

**VISUALISATION ET ANALYSE
DE PRODUCTIONS VOCALES
(Applications en analyse linguistique et
dans l'enseignement du français)**

Jean-Louis MALANDAIN

Présentation du module ORPHON, conçu et réalisé au Laboratoire d'Informatique du Conservatoire National des Arts et Métiers par Frédéric SOURDILLAT (CNAM, 292 rue Saint-Martin, 75141 PARIS CEDEX 03 Tél. 42 71 24 14 poste 678).

Maquette en cours d'expérimentation, dérivée du module ORPHY (cf. Bulletin de l'EPI N° 32 de décembre 1983, pp 89-93), en cours d'expérimentation dans le cadre des travaux du groupe EVARISTE (cf. Actes des Sèmes journées de synthèse et d'études, 21-22 janv. 87, disponibles à la librairie du CNAM).

Plaidoyer pour le développement et l'utilisation du module ORPHON rédigé par Jean-Louis Malandain, chargé d'études au Bureau pour l'Enseignement de la Langue et de la Civilisation du Français (BELC, 8 rue Malebranche. 75005 PARIS. Tél 43 54 41 51).

LA PAROLE, OBJET D'UNE ANALYSE PHYSIQUE

Donner une équivalence graphique de phénomènes sonores n'est pas une nouveauté pour les physiciens, acousticiens et autres spécialistes : les vibrations sonores, transmises mécaniquement puis traduites en intensités électriques, sont analysées grâce à des appareils comme l'oscilloscope ou le sonographe dans les laboratoires. Mais leur prix élevé en rend l'usage pratiquement impossible ailleurs.

Ce qui est nouveau, grâce à la micro-informatique et aux travaux des spécialistes pour débusquer de nouvelles applications, c'est la mise à la portée des enseignants d'un appareillage simple et bon marché capable d'émuler (comme on dit maintenant pour "faire la même chose que") les instruments des spécialistes. Ce n'est pas seulement l'appareil qui entre

LE BULLETIN DE L'EPI N° 46 VISUALISATION ET ANALYSE VOCALES

dans la classe, c'est aussi un état d'esprit qui "émule" l'attitude scientifique d'observation des phénomènes : la vraie nouveauté est peut-être là, s'agissant de l'enseignement des langues et du français en particulier.

Pour l'utilisateur, le module ORPHON est une "boîte noire" au sens métaphorique du terme ; en réalité un coffret blanc de 8 x 6 x 15 cm qui se branche à l'arrière d'un micro-ordinateur Thomson. Dans sa configuration actuelle, ce module permet de capter, à la façon d'un vu-mètre, les variations d'intensité venant d'un enregistrement lu par le LEP (Lecteur Enregistreur de Programme) que commandent les instructions MOTOR ON et OFF ¹.

Tout l'intérêt de cette interface et de la routine en langage machine qui l'accompagne est de permettre l'inscription des variations et d'en garder la trace sur l'écran. Trois réglages sont prévus pour tenir compte de la durée de la séquence sonore analysée et du bruit de fond.

Il s'agit donc d'une utilisation particulière d'un capteur qui n'a pas la prétention de supplanter les appareils de mesure spécifiques : pour être précis, il faudrait parler de « simulation » un autre terme à la mode pour "faire comme si..."

UN OUTIL POUR L'ANALYSE LINGUISTIQUE

Dans toutes les situations où des enseignants sont amenés à observer d'un peu près les productions orales, pour une recherche ou en fonction d'objectifs didactiques ², le module ORPHON apporte une solution commode pour la mise en évidence des phénomènes les plus marquants de l'ordre du supra-segmental (tous les phénomènes sonores qui accompagnent la suite des unités lexicales).

Du fait que leur perception et leur intégration dans la compréhension du message sonore sont immédiates, ces faits de prosodie restent le plus souvent inconscients. D'où la difficulté du repérage à l'oreille de phénomènes comme la pause, le rythme ou l'intonation. Elle vient autant de la maladresse à les noter "à la volée" (même avec un oscilloscope) que

1 cf. *Bulletin de l'EPI* n° 41, mars 1986, p. 163-164.

2 Sur ces problèmes, consulter : *Pour enseigner l'oral*, J.MOUCHON & F.FILLOL CEDIC PARIS 1980.

Décrire et découper la parole, M. LEBRE-PEYTARD & J-L MALANDAIN Fascicules 1 et 2, EELC PARIS 1982

des déviations subjectives que l'observateur le plus averti ne maîtrise pas.

Même si les résultats obtenus avec le module ORPRON n'ont pas la finesse de ceux des appareils de laboratoire, le fait de pouvoir figer une représentation des phénomènes est un progrès indéniable.

Pour des étudiants Français ou Étrangers amenés à se spécialiser dans le domaine de l'oral (par exemple de futurs professeurs de français), ce module est un remarquable instrument d'initiation à l'analyse.

QUELS REPÉRAGES PEUT-ON FAIRE ?

Les deux priorités demandées au physicien-informaticien par le professeur de français confronté à l'analyse de la parole ont été, pour un document d'environ une minute :

a) la visualisation et la sonorisation des pauses, dont l'occurrence et la durée ont des fonctions significatives dans le découpage et l'organisation de l'énoncé ³.

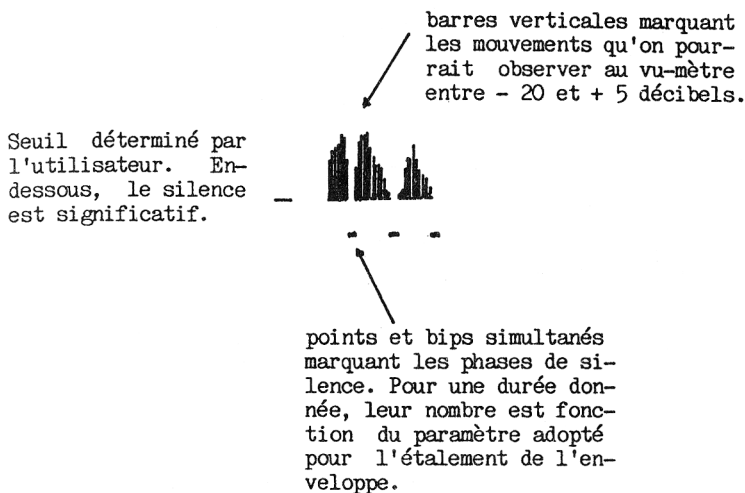
Lorsque l'émission vocale passe au-dessous d'un seuil déterminé (pause volontaire et/ou respiration), un ou plusieurs points s'inscrivent en bas de l'écran en même temps qu'un ou plusieurs bips se font entendre, selon la durée de la pause.

On peut ainsi lire une transcription et y porter les pauses sans avoir à lever les yeux ni à s'interroger sur la réalité de la pause puisque le seuil choisi reste constant pour un même document ; la valeur relative des pauses est immédiatement repérable en comptant les bips ou, mieux, les points inscrits à l'écran. Les linguistes habitués à ce type d'analyse apprécieront.

b) l'intensité de l'émission vocale dont les sommets sont des éléments structurant l'énoncé, intentionnels (accents d'insistance) ou non.

Chaque variation produit l'inscription de lignes montantes qui remplissent l'écran de gauche à droite et dont l'enveloppe constitue l'allure générale du document.

³ En théorie, la valeur de l'énergie sonore pourrait être captée plusieurs centaines de fois par seconde, mais le transfert jusqu'au programme d'affiche produit des retards ; de plus, pour l'usage qui en est fait, ces données ne seraient pas pertinentes.



Pour des productions vocales très brèves et dans des situations didactiques particulières, la visualisation peut porter sur quelques mots, voire une syllabe ou un son. Dans ce cas, grâce à l'étalement des repères, on voit apparaître des traces qui rendent compte de différences significatives au plan phonétique. Dans la configuration actuelle, l'utilisation du microphone "en direct" est possible : au lieu d'enregistrer au préalable les documents à analyser, on obtient une visualisation en temps réel. On voit bien le parti qu'on pourrait en tirer pour la rééducation en orthophonie ou la correction de la prononciation dans l'enseignement des langues.

POUR QUELLES EXPLOITATIONS DIDACTIQUES ?

Avant d'en décrire quelques unes et sans empiéter sur les découvertes que ne manqueront pas de faire les pédagogues munis de cet outil, il est possible de dégager le fondement de l'apport en didactique : transférer des phénomènes sonores dans le champ du visuel et donner une réalité concrète et tangible à des phénomènes difficiles à décrire ou dont on n'a pas une conscience claire. C'est souvent le cas pour le rythme et la prosodie l'élève n'en perçoit pas toujours les éléments significatifs et n'a pas les moyens de vérifier s'il les a produits. La visualisation et la persistance des traces permet l'observation et l'interaction. Ainsi, l'écart par rapport à une norme n'est pas seulement énoncé par le professeur ("Ce n'est pas ça !") mais constaté par l'élève qui "voit" la différence et peut localiser la non-conformité.

A titre indicatif, on peut suggérer quelques situations où l'utilisation du module ORPHON apporterait une aide appréciable, pour des publics et à des niveaux différents :

1) Jeunes élèves en phase d'apprentissage :

- Entraînement à la lecture expressive, à la récitation ou à la diction par comparaison avec un modèle de référence à l'écran.

2) Situation de rééducation (orthophonie) ou d'apprentissage des langues (correction phonétique) :

- aide à un meilleur contrôle des sons émis, de l'intonation et du rythme.

3) Apprentissage et formation continue :

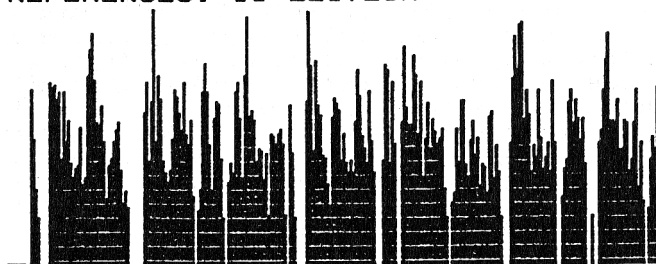
- Aide à la maîtrise de la voix et de l'élocution. Analyse fine des faits de prosodie et d'organisation de la parole par le repérage des éléments significatifs (pauses, accents d'insistance...).

4) Niveau universitaire : perfectionnement ou initiation à la recherche.

- Étude des relations entre le segmental (chaîne des unités lexicales) et le suprasegmental (faits de prosodie). Aide à la transcription exhaustive des productions orales et à l'étalonnage des caractéristiques du discours (par exemple la valeur des pauses).

EXEMPLE DE REPRÉSENTATION A L'ÉCRAN

REFERENCES: 18 LECTEUR



SEUIL, PAS 35, .35

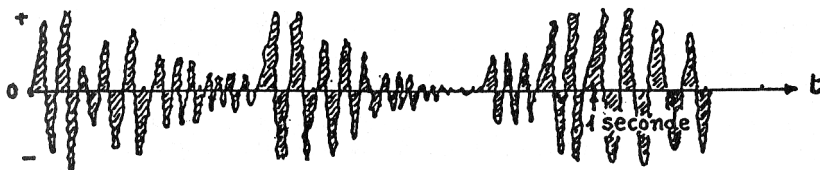
C'est grâce à cette représentation, qui traduit les données physiques recueillies par le module (intensité, courbe mélodique, silence ou pause...), qu'on peut envisager un travail d'analyse ou de comparaison entre des énoncés, mettre en évidence des phénomènes rythmiques et le fonctionnement des pauses dans la structuration des messages oraux ou la détermination des registres de langue.

ET COMMENT ÇA MARCHE ?

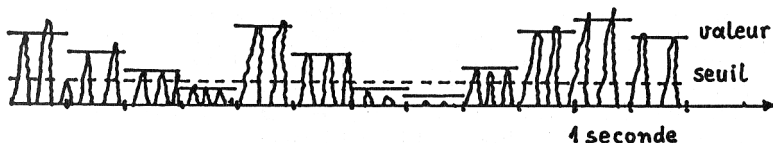
Pour comprendre, il faut considérer une production vocale comme une émission sonore et, pour simplifier encore, admettre que le son est une vibration. Bien sûr, rien de tout cela n'est tout à fait vrai ! C'est du moins le schéma qui permet la reproduction du son grâce au microphone et au haut-parleur.

A. En fixant un stilet à la membrane d'un haut-parleur, on peut graver la trace des oscillations sur un tambour mobile ; cette trace rend compte des mouvements dont le cumul représente l'énergie dépensée.

B. Si on considère que la membrane se déplace également de part et d'autre de sa position de repos et qu'elle engendre un signal électrique, l'énergie sonore peut être représentée par les surfaces en grisé sur la figure :



C. L'énergie "positive" étant équivalente à l'énergie "négative", on peut supprimer la trame négative. Le signal ainsi "tronqué" est passé à travers un filtre de coupure "passe bas" ; on obtient alors une sorte de lissage qui traduit l'énergie moyenne à un instant donné. Le filtre ralentit les réactions aux changements ; il conserve l'énergie qui ne retombe à zéro que pour des périodes significatives sans vibration. Il est réglé par un potentiomètre à la disposition de l'utilisateur.



D. L'énergie électrique est "découpée" en 255 tranches de 0,02 volts. C'est l'opération d'échantillonnage qui permet de transmettre, selon la tranche atteinte par le signal à un instant t , une valeur comprise entre 0 et 255 (contenu d'un octet).

E. Par ailleurs, un seuil minimal du signal filtré en B est choisi dans le logiciel pour déterminer la valeur significative du silence dans l'environnement sonore du document.

P. L'utilisateur fixe le pas de la boucle qui détermine la durée pendant laquelle le document sonore sera analysé à raison d'environ une saisie par dixième de seconde ³.

G. La valeur recueillie conditionne la représentation graphique à l'écran : un point blanc et un bip en deçà du seuil, une barre verticale colorée proportionnelle à la valeur au-dessus du seuil.

ON ENTEND TOUT, MAIS QU'EST-CE QU'ON VOIT AU JUSTE ?

Dans l'écoute "naturelle" (c'est-à-dire sans effort d'analyse), nous percevons l'émission de voix comme un continuum ; pourtant, ce flux sonore est une succession d'éléments discontinus où notre oreille puise les repères nécessaires à la compréhension : essentiellement les phonèmes qui sont utilisés pour constituer les mots, séquences phoniques auxquelles l'auditeur attribue un sens en référence à un " dictionnaire cérébral" ; à un niveau plus conscient, apparaissent les ruptures (pauses et silences) et les variations mélodiques elles-mêmes porteuses de signification :

"Tu prends le sac !" ≠ "Tu prends le sac ?"

Il ne faut pas penser cependant que tous ces éléments sont discontinus dans leur réalité physique : le passage de u à ou est une variation de timbre évidente pour un locuteur français car elle porte une différence fonctionnelle bien répertoriée :

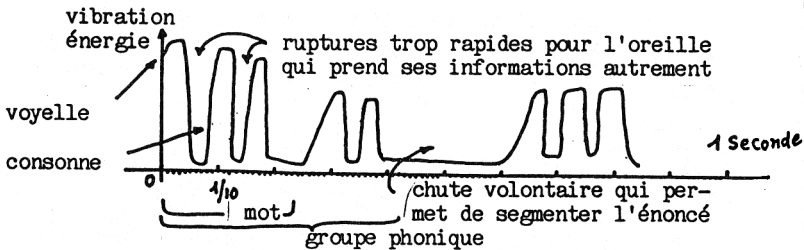
"Tu vas mal." ≠ "Tout va mal."

mais difficile à détecter ou ignorée dans une autre langue.

Il serait donc illusoire d'attendre que la machine fasse pour nous des différences qui sont dans notre tête (ou dans notre culture) : dans l'état actuel de la technique et du matériel utilisé, nous sommes encore imbattables pour repérer les éléments "discrets" du langage.

Ce que peut faire une machine, c'est traduire des événements tels que plus d'énergie, moins d'énergie, pas d'énergie développée par le flux sonore, et cela n fois par seconde. Or, dans la parole, la durée d'un phonème, en gros une voyelle ou une consonne, est de l'ordre de 50 millièmes de seconde. Comme le capteur ORPHON fait une saisie tous les 100 millièmes de seconde, il ne rendra pas compte de la succession voyelle-consonne qui, au "microscope sonore" (!) se traduirait par vibration-absence de vibration (ou moindre vibration pour les consonnes sonores ou nasales). ORPHON n'est pas un analyseur vocal et n'a pas

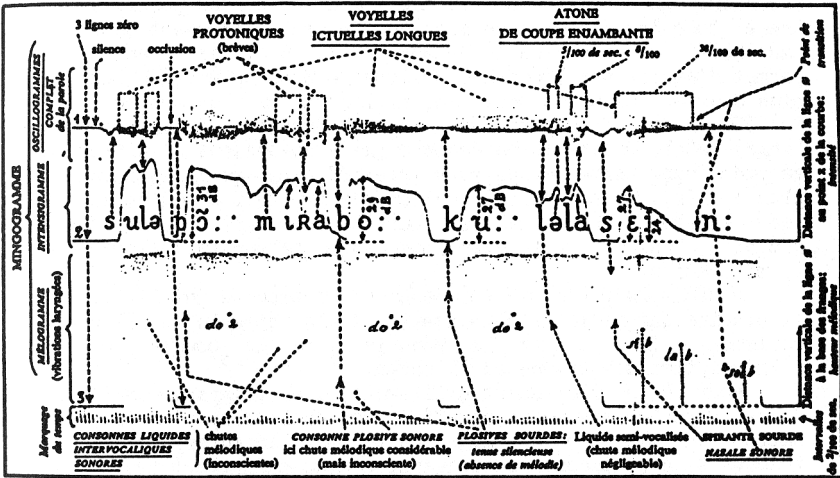
pour objectif la reconnaissance de la parole ! Disons qu'il a les mêmes capacités que l'oreille à saisir les variations d'intensité et les ruptures qui accèdent au niveau de la conscience de l'auditeur, d'où son intérêt pédagogique dans la représentation des macro-structures : il fixe des phénomènes qui peuvent devenir conscients, à plus forte raison les découpages et les variations qui sont le fait d'actes volontaires ou de nécessités physiologiques qui se traduisent par une tension articulatoire, une montée mélodique, un relâchement etc.



Ainsi, la pause est un arrêt des vibrations ou de l'énergie sonore perçu comme tel si sa durée dépasse 100 millièmes de seconde et jusqu'à environ deux ou trois secondes. En deçà, c'est l'analyse au microscope des différences voyelle-consonne puis des ruptures éventuelles entre les mots ou les groupes phoniques, au delà c'est le silence.

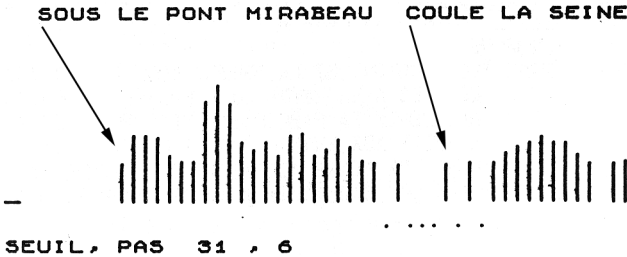
C'est ce qui apparaît nettement dans ce mingogramme qui présente plusieurs tracés simultanés et complémentaires : oscillogrammes (amplitude et fréquence), intensigramme (résultante "redressée"), mélogramme (hauteur mélodique), constante de temps (unités de 20 millièmes de seconde) ⁴

⁴ Extrait de Henri MORIER : "Dictionnaire de poétique et de rhétorique", P.U.F., Paris, 1961 et 1981.



Sous le pont Mirabeau coule la Seine

On peut le comparer aux performances d'ORPHON pour le même énoncé :



QUELQUES MOTS DU PROGRAMME DE REPRÉSENTATION GRAPHIQUE

Le cœur du programme est une boucle de temporisation entre MOTOR ON et MOTOR OFF. C'est une boucle de travail dont le frein est la représentation graphique de la valeur transmise par le capteur. On utilise pour cela une bande de 300 pixels parmi les 320 disponibles sur la largeur de l'écran. Au pas de 1, la boucle fait 300 tours : à chaque tour, la valeur envoyée par le capteur est utilisée pour tracer un point et faire un bip (cas de la pause) ou tracer une barre de hauteur proportionnelle à l'énergie sonore ; à ce pas, la boucle atteint son terme après environ 30 secondes : on obtient ainsi le maximum de notations possible, une trace tous les dixièmes de seconde. On peut évidemment serrer le pas (0.9, 0.5 etc.) ou le desserrer (2, 5 etc.) ; c'est ce qui permet d'étaler sur l'écran

l'analyse de documents plus longs ou plus courts. Mais, de toute façon, l'écran ne peut contenir plus de 300 traces (donc perte de notations dans les documents plus longs) et le programme ne descend guère au-dessous du dixième de seconde (donc perte des valeurs intermédiaires pour les documents plus courts). On ne pourrait éviter ces limitations qu'en entrant les valeurs dans un tableau pour obtenir une représentation graphique "en différé"; cette fois, la perte serait de l'ordre de la motivation.

L'ossature présentée ci-dessous ne comporte pas la routine en binaire qui permet l'alimentation de la variable Y. Il s'agit seulement de montrer le mécanisme de la boucle et de l'inscription des traces sur l'écran.

Pour être facilement utilisable, le programme définitif devra comporter des menus et offrir des choix comme la représentation en barres ou en nuages de points, la superposition des traces de deux documents successifs (avec changement de couleur pour permettre les comparaisons), l'effacement pour analyser un document continu par saccades de 30 secondes ou le choix par la durée de la séquence analysée etc.

```

10  MOTORON:P = 1
20  FOR X = 10 TO 300 STEP P
30  EXEC Z:Y = PEEK(ZB) REM Valeur venant du capteur.
40  IF Y<SEUIL THEN PRINT CHR$(7):PSET(X,185),7:GOTO 60
50  LINE(X,180)-(X,180*(1-Y/255)),2 REM Barre montante
60  NEXT X
70  MOTOROFF
80  PRINT "CONCLUSION: SI LE PROJET VOUS PARAIT INTERESSANT,
        ECRIVEZ A LA REVUE OU A Frédéric SOURDILLAT POUR
        ACCELERER LA FAERICATION D'ORPHON ET FAIRE TOMBER
        LE PRIX AUX ENVIRONS DE 1500 Francs."
90  PRINT "MERCI DE VOTRE ATTENTION !":PRINT "FIN"
100 END REM (un peu d'anglais pour réaliser le mot français !

```

Jean-Louis MALANDAIN

07-04-1987