'yes' and 'no' fragments were present in the other three nap conditions, and therefore they could be analyzed. I will refer to these conditions in the rest of the article as: Nap Speech, Nap Soft mask and Nap Loud mask. And the data recorded in response to the speech sound in the awake session, I will refer to as Awake Speech. For all four conditions a prototype EEG response was made for the yes condition, and one for the no condition. For an overview of the conditions used in the EEG analyses, see Table 3.

Table 3: Overview of conditions used in EEG analyses, and the number of trials used for the global averages per audio event and per condition.

Condition	Audio event	# Trials	# Trials Total
1 Nap Speech	'yes'	1112	2210
	'no'	1098	
2 Nap Soft mask	'yes'	797	1596
	'no'	799	
3 Nap Loud mask	'yes'	930	1854
	'no'	924	
4 Awake Speech	'yes'	1561	3124
	'no'	1563	

Fieldtrip, an Open Source Matlab software toolbox for MEG and EEG analysis (Oostenveld et al., 2011) was used to preprocess and analyze the data. The expectation was that the strongest responses in the ERP would visible in Awake Speech, a bit less strong in Nap Speech, and with increasing masking, decreasing EEG peaks. To be able to read in the data, the raw data had to be converted to .edf format using a raw to edf converter (TEMEC Instruments).

Temporal alignment During the test days the EEG was continuously recorded, but only a fraction of the EEG data was relevant. We saved relevant chunks of data in separate matrices.

The microphone channel of one chunk was loaded, and the starting points of the pulses were detected. One set of pulses contained three peaks, and was always 156 samples long. A peak detector was made to detect the peaks of the pulses. The average sample number of every three sequential peaks was calculated, the starting point of the pulses was 78 samples before this midpoint, which corresponded with the starting point of the 'yes' and 'no' audio files. If the pulses were clipped (the peaks are not sharp, but flat), then cross-correlation was used to determine where the pulses started.

When all starting samples of the pulses were determined, the intervals in between them were checked if they were as expected. When the speech sound was made, the sample numbers of where the yes's and no's were placed were saved. However, because the sampling rate of the audio was 44100 samples per second, and EEG was sampled at 256 Hz, the indices had to be down sampled to 256 samples per second. Less than three samples difference was counted as a correct interval. If an interval deviated from expected then too many peaks were detected by the peak detector. Those peaks had to be removed. The pulses were aligned and plotted to make the alignment visible. If the alignment was correct, the indices were saved. The indices belonging to yes's were saved separately from the indices belonging to the no's.

The found starting samples in the chunks were converted to sample numbers usable in the full data. Then the triggers for the audio events ('yes' or 'no') were known. Only the triggers between lights off and 90 s after sleep onset were used. Then trial tables were made with the begin samples of the trials (0.5 s before the trigger), the end samples of the trials (2 s after the trigger), and the trigger offset (0.5 s after the begin of the trial). The total duration of one trial is 2.5 s, this was chosen because ERPs can be expected within this period.

All the trial tables of one participant were read, and used to load only the trial data from the full data. The trial data was put in a $time \times channels \times trials$ matrix. Information about the condition number was added, to be able to separate the data later again in four conditions.

Preprocessing and averaging Preprocessing was applied to remove the noise from the trial data. The power line noise was removed using a discrete Fourier transform filter. The trials which contained obvious movements or spikes were removed using Fieldtrip's databrowser. Then independent component analysis (ICA) was applied to remove chin movements and EOG artifacts. After that, baseline correction, detrending to remove slow drift, and a 15 Hz low-pass filter was applied.

For every participant eight averages were calculated, one for every condition, to make prototype EEG responses. The number of usable trials for averaging was dependent on how many trials were collected and on how noisy the data was. Therefore the averages were not based on the same number of trials. For example, the average response to 'yes' of participant 9 in the Nap Speech condition was based on 156 trials, and the average response to 'no' of participant 15 in the Nap Soft mask condition was based on 60 trials. The averages for yes and no were combined and used in the similarity measurements.

Also global (cross-subject) averages for all conditions were calculated. The global averages for yes and no were also combined, to obtain one global average for every condition. In Table 3 can be found on how many trials these global

averages were based.

Similarity measurements　The EEG prototype responses from the four conditions were compared with each other. We measured the similarity between the awake session and the nap conditions, and between Nap Speech and the two mask conditions (see Table 4. The transposed average trial data matrix of the Awake Speech was multiplied by the average trial data matrix of Nap Speech, and divided by the normalized Awake Speech matrix. This resulted in a similarity measurement of Awake Speech and Nap Speech. The higher the score, the more similar the EEG responses were in the two conditions. These similarity scores were analyzed in SPSS using repeated measures ANOVAs.

Table 4: Overview of conditions that were assessed on similarity.

Awake Speech	vs.	Nap Speech
Awake Speech	vs.	Nap Soft mask
Awake Speech	vs.	Nap Loud mask
Nap Speech	vs.	Nap Soft mask
Nap Speech	vs.	Nap Loud mask

4　Results

4.1　Perceived disturbance and intelligibility

4.1.1　Disturbance

The participants assessed if the sound conditions were disturbing on a scale of seven items from 'totally disagree' to 'totally agree', this resulted in a score between 1 and 7. The mean subjective disturbance for each sound condition and associated standard deviations can be found in Table 5 and are visualized in Figure 5.

Sound condition had a significant effect on the subjective disturbance ($F_{3,11} = 28.621$; $p = 0.000$; $\eta_p^2 = 0.886$). The Bonferroni adjusted pairwise comparisons showed significant differences between all sound conditions, except between Speech and Soft mask ($p = 0.357$).

Speech was rated as more disturbing than No Sound, and only the Loud mask could make the speech less disturbing. If we only look at the mean subjective disturbance scores for Speech, Soft mask, and Loud mask, we can clearly see a downward trend, but the Soft mask did not decrease the disturbance enough.

4.1.2　Intelligibility

The participants assessed if the speech in the sound conditions was intelligible on a scale of seven items from 'totally disagree' to 'totally agree', resulting in a score between 1 and 7. The mean subjective intelligibility for each sound condition and associated standard deviations can be found in Table 6 and are visualized in Figure 6.

Table 5: The mean subjective disturbance for each sound condition and their standard deviations.

	No Sound	Speech	Soft mask	Loud mask
N	14	14	14	14
Mean	1.14	5.36	4.57	3.43
Std. Deviation	0.54	1.22	1.56	1.87
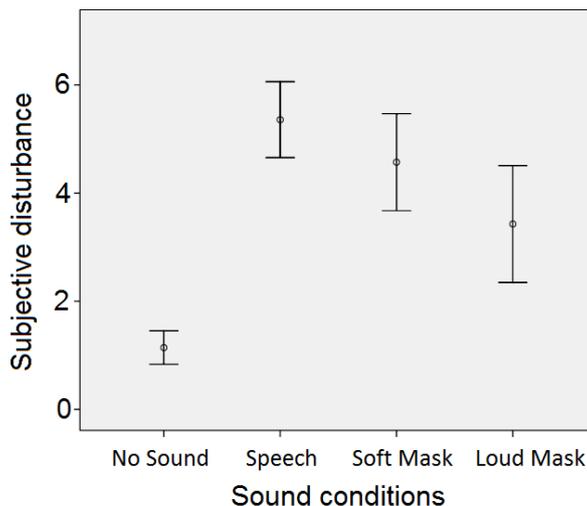
Figure 5: The mean subjective disturbance for each sound condition with error bars (95% confidence interval for mean).

Sound condition had a significant effect on the subjective intelligibility ($F_{3,11} = 46.698$; $p = 0.000$; $\eta_p^2 = 0.927$). The Bonferroni adjusted pairwise comparisons showed significant differences between all sound conditions, except between Speech and Soft mask ($p = 0.101$).

So the same pattern was found for intelligibility as for disturbance. Speech was rated as more intelligible than No Sound (which was obvious, since there was not speech in No Sound), and only the Loud mask made the speech less intelligible. Here was also a downward trend found: the more masking, the less intelligible the speech. But in the Soft mask condition there was too much variance between participants to make the difference significant. Participants rated the intelligibility in the Speech condition just above average ($4.71 > 4$), meaning that they thought the speech was not clearly intelligible, but also not unintelligible. The speech in the Loud mask condition was rated as almost unintelligible (1.86).

4.2 Sleep Onset Latency results

Participants had four opportunities to nap, and it was measured how long it took them to fall asleep based on analysis of PSG. A repeated-measures ANOVA

Table 6: The mean subjective intelligibility for each sound condition and their standard deviations.

	No Sound	Speech	Soft mask	Loud mask
N	14	14	14	14
Mean	1.00	4.71	3.79	1.86
Std. Deviation	0.00	1.14	1.53	0.77
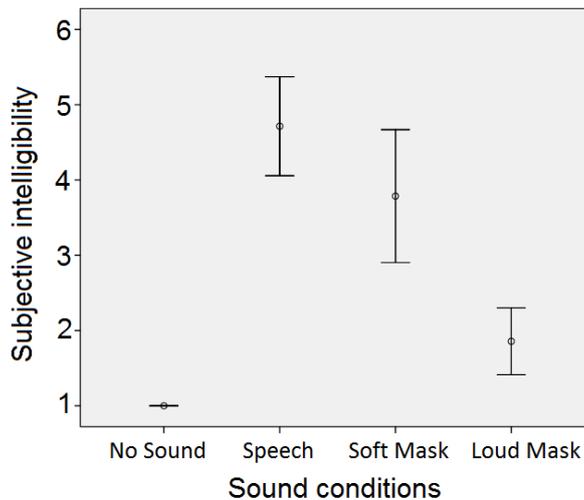
Figure 6: The mean subjective intelligibility for each sound condition with error bars (95% confidence interval for mean).

was used to determine if there was a time effect. The time when the nap was done did not have a significant effect on the SOL ($F_{3,11} = 0.935$; $p = 0.457$; $\eta_p^2 = 0.203$).

The mean SOL for each sound condition and associated standard deviations were calculated, and can be found in Table 7 and are visualized in Figure 7. As can be seen, there was a trend found that SOL was increased when adding speech if compared with no sound, and adding a rain mask was decreasing the SOL again. The soft mask decreased the SOL more than the Loud mask. But there was also a lot of variance found in all sound conditions.

The means were compared using a repeated measures ANOVA to check the effect within participants. The sound condition did not have a significant effect on the SOL ($F_{3,11} = 1.879$; $p = 0.192$; $\eta_p^2 = 0.339$). Speech did not significantly increase the SOL in comparison with No Sound, and the two masks did not significantly reduce the SOL. But there was a strong effect size found.

4.3 EEG results

The pulses were used for temporal alignment, the starting sample of one set of pulses was used as trigger for one trial. Each trial started 0.5 s before the

Table 7: The mean Sleep Onset Latency for each sound condition and their standard deviations.

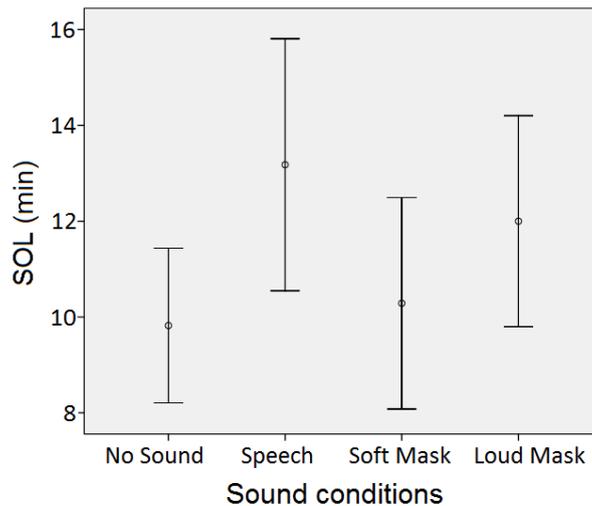|	No Sound	Speech	Soft mask	Loud mask
N	14	14	14	14
Mean	9.821	13.179	10.286	12.000
Std. Deviation	6.037	9.842	8.257	8.239

Figure 7: The mean Sleep Onset Latency for each sound condition with error bars (one standard error).

trigger, and ended 2 s after the trigger. The same pattern of peaks was visible in many trials, so therefore all trials were averaged to make prototype EEG responses. They were made for all conditions, and for all participants. All the prototype responses were plotted to see if any ERPs were visible in the EEG channels F3, F4, C3, C4, O1 and O2.

Visually, we observed a few expected peaks right after audio onset, so the audio appeared to have influence on the EEG. But also a lot of alpha waves (8-15Hz) were visible, which makes sense because alpha waves were there because the participants had their eyes closed most of the time during recording of the data. In the awake session participants were asked to keep their eyes closed, but when it was difficult to stay awake they could open their eyes. This meant that the participants blinked more with their eyes.

To remove more noise and to get a better visualization of the data, global averages were made. The global average response to 'yes' can be found in Figure 8. Starting at time point zero, the point of the trigger onset, until 1.5 s after the trigger, to get a better view of the peaks. The dotted vertical line at 0.3 s indicates where the audio of 'yes' or 'no' started.

Negative peaks are visible after 0.4-0.6 s (100-300 ms post audio onset (PAO)), and positive peaks after 0.6-0.9 s (300-600 ms PAO) in channels F3,
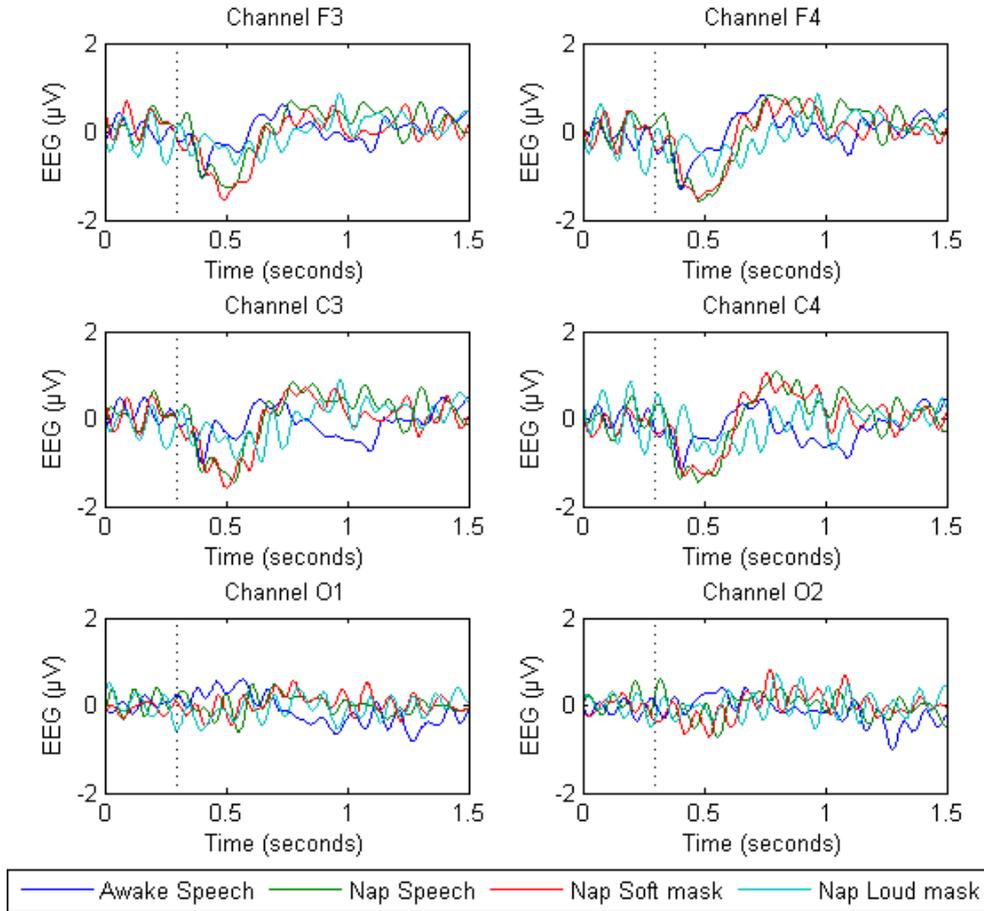
Figure 8: Prototype EEG responses to 'yes', global average, low-pass cut-off = 15 Hz.

F4, C3 and C4, but not in the occipital channels O1 and O2. Auditory and cortical responses were expected in the anterior sites, and not in the posterior sites. So, the responses were as expected.

The prototype EEG responses to 'no' looked the same as the ones for 'yes'. So it was decided to add the trials for yes and no together and to make one response for every condition, to make the signal smoother. We also used a low-pass cut-off of 8 Hz instead of 15 Hz to remove the alpha waves. See Figure 9 for the global average prototype EEG responses to 'yes' and 'no' combined for all four conditions.

In the anterior channels there were negative peaks found around 100 ms and 200 ms PAO, and a positive peak 300-600 ms PAO. These peaks match with the N100, N200, and P300 ERP components, which are AEPs involved in the perception of auditory stimuli and in language processing. This pattern was visible in Nap Speech and in Nap Soft mask, but was not there in Nap Loud mask. For all three nap conditions there were no big peaks visible in the
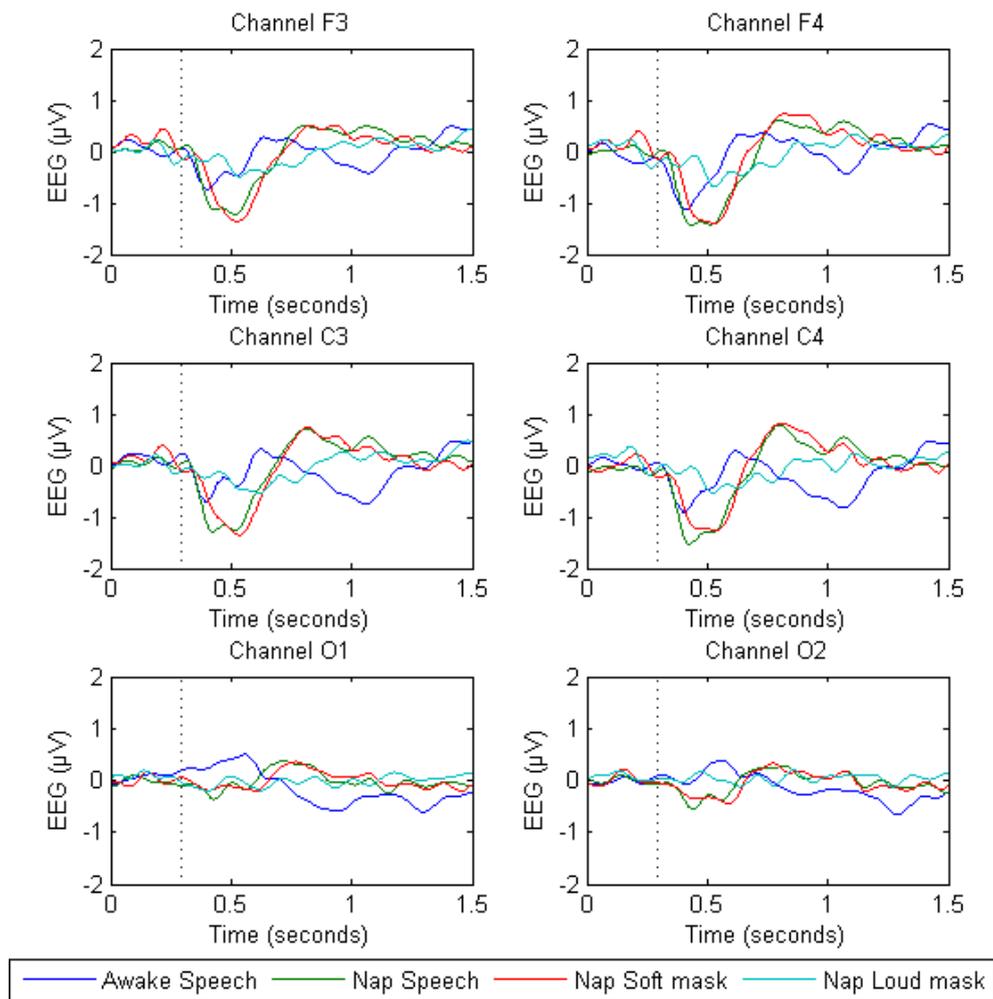
Figure 9: Prototype EEG responses to 'yes' and 'no' combined, global average, low-pass cut-off = 8 Hz.

occipital channels.

The N100, N200 and P300 waveforms were also visible in the Awake Speech condition, but were not as strong as in the nap conditions. In Awake Speech also a negative peak was found around 1 s post trigger (400-900 ms PAO), which could be a N400 ERP component.

Per participant the four prototype EEG responses (yes and no combined, low-pass cut-off = 8 Hz) were compared with each other using the similarity measurements. Only participants who had more than 50 trials in each of the conditions were used in this analysis.

Awake Speech was compared with Nap Speech, Nap Soft mask, and with Nap Loud mask. The similarity between Awake Speech and Nap Speech was on average 10.16 ± 18.15, which means that the two responses were quite similar to

each other. The similarity between Awake Speech and Nap Soft mask was on average 7.03 ± 14.15, which means that the two responses were less similar to each other. The similarity between Awake Speech and Nap Loud mask was on average 0.47 ± 9.90, which means that the two responses were not that similar to each other. The differences between these three similarities was marginally significant ($F_{2,8} = 3.187$; $p = 0.096$; $\eta_p^2 = 0.443$).

We also compared Nap Speech with Nap Soft mask, and with Nap Loud mask. The similarity between Nap Speech and Nap Soft mask was on average 17.71 ± 13.31, which means that the two responses were very similar to each other. The similarity between Nap Speech and Nap Loud mask was on average 2.12 ± 8.57, which means that the two responses were not that similar to each other. The difference between these two similarities was significant ($F_{1,9} = 19.293$; $p = 0.002$; $\eta_p^2 = 0.682$). Thus the EEG response in Nap Speech was significantly less similar to the response in Nap Loud mask, than to the response in Nap Soft mask.

5 Discussion

Disturbing speech can have negative effects on the SOL, a sleep study was conducted to determine the impact of disturbing speech sound and speech that was masked with rain. SOL was objectively measured using polysomnography (PSG) and electroencephalography (EEG) responses were time-locked to the yes's and no's in the disturbing speech. For all sound conditions subjectively perceived disturbance and speech intelligibility were measured. Prototype EEG responses were calculated for every condition, and compared using similarity measurements. Global EEG responses were visualized, and the found ERPs were analyzed.

The analyses showed that sound condition did not have a significant effect on the sleep onset latency. Although Speech was evaluated as more disturbing and more intelligible than No Sound, Soft mask as disturbing and as intelligible as Speech, and Loud mask less disturbing and less intelligible than Speech. The EEG analyses showed that the yes's and no's in the speech evoked ERPs. In Awake Speech the ERP components N100, N200, P300, and N400 were visible. In the nap conditions the N400 component was not visible anymore. The N100, N200 and P300 components were stronger in Nap Speech and in Nap Soft mask, and were gone in the Nap Soft mask condition. The EEG response in Nap Speech was significantly less similar to the response in Nap Loud mask, than to the response in Nap Soft mask. These results were in line with the perceived disturbance and intelligibility. The results can now be used to formulate answers to the research questions.

Can masking speech noise with rain sound reduce the sleep onset latency (SOL)? This study showed that masking speech noise with the used rain masks could not significantly reduce the SOL. However, a strong effect and a trend were found. Speech increased the SOL if compared with No Sound. The Soft mask decreased the SOL if compared with Speech, and the Loud mask also decreased the SOL if compared with Speech, but not as much as Soft mask.

This contradicts with the perceived disturbance and intelligibility: The Loud mask was found less disturbing and intelligible than Soft mask. A contradiction

between subjective and objective disturbance was also found in previous studies. For example, Haka et al. (2009) found that subjective disturbance ratings showed a more linear relation with STI, than the objective measurements.

Does masking speech noise with rain sound reduce the EEG response strength? In the Speech condition the ERP components N100, N200, and P300 were visible, they were also visible when using the Soft mask. But when using loud rain to mask speech, the EEG responses were strongly decreased. This means that masking speech with rain sound can reduce the EEG response strength. The known relation between SNR of stimulus and the amplitude of ERP, is therefore confirmed in the current study.

The ERP components N100, N200, and P300 are all involved in the perception of auditory stimuli and language processing. So in the Speech and Soft mask the auditory stimuli ('yes' and 'no') are perceived, and the language is processed. In the Loud mask condition there are no clear ERPs visible anymore, implying the yes's and no's are not perceived anymore. The Loud mask masks the speech enough to let the participants only focus on the rain. The auditory stream is not evaluated as speech anymore.

In the awake session the N100, N200, and P300 components were smaller, which could mean that because participants paid more attention to the speech, the yes's and no's became more predictable. There was also a late negative peak found 400-900 ms PAO, this could be a N400 ERP component, which is elicited when processing the meaning of words (Kutas & Federmeier, 2000). The absence of N400 in the nap conditions, could mean that the meaning of the words is not processed when trying to fall asleep.

Can you predict if the participant was disturbed by the sounds from the measured EEG? The EEG response strength is reduced by the Loud mask and not so much by the Soft mask. This pattern is also found in the perceived intelligibility evaluations. In these evaluations, a big difference was also found between Soft mask and Loud mask. So the more intelligible the words are, the stronger the EEG responses.

Perceived disturbance is related to the intelligibility of speech (Haka et al., 2009). With the EEG response strength, the SNR and then also the intelligibility can be predicted, and therefore the subjective disturbance as well. This makes the EEG response strength a predictor for subjective disturbance.

To summarize, masking speech with rain cannot reduce the SOL, but can reduce the perceived disturbance and intelligibility. It can also reduce the EEG response strength, which implies that it can be a predictor for disturbance.

A strong effect size in the SOL analysis was found ($\eta_p^2 = 0.339$), meaning that 33.9% of the variance was caused by the within-subjects factor sound condition. This indicates that the effect could be significant when the experiment is run with more participants. The biggest difference in SOL was between no sound and speech. A post-hoc power calculation was done with GPower that showed that the difference between No Sound and Speech would have been significant if the sample size had been 36.

It took the participants longer to fall asleep in the Loud mask condition, than in the Soft mask condition. This contradicts the subjective results. The SPL

in the Loud mask condition was 9 dBA higher than in the Soft mask condition, it can be hypothesized that this higher volume had a negative influence on the SOL.

A lot of factors other than sound condition are influencing the SOL, as described by Ouis (2005). The factors gender, sleepiness, vigor, affect, and noise sensitivity were measured during the experiment, but not analyzed in this study. In further research, these factors can be analyzed.

The subjective evaluation of Soft mask was not significantly different to the evaluation of Speech, not less disturbing and not less intelligible. This was not expected. Explanations for this can be that the soft rain was not recognized as rain, or that the level of speech transmission, which was 0.40 in Soft mask, was still high enough to be perceived as disturbing. The Loud mask, with STI = 0.25, did have an effect on the subjective results. More values of STI can be tested in further EEG research.

Noise masking still has potential to reduce the negative effects of disturbing sounds on the SOL. More research is needed to determine if noise masking can be used to reduce the SOL. Repeating this study for another 22 participants (this makes a total of 36 participants), can be useful to determine if speech masking with soft or loud rain can reduce the SOL.

Further, instead of doing more MSLT studies, EEG responses to several sound fragments can be compared, to decide which masking can be used at best to reduce the SOL.

To conclude, the EEG response strength does not give a good prediction for the SOL. As said before, this study showed that masking speech noise with the used rain masks could not significantly reduce the SOL, only a trend was found. In that trend the SOL in the Soft mask condition was shorter than the SOL in Speech, the SOL in Loud mask was also shorter than the SOL in Speech, but longer than Soft mask. The pattern in this trend is not explainable using only the EEG responses. More participants are needed to make the found differences in SOLs significant, and more research is needed to find the most optimal mask.

As argued before, other factors probably have influence on the SOL as well. More variables were measured during the experiment, the effect of these variables on the SOL can be studied in further research. Also the effect on the EEG responses can be investigated.

The EEG response strength does give a good prediction for the SNR, for the intelligibility of speech, and therefore also for the perceived disturbance. This makes the EEG response strength a predictor for subjective disturbance.

6 Acknowledgements

Without them, the experiment would not have been possible.

And last but not least, I want to thank Dirk, my parents, my family, and my friends who supported me not only during the last year, but during my entire study.

References

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal Neuroscience*, *52*, 29–37.

Aliakseyeu, D., Bruin, W., & Kessels, A. (2009a). *Undisturbed sleep: evaluation of the relative effectiveness of sound masking* (Tech. Rep. No. 2009/00520). Philips Research Technical Note. (Restricted to Philips internal use.)

Aliakseyeu, D., Bruin, W., Skowronek, J., & Abtahi, S. (2009b). *Undisturbed sleep: reducing the sleep onset latency with noise masking* (Tech. Rep. No. 2009/00710). Philips Research Technical Note. (Restricted to Philips internal use.)

Aliakseyeu, D., Kessels, A., Bruin, W., Chen, W., & Loenen, E. (2009c). *Undisturbed sleep: evaluation of the relative effectiveness of sound masking* (Tech. Rep. No. 2009/00228). Philips Research Technical Note. (Restricted to Philips internal use.)

Andringa, T., & Lanser, J. (2011). *Towards causality in sound annoyance.* (Presented at the Internoise 2011, Osaka, Japan)

Buysse, D., Reynolds, C. F. I., Monk, T., Berman, S., & Kupfer, D. (1988). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Research*, *28*, 193–213.

Carskadon, M., Dement, W., Mitler, M., Roth, T., Westbrook, P., & Keenan, S. (1986). Guidelines for the multiple sleep latency test (mslt): a standard measure of sleepiness. *Guidelines for the Multiple Sleep*, *9*(4), 519–524.

Forquer, L., & Merle-Johnson, C. (2007). Continuous white noise to reduce sleep latency and night wakings in college students. *Sleep and Hypnosis*, *9*(2), 60–66.

Griefahn, B., & Spreng, M. (2004). Disturbed sleep patterns and limitation of noise. *Noise & Health*, *6*(22), 27–33.

Haka, M., Haapakangas, A., Keränen, J., Hakala, J., Keskinen, E., & Hongiste, V. (2009). Performance effects and subjective disturbance of speech in accoustically different office types - a laboratory experiment. *Indoor Air*, *19*, 454–467.

Halliwell, R., Nightingale, T., Warnock, A., & Birta, J. (1998). Gypsum board walls: Transmission loss data. *National Research Council Canada Internal Report*, *IRC-IR-761*, 36.

Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*(12), 463–470.

Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on n1 and the mismatch negativity to speech sounds /ba/ and /da/. *Journal of Speech, Language, and Hearing Research*, *42*(2), 271–286.

Monk, T. (1988). A visual analogue scale technique to measure global vigor and affect. *Psychiatry Research*, *27*, 89–99.

Muzet, A. (2007). Environmental noise, sleep and health. *Sleep Medicine Reviews*, *11*, 135–142.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859.

Nelson, P. (1987). *Transportation noise reference book*. Cambridge: Butterworths.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*.

Ouis, D. (2001). Annoyance from road traffic noise: a review. *Journal of Environmental Psychology*, *21*, 101–120.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152.

Schaefer, R., J., F., Blokland, Y., Sadakata, M., & Desain, P. (2011). Name that tune: decoding music from the listening brain. *NeuroImage*, *56*(2), 843–849.

Stanchina, M., Abu-Hijleh, M., Chaudhry, B., Carlisle, C., & Millman, R. (2005). The influence of white noise on sleep in subjects exposed to icu noise. *Sleep Medicine*, *6*, 423–428.

Stapells, D. (2005). *What are auditory evoked potentials?* (Available from http://www.courses.audiospeech.ubc.ca/haplab/aep.htm)

Steeneken, H. J. M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, *67*(1), 318-326.

Violani, C., Devoto, A., Lucidi, F., Lombardo, C., & Russo, P. (2004). Validity of a short insomnia questionnaire: the SDQ. *Brain Research Bulletin*, *63*, 415–421.