ELECTROMYOGRAPHICAL CORRELATES OF SHOUTED AND WHISPERED VOICE

JEAN-FRANCOIS P. BONNOT

Laboratoire de Phonétique et Centre de Recherches en Informatique
et Combinatoire, UA CNRS 1099
Université du Maine, 72017 Le Mans Cedex, France

ABSTRACT

L'examen de l'activité électromyographique des muscles orbiculaire des lèvres et élévateur du voile lors de la production en voix normale de logatomes CVCVCV, nous a permis de mettre en évidence un modèle d'encodage hiérarchisé. L'émission en voix criée et en voix chuchotée de séquences du même type entraîne une restructuration de l'organisation temporelle (2 sujets; muscles étudiés: orbiculaire inférieur, élévateur, palatopharyngien). Ces conditions exceptionnelles présentent un certain nombre de caractères communs: il sèmble en particulier que la programmation opère à plus court terme, par unités de la taille de la syllabe ou du segment; les instructions motrices seraient donc fragmentées, au lieu d'être émises sous forme de "liste". D'autre part, l'élévateur du voile est plus sensible que le palatopharyngien aux modifications dans la situation de production. Une explication physiologique et linguistique est proposée.

INTRODUCTION

In two recent publications (Bonnot |1| and Bonnot et al.|2|), we brought forward a certain amount of experimental evidence supporting the concept of a temporal hierarchical organization of speech production. The utterance (CVCVCV nonsense words) was partly preprogrammed and C1 constituted an encoding reference for the whole item. A local reappraisal of timing arose during phonation, determining a re-structuration of the electromyographical activity (orbicularis oris sup.: OOS, levator veli palatini: LP) on an intrasegmental level. The basic motor controls were thus governed by two components operating in two different temporal fields: the sequencing was in charge of the seriation of the units and depended on the macro-structure. The phasing was related to the micro-structure. Its role was to produce the necessary adjustments and to protect the fluency (Kent |3|, Glencross |4|). This theory, which implies that time is a controlled variable, is compatible with a structural linguistic description because it accounts properly for the translation between an abstract dimension and a superficial one. The preprogrammed component carries out the choice and the transfer of units from the phonemic level to the phonetical level. This process is followed by allophonic specifications.

The model does not exclude biomechanical effects, but subordinates them to the programming requirements of the voluntary movement.

These experiments can be integrated to the framework of a normal use of the possibilities of the vocal tract. As is pointed out by Lubker |5|, apropos of the velopharyngeal mechanism, it is tempting to take up a teleological standpoint. The muscular activity and the articulatory gestures are organized and directed toward the goal of communication. Of course, speech production depends on temporal and physiological "boundary limits". The performer has to take into account the constraints peculiar to the implemented structures. The velocity and the accuracy of the various articulators or of parts of an articulator vary very much, as was shown by Eek |6| and Bothorel |7| among others. Furthermore, the motor task has to be carried out in a well-defined period. For Lubker |5|, "within these boundary limits, speakers have a great deal of variability open to them in their use of the velo-pharyngeal system." This variability, within or between subject(s), can also be linked to specific configurations of the tractus (post-operative patients, dental prosthesis ...) or just to unusual circumstances, such as local anaesthesia or shouted and whispered voice.

Both latter cases belong, like the pathological ones, to the "extrinsic variability", which is partly independent of the structure of the phonological system and of the "physiological weight" of the articulatory units. However, the point here is that we are within a natural use of the possibilities of the phonatory apparatus and of its motor controls. It can be proposed that the model which is described above undergoes a drastic restructuring when it comes to encounter these requirements.

RESULTS AND DISCUSSION

In order to test this hypothesis, we recorded two male native speakers of French (DA,JFB). The subjects were instructed to read nonsense CVCVCV words in normal, whispered and shouted voice. The consonant was |p t k g R| and the vowel |i u|.

We selected the following parameters: acoustical duration of CVCVCV; duration of the activity of the orbicularis oris inf.(OOI), LP, and palatopharyngeus (PPH); latency time of the same muscles (in the present case, interval between the onset of electromyographical activity and the first periodic oscillation for the initial vowel: |p t k| were of course

voiceless. In view of comparison, an identical procedure was used for the voiced |g| and |R|). Student's t test and, in some cases, Cochran's test were applied. The coefficient of variation (C.V.= 100.D/x) were calculated. With JFB, we could not obtain a good signal for LP during this session.

We first noticed that PPH and LP were acting in a very different way for subject DA. Whereas LP was very sensitive to the three conditions of production, the pattern of PPH remained steadier: in table 1, it can be seen that the normal, whispered and shouted mean values of LP are perfectly separated. On the contrary, there is considerable overlapping for PPH. The statistical comparisons reached a significant threshold in 12 cases out of 15 for LP, but in only 3 cases out of 15 for PPH.

Total duration of LP (in msec). Subject DA

	Normal		Shouted		Whisp.	
	$\bar{X}$	C.V.	$\bar{X}$	C.V.	$\bar{X}$	C.V.
|p	1545	11.52	1417	3.82	1216	11.73
t	1521	9.69	1403	3.09	1226	11.71
k	1529	10.19	1424	5.54	1260	13.48
g	1698	12.56	1498	5.97	1163	12.53
R|	1549	12.93	1450	5.29	1139	10.15

|p	Shouted vs. Normal	ddl: 18 NS
	Shouted vs. Whisp.	ddl: 18 $p < 0.05$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$
t	Shouted vs. Normal	ddl: 18 $p < 0.05$
	Shouted vs. Whisp.	ddl: 18 $p < 0.05$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$
k	Shouted vs. Normal	ddl: 17 NS
	Shouted vs. Whisp.	ddl: 17 $p < 0.05$
	Whisp. vs. Normal	ddl: 16 $p < 0.01$
g	Shouted vs. Normal	ddl: 18 $p < 0.05$
	Shouted vs. Whisp.	ddl: 18 $p < 0.001$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$
R|	Shouted vs. Normal	ddl: 18 NS
	Shouted vs. Whisp.	ddl: 18 $p < 0.001$
	Whisp. vs. Normal	ddl: 18 $p < 0.001$

TABLE 1 A

Total duration of PPH (in msec). Subject DA

	Normal		Shouted		Whisp.	
	$\bar{X}$	C.V.	$\bar{X}$	C.V.	$\bar{X}$	C.V.
|p	1168	6.91	1236	6.58	1226	7.70
t	1156	16.64	1302	11.94	1094	7.38
k	1385	7.68	1464	1.96	1317	8.27
g	1502	9.04	1512	6.95	1354	1.61
R|	1193	15.37	1395	7.61	1167	5.26

|p	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 8 NS
	Whisp. vs. Normal	ddl: 8 NS
t	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 8 $p < 0.05$
	Whisp. vs. Normal	ddl: 8 NS
k	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 7 NS
	Whisp. vs. Normal	ddl: 8 NS
g	Shouted vs. Normal	ddl: 8 NS
	Shouted vs. Whisp.	ddl: 8 NS
	Whisp. vs. Normal	ddl: 8 NS
R|	Shouted vs. Normal	ddl: 8 $0.10 > p > 0.05$
	Shouted vs. Whisp.	ddl: 8 $p < 0.01$
	Whisp. vs. Normal	ddl: 8 NS

TABLE 1 B

NB: the smaller number of items for the "total duration of PPH" is due to the following fact: in most cases, for subject DA, te activity of PPH was absent or very weak for the nonsense words CiCiCi. Consequently, we took only into account utterances with |u|. For a detailed discussion,cf.Bonnot |1|.

It can thus be suggested that some muscles are more directly sensitive to those kinds of extrinsic constraints. It can be recognized that PPH plays a role in the narrowing of the velopharyngeal Isthmus (Fritzell |8|; Legent, Perlemuter and Vandenbrouck |9|), but there is no denying that LP is the only one which is responsible for the upward gesture of the velum, and to a great extent for the holding of the closure of the port (see for example Bell-Berti |1|). Even if we consider that Halle's model |11| describing the velar functioning is far from being adequate, we agree with his suggestion that "the distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems."

For subject DA, an increase in the acoustical duration was not accompanied by a concomitant lengthening of the electromyographical activity. Whereas the durations were mostly shorter for the normal nonsense words on the acoustical level, on the contrary, they were systematically higher when considering the activity of LP and OOI.

It seems thus that a greater duration is not always straightforwardly correlated with a higher "force of articulation". The datas obtained from speaker JFB brought some support: here it is true that both the acoustical duration and the electromyographical activity of PPH increased from normal voice to shouted voice and finally to whispered voice. However, in both cases, significant differences were found between mean values for whispered vs. normal voice and for shouted vs. normal voice, but never for shouted vs. whispered voice. For example, for the nonsense words with |p t k|, the activity of PPH varied as follows (durations in msec.): normal voice: 1287-1309; shouted voice: 1594-1609; whispered voice: 1644-1729. The superior and inferior limits were separated by 285 msec. for normal voice vs. shouted voice, but by only 35 msec for shouted vs. whispered voice. With the |RVRVRV| items, the differences were 213 and 15 msec.

It must be added that the activity of PPH was remarkably similar for the normal and whispered utterances: the signal was poor and of a very limited amplitude; the shouted items were characterized by a much richer pattern.

This phenomenon underlines again the separation of the levels and suggests that duration is highly conditioned by the constraints inherent to the temporal programming of the sequence. Furthermore, the values of the C.V. were smaller for shouted voice and, to a lesser degree, for whispered voice: it could be that the speaker was "obliged" to reconsider partly his program, and to reduce to a minimum the area of variability.

Total duration of PPH (in msec). Subject JFB

	Normal		Shouted		Whisp.	
	$\bar{X}$	C.V.	$\bar{X}$	C.V.	$\bar{X}$	C.V.
|p	1309	16.83	1609	4.62	1670	11.78
t	1294	13.75	1594	5.10	1644	8.27
k	1287	16.47	1604	6.04	1729	13.57
R|	1458	12.06	1671	6.51	1686	7.49

|p	Shouted vs. Normal	ddl: 16 $p<0.05$
	Shouted vs. Whisp.	ddl: 16 NS
	Whisp. vs. Normal	ddl: 18 $p<0.01$
t	Shouted vs Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 16 NS
	Whisp. vs. Normal	ddl: 20 $p<0.001$
k	Shouted vs. Normal	ddl: 16 $p<0.05$
	Shouted vs. Whisp.	ddl: 26 NS
	Whisp. vs. Normal	ddl: 28 $p<0.001$
R|	Shouted vs. Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 16 NS
	Whisp. vs. Normal	ddl: 20 $p<0.01$.

TABLE 2

For DA, the latency times of implementing of LP were shorter in shouted and whispered voice, in comparison with normal voice (12 comparisons out of 15 were significant). For DA and JFB, OOI varied precisely in the same manner, even if all the comparisons did not reach the significant threshold of $p < 0.05$. As could be predicted on the basis of the behaviour of the total durations, the modifications in the latency times of PPH were scarcely noticeable although they followed the same pattern.

It can be concluded that:

(a) A stronger articulatory energy does not necessarily manifest itself through an earlier implementing of muscular activity.

(b) The shouted and whispered utterances can probably be joined together under the same head.

Latency time of LP (in msec). Subject DA

	Normal		Shouted		Whisp.	
	$\bar{X}$	C.V.	$\bar{X}$	C.V.	$\bar{X}$	C.V.
|p	629	24.	405	12.75	370	28.51
t	630	21.91	393	9.98	373	28.90
k	592	23.29	399	13.90	334	22.72
g	717	24.40	491	15.71	336	13.83
R|	476	33.89	296	21.56	357	35.98

|p	Shouted vs. Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 18 NS
	Whisp. vs. Normal	ddl: 18 $p<0.001$
t	Shouted vs. Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 18 NS
	Whisp. vs. Normal	ddl: 18 $p<0.001$
k	Shouted vs. Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 17 $p<0.05$
	Whisp. vs. Normal	ddl: 17 $p<0.001$
g	Shouted vs. Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 18 $p<0.001$
	Whisp. vs. Normal	ddl: 17 $p<0.05$
R|	Shouted vs. Normal	ddl: 18 $p<0.05$
	Shouted vs. Whisp.	ddl: 18 NS
	Whisp. vs. Normal	ddl: 18 $0.10>p>0.05$

TABLE 3

Instead of assuming that in uncommon circumstances, the latency time essentially reflects the global programming of the nonsense word, as it seems to be the case under normal conditions, we suggest that the encoding system works within a shorter temporal window (see also for an acoustical study of French CVC syllables: Rostolland et al. |12|). In our interpretation, the initial latency partly expresses the read-out time of generative encoding rules which are specific to the language in question. However, these rules are dependent on a rhythmic and accentual frame which is modified by the constraints inherent to the shouted and whispered phonation.

CONCLUSION

"Extrinsic variability" constitutes an essential part of a model of speech production, since the encoding modalities are conditioned by the constraints exerted on the muscular (sub-)system(s) and on the articulatory organs. Our conclusions allow to extend the notion of "syllabic segregation" applied by Kent and Rosenbek |13| to "apraxia of speech", and by Kent |14|, to the acquisition of language by children. Close connections can also be established between our results and some works on "expressivity". Fónagy |15| studied the articulatory manifestations of hatred and anger. These sentiments were rendered in the same way in French and Hungarian, by a series of jerky movements (abrupt transitions). In a cineradiographic study, Flament |16| looked into the question of "stylistic emphasis" in French. He came to the conclusion that there was a marked individualization of the articulatory units, in comparison with a neutral context: the coarticulatory link between successive segments was strongly weakened.

It appears that our analysis can be integrated into a more comprehensive body of facts, regrouping a great number of pathological and exceptional conditions. Of course, the patterns will be different according to the severity of the disease or to the weight of the constraint.

All these productions have probably in common to provide the subject with feedback information which is particularly difficult to handle. It could be that the (normal) subjects "tend to achieve some kinethetic-tactile feedback by finding articulatory landmarks" as was proposed by Rothman |17| for deaf adult speakers. Therefore, it can be suggested that the instructions are being split up, instead of being issued in the form a "list".

However, it must be stressed that there is a great variety of possible "amendment procedures" (Glencross |18|) and, consequently, a high flexibility of the matching of various kinds of feedback with the contextual requirements.

ACKNOWLEDGEMENTS

This work was supported in part by the Université du Maine (Le Mans), and in part by the Unité 3 INSERM "Physiologie et pathologie cérébrales". We are very indebted to Mme C.Chevrie-Muller, to M.B.Maton, and to Professor G.F.Greiner for the scientific and logistic support they provided us in the carrying out of the experiments. We also wish to thank our colleague, Mme C.Guillerme, for her able assistance in the english translation of this paper.

REFERENCES

|1| Bonnot J-F.P. Contribution à l'étude phonétique et phonologique de l'organisation temporelle de l'activité électromyographique labiale et vélaire. Coarticulation et processus d'encodage moteur. Thèse de doctorat d'Etat, Strasbourg II, 1986, 2 volumes, 704 pages.

|2| Bonnot J-F.P., Chevrie-Muller C., Arabia-Guidet C., Maton B. and Greiner G.F. "Coarticulation and Motor Encoding in CVCVCV Nonsense Words" Speech Communication, 5, 1986, 83-95.

|3| Kent R.D. "The Segmental Organization of Speech" in P.F.Mac Neilage (ed), The Production of Speech, Springer Verlag, New York - Heidelberg, 1983, 57-89.

|4| Glencross D.J. "Temporal Organization in a Repetitive Speed Skill" Ergonomics, 16, 1973, 765-776.

|5| Lubker J.F. "Some Teleological Considerations of Velopharyngeal Function", in Daniloff R.G.(ed), Articulation Assessment and Treatment Issues, College-Hill Press, 1983, 179-193.

|6| Eek A. "Articulation of the Estonian Sonorant Consonants I. |n| and |l|", Eesti NSV Teaduste Akadeemia Toimetised 19, Koïde, Uhiskonnateadused, 1, 1970, Institute of Language and Literature, Estonia, USSR.

|7| Bothorel A. Etude phonétique et phonologique du breton parlé à Argol (Finistère-Sud). Thèse de doctorat d'Etat, Strasbourg II, 1978. Atelier national de reproduction des thèses, Lille III, diffusion Breizh, Spezed, 1982, 514 pages.

|8| Fritzell B. "The Velopharyngeal Muscles in Speech", Acta Otolaryngologica, Supplementum 250, Göteborg, 1969, 79 pages.

|9| Legent F., Perlemuter F. et Vandenbrouck C. Fosses nasales, pharynx, Cahiers d'anatomie, Masson, Paris, 1974.

|10| Bell-Berti F. "An Electromyographical Study of Velopharyngeal Function in Speech", Journal of Speech and Hearing Research, 19, 1976, 225-240.

|11| Halle M. "On Distinctive Features and Their Articulatory Implementation" Natural Language and Linguistic Theory, 1, 1983, 91-105.

|12| Rostolland D., Parant C., Takahashi A. et Pandales E. "Durée vocalique intrinsèque et co-intrinsèque en français: contraintes physiologiques et variations temporelles dans des syllabes CVC", Actes des 14e Journées d'Etudes sur la Parole, Paris GALF et E.N.S.T., 1985, 179-182.

|13| Kent R.D. and Rosenbek J.C. "Prosodic Disturbances and Neurologic Lesion" Brain and Language, 15, 1982, 259-291.

|14| Kent R.D. "Sensorimotor Aspects of Speech Developments" in Aslin R.N., Alberts J.R. and Peterson M.R. (eds), The Development of perception: Psychobiological Perspectives, Academic Press, New York - London, 1982, 161-189.

|15| Fónagy I. La vive voix. Essais de psycho-phonétique, Payot, Paris, 1983, 346 pages.

|16| Flament B. Recherches sur la mise en relief en français. Thèse de doctorat d'Etat, Strasbourg II, 2 volumes, 1984, 1175 pages.

|17| Rothman H.B. "A Spectrographic Investigation of Consonant-Vowel Transitions in the Speech of Deaf Adults" Journal of Phonetics, 4, 1976, 129-136.

|18| Glencross D.J. "Levels and Strategies of Response Organization" in Stelmach G.E. and Requin J. (eds), Tutorials in Motor Behavior, North-Holland Publishing Company, Amsterdam, 1980, 551-566.