

Э. КЮННАП

УДК 537.782 : 311.21 : 801.41

СТАТИСТИКА ЭСТОНСКОЙ УСТНОЙ РЕЧИ

(Представлена Н. Алумяэ)

Как известно, устная речь состоит из последовательности звуков, которые произносятся, как правило, слитно, с паузами только после отдельных слов или групп звуков. По объему фраз, т. е. по количеству слов во фразе, устная речь отличается от письменной [1]. Каждый звук имеет индивидуальный характер и оказывает в процессе речи некоторое влияние на конечное состояние предыдущего и начальное состояние следующего за ним звука. Очень часто в устной речи окончания слов произносятся с фонетическими ошибками либо не произносятся совсем, нередко в речь вкрадываются слова, лишенные какого-либо смысла. В данной работе излагаются статистические данные о частоте появления в речи отдельных звуков, а также двух- и трехзвуковых сочетаний и сравниваются в этом плане устная и письменная формы эстонского языка.

Анализируемый материал

Произвольно выбранные отрезки речи между двумя или более лицами, услышанные на собраниях, в магазинах, по радио или телевидению, были записаны на магнитную ленту, переведены на язык фонетических символов, перфорированы и проанализированы на ЭВМ. Анализ был проведен по синтагмам, т. е. по паузам в слитной речи. Выборка содержала 105 942 фонетических символа (звукотипа), образующих в совокупности 19 620 слов и 4923 синтагмы.

Чтобы в тексте отличать звукотипы от букв, первые заключены в скобки, а вторые выделены курсивом.

В алфавите эстонского языка 23 буквы. При написании иностранных имен и выражений используются еще *c, č, f, q, š, z, ž, w, x, y*, из них наиболее часто встречается *f*. Звукотипов в нашем материале больше — их насчитывается 32 (см. табл. 1). Это значит, что на письме разные звукотипы обозначаются одними и теми же буквами. В эстонском языке согласные *l, t, n, s, d* могут быть как палатализованными, так и непалатализованными. Первые отмечены нами апострофом. Наиболее часто палатализируются *l'* и *t'* — 31,1 и 29,7% соответственно, реже *n'* и *s'* — 19,7 и 16,4% соответственно и редко *d'* — 3,3%. В «Oigekeelsussõnaraamat» («Орфографический словарь», Таллин, «Валгус», 1976), содержащем более 100 000 слов, есть 61 пара таких слов, которые пишутся одинаково, но вследствие палатализации имеют различные значения. В тех случаях, когда /u/ в комбинациях /au/ и /ue/ произносится как нечто среднее между /u/ и /v/, он выделен нами в самостоятельный звукотип и отмечен как /w/.

Таблица 1

Звук	Характерное слово		Частотность		Доверительные границы при 99%-ной вероятности
			абсолют.	относит., %	
/a/	kana	курица	12304	11,61	11,36 — 11,87
/e/	neli	четыре	12215	11,53	11,28 — 11,78
/i/	pime	темно	10465	9,88	9,65 — 10,12
/t/	ketas	диск	8204	7,75	7,53 — 7,97
/s/	võsa	поросль	6731	6,35	6,15 — 6,55
/l/	väle	пряткий	5777	5,45	5,27 — 5,63
/u/	tuli	огонь	5446	5,14	4,97 — 5,32
/n/	pinu	поленница	5425	5,12	4,95 — 5,30
/k/	lake	пойло	5056	4,77	4,60 — 4,94
/d/	ida	восток	4202	3,97	3,82 — 4,13
/m/	lame	плоский	4193	3,96	3,81 — 4,12
/o/	logi	лаг	3786	3,58	3,44 — 3,73
/z/	tagasi	назад	2937	2,77	2,64 — 2,90
/r/	tara	забор	2527	2,39	2,27 — 2,51
/j/	maja	дом	2422	2,29	2,17 — 2,41
/v/	kava	план	2397	2,26	2,14 — 2,38
/g/	puga	нож	2295	2,17	2,06 — 2,29
/p/	hapu	кислый	1732	1,64	1,54 — 1,74
/h/	rahe	град	1604	1,51	1,42 — 1,61
/ä/	nägu	лицо	1505	1,42	1,33 — 1,52
/õ/	kõle	нежилой	1460	1,38	1,29 — 1,48
/b/	tuba	комната	1085	1,02	0,942 — 1,10
/ü/	püha	святой	956	0,902	0,827 — 0,977
/l'/	kull	ястреб	306	0,288	0,241 — 0,336
/õ/	lõga	жижа	230	0,217	0,180 — 0,254
/η/	rang	ведро	171	0,161	0,130 — 0,193
/s'/	kass	кошка	142	0,134	0,105 — 0,163
/n'/	kann	игрушка	132	0,124	0,097 — 0,152
/i/	faas	фаза	104	0,098	0,073 — 0,123
/w/	kaua, uuesti	долго, снова	69	0,065	0,044 — 0,085
/t'/	kott	мешок	41	0,039	0,023 — 0,054
/d'/	lodi	барка	13	0,012	0,004 — 0,021
Всего			105942	100,000	

Таблица 2

Класс звуков	Частота появления, %
Гласные	
передние делябиализованные /i, e, ä/	22,83
передние округленные /ü, õ/	1,12
задние делябиализованные /õ, a/	12,99
задние округленные /u, o/	8,72
Всего 45,66	
Согласные	
глухие взрывные /p, t, t', k/	14,20
глухие фрикативные /f, s, s', h/	8,09
полувзрывные /b, d, d', g/	7,17
звонкие фрикативные /v, z, w/	5,09
носовые /m, n, n', η/	9,37
латеральные /l, l'/	5,74
дрожащий /r/	2,39
полугласный /j/	2,29
Всего 54,34	

	/a/	/b/	/d/	/d'/	/e/	/f/	/g/	/h/	/i/	/j/	/k/	/l/	/l'/	/m/	/n/	/n'/
/a/		341	692	3	366	13	572	482	351	364	832	1082	127	699	756	20
/b/	177		1	0	113	1	1	0	115	16	56	40	0	34	32	0
/d/	1102	0		0	485	2	2	0	330	50	168	120	0	137	115	0
/d'/	0	0	0		0	0	0	0	13	0	0	0	0	0	0	0
/e/	693	157	630	0		19	504	167	1101	141	791	1504	43	708	754	65
/f/	7	0	0	0	12		0	0	38	0	0	1	0	0	0	0
/g/	717	0	1	0	247	0		0	553	3	6	11	0	12	16	0
/h/	242	0	1	0	456	0	0		74	32	89	1	0	14	22	0
/i/	134	154	1007	0	128	6	256	96		318	937	599	3	666	1159	6
/j/	1452	0	0	0	222	0	0	0	0		0	0	0	0	0	0
/k/	965	4	0	0	417	0	0	1	330	8		56	0	22	11	0
/l/	535	9	177	0	1761	2	87	6	986	52	110		0	208	103	0
/l'/	1	3	8	0	2	0	13	0	20	210	5	1		21	1	0
/m/	1122	61	0	0	946	1	13	1	1135	7	18	12	0		63	0
/n/	693	2	602	0	963	15	63	0	808	54	150	31	0	95		0
/n'/	0	0	57	0	0	1	0	0	40	0	0	0	0	0	0	0
/ŋ/	0	0	0	0	0	0	161	0	0	0	10	0	0	0	0	0
/o/	13	37	99	10	36	3	95	282	63	32	76	937	5	232	1097	6
/p/	352	0	0	0	357	1	0	0	257	1	5	33	0	4	1	0
/r/	530	13	145	0	272	4	83	0	541	139	32	5	0	62	21	0
/s/	619	1	2	0	1583	5	25	0	864	54	285	34	0	65	46	0
/s'/	1	0	0	0	1	0	0	0	32	70	16	1	0	0	1	0
/z/	152	2	0	0	936	10	6	0	501	27	98	44	0	93	66	0
/t/	1455	3	0	0	1602	10	0	1	535	69	238	236	0	186	136	1
/t'/	0	0	0	0	1	0	0	0	9	0	4	0	0	1	0	0
/u/	73	199	618	0	35	6	278	136	406	61	206	612	11	199	332	34
/v/	887	0	0	0	195	1	0	0	301	4	11	5	0	2	2	0
/w/	34	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0
/ɔ/	0	10	7	0	46	0	0	82	667	11	3	43	1	20	104	0
/ä/	0	72	12	0	113	0	122	208	306	0	57	63	88	2	49	0
/ö/	5	5	30	0	45	1	6	0	7	1	8	16	1	6	1	0
/ü/	1	7	109	0	0	0	8	138	4	6	110	201	27	155	9	0

11962 1080 4198 13 11375 101 2295 1600 10387 1730 4321 5688 306 3643 4897 132

	/a/	/b/	/d/	/d'/	/e/	/f/	/g/	/h/	/i/	/j/	/k/	/l/	/l'/	/m/	/n/	/n'/	/ŋ/
	166	906	0	666	7	596	191	133	1293	920	508	1	996	651	0	0	
285		0	0	134	0	0	0	140	0	4	9	3	61	2	0	0	
595	1		0	524		1	0	886	0	0	176	0	0	582	56	0	
3	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	
364	108	446	0		12	233	427	128	219	406	1574	2	903	874	0	0	
13	1	2	0	19		0	0	6	0	0	2	0	1	15	0	0	
572	1	2	0	495	0		0	256	0	0	87	13	58	0	161	95	
333	0	0	0	165	0	0		96	0	1	6	0	1	0	0	0	
346	98	286	10	1075	38	453	73		0	305	933	20	1124	674	35	0	
364	16	50	0	141	0	3	32	318		8	52	210	7	54	1	0	
830	56	168	0	787	0	6	89	904	0		110	5	18	150	0	10	
1023	40	120	0	1425	1	11	1	573	0	56		1	12	31	0	0	
127	0	0	0	43	0	0	0	3	0	0	0		0	0	0	0	
689	34	137	0	680	0	12	14	666	0	22	201	21		95	0	0	
725	32	115	0	741	0	16	22	1108	0	11	103	1	63		0	0	
20	0	0	0	63	0	0	0	6	0	0	0	0	0	0	0	0	
17	0	0	0	1	0	0	0	112	0	0	0	0	0	0	0	0	
253	83	155	0	291	29	10	11	354	40	775	230	1	104	287	0	0	
287	22	85	0	284	1	3	2	163	0	9	53	0	27	55	0	0	
465	27	37	0	405	0	36	0	139	0	42	11	0	1	17	0	0	
982	74	176	0	876	0	15	14	1178	0	714	71	1	19	230	10	0	
113	0	0	0	4	0	0	0	4	0	0	0	0	0	1	0	0	
443	0	6	0	384	0	0	0	989	0	0	24	0	8	6	0	0	
1126	39	112	0	1139	0	4	328	848	0	487	15	15	20	199	18	0	
5	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	
150	26	340	0	26	4	518	59	30	476	816	232	1	290	406	0	0	
423	43	132	0	425	0	4	54	251	0	21	57	1	20	112	0	0	
0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	
33	4	15	0	39	0	1	0	26	41	185	49	0	137	53	0	0	
14	37	14	0	20	1	0	0	16	154	110	151	0	80	182	0	0	
8	3	0	0	5	1	0	0	20	0	1	5	0	12	6	0	0	
91	18	34	0	49	10	102	0	36	1	181	43	1	19	122	0	0	

Таблица 3

	/ŋ/	/o/	/p/	/r/	/s/	/s'/	/z/	/t/	/t'/	/u/	/v/	/w/	/õ/	/ä/	/õ/	/ü/	
17	253	287	470	990		113	524	1156	7	150	433	0	33	14	8	91	11246
0	83	22	27	74		0	0	39	0	26	43	0	4	37	3	18	962
0	155	85	37	176		0	6	112	0	347	132	0	15	14	0	34	3624
0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	13
1	291	284	427	895		5	475	1167	7	26	430	1	39	20	5	49	11399
0	29	1	0	0		0	0	0	0	4	0	0	0	1	1	10	104
0	10	3	36	15		0	0	4	0	540	4	0	1	0	0	102	2281
0	12	2	0	14		0	0	339	0	61	54	0	0	0	0	0	1413
112	355	164	139	1209		4	1070	872	0	30	253	0	26	16	20	36	9775
0	41	0	0	0		0	0	0	0	510	0	0	41	154	0	1	2421
0	775	9	42	804		0	0	96	0	834	21	0	185	110	1	181	4872
0	231	53	11	71		0	24	562	0	237	57	0	49	151	5	43	5530
0	1	0	0	1		0	0	15	0	1	1	0	0	0	0	1	305
0	104	27	1	19		0	8	20	0	294	20	0	137	80	12	19	4120
0	295	55	17	230		1	6	200	0	411	112	0	53	182	6	122	5166
0	0	0	0	10		0	0	22	0	0	0	0	0	0	0	0	130
0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	171
10	46	326	98			1	85	123	17	6	39	0	1	2	0	3	3780
0	149	117	34			0	0	15	0	151	1	0	75	142	10	13	1718
0	154	10	33			0	7	123	0	115	128	0	13	46	0	16	2492
0	183	152	7			0	0	1763	0	397	98	0	111	4	9	50	6357
0	1	0	0	0			0	16	0	0	1	0	0	0	0	1	141
0	146	50	23	137		0		78	0	133	73	0	7	11	1	10	2604
0	282	95	177	609		0	0		0	1010	126	0	80	216	144	70	7281
0	1	0	0	17		0	0	1		0	2	0	0	0	0	1	37
5	44	76	255	414		14	645	419	1		127	67	7	8	3	7	5298
0	41	1	0	7		0	4	3	0	52		0	559	291	0	3	2369
0	0	0	0	0		0	0	0	0	0	0		0	0	0	0	69
2	0	148	55	6		0	10	147	0	90	8	0		0	0	0	1460
24	3	19	295	26		1	4	34	2	8	4	0	0		0	0	1512
0	2	2	15	5		0	4	57	0	0	3	0	0	0		3	223
0	0	6	16	18		3	64	66	7	0	0	1	0	0	0		956
171	3641	1597	2493	5912		142	2936	7449	41	5433	2170	69	1436	1499	228	884	99829

Таблица 4

/o/	/p/	/r/	/s/	/s'/	/z/	/t/	/t'/	/u/	/v/	/w/	/õ/	/ä/	/õ/	/ü/	
12	349	484	605	1	141	1357	0	73	879	29	0	0	5	1	/a/
36	0	11	1	0	2	3	0	179	0	0	10	66	4	7	/b/
98	0	136	2	0	0	0	0	407	1	0	7	12	27	103	/d/
10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	/d'/
36	357	261	1478	1	918	1538	1	35	193	34	46	110	45	0	/e/
3	1	4	5	0	10	10	0	6	1	0	0	0	1	1	/f/
95	0	83	25	0	6	1	0	278	0	0	0	122	6	8	/g/
248	0	0	0	0	0	0	0	136	0	0	82	205	0	138	/h/
57	256	524	859	29	417	491	9	402	292	0	600	304	7	4	/i/
31	1	139	54	69	27	69	0	61	3	0	11	0	1	6	/j/
76	5	31	285	16	98	238	0	202	11	0	3	57	8	110	/k/
929	33	5	34	1	44	236	0	587	5	0	43	62	14	179	/l/
5	0	0	0	0	0	0	0	10	0	0	1	88	1	27	/l'/
232	4	62	65	0	93	185	1	197	2	0	19	2	6	155	/m/
988	1	20	46	0	66	136	0	328	2	0	104	49	1	9	/n/
6	0	0	0	0	0	1	0	34	0	0	0	0	0	0	/n'/
10	0	0	0	0	0	0	0	5	0	0	2	24	0	0	/η/
46	149	154	179	2	146	282	1	44	41	0	0	3	2	0	/o/
323	116	10	152	0	50	95	00	76	1	0	148	19	2	5	/p/
1	0	0	0	0	23	177	0	252	0	0	55	295	15	16	/r/
83	0	7	0	0	137	600	17	406	7	0	6	26	5	18	/s/
119	15	120	1564	15	78	0	0	14	0	0	0	1	0	3	/s'/
17	0	0	0	0	0	0	0	1	0	0	0	2	0	7	/t/
5	150	109	395	0	126	979	0	52	0	89	0	0	0	0	/t'/
38	1	122	98	1	73	126	2	126	0	8	4	3	0	0	/u/
0	0	0	0	0	0	0	0	67	0	0	0	0	0	0	/v/
1	75	13	111	0	7	80	0	7	559	0	0	0	0	0	/w/
0	141	46	4	0	11	216	0	8	291	0	0	0	0	0	/õ/
0	10	0	9	0	1	137	0	3	0	0	0	0	0	0	/ä/
3	13	16	50	1	10	70	1	7	3	0	0	0	3	0	/õ/
															/ü/

	/a/	/b/	/d/	/d'/	/e/	/f/	/g/	/h/	/i/	/j/	/k/	/l/	/l'/	/m/	/n/
/a/		340	692	3	364	13	414	464	337	360	822	1066	127	694	747
/b/	176		1	0	111	1	1	0	113	16	56	40	0	34	32
/d/	1102	0		0	485	2	2	0	329	50	168	120	0	137	115
/d'/	0	0	0		0	0	0	0	13	0	0	0	0	0	0
/e/	688	156	628	0		18	486	164	994	140	764	1493	42	707	742
/f/	7	0	0	0	12		0	0	38	0	0	1	0	0	0
/g/	717	0	1	0	247	0		0	553	3	6	11	0	12	16
/h/	242	0	0	0	455	0	0		74	32	89	1	0	14	22
/i/	134	154	1007	0	128	6	244	96		318	932	585	3	666	1149
/j/	789	0	0	0	222	0	0	0	0		0	0	0	0	0
/k/	854	4	0	0	334	0	0	1	318	8		52	0	22	11
/l/	525	9	177	0	1747	2	87	6	976	52	110		0	208	103
/l'/	1	3	8	0	2	0	13	0	20	210	5	1		21	1
/m/	1009	61	0	0	894	1	12	1	835	7	18	12	0		63
/n/	654	2	602	0	908	15	63	0	657	54	150	31	0	95	
/n'/	0	0	57	0	0	0	0	0	40	1	0	0	0	0	0
/m/	0	0	0	0	0	0	161	0	0	0	10	0	0	0	0
/o/	13	33	99	10	36	3	94	277	61	32	75	903	5	213	1033
/p/	315	0	0	0	328	1	0	0	250	1	5	32	0	4	1
/r/	517	13	145	0	269	4	83	0	536	139	32	5	0	62	21
/s/	581	1	2	0	1269	5	25	0	649	54	285	34	0	65	46
/s'/	1	0	0	0	1	0	0	0	32	69	16	1	0	0	1
/z/	152	2	0	0	936	10	6	0	501	27	98	44	0	93	66
/t/	1408	3	0	0	1536	10	1	0	534	69	238	236	0	186	136
/t'/	0	0	0	0	1	0	0	0	9	0	4	0	0	1	0
/u/	73	199	617	0	35	6	278	134	406	61	204	612	11	195	332
/v/	830	0	1	0	187	1	0	0	283	3	11	5	0	2	2
/w/	34	0	0	0	35	0	0	0	0	0	0	0	0	0	0
/õ/	0	10	6	0	46	0	0	79	660	11	3	42	1	18	104
/ä/	0	72	12	0	113	0	122	208	306	0	57	61	88	2	49
/ö/	5	5	30	0	43	1	6	0	7	1	8	16	1	6	1
/ü/	1	7	109	0	0	1	8	113	4	6	90	187	26	155	9

Эстонский /s/ произносится обычно только с шумом. Но иногда, если он стоит между гласными или после сонорного согласного, в его произношении принимают участие и голосовые связки, в результате чего шум сопровождается тональным компонентом. В этих случаях /s/ воспринимается как полувзвонный и обозначается здесь как /z/.

В эстонском языке /п/ может быть обычным, палатализованным и назальным. Эти звуки выполняют смыслоразличительную функцию. Обычный встречается в любом фонетическом окружении, а назальный — лишь в комбинациях /ng/ и /nk/. Он отмечен здесь как /ŋ/. Гласный /i/ и полугласный /j/ в фонетическом отношении почти одинаковы, только /i/ произносится сильнее, яснее и имеет все степени долготы. /j/ встречается только в начальной позиции слога и слова. Здесь они рассматриваются как два разных звука.

В данной работе долгота звуков, которая также изменяет смысл слов, во внимание не принимается.

В эстонском языке *b*, *d* и *g* используются для индикации кратких форм /p/, /t/ и /k/. Однако в некоторых случаях /b/, /d/ и /g/ отличаются от /p/, /t/ и /k/ не только уровнем интенсивности, но и спектром и способом произношения. Глухие /p/, /t/ и /k/ произносятся путем кратковременного полного смыкания голосового тракта и затем быстрого раскрытия ретрограда. В зависимости от манеры раскрытия рта произносятся /p/, /t/ или /k/. Если /b/, /d/ или /g/ стоят перед гласной, голосовые связки дают короткие импульсы в пределах основного тона, которые быстро исчезают до окончания произношения. Поэтому /b/, /d/ и /g/ приняты здесь за разные звукотипы /p/, /t/ и /k/ и условно названы полувзвонными взрывными. Итак, в эстонской устной речи нами выде-

Таблица 5

/n'/	/η/	/o/	/p/	/r/	/s/	/s'/	/z/	/t/	/t'/	/u/	/v/	/w/	/õ/	/ä/	/ö/	/ü/
20	17	252	284	457	982	112	521	1154	7	146	429	0	33	14	8	91
0	0	83	22	27	74	0	0	39	0	26	43	0	4	37	3	18
0	0	154	85	37	176	0	6	112	0	347	132	0	15	14	0	34
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	1	291	284	412	889	5	460	743	7	25	430	1	39	20	5	49
0	0	27	1	0	0	0	0	0	0	4	0	0	0	1	1	9
0	0	10	3	36	15	0	0	4	0	540	4	0	1	0	0	102
0	0	12	2	0	14	0	0	339	0	61	54	0	0	0	0	0
6	112	355	164	139	1206	3	1062	872	0	30	253	0	26	16	20	36
0	0	39	0	0	0	0	0	0	0	494	0	0	38	145	0	1
0	0	701	9	41	804	0	0	96	0	564	21	0	165	104	1	164
0	0	224	53	11	71	0	24	562	0	234	57	0	45	136	5	42
0	0	1	0	0	1	0	0	15	0	1	1	0	0	0	0	1
0	0	102	27	1	19	0	8	20	0	258	20	0	121	76	12	17
0	0	123	55	17	230	1	6	200	0	408	112	1	49	149	6	100
0	0	0	0	0	10	0	0	22	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	9		43	323	95	0	85	114	15	6	39	0	1	0	0	3
0	0	144		103	34	0	0	15	0	142	1	0	63	137	10	9
0	0	151	10		33	0	7	123	0	115	128	0	12	38	0	0
0	0	178	151	7		0	0	1760	0	371	98	0	106	2	8	47
0	0	2	0	0	0		0	16	0	0	1	0	0	0	0	0
0	0	146	50	23	137	0		78	0	133	73	0	7	11	1	10
1	0	274	95	174	607	0	0		0	980	126	0	72	167	138	69
0	0	1	0	0	17	0	0	0		0	2	0	0	0	0	1
34	5	44	76	252	413	14	644	418	1		126	64	7	8	3	7
0	0	37	1	0	7	0	4	3	0	52		0	453	260	0	3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	138	55	6	0	10	147	0	90	8	0		0	0	0
0	0	3	19	293	22	1	4	34	2	0	4	0	0		0	0
0	0	2	2	15	5	0	4	57	0	0	3	0	0	0		3
0	0	0	6	15	18	3	64	55	7	0	0	0	0	0	0	

лены следующие 32 звукотипа: /a, b, d, d', e, f, g, h, i, j, k, l, l', m, n, n', η, o, p, r, s, s', z, t, t', u, v, w, õ, ä, ö, ü/.

Таблицы

В табл. 1 приведены перечисленные выше звукотипы и слова, их характеризующие. Из 32 звукотипов относительная частота появления девяти гласных составляет 45,66%, а из всех гласных наиболее употребительны /a, e, i/ — 33,02%, причем /a/ встречается в 53 раза чаще, чем /õ/. Гласный /i/ появляется в 4 раза чаще полугласного /j/ (на письме в 6 раз чаще). Обращает на себя внимание тот факт, что половина встречаемости всех гласных приходится на гласные /a, e/. Частота появления /k, l, n, s, t/ составляет больше половины встречаемости всех остальных согласных, вместе взятых. Среди перечисленных пяти звукотипов наиболее распространен /t/ (на письме s). Из взрывных согласных /t, t'/ появляются примерно в 5 раз чаще, чем /p/, и в 2 раза чаще, чем /k/. В табл. 2 звукотипы сгруппированы согласно их артикуляции и частоте появления в речи.

В табл. 3 приведены данные о частотности двухзвуковых сочетаний (диграмм). По вертикали выписаны первые и по горизонтали вторые звуки диграмм. Интересно отметить следующее. В отдельных словах печатного текста нами было выявлено 25 разных дифтонгов, причем на первом месте могут стоять все гласные, а на втором — гласные a, e, i, o. В устной же речи количество различных гласных диграмм в синтагмах увеличивается до 53, и в первой позиции могут находиться все гласные. Теоретически из девяти гласных можно образовать 72 сочетания. Сле-

довательно, в синтагмах встретилось из них 73,61%. В нашем материале зафиксировано 586 разных диграмм из возможных 992, или 59,07%. Из всех 99 829 диграмм более 10 раз встретилась 99 241 диграмма в 416 комбинациях, более 5 раз 516 диграмм в 106 комбинациях и по одному разу 72 диграммы. Пять наиболее употребительных диграмм: /st/ — 1763, /le/ — 1761, /te/ — 1602, /se/ — 1583, /el/ — 1504 (на письме из 94 010 диад, образованных из 102 307 букв, наиболее используемыми были следующие: *se* — 2624, *is* — 2011, *st* — 1751, *te* — 1708, *es* — 1566).

Теоретически из 32 звукотипов можно создать 29 760 трехзвуковых сочетаний (триграмм) без повторения рядом одинаковых звуков. Практически же удалось выявить только 3026 таких сочетаний, или 10,16%. Перечислим пять с высокой частотностью: /ele/ — 526, /ist/ — 489, /sel/ — 464, /est/ — 400, /ole/ — 398 (на письме: *ste* — 266, *sel* — 230, *use* — 224, *est* — 216 и *mis* — 197). По одному разу встретилось 950 триграмм. В табл. 4 приведены правые и в табл. 5 левые соседи центральных звуков. Последние выписаны на горизонтальной линии таблиц.

Относительная частота появления звуков в процентном выражении вычисляется по формуле $z_i = x_i/n \cdot 100$, где x — частотность данного звука i среди всех звуков n , вероятность появления p_i которого неизвестна. Располагая большим количеством n , можно предполагать, что $p_i \approx z_i$. Поскольку z_i является лишь одним значением случайной величины, желательно знать ее доверительные границы. При 99%-ной доверительной вероятности изменчивость переменной величины выражается формулой

$$p_{1,2} = x/n \pm 2,576 [x(n-x)/n^3]^{1/2}.$$

Результаты расчета приведены в табл. 1. Например, для /a/, абсолютная частотность которого 12 304 из 105 972 звуков и относительная частотность 11,61%, доверительные границы равны 11,36—11,87%. Следовательно, с 99%-ной вероятностью можно прогнозировать, что в эстонской устной речи /a/ встретится от 11 360 до 11 870 раз на каждые 100 000 звуков.

Заключение

Проведен статистический анализ частотности появления отдельных звуков, двух- и трехзвуковых сочетаний в произвольно выбранных отрезках эстонской устной речи. Установлено, что наиболее употребительными являются звук /a/, диграмма /st/ и триграмма /ele/ (в печатном тексте буква *a*, диада *se* и триада *ste*). Слово в устной речи состоит в среднем из 5,4 звука и синтагма из 22,5 звука (в печатном тексте эти данные равны соответственно 7 и 35 буквам). Синтагма содержит сочетания звуков, отсутствующие в отдельных словах. Результаты исследования могут быть использованы при разработке правил управления синтезатором речевых сигналов.

ЛИТЕРАТУРА

1. Кюннп Э., Изв. АН ЭССР, Физ. Матем., 26, № 4, 424—430 (1977).

Институт кибернетики
Академии наук Эстонской ССР

Поступила в редакцию
22/XII 1978

E. KUNNAP

EESTI KÕNEKEELE STATISTIKA

Artiklis on analüüsitud eesti kõnekeelt ja toodud andmeid häälikute ning nendest koosnevate digrammide ja trigrammide esinemissageduse kohta. Kokku on uuritud 105 942 häälikut, mis moodustavad 19 620 sõna ja 4923 süntagmat.

E. KUNNAP

ON THE STATISTICS OF SPOKEN ESTONIAN

Different combinations of sounds are used to synthesize fluent speech. Every sound in a syllable has some influence upon the final state of the preceeding sound as well as on the following sound. Therefore, in order to investigate the rules of the control of a synthesizer of human speech, it is necessary to have data about the probability of the appearance of the sounds in fluent speech, to find out which combinations are more frequent and which combinations do not exist at all.

Conversation differs from written text by both the duration of a phrase and the composition of words in the phrase. Very often in conversation the endings of words are pronounced with phonetic mistakes or are not pronounced at all, sometimes some parasite words which have no sense and do not exist in written text, are added. The other aim of this work is to obtain the statistical data on spoken Estonian and the differences between the latter and the written text in both the frequency of occurrence of sounds and the composition of digrams and trigrams. The analysis was made by syntagmas, i.e. from pause to pause in real speech. The material is based on random fragments from the speech between two or more persons, from the conversation held at meetings, in shops, radio and TV announcements. All in all 105 942 sounds were investigated forming 19 620 words and 4923 syntagmas. In Tables 3—5 the frequency of occurrence of sounds as well as digrams and trigrams are shown.