

Production et perception des phénomènes prosodiques dans la parole : comment identifie-t-on l'insistance malgré la variabilité inter-locuteurs ?

D. DAHAN et J.-M. BERNARD*

Laboratoire de Psychologie Expérimentale, Université Paris V, URA 316 du CNRS, 28 rue Serpente, 75006 Paris, France

** Groupe Mathématiques et Psychologie, Université Paris V, URA 1201 du CNRS, 12 rue Cujas, 75005 Paris, France*

abstract: This research deals with the production and perception of emphatic stress in French. The central issue is to determine which prosodic variations produced by several speakers allow the identification of emphasis by listeners. In the production experiment, four speakers were recorded while reading 24 sentences both with and without emphasis. Temporal, intonative, and intensive variations were measured and analyzed. Although the speakers adopted different speech rates, they produced a similar contrast in the speech rates between the stressed target and its contexts. Two of the four speakers modified the melodic contour on the target, i.e. by stressing the first syllable rather than the last one. The correlations between the percentage of emphasis perception and the prosodic variations suggest that the listeners may use different prosodic variations depending on the speaker and the cues he provides them with. Nevertheless, the intonative variations seem to be the most crucial cue for the perception of emphatic stress.

1. INTRODUCTION.

Nos recherches portent sur la production et la perception des aspects prosodiques de la parole, en particulier sur un phénomène stylistique dans le discours : l'accent d'emphase ou d'insistance. Cet accent rend plus proéminent le mot sur lequel il porte par rapport aux autres mots de l'énoncé. Cette proéminence focalise ainsi l'attention de l'auditeur sur ce mot, le processus de compréhension de l'énoncé pouvant être modifié [1, 2]. Le but de cette recherche est l'étude, en langue française, de la production d'emphase sur un mot dans un énoncé, ainsi que sa perception et reconnaissance par des auditeurs. Les modèles de perception de la parole supposent que l'auditeur effectue une comparaison entre les variations prosodiques de l'énoncé entendu et une référence pré-stockée des évolutions prosodiques habituellement associées à cet énoncé dans sa langue. Les écarts extraits de cette comparaison permettraient l'identification du phénomène prosodique réalisé [3]. Nous cherchons à préciser les variations que réalisent des locuteurs lors de la production d'insistance, et le rôle de ces variations dans l'identification de l'insistance par des auditeurs.

L'allongement de durée du mot accentué ainsi que l'augmentation de sa fréquence sont des variations fréquemment observées [4, 5, 6]. L'importance des variations de F0 dans la perception d'une accentuation a été également mise en évidence [7, 8, 9]. Cependant, les résultats des analyses acoustiques sont parfois contradictoires: certaines recherches ont mis en évidence des variations de durée limitées au mot accentué [5, 6], alors que d'autres [10] ont mis en évidence des allongements de durée sur les segments au voisinage du mot accentué, l'importance de l'allongement étant supposée inversement proportionnelle à la distance entre le segment mesuré et le mot cible. Nous faisons l'hypothèse que cette divergence témoigne de l'existence de différentes stratégies selon les locuteurs.

Dans cette recherche, nous nous sommes focalisés sur cette variabilité inter-locuteurs, pour tenter de l'explicitier et de voir comment l'auditeur la traite. En quoi consiste la variabilité des réalisations ? Comment l'emphase peut-elle être identifiée malgré la diversité des réalisations prosodiques possibles ? Nous pouvons supposer que l'auditeur extrait un élément commun à toutes ces réalisations, la variation d'un ou de plusieurs paramètres prosodiques, qui lui indiquerait la présence d'une insistance: ou bien que l'auditeur est capable d'utiliser les variations d'un indice ou d'un autre, selon leur disponibilité.

2. METHODE.

2.1. Expérience 1 : Production.

Nous avons enregistré quatre locuteurs (deux hommes, deux femmes) lisant 24 phrases différentes, dans un premier temps sans emphase, dans un second temps en réalisant une emphase sur un mot cible mis en relief (en gras et souligné). Ces productions ont été digitalisées puis analysées grâce à des éditeurs de signaux de parole (UNICE-VECSYS sur PC, et OEDIPE, sur NeXT).

Nous avons mesuré et calculé la durée syllabique moyenne, témoignant de la **vitesse de parole**, sur le mot-cible et sur la portion de parole précédant et suivant la cible. Nous avons mesuré l'**intensité** maximale sur la cible, et sur les 200 ms précédant et suivant la cible, ce qui correspond à la durée d'une syllabe environ, nous permettant d'évaluer la proéminence de la cible par rapport à son contexte immédiat. Enfin, nous avons mesuré les valeurs maximales de **F0** sur les voyelles de chaque syllabe du mot cible (toujours bisyllabique), ainsi que la valeur maximale de F0 sur les 200 ms précédant et suivant la cible. Nous avons ainsi analysé la proéminence intonative de la cible par rapport à son contexte immédiat, ainsi que les éventuelles variations du pattern accentuel du mot lors de la production d'emphase (la comparaison des valeurs de F0 sur les premières et deuxième syllabes nous permettant de déterminer la syllabe accentuée).

2.2. Expérience 2 : Perception.

Pour étudier l'identification des emphases réalisées, nous avons présenté les énoncés produits à 253 auditeurs, chaque auditeur n'entendant qu'une seule version de chacune des 24 phrases. Leur tâche était de décider si une insistance était produite dans l'énoncé entendu, et si oui, de préciser sur quel mot de l'énoncé. Nous avons calculé le pourcentage de perception d'emphase pour chaque phrase et chaque version. Si les volontés d'insistance des locuteurs étaient "bien" interprétées par les auditeurs, le pourcentage de perception d'emphase devait être faible pour les énoncés en condition neutre et élevé en condition d'emphase.

3. RESULTATS.

3.1. Production.

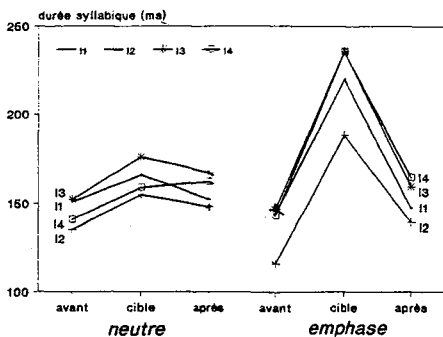


Figure 1: Durées syllabiques

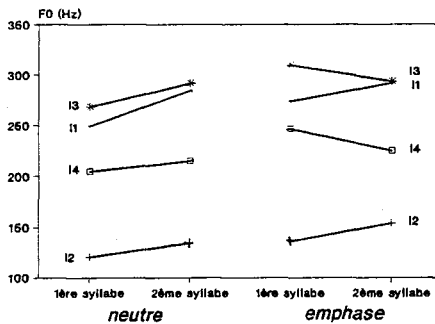


Figure 2: Contours mélodiques

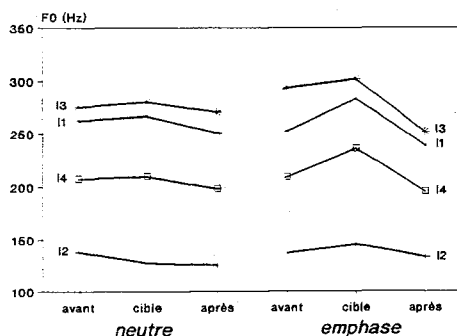


Figure 3: Variations intonatives

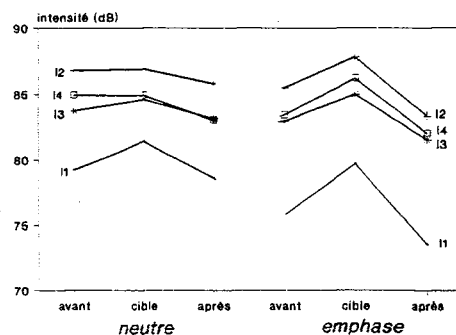


Figure 4: Variations intensives.

La figure 1 présente les variations de durée syllabique pour chaque moment de la phrase (contexte précédant, cible, contexte suivant), pour chaque locuteur (de 11 à 14). la partie gauche présentant les résultats en condition neutre, la partie de droite, en condition avec emphase. Les locuteurs 1, 3, et 4 ralentissent fortement leur élocution sur la cible, sans la modifier sur les contextes. En revanche, le locuteur 2 accélère sa vitesse de parole avant et après la cible, et allonge l'élocution de la cible elle-même, mais dans une moindre mesure que les autres locuteurs. Si on analyse le *contraste* que les locuteurs réalisent entre leur élocution sur la cible et autour d'elle, les 4 locuteurs créent un contraste équivalent -les courbes sont parallèles en condition d'emphase, mais en adoptant des stratégies différentes. La figure 2 présente les contours mélodiques des cibles. En condition neutre, les contours sont montants, l'accent tonique portant le plus souvent sur la dernière syllabe du mot en français. En condition avec emphase, seuls les locuteurs 3 et 4 *modifient* le pattern accentuel des mots en faisant porter l'accent d'emphase sur la première syllabe. Enfin, les résultats concernant les variations intensives et intonatives (figures 3 et 4) ont suggéré que les locuteurs augmentent la prééminence de la cible accentuée par rapport à ses contextes immédiats.

3.2. Perception.

Tableau 1: Pourcentage de perception d'emphase

	a1	a2
11	27.5	67.8
12	7.0	60.4
13	32.0	78.9
14	11.9	88.6

Le tableau 1 présente le pourcentage de perception d'emphase pour chaque locuteur (de 11 à 14) et chaque condition (a1:neutre, a2:emphase). Les valeurs montrent que les auditeurs ont réussi à identifier une emphase lorsque telle était l'intention des locuteurs. Le locuteur 4 est celui dont les intentions ont été les mieux interprétées, particulièrement dans la condition d'emphase (près de 90 % d'identification juste).

Le tableau 2 présente, pour chaque locuteur, les valeurs de corrélations entre les variations de chaque indice et celles de la variable "perception d'emphase", uniquement pour la condition d'emphase. En ce qui concerne la durée syllabique, la fréquence et l'intensité, nous avons pris comme indice le contraste entre la valeur sur la cible et les valeurs regroupées des contextes; pour le contour, nous avons utilisé la différence des valeurs de F0 entre la première et la seconde syllabe. De ces calculs de corrélations, il est bien sûr impossible d'affirmer que certaines variations *déterminent* l'identification d'une emphase, mais ces résultats peuvent fournir des éléments à partir desquels une investigation contrôlée pourrait être élaborée.

Tableau 2: Valeurs des corrélations (r) pour chaque indice et chaque locuteur (ddl: 22).

	durée syllabique	fréquence	intensité	contour
11	0.007 NS	0.365 NS	0.280 NS	0.435 p < .05
12	0.151 NS	0.425 p < .05	0.072 NS	0.120 NS
13	0.430 p < .05	0.459 p < .05	-0.020 NS	0.267 NS
14	0.225 NS	0.639 p < .001	0.223 NS	0.120 NS

Les résultats montrent que certaines variations *produites* sont très peu corrélées avec l'indice de perception. En particulier, l'intensité est un indice qui varie avec la production d'insistance, mais il semble déterminer peu son identification. En revanche, la proéminence intonative du mot cible par rapport à son contexte semble jouer un rôle primordial : le locuteur 1 mis à part, ses variations sont corrélées significativement avec la perception d'une emphase par les auditeurs. Les résultats indiquent également une utilisation des variations d'indices différentes selon les locuteurs.

4. CONCLUSION.

La réalisation d'un accent d'emphase implique certaines variations prosodiques, plus ou moins utilisées selon les locuteurs. Certaines ne semblent pas jouer de rôle dans la reconnaissance de l'emphase, d'autres au contraire semblent avoir un rôle crucial. L'auditeur ne semble pas extraire un élément commun à toutes ces réalisations pour l'utiliser dans sa reconnaissance de l'emphase, quoique la proéminence intonative de la cible par rapport à son contexte pourrait jouer ce rôle. Il est cependant plus vraisemblable que l'auditeur utilise certaines variations plutôt que d'autres selon les indices fournis par les locuteurs. D'autres expériences, contrôlant systématiquement les variations de certains indices prosodiques, en particulier le pattern accentuel du mot cible et les variations de la vitesse de parole, pourront permettre la mise en évidence de leurs rôles respectifs dans la perception et la reconnaissance de l'emphase.

5. BIBLIOGRAPHIE.

- [1] Kimelman M.D.Z. and MacNeil M.R., *J. Speech Hear. Res.* **30** (1987) 295-300.
- [2] Bean C., Folkins J.W. and Cooper, W.E., *J. Speech Hear. Res.* **32** (1989) 707-712.
- [3] Calliope, La parole et son traitement automatique (Masson, Paris, 1989).
- [4] Séguinot A., *Studia Phonetica* **12** (1977) 1-58.
- [5] Cooper W.E., Eady S.J. and Mueller P.R., *J. Acoust. Soc. Am.* **77** (1985) 2142-2156.
- [6] Eady S.J. and Cooper W.E., *J. Acoust. Soc. Am.* **80** (1986) 402-415.
- [7] Rigault A., "Rôle de la fréquence, de l'intensité et de la durée vocalique dans la perception de l'accent en français", Proceedings of the 4th International Congress of Phonetic Sciences, A. Sovijärvi and P. Aalto Eds, Helsinki 1961 (Mouton, The Hague, 1962) pp 735-748.
- [8] Terken J., *J. Acoust. Soc. Am.* **89** (1991) 1756-1776.
- [9] Hasegawa Y. and Hata K., *Language and Speech* **35** (1992) 87-98.
- [10] Weismer G. and Ingrisano D., *J. Speech Hear. Res.* **22** (1979) 516-533.