Phoneme coarticulation in Russian hard and soft VCV-utterances with voiceless fricatives

Derkach, M. and Fant, G. and de Serpa-Leitao, A.
I. SPEECH ANALYSIS AND SPEECH PRODUCTION

A. PHONEME COARTICULATION IN RUSSIAN HARD AND SOFT VCV-UTTERANCES WITH VOICELESS FRICATIVES

M. Derkach*, G. Fant, and A. de Serpa-Leitão

General

Hard and soft distinctions known in the distinctive feature terminology as sharp/plain are important phoneme characteristics in the Russian language. It is generally accepted to allocate the hard-soft distinction to the consonantal system of Russian phonemes dividing them into hard (unpalatalized) and soft (palatalized), respectively, see Shupljakov et al (1966). It was, however, not clear enough how far the soft-hard distinctions are spread in the whole syllabic structure. This article is devoted to the spectrographic study of these distinctions in Russian VCV-utterances in the time domain.

Experimental Material

The experimental material included 300 VCV-disyllables recorded in an anechoic chamber with a male speaker of Modern Standard Russian with some influence of West-Ukrainian accent. These represented hard and soft disyllables with the Russian vowels [u], [o], [a], [ɛ], [y], [i] and the voiceless fricatives [s], [ʃ], [f], [h] combined into symmetrical and asymmetrical utterances.

Spectrograms were made by means of a Voice Print spectrograph, type 4691, using the broad band-pass filter of 300 Hz and pre-emphasis of 6 dB/octave between 300 Hz and 3000 Hz.

A direct comparison between spectrograms of hard and soft VCV-utterances showed that they were based on two different types of syllabic organization in the time domain. In Fig. 1-A-1 spectrograms of hard disyllables are presented. All of them initiate on a vowel [a] and terminate on [o], including different voiceless fricatives.

Formant transitions reflecting time changes of vocal tract configuration common for all hard utterances with the same adjacent vowels are apparent. They pass through the entire consonant reflecting the coarticulation influence of the second vowel on the preceding part of the sequence.

* Guest researcher from Lvov University, USSR, from February 17, 1970.
Fig. 1-A-1. Spectrograms of hard [aCo]-words with the voiceless fricatives [s], [ʃ], [f], and [h]. (The symbol C indicates a fricative consonant in general.)
Spectrograms of soft disyllables are presented in Fig. I-A-2. They consist of the same vowels and consonants as those in the previous figure except that they have been pronounced with the soft articulation. It is apparent that the second vowel in all the soft utterances is initiated by a compulsory [i] coarticulation so that the formant transitions pass through the consonantal part towards an [i] position at the CV boundary independently of the proper phoneme value of the second vowel. The further transitions to the second vowel, accordingly, turn it into an [i] initiating diphthong. In Fig. I-A-3 the spectrograms of a hard disyllable [isu] and a soft one [is'u] are presented in direct contrast in order to show the typical formant transitions in the soft utterance at the beginning of the second vowel.

Statistical study of phoneme coarticulation in VCV-utterances

Phoneme coarticulation was studied in four points of time in the sequences shown in Fig. I-A-4: at the very beginning of the first vowel (A), at its very end at the VC boundary (B), at the beginning of the second vowel (C), and 150 msec later when the V2, as a rule, reaches its stationary position (D).

The spectrographic material referred to in this chapter was represented by 120 spectrograms of VCV-disyllables, 60 hard and 60 soft. Combinations of three initial vowels [u], [a], [i], four voiceless fricatives [s], [ʃ], [f], [h], and five terminal vowels [u], [o], [a], [ɛ], [y] (in soft syllables [i]) provided variability of the syllables required for a statistical study. The utterances were pronounced in pairs in direct hard-soft contrast to each other.

The main questions formulated in this study were the following: 1) How do inter-phoneme distinctions in the F-pattern develop in time in the VCV-words and when do they reach maximal values for each type of phoneme? 2) How do the distinctions between soft and hard utterances develop in the time domain, what phonemes and what points in time are responsible for these and when do they reach their maximal values?

The following statistical measures were applied to answer these questions: (a) mean values $M$ of the formant frequencies $F_1$, $F_2$, and $F_3$ (b) their standard deviations $m$ calculated by the formula: 
Fig. I-A-2. Spectrograms of soft [aC'o]-words with the voiceless fricatives [s], [ʃ], [f], and [h].
Fig. I-A-3. Hard and soft words [isu]-[is'u] in a direct contrast.
Fig. 1-A-4. The four points used for measurements of formant frequencies in VCV-utterances.
where $X_i$ is the correspondent formant frequency,

$$m = \frac{n}{n(n-1)} \sum_{i=1}^{n} (X_i - M)^2$$

where $X_i$ is the correspondent formant frequency,

\[i = 1 \div n,\]

$n$ is the number of spectrograms in each group.

They have been calculated for the points A, B, C, and D in different groups of disyllables, and an interval $M \pm 3m$ defining the limits of the distribution of the mean formant frequencies with the reliability 0.99 * was established in each case. These intervals are shown in the graphs which follow below.

In the first series of experiments 120 VCV-words, 60 hard and 60 soft, were classified into three groups according to the initial vowels [u], [a], and [i]. The formant frequencies $F_1$, $F_2$ and $F_3$ were measured and presented in Figs. I-A-5, I-A-6, and I-A-7, respectively, for the points A, B, C, and D from top to bottom of these figures. Mean values and limits of their statistical variation $M \pm 3m$ are shown in each of the graphs. The figures with positive polarity represent hard disyllables and those with negative polarity are the ones for soft utterances.

The coarticulation phenomena are most apparent in the movements of the second formant which is a primary correlate of the vocal tract configuration. As it follows from the graphs I-A-5, I-A-6, I-A-7 the first vowels are phonetically best defined in their beginning (point A). The coarticulation effect with following phonemes is already noticeable in the end of the first vowels (point B) as a decrease of differences between their mean formant frequencies. Information about the first vowel decreases with distance in a word so that no statistically significant information about its phonological type is noticed in the second vowel of an utterance (points C and D). The second fact which is to notice in these graphs is the statistically significant division of hard and soft disyllables into two groups in any formant position in point C of the word.

The second series of experiments consisted in measuring formant frequencies for 60 hard and 60 soft VCV-words divided into four groups according to the fricatives [s], [ʃ], [f], [h]. The same variability of

* A normal-type of statistical distribution has been anticipated.
Fig. I-A-5. Statistical distribution of the F1 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables initiating on the vowels [u], [a], [i].
Fig. I-A-6. Statistical distribution of the F2 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables initiating on the vowels [u], [a], [i].
Fig. I-A-7. Statistical distribution of the F3 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables initiating on the vowels [u], [a], [i].
adjacent vowels takes place in each of the groups. The correspondent statistical data is shown in Figs. I-A-8, I-A-9, and I-A-10.

It can be seen in the graphs that information about the consonants in the utterances exists already at the end of the first vowel (point B), thus reflecting the consonantal effect of coarticulation in the first vowel. The phonetical type of consonants is best defined by $F_1$, $F_2$, $F_3$ in the point C as well for the hard as for the soft disyllables. On the other hand, there is not enough information in the adjacent vowels to provide an accurate defining of all four types of voiceless fricatives. This stresses the importance of distinctive cues laying in the consonants themselves for their phonetical and phonological classification.

What concerns the hard-soft distinctions they reach their maximal values in point C of the word. Information about the phonetical type of the consonant as well as that of hard-soft distinctions is vanishing below a threshold of statistical significance in the end of the second vowel (point D).

In the third series of experiments 60 hard and 60 soft VCV-words were classified into five groups each according to the terminal vowels [u], [o], [æ], [ɛ], and [y]. The data about statistical distinctions between their formant positions in the points A, B, C, and D are presented in Figs. I-A-11 ($F_1$), I-A-12 ($F_2$), and I-A-13 ($F_3$).

It seems to be worth-while to notice two facts reflected in these figures. The first one is about the soft-hard distinctions still maintaining the highest values in point C in the utterance and the gradual release in holding these differences in the second vowel itself. With the decreasing of the soft-hard distinctions along the second vowel the accuracy for defining its phonetical type becomes higher.

The second fact reflects a coarticulation effect of the second vowel starting already at the end of the first vowel. It is noticeable in the increase of differences between the disyllables in the $F_2$-location in point B.

The general conclusions referring to hard-soft relations in VCV-words are as follows. The mentioned distinctions start at the end of the first vowel (point B) and are also specifically influenced by the consonantal-type of coarticulation and with the second vowel confirming the model of coarticulation proposed by Ohman (1956, 1967). As a rule the very beginning of
Fig. I-A-8. Statistical distribution of the F1 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables with intermediate fricatives [s], [ʃ], [f], [h].
Fig. I-A-9. Statistical distribution of the F2 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables with intermediate fricatives [s], [ʃ], [ʃ], [h].
Fig. 1-A-10. Statistical distribution of the F3 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables with intermediate fricatives [s], [ʃ], [f], [h].
Fig. I-A-11. Statistical distribution of the F1 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables terminating on the vowels \{u\}, \{o\}, \{a\}, \{e\}, \{y\}. 
Fig. I-A-12. Statistical distribution of the F2 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables terminating on the vowels [u], [o], [a], [ɛ], [y-i].
Fig. 1-A-13. Statistical distribution of the F3 frequency in hard (positive polarity) and soft (negative polarity) VCV-disyllables terminating on the vowels [u], [o], [a], [e], [y-i].
the second vowel (point C) deviates from its target and the process of its phonetical establishment continues in the vowel itself. At the same time point C carries the essential information on the hard-soft distinctions within a word. All the soft words hold the [i]-like configuration of the vocal tract until the beginning of the second vowel (point C) independently of their phonemic composition. The [i]-like beginning of the second vowel turns it into a typical diphthong constituting a stable and highly statistically reliable spectrographic cue of softness in the word. Differences between soft and hard utterances disappear by the end of the second vowel (point D) where the vowels return back to their proper formant positions according to their phonological categories.

Perceptual cues of soft-hard distinctions in VCV-utterances

Perceptual experiments have been accomplished in order to study the importance of formant transitions in a second vowel for a direct perception of hardness and softness of a disyllable. Experimental utterances were distributed in random order within different dummy VCV-words and recorded on magnetic tape appearing 14 times in a test presented to a single listener. Two Russian listeners participated in the experiments. Headphones of type Sennheiser MD 414 were used.

Six soft syllables, namely symmetrical [as’a], [af’a], [ah’a] and asymmetrical [as’u], [af’u], [ah’u] were used as control representative utterances. Spectrograms of these disyllables are shown in Figs. I-A-14 and I-A-15. All of them were perceived correctly as soft words. In the first experiment a part of the second vowel (starting in point C and 100 msec long) was cut off, thus removing the typical [i]-like beginning. The remaining parts of the word were joined and spliced together. Spectrograms of these sequences are presented in Figs. I-A-16 and I-A-17. The percentages of correct hard-soft identification are indicated in the figures.

Data obtained in this series of experiments show that in general 80% of the soft disyllables with the removed beginning of the second vowel were perceived as hard. Probably the second formant carried the main information about their softness. Therefore, in the case of the consonants having very weak F2, as for instance in [s], the information about softness is mainly related to the F2-transition in the second vowel. Removing these transitions from the disyllables will remove the cues about softness and
Fig. I-A-14. Spectrograms of intact symmetrical soft utterances [aC' a]. The soft-hard identification score is indicated.
Fig. I-A-15. Spectrograms of intact asymmetrical soft utterances [aG' u].
The soft-hard identification score is indicated.
Fig. I-A-16. Spectrograms of symmetrical utterances [aG' a] with removed beginning of the second vowel and hard-soft identification scores.
Fig. 1.-A-F. Spectrograms of asymmetrical soft utterances [ə[ə]u] with removed beginning of the second vowel and hard-soft identification scores.
the experimental sequences will be perceived as hard. In other cases, when the F2-traces are intensive enough in the consonant itself, they provide some amount of perceptual information about the softness in addition to that being contained in the formant transitions of the second vowel. As it is seen in the words [af' u], [ah' u], and [ah' a], cutting the [i]-like transitions of the second vowel does not remove all information about the softness of the utterance.

In the second series of perceptual experiments the relative role of F2- and F1-transitions in the perception of softness was studied. The experimental sequences have been filtered above 1.5 kHz by means of three low-pass filters connected in succession attenuating the higher regions of spectrum on about 50 dB. Spectrograms and perceptual data about the filtered utterances are presented in Figs. I-A-18 and I-A-19.

The perceptual role of the F1-transition is evident from the data shown in Fig. I-A-18, where the soft words with removed F2 have been perceived correctly as being soft.

It is interesting to notice the data shown in Fig. I-A-19. There is some ambiguous information about a type of F2-transitions in the beginning of the second vowel in the filtered sequences [as' u], [af' u], and [ah' u] and the identification of softness decreased, respectively, in these cases too.

The third series of perceptual experiments presented in Figs. I-A-20 and I-A-21 consisted in substituting the second vowel in the hard utterances by the second vowel (with its [i]-like beginning) from the soft word [as' a]. All the disyllables rebuilt in this way have been perceived as soft. The "hard-like" F2-transitions passing through the consonantal part of the word could not overwhelm the "soft-like" information carried by the [i]-like diphthong of the second vowel, even though some peculiarity in sounding can be noticed by a trained listener.

Conclusions

The data obtained in this study show that the hard-soft distinctions in a VCV-utterance are spread along the whole word according to the unpalatalized or palatalized articulation. The first three formants are responsible for the acoustical representation of this control, F2 carrying the most apparent part of the distinction. F1 appears to be next in importance. F3 also
Fig. I-A-18. Spectrograms and perceptual data about softness-hardness for the symmetrical soft disyllables [aC’a] filtered above 1.5 kHz.
Fig. I-A-19. Spectrograms and perceptual data about softness-hardness for the asymmetrical soft disyllables [aC'u] filtered above 1.5 kHz.
Fig. I-A-20. Spectrograms and perceptual data about softness-hardness for the sequences combined from the hard VC-part taken from the symmetrical disyllables [a_2 a], and soft V2-part taken from the disyllable [as' a].
Fig. I-A-21. Spectrograms and perceptual data about softness-hardness for the sequences combined from the hard VC-part taken from the asymmetrical disyllables [asu], and soft V2-part taken from the disyllable [as'a].
contributes to the distinction. As a rule the second vowel of a VCV-utterance is of main importance for the hard-soft distinction. Holding the [i]-like shape of the vocal tract until the beginning of the second vowel provides the main cue for softness in Russian syllables. A physiological reason for this articulatory delay is likely to be the necessity to increase the reliability of the auditory cues for the [i]-like position of the vocal tract which are acoustically and perceptually enhanced in vowels.

Acknowledgments

The authors are indebted to Doc. B. Lindblom for the critical discussions, to Mrs. S. Felicetti for her editorial assistance, and to Mrs. G. Karlsson for the drawings.

References:

