

# The Impact of Audiovisual Input upon the Perception and Comprehension of Speech in Second Language Acquisition

Smiljana Komar

University of Ljubljana, Slovenia

E-mail: smiljana.komar@guest.arnes.si

## ABSTRACT

The paper addresses the impact of audiovisual stimuli upon the perception, comprehension and retention of information retrieved from spontaneous speech in English as L2. Speech production is linear while its perception and comprehension are more complex and depend on the linguistic as well as meta-linguistic context. We will present the results of a listening comprehension experiment carried out with two groups of students of English. One group was exposed to a video recording of a BBC TV cooking show, whereas the other group only listened to it. Both groups had to answer the same questions. The results show that the visual input eases comprehension only when it directly supports information retrieved from the acoustic input. The visual input has a stronger impact than the acoustic which may result in the loss of information retrieved from the acoustic input.

## 1. INTRODUCTION

Psycholinguists generally agree that the recognition of spontaneous speech is not a linear process. Instead it is governed by a very complex cognitive process. Lexical processing depends on two broad classes of information: information retrieved from acoustic features and representations constructed from the context (i.e. lexical, syntactic, semantic, pragmatic). A previous study of perception of spontaneous speech [2] has shown that in listening comprehension various contexts can help listeners to catch the meaning of a speech sequence, helping them to overcome obstacles arising from unfamiliar sounds and words as well as the background noise. The study by Leoni and Cutugno [1] has shown that there should be a balance between the information retrieved from the acoustic signal and the contextual information. If the balance is not achieved, the communication process can fail.

A number of textbooks for the teaching of listening comprehension in L2 [3, 4] stress the importance of guessing and developing grammatical and lexical expectations. Recent publications [5] have begun to stress the importance of visual materials supporting the acoustic stimulus in the comprehension of speech, claiming that the presence of visual materials is immensely important in contextualizing the listening situation.

In the process of teaching listening skills to future

translators and interpreters, we have found out that their understanding of spontaneous speech varies significantly. Since all the training is supported by visual input, we decided to have a closer look at what happens when listeners are exposed to acoustic and visual inputs. In order to analyse the impact of visual material on the comprehension of spontaneous speech, we carried out an experiment with 90 third-year students whose mother tongue is Slovene.

## 2. EXPERIMENT

For the experiment we recorded a 30-minute BBC Prime TV cooking show (*Ready, Steady, Cook*). The show was chosen for three reasons: firstly, its main characteristic is speech-in-action (i.e. cooking and speaking at the same time), secondly, it hosted chefs and contestants who all used standard English pronunciation coloured to some extent by their regional dialects and finally, it was amusing enough to increase listeners' motivation.

The students (90) were divided into two equal groups. One group was exposed to the visual and acoustic material, while the other only listened to the show without being able to see what was going on in the studio. Both groups were given the same 58 comprehension questions which they had to answer as the show proceeded.

The questions were explanatory or WH-questions and quite a number of them were after several pieces of information (e.g. asking students to enumerate the items of food used for the preparation of a dish).

The questions can be divided into three groups: there were 29 questions referring to information which could be retrieved from the acoustic input supported by the same visual input; 12 questions referred to information which could be retrieved from the acoustic input only (the visual input was thought to be disturbing because it provided different information from that of the acoustic input); 17 questions referred to information which could be retrieved from the acoustic input, but the visual input was neither disturbing nor supporting it.

## 3. ANALYSIS AND EVALUATION OF RESULTS

Looking at the results of the listening comprehension

experiment, we can make a general conclusion that the students were not very successful. Table 1 shows that the students who listened and watched (L+W) the show answered correctly to only 51% of all questions, 20% were wrong answers, while 29% of questions were not answered at all.

The students who were exposed only to the acoustic input (i.e. listening L), performed even worse, answering correctly to only 38.9% of questions, giving wrong answers to 21.9% and not answering 38.3% of questions.

	CORRECT	WRONG	NO ANSWER
L+W	51%	20%	29%
L	39.8%	21.9%	38.3%

**Table 1:** Overall results.

The results show a slight increase (11.2%) in correct answers in the group exposed to audiovisual input compared to the results of the group exposed to acoustic input only. While the two groups exhibit almost identical percentage of wrong answers, there is a slight difference (9.3%) between them in failing to provide an answer at all.

One half of the questions (29) referred to information which the students could retrieve from the acoustic as well as visual input. In other words, the action on the screen was supported by the talk about it. Table 2 presents the results that the students obtained in these 29 questions.

	CORRECT	WRONG	NO ANSWER
L+W	58.4%	17.3%	24.3%
L	38.4%	24%	37.6%

**Table 2:** Results in questions referring to information retrieved from the visual input supported by the acoustic input.

A significant increase (20%) in correct answers achieved by the students who were exposed to the acoustic as well as visual input is noticed immediately. The students who only listened, failed to provide an answer in almost the same percentage as they answered correctly, but almost a quarter of answers were wrong.

There were 12 questions which referred to information provided only by the acoustic input; the visual input was different and thus possibly disturbing. Table 3 confirms this.

	CORRECT	WRONG	NO ANSWER
L+W	38.6%	25.2%	36.2%
L	42%	21.8%	36.2%

**Table 3:** Results in questions referring to information retrieved from the acoustic input only (disturbing visual input).

The results show an insignificant increase (3.4%) in correct answers obtained by the students who only listened to the show and were thus not disturbed by the action on the screen. What is surprising, is the fact that the percentage of unanswered questions is the same in both groups.

The remaining 17 questions referred to information retrieved from the acoustic input with no disturbing visual input. In other words, while the speakers were talking there was no distracting action on the screen. Table 4 shows the results.

	CORRECT	WRONG	NO ANSWER
L+W	43.6%	20.1%	36.3%
L	49.5%	9.7%	40.8%

**Table 4:** Results in questions referring to information retrieved from the acoustic input only (no disturbing visual input).

An insignificant increase (5.9%) in correct answers obtained by the students who only listened can be noticed again. However, the percentage of unanswered questions in this group is surprisingly high (40.8%).

The results presented in the above four tables are not convincing enough to lead us to a conclusion on the impact of visual input upon the perception and comprehension of speech in L2. In order to clarify the influence of picture, we shall present in Tables 5, 6, 7 and 8 the results to four questions in which the visual impact was particularly important. The first two questions referred to information retrieved from the acoustic as well as visual input. We expected the students who watched and listened to the show to obtain better results. Results presented in Tables 5 and 6 confirm our expectations.

Answer	L+W	L
2	59.6%	11.7%

**Table 5:** Results to the question *How many dishes has Tony prepared for Rachel?*

Answer	L+W	L
3	77%	41.2%

**Table 6:** Results to the question *How many dishes has Nick prepared for Gillian?*

Results to the above two questions show a significantly better performance by those students who watched and listened to the show and who could actually see the number of dishes on the screen. The listening group could only guess about the exact number of dishes. These results confirm the assumption that the supporting visual material has a positive impact on the perception of speech and enhances its comprehension.

Let us now have a closer look at the answers to two other questions in which the acoustic input was disturbed by the visual. Table 7 shows the results to the question *What did Tony do with one piece of cod?* It has to be mentioned that while Tony is explaining what he did with a piece of cod, the cod is already frying in a pan and Tony is busy taking care that it does not get overfried.

Answer	L+W	L
Removed the skin, put in flour, egg, crumbs	17.3%	33.3%

**Table 7:** Results to the question *What did Tony do with one piece of cod?*

Table 8 presents the results to the question *How did Tony make the chips?* Here, while Tony is explaining the procedure, he is covering a plate with a paper cloth which is then accidentally set on fire.

Answer	L+W	L
Blanched and fried them	46.2%	78.4%

**Table 8:** Results to the question *How did Tony make the chips?*

In answers to both questions we can immediately observe a significant increase in correct answers in the listening group, while the listening and watching group performed worse. This confirms our initial assumption that the visual input can have a distracting or even disturbing influence upon successful comprehension of speech if it does not agree with the acoustic input.

#### 4. CONCLUSIONS

Although the overall success in the listening comprehension test was not very high, we can nonetheless conclude that the impact of visual stimuli upon the

perception and comprehension of speech in L2 is twofold:

- (i) it eases comprehension only when it directly supports information which can be retrieved from the acoustic input;
- (ii) the visual input has a stronger impact on the listener than the acoustic input which very often results in the loss of information present in the acoustic input only.

On the basis of these two conclusions we can safely claim that the teachers who teach listening comprehension in L2 should carefully choose the visual support to help their students in listening comprehension tasks. They should choose such audiovisual material where the acoustic and visual inputs support each other. Any distracting visual input which presents information different from that provided by the acoustic input will have a negative impact on comprehension and will confuse the listeners which will in the end result in the loss of information.

#### REFERENCES

- [1] F. Albano Leoni and F. Cutugno, "The role of context in spontaneous speech recognition," in *Proceedings of The XIVth International Congress of Phonetic Sciences*, vol. 2, pp. 861-864, San Francisco, 1999.
- [2] S. Komar, "Perception of continuous speech in second language acquisition," in *Proceedings of The XIIIth International Congress of Phonetic Sciences*, vol. 1, pp. 298-301, Stockholm, 1995.
- [3] M. Rost, *Introducing Listening*, London: Penguin, 1994.
- [4] P. Ur, *Teaching Listening Comprehension*, Cambridge: Cambridge University Press, 1984.

