What you Hear is what you See – a study of visual vs. auditive noise

Anna Berg & Annelie Brandt
Department of Linguistics, Stockholm University

Abstract

Is visual noise as much a distractor as audtive noise? From six video sequences with increasing S/N ratio, we have investigated the target of the subjects focus and the relationship between visual and auditive noise, while performing a task. This was done using an eye-tracking system.

The results show that one tends to put the energy into the performance of the task. The auditive noise seems to be more of a trigger for recognizing the visual background disturbance and is therefore a greater distractor.

Introduction

The fact that visual information is decisive when the signal to noise (S/N) ratio is low, is common knowledge (Sumby & Pollack, 1954; Erber, 1969). When the speech signal is corrupted by noise, one tends to use visual cues as compensation (Lansing & McConkie, 1999). Occasionally, visual cues are impaired by other visual interference which we will refer to as visual noise. By using video equipment it is possible to examine the interaction between the auditive and visual signals.

The goal of this experiment is to try to determine the S/N ratio where the observer starts relying on visual cues for information extraction. The study will also investigate the target for the observers' visual focus as well as the relationship between visual and audtive noise.

Our expectations are that the subjects will try to get acquainted with the situation during the first video sequences and therefore will focus on both the characters and the background. As they get more familiar with the situation the subjects tend to focus more and more on the task. When the S/N ratio decreases, visual cues get automatically more important and increased focus on the monitor is expected.

Method

In this experiment a short video recording was used containing two characters describing a practical task. Two actors, A and B, where filmed while actor A verbally explained how actor B should solve a puzzle in form of paper figures. The two actors were not able to see each other. Visual background noise was recorded separately from a shopping mall and subsequently applied to the video. The existing sound in the mall was also used as auditive noise to create a natural atmosphere of competing sounds. The recording was divided in six sequences with decreasing S/N ratio: no noise, 30 dB, 18 dB, 7 dB, 0 dB, - 6 dB.

Twelve subjects participated and were asked to perform the same task as actor B with the instructions from actor A, while watching the video on a monitor. The eye movements were recorded using the software Clear View 2.7.0 and the eye-tracking system Tobii 1750, both developed by Tobii Technology AB.

Results

The results are calculated from nine out of the twelve subjects. Three of the subjects were discarded due to technical problems with calibration of the eye-tracking system.

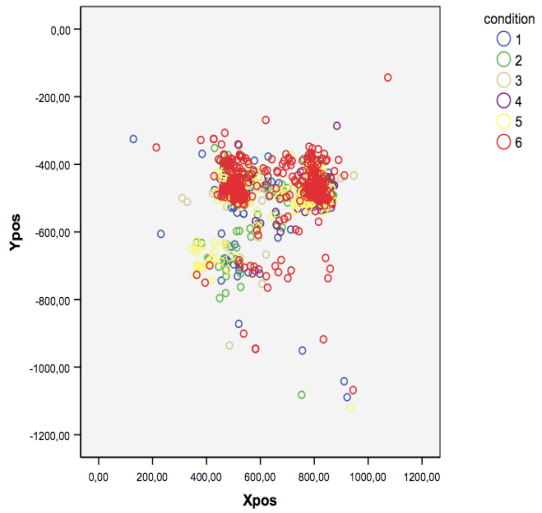

Figure 1. Condition 1-6 refers to the six sequences in the video with decreasing S/N ratio: no noise, 30 dB, 18 dB, 7 dB, 0 dB, - 6 dB. The coloured spots indicate 20 ms fixations when the eye movements are steady within 30 pixels/100 ms.

In figure 1 the values on the x- and the y-axes are coordinates that shows positions of fixation on the monitor. The coloured spots indicate 20 ms fixations when the eye movements are steady within 30 pixels/100 ms. As the results show, one tend to increase the focus on the visual information when the S/N ratio decreases. The fixations in sequence 6 (condition 6) form two wide spread areas where the faces of actor A and B appears on the monitor screen. In this condition where the noise exceeds the speech signal, there are more fixations than in the earlier sequences with more favourable S/N (Fig. 1).

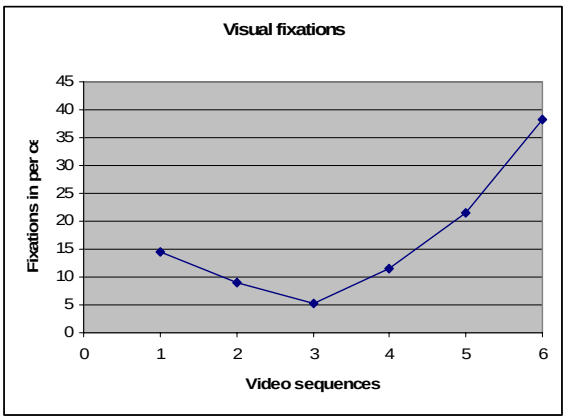

Figure 2. The total amount of visual fixations over the six video sequences are shown in per cent.

Figure 2 shows a large variation of the total amount of informant fixations during the video sequences. During sequence 3 the amount of fixations is at its lowest, increasing thereafter.

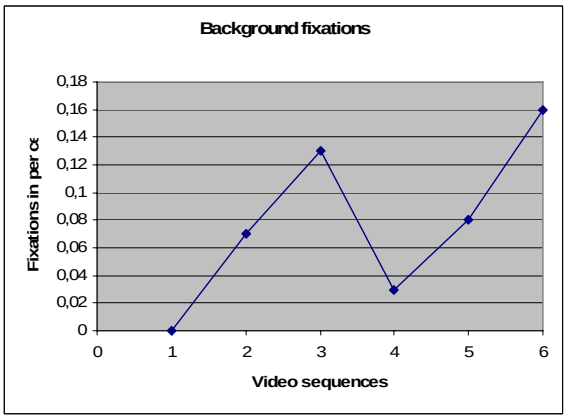

Figure 3. The amount of visual background fixations over the six video sequences are shown in per cent.

Figure 3 shows the amount of background fixations in per cent out of the total fixations shown in figure 2. Over all, the amount of background fixations is very low, with the largest proportion in sequence 3 and 6.

Discussion

Over all, the target of the observers' visual focus tended to be the faces of the actors in the video, just as expected (Fig. 1). The curve in figure 2 confirms our expectations that the subjects would pay more attention to the monitor at the beginning, an attention that was expected to decrease as familiarity with the task increases. This behaviour was observed for S/N up to 18 dB but after that the subjects apparently needed more visual information to solve the problem. For S/N lower than 7dB, there was a significant increase in visual fixations at the monitor. Although the subjects had to rely more on visual cues to extract information as the S/N ratio decreased, they still focused on solving the task without giving up.

In the beginning of the experiment we assumed that most of the background fixations would appear in the first sequences as the auditive signal was fully understandable. The results showed otherwise: the fixations on the background increased at the final sequences when the auditive noise was most disturbing. The reason could be that the natural auditive noise triggers the observer to react and notice the visual noise. If this is the case, the auditive

noise would therefore be more distracting than the visual noise. All in all the amount of background fixations were considerably fewer than expected. In figure 3 there is a peak in the curve at sequence 3 which could seem rather odd. One explanation could be the fact that the total amount of fixations in sequence 3 is very low, resulting in a proportionally higher score of background fixations.

Our attempt to establish at which S/N ratio the observer would start to rely on visual cues to extract information, could not be answered by this experiment. To address that question the experiment should perhaps have used a mixed order of S/N, rather than the systematic decreasing S/N used in the current design.

Conclusion

Confronted with a task to solve, the subjects appear to put all their energy in succeeding with the problem and do not seem to be easily distracted by spurious information, at least in a situation where they are allowed to adapt to the progressively deterioration of the communicative conditions. In this context, visual noise happened to play a minor role in comparison with the auditive noise.

Acknowledgements

Student work done within the scope of the course *Production and perception of speech 2* using the facilities provided by MILLE and CONTACT projects at the phonetics laboratory at Stockholm University.

The authors are grateful for all assistance and advise given by Ellen Marklund and Francisco Lacerda.

References

Erber N (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of speech and hearing research*, 12: 423-425.

Lansing C R & McConkie G W (1999). Attention to the facial regions in segmental and prosodic visual speech perception tasks. *Journal of speech, language and hearing research*, 42: 526-539.

Sumby W H & Pollack I (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of America*, 26: 212-214.