Phoneme restoration in degraded speech communication

Slobodan T. Jovičić 1, Sandra Antešević 1 and Zoran Šarić 2

1 Faculty of Electrical Engineering, University of Belgrade, Serbia and Montenegro
2 Institut of Security, Belgrade, Serbia and Montenegro
jovicic@etf.bg.ac.yu

Abstract
This paper presents the experimental results of investigation in a research on the phoneme restoration effects for the case of degraded speech communication. The first experiment examined the role of coarticulation in consonant restoration inside CV syllables. In the second experiment the phonological effects in initial consonant restoration around the words were demonstrated. The third experiment demonstrated consonant restoration effects inside the sentences either grammatically and semantically correct or incorrect. Finally, the forth experiment showed an integral effects of coarticulation, phonology, syntax and semantics on phoneme restoration in spontaneous speech communication.

1. Introduction

Human auditory mechanism is able to restore the deleted or damaged phonemes in the case of degraded speech communication. This phoneme restoration effect demonstrates the robustness of speech perception. Two main information sources rely on this effect: (i) at the acoustic and phonetic level, this is the coarticulation, and (ii) at the linguistic and cognitive level, these are phonologic, syntactic and semantic effects.

The phoneme identification in speech perception is highly related to the coarticulation effects [1, 2]. The articulation transformation from one phoneme to another is performed in junction between them and is a consequence of the inertia of muscles of articulators’ organs. The hard boundaries between phonemes are not possible because many acoustic features of adjacent phonemes penetrate each other. Acoustic basis for perceptive phoneme restoration was found in these coarticulation effects.

The phenomenon of phonemic restoration related to the coarticulation effects was examined by removal of the release burst in plosives [3], by deleting of initial and final plosives in CV or VC syllables (C – consonant, V – vowel) [4], or by perception of deleted phonemes in wider linguistic context [5, 6], including the word recognition in fluent speech [7]. Another aspect in phonemic restoration is the role of higher cognitive levels. It was shown that these levels highly contribute to the phoneme identification [8, 9], and phoneme restoration in speech communication in very noisy environments [10].

The goal of this investigation is to shed more light on the coarticulation effects and the higher linguistic and cognitive levels in consonant restoration and enhancement of speech communication. Four experiments were performed: the first one, which examined the role of coarticulation in consonant restoration inside CV syllables; the second one, which demonstrated the phonological effects in initial consonant restoration around the words; the third one, which demonstrated the consonant restoration effects inside the sentences either grammatically and semantically correct or incorrect, and finally the forth experiment, which showed an integral effects of coarticulation, phonology, syntax and semantics on phoneme restoration in spontaneous speech communication.

2. Experiment I: Coarticulation effects

A method accepted for evaluation of the perceptual effects of the coarticulation in phoneme restoration is deletion of various portions of consonant, including the coarticulation segment of vowel, and examination how well the deleted consonant can still be identified. This experiment shows the identification functions for all 25 Serbian consonants expressed as the consonant intelligibility versus size of consonant deletion. The term “consonant restoration” was used in the perceptual meaning, while the term “consonant identification” was used as an experimental indicator represented by measure of intelligibility.

![Figure 1: An example for cutting points in syllable /ša/.](image)

10.21437/Interspeech.2004-213
cutting 75%, then 87.5%, 93.75%, and finally cutting the whole consonant; in the last two conditions one and two pitch periods of subsequent vowel were deleted in the case of initial fricatives and affricates, and one and four pitch periods for plosives, nasals and semi-vowels. As example, Fig. 1 shows CV syllable /ša/ with methodology of cutting. Experimental condition 5, as bold marked in Fig. 1, served as junction point between the consonant and the vowel. This point was subjectively defined by visual inspection of waveform and sonogram of syllable in question.

Figure 2: The identification functions of restored consonants with a) high, b) medium, and c) low restoration level.

2.1.2. Subjects and procedure
The 5 subjects volunteered in this experiment as the listeners. Experiment was performed in quiet laboratory environment. The stimuli were presented to the subjects through the loudspeaker from a distance lower than 2 m.

2.2. Results and discussion
Fig. 2 shows the intelligibility of restored consonants averaged over all vowels versus seven experimental conditions. The consonants are divided in three groups according to the level of the intelligibility: group I (consonants /p, m, n, nj, j, l, r/) with the highest intelligibility, group II (consonants /t, b, d, g, v, z, dž, lj/) with medium intelligibility, and group III (consonants /k, f, h, s, š, Ć, č, ĉ/) with very low intelligibility.

It is obviously that greater consonant sonority (feature of group I) generally implies higher intelligibility, or that consonants with predominantly transitional cues are more robust on deletion. On the other hand, the consonants with more invariant cues (feature of group III) are more independent from the following vowels in the process of articulation, and it is more difficult to restore them after deletion. The average intelligibility score for all three groups at experimental condition 5, where the whole consonant is deleted, is the following: 81% for group I, 37.5% for group II, and 3.9% for group III.

3. Experiment II: Phonological effects
The purpose of this experiment was to test the perception of the isolated words with deleted initial consonant in function of the kind of deleted consonant and the phonologic structure of the words. The intelligibility of words was analyzed for three types of words as stimuli: mono-syllabic, two-syllabic and three-syllabic words.

3.1. Stimuli and method
The stimuli were 25 mono-syllabic, 25 two-syllabic and 25 three-syllabic words. All words contain CV syllable in initial position which start with one of 25 consonants in Serbian language. A male speaker with healthy articulation spoke the words. After editing the words, the initial consonant in each word was deleted to the junction point with following vowel. The 11 subjects (7 males and 4 females), with healthy hearing, participated in this experiment as the listeners. All other experimental conditions were equal as in Experiment I.

3.2. Results and discussion
The results are shown in Fig. 3. All responses are divided according to the three consonant groups, referred to Experiment I. Two- and three-syllabic words were correctly restored in majority of responses and show nearly equal intelligibility of about 89%. On the other hand, the monosyllable words have lower restoration effect and are similar to CV syllables in Experiment I.

As it could be expected, there is greater possibility for longer stimuli (two- and three-syllabic words) to provide needed phonologic information for missing phoneme restoration. Obviously, undeleted part of stimuli was generally sufficient to start process of association with some meaning, in most cases with the meaning of original word.
4. Experiment III: Syntactic-semantic effects

It is well known that prediction of speech message highly contributes to its intelligibility, thanks to the syntax and semantics relations [1, 8]. The object of this experiment was to analyze influence of these higher linguistic levels on the perceptual restoration of deleted phonemes. Experiment was performed with sentences with different grammatical rules and meaning.

4.1. Stimuli and method

The stimuli material consisted of three categories of 21 sentences each.
Category I: contains correct sentences in syntactic and semantic sense.
Category II: contains sentences with correct syntax, but undefined semantics.
Category III: contains strings of unrelated words (incorrect both semantically and syntactically).

Stimuli of all three categories are four-word sentences. This length was selected as being enough to permit syntax and semantics to play role, but not so large to cause difficulty for listeners to retain strings of unrelated words. All words were two-syllabic. The key word with deleted initial consonant was one of the inner words in sentence environment. The key words over all sentences were selected to start with CV syllables (like in experiment II) with 21 different consonants in initial positions. The same key words were repeated through all three categories of sentences.

4.2. Results and discussion

Fig. 4 shows the results of this experiment. The results demonstrate expected differences between three categories of context: the restoration effect was most obvious in sentences of the category I (correct sentences), and the lowest restoration score (the intelligibility) was obtained for the category III (incorrect sentences). (A sentence was scored as correct if all constitutive words were correctly perceived.)

The intelligibility of key words in correct sentences (95%, category I) in comparison to two-syllabic isolated words (89%, Experiment II, Fig. 3) supports starting assumption that linguistic environment of the missing consonant contain additional information that can help its restoration. On the other hand, the intelligibility of the key words in sentences of category II (88%) is almost equivalent to the intelligibility of isolated words in experiment II. It seems that syntactically correct context, but without rationale meaning, has no influence on perception of belonging words. Finally, the intelligibility of the key words in incorrect sentences both syntactically and semantically (category III) show lower level of phonemic restoration than for the isolated words. This denotes negative effect of linguistically incorrect context where listeners were confused finding together completely unrelated words inside the sentence.

5. Experiment IV: Communication effects

The goal of this final experiment was the examination of an integral effect of coarticulation, phonologic, syntactic and semantic effects on phoneme restoration in speech communication.

5.1. Stimuli and method

The stimuli were 20 sentences composed by four two-syllabic words. Each consonant was deleted (edited and set to zero) between two coarticulation junction points, as is shown in one example on Fig. 5. Figure shows one Serbian sentence with transcription.
Two types of sentences were formed: (i) sentences with deleted consonants and high level of interruption effect, and (ii) sentences with deleted consonants masked by continuous white noise at speech-to-noise ratio (S/N) of 10dB. At this S/N the intelligibility of speech is not degraded by noise [1], but pauses between vowels (Fig. 5b) are filled with noise enough to mask interruption effect.

5.2. Results and discussion

Intelligibility of interrupted words and sentences without and with masking noise is shown in Fig. 6.

![Figure 6: Intelligibility of interrupted words and sentences without and with masking noise.](image)

The intelligibility of words and sentences without masking noise is significantly lower in comparison to correct sentences in Experiment III (Fig. 4, I. category). This result is the consequence of all consonants deletion. Words meaning names and words with lower frequency in lexicon show lower intelligibility. These words show lower prediction with sentence’s context.

When the sentences were masked by noise the intelligibility got significantly higher. It is obvious that noise connects the separate parts of staccato sentences, in this case the vowels, performing continuity in speech signal. In both conditions (Fig. 6) the residual consonant intelligibility (Experiment I, Fig. 2, condition 5) with context information is enough to increase the speech intelligibility.

6. Conclusions

The speech is highly redundant signal. This fact helps the human perceptual mechanism to retain fluent speech communication intelligible even if the speech is highly degraded. The experimental results reported in this paper quantitatively demonstrate how the coarticulation, phonologic, syntax and semantics effects contribute to the speech intelligibility. The results are important from practical communication point of view, as well as for theoretical investigation in speech perception.

7. Acknowledgements

This work was partially supported by Ministry of Science, Technology and Development under Grant number OI-1784.

8. References