

Univerzita Palackého v Olomouci
Filozofická fakulta

DIPLOMOVÁ PRÁCE

2008

Michal Kubánek

Katedra anglistiky a amerikanistiky

Filozofická fakulta

Univerzita Palackého v Olomouci



Michal Kubánek

Context Knowledge Utilization in Word Recognition of Czech
Students of English

Vedoucí práce: Mgr. Šárka Šimáčková, Ph.D.

Olomouc 2008

Prohlášení

Prohlašuji, že jsem tuto magisterskou diplomovou práci zpracoval samostatně pod vedením Mgr. Šárky Šimáčkové, Ph.D. a že jsem uvedl odkazy na všechny použité zdroje.

V Olomouci dne 27. 7. 2008

Acknowledgements

I would like to express my gratitude to Šárka Šimáčková for the initial stimulus and insightful suggestions throughout the thesis processing as well as to Václav Jonáš Podlipský for his valuable comments, assistance in organizing the experiments and help with statistical analyses. Many thanks go to Bronislava Grygová and a group of teachers at Jakub Škoda Grammar School in Přerov for their kind help with subject recruitment.

Contents

1	Introduction.....	1
1.1	Levels of Contextual Influence and Contribution to Perception.	1
1.2	Do non-native listeners utilize context information as well?.....	5
1.3	State-of-the-art technology for investigation.....	8
1.4	A review of related studies.....	10
1.5	Research questions and experiment design.....	17
2	Methods.....	23
2.1	Subjects.....	23
2.2	Materials.....	24
2.2.1	Spoken stimuli preparation.....	27
2.2.2	Recording and manipulation.....	29
2.2.3	Visual stimuli preparation.....	32
2.3	Procedure.....	33
2.3.1	Experiment 1.....	34
2.3.2	Experiment 2.....	35
2.3.3	Experiment 3.....	36
3	Results.....	37
3.1	Experiment 1.....	38
3.2	Experiment 2.....	43
3.3	Experiment 3.....	48
3.4	Native listener.....	52
4	Discussion.....	54
4.1	Experiment 1.....	55
4.2	Experiment 2.....	57
4.3	Experiment 3.....	59
4.4	Cross-experiment analysis.....	61
4.5	Substitutions.....	62
4.5.1	Experiment 1.....	63
4.5.2	Experiment 3.....	66
5	Summary.....	70
	Appendix A.....	79
	References.....	80

1 Introduction

Speech perception represents one of the most exciting and at the same time challenging areas of research in the field of linguistics. The fact that people can communicate in spite of many obstacles observed by the researchers is still surprising. This thesis deals with the complex interplay of different factors which influence speech perception, namely with context utilization of Czech learners of English. The introductory section deals briefly with the problems arising from the nature of spoken communication in general and discusses contextual influences at different language levels. The following parts concentrate on non-native listeners and present relevant studies and techniques which deal with this topic. The crucial part of this section introduces the two main observations which served as a stimulus for this study. Furthermore, it presents the study question, hypotheses and outlines the experiments which are intended to further investigate the topic and approach possible answers.

The experiment itself is described in the following section in detail including description of materials, subjects and procedure. Next, results obtained in the experiments are presented followed by the discussion in which we tried to comment on the results and relate them to our original hypotheses. The thesis endeavors not only to relate the findings to general discussion in this field of research, but also to assess possible deficiencies in methodology and to direct investigation in the future with recommended enhancements.

1.1 Levels of Contextual Influence and Contribution to Perception

An informative overview of issues connected with speech perception such as the overlap of phonemes in speech signal and

the problem of segmentation, the lack of invariance, or the need of normalization across speakers can be found in every book on phonology (see Pickett, 1999, for example).

Several models and theories of speech perception have been developed over the course of study in this field. Traditionally, the approaches are divided into two groups based on their main focus and the mechanism of perception they stress. The first group consists of models which see speech perception as a bottom-up process. According to this approach, listener almost exclusively relies on a thorough analysis of the incoming acoustic signal which comprises of its segmentation and extraction of relatively discrete units such as distinctive features, phonemes, syllables, or words. The other approaches maintain that the process of speech perception is largely aided by information available in higher levels of linguistic competence. Thus listeners make use of their knowledge of morphology, lexicon, syntax, and semantics of their language and also of contextual information originating in the situation of a particular speech act as well as their world knowledge and previous experience.

The influence of context on speech perception was demonstrated by many experiments. Felty (2007) tested the effects of morphological level on word processing. As he predicted, the phonemes in nonwords were recognized more independently of one another than in words. It also took longer time to recognize bimorphemic words than monomorphemic ones of the same length and lexical frequency. This observation is in accordance with the prediction that morphemes are stored separately in the mental lexicon. A monomorphemic word is recognized more readily than bimorphemic one in which case the listener has to wait for the acoustic information about the second morpheme.

The preference of words over nonwords and thus the utilization of lexical level in speech perception was attested in a classical experiment by Ganong (1980). He created continua of words / nonwords based on manipulation of the initial phoneme voice onset time (“kiss” / “giss”; “gift” / “kift”). The ambiguous stimuli at the boundary between the two categories (voiced / voiceless) were judged by the listeners with preference to the word over non-word status.

Another fundamental experiment demonstrates the effect of meaning of the whole sentence on a phoneme restoration. Warren and Warren (1970) obscured the initial phoneme of a word-like phoneme sequence /_il/ with cough and presented it to the listeners embedded into sentences beginning with “It was found out that the _eel was on the ____.” The last word of the sentence was one of the set: axle, shoe, orange, and table. Listeners compensated for the obscured sound with respect to the meaning of the final word. The experiment thus yielded matches: wheel, heel, peel, and meal respectively.

A more intricate manipulation of the initial phoneme was used by Garnes and Bond (1976). They created a continuum varying the cue for place of articulation of voiced plosives obtaining thus three tokens of ideal words (“bait”, “date”, “gate”) and other tokens in between these categories. These stimuli were presented to listeners embedded in carrier sentences each of them matching semantically with only one of the target words (for example, “Bring the fishing gear and the bait.”). The results again proved that the listeners judged ambiguous stimuli (where the quality of the initial phoneme was at a boundary between two categories) regarding the semantics of the whole sentence.

Lexical status of a word plays an important role as well. Apart from the most fundamental task involving the decision

between word versus non-word status of a stimulus, there are experiments assessing the influence of other statistical features. One of them is lexical frequency which represents the relative measure of usage of a lexical unit in a language and can be established with the help of specialized corpora (for example, the British National Corpus or the Brown Corpus of Standard American English). The other is neighborhood density which specifies the number of phonologically similar words which differ in one phoneme only from the target word. Results of a study by Benki (2002) show that while non-words are perceived as clusters of independent phonemes, words are accessed as integrated units. Moreover, high frequency words are recognized more easily as well as words in sparse neighborhood density as predicted by the neighborhood activation model developed by Luce and Pisoni (1998).

Syntactic level influence on speech perception was investigated for example by Mattys, Melhorn, and White (2007). They were interested in the effect of expected syntactic agreement in number (singular / plural) between subject and verb of a sentence on the segmentation of words based on acoustic cues. Stimuli pairs were selected according to the pattern in which the phoneme /s/ might be either the last one of the first word and at the same time the morpheme of 3rd person singular verb agreement, or the first phoneme of the following word (“takes pins” / “take spins”). These pairs were presented to the listeners in singular and plural contexts (“This woman ...” / “Those women ...”). The authors detected facilitation of segmentation by syntactic context.

Moving on to the discourse level, Van Berkum and his colleagues (2003) investigated information integration over the course of several sentences. Their experiments proved that the subjects are gradually building semantic frame and are sensitive

to semantically incongruent stimuli. For example, if three sentences are coherent in indirect presentation of an activity as being exceptionally fast, the listeners are disturbed when the target word (the very last one in the discourse) directly denotes the activity as slow (as being of completely opposite quality).

1.2 Do non-native listeners utilize context information as well?

All the so far mentioned studies were carried out with native listeners of the target language. The investigation of top down processes employment in non-native listeners poses several challenges and the results of the studies presented are much less definite. Jennifer Jenkins in her book *The Phonology of English as an International Language* (Jenkins, 2000) points out the differences between native and non-native speakers in comprehending spoken communication. She uses the umbrella term “speakers” to refer to all language skills including listening. Native speakers have quite deep even though subconscious knowledge of their L1 language, its structure, functions, probability patterns, or constraints. In the process of socialization they also acquired a system of shared knowledge about the world and appropriate communication itself. Non-native speakers lack this kind of background contextual information of cultural, social, or linguistic sort. Jenkins also maintains that most communication by non-native speakers of English happens in non-native environment where none of the speakers is a native English speaker. Speech production of these speakers is more careful and suppresses processes of assimilation and coarticulation, which consequently favors bottom-up processing on the part of the listeners. All these claims together with her own experience led her to a relatively strong statement:

I am suggesting that NBESs¹, even at relatively high levels of competence, still process speech using a predominance of bottom-up strategies. As listeners they seem to find it difficult to make much use at all of the context underlying and surrounding the speech they receive, at both linguistic and extralinguistic levels. (Jenkins, 2000, p. 80)

The problem of non-native speakers is that when they do not recognize a word, they are not sure whether the speaker has used a word they do not know, whether they misheard, or whether the speaker made a pronunciation mistake. Because they cannot rely on contextual information, they devote much of their mental capacity to processing of the acoustic signal itself.

Jenkins gives several examples of listeners' failures to use obvious linguistic and extralinguistic context. In one case, the last word in the sentence "The table is surrounded by chairs" was interpreted quite illogically as "chess". In a couple of other examples, listeners were not able to make use of the established conversational context, or schema as Jenkins calls it (football match, description of a person, etc.). Much fewer occasions demonstrate the utilization of contextual information in listening. Thus the overall summary regarding usage of top-down processing for compensation of insufficient acoustic information might be that "not only are these strategies very rare, but the subjects despite being of upper-intermediate to low-advanced proficiency levels do not seem to use them with any great degree of confidence." (Ibid., p. 89)

Jenkins herself mentions that there are also different opinions among researchers. One of them is John Field, who has been concerned with research of listening in English language

¹ Non-bilingual English speakers (our note) – a term which Jenkins uses for speakers of English as their L2 who are not fluent, or native-like, in this language. (Jenkins, 2000, 6-11)

teaching (ELT) for a long time. In one of his tests (unpublished) the learners of English at lower intermediate level listened to an authentic narrative passage from a book. Several pauses were inserted into the passage and the listeners were asked to write down last four words before each pause. According to his observations more than 75 % of the group of listeners failed to recognize many common words. Nevertheless, as the author states, “[t]o compensate for this lack of adequate ‘bottom-up’ information, L2 listeners form inferences: they use their knowledge of the context to make intelligent guesses about the ideas which link the sometimes dislocated words which they have been able to recognize.” (Field 1998)

Further on, Field argues for a trade-off between information available in the acoustic signal and contextual clues. Because these compensatory strategies are quite common in the non-native listener’s L1, it should be possible to teach him or her to employ them also in listening to L2. Although there are several problems connected with this effort (difficulty with assessing the results of the training, unconscious nature of the strategies, individual differences among listeners), new methodology for listening training is called for. Most importantly, learners should work with authentic material and they should be encouraged to make informed guesses, use their knowledge of the world, topic, speaker, the previous text, and also check their hypotheses against the following passage. Field believes that “[a]t the later stages of learning, some compensatory strategies will develop into good ‘top-down’ listening techniques of the kind used by native listeners (for example, making bridging inferences, establishing expectations as to what will be heard, or switching between levels of generality).” (Field, 1998, p. 115)

These two views seem to be quite contradictory. While Jenkins claims that the listeners are in fact overwhelmed by acoustic signal and thus lack processing capacity for context utilization, Field believes that they are quite good in extracting information from context. Moreover, Jenkins posits that the situation she observed is not likely to change profoundly even at higher levels of proficiency while Field states that the strategies used in L2 will become almost identical with those typical for L1. Taking these conflicting viewpoints into consideration, the aim of this thesis is to investigate the strategies for word recognition used by Czech learners of English at different levels of proficiency.

1.3 State-of-the-art technology for investigation

Before we embark on reviewing some more studies which can provide further insight into this controversy, two methods of investigation will be introduced to illustrate the possibilities of precise research. Both have been brought about by relatively recent development in technology and enable researchers to follow mental processes in real time. The first is a so called eye-tracking experiment (see, for example, Pelz et al., 2000) in which a light-weight helmet with a camera is fixed on the subject's head. The camera follows movements of the eye and a computer integrates these data with the scene at which the subject is looking. Researcher can subsequently evaluate which parts of the scene received more attention and in which order. This method has been extensively used in psychology, cognitive linguistics and also marketing and product design. In language processing the experiment is used for observation of reading strategies or it can show the relation between audio input and a subject's reaction to it in the visual world (investigating a picture).

Another sophisticated method is called event-related brain potentials (ERP) measuring. It has a large field of application in psycholinguistic and neurolinguistic research. Small electric potentials in brain are caused by internal or external stimuli and are measured using electroencephalography (EEG) devices which consist of several sensors evenly distributed on the subject's scalp. Researchers have established several characteristic components of EEG one of them being N400. This abbreviation designates a negative deflection with its amplitude at about 400 ms after the onset of a typically linguistic stimulus which is semantically deviant from its context and thus contradicts subjects' expectations. A thorough overview of this method can be found in Rugg and Coles (1995). Anja Hahne (2001) used this method to investigate differences in language processing between native speakers of German and Russian learners of German. The results led her to the conclusion that semantic integration is slower and consumes more mental resources for non-native listeners. Nevertheless, this integration is done in a comparable manner. It means that the difference in performance of these two groups is merely quantitative. On the other hand, when syntactic integration was investigated, the two groups differed qualitatively. In this case, syntactic violation produced alteration in the shape of EEG waveform for native listeners but not for non-native ones. The author concludes that syntactic processing is of quite robust and rather automatic nature for native listeners when compared to non-natives.

The technology is progressing continuously so that faster, more accurate and more direct means of investigation are available. EEG procedures are being replaced by magnetoencephalography (MEG) which offers more fine-grained observations of the neural activities in human brains. The N400 component itself

can be divided into several activity peaks. Researchers argue that these 400 ms represents a period long enough to contain several separable traces of different linguistic processes which have their own significance in the process of speech perception. (Pylkkanen and Marantz, 2003).

1.4 A review of related studies

Many studies have been carried out to compare native and non-native listeners' strategies in spoken word recognition. In their introduction, Andrea Weber and Anne Cutler (2004) point out that non-native listeners who started to learn a second language later on in their life are in quite a different position than the infants learning their L1. Paradoxically enough, their learning conditions can be seen from a certain perspective as more advantageous; they know how a language works from their previous L1 experience; they are familiar with the function of words, their construction out of sounds, spelling and they understand vocabulary structure and processes. They also have much larger social competence and know how to elicit information they need for their further learning. They are often subjects of intensive language training. In spite of all these, or maybe better because of this multiplicity of sources other than natural spoken input, the L2 learners are in fact disadvantaged.

As evidenced in these two authors' experiments using eye-tracking, Dutch listeners whose L1 does not distinguish /æ/ and /ɛ/ phonemically were fixating their eyes on pictures representing words containing both these phonemes ("pencil" X "panda") if the target word contained /æ/, i.e. the one which does not exist in their native language but not vice versa. The authors conclude that non-native listeners store the difference in their lexical representations; nevertheless, their acoustic-phonemic processing of English is not so successful to map the input directly on the

abstract representations. In a study by Escudero et al. (2008), the researchers found out that L2 learners indeed incorporate spelling knowledge into their abstract lexical representation. When learning a new word, the subjects without access to its written form were not likely to make use of phonemic difference which does not exist in their L1.

Incorporating semantic level, the researchers have modeled basically two types of experiments in which they look for correlation between semantic information of a carrier sentence and the target, usually final word. In the first type the criterion is so called predictability of the intended word from the context of its embedding sentence, i.e. the possibility to use the meaning of the sentence to logically “guess” its completion. For example, it is quite easy to complete the sentence “February has twenty-eight ...”, whereas the sentence “There are many ...” can be concluded with almost any countable noun. These two types of sentences were actually used by Bradlow and Alexander (2007) who wanted to investigate the relative contribution of sentence semantics and acoustic information to the final word recognition. They presented both native and non-native listeners with high and low predictability sentences (semantic enhancement) spoken in either causal or clear speaking style (acoustic enhancement). Each sentence was mixed with noise at a specific signal-to-noise ratio. The experimenters found out that whereas native listeners benefited from both types of enhancements, non-native listeners’ performance was significantly improved when these two factors worked in combination. This observation suggests that in order to enable non-native listeners to employ top-down processing, they first need clearly defined acoustic information. Our experiment reflects this in its procedure in which only the target words are masked by noise whereas the preceding carrier sentences are

undisturbed to provide easily recognizable semantic priming for the targets.

In the second type of experiment, the carrier sentence is always semantically rich resembling the high predictability sentences of the previous experiment type. The target word then is semantically congruent with the sentence (“John cannot pay his bills because he doesn’t have any money.”) or semantically anomalous (“...because he doesn’t have any trousers.”). Ian FitzPatrick and Peter Indefrey (2007) developed this paradigm and added two more conditions to investigate lexical access of non-native listeners: first, the initial phonemes of the final word are the same as those of the most logical sentence completion (“...because he doesn’t have any muffin.”), and second, the initial phonemes are identical (or at least quite similar) with the phonemes of the most logical completion in the listener’s L1 (“...because he doesn’t have any pencil.” – in our example for Czech “peníze”). Their results brought by ERP measuring indicate that non-native listeners are sensitive to both semantically incongruent conditions in L2. Their specific reaction on initial phoneme overlap proves that the listeners start semantic integration quite early, even before complete phonetic information about the word is available. In contrast, they do not consider initial phoneme overlap with words from their L1 which suggest that the listeners keep the two lexicons separated. Our study draws from these findings in the design of experiment 1 with gradual masking. Based on the reported results and conclusions we also decided to set aside investigation of cross-language lexicon access because it did not prove relevant.

Similar results were reported earlier by Anja Hahne (2001) whose study has already been mentioned in the previous section. She employed two types of violation in the sentences: semantic

and syntactic. Whereas semantic violation produced comparable ERP reaction in both groups with a slight difference in its intensity and timing, reactions on syntactic violation differed qualitatively. The process of syntactic integration seems to be rather automatic and requires minimum resources in case of native listeners and thus any violation causes rapid reaction. Non-native listeners employ mental resources for the same process more consciously and continuously regardless of the syntactic correctness and that is why no additional activation could be observed.

Apart from semantic integration confined to the linguistic domain, the listeners might use originally non-linguistic sources of information. One of Jenkins' observations which led her to her claims about non-bilingual English speakers being not able to use contextual information concerns visual source of information. She presents the task in which there are two discussion partners. One of them describes a picture for the other so that he or she can identify it out of a set of six pictures. As the subsequent rather informal analysis showed, the listener had troubles recognizing the word "red" pronounced by his Japanese partner more like [let] in a phrase with the word "cars". Even though only one picture contained any cars and all of them were red, the listener was not able to use this information to compensate for confusing pronunciation. Instead, the listener was stubbornly elaborating on the acoustic input and tried to find the slightest evidence of some "cars for hire / to let". (Jenkins, 2000, 81-2)

Cross-modal integration in speech perception is evidenced in its purest form in the so called McGurk effect. This phenomenon which is popularly known as lip-reading has been thoroughly explored also for L2 perception both on the level of phonemic contrasts in non-words (Hazan et al., 2006) and word

recognition in different speech styles (Hardison, 2005). Both studies proved that non-native listeners are able to integrate these two modes of information for recognition of phonemes which exist in their L1 as well. Visual cues also generally facilitated word identification.

Nevertheless, our study is more interested in the ability of non-native listeners to use and integrate visual information which is not directly connected with speech production. Tanenhaus et al. (1995) used eye-tracking to examine the interplay between visual and linguistic information in spoken form. The paper reports several interesting observations originating in their previous studies in which the subjects were asked to manipulate with objects while the movement of their eyes was recorded. The results indicate that language processing is incremental because it took longer time for the subjects to fixate their eyes on the correct object when two different objects beginning with the same phonemes were present on the scene (“candy” X “candle”). The subjects also looked at the correct object as soon as sufficient information was available for unambiguous identification (for example after the word “starred” in the sentence “Touch the starred yellow square.” when there was only one starred object or after the word “square” if only the shape differentiated two possible objects). The authors extended their investigation to see whether syntactic parsing can be modified by visual scene configuration. They used potentially ambiguous sentences of the type “Put the apple on the towel in the box.” in which it is not clear whether the phrase “on the towel” modifies the apple, i.e. where is it located now, or whether it specifies the destination, i.e. where should it be put. The eye movements revealed that interpretation of is in these cases determined by what listeners in reality see.

Another experiment by Kamide et al. (2003) used eye-tracking to follow anticipatory eye movements to investigate whether preceding linguistic input can provide basis for a sentence completion prediction. The researchers were interested in semantic constraints of relations which are established between Agent, a transitive verb and Theme of a sentence. Subjects were presented with pictures containing two Agents (for example, a girl and a man), two possible Themes (a carousel and a motorbike) and several distracting objects. The spoken sentences corresponding with the pictures contained one verb over which the Agent selected its Theme. To continue with the above started example, the sentences would be “The girl will ride the carousel.” and “The man will ride the motorbike.” As hypothesized, the subjects moved their eyes in anticipation of the completion of the sentence to the appropriate object in the picture before the onset of the word which denotes the object. The conclusion is that the listeners use gradually revealing constraints to compute probable relationships between entities already referred to in the linguistic input and those available in the visual context at the earliest possible moment. These two studies indicate that native listeners are able and to a significant degree do integrate linguistic and visual information in the process of speech comprehension. In our study we aim for investigation of this process for non-native listeners.

Before introducing our own research questions and experiment designs, let us briefly review the highest level of information integration and compensation for lower listening competence of the listeners in their L2. In his above presented article, Field (1998) writes about the ability of his non-native students to make use of their real world knowledge to bridge over the information gaps caused by deficiency in their bottom-up processing and mis-

hearing in less than ideal listening conditions. Chiang and Dunkel (1992) present a thorough review of different approaches to L2 speech comprehension influenced by prior knowledge and topic familiarity. In their own study, they presented Chinese students of two proficiency levels in English with short English lectures on Confucius (a familiar topic) and on Amish people (unfamiliar topic). The authors also prepared a modified version of each lecture with higher semantic redundancy and repetition. They found out that only the students in the group of high-intermediate listening proficiency were able to benefit from more redundant speech. More importantly, the subjects were asked to complete a multiple choice post lecture test with two types of questions: either passage independent about general knowledge of the topic or passage dependent about specific information contained in the respective lectures. Those who listened to the familiar topic lecture achieved significantly higher scores for passage independent than for passage dependent questions. No such difference could be observed in those who listened to the unfamiliar topic lecture. When performance levels of those who listened to familiar and those who listened to unfamiliar topic is analyzed, it can be seen that there is significant difference in the passage independent question scores; nevertheless, there is no significant difference in passage dependent question scores. According to us, this suggests that the listeners in fact are not able to use their prior knowledge to compensate for their poorer listening competence because their prior knowledge did not help them to achieve better comprehension of information specific to the particular passage even though they knew “what it was generally about”.

On the other hand, in a more recent study by Sadighi and Zare (2006), the authors reported that they found significant difference in the performance of a group of listeners which

received some treatment before listening when compared to a control group. The experimental group was instructed to do a research on several selected topics on their own and then they discussed these topics in class before listening. The topic activation improved the results in a TOEFL listening comprehension test. Although these studies are fruitful and enriching especially in the field of English language teaching methodology, we see some shortcomings in the administration of them as well. Procedures employ rather loosely defined methodology. For example, it is quite difficult to separate the pieces of information which a listener obtained from the lecture or listening itself and which he or she knew before from his or her prior experience or the pre-listening discussions. In other words, we can assume that the listeners would be able to answer passage independent questions even without listening to the lectures with similar scores. That is why we decided to adhere to a more strictly controlled testing paradigm which should yield more precisely measurable results.

1.5 Research questions and experiment design

The review of several approaches at different levels of contextual influence on speech comprehension offered above proves that this topic covers rather broad area of interest. In order to keep a compact scope of our study, we decided to pursue the line of research more closely related to linguistics, especially the interaction of phonetics and semantics. World knowledge utilization and gap bridging via informed guesses which are the topics of the two final studies presented in the previous section seem more closely related to psychology, information processing and theory of cognition. Moreover, it is difficult to devise precise and operative methodological procedures to measure these phenomena. Our research is anchored in the word and sentence

level with the fundamental aim of exploring the interplay between information incoming to the subjects in the form of acoustic signal (spoken utterance) and semantic information which can be derived from the meaningful structure of a sentence (linguistic context) and/or a picture (visual non-linguistic context).

Altogether, three experiments were designed. The purpose of experiment 1 is to find out whether non-native listeners derive semantic information from the onset of a spoken sentence throughout its duration and establish semantic frame towards its end so that the final word recognition is influenced by the semantic context of the sentence. As mentioned before, non-native listeners require clear acoustic information to allow operation of top-down processing (Bradlow and Alexander, 2007). That is why the whole sentence apart from its final word is left undisturbed by any manipulation to provide a sound basis for the listeners to establish semantic context. The final words can be divided into three categories with respect to their relation to the semantic context of their carrier sentences: they are either semantically congruent with them, semantically incongruent or semantically incongruent but sharing initial phoneme with the most probable congruent word completing the sentence. The final word is mixed with noise of gradually increasing intensity so that listeners should be able to recognize the initial phoneme more easily. If listeners establish semantic context and try to relate the final word to it, then their performance should be significantly better with semantically congruent target words than with incongruent ones where the semantic information would be in conflict with acoustic information which should cause processing disruption and ultimately lower accuracy and higher reaction time. On the other hand, if listeners recognize each word more or

less separately without trying to set them in the overall sentence context, then there should be no significant difference in performance, i.e. the semantically incongruent words should be recognized with accuracy and reaction time more or less similar to the semantically congruent condition. If listeners relate the words to the semantic context, the recognition of the final word in the third group might be biased toward its more probable, semantically congruent competitor beginning with the same phoneme. Thus the cases in which subjects submit a probable competitor sharing initial phonemes with the actually pronounced word should be treated as signs of top-down semantic intervention for compensation of acoustic distortion.

Experiment 2 connects linguistic and visual modality. Subjects are presented with a simple picture containing several objects of standardized style which are in relation with each other. A certain time span is provided for picture inspection before the onset of a sentence. The sentence semantics might be either fully congruent with the visual information of the picture (the sentence describes the picture), partially congruent but deviating in one detail, or completely different from what the picture shows. Subjects are asked to judge whether the sentence talks about the picture or not. This experiment is concerned with the question whether non-native listeners are able to process acoustic signal together with visual information and in what way do they manage it. The prediction is that if non-native listeners are able to process these two modes of presentation simultaneously, then accuracy should be comparable in congruent and completely incongruent conditions, whereas subjects might make more mistakes when the difference rests in one detail only (partially incongruent condition) in which case it can be easily overlooked. As for reaction time, the prediction is that it may be

significantly lower in completely incongruent condition when compared to the other two because the answer might be arrived at even before the offset of the sentence. If listeners use adaptive strategies with respect to the task demands, they may skip processing of the rest of the sentence after finding early incongruity thus performing at quite short reaction time. On the other hand if listeners do not pragmatically accommodate for the assignment (similar to Jenkins's observation), then the reaction time will be comparable with the congruent condition. As for partially incongruent condition, reaction time might be longer if listeners evaluate linguistic and visual information separately and subsequently.

The last experiment 3 resembles experiment 1 in the gradual masking of the sentence final word which, however, is always semantically congruent with its carrier sentence. Nevertheless, it may be congruent with visual information provided by a simultaneously presented picture, or it may differ from it. There are again two degrees of incongruity: partial and complete. In partially incongruent condition, the carrier sentence can be related to the presented picture (the sentence names relevant objects, their properties and setting) but the target word denotes an object which is not present in the picture. In completely incongruent condition, the sentence describes a completely different picture. The research question covered by this experiment is whether non-native listeners are able to utilize information presented in visual mode to compensate for noise in acoustic signal. The prediction is that if listeners do integrate these two modes of presentation and follow the pictures to arrive at the target word recognition, accuracy should be significantly higher and reaction time significantly lower in completely congruent condition in comparison with incongruent condition. Partial incongruity should bring

distortion to the listeners increasing their error rate and reaction time. As in experiment 1, the substitution of what was actually pronounced with the most probable completion presented in the picture should be treated as a sign of cross-modal integration.

A further refinement of this experimental pattern was originally considered, namely to separate the partially incongruent condition into two categories. In the first one, the name of the object in the picture corresponding to the target word would begin with the same phoneme as the target word actually pronounced (the picture would show a girl sitting in a window and the sentence would be “There is a girl sitting in a wheelchair”). In the second one, there would be no phonetic relation between the name of the object and the actual target word. The hypothesis might be that if listeners make use of visual information thoroughly and activate names of the objects in the picture in L2 immediately, then the cases in which the pronounced target word is substituted with the name of the object presented in the picture would be more numerous in the first subcategory. Nevertheless, developing such an intricate system of carrier sentences, target words, pictures and objects proved quite difficult with respect to limited vocabulary size of the beginners, as well as to limitations of picture bank and growing number of stimuli. Moreover, subsequent analysis would impose high demands as well, so that this further extension was not implemented.

This design should also allow for cross-experiment analysis, especially comparing the results of experiments 1 and 3. Recognition accuracy should be even higher in otherwise comparable congruent conditions when visual information is presented. On the other hand, if visual information processing and integration represents too high processing load, then the results will be worse for cross-modal condition. Another

comparison can be done between experiments 2 and 3: while subjects are forced to relate acoustic and visual information to arrive at a decision in experiment 2, it is possible to ignore visual information completely and concentrate purely on the incoming acoustic signal in experiment 3. The prediction is that if listeners suppress considering visual information in experiment 3, no trace of disturbance will be found in their performance in partially incongruent condition. On the other hand, if the result patterns of congruent and partially incongruent conditions are similar when these two experiments are compared, listeners might be said to be evaluating visual information consistently across these experiments.

2 Methods

This section presents the study itself, more specifically its three constituent parts experiments 1, 2 and 3. They are independent of each other to a certain level, yet they were prepared to form a meaningful complex and allow proper investigation of the research questions. Individual parts deal with selection of the subjects, material preparation, test administration, data collection and analysis procedures. All the supplementary material and data presenting individual components of the experiments in detail as well as the experiment files and scripts themselves together with software used to carry out this study can be found in electronic version on a CD which is a part of this thesis. For the CD content overview, see Appendix A.

2.1 Subjects

The two sources which form the basis for this investigation mention different levels of their subjects' language competence, and curiously enough, the relation between this proficiency level and performance or strategies employed seems to be contrary to what could be expected. While Field talks about "learners at lower intermediate level" (Field, 1998, p. 115) as being able to apply top-down processes to make efficient guesses, Jenkins claims that listeners "at relatively high levels of competence" (Jenkins, 2000, p. 80) still do not make much use of the context to bridge pronunciation deficiencies.

With this slight discrepancy in view, we decided to form two groups of subjects with respect to their level of competence in English. It was not feasible to perform private language performance tests for excessive methodological and administrative demands or to make use of some kind of standardized testing so that the only decisive criterion for subject selection was the

number of years they had been learning English. Both groups contained 15 subjects. Group 1 consisted of grammar school students who had learnt English for three or four years at school (three 45 minute lessons per week) and Group 2 consisted of first-year university students majoring in English for business purposes who had studied English for at least nine years. For summarized group data see Table 2.1.

	years of learning English			age			male	female
	min	max	mean	min	max	mean		
Group 1 lower- intermediate	3	4	3,33	12	16	13,6	7	8
Group 2 upper- intermediate	9	14	10,87	20	24	20,73	6	9

Table 2.1: Summarized data describing main features of the two experimental groups.

None of the subjects reported any known hearing or language difficulties. Three subjects in group 2 had stayed in an English speaking country for over six months. Details can be found on the CD (see Appendix A).

2.2 Materials

Preparation of materials for a speech perception test is a complicated task requiring much attention to several factors. Inspiration can be taken from Kalikow and his colleagues (Kalikow, Stevens, Elliott, 1977) who endeavored to develop a universal test for assessing speech intelligibility in noise (SPIN test). The authors claim that “[a]n especially difficult aspect of the design of sentence tests is the generation of sentences that are reasonably natural and that are at the same time well controlled for word predictability, phonetic content, and other attributes. It is also desirable that the audiometric test be easily administered as

well as reliable and valid.” (p. 1338). They identified several factors which should be taken into consideration and properly managed during test preparation. Even though their test was intended exclusively for native listeners, it might be useful to go through these factors and briefly comment on them with regard to material preparation for this thesis experiments.

Generally speaking, the crucial difference and obstruction when tests for natives and non-natives are compared is the limited vocabulary size of the latter group which makes it much more difficult to meet most of the requirements. It was necessary to ensure that the subjects were familiar with the target words and presumably with all the words in carrier sentences as well. Since all the subjects in group 1 were attending the same school and following the same English course, all the target words were taken from a dictionary accompanying their course book *New Opportunities*². All the selected words were pre-tested with the subjects of this group.

The total of 48 words (experiment 1 and 3 combined together) were randomly distributed into five mutually exclusive groups of 9 or 10 words. Another group of 10 words was constructed using the same source textbook. In this case, no special criterion for word selection was employed. These words were mixed with the words in each of the five groups and served as fillers to distract subjects attention from the words actually used as targets. Subjects were presented with one of the groups consisting of 9 to 10 randomly selected targets and 10 fillers and asked to point out any unknown word. Thus, each target word was presented three times (5 groups for 15 subjects). Subjects were given Czech equivalents of any words marked as unknown.

² Hartus, Michael, David Mower and Anna Sikorzyńska: *New Opportunities – Elementary*. London: Longman, 2006.

None of the words was marked unknown by all three subjects it was presented to and therefore none of the originally selected words was dropped and substituted. It was supposed that the more advanced group 2 should also know these words, thus no such pre-testing was done with this group.

One factor with which Kalikow's team deals extensively falls into the field of phonetics and prosody. Namely, it is true that some groups of sounds are affected by noise masking more profoundly than others and thus it would be desirable to have words containing in total the same number of sounds of each group. Limited size of vocabulary, however, did not allow for control of this factor. Moreover, the testing design with gradual masking required longer words, so while the SPIN test uses monosyllabic words, our experiments 1 and 3 employ words containing at least two syllables which again significantly limits the number of candidate words. In addition, to prevent reduction in pronunciation and enable proper activation processes in the testing paradigm of experiment 1 (words beginning with the same phoneme) it was required that the stress falls on the initial syllable for the target words in this experiment (with the exception of "computer").

Another factor mentioned by Kalikow is word familiarity. The authors used a list of words indicating their frequency and decided to use only those with frequency counts ranging from 5 to 150 per million words. This criterion is not that relevant for learners of English because the frequency in everyday language does not necessarily correspond to usefulness for learners of a foreign language. For example, target words used in this experiment such as "dentist", "elephant", or "headache" are not listed in the British National Corpus frequency list³ going down

³ The corpus can be accessed through web page <http://ucrel.lancs.ac.uk/bncfreq/>

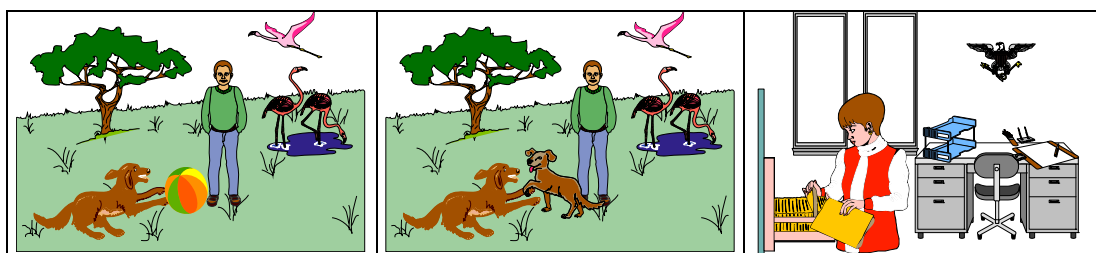
to frequency count of 10 occurrences per million words. Still these words are quite often included in textbooks for beginners.

2.2.1 Spoken stimuli preparation

Ensuring semantic predictability was more difficult as well, especially when two other requirements are considered: comparable length of the carrier sentences and their naturalness. While Kalikow and his colleagues imposed relatively strict regulations on these factors and carried out even predictability judgment pre-tests, our experiments did not control the sentence length so rigorously in favor of semantic priming and natural sounding. Each carrier sentence in experiment 1 contains at least two words pointing semantically to the congruent target. Pairs of targets sharing initial phonemes were created and set in their carrier sentences. Different conditions were achieved through a system of carrier sentence exchange. For example, the target words “author” and “autumn” were embedded in their congruent carrier sentences “The *book* was *written* by two authors.” and “The *season* which comes after *summer* is autumn.” respectively. The words printed in italics are intended as semantic pointers. For the condition in which the initial phoneme was intended to be the same as in the most probable congruent sentence completion, the two target words were swapped. A sentence originally carrying phonetically unrelated target word and lacking any semantic priming was taken for the completely incongruent condition. Replacing the word “uniforms” we thus obtained an incongruent sentence “In some British schools children have to wear author.”

In the visual mode of experiments 2 and 3, there are at least three words pointing to relevant objects in the picture in congruent condition including the target word. In this case the sentence remains the same, different conditions are achieved

through picture manipulation (see Figure 2.1 for illustration). Again, the size of vocabulary was quite limiting.



Experiment 2: My *brown dog* loves to play with a *ball* in the *park*.

Figure 2.1: Left picture is congruent with the sentence using the grass, tree and birds with a small pool to create semantic frame. Middle picture is partially incongruent because the ball was replaced with another dog. Right picture is completely incongruent. The words printed in italics point to the semantic frame of the congruent picture. Similar pattern applies to experiment 3 as well in which the object being replaced in partially incongruent condition is always the one referred to by the target word.

For experiment 1, a set of 12 target word pairs sharing initial phoneme was created. Each of these 24 words appeared in all three conditions (congruent – CO, incongruent but sharing initial phoneme – IP, incongruent – IC) which produced 72 sentences. Grammatical number adaptations (singular/plural) were made. Another set of 32 sentences was created for experiment 2. Half of them were intended to be congruent with the picture, the other half either partially or completely incongruent (in 1:1 ratio). Finally, another set of 24 target words different from those of experiment 1 was created in experiment 3 and set into their carrier sentences which contained other pointer words referring to their congruent pictures. Trial samples were created for each condition in each experiment which gave 9 (3 x 3) more sentences. These stimuli met the same requirements as the test ones. In total, 137 sentences were prepared to be recorded (72 for exp. 1, 32 for exp. 2, 24 for exp 3, 9 trials).

2.2.2 Recording and manipulation

A trained native speaker with unmarked British English accent was asked to read each sentence twice at normal conversational speech rate. The recording was done on a laptop computer in a quiet office using condenser microphone Monacor ECM-100. Software used for recording and subsequent modifications was Audacity 1.3, an open source audio processing program⁴. Recording parameters were set to 22 kHz sampling rate, 32-bit float resolution, one channel (mono), and wav file format for saving.

Using the same software, the recordings were normalized to suppress as much as possible varying loudness resulting from slight changes of the speaker's distance from the microphone and his general inclination to talk louder towards the end of every recording section. Each stimulus was cut out and saved as a separate file choosing the one of the pair which subjectively exhibited more even speech rate and more natural intonation curve.

To establish proper signal to noise ratio (SNR) for target words in experiments 1 and 3, several studies were consulted and pilot testing was carried out. Developing the speech reception threshold (SRT) paradigm, Nilsson and his team (1994) reported in their article that the mean SNR across their English subjects was -2,91 dB. At this level, the subjects were able to repeat a whole sentence masked with spectrally matched noise correctly. Kalikow and his colleagues present their results using multi-talker bubble masking in a figure (p. 1346) which reveals that for young listeners, the biggest difference between perception accuracy for high and low predictability sentences is at SNR of 0 dB where high predictability accuracy reaches 90 % and low

⁴ Audacity 1.3.5 (beta) including necessary filters and generators obtainable at <http://audacity.sourceforge.net/>

predictability is slightly lower than 50 %. Additional useful information can be found in Adachi et al. (2006) who investigated effects of white and pink noise on phoneme discrimination by native (American English) and non-native (Japanese) listeners. Furthermore, Trimmis et al. (2007) reported their results using again speech shaped noise to mask Greek bisyllabic words which were subsequently presented in isolation to native and non-native speakers of Greek. Taking all these sources together with the piloting into consideration, we decided for SNR at 0 dB for experiment 1 and -3 dB for experiment 3 in which the sentence itself is always semantically congruent making the recognition easier. The SNR is more favorable than in the studies with native listeners only but less favorable than in the studies with non-native listeners because subjects in our experiments can profit from undistorted carrier sentences and gradual masking of the targets. Pink noise (also called 1/f noise) was chosen because it covers all octaves with equal noise power, thus masking the whole spectrum evenly.

Even though initial phoneme congruity is being mentioned throughout the thesis, some of the target word pairs share more than one phoneme (“instrument” / “interview”, “lemon” / “lesson”). These were intended to activate their semantically congruent counterparts more profoundly. Again, it would be reasonable to control the phonetic structure of the words because it is true that the most relevant phonetic cues might be distributed differently across the words, nevertheless, this task would be quite difficult to achieve with such a restricted set of possible words. Since it was not necessary for the purpose of the study to spread the gradual masking exactly over the initial phoneme exclusively, it was set to the first fifth of each target word’s duration, a measure which again was based on piloting. As a

result, longer, more difficult words obtained clearer acoustic information in their initial portions when compared with shorter and easier ones.

Each sentence for experiments 1 and 3 was re-loaded into Audacity and an exact copy of the track was made. The final word was selected as precisely as possible using spectrum instead of waveform and zooming in as necessary. Finding word boundaries was not difficult in most cases, yet combinations of two vowel sounds caused some problems (“grey elephant”). Nevertheless, when the length of the masking gradation is compared with the range of vowel overlap which is much shorter, possible inaccuracy is not significant and could not influence the results in a serious way.

Unfortunately, Audacity cannot measure mean intensity of a given sample. PRAAT, free software for doing phonetics by computers⁵, was used for this purpose. First, mean intensity of a 10 second sample of pink noise with maximum amplitude generated in Audacity was measured and stored. Subsequently, mean intensity of each final word selection was measured. The difference between these two values was used as the parameter for adjusting intensity of the masking pink noise. It was generated with maximum amplitude in one of the two tracks in the same position and with the same length as the target word which remained undisturbed in the other track. After adjusting noise intensity, the selection was reduced to its initial fifth and fade-in effect was applied to it to achieve the intended gradation of masking. Finally, the two tracks were merged into one and saved into wav file format. For summary information on target word manipulation, see Table 2.2. Details can be found on the CD.

⁵ PRAAT was developed by Paul Boersma and David Weenink at Institute of Phonetic Sciences, University of Amsterdam and its most recent version can be downloaded from <http://www.fon.hum.uva.nl/praat/>

Parameter	Min	Max	Mean
Duration (ms)	402,742 (lemon)	911,332 (uniform)	644,340
Fade-in duration (ms)	80,548	190,867	128,868
Mean intensity (dB)	63,301	79,203	70,562
Intensity adjustment (dB)	-1,208	-20,172	-12,162
10s noise sample mean int. (dB)	83,473		

Table 2.2: Summary of parameters measured for manipulation of target words in experiments 1 and 3

2.2.3 Visual stimuli preparation

Visual mode stimuli were all prepared in Print Artist 3.0, software for creating graphics⁶. It contains a large collection of vector graphics which can be easily combined, adjusted, shaped and colored. Most importantly, its unified visual style ensured that individual pictures were comparable with respect to mental processing load. Each picture is composed of approximately five objects which create a semantic frame. The sentence which is congruent with the picture contains at least three words which point to this frame including the target word in experiment 3. The two target objects within a pair the swap of which makes the difference between completely and partially incongruent conditions in both experiments 2 and 3 are always approximately of the same size and they are located in the same place. See Figure 2.1 above for illustration. All stimuli can be found on the CD. After their composition, the pictures were saved in gif file format and named appropriately to match them easily with corresponding sound files.

⁶ Print Artist is commercial software developed by Sierra On-Line, Inc. © 1996

2.3 Procedure

Stimuli were presented to the subjects using Alvin 2, a free program for controlling experiments.⁷ Scripts for running the three experiments were obtained by modifying some of those provided with the program to suit our experimental purposes and allow saving responses in a convenient way. Group 1 took the experiments in pairs on two computers in a quiet teacher's office directly at their grammar school, group 2 were tested in sets of four to six subjects in a large and quiet language laboratory at the university. They were seated far enough from each other to minimize disturbance. Stimuli were presented binaurally via headsets with microphones for recordings. In accordance with the aim of the study, the experiment scripts did not allow replying of the stimuli, each stimulus was played only once and then a response was awaited. Only subjects who reported no hearing or language difficulty were selected and subsequently asked to fill in brief forms with their initials and several other statistical details prior to testing (data collected are summarized in subsection 2.1 Subjects above). Listeners in group 1 were presented with vocabulary pre-tests to make sure they knew all the target words.

Instructions before each experiment were given in Czech to ensure that the subjects of group 1 understood the task. The sessions with trial stimuli were taken prior to test sessions to make the subjects familiar with the experimental pattern, controls and to enable them to adjust loudness of playback to suit them. The sequence of experiments was changed randomly to eliminate bias caused by learning, tiredness or any other interference among experiments. Stimuli within one experiment were presented in random order as well. It took approximately

⁷ Alvin 2.0.12, developed by James M. Hillenbrand and his colleagues at Western Michigan University, downloaded from <http://homepages.wmich.edu/~hillenbr/>

30 minutes to complete all the three experiments including form filling, vocabulary pre-testing (group 1 only), instructing, and trials.

Each experiment was prepared in three varieties A, B, and C. Each variety was taken by the same number of subjects in each group. Each target word and each carrier sentence (experiment 1) or picture (experiments 2 and 3) appeared only once within one variety. It was reasoned that subjects might be biased (prompted by a previous token) if they heard or saw the same stimulus component (carrier sentence, target word, or picture) in different conditions. Details for individual experiments follow.

2.3.1 Experiment 1

Subjects were told that they were going to hear sentences the final word of which would be masked by noise. Their task was to repeat the final word only and then click on “Next” button for the following stimulus. In case subjects did not recognize the target word, they were asked to say “nevím”, Czech equivalent for “I don’t know”. The program started recording immediately after the offset of the sentence and stored the response in a wav file for subsequent analysis. Each subject heard all the 24 target words evenly distributed into the three conditions. Table 2.3

Variety	Sentence	Target	Condition	Table 2.3: Pattern for combining carrier sentences S1-3 with target words T1-3 to ensure that subjects within one variety hear each carrier sentence and each target only once. Each target word appears in different condition for each variety.
A	S1	T1	CO	
B	S2	T1	IP	
C	S3	T1	IC	
A	S3	T2	IP	
B	S1	T2	IC	
C	S2	T2	CO	
A	S2	T3	IC	
B	S3	T3	CO	
C	S1	T3	IP	

shows combinations of targets with carrier sentences which produced different conditions and their distribution into varieties A, B, and C.

2.3.2 Experiment 2

First, a picture appeared on the computer screen and after two seconds a sentence was played. This time span was intended to let the subjects inspect the picture so that they could follow the sentence and relate it to the picture more easily. The subjects were asked to decide whether the sentence was describing the picture or not and click on the respective “True” or “False” button after which a new stimulus appeared. The choice was recorded and written to a text file together with reaction time which was measured from the offset of the sentence.

Each subject heard all the 32 sentences out of which 8 were congruent with the picture and shared across the three varieties. This set served to counterbalance negative responses and prevent subjects’ urge to respond positively in the dominance of negative responses. These stimuli were not included in statistical analyses. The remaining 24 sentences were combined with pictures across the three varieties in a manner which ensured that each variety obtained a given sentence with a different picture (producing thus one of the three conditions, see Table 2.4).

Variety	Sentence	Target	Condition	Table 2.4: Pattern for combining sentences S1-3 with pictures P1-3 and a picture distracter D which remained the same for three sentences depicting something completely different from their content. Pictures marked with asterisk were created from their unmarked (congruent) counterparts by changing one object.
A	S1	P1	CO	
B	S1	P1*	IP	
C	S1	D	IC	
A	S2	D	IC	
B	S2	P2	CO	
C	S2	P2*	IP	
A	S3	P3*	IP	
B	S3	D	IC	
C	S3	P3	CO	

2.3.3 Experiment 3

As in experiment 2, a picture appeared first. Sentence with its final target word masked by noise was played after two seconds which were intended again for visual inspection. Subjects were asked to repeat the final word only (or “nevím” to signal failure of recognition) and then click on “Next” button for the following stimulus. As was the case in experiment 1, the control program started recording immediately after the offset of the sentence. The response was saved in a wav file. Subjects were presented with all the 24 target words evenly distributed into the three audio/visual congruency conditions. Target words were always semantically congruent with their carrier sentences. Experimental pattern for presenting sentences with pictures was the same as in experiment 2 (see Table 2.4). The object denoted by the target word was replaced with a different one in pictures marked with asterisk producing partial incongruity. This pattern again ensures that none of the subjects heard the same sentence or saw the same picture more than once, while each sentence appeared in all three conditions across the varieties with respect to congruency with the visual mode of presentation.

3 Results

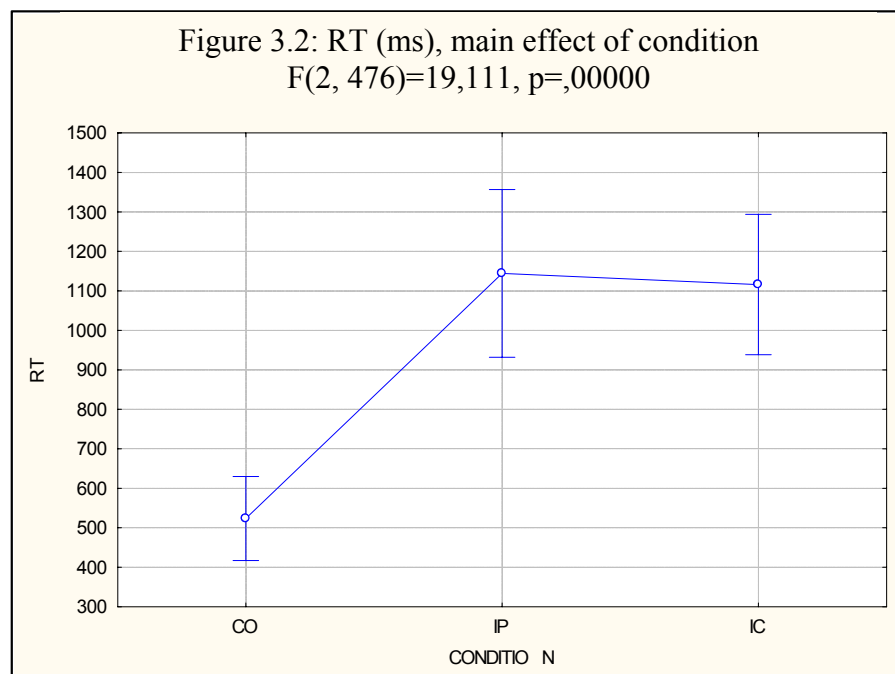
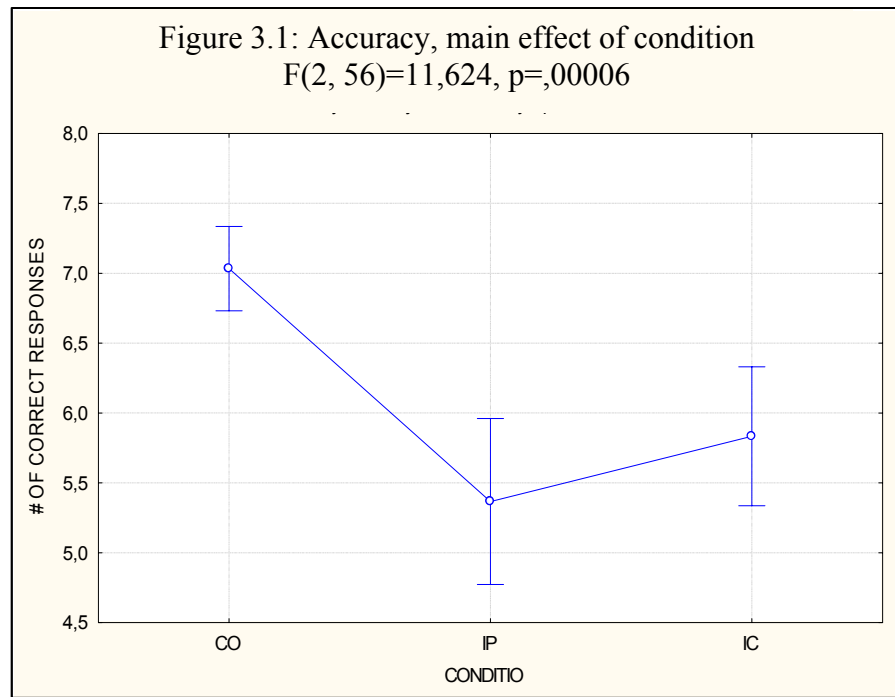
Two factors were measured in all three experiments: accuracy and reaction time. To obtain values for experiment 1 and 3 all the recordings of subjects' responses were opened in PRAAT and played to establish whether the response was accurate or not. In a spreadsheet, "true" was marked for accurate recognition and "false" for both inaccurate and "nevím" responses. Any interesting inaccurate response was recorded in written form together with notes about the condition in which it appeared and which target word was substituted by it for further analysis. As discussed in section 1.5 Research questions and experiment design, responses toward congruent target words were expected especially in partially incongruent condition.

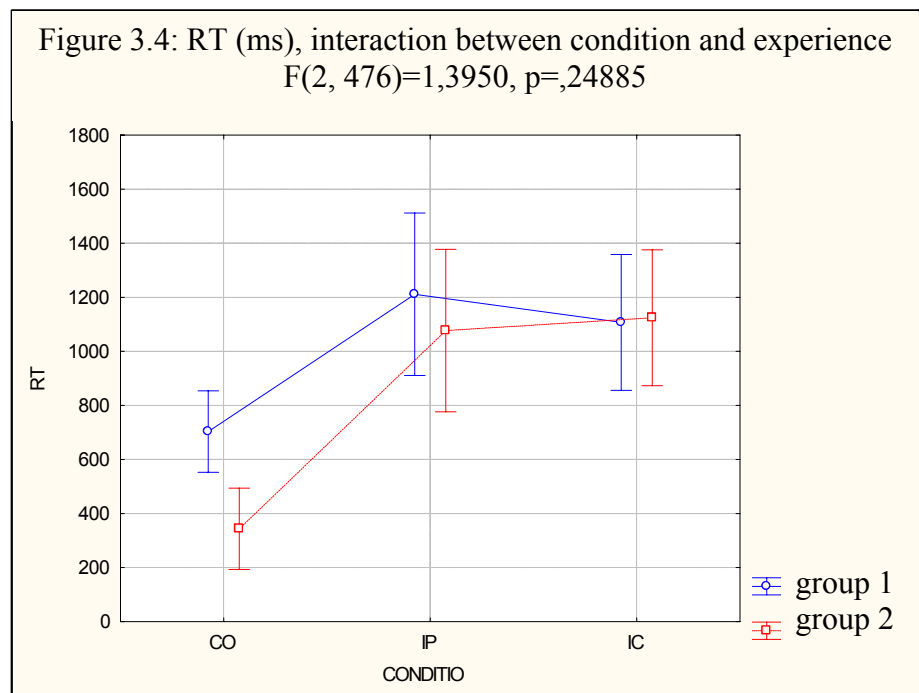
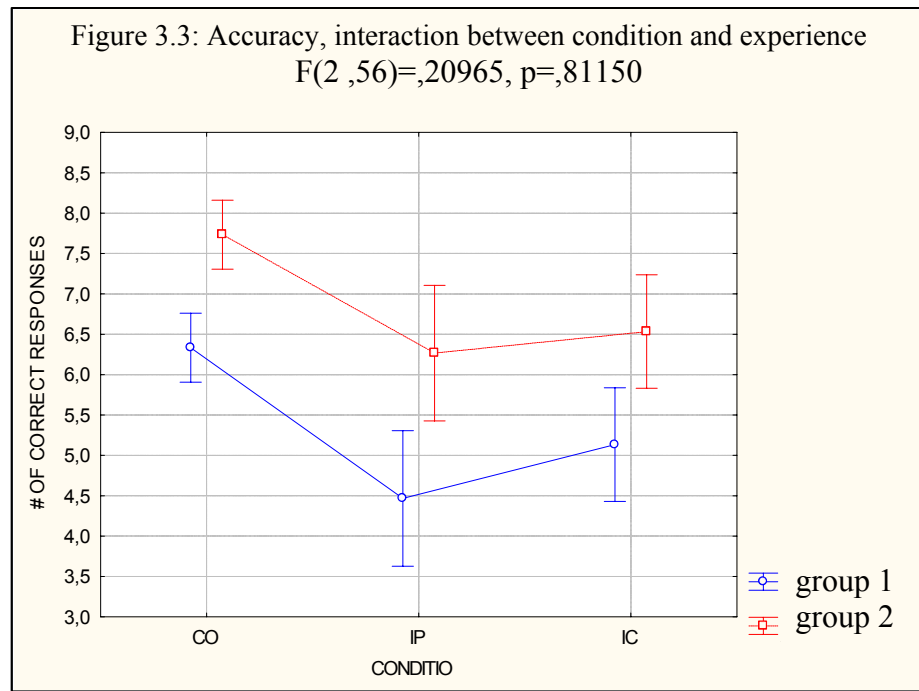
Reaction time was measured from the beginning of each recording to the onset of a relevant response regardless of the response accuracy. The values in milliseconds were again stored in a spreadsheet. Results from experiment 2 could be obtained more easily directly from Alvin output files generated for each subject. Accuracy was obtained by comparing subject's button response against intended response ("true" for congruent, "false" for both partially incongruent and completely incongruent conditions). The results of a set of eight congruent sentences which were intended to counterbalance the number of "false" responses (for psychological reasons only) were not included in statistical analyses.

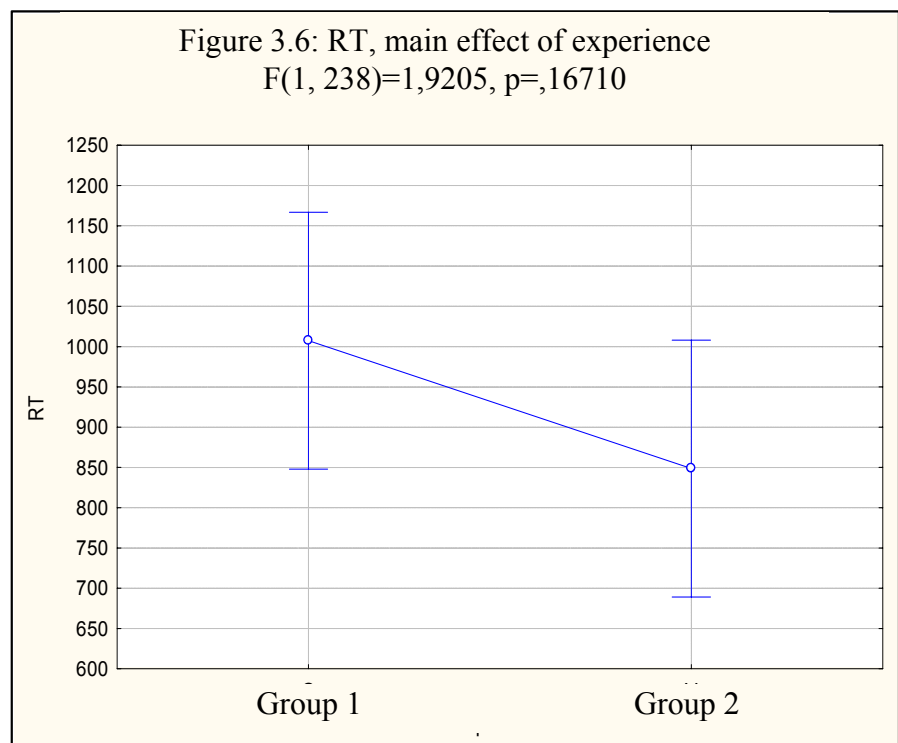
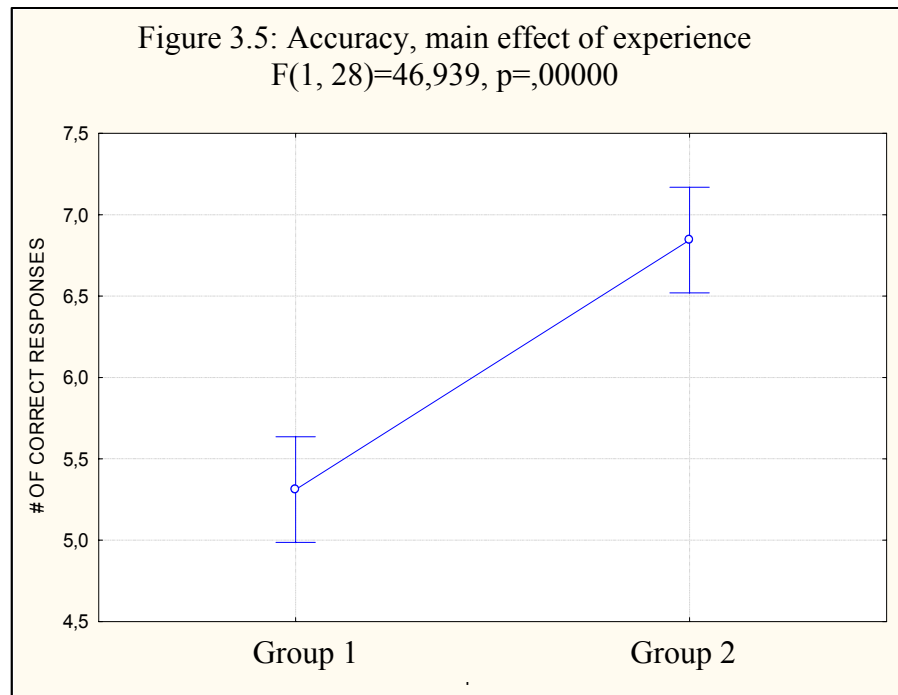
3.1 Experiment 1

Significant main effect of condition was found for both accuracy ($F(2, 56)=11,624$, $p<,001$, Figure 3.1) and reaction time ($F(2, 476)=19,111$, $p<,001$, Figure 3.2). No interaction was found between condition and experience neither for accuracy nor for reaction time (Figure 3.3 and Figure 3.4). When results across conditions are summed up, the more experienced group 2 performed significantly better with respect to accuracy only ($F(1, 28)=46,939$, $p<,001$, Figure 3.5). Reaction time was not significantly different (see Figure 3.6).

Post hoc Scheffe tests and the less strict Tukey HSD tests were conducted to discover which conditions differ significantly. With respect to accuracy, it is the congruent and initial phoneme congruent conditions for group 1 ($p<,05$, Table 3.1). Group 2 approaches significance for this comparison in Tukey HSD test ($p=,055703$, Table 3.2). When reaction time is taken into account, the difference between congruent and both incongruent conditions is significant for group 2 in both tests ($p<,001$, Table 3.3 and Table 3.4], whereas the only significant difference for group 1 can be found in Tukey HSD test when congruent and initial phoneme congruent conditions are compared (Table 3.4).







Exp:		1			2		
Exp.	Cond.	CO	IP	IC	CO	IP	IC
1	CO		0,027736	0,354208	0,046603	0,999997	0,999264
	IP	0,027736		0,880497	0,000000	0,004921	0,003316
	IC	0,354208	0,880497		0,000071	0,331572	0,046603
2	CO	0,046603	0,000000	0,000071		0,151796	0,354208
	IP	0,999997	0,004921	0,331572	0,151796		0,997894
	IC	0,999264	0,003316	0,046603	0,354208	0,997894	

Table 3.1: Scheffe post-hoc test for accuracy, significant difference ($p < ,05$) in red.

Exp:		1			2		
Exp.	Cond.	CO	IP	IC	CO	IP	IC
1	CO		0,006336	0,181376	0,013598	0,999992	0,998188
	IP	0,006336		0,771870	0,000125	0,001054	0,000561
	IC	0,181376	0,771870		0,000129	0,162598	0,013598
2	CO	0,013598	0,000125	0,000129		0,055703	0,181376
	IP	0,999992	0,001054	0,162598	0,055703		0,994830
	IC	0,998188	0,000561	0,013598	0,181376	0,994830	

Table 3.2: Tukey HSD test for accuracy, significant difference ($p < ,05$) in red, approaching significance in bold.

Exp:		1			2		
Exp.	Cond.	CO	IP	IC	CO	IP	IC
1	CO		0,076142	0,275340	0,657137	0,463693	0,320923
	IP	0,076142		0,994719	0,000176	0,993574	0,998478
	IC	0,275340	0,994719		0,001891	0,999992	1,000000
2	CO	0,657137	0,000176	0,001891		0,000989	0,000310
	IP	0,463693	0,993574	0,999992	0,000989		0,999887
	IC	0,320923	0,998478	1,000000	0,000310	0,999887	

Table 3.3: Scheffe post-hoc test for reaction time, significant difference ($p < ,05$) in red.

Exp:		1			2		
Exp.	Cond.	CO	IP	IC	CO	IP	IC
1	CO		0,019112	0,118062	0,458404	0,261050	0,148712
	IP	0,019112		0,987263	0,000028	0,984696	0,996204
	IC	0,118062	0,987263		0,000180	0,999979	0,999999
2	CO	0,458404	0,000028	0,000180		0,000087	0,000036
	IP	0,261050	0,984696	0,999979	0,000087		0,999711
	IC	0,148712	0,996204	0,999999	0,000036	0,999711	

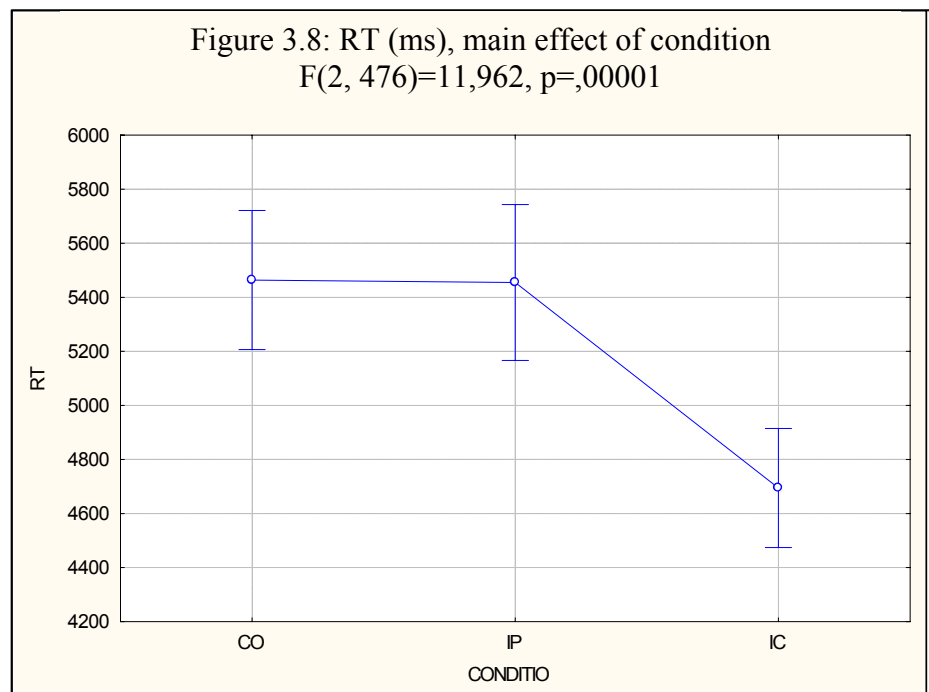
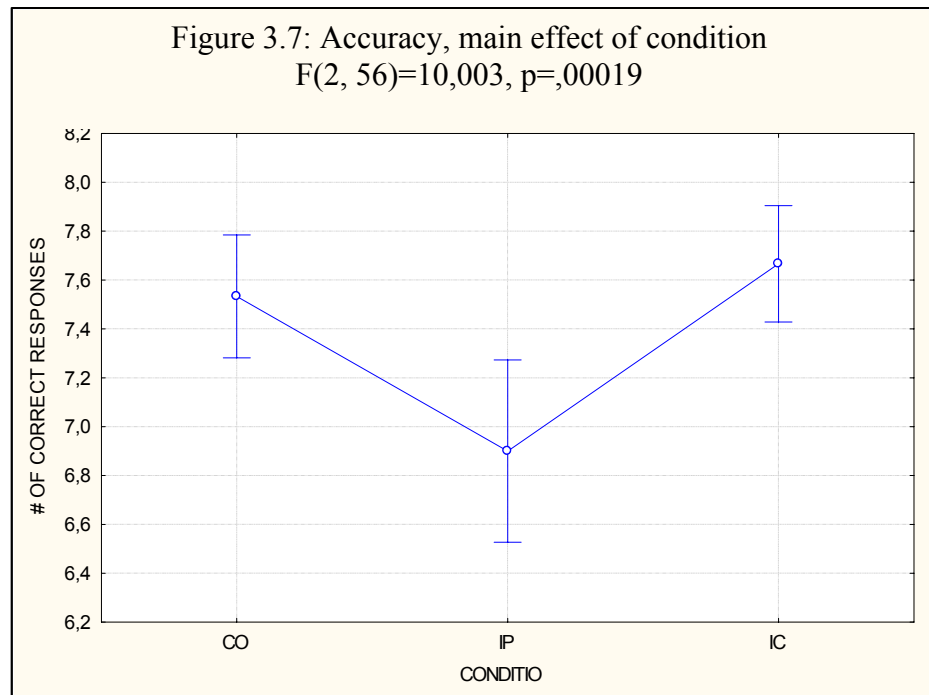
Table 3.4: Tukey HSD test for reaction time, significant difference ($p < ,05$) in red.

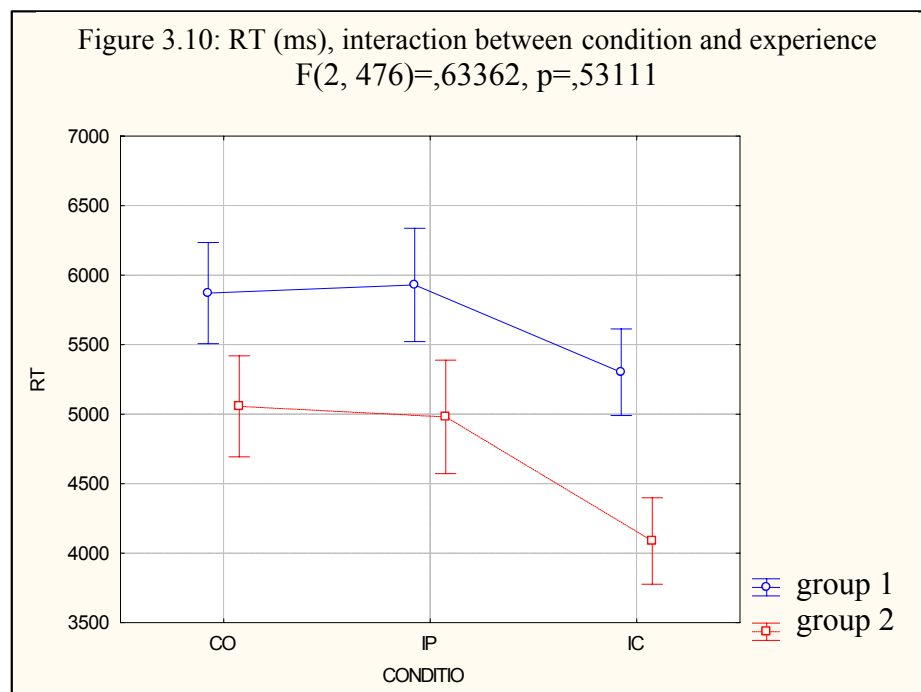
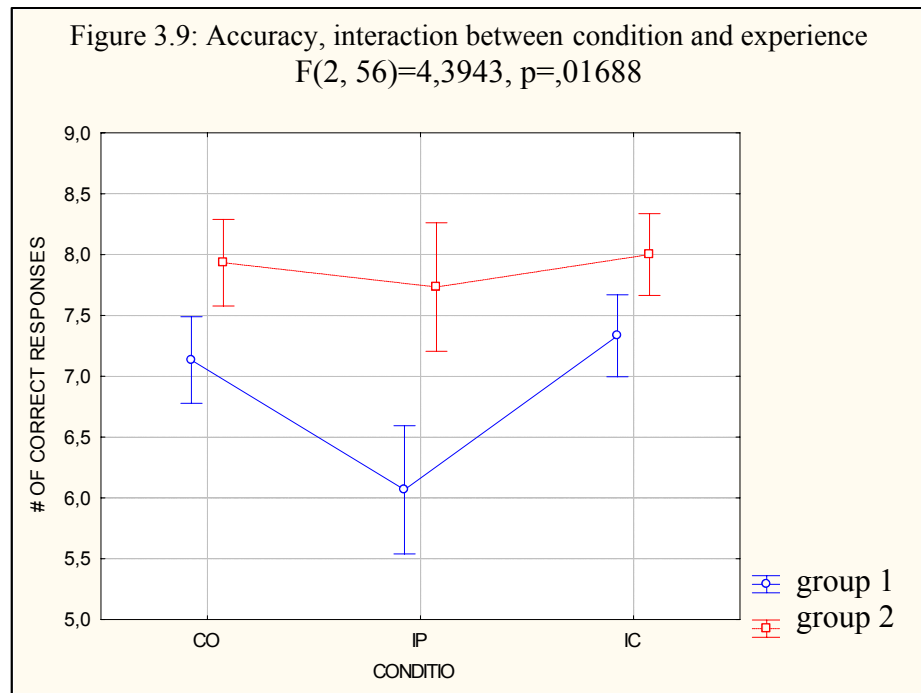
3.2 Experiment 2

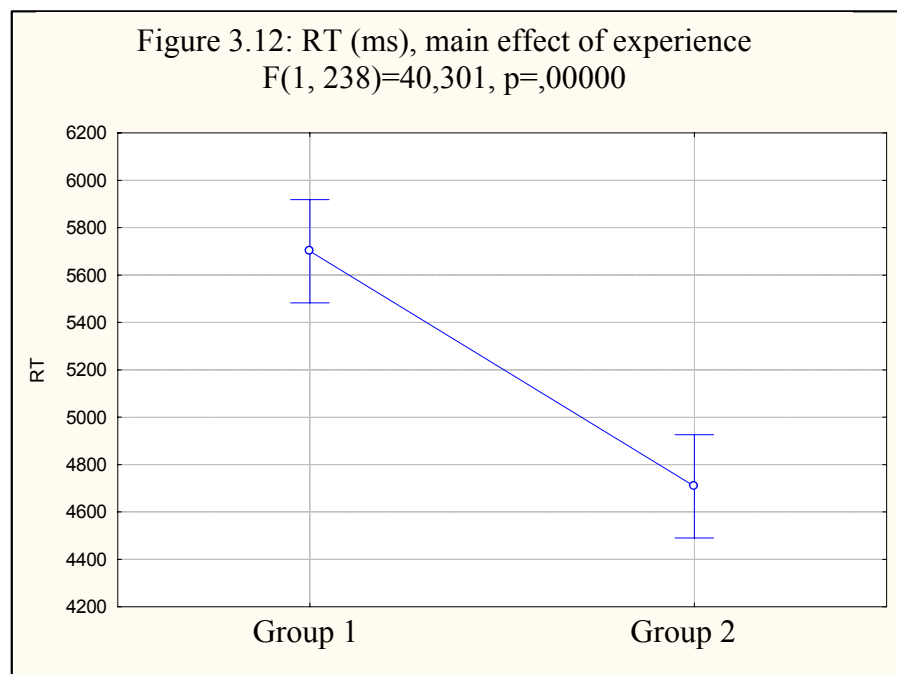
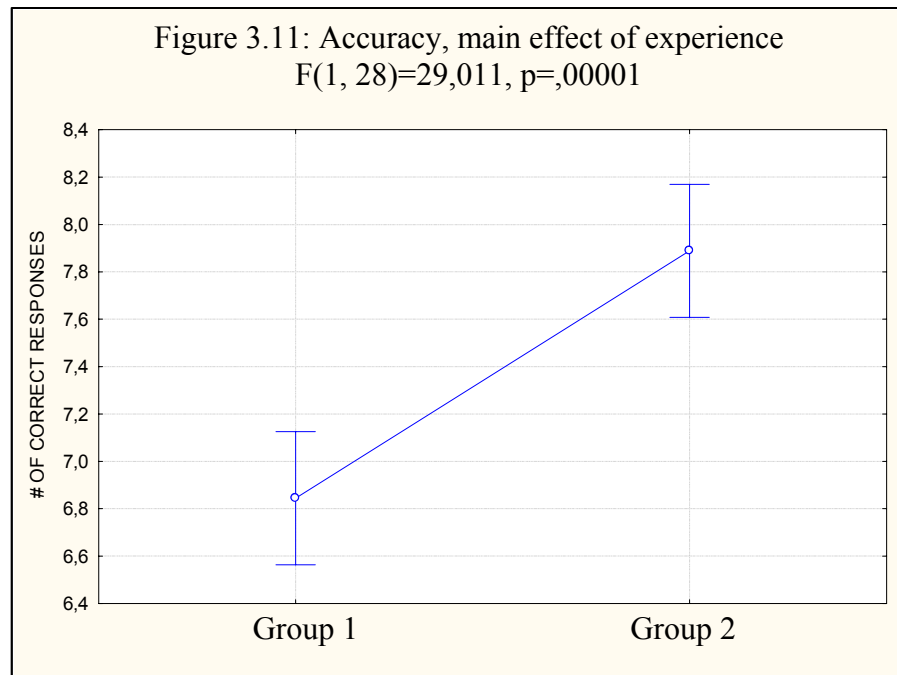
Significant main effect of condition was found for both accuracy ($F(2, 56)=10,003$, $p<,001$, Figure 3.7) and reaction time ($F(2, 476)=11,962$, $p<,001$, Figure 3.8). Unlike in experiment 1 and 3, there is interaction between condition and experience but only for accuracy ($F(2, 56)=4,3943$, $p<,05$, Figure 3.9), not for reaction time (Figure 3.10). When results across conditions are summed up, the more experienced group 2 performed significantly better both with respect to accuracy ($F(1, 28)=29,011$, $p<,001$, Figure 3.11) and reaction time ($F(1, 238)=40,301$, Figure 3.12).

Scheffe post hoc tests reveal that group 1 performed with respect to accuracy significantly worse in partially incongruent condition compared to both congruent and completely incongruent conditions ($p<,05$, Table 3.5). There was no significant difference between the three conditions with respect to accuracy for group 2.

On the other hand, when reaction time is taken into consideration, it is only the more experienced group 2 which shows significant difference. This group performed significantly faster in completely incongruent condition compared to both congruent and partially incongruent conditions ($p<,05$, Table 3.6). Reaction times of group 1 were not significantly different among these three conditions.







Exp:		1			2		
Exp.	Cond.	CO	IP	IC	CO	IP	IC
1	CO		0,009547	0,987598	0,365320	0,502161	0,117582
	IP	0,009547		0,001039	0,000002	0,002335	0,000001
	IC	0,987598	0,001039		0,502161	0,855185	0,567155
2	CO	0,365320	0,000002	0,502161		0,987598	0,999937
	IP	0,502161	0,002335	0,855185	0,987598		0,956047
	IC	0,117582	0,000001	0,567155	0,999937	0,956047	

Table 3.5: Scheffe post-hoc test for accuracy, significant difference ($p < ,05$) in red.

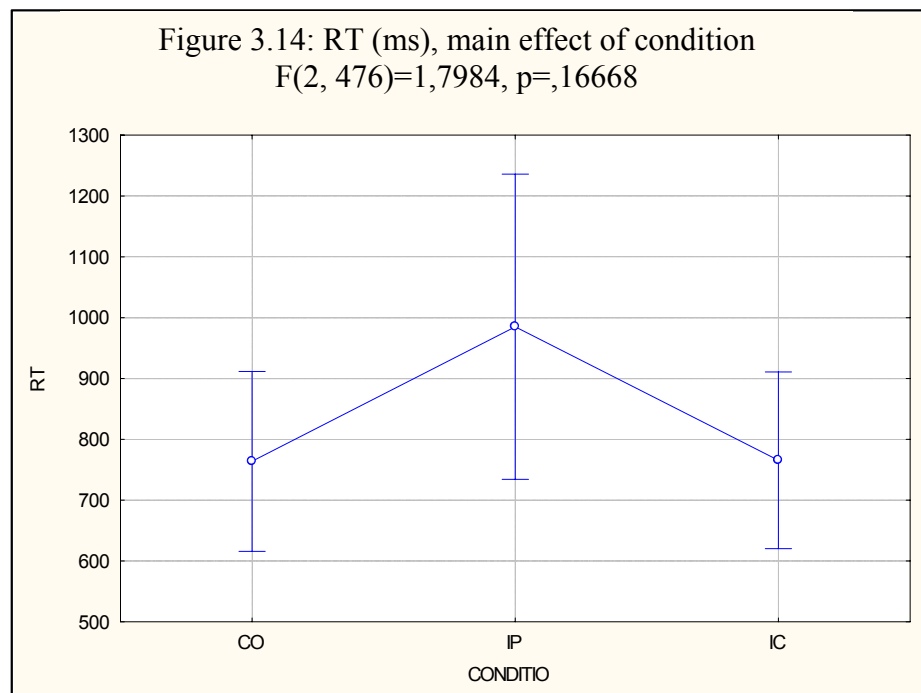
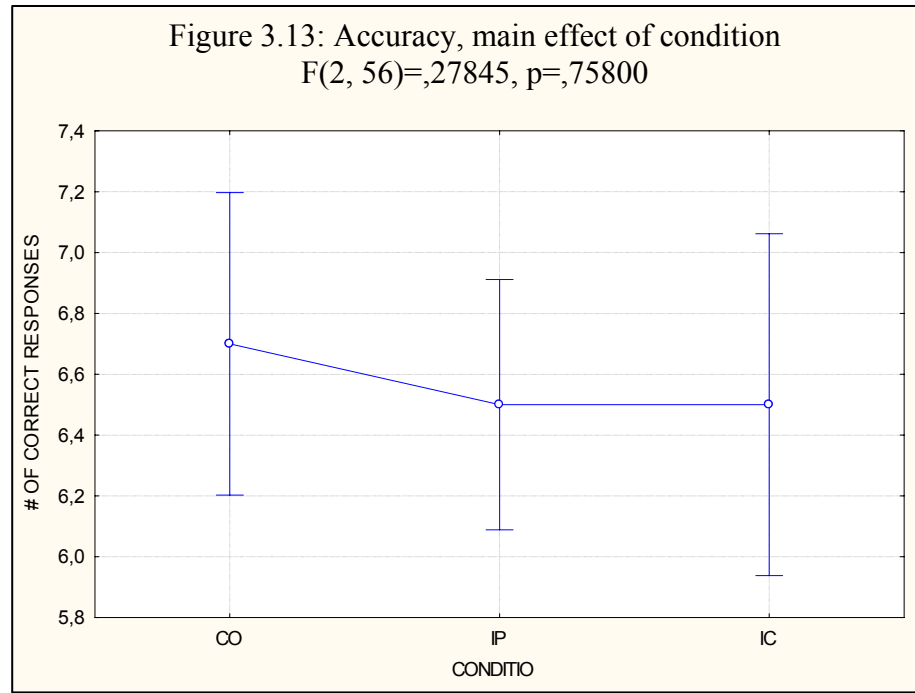
Exp:		1			2		
Exp.	Cond.	CO	IP	IC	CO	IP	IC
1	CO		0,999966	0,420424	0,111933	0,040591	0,000000
	IP	0,999966		0,303144	0,048448	0,033978	0,000000
	IC	0,420424	0,303144		0,971509	0,910867	0,001609
2	CO	0,111933	0,048448	0,971509		0,999875	0,014196
	IP	0,040591	0,033978	0,910867	0,999875		0,033404
	IC	0,000000	0,000000	0,001609	0,014196	0,033404	

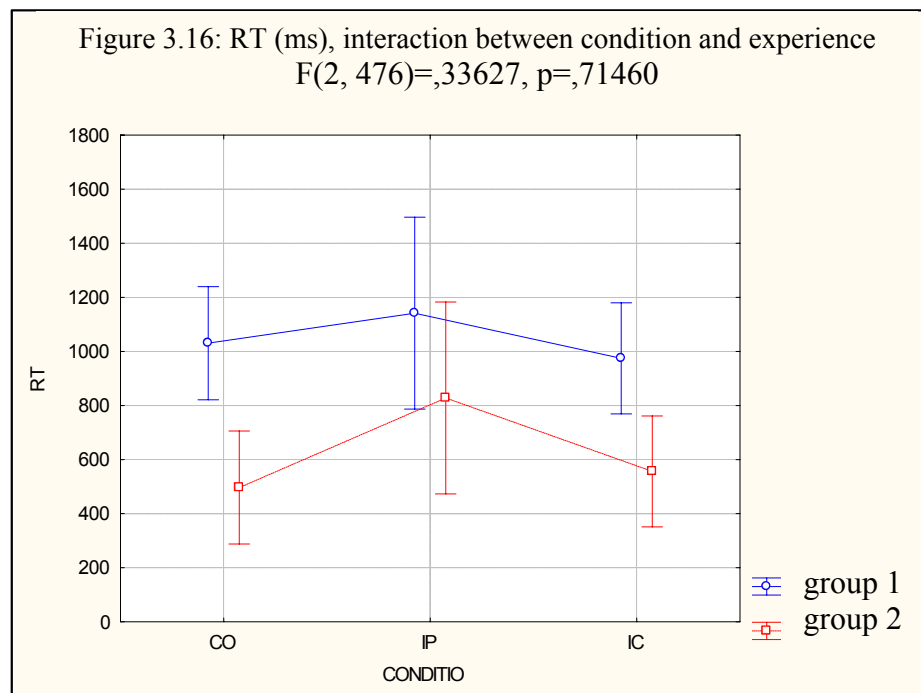
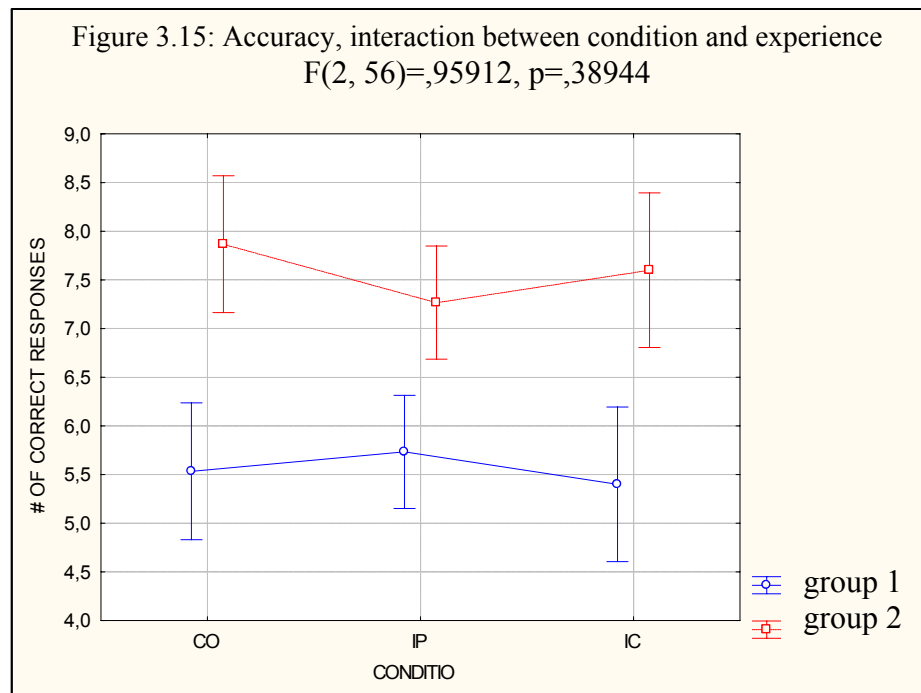
Table 3.6: Scheffe post-hoc test for reaction time, significant difference ($p < ,05$) in red.

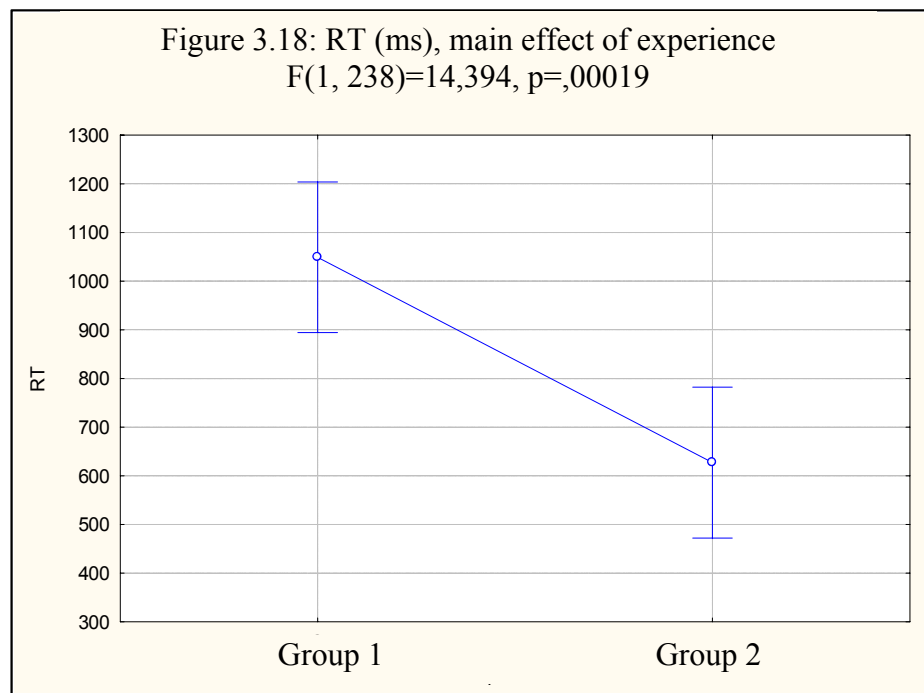
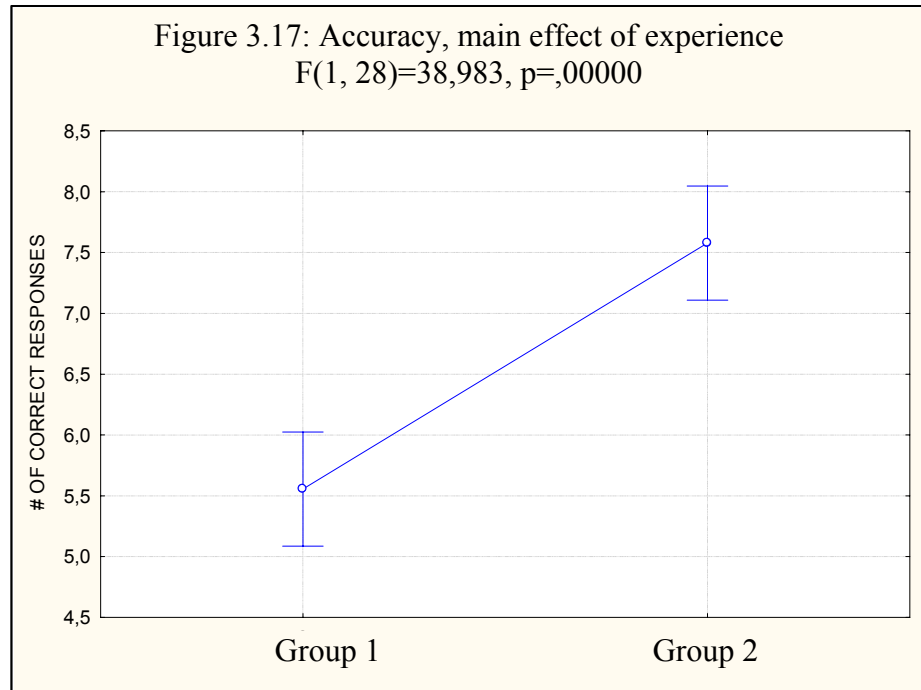
3.3 Experiment 3

Unlike in the previous two experiments, significant main effect of condition was found neither for accuracy (Figure 3.13) nor for reaction time (Figure 3.14). There was no interaction between condition and experience for both accuracy and reaction time (Figure 3.15 and Figure 3.16). Finally, as was the case in both previous experiments, main effect of experience was found for both accuracy ($F(1, 28)=38,983, p<,001$, Figure 3.17) and reaction time ($F(1, 238)=14,394, p<,001$, Figure 3.18).

Since there was no significant main effect of condition found in this experiment, Scheffe post hoc tests could not reveal any difference between individual conditions.







3.4 Native listener

The experiments were not intended for comparison with native listeners because of its design and SNR level (see section 2.2 Materials for details). Also, the hypothesis is stated exclusively for non-native listeners and its validation or rejection could be based on comparison of performance in the three experimental conditions. In spite of all these, one native listener was asked to participate to report any flaws of the materials and procedure.

As expected, the subject reached the ceiling level with respect to accuracy with 8 correct (100 %) responses for each of the three conditions in all three experiments. The same can be said about reaction times which reveal significant main effect of condition in none of the three experiments. Figure 3.19 captures slight tendency toward shortest mean reaction times in congruent condition. The patterns are somewhat similar with the results of the non-native listeners in experiments 1 and 3 but different in experiment 2 in which the mean reaction time is longer in completely incongruent condition. These observations, however, make no claim for validity because of insufficient numbers of native listener measurements and unsuitable experimental parameters for native listeners.

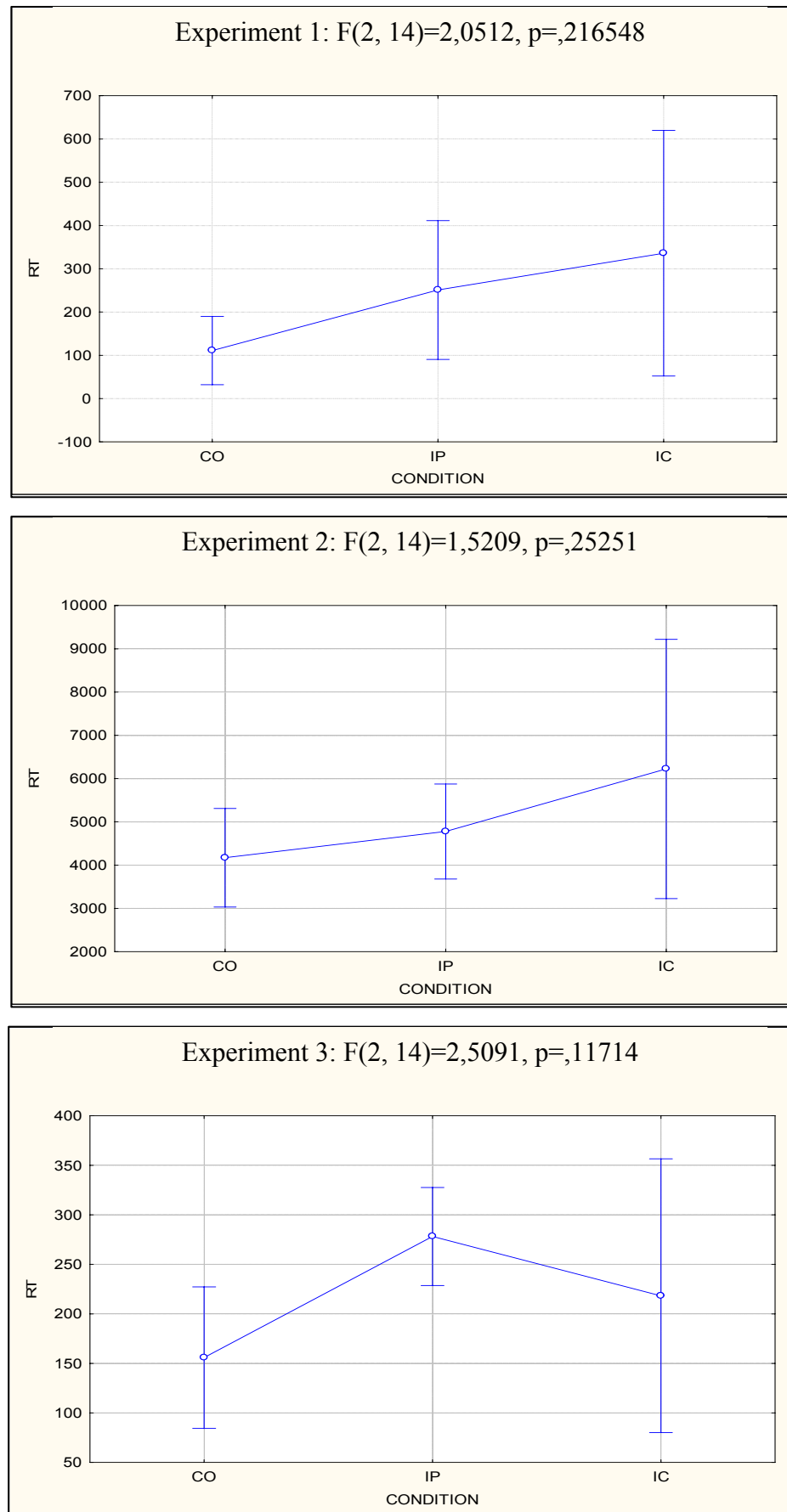


Figure 3.19: Reaction time (ms), main effect of condition, one native listener.

4 Discussion

In this section, we will try to relate results presented in the previous section to our initial hypotheses stated in section 1.5 Research questions and experiment design. For the sake of clarity, the hypotheses are summed up in Table 4.1.

Hypothesis 1: If the listeners do utilize semantic context available in audio and/or visual mode, then ... (H1)	
Hypothesis 0: If the listeners do not utilize semantic context available in audio and/or visual mode, then ... (H0)	
Comparing congruent and both incongruent conditions	
Exp. 1 & 3	<p>H1: accuracy is significantly higher and RT is significantly lower in CO than in IC condition.</p> <p>H0: there is no significant difference in accuracy or in RT when CO and IC conditions are compared because subjects recognize each word separately.</p>
Exp. 2	<p>H1: accuracy is comparable in CO and IC condition whereas RT is significantly lower in IC.</p> <p>H0: there is no significant difference in accuracy or in RT when CO and IC conditions are compared because subjects take no advantage of simultaneous processing.</p>
Effect of partially incongruent (IP) condition	
Exp. 1	<p>H1: some targets in IP condition might be substituted by the CO target sharing initial phoneme. Subjects compensate for noise masking using their semantic knowledge and audio priming (fade in masking pattern brings clearer information about the initial part of the target word).</p> <p>H0: no such attraction toward CO target is observed.</p>
Exp. 2	<p>H1: RT comparable with CO condition in combination with lower accuracy show that subjects are satisfied with partial congruence and skip evaluating details making use of quick simultaneous processing.</p> <p>H0: longer RT in combination with comparable accuracy show that subjects inspect the picture more closely to compare all details mentioned in the sentence with the picture after listening.</p>
Exp. 3	<p>H1: some targets in IP condition might be substituted by the name of the object which is in the picture in the position of the described object. Subjects compensate for the masking by expecting that the picture is completely congruent with the phrase.</p> <p>H0: subjects recognize the targets correctly disregarding the incongruence between picture and the phrase.</p>

Table 4.1: Summary of the study hypotheses.

4.1 Experiment 1

The results of experiment 1 for both groups follow a very similar pattern with highest accuracy and lowest reaction time in congruent condition. Statistical analysis proved significant main effect of condition which allows us to discard the null hypothesis (H0) that there is no influence of condition and that the listeners recognize words more or less separately from each other. Apparently, subjects were able to utilize semantic context of the carrier sentence to compensate for masking of the target word.

Post hoc tests revealed that the most clear significance was found when reaction times between congruent and both incongruent conditions were compared for group 2. This would suggest that more experienced listeners use the top-down processing more effectively. Paradoxically, in post experimental discussions the subjects especially in this group confessed that they were initially perplexed when they heard incongruent stimuli but tried to accommodate for such conditions throughout the experiment and finally expected that the sentence does not necessarily have to make sense. This might be seen as a flaw of the procedure methodology; nevertheless, the results still show very clearly the effect of condition on reaction time. The subjects did not refrain from evaluating the semantic message of the sentence altogether and retained the advantage of semantic congruency.

Reaction times did not differ so clearly for group 1 which might suggest that the less experienced listeners might process the sentence in more word-by-word manner. Nevertheless, their reaction times were significantly longer in partially incongruent condition in which the initial phoneme was congruent with the most probable completion. This observation supports the idea that listeners were in fact disturbed by target words phonetically related to their alternatives semantically congruent with the rest

of the sentence which means that they did evaluate semantic message of the carrier sentence.

This reasoning may be taken further when accuracy is considered. The difference in accuracy actually reached significant level only when congruent and initial phoneme congruent conditions were compared for group 1. The more experienced group 2 approached this level. When congruent and completely incongruent conditions were compared, the difference was not significant. This might be caused by heavier processing load in partially incongruent condition because in addition to semantic discrepancy between the carrier sentence and the target word there is the phonetic similarity of what the listeners expect to hear and the actual target. Exact analysis of substitutions and their effect on the final results was not carried out, nevertheless, see section 4.5 Substitutions below for several examples of such replacements.

When the two experimental groups are compared, significant main effect of experience is revealed. The more experienced group 2 performed with significantly higher accuracy in all three conditions than group 1 while the reaction time was not significantly different from group 1. A possible explanation may be that listeners already at lower experience level are used to allow themselves quite specific amount of time for processing information and when this interval is over they feel an urge to give a response regardless of accuracy. So while accuracy is higher for the more experienced group 2 the listeners of which have been trained in listening to English for much longer time, reaction times do not show significant difference between these two groups.

Since accuracy was lower and reaction times higher in both incongruent conditions and especially in initial phoneme

congruent condition, we can conclude that listeners are indeed able to make use of semantic information and that they are disturbed by phonetic versus semantic discrepancy.

4.2 Experiment 2

Results of this experiment are more difficult to assess with respect to our hypotheses. Analysis showed significant main effect of condition for accuracy. As expected, post hoc tests revealed for both groups that accuracy is not significantly different when congruent and completely incongruent conditions are compared. Subjects judged these stimuli with comparable and relatively high accuracy. On the other hand, accuracy drops down significantly for group 1 (and somewhat also for group 2) when partially incongruent condition is compared with both congruent and completely incongruent conditions. The expectation was that this is a sign of quick processing without much regard for details. Nevertheless, analysis of interaction between condition and experience revealed that the two groups responded in a significantly different manner with respect to partially incongruent condition compared to the other two. This suggests that the reason for worse performance of group 1 in partially incongruent condition lies in the fact that they are less experienced listeners and simply missed the crucial detail. The drop in accuracy performance was not significant for group 2. Therefore, it should not be regarded as a proof against the null hypothesis.

On the other hand, when reaction time is considered, no significant difference between congruent and partially incongruent conditions was found for both groups which is in accordance with our prediction. We also hypothesized that reaction time should drop significantly for completely incongruent condition. Even though the reduction reached significant level only for the more experienced group 2, the

reaction time pattern is very similar for both groups and, unlike in the case of accuracy, there is no significant interaction between condition and experience. Therefore, the fact that group 1 did not perform with significantly lower reaction time in this condition should not be attributed to their lower experience. The reason might be in uneven experience level of individual subjects of this group which produces higher result variance. While group 2 consisted of university students focusing on English, grammar school students do not necessarily have to be excellent in languages. Still, there is an obvious tendency towards shorter reaction times in completely incongruent condition also for group 1.

With these observations in mind, we conclude that the listeners are indeed evaluating a sentence simultaneously with a picture and thus they can discover complete discrepancy quite early. On the other hand, they need approximately same time for decision when the context of the sentence is congruent with the picture except for one detail. The processing load is the same as in completely congruent condition which is understandable because the subjects concentrate on the details in these two conditions in the same way.

As already mentioned and explained, accuracy showed significant interaction between condition and experience. Less experienced listeners of group 1 had more troubles identifying the details in partially incongruent condition and thus sometimes judged the sentences as completely describing the pictures. Otherwise, listeners of group 2 were not significantly more accurate in the two remaining conditions. This shows that when listeners are forced to judge congruency of spoken and visual information, experience plays important role in details recognition

whereas pragmatic accommodation to overall task demands can be found already in less experienced listeners.

4.3 Experiment 3

This experiment failed to bring significant main effect of condition both on accuracy and on reaction time for both groups. The results support null hypothesis because no difference was found in listeners' performance. Nevertheless, the problem might be in the experimental design itself. Although SNR was more adverse in this experiment than in experiment 1, subjects were not forced enough to make use of the pictures to recognize the target word. Carrier sentences were always semantically congruent with the target words which were relatively highly predictable from the semantic context only. For illustration, in the sentence "My father doesn't have time for lunch at work, he eats only a hamburger." the words "lunch" and "eats" semantically prime the target word "hamburger". Listeners might have ignored the pictures completely because only one third of them were completely congruent with the information incoming in the acoustic mode, the rest of them being deviant from it.

In a future experiment, it would be advisable to employ low predictability sentences. For example, carrier sentence could be of the type "What is the Czech word for ___?" with the object representing the target word present or absent in a picture. However, the link between acoustic and visual information would be quite weak.

Another possibility is to develop the paradigm originally used by Garnes and Bond (1976) in acoustic mode only. It would require finding pairs of target words differing in one phoneme only (minimal pairs), which is unfortunately not a simple task with limited vocabulary of non-native listeners. Several tokens of the target words could be created through the manipulation of

the acoustic property cueing the particular distinctive feature of the phoneme in which the two original words differ along a continuum. In this manner, we could arrive at sentences as “There is a pin / bin on the floor.” combined with pictures containing one of these two objects. The aim would be to find out whether the boundary between the two phonemes shifts towards the target presented in the visual mode, in other words, whether there is a trade-off between acoustic and visual information.

Although no statistically significant evidence for the original hypothesis was found in this experiment, the following lines discuss the slight differences in the mean values. Since the failure might be caused by the methodological shortcomings reported above, these observations might be useful for future enhancements.

Accuracy is relatively highest for group 2 in congruent condition followed by completely incongruent and lowest in partially incongruent condition. Group 1 performed unexpectedly in an opposite manner with relatively highest mean accuracy in partially congruent condition. As already mentioned, the visual stimuli do not facilitate target word recognition in this experiment design. On the other hand, when reaction time is considered, relatively highest mean values were obtained for partially incongruent condition for both groups which suggests that this condition caused highest disturbance to the subjects and that they might have considered the pictures to some extent.

While group 1 performed with almost the same mean reaction time in both completely congruent and completely incongruent conditions, group 2 had mean reaction time slightly shorter in the congruent one. These observations support the idea that the effect of the visual mode can more likely be termed as disturbance. When the sentence named relevant objects in the

picture but differed in the target word which could not be found in visual mode, subjects could be confused and their reaction time was higher (and accuracy lower for group 2). When the two modes of presentation were congruent they had no troubles finding and validating their response in the picture. And on the other hand, when there was complete incongruence, they discarded evaluating the picture quite early in the initial portion of the sentence. Indeed, this reasoning was supported by post experimental discussions in which listeners reported their strategy of ignoring the picture when it was apparently incongruent with the sentence and concentrating purely on the acoustic mode. They took advantage of the fact that all the carrier sentences were semantically congruent with the target words which was relatively highly predictable so this was a safe way for them.

The more experienced group performed overall significantly better both with respect to accuracy and reaction time. Although we reported above slightly different pattern of results for accuracy when the two groups were compared, no interaction was found between condition and experience for accuracy or for reaction time

4.4 Cross-experiment analysis

We hypothesized that when results in congruent condition in experiment 1 and 3 are compared, recognition accuracy should be higher and reaction time lower when sentence is accompanied by a congruent picture. Nevertheless, as mentioned already in discussion of experiment 3 results, it seems more likely that this reasoning cannot be substantiated because mean accuracy is only very slightly (non-significantly) higher for group 2 and even slightly lower for group 1 in experiment 3. Furthermore, both groups performed with longer reaction times in experiment 3 which clearly suggests that processing the two modes of infor-

mation simultaneously brings higher processing load and poor or no benefits.

Comparison of experiments 2 and 3 in order to reveal whether subjects are considering visual information in the same way is quite difficult. Accuracy is quite flat for both groups in experiment 3 and drops significantly for group 1 in experiment 2 which has already been explained as a result of lower experience. Reaction time is slightly higher in partially incongruent when compared to congruent condition in experiment 3 but no difference can be found for the same comparison in experiment 2. As already explained, we consider it a sign of comparable demands of these two conditions (listening for detail) in experiment 2 and higher processing load and disturbing effect in experiment 3.

4.5 Substitutions

Response processing yielded several interesting tokens of misunderstandings and substitutions of the target word with a different word phonetically and/or semantically related to it. As Jenkins does in her chapter which served as a starting point for this thesis, a few of these tokens will be presented and briefly commented on. They serve only for illustrations and were not statistically analyzed.

It would be difficult to define precisely closeness of phonetic and semantic relationship between the target word of a stimulus and the response given by a subject. When phonetic relationship is concerned, the decisive factor for vowels is their relative location in the vowel space (high/middle/low, front/central/back) and for consonants it is the place and manner of articulation together with voicing. When judging phonetic closeness, differences between English and Czech phoneme inventories which may be the source of misunderstanding are con-

sidered as well. See Ladefoged (2001) for an overview of English phonemic inventory, and Palková (1994) for Czech phonemic inventory. There are means of analyzing words semantically based on their constituent semantic features which allow us to group them into semantic classes. Nevertheless, for the purpose of this overview, semantic relations were considered predominantly on our “common sense” judgments.

4.5.1 Experiment 1

Substitutions of the target word with its most probable competitor semantically congruent with the carrier sentence were expected in the initial phoneme congruent condition. Here the gradual masking and thus clearer initial acoustic information was intended to activate the competitor. Indeed, even a balanced pair of such substitutions was observed in both groups, where “autumn” was recognized as “author” by one subject and vice versa by another one (ex. 1 in Table 4.2). These two words share more than the initial phoneme ([ɔ] + alveolar/dental obstruent + [ə] /+nasal). Two subjects in group 2 showed semantic influence as well (ex. 2). They both provided “awful” instead of “author”, these two words again being phonetically close ([ɔ] + labial/dental fricative, only the labial one exists in Czech + [ə] /+ [ɸ]). Example 3 from the same condition again represents tokens of semantic influence because all three “electrons”, “electronic” and “elements” are semantically related to the carrier sentence and the priming of the initial chain [ɛɛ...] can be seen as well.

Group 2 provided two more examples (4 and 5) of substitutions which were expected to happen. As a result of phonetic priming, “instrument” was recognized instead of “interview”. Moreover, the phrase “musical instrument” apparently forms a semantically tight unity. The second example can be attributed to

the initial chain of three shared phonemes. Example 6 does not offer the substitution which was intended but one which is still semantically congruent and phonetically closer to the actual target word ([d] + [p]/[ɔ] instead of intended [ɛ]/[ɔ]). Example 7 presents three tokens of substitutions done by one subject going from the most phonetically motivated to the least which happened already in completely incongruent condition where there was no initial phoneme priming. Different degrees of semantic enhancement can be seen as well: while “authors” substituted with “uniforms” makes a big step towards semantic congruity, to arrive at “government” saving the football match instead of “goalkeeper” requires much wider semantic imagination.

All these were examples in which semantic relationship might be said to influence subjects’ final decision. Several cases appeared as well in which subjects provided answers without apparent reflection of the overall semantics of the stimulus as a whole. Example 8 shows a few tokens from completely incongruent condition in which no particular substitution based on phonetic priming was expected. From this point of view, replacing an incongruent target word with another incongruent word as a response without any noticeable semantic enhancement should not be surprising. Apart from the fact that the word does not exist in English (it is probably a token of morphological transfer from Czech), going to watch a movie to the theoretist is as good as going to watch it to the daughters. Most of the substitutions can be ascribed to phonetic closeness especially in the initial portion of the two words in question (“headache” / “heading”).

Nevertheless, as example 9 illustrates, several subjects did not recognize words in congruent conditions and substituted them with semantically less congruent words instead. Whereas

drinking tea with juice from “eleven” makes sense as an elliptical sentence and shows interesting phonetic processes connected with stress placement and vowel reduction, the other two sentences are semantically quite poor. However, substituting “energy” with “television” might indeed be a proof of semantic influence completely dominating acoustic information as is the case of example 10. Discussion of these cases leads to a more general question of psychology and decision making which is far beyond the scope of this thesis.

Generally speaking, several examples of substitution in the direction of semantic congruency were found mostly in initial phoneme congruent condition illustrating that the subjects were taking semantics of the carrier sentence and the target word in these examples into account.

Ex.	Subject	Group	Stimulus presented in the experiment (response)	Cond.
1	KH / MO	1 / 2	The season which comes after summer is author. (autumn)	IP
	OL / SV	1 / 2	The book was written by two autumns. (author)	IP
2	PM1+CT	2	The season which comes after summer is author. (awful)	IP
3	TV	1	Computers and other machines need lot of electric elephants. (electrons)	IP
	JS	1	Computers and other machines need lot of electric elephants. (electronic)	IP
	JC	2	Computers and other machines need lot of electric elephants. (elements)	IP
4	RT	2	Piano is my favorite musical interview. (instrument)	IP
5	PC	2	In some British schools children have to wear universe. (uniforms)	IP
6	KS	2	My tooth aches, I have to go to the daughter. (doctor)	IP
7	PM	2	My favorite animal in the zoo is a big grey energy. (elephant)	IP
			The last football match saved our excellent gallery. (government)	IP
			In some British schools children have to wear authors. (uniforms)	IC
8	TV	1	Stars and planets together form the bathroom. (basket)	IC
	TV	1	I have to study hard for my English headache. (heading)	IC
	PC	2	His friends gave him many gifts for his twentieth birthday. (thirty four)	IC
	CT	2	The book was written by two birthdays. (birth case)	IC
	CT	2	I don't like watching movies at home, I go to the daughters. (theoretist)	IC
9	MC	1	Computers and other machines need lot of electric energy. (television)	CO
	JK / IF	1 / 2	The book was written by two authors. (office)	CO
	SV	2	I like tea with sugar and juice from a lemon. (eleven)	CO
10	PC	2	My favorite animal in the zoo is a big grey interview. (bear)	IC

Table 4.2: Examples illustrating semantic influence on subjects' responses.

4.5.2 Experiment 3

As was the case of experiment 1, several interesting interactions this time between acoustic and visual information were recorded. The analysis is based on phonetic comparison of the target word and the response provided and on picture inspection aiming to find possible explanations for individual substitutions.

These substitutions were most likely to happen in partially incongruent condition in which the picture could be related to the context of the carrier sentence apart from the target word itself. Several subjects substituted “palace” for “Paris” (one of them even in congruent condition, example 1). This was not intended during the stimuli preparation phase in which we decided not to search for objects the names of which begin with the same phoneme as the target words (limited vocabulary and picture bank reasons, see section 1.5 above for detailed explanation). Nevertheless, the picture by chance depicts the Eiffel Tower, a symbol of Paris. Moreover, these two words are phonetically very close and the indefinite article could easily be overheard due to vowel reduction.

Example 2 illustrates two more proofs of picture evaluation in partially incongruent condition. Again, quite unexpectedly, one subject responded with “apples” instead of “airport” referring probably to the tree depicted in the picture. Although the phrase “modern apples” is not quite congruent, it resembles “airport” ([æp...]/[ɛəp...]) much more than the word “museum” which provided another subject this time disregarding acoustic information completely and referring clearly to the building with a dinosaur on its roof.

Another substitution in partially incongruent condition is presented in example 3 in which one subject replaced “cabbages” with “cottages”. A cottage can be found in the picture;




nevertheless, planting them does not make much sense. It seems more likely that the acoustic information played the more prominent role, especially when the second subject's reaction in this example is considered: although in congruent condition (picture depicting cabbages) the response was "oranges". All these three words share the middle vowel [ɪ] and the final cluster [dʒɪz] which should be masked by the noise, nevertheless, as a study by Lacumberri and Cooke (2006) suggests, affricates are quite resistible to masking.

Examples 4, 5 and 6 capture interesting substitutions in completely incongruent condition based on visual information. In the first one, the subject's response was probably primed phonetically by [m] + vowel + [ʃ]. Taking the objects in the picture into consideration, responding with "machines" instead of "mushrooms" seems quite logical. The following example 5 lacks any phonetic priming at all, in this case the subject decided to simply name what the picture depicts, even though the sentence semantics is quite poor. In the last example in this set the subjects provided different endings for the actual target "colonies" under the unintended influence of the picture which was created only as a distracter sharing no objects with the sentence. Nevertheless, the scene may indeed be seen as portraying a colonist or the time of colonialism (rather than the incorrectly formed "colonism" provided by one subject).

As was the case in experiment 1, examples 7 and 8 illustrate responses in congruent conditions in which the subjects substituted the target word which was depicted in the picture by a different word. Whereas the response in the first example is probably influenced by the phonetic cluster [laʊ] shared in the middle portion by both words, the second example response shows no phonetic relatedness at all. On the other hand, the apple

in the picture of example 8 might be mistaken with a tomato which is a possible ingredient for cooking as well, while there is no sign of a cloud in the picture of example 7 nor would it be a nice gift for anyone's birthday.

Even though it is by no means possible to generalize these examples, they might serve as illustrations of subjects' more or less successful attempts to integrate acoustic and visual information and semantic context of the sentences. The analyses make no claims to scientific precision; they are rather brief but potentially interesting discussions of substitutions discovered during results processing.

Example			Stimulus presented in the experiment	Visual stimulus
Subject	Group	Cond.		
Response				
1			I brought some souvenirs from the holiday and a postcard with a palace.	
IF + AF	2	IP		
RT	2	CO		
Paris				
2			The old church is situated near the modern airport	
OL	1	IP		
MA				
museum				
apples				
3			My grandfather has a big farm in the country where he plants cabbages.	
IF	2	IP		
MA	1	CO		
cottages				
oranges				

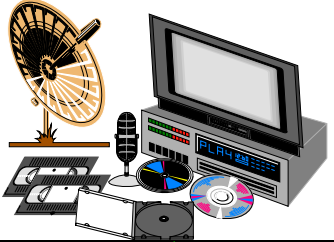

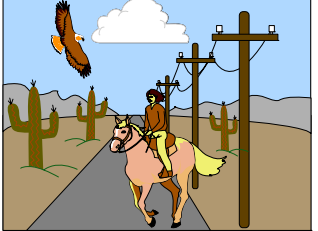


4			To prepare this meal you need cheese, apples, salt and some mushrooms.	
JS	1	IC		
machines				
5			I want to watch a film on TV but I cannot find the TV magazine.	
GD	1	IC		
wedding				
6			In this lesson we will learn more about the British colonies.	
IF + KS	2	IC		
PC				
colonist colonism				
7			I went to Jane's birthday party yesterday and I gave her some flowers.	
GD	1	CO		
clouds				
8			To prepare this meal you need cheese, apples, salt and some mushrooms.	
NK	+	1		
SH				
tomatoes				

Table 4.3: Examples illustrating influence of the pictures on target word recognition

5 Summary

This thesis tried to insert another piece into the complex mosaic of phenomenon called speech perception. After introducing fundamental challenges and findings in this field, we reviewed current state of progress in the discussion of bottom-up and top-down processes interplay. We focused especially on the ability of non-native listeners to employ higher linguistic and extralinguistic (contextual) levels information to facilitate their speech perception.

We took these findings into consideration together with slightly contradictory statements by two ELT scholars and designed three experiments to inquire into the topic. Hypotheses were formulated in connection with results predictions. In principle, if listeners are able to use supplementary information in the form of congruent sentence semantics or visual context, higher accuracy and lower reaction times were expected for congruent stimuli.

The results suggest that non-native listeners on both levels of experience do relate words of a sentence to each other semantically and establish a semantic frame with particular congruency expectations. Worse results were obtained in sentences in which the final target word semantically deviated.

Results of experiments inquiring into visual mode utilization are less straightforward. Both experimental groups proved their ability of pragmatic adaptation to task demands, even though worse results were found for the less experienced group in listening for details. Although the final experiment did not prove our hypothesis suggesting that non-native listeners do not utilize visual context for word recognition, the results suggest that they might consider pictures in the process of perception. Nevertheless, the effect is more likely to be adverse because

processing acoustic and visual information at the same time seems to be demanding and causes high processing load for non-native listeners. From this point of view, visual information is more likely to bring disturbance to perception if incongruent with acoustic information rather than its facilitation if congruent with it.

Souhrn

Zkoumání procesu percepce řeči je značně rozsáhlé téma, které s sebou nese mnoho otázek. Jednotlivé hlásky v proudu řeči běžně podléhají koartikulaci a asimilaci, navíc řečové akty často probíhají v prostředí s množstvím ruchů. Lidé se však s těmito problémy dokáží vypořádat. Vědci vypracovali několik teorií percepce řeči, které se obvykle rozdělují podle toho, zda zdůrazňují význam především důkladné analýzy akustického signálu, ze kterého posluchač vyzíská jednotlivé konstituční řečové prvky, a ty pak skládá do vyšších celků (postup zdola nahoru, bottom-up), nebo zda berou v potaz i významný vliv vyšších jazykových rovin (morfologické, lexikální, syntaktické, sémantické) a pragmatických zkušeností posluchačů s okolním světem (postup shora dolů, top-down). Současné modely percepce řeči obvykle kombinují prvky obou těchto základních přístupů.

Výzkumníci již prokázali, že posluchači jsou skutečně při porozumění řeči ovlivněni znalostmi zákonitostí vyšších jazykových rovin. Pro tuto diplomovou práci mají největší význam studie zabývající se vlivem sémantiky na výslednou percepci. Ganong (1980) například dokázal, že posluchači mají sklon kompenzovat změny v manipulovaném akustickém signálu tak, aby výsledkem percepce bylo smysluplné slovo. Dále se rovněž ukázalo, že posluchači při manipulaci prvního fonému slova zasazeného do smysluplné věty mají tendenci dané slovo rozpoznat tak, aby sémanticky do zbytku věty zapadalo (Warren a Warren, 1970; Garnes a Bond, 1976). Pro zpřesnění pozorování se v posledních letech využívají nejmodernější výzkumné technologie, jako je elektroencefalografie a magnetoencefalografie pro záznam mozkové aktivity v reálném čase, a také sledování pohybu očí pro průzkum integrace informací přicházejících v akustické a vizuální podobě. I tyto metody prokázaly, že

posluchači reagují velice brzy na sémantické, ale i syntaktické neshody v mluvené řeči (Hahne, 2001).

V otázce, zda rovněž posluchači učící se cizímu jazyku dokáží využívat informací z těchto vyšších jazykových rovin pro porozumění řeči, nepadá mezi odborníky shoda. Jennifer Jenkins (2000) ve své knize uvádí, že podle jejích zkušeností z práce lektorky anglického jazyka používají studenti při percepci řeči převážně postup zdola nahoru, a to i když dosáhnou poměrně vysoké úrovně zvládnutí daného jazyka. Dále tvrdí, že vyšší mentální zátěž při zpracování příchozího akustického signálu a vlastní nejistota v cizím jazyce jim nedovolují využít kontextové informace, lingvistické ani extralingvistické, k překlenutí možných nedokonalostí v tomto signálu. Udává několik příkladů, kdy posluchači evidentně příliš spoléhali na samotný signál obsahující obvykle nějaký druh chyby (např. ve výslovnosti), přičemž sémantika dané věty či vizuální informace z obrázku by jim tento problém pomohly odhalit a překonat.

Na druhou stranu John Field (1998) tvrdí, že studenti už na poměrně nízké jazykové úrovni dokáží dovozovat a odhadovat slova, kterým z akustického signálu v důsledku nedokonalostí v postupech typu zdola nahoru nerozuměli, a to za pomoci znalosti kontextu a celkové sémantiky dané výpovědi. Jelikož jsou tyto kompenzační strategie běžné při komunikaci v rodném jazyce, Field má za to, že by nemělo být obtížné naučit studenty cizího jazyka používat je i v této oblasti. Zatímco tedy Jenkins vidí problém při poslechu cizího jazyka v zahlcení procesní kapacity samotným akustickým signálem, Field naopak tvrdí, že postupem času si posluchači vytvářejí kompenzační postupy typu shora dolů, které jim umožní nedokonalosti ve zpracování signálu překonat

Se zřetelem k těmto dvěma názorům se tato diplomová práce tedy zabývá zkoumáním toho, jak a zda vůbec čeští studenti angličtiny dokáží využít kontextu při porozumění řeči. Problematiku zkoumáme ve třech do jisté míry nezávislých experimentech, jejichž výsledky však tvoří smysluplný celek. V úvahu bereme kontext dvojího druhu: větně sémantický a obrazový. Základem všech tří experimentů je manipulace se sémantickou/obrazovou shodou mezi jednotlivými složkami stimulu. Vyskytuje se buď úplná shoda, nebo částečná, či úplná neshoda. Pro nalezení odpovědí na výzkumné otázky se měří a analyzuje rychlost a přesnost odpovědí posluchačů.

Experiment 1 zjišťuje, zda si posluchači vytvářejí sémantický rámeček od začátku věty a zda je tudíž percepce posledního slova věty, které je maskováno šumem, ovlivněna sémantickou shodou mezi ním a zbytkem věty. Jejich úkolem je zopakovat toto poslední slovo, které může do zbytku věty zcela zapadat, nebo nezapadat. Tato druhá možnost se dále dělí na případ, kdy skutečně vyslovené slovo začíná stejným fonémem jako nejpravděpodobnější sémanticky vhodné dokončení dané věty, a na případ, kdy se žádná takováto zvuková podoba nevyskytuje. Účelem zvukové podoby v úvodní části slova je zjistit, zda posluchači skutečně v některých případech nahradí slovo vyslovené slovem sémanticky odpovídajícím. K posílení této atrakce je náběh šumu u všech těchto maskovaných slov postupný v průběhu první pětiny trvání slova. Pokud tedy posluchači sémantiku věty zvažují, měli by reagovat rychleji a s větší přesností v případě, že slovo sémanticky do věty zapadá. Vhodné substituce v případě se zvukovou atrakcí se považují opět za projev využití sémantického kontextu.

Experiment 2 zkoumá propojení informace přicházející v akustické a vizuální podobě. Úkolem posluchačů je rozhodnout,

zda věta, kterou slyší, přesně popisuje obrázek, který vidí. Ten může být opravdu zcela shodný s větou, lišit se v jediném detailu, nebo být zcela odlišný. Pokud posluchači dokáží integrovat tyto dva informační kanály průběžně, měly by jejich reakce být nejrychlejší v posledním případě, kdy neshodu lze odhalit velmi brzy. Naopak pokud je rozhodujícím prvkem pouze jediný detail, může být tento snadno přehlédnut a přesnost odpovědí klesá.

V experimentu 3 posluchači současně slyší větu, jejíž poslední slovo je opět zašuměné, a vidí obrázek. Věta samotná se svým posledním slovem je vždy sémanticky v pořádku. Avšak obrázek může buď plně odpovídat větě, lišit se v jediném detailu, který právě odpovídá poslednímu slovu věty, nebo může být zcela odlišný. Úkolem posluchačů je zopakovat toto poslední slovo věty. Pokud dokáží využít obrázek ke kompenzaci šumu v posledním slově, měly by jejich reakce být nejpřesnější a nejrychlejší, když se obrázek s větou zcela shoduje. Podobně jako v případě experimentu 1, náhrady skutečně vysloveného slova slovem pojmenovávajícím objekt na obrázku, kterým se tento od věty liší, lze opět považovat za projev kompenzační strategie.

Jelikož autoři, jejichž názory sloužily jako výchozí bod pro tuto práci, hovoří u pozorovaných studentů o rozdílných úrovních zvládnutí jazyka, rozhodli jsme se otestovat rovněž dvě skupiny posluchačů. Jediným kritériem byl počet roků studia angličtiny: skupinu 1 tvořili studenti gymnázia po 3 až 4 letech studia a skupinu 2 vysokoškolští studenti angličtiny se zaměřením na ekonomii po minimálně 9 letech studia. Každá skupina obsahovala 15 studentů.

Sestavení experimentálních materiálů bylo poněkud komplikované s ohledem na stanovené požadavky (dvojslabičná slova, pro experiment 1 navíc páry začínající stejnou hláskou s přízvukem na první slabice) na jedné straně a omezená velikost

slovní zásoby skupiny 1 na straně druhé. Slova byla vybrána ze slovníku k učebnici, ze které se studenti skupiny 1 učí, a před samotným experimentem předtestována. Tato slova byla zašuměna růžovým šumem (1/f), a to s odstupem od signálu 0 dB pro experiment 1 a -3 dB pro experiment 3. Rovněž na obrazovou prezentaci byly kladeny nároky v podobě jednotného vizuálního stylu a snadné interpretace.

Každá sémanticky neporušená věta v experimentu 1 obsahovala alespoň 2 slova vztahující se významově k poslednímu slovu. V experimentech 2 a 3 zase v případě, že obrázek zcela odpovídal větě, na něm bylo možné najít minimálně 3 objekty, které věta pojmenovávala.

Každý posluchač v rámci každého experimentu slyšel 24 vět rovnoměrně rozdělených do tří výše popsaných úrovní shody. Pouze v experimentu 2 bylo přidáno 8 dalších vět přesně popisujících obrázek, aby byly kladné odpovědi vyváženy se zápornými. Systém rozložení stimulů zajistil, že každý stimul se vyskytoval ve všech třech úrovních sémantické/obrazové shody a zároveň že žádný posluchač neslyšel tutéž větu nebo poslední slovo ani neviděl tentýž obrázek dvakrát. Posluchači měli k dispozici zkušební věty pro seznámení se s požadavky a pro nastavení hlasitosti.

Výsledky experimentu 1 pro obě skupiny prokázaly, že míra sémantické shody měla signifikantní vliv na přesnost a rychlost reakcí posluchačů. Ty byly nejpřesnější a nejrychlejší právě v případě, kdy poslední slovo věty do ní významově zcela zapadalo. Nejvýznamnější rozdíly byly zjištěny při srovnání případů s úplnou shodou a případů, kdy skutečně vyslovené slovo začínalo na stejnou hlásku jako nejpravděpodobnější dokončení věty. Zde posluchači nejvíce chybovali a rovněž jejich reakce byly nejpomalejší. To dále dosvědčuje, že sémantická neshoda

společně se zvukovou atrakcí způsobily největší problémy při percepci. Zkušenější skupina 2 odpovídala v porovnání se skupinou 1 statisticky významně přesněji, což se dalo očekávat, ovšem ne významně rychleji. Zdá se, že posluchači již na nižší úrovni zvládnutí cizího jazyka dodržují určitý časový limit pro zpracování signálu, po jehož uplynutí reagují bez ohledu na přesnost reakce.

Interpretace výsledků experimentu 2 je poněkud složitější. Míra shody obrázku s větou celkově měla signifikantní vliv na přesnost reakcí při srovnání částečné shody s úplnou shodou i neshodou. Zde se ukázalo, že posluchači často nedokázali identifikovat detail, ve kterém se věta od obrázku lišila. Původně zamýšlená interpretace tohoto jevu jako důkaz rychlého současného zpracování akustické a obrazové informace se nepotvrdil, jelikož méně zkušená skupina 1 podala významně horší výkon než skupina 2. Jde tedy zřejmě spíše o důsledek míry jazykových dovedností. Na druhou stranu posluchači skutečně reagovali signifikantně rychleji, když obrázek byl zcela odlišný od věty, což potvrzuje, že jsou schopni přizpůsobit se požadavkům experimentu a poměrně rychle vztáhnout obrazovou informaci k postupně přicházející akustické. Doba reakce se při srovnání úplné a částečné shody významně nelišila, jelikož posluchači museli zvážit každý detail ve větě a na obrázku. Ukázalo se tedy, že pokud o správnosti odpovědi rozhoduje detail, je zásadním faktorem zkušenost posluchačů, zatímco přizpůsobit se základnímu požadavku tohoto experimentu v zásadě dokáží obě skupiny.

V experimentu 3 neměla míra shody obrázku s větou významný vliv ani na míru přesnosti, ani na rychlost reakce pro žádnou z obou skupin. Nepotvrdilo se tedy, že posluchači dokáží využít obrazové informace ke zlepšení percepcie. Tyto

výsledky však mohou být způsobeny nedostatky v samotné metodologii tohoto experimentu, kdy už samotná věta je vždy sémanticky smysluplná a posluchači kromě výraznějšího zašumění nejsou příliš nuceni obrazovou informací zvažovat. Zde by byla na místě zásadní změna návrhu experimentu v podobě užití sémanticky chudých vět, případně užití minimálních párů a akustické manipulace distinktivního rysu daných fonémů za účelem odhalit, zda se hranice mezi nimi pod vlivem vizuální informace posune daným směrem.

Pokud tedy vezmeme v potaz pouze průměrné hodnoty, tak skupina 2 dosáhla nejvyšší přesnosti v případě, kdy se obrázek zcela shodoval s větou a vyobrazoval i objekt odpovídající poslednímu slovu dané věty. Horší průměrné výsledky byly dosaženy v případě nulové shody a nejhorší tehdy, když se obrázek lišil v detailu. Skupina 1 však poněkud nečekaně dosáhla v tomto případě relativně nejlepšího výsledku. Je tedy zřejmé, že shodná obrazová informace nepřinesla významné zlepšení porozumění. Spíše je patrné, že efekt vizuální informace je právě opačný, tedy že její zpracování přináší zvýšenou mentální zátěž. Důkazem toho je relativně nejdelší doba reakce pro obě skupiny v případě, kdy se věta lišila od obrázku pouze v posledním slově.

Výsledky tedy prokázaly, že nerodilí mluvčí na obou úrovních zvládnutí jazyka jsou schopni využít kontextové informace v podobě větné sémantiky pro zlepšení percepce slov maskovaných šumem. Jsou rovněž schopni pragmaticky se přizpůsobit požadavkům experimentu při integraci akustické a obrazové informace. Na druhou stranu neprokázalo se, že by obrazová informace měla významný pozitivní vliv na percepci, spíše naopak, jedná se o zvýšenou procesní zátěž.

Appendix A

Throughout the process of experiment preparation, several supplementary materials were created. They were not included in the thesis because of their lower importance in combination with sometimes extensive space demands. Nevertheless, they are stored in a digital format on the CD which is accompanying the thesis. We thought it might be useful and interesting to include the experiments and results as well together with software which was used to perform this study. The table below summarizes the content of individual folders on the CD.

Folder	Content description
Experiments	Complete experiments 1, 2, 3 with all the stimuli and control files for individual varieties. It can be run in Alvin 2.
Materials	Stimuli sets for individual experiments and their varieties together with data measured for noise masking in experiments 1 and 3.
Responses	Subject details and compressed raw subjects' responses obtained in all three experiments including recordings in experiments 1 and 3. The order in which the stimuli were presented can be seen from Alvin configuration files (cfg).
Results	Data obtained in response processing.
Software	Free programs which were used to prepare and administer the experiments and for the subsequent response processing.

References

- Adachi, T., Akahane-Yamada, R., and Ueda, K. (2006). "Intelligibility of English phonemes in noise for native and non-native listeners". *Acoustical Science and Technology* **27(5)**: 285–289.
- Benki, J. R. (2003). "Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition". *Journal of the Acoustical Society of America* **113(3)**: 1689–1705.
- Bradlow, A. R. and Alexander, J. A. (2007). "Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners". *Journal of the Acoustical Society of America* **121(4)**: 2339–2349.
- Chiang, C. S., and Dunkel, P. (1992). "The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning". *TESOL Quarterly* **26(2)**: 345–374.
- Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). "Novel second-language words and asymmetric lexical access". *Journal of Phonetics* **36**: 345–360.
- Felty, R. (2007). "Confusion patterns and response bias in spoken word recognition of German disyllabic words and nonwords". Poster presented at the *International Congress of the Phonetic Sciences*, Saarbrücken, Germany, August 6–10.
- Field, J. (1998). "Skills and strategies: towards a new methodology for listening". *ELT Journal* **52(2)**: 110–118.
- FitzPatrick, I., and Indefrey, P. (2007). "Effects of sentence context in L2 natural speech comprehension". *Nijmegen CNS Journal* **2(1)**: 43–56.
- Ganong, W. F. (1980). "Phonetic categorization in auditory word perception". *Journal of Experimental Psychology: Human Perception and Performance* **6**: 110–125.
- Garnes, S. and Bond, Z. S. (1976). "The relationship between acoustic information and semantic expectation". *Phonologica*, 285–293.
- Hahne, A. (2001). "What is different in second-language processing? Evidence from event-related brain potentials". *Journal of Psycholinguistic Research* **30(3)**: 251–266.
- Harrison, D. M. (2005). "Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style". *Speech Communication* **46**: 73–93.
- Hazan, V., Sennema, A., Faulkner, A., and Ortega-Llebaria, M. (2006). "The use of visual cues in the perception of non-native consonant contrasts". *Journal of the Acoustical Society of America* **119(3)**: 1740–1751.
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Kalikow, D. N., Stevens K. N., and Elliott L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability". *Journal of the Acoustical Society of America* **61(5)**: 1337–1351.

- Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003). "The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements". *Journal of Memory and Language* **49**(1): 133–156.
- Ladefoged, P. (2001). *A Course in Phonetics* (4th ed.). Heinle & Heinle Publ.
- Lecumberri, M. L. G., and Cooke, M. (2006). "Effects of masker type on native and non-native consonant perception in noise". *Journal of the Acoustical Society of America* **119**(4): 2445–2454.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model". *Ear Hear* **19**: 1–36.
- Mattys, S. L., Melhorn, J. F., and White, L. (2007). "Effects of syntactic expectations on speech segmentation". *Journal of Experimental Psychology: Human Perception and Performance* **33**(4): 960–977.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise". *Journal of the Acoustical Society of America* **95**(2): 1085–1099.
- Palková, Z. (1994). *Fonetika a fonologie češtiny s obecným úvodem do problematiky oboru*. Praha: Karolinum.
- Pelz, J., Canosa, R., Kucharczyk, D., Babcock, J., Silver, A., and Konno, D. (2000). "Portable eyetracking: a study of natural eye movements". *Proceedings of SPIE: Human Vision and Electronic Imaging* **3959**: 566–582.
- Pickett, J. M. (1999). *The Acoustics of Speech Communication. Fundamentals, Speech Perception Theory, and Technology*. London: Allyn and Bacon.
- Pylkkanen, L. and Marantz, A. (2003). "Tracking the time course of word recognition with MEG". *TRENDS in Cognitive Sciences* **7**(5): 187–189.
- Rugg, M. D., and Coles, M. G. H. (Eds.). (1995). *Electrophysiology of Mind: Event-related Brain Potentials and Cognition*. New York: Oxford University Press.
- Sadighi, F., and Zare, S. (2006). "Is listening comprehension influenced by the background knowledge of the learners? A case study of Iranian EFL learners". *The Linguistics Journal* **1**(3): 110–126.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). "Integration of visual and linguistic information in spoken language comprehension". *Science* **268**(5217): 1632–1634.
- Trimmis, N., Markatos, N., Malaperdas, K., and Papadeas, E. (2007). "Word recognition scores by native and non-native speakers of modern Greek language". Paper presented at 8th EFAS Congress / 10th Congress of the German Society of Audiology, Heidelberg, Germany, June 6–9.
- Van Berkum, J. J. A., Zwitserlood, P., Brown, C. M., and Hagoort, P. (2003). "When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect". *Cognitive Brain Research*, **17**: 701–718.
- Warren, R. M., and Warren, R. P. (1970). "Auditory illusions and confusions". *Scientific American* **223**: 30–36.
- Weber, A., and Cutler, A. (2004). "Lexical competition in non-native spoken-word recognition". *Journal of Memory and Language* **50**(1): 1–25.