ACOUSTICS AND PERCEPTION OF SPEECH IN VARIOUS MODES OF ARTICULATION

B.M.Kolesnikov L.M.Zakharov

Dept. of Philology Moscow State University
Moscow, USSR, 119899

ABSTRACT

This paper presents preliminary data and observations from research on acoustic modifications of speech in various modes of articulation. We consider acoustic variables of speech within a general model of speech production in which the mode of articulation (MA) is an independent source of acoustic change. Intelligibility scores of the word-lists heard under conditions of noise differ due to the degree of physical manifestation of phonetic features which is the highest in forced speech and the lowest in sloppy speech.

INTRODUCTION

Quite a number of papers have recently been published on acoustic properties of speech and its perception. The bulk of this research deals with speech in a comparatively narrow range of acoustic change. However, speech often may differ acoustically due to external conditions of communication or to the internal state of the speaker. One example is forced speech (FS), speech which is produced in the mode of forced articulation. As a rule FS is louder than normal speech (NS) and therefore less subject to distortions and more intelligible due to its more effective use of the hearer's attention. FS occurs in many situations of everyday life and is a universal means of overcoming distance, ambiant noise, an interlocutor's dullness or a child's disobedience (in the latter two cases FS is not logically motivated and has negative emotional connotations).

Another modification of speech has not yet been sufficiently analysed. It is the speech produced in a state of extreme weariness or intoxication. This variety of speech with slurred articulation is commonly called sloppy speech (SS). It's acoustics are markedly different from those of normal speech (NS) and forced speech. We distinguish three modes of articulation (MA) and three corresponding modifications of speech: normal, forced and sloppy respectively.

METHOD

In our experiments acoustic peculiarities and perception of FS and SS as compared with NS have been the main object of interest. Word-lists of 14 words have been read by 6 speakers possessing certain dramatic skills. Speakers were asked to imagine a situation where it was necessary to "out-voice" ambient noise or imitate the speech of an operator exhausted by 48 sleepless hours.

Recordings thus obtained were presented to a group of subjects who were asked to describe the speaker's condition and any peculiarities of his speech. Here follow some examples of such descriptions: "neutral speech" (about normal articulation), "speech most likely produced at a meeting (about forced articulation), "indifference, or rather weariness near to drowsiness" (about slurred articulation). The recordings were used for further acoustic analysis.

The sonagrams of the recorded stimuli were made by means of the "Kay Elemetrics" Sona-graph. The measurements of fundamental frequency, duration and spectral characteristics of the stimuli were made on the sonagrams. The main data are presented in the following tables.

In table I FS and SS are compared with NS. A plus-sign stands for increased measured duration of a segment in various MAs as compared with NS, a minus-sign stands for decreased measured duration and ∅ stands for equality of measured duration.

The table shows that while a word in general becomes longer in FS, stressed vowels and to a lesser extent preceding consonants are consistently lengthened. The remaining segments (unstressed vowels and other consonants are not necessarily lengthened. Other consonants are lengthened 55% of the time for speaker I and 82% of the time for speaker II. Unstressed vowels are lengthened 64% of the time for speaker I and 86% of the time for speaker II.

In SS the character of duration change is different for both speakers. Thus speaker II lengthens half of all the words, and in only 29% of these cases lengthens vowels stressed or unstressed. Speaker I, who lengthens 40% of the words, constantly makes this by lengthening stressed vowels. He lengthens unstressed vowels only 62% of the time. He does not lengthen preceding consonants. However, consonants other than preceding ones are lengthened more consistently - 100% of the time for speaker I and 91% of the time for speaker II.

The tendency to lengthen consonants was confirmed in experiments on a larger scale employing

5 and 6 speakers and a greater number of words. The consonants in FS do not always increase in duration in unstressed syllables. Lengthening occurs as follows: sonants - 42%, fricatives - 13%, plosives and affricates - 59% as compared to their duration in NS. In FS for consonants preceding stressed vowels lengthening occurs thus: sonants - 61%, fricatives - 43%, plosives and affricates - 52%.

Table I. Duration change in FS and SS as compared with NS

Stimulus word / speaker		forced speech: stressed vowel	unstressed vowel	preceding consonant	other consonants	sloppy speech: general	stressed vowel	unstressed vowel	preceding consonant	other consonants
ia-ma	II	+	+	+	–	+	–	ø	ø	+
	I	+	+	+	+	+	+	–	–	+
azu-	II	+	+	+		+	+	+	–	
	I	+	+	+		+	+	–	–	
a-lpha	II	+	+	+	–	+	–	+	+	+
	I	+	+	+	+	+	+	+	+	+
uzda-	II	+	+	+		+	+		–	
	I	+	+	–		+	+		–	
i-m'a	II	+	+	+	+	–	–	–		+
	I	+	+	+	ø	+	+	–		+
uzhe-	II	+	+	+		–	+	+	–	
	I	+	+	+		+	+	+	+	
e-ta	II	+	+	+	+	–	–	–		+
	I	+	+	+	ø	+	+	+		+
ke-pka	II	+	+	+	(ø)	+	–	–	– (+)	+
	I	+	+	–	(+)	+	+	+	+ (ø)	+
i-kry	II	+	+	+	+	+	–	+		+
	I	+	+	+	+	+	+	+		+
t'o-sh'a	II	+	+	+	(+)	+	+	–	– (+)	+
	I	ø	+	–	(+)	+	+	+	+ (+)	+
kho-lodno	II	+	+	–	+	–	–	–		–
	I	+	+	–	+	+	+	–		+
za-ponki	II	+	+	+	+	+	–	+	–	– +
	I	+	+	+	+	+	+	–	+	+
pała-tka	II	+	+	+	+	+	–	ø	–	ø –
	I	+	+	–	+	–	+	+	+	ø +
analgi-n	II	+	+	+	ø	+	+	+	–	+ +
	I	+	+	+	ø	–	+	+	+	ø +

In SS consonants other than preceding ones tend to lengthen thus: sonants - 69%, fricatives - 67%, plosives and affricates - 76%. The preceding sonants in SS tend to maintain their duration or shorten in the following manner: sonants - 57%, fricatives - 69%, plosives and affricates - 48%.

Table 2 shows that FS is characterized by lengthening of vowels. However, such lengthening depends on a number of circumstances. Thus for example only 4 out of 6 speakers lengthen stressed vowels. One speaker might shorten all the vowels, whereas another might lengthen the stressed /u/ and /i/ and shorten all the rest. One circumstance might be the tempo chosen by a speaker - vowel lengthening is characteristic of slower tempo. Another might be the degree to which articulation is forced. Curiously enough speakers I and IV are the same person. The second set of his recordings were made after an interval of 6 months. The data are totally different: the first recordings display shortening of almost all vowels, but the second recordings display lengthening of the stressed vowels. It is worth mentioning that the tempo in the first case was slower than in the second. Five out of six speakers lengthen the stressed /i/ and /u/ to a greater degree than /a/, /o/, /e/ in FS.

As far as the unstressed vowels are concerned, there are not consistent differences between various MAs. Only one speaker out of six constantly lengthened unstressed vowels and only one of six constantly shortened the unstressed vowels (the same speaker also shortened the stressed vowels).

The duration of unstressed vowels was not found to depend on their identity, position relative to stress or consonant environment. The duration change of unstressed vowels did not reveal any evident regularity.

The SS is characterized by an irregular and individualized manner of vowel duration change. 2 speakers (I and III) tend to lengthen almost all vowels, three speakers (II, V, VI) tend to shorten almost all vowels. One speaker lengthens almost all stressed vowels and shortens almost all unstressed vowels. Moreover, the character of vowel duration change in SS does not depend on the speaker's chosen tempo. The duration of all vowels was not found to depend on their identity, position relative to stress or consonant environment.

One can point out two contrasting tendencies in intensity characteristic of FS as compared with NS.

I. Greater prominence of the stressed vowel (the difference between maximum intensity (Imax) for stressed and unstressed vowels being less in NS). When forcing is very strong a speaker just "shouts out" the stressed syllable while the rest of the word is almost inaudible.

2. The levelling of intensities of stressed and unstressed vowels (the difference between Imax of the stressed vowel and Imax of the unstressed ones in FS is less than in NS). In the extreme case of FS a speaker begins verbally to scan all the syllables.

Some speakers displayed the first tendency, others displayed the second, and some displayed both tendencies. However, speakers more often displayed only one tendency with polysyllabic word. The first tendency is predominant in forced speech if in the normal speech the stressed vowel is not distinguished by its intensity. Such is the case if the stressed vowel in a word is /i/ or /u/ and the preceding vowel is /a/.

In SS also both tendencies are displayed. However, the second one (levelling of intensities) is more frequent. In such cases the intensity contrast may disappear, that is, vowel intensity may coincide with consonant intensity, the consonants being sonants as well as voiced and voiceless fricatives. It is not infrequent

that all sounds in a word are of equal intensities.

Table 2. Duration change of vowels in FS and SS as compared with NS (msec)

vowel / speaker MA		stressed /a/	/o/	/e/	/u/	/i/	pretonic /a/	/u/	posttonic /a/	/'a/	/ɨ/	/i/	/o/
I	NS	200	185	210	260	180	110	120	140	160	150	140	65
	FS	+25	+15	+40	+30	+40	+10	–10	+10	+15	+30	–25	+10
	SS	+65	+50	+90	+95	+65	+10	+40	+15	ø	+55	–10	–10
II	NS	145	130	130	105	105	95	90	75	95	80	75	50
	FS	+25	+25	+25	+55	+30	+25	+25	+30	+30	+15	ø	+25
	SS	–10	–25	–15	+15	–25	–15	+25	–10	–25	+55	–70	–10
III	NS	140	120	145	160	140	75	105	80	80	95	65	50
	FS	+55	+55	+75	+55	+55	+15	ø	+10	+15	–15	+10	+10
IV	NS	230	210	235	265	230	160	130	160	175	130		
	FS	–30	–15	–25	–65	–80	+15	–25	–40	–40	–10		
	SS	+65	+40	+55	–15	+65	+25	+55	–10	–25	–30		
V	NS	195	195	175	200	200	120	130	130	170	120		
	FS	+25	+40	+40	–15	+15	+30	–15	–25	–40	+10		
	SS	–30	–75	–40	–40	–105	+25	–50	ø	–75	–15		
VI	NS	160	155	195	185	140	170	115	95	80	90		
	FS	–10	–30	–10	+10	+25	–30	+40	–15	–10	+55		
	SS	–10	–15	–40	+25	–10	–15	–30	–25	ø	+25		

Table 3. Intensity of vowels in FS and SS as compared with NS

stimulus word / speaker	forced speech: II	IV	I	III	V	VI	sloppy speech: II	IV	I	III	V
pala-tka	+–	++				++	+–	+–			
za-ponki	ø–	+–				++	+–	––			
a-lpha	–	–	–	+	+	–	–	–	–	+	–
uzda-	–	ø	+	–	+	+	–	+	+	+	–
ia-ma	–	–	+	+	–	+	+	+	–	+	–
kho-lodno	––	++				++	–+	++			
t'o-sh'a	ø	+	+	+	–	+	ø	–	+	ø	–
uzhe-	+	+	–	+	–	++	–	+	–	ø	–
e-ta	–	–	–	+	+	+	–	–	+	–	–
ke-pka	–	+				+	+	–			
azu-	+	+	+	–	+	+	–	+	+	+	+
analgi-n	ø+	++				++	ø–	++			
i-kry	––	++	+	–	+	++	++	ø+	+	+	–
i-m'a	–	–	–	–	+	+	–	+	–	–	+

\+ (stressed) vowel more intensive
– (stressed) vowel less intensive
ø (stressed) vowel equally intensive

SPECTRUM

Only visual estimates of the spectrum have been made. Due to these estimates there are no regular substantial changes in the spectrum in FS as compared with that of NS. F_1 and F_2 of the vowels remain at the same frequencies and retain their normal intensity values. It is important that F_1-F_2 difference does not change perceptibly in FS even for /i/, though this contradicts some observations reported in the literature. Despite increasing intensities of higher formants, errors in perception can be more satisfactorily explained by errors in the horizontal rather than the vertical position of the tongue in articulation, that is, information about F_1 is more easily perceived than that about F_2 (cf. frequent substitutions /u/—/i/—/ɨ/ and /o/—/e/). However, in FS there are occasional peculiarities of spectrum that serve to increase its intelligibility as compared with NS. The peculiarities are as follow:

I. Vowel formants occupy the most characteristic frequencies in the spectrum (e.g. F_1 for /a/ and F_2 for /u/ are higher than in normal speech).

2. Consonant noise is amplified at more

characteristic frequencies than in NS.

3. The formants of sonants are physically more distinct (e.g. better physical manifestation of nasal formant, etc.).

In general these peculiarities may, together with the lengthening of sounds frequent in FS, explain the increase in the intelligibility of FS as compared with NS.

FREQUENCY

Table 4. Fo_{max} and Δ Fo of the stressed vowels in various MAs

Vowel	MA speaker	NS Fo_{max}	NS ΔFo	FS Fo_{max}	FS ΔFo	SS Fo_{max}	SS ΔFo
/a/	I	I32	I5	200	52	II7	3
	II	I30	I3	I90	38	II3	20
	IV	I3I	I6	I48	44	I06	9
	V	I29	6	272	I02	II7	II
	VI	I4I	I9	I73	34	-	-
/o/	I	I35	I3	208	55	I37	8
	II	I40	I0	202	35	I37	25
	IV	I34	I7	I56	40	II9	I2
	V	I36	9	280	80	II6	II
	VI	I40	2I	I74	34	-	-
/u/	I	I22	5	2I0	56	II7	2
	II	I42	I2	I98	42	I27	I5
	IV	I44	I9	I73	6I	I08	8
	V	I38	9	276	I06	II3	8
	VI	I43	20	I93	40	-	-
/t /	I	I50	20	240	85	I35	I5
	II	I50	I0	2I5	30	I40	25
	IV	I40	I7	I55	28	II3	8
	V	I40	I3	283	75	I20	I0
	VI	I4I	23	I87	48	-	-
/i/	I	I33	I3	235	55	II8	I0
	II	I43	I5	I95	50	I23	23
	IV	I30	I4	I70	43	III	3
	V	I45	9	285	I06	II8	9
	VI	I4I	I5	I85	40	-	-
/e/	I	I28	5	2I8	50	I23	7
	II	I40	I3	202	38	I32	20
	IV	I38	20	I55	47	I03	7
	V	I32	II	28I	88	II7	8
	VI	I38	I5	I80	37	-	-

As could be seen in table 4, an increase of Fo of 50% in the average is characteristic of FS (the range of Fo-increase is from II% for speaker IV to II3% for speaker V). In SS Fo_{max} increases by about I3% (the range of Fo-increase is from I% for speaker I to 25% for speaker IV). In addition, there is difference between Fo_{max} and Fo_{min} over the same vowel four times greater in FS than in NS (the range is from I,6 times for speaker VI to I7 times for speaker V).

These speech events are consistent and reproducable and as such can serve to distinguish between various modes of articulation.

It is probable that the degree of forcing determines Fo values. The intensity increase and the rise of Fo (maximum and change) are absolute indicators of FS, while the intensity decrease and the fall of Fo (maximum over stressed vowels) are absolute indicators of SS.

PERCEPTION

To investigate perception of speech in various MAs four word-lists containing 3I stimuli each were read by 3 speakers. The record level was adjusted so that all stimuli were equal in intensity. Thus the factor of intensity was excluded since it effects perception greatly, FS being about 3 times more intensive and SS two times less intensive than NS. Three groups of subjects listened to the recordings under the conditions of noise. The intelligibility score for each word-list has been calculated. To neutralize memorizing the order of presentation was as follows: NS, FS, SS.

The average intelligibility scores were 58% for NS, 66% for FS and 48% for SS. The factor of intensity being neutralized three main variables: fundamental frequency, duration and spectrum-determined speech intelligibility. Energy distribution among vowels effects perception as well. Duration, on the other hand, does not effect intelligibility as such (in SS duration is often greater than in NS without any evident effect). The accuracy of articulation is an important factor in the increase of intelligibility in FS and reduction of intelligibility in SS.

It is not infrequent in FS that the better physical manifestation of formants and formant transitions accounts for better vowel identification and better identification of place for adjacent consonants. Consonants in FS are characterized by amplified parts of the spectrum relevant for their identification.

There are certain errors in FS, such as substitution of fricatives for plosives, which may result from release lengthening in plosives (substitution KH for K, etc.) and inserted vowels (/cɨ-kl/ is perceived as /tɨ -pel/, /ku-pol/, etc.). These errors, however, are compensated for by better recognition of other sounds, final vowels and consonants in particular.

Thus the increased intelligibility of FS as compared with NS and SS is explained not only by its being louder but by the change in other parameters like duration, fundamental frequency and spectrum. On the other hand, the increased loudness and more "imperative" sound of FS mobilize the hearer's attention to a greater degree. This may be one reason for the inappropriate use of FS to overcome an interlocutor's dullness or a child's disobedience.

CONCLUSION

FS and SS may be considered as special varieties of speech characteristic of everyday speech communication and, as such, may be of theoretical and practical interest. The peculiarities of FS and SS may prove useful for automatic speech recognition and high quality synthesis of speech.