Acoustical features of hard and soft Russian consonants in connected speech: A spectrographic study

Shupljakov, V. and Fant, G. and de Serpa-Leitao, A.

journal: STL-QPSR
volume: 9
number: 4
year: 1968
pages: 001-006

http://www.speech.kth.se/qpsr
I. SPEECH ANALYSIS

A. ACOUSTICAL FEATURES OF HARD AND SOFT RUSSIAN CONSONANTS IN CONNECTED SPEECH:
A SPECTROGRAPHIC STUDY

V. Shuplyakov, G. Fant, and A. de Serpa-Leitão

The distinction between soft and hard consonants is a fundamental characteristic of the Russian consonant system. The physiological basis for this distinction is considered to be the presence versus absence of a palatalization of the consonant. In distinctive feature terminology the distinction between soft (palatalized) and hard (unpalatalized) consonants is referred to as sharp/plain*, the typical acoustic cue being a rise of $F_2$ and also of $F_3$ as the tongue approaches the [i]-position, Fant (1960), p. 220. However, according to the same source, the plain member of the opposition is not merely "neutral" from a phonetic point of view since there is a clear tendency of the tongue to approach an "[u]-like" position in the hard member of the opposition which becomes especially apparent in labials, where the tongue-body has a greater freedom of movement. This effect showed up in the study of Fant (1960) and was apparently not due to coarticulation with the following vowel [əː]. The subject in question had a rather long vocal tract which correlates with the relatively low $F_2$ of his [j] consonant, ($F_2 = 1850$ Hz and $F_3 = 3000$ Hz). The following measured data for the subject illustrate the great $F_2$-span in labials.

<table>
<thead>
<tr>
<th></th>
<th>$F_2$</th>
<th>$F_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[v]</td>
<td>850</td>
<td>2100</td>
</tr>
<tr>
<td>[v̂]</td>
<td>1650</td>
<td>2150</td>
</tr>
<tr>
<td>[f]</td>
<td>1100</td>
<td>2200</td>
</tr>
<tr>
<td>[f̃]</td>
<td>1650</td>
<td>2350</td>
</tr>
<tr>
<td>[b]</td>
<td>900</td>
<td>2150</td>
</tr>
<tr>
<td>[b̃]</td>
<td>1700</td>
<td>2100</td>
</tr>
<tr>
<td>[p]</td>
<td>950</td>
<td>2050</td>
</tr>
<tr>
<td>[p̃]</td>
<td>2050</td>
<td></td>
</tr>
</tbody>
</table>

* Cf. the improvement suggested recently by Chomsky and Halle (1968) who regard palatalization as the superposition of an [i] vowel on the feature complex of the palatalized consonant and replace sharp/plain by the features needed for the vowels: high, low and back.
The difference in $F_3$ was small and not consistent. These effects were seen also in the laterals

\[
\begin{array}{cc}
F_2 & F_3 \\
[1] & 800 & 2800 \\
[1_1] & 1600 & 2300 \\
\end{array}
\]

Here the $F_3$ of the hard member is higher than that of the soft member which can be explained from the typically high $F_3$ of a mid-vocal tract (uvular) point of maximal tongue constriction, see the three-parameter model of the vocal tract, Fant (1960), p. 82.

Finally we can quote the $F_2$, $F_3$ of

\[
\begin{array}{cc}
F_2 & F_3 \\
[z] & 1200 & 2200 \\
[z_1] & 1600 & 2300 \\
\end{array}
\]

In a study performed by Shupiljakov (1966) it was found that subjects perceived isolated stationary consonants [s] and [ʃ] as soft when $F_2$ was higher than 1645 Hz but as hard consonants otherwise. Öhman (1964) stresses the tendency of Russian hard and soft consonants to resist coarticulation with adjacent vowels.

It is possible to formulate three partially equivalent hypotheses about the perceptual invariance of the soft-hard distinction.

(1) The extreme $F_2$-location in the consonant is always higher in a soft than in a corresponding hard consonant pronounced in the same context by the same speaker.

(2) All soft consonants have a higher $F_2$ than all hard consonants independent of vowel context.
   a) Validity for all speakers irrespective of "size-factor".
   b) Validity for a single speaker only.

(3) The $F_2$ of the soft consonant is always higher than that of following and preceding vowels, i.e. the transition is positive in VC and negative in CV sequences.

Hypotheses (1) and (2) above stress the ensemble aspect with respect to a selection of alternative $F_2$-loci and hypothesis (3) the identification with respect to an invariance in the time function of the pattern, i.e. a transition of specified direction.

The purpose of the following study is to contribute to the evaluation of these various pattern aspects and to study $F_3$ as well as $F_2$. 
Experimental procedure

Three female and two male speakers of Modern Standard Russian took part in the experiments. The subjects pronounced lists of symmetrical sequences to VCV (vowel-consonant-vowel) type. These lists contained all Russian vowels and consonants, in all 180 syllables.

This speech material was recorded on an Ampex magnetic tape-recorder. Spectrograms were made with a Kay Electric Sona-Graph, type 4691, operated with a broad filter of 300 Hz and a high frequency emphasis of 6 dB per octave above 1000 Hz.

Measurements were made of $F_2$ and $F_3$ within the consonant sampling at places of extreme values. If these were obscured, measurements were taken at the very beginning of a transition from the consonant to the vowel. The $F_2$ and $F_3$ data were quantized to nearest 100 Hz values for the purpose of constructing diagrams.

The direction of the transitions were observed.

Results and conclusions

It was soon verified that hypothesis (3) did not hold. $F_2$-transitions from adjacent vowels to the soft consonant were generally positive, see Fig. I-A-1, but could also be neutral or slightly negative, see the word [əl'æ] of Fig. I-A-2, where $F_2 = 2100$ Hz in [ə] as well as in the [l]. The sequence [əlæ] of Fig. I-A-2 exemplifies a typical negative transition from the $F_2 = 1900$ Hz of the [ə] to the $F_2 \text{ min} = 830$ Hz of the dark [l].

However, in [odo] and [aza], Fig. I-A-2, the transition from the vowel to the hard consonant is neutral or positive.

Although the direction of transition can be an important perceptual parameter, Stevens (1967), Chistovich (1968), it apparently does not provide a simple invariance rule for the soft/hard distinction.

The ceteris paribus hypothesis (1) was consistently found to hold. The $F_2$ of a soft consonant is always higher than the $F_2$ of a hard consonant in the same context and assuming the same speaker. The next step was to see whether there existed an absolute criterion in $F_2$ independent of vowel context as formulated in hypothesis (2) above. Fig. I-A-3 and Fig. I-A-4 show the distributions for each of the five speakers. Apparently the $F_2$
Fig. I-A-1. Spectrogram of the syllable uh' u.
Fig. 1-A-2. Spectrograms of different syllables.
Fig. 1-A-3. The frequency of the second formant in hard (solid lines) and soft (dashed lines) consonants, pronounced by two male subjects.
The frequency of the second formant in hard (solid lines) and soft (dashed lines) consonants, pronounced by three different female subjects.
distributions of hard and soft consonants show very little overlap. Their crossover points occur at 1700 Hz for the two male speakers and 1900 Hz, 2000 Hz, and 2000 Hz in the distributions of the female speakers. The number of mistakes that would occur when using a single threshold value in $F_2^t$, individually determined for each speaker would be 1.9% in the 360 utterances of the male subjects and 4% in the 540 utterances of the female group. A similar tendency was found in the results from the experiments on the perception of hard and soft stationary fricative consonants, Shupljakov (1968).

An analysis of the overlap in the $F_2$ distributions shows that 18 out of the 29 cases originate from the sequences [mat] [ʌhə] [uɡ'u] [uʃ'u][oʃ'o] which are rare or unnatural in Russian. The front vowel contexts ([i] [ʌ][æ] and [e] account for 17 of the overlap points.

The finite amount of overlap can be further reduced by taking into account the $F_3$ distributions, see Figs. 1-A-5 - 1-A-9. This cue, however, is effective mainly in front vowel contexts, where $F_3$ of the soft consonant is generally higher than in hard consonants. In [a] contexts the overlap is considerable. The same is found in [o] and [u] contexts but it is interesting to note that for one of the male subjects (M-n) there is a clear tendency of a lower $F_3$ in soft than in hard consonants. This effect could be attributed to a coarticulation of the hard consonant with "mid-vocal tract" place of articulation conditioning a high $F_3$.

The crossover points of $F_3$ in front vowel contexts were found to be 2600 Hz for the male speakers and 3100 Hz, 3100 Hz, and 3000 Hz for the female speakers.

The question now arises whether the critical threshold values of second and third formants separating hard from soft consonants can be inferred from some general property of the speaker. One such parameter is the total length of the speaker's vocal tract which is inversely proportional to the average $F_3$, Fant (1959). The average $F_3$ of each speaker's front vowels [i] [ʌ][æ] and [e] was accordingly calculated and was found to coincide with the hard/soft $F_3$ boundary. The hard/soft $F_2$ boundary was found to be 0.53 of the male speaker's average $F_3$ and within the female group the $F_2$ boundary was 0.59 of the speaker's average $F_3$. Thus

$$F_2^i = 0.53 \overline{F_3} \quad \text{(males)}$$
$$F_2^t = 0.59 \overline{F_3} \quad \text{(females)}$$
$$F_3^t = \overline{F_3} \quad \text{(males and females)}$$
Fig. I-A-5. The frequency of the third formant in hard (solid lines) and soft (dashed lines) consonants, followed by different vowels (Male speaker: M-n).
Fig. 1-A-6. The frequency of the third formant in hard (solid lines) and soft (dashed lines) consonants, followed by different vowels (Male speaker: B-r).
Fig. I-A-7. The frequency of the third formant in hard (solid lines) and soft (dashed lines) consonants, followed by different vowels (Female speaker: H-a).
Fig. 1-A-8. The frequency of the third formant in hard (solid lines) and soft (dashed lines) consonants, followed by different vowels (Female speaker: M-a).
Fig. I-A-9. The frequency of the third formant in hard (solid lines) and soft (dashed lines) consonants, followed by different vowels (Female speaker: P-a).
The normalizing function of the third formant determining $F_2$ boundaries in vowel identification has been studied by Fujisaki and Kawashima (1967 and 1968). From their data and those of Fant (1959) it is known that the perceptual importance of $F_3$ is much greater in front vowels than in back vowels. This general rule is related to our finding that $F_3$ adds to the phonetic separation of soft and hard consonants in front vowel context only. In this connection it is also worth noting that in the experiments on the perception of sustained fricatives, Shupljakov (1968), the $F_2t$ was constant as long as the part of the noise spectrum above 2500 Hz was kept the same.

Summary. Perceptual model

Our final conclusion is accordingly that the speech wave characteristics separating Russian hard and soft consonants can be expressed by a single threshold in the consonant $F_2$ with a support of a similar normalized threshold in $F_3$ if $F_2$ is high. These thresholds are simply related to the speaker's average formant pattern. Consonants with formant frequencies above the thresholds are soft and those with formant frequencies below the thresholds are hard. The speech production correlate is palatalization as opposed to a neutral or "mid tube" place of tongue-body articulation. Therefore a listener's hard/soft identification could be based on a formant frequency boundary categorization as the primary cue as was suggested for vowels by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966).

More detailed data on the F-pattern of various hard and soft consonants in different contexts will be given in a forthcoming publication.

Acknowledgments

We are thankful to Mrs. S. Felicetti and Dr. B. Lindblom for the help during the preparation of the paper and to Miss E. Agelfors for the drawings.

References


references, cont.


