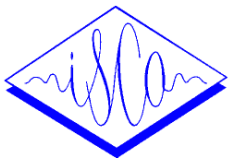

ESCA Tutorial and Research Workshop on
Speech Processing in Adverse Conditions
Cannes-Mandelieu, France
November 10-13, 1992

On the influence of noise on speech production and perception

D. Rostolland

Institut de Linguistique et Phonétique Générales et Appliquées,
CNRS/URA 1027, 19 rue des Bernardins, 75005 Paris, France

An experiment was conducted to study the variations of vowel duration as a function of voice intensity level. Acoustical analyses were carried out on a set of utterances produced in noise. The results showed differences in the durations of vowels as a function of noise, post-vocalic consonant and of vowel aperture. Perceptual implications might be important.

1. INTRODUCTION

It is known for a long time that the intelligibility of normal spoken voice much depends on the speaker, especially when the listener is in a disturbing noise (1). If the speaker itself is in a loud enough noise, the vocal signal is modified at the root in intensity, frequency and duration. As a matter of fact, the speaker is obliged to raise his voice, to try to control to be heard and understood. This vocal effort consequently leads to various alterations of the acoustic signal which become in particular more high-pitched : people raises one's voice for we cannot shout in a low-pitched register. The rhythm and the timing of the signal are changed too because we generally talk slower. It is this latter point that we will approach here, with a simple verbal material (CVC syllables).

The variations of the speech signal in terms of vocal effort were relatively little studied on a phonemic scale, how Traunmüller noted this in his paper on the paralinguistic variations and invariance of the vowels (2) : "The articulatory, acoustic and perceptual consequences of variations in vocal effort have only recently been subjected to thorough experimental investigation (3-4). The effects of variations in vocal effort have not been considered in treatises on theories of speech production and perception such as action theory and motor theory of speech perception to which these effects appear to be crucially relevant".

When we raise one's voice, acoustic characteristics and phonetic features change more or less according to the conditions (3-4-5-6-7-8). At the limit, the deformations occured as one shouts decrease more or less the intelligibility too, by a signal / noise ratio maintained constant (9-10).

The study of relationships between the intensity, frequency and duration, in continuous noisy speech or for isolated words, is relatively recent (2-3-4-5-6-7-8-10-11-12-13-14). These research especially carried out about english language produced and perceived in noise, show that this field begin to be explored phonemically, while suprasegmental

studies are ancient (15). It is as if a large field of phonetics in quiet must be now studied in noise.

2. METHOD

The verbal material is made up of monosyllables placed in stressed position within a frame sentence (he said "C1 V C2" three times). One set C1 = /k/, and V = /a/ or /i/. In order to study the influence of the mode, the manner and the place of articulation of the final consonant, we select C2 = / p k b g f ʃ v ʒ /.

The recording were realized in an audiometric cabin, at a rate of one sentence every 4 sec. Six male speakers from the parisian region were recorded. The same series of sentences are pronounced in 3 levels of uniform noise : 50 dB (quiet), 80 dB (medium noise) and 85 dB (loud noise). The 3 corpus differ from one another by only the type of voice : normal (L1), loud (L2) and shouted (L3).

The reproduction of the magnetic tapes were made using the same recorder and a graphic sound level recorder. This recorder is set in a writing speed of 100 cm/s, and a logarithmic potentiometer is used. For the difficult segmentations (with / f ʃ v ʒ /), a Sonagraph and a dual time base memory oscilloscope, for expend the transitions, were used.

The data as a whole include 6 (subjects) * 2 (vowels) * 8 (consonants) * 3 (levels) * 10 (repetitions) = 2880 measures of duration of V. In order to segment, we found upon the informations brought by the different devices and we carry the criterions defined by several authors (in 16) and which we already have had the opportunity to make use (3-5-7-10). You particularly encounter problems of segmentation towards the end of the vowels (16).

3. RESULTS

If the noise varies from L1 to L2 the duration of /i/ more increases than the one of /a/ (33 % and 9 %). From L2 to L3, the increases are of 17,5 %. Between L1 and L3, /i/ increases by 56 % and /a/ by 29 % (table 1). This table also shows that the variation between subjects increases at L2 (141 % for /a/ and 128

Table 1
Vocalic duration (ms) in function of ambient noise and speaker

		/a/			/i/		
Levels		L1	L2	L3	L1	L2	L3
Subject	S1	226.9	217.1	201.0	157.1	195.9	188.0
	S2	134.8	185.6	298.8	91.7	120.5	217.6
	S3	258.4	318.5	317.5	181.8	274.3	248.4
	S4	140.9	132.1	183.8	103.0	132.4	183.2
	S5	145.9	152.1	209.9	101.3	132.9	184.6
	S6	168.1	172.5	176.4	120.8	146.0	155.1
Mean	:	179.2	196.3	231.2	125.9	167.0	196.1

Table 2
Vocalic duration (ms) in function of ambient noise and consonant

		/a/			/i/		
Levels		L1	L2	L3	L1	L2	L3
Consonant	p	125.7	153.2	194.3	86.1	131.8	166.5
	k	137.8	167.2	203.1	92.6	141.8	169.9
	b	165.2	189.5	221.4	110.4	159.6	190.4
	g	186.3	205.8	236.2	124.3	174.5	201.1
	f	164.8	183.8	224.0	112.4	157.1	186.6
	ʃ	182.9	199.7	238.0	123.3	161.8	195.6
	v	213.9	218.7	252.8	157.8	188.3	217.8
	ʒ	256.8	252.4	279.9	200.6	221.3	241.2

% for /i/) but that at L3 it is less (80 and 60 %) than the one at L1 (92 and 98 %). In other respects, the duration of /a/ exceeds the one of /i/ by 42 % at L1 but by 18 % only at L2 and L3. The maximum variation of duration according to the noise, the subjects and the two vowels, is of 246 % (317.5 ms / 91.7 ms).

Vocalic duration often much depends on the post-vocalic consonant and the noise in which we are speaking. The table 2 shows that the variation between consonants decreases with the noise : 104 % at L1, 65 % at L2 and 44 % at L3 with /a/ ; 133 %, 68 % and 45 % with /i/. The maximum variation of duration according to the noise, the consonants and the vowels, is of 225 % (279.9 ms / 86.1 ms). The detail of variations per subject is indicated on figure 1 (for the extreme durations of /kip/ and / kaʒ /.

The maximum variations of duration for a given vowel and consonant when we speak in quiet and then in noise (from L1 to L3), arises for /kip/ (93 %) and for /kik/ (84 %). It is at L1 that we found the maxi-

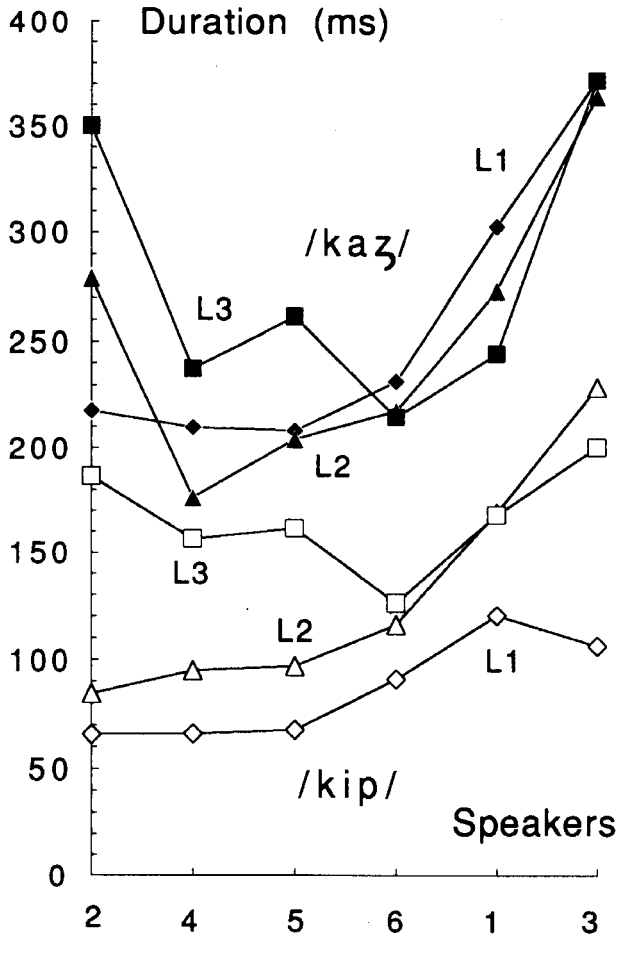

Figure 1. Vocalic duration for /kip/ and / kaʒ / in normal voice (L1), in loud voice (L2) and in shouting (L3)

Table 3
Effect of consonant category (%)

	/a/			/i/		
	L1	L2	L3	L1	L2	L3
Mode	35	23	16	43	28	18
Manner	34	22	18	45	24	16
Place	15	12	8	16	11	6

mum variation between the 2880 individual measures, with a ratio of 9.7 (435 ms for /a/ in / kaʒ / versus 45 ms for /i/ in /kip/). At L2 and L3, the corresponding ratios are of 5.6 and 3.7 only.

While dividing the consonants in categories (voiceless/voiced, stops/fricatives, labials/velars), we see that the effect of the voicing, of the manner and of the place of articulation of the post-vocalic consonant clearly decreases according to the noise (table 3).

4. CONCLUSION

When the noise increases from L1 to L2, we observe a clear augmentation of the vocalic duration, especially for /i/, and of the variation between subjects, with /a/ and /i/. From L2 to L3, the durations still increases (as much for /a/ as for /i/) but the variation between subjects much decreases. The noise decreases the variation of duration which the vocalic aperture deserves, and the one which the retroactive effects of consonants deserve. This tendency to the timing uniformisation probably reduces the intelligibility of noisy speech, independently of the other possible causes (1-3-9-10).

REFERENCES

1. J.P. Egan, Articulation Testing Methods, Laryngoscope 58 (1948).
2. H. Traunmüller, Paralinguistic Variation and Invariance, Phonetica N° 45 (1988) 29.
3. D. Rostolland, Acoustic-Phonetic Features of Shouted Voice, Acustica N° 50-51 (1982) 118.
4. R. Schulman, Articulatory Dynamics of Loud and Normal Speech, JASA N° 84 (1988).
5. D. Rostolland et al., The effect of Voice Level on Vowels Duration, 10th ICPS, Utrecht (1983).
6. D. Pisoni et al., Some Acoustic-Phonetic Correlates of Speech, ICASSP (1985) 1581.
7. D. Rostolland, Effet du bruit sur la variabilité des locuteurs Séminaire SFA, Marseille (1989).
8. W. Summers et al., Effects of Noise on Speech Production, JASA N° 84 (1988) 917.
9. J.M. Pickett, Effects of Vocal Force on the Intelligibility of Speech, JASA N° 28 (1956) 902.
10. D. Rostolland, Intelligibility of Shouted Voice, Acustica N° 57 (1985) 103.
11. M. Grenié et al, Acoustic-Phonetic Analysis of Speech under Noise, Eurospeech, Paris (1989).
12. J.C. Junqua et al, A Comparative Study of Cepstral Lifters and Distance Measures, ICASSP (1989)
13. Z. Bond et al., Some Phonetic Characteristics of Sentences, Spring Meet., Washington (1988).
14. B.J. Stanton et al., Acoustic-Phonetic Analysis of Loud and L. Speech, ICASSP (1988) 331.
15. T.D. Hanley et al., Effect of Level of Distracting Noise, Jour Speech-Hearing Disorders (1949).
16. A. Di Cristo, De la microprosodie à l'intonosyntaxe, Doctorat d'Etat ès-Lettres, Aix-en-Provence (1978).