

Universities of Aarhus and Copenhagen

October 10-12, 1996

Roman JAKOBSON
Centennial Symposium

MODÈLES MORPHODYNAMIQUES
de
CATÉGORISATIONS PHONÉTIQUES

Jean PETITOT
EHESS, 54 Bd. Raspail, 75 006 Paris
petitot@poly.polytechnique.fr

INTRODUCTION

A l'occasion de ce symposium du centenaire, j'aimerais revenir sur les modèles morphodynamiques de perception phonétique que j'avais présentés au début des années 80 dans *Les Catastrophes de la Parole*, ouvrage dont le sous-titre était "de Roman Jakobson à René Thom" et qui, en hommage à *La Charpente phonique du langage*, était précisément dédié à la mémoire de Roman Jakobson.

Le problème motivant ces recherches est le suivant :

Comment comprendre sur des bases phonétiques — et donc acoustiques, psychophysiques et cognitives — les phonèmes, au sens linguistique et structuraliste du terme, c'est-à-dire comme unités fonctionnelles abstraites, comme entités purement relationnelles et discriminantes.

Ce problème théorique est extrêmement délicat car si l'on vise une théorie phonétique intégrée, il faut alors arriver à concilier deux conceptualités théoriques :

- côté "forme" (phonologie) : une conceptualité *structurale* à la Saussure-Jakobson où il est question d'identité relationnelle, de différence, de détermination réciproque, de valeur positionnelle, etc.
- côté "substance" (phonétique) : une conceptualité *psycho-physique* où il est question de spectres, de formants, d'indices acoustiques, de déformation de spectres, de contrôle des spectres par les paramètres articulatoires, de catégorisation, etc.

Il s'agit là de l'aspect phonétique d'un problème plus général qui est commun au structuralisme et au cognitivisme.

Le fonctionnalisme cognitiviste (disons à la Fodor-Pylyshyn) reprend en effet

certaines thèses du structuralisme formaliste (disons à la Hjelmslev), et en particulier la thèse que la forme est indépendante de la substance et seulement "implémentée" dans la matière.

Ce formalisme structuraliste-fonctionnaliste n'est pas tenable car la forme est au contraire le phénomène de l'organisation de la substance. Elle est certes descriptible symboliquement, mais, en ce qui concerne sa *genèse*, elle *émerge* de l'organisation dynamique de la matière sous-jacente. Jakobson en était très conscient. Je renvoie pour cela aux remarquables analyses de Elmar Holenstein.

En sciences cognitives l'opposition substance/forme s'identifie à celle entre le physique et le symbolique. L'un des problèmes centraux est de comprendre comment l'on peut passer de continuums physiques (par exemple acoustiques) à des entités symboliques discrètes symboliquement manipulables pouvant être le support d'un code (par exemple phonologique).

Le problème est avant tout théorique et mathématique.

THESE : *il est impossible d'élaborer une théorie mathématique plausible des rapports entre le continu physique et le discret symbolique en sciences cognitives sans passer par la médiation de modèles morphodynamiques.*

A la suite de Husserl et de Thom, nous appelons morphologique les formes définies par des ensembles de discontinuités qualitatives sur des espaces substrats. Et nous appelons modèles morphodynamiques les modèles dynamiques expliquant la genèse et la géométrie de telles discontinuités qualitatives. Nous voudrions établir le lien entre ces modèles et le fait bien connu que, en phonétique, la médiation fondamentale entre le physique (acoustique) et le symbolique (phonologique) est assurée par *les processus de catégorisation*.

I. RAPPELS SUR LES *PRELIMINARIES*

Je rappelle très brièvement le problème des traits distinctifs phonologiques dans les célèbres *Preliminaries to Speech Analysis* de Roman Jakobson, Gunnar Fant, Morris Halle (MIT Press, 1952) et *La Charpente phonique du langage* de Roman Jakobson et Linda Waugh (Minuit, 1980). Je ne prends en considération ni les traits prosodiques comme l'accent, le rythme, la hauteur, la longueur, etc. ni les traits configurationnels assumant une fonction grammaticale comme la montée finale d'une phrase interrogative, ni les traits expressifs comme l'intonation.

Les phonèmes sont des types et leurs allophones ou variantes combinatoires sont des tokens. Ils sont définis comme faisceaux de traits distinctifs. Ceux-ci s'organisent en un système hiérarchisé et stratifié qui a été dégagé à partir d'une masse énorme de données empiriques sur les différentes langues.

Le système jakobsonien repose, on le sait, sur un principe fondamental de binarisme :

“the dichotomous scale is the pivotal principle of the linguistic structure. The code imposes it upon the sound”. (p. 9).

Cela correspond à une évaluation booléenne des phonèmes. On cherche un ensemble universel de traits $T_i, i = 1, \dots, n$ à valeurs dans l'espace $B = \{+, -, 0 = \text{neutre}\}$. Étant donné l'ensemble P des phonèmes d'une langue quelconque, on définit l'application d'évaluation $\Phi : P \rightarrow S = B^n$ de composantes $\Phi_i : P \rightarrow T_i$ par le fait que $\Phi_i(p) = +, -, 0$ suivant que le phonème $p \in P$ est positif, négatif ou neutre relativement à T_i . Cette application d'évaluation doit être injective et n doit être minimal.

Les *Preliminaries* distinguent, on le sait, deux types de traits distinctifs concernant la source et les résonateurs:

- Traits concernant la source
 - Vocalique / non-vocalique
 - Consonantique / non-consonantique
 - Continu / discontinu
 - Glottalisé / non-glottalisé
 - Strident / mat
 - Voisé / non voisé
- Traits concernant les résonateurs
 - Nasal / non-nasal
 - Compact/diffus
 - Tonalité :
 - grave / aigu
 - bémolisé / non-bémolisé
 - tendu / lâche.

II. LE PROBLÈME CENTRAL DE LA CATÉGORISATION

Le problème central de la phonétique oppose les approches “form-based” à celles “substance-based”. Comme le dit P. Ladefoged (1972) :

“One of the major difficulties (...) is in relating the essentially continuous nature of speech with the essentially discontinuous nature of linguistic description”.

Et D. Pisoni ajoute :

“This lack of correspondance between attributes of the acoustic signal and units of linguistic analysis has been, and still currently is, one of the most important and controversial issues in speech perception”.

La réponse aux problèmes posés par la compréhension des relations entre forme discrète articulée (code phonologique) et matière continue indifférenciée (flux audio-acoustique) est donnée par le concept de catégorisation.

Comme le disait Bertil Malmberg (*Manuel de Phonétique générale*, 1974), le phonème est

“une unité unique différenciatrice et sans qualités concrètes mais manifestée, dans la parole, par un allophone ayant des qualités physiques qui traduisent dans le monde des réalités physiques les qualités différentielles”.

De façon générale, une catégorisation consiste en la donnée d'un espace W (le support de la catégorisation : ici un espace d'indices acoustiques et/ou de paramètres articulatoires) décomposé en domaines (en catégories) par un système K d'hypersurfaces de codimension 1.

Le passage au discret consiste essentiellement à prendre les composantes connexes de $W-K$, à y considérer des valeurs centrales et à les étiqueter par des symboles. Le binarisme est donc localement et génériquement justifié et universel. En effet, localement, une hypersurface de codimension 1 partitionne l'espace support W en deux demi-espaces et dans toute direction transverse on a bien une opposition binaire +/-.

L'erreur principale du structuralisme combinatoire à la Jakobson est *dans le passage du local au global*. Il fait en effet l'hypothèse que catégorisation est du type “intersection transversale d'hyperplans” (comme les hyperplans de coordonnées dans un espace \mathbb{R}^n). Cela est évidemment totalement faux. Par exemple en dimension 2, on a génériquement non pas des partitions en quadrants, mais des points triples et des points d'arrêts.

Toutefois, une fois rectifiée cette erreur, l'approche structuraliste redevient descriptivement correcte et tout le problème est alors de comprendre les *mécanismes générateurs* des interfaces K dans les espaces supports W . De quels concepts, théories et modèles mathématiques a-t-on besoin pour résoudre ce problème ?

III. LA PERCEPTION CATÉGORIELLE EN PHONÉTIQUE

Le phénomène de catégorisation et, en particulier, celui de perception catégorielle (CP) est central pour toute la cognition car il est une condition de possibilité de nombreux actes cognitifs de haut niveau . Ce point a été particulièrement bien expliqué par Stevan Harnad dans son recueil de référence *Categorical Perception*.

"One of the most basic questions of cognitive science is : how do organisms sort the objects of the world into categories ?".

“In CP there is a quantitative discontinuity in discrimination at the

category boundaries of a physical continuum, as measured by a peak in discriminative acuity at the transition region for the identification of members of adjacent categories”.

“CP occurs when the continuous, variable and confusable stimulation that reaches the sense organs is sort out by the mind into discrete, distinct categories”.

“The CP phenomenon can be seen as an analog -to- digital transformation that recodes a continuous region of physical variation as a discrete, labeled equivalence class”.

Au niveau le plus élémentaire, la perception catégorielle est définie de la façon suivante. Les sonogrammes permettent d'abord de dégager des indices acoustiques (acoustic cues). Dans un diagramme Fréquence/Amplitude, on constate immédiatement qu'il existe un spectre continu, dépendant de la forme des résonateurs du tractus vocal qui module le spectre harmonique produit par les vibrations des cordes vocales. Les pics de ce spectre continu sont appelés les *formants*. Ils constituent l'indice acoustique principal des sons stationnaires (les voyelles). En ce qui concerne les sons non-stationnaires (transients) comme les occlusives (stops), les fricatives, etc. les principaux indices acoustiques sont le voisement, la turbulence et les transitions de formants. L'indice de voisement (VOT = voice onset time) est défini comme l'excitation de l'harmonique fondamental (trait voisé) VS un cut-off du premier formant (trait non-voisé). La turbulence est caractéristique de l'explosion des occlusives et des fricatives. Quant aux transitions de formants, elles constituent les principaux indices acoustiques de la place d'articulation (labial/dentale/vélaire).

On considère alors un "continuum" le long duquel varie un indice (disons le voisement) : par exemple de /ba/ à /pa/. On fait deux types d'expériences : des tests d'identification et de discrimination.

(i) Les tests d'identification produisent ce que l'on attend : on obtient deux courbes en X définissant un bord K . D'un côté de K la syllabe est identifiée comme /ba/ et de l'autre côté comme /pa/.

(ii) Le phénomène caractéristique de la perception catégorielle est que la discrimination intracatégoriale dégénère. A la limite (CP idéale), deux stimuli ne sont *discriminés* que s'ils sont *identifiés* comme différents. La discrimination est donc subordonnée à l'identification. Comme l'ont souligné Michael Studdert Kennedy et Alvin Liberman

“categorical perception refers to a mode by which stimuli are responded to, and can only be responded to, in absolute terms”.

Bien sûr, la CP pure n'existe pas. Mais il existe de nombreux exemples perceptifs de CP approximative. D'où l'intérêt de comprendre la possibilité de ces phénomènes.

La CP montre que les *espaces de contrôles* (indices acoustiques, paramètres

articulatoires) sont partitionnés en domaines (classes d'équivalences) par des bords, i.e. catégorisés. Il s'agit là de l'analogie perceptif de *phénomènes critiques* comme ceux de transitions de phases.

La CP possède entre autres les caractéristiques suivantes.

- (i) *Innité* (Eimas, Mehler, Jusczyk). Une catégorisation "initiale" K_0 est transformée lors de l'apprentissage du système phonologique par déplacement, dédoublement et effacement de seuils.
- (ii) *Nécessité fonctionnelle*. C'est une condition de possibilité de l'apprentissage. Elle explique la discrétisation perceptive du flux audio-acoustique. Sans catégorisation perceptive, le flux audio-acoustique ne pourrait pas être un signal support du code phonologique.
- (iii) *Spécificité*. La perception des parties consonantiques du signal acoustique (chirps ressemblant à des bruits non linguistiques) est, elle, continue.
- (iv) Comment l'expliquer théoriquement? L'hypothèse dominante est celle, réductionniste, des détecteurs de traits (feature detectors), récepteurs neuro-sensoriels répondant sélectivement à des domaines précis de valeurs d'indices acoustiques (cf. les travaux de Hubel et Wiesel en perception visuelle). Elle n'est pas satisfaisante car elle est trop réductionniste et conduit à faire proliférer les détecteurs.
- (v) Il s'agit en fait de mathématiser le concept *structural de paradigme* (d'espace catégorisé) au moyen de modèles à la fois structuraux ("forme" phonologique) *et* psychophysiques ("substance phonétique"). Il faut comprendre comment un système contrôlé S peut catégoriser son espace de contrôle en y engendrant un système d'interfaces, de seuils, de discontinuités.

La CP a suscité un certain nombre de controverses. Plusieurs types de théories explicatives s'opposent.

- Pour certains (Pastore, Eimas), elle serait innée et correspondrait à des dispositions sensorielles naturelles et générales.
- Pour d'autres (Liberman), elle serait innée mais de nature cognitive, spécifique de la perception de la parole.
- Pour d'autres encore, elle serait cognitive mais non innée et liée à l'apprentissage de prototypes conventionnels.

Elle a été qui plus est fortement remise en cause : par exemple à propos des célèbres expériences de Cutting et Rosner (1974) concernant la perception catégorielle des timbres (les stimuli sont des sons en dents de scie avec une attaque variable : il y a perception catégorielle entre corde pincée/frottée avec seuils à 40 ms). Mais il semble que ces critiques soient biaisées par la non prise en compte du fait que des dimensions caractéristiques du timbre comme l'attaque sont perçues logarithmiquement.

IV. LES MODÈLES DE CATÉGORISATION

Les modèles de catégorisation fondés sur le concept de discontinu sont des modèles de décomposition cellulaire d'espaces en domaines (catégories). Ils donnent une version géométrique du concept ensembliste de classe d'équivalence.

Il en existe de plusieurs natures différentes.

1. Les modèles de prototypes : cognitifs, top-down et métriques.
2. Les modèles d'apprentissage compétitif non supervisé à la Kohonen.
3. Les modèles connectionnistes : perceptifs, ascendants et dynamiques, mais à un seul niveau de dynamique.
4. Les modèles morphodynamiques : perceptifs, ascendants et dynamiques, mais à deux niveaux de dynamique.
5. Les modèles dynamico-symboliques à deux niveaux : dynamique (lui-même à deux niveaux) et cognitif (on retrouve alors des modèles de prototypes).

1. Les modèles de prototypes

Dans "l'espace" P (difficile à définir) des percepts, on suppose qu'il existe des prototypes p_1, \dots, p_n définis par des valeurs typiques des indices acoustiques. Un peu comme des capitales pour les régions d'une géographie. Etant donné un stimulus p , on suppose qu'un dispositif optimise la distance de p aux prototypes p_i . p est alors identifié comme allophone (token) du p_i optimal, i.e. le plus proche.

Ce modèle n'est pas satisfaisant car il ne peut produire que des interfaces qui sont des morceaux d'hyperplans médiateurs, et il relève d'autre part de processus top-down alors que le processus de perception catégorielle est en grande partie bottom-up et modulaire.

2. Les modèles d'apprentissage compétitif non supervisé à la Kohonen

• Dans ces modèles on considère l'espace des états V^n d'un réseau à n unités u_1, \dots, u_n . On ajoute un sur-réseau classifiant A , composé d'unités A_1, \dots, A_N . Chaque A_r est relié aux u_i par des poids w_{ri} . Le vecteur $w_r = (w_{ri})_{i=1, \dots, n} \in V$ représente A_r dans V .

Supposons alors que le N -vecteur $w = (w_1, \dots, w_N)$ soit donné. On considère la *stratification* de V définie comme précédemment par les hyperplans médiateurs. Autrement dit, on définit des cellules $F_r(w)$ de W centrées sur les w_r .

$$F_r(w) = \left\{ v \in V \mid \forall s \ \|v - w_r\| \leq \|v - w_s\| \right\}$$

On adapte alors la stratification aux stimuli. Autrement dit, si v est un input, il change le vecteur w (transition $w \rightarrow w'$) suivant la loi :

$$w'_r = (1 - \varepsilon h_{rs})w_r + \varepsilon h_{rs}v$$

i.e.

$$\delta w_r = w'_r - w_r = \varepsilon h_{rs}(v - w_r)$$

où ε est un step size, s est le s du A_s et du w_s tels que $v \in F_s$ et h_{rs} est une courbe en cloche centrée sur $r = s$.

L'application $w \rightarrow w' = T(w, v, \varepsilon)$ définit un processus de Markov. Si $P(v)$ est une distribution de probabilité d'inputs, on a la loi de transition:

$$Q(w, w') = \sum_r \int_{F_r(w)} \delta(w' - T(w, v, \varepsilon)) P(v) dv$$

On étudie alors les propriétés de convergence, de stabilité, etc. de l'application $V \rightarrow A$ ainsi définie.

3. La catégorisation dans les réseaux neuro-mimétiques

Sous sa forme la plus simple, un réseau de neurones formels consiste en la donnée de N unités u_i dont l'état d'activation y_i varie dans un certain espace d'états S . Les cas les plus utilisés sont

$$S = \{0, 1\}, \{-1, 1\}, [0, 1].$$

Un état global instantané du réseau est donc décrit par le vecteur $\mathbf{y} = (y_i)_{i=1, \dots, N}$ de l'espace de configurations $M = S^N$. M possède le statut de ce que l'on appelle un espace interne dans les modèles morphodynamiques.

Les unités u_i sont connectées entre elles par des connexions de poids synaptique w_{ij} .

Les $w_{ij} > 0$ correspondent à des connexions excitatrices.

Les $w_{ij} < 0$ à des connexions inhibitrices.

On a en général $w_{ij} = 0$.

Les w_{ij} déterminent le programme de calcul du réseau, sa compétence. Cette compétence est distribuée et encodée dans des valeurs numériques, celles des poids synaptiques.

Le réseau "calcule" de la façon suivante. Chaque neurone u_i reçoit des signaux afférents venant de ses neurones présynaptiques. Il "calcule" (i.e. change d'état interne en fonction d'une loi de transition) et envoie un signal efférent à ses neurones postsynaptiques.

On définit en général l'input de l'unité u_i comme la somme pondérée des signaux afférents:

$$h_i = \sum_{j=1}^{j=N} w_{ij} y_j, \text{ i.e. } \mathbf{h} = \mathbf{w}\mathbf{y}.$$

Les neurones u_i sont traités comme des automates à seuil dont l'état interne est régi par une loi de transition locale du type:

$$y_i(t+1) = g(h_i(t) - T_i), \text{ i.e. } \mathbf{y}(t+1) = g(\mathbf{h}(t) - \mathbf{T})$$

où T_i est un seuil et g une fonction gain.

On a typiquement:

- $g =$ fonction de Heaviside si $S = \{0, 1\}$,
- $g =$ fonction signe si $S = \{-1, 1\}$,
- $g =$ sigmoïde $= 1/(1+e^{-x})$ si $S = [0, 1]$.

Les w_{ij} et les T_i parcourent un espace de contrôle W ayant le statut de ce que l'on appelle un espace externe dans les modèles morphodynamiques. La dynamique *globale* du réseau s'obtient en agrégeant les lois de transition locales et en les itérant. Elle définit un endomorphisme $T : M \rightarrow M$ qui caractérise le réseau comme calculateur.

Dans la limite d'un temps continu, on obtient un grand système d'équations différentielles du type:

$$\dot{\mathbf{y}} = -\mathbf{y} + g(\mathbf{w}\mathbf{y} - \mathbf{T}).$$

◆ Sous l'hypothèse d'un feed-back complet (bouclage des entrées sur les sorties) on doit itérer l'endomorphisme T . On obtient ainsi un *système dynamique* sur l'espace interne M . Ce sont alors les états asymptotiques stables du système — ses *attracteurs* — qui sont significatifs. Ils définissent les *états internes* du réseau. Le phénomène dynamique de base est alors la capture asymptotique d'un état global instantané \mathbf{y}_0 du réseau par un attracteur A . Les réseaux neuro-mimétiques calculent donc d'une façon radicalement différente de celle d'une machine de Turing-von Neumann. Ce sont des calculateurs dynamiques bifurquant d'attracteurs en attracteurs.

On applique alors les concepts fondamentaux de la théorie qualitative des systèmes dynamiques (Poincaré, Birkhoff, Smale, Palis, Peixoto, Guckenheimer, Kolmogorov, Anosov, Arnold, Sinaï, Thom, Douady, etc...). Elle montre que les attracteurs asymptotiquement stables capturent toutes les trajectoires d'un de leur voisinage en les y confinant. Si A est un attracteur et $B(A)$ son bassin d'attraction, il existe des fonctions de Lyapounov f sur les complémentaires $B(A)-A$, c'est-à-dire des sorte de fonctions énergie

telles que:

- (i) $f \equiv 0$ sur A ,
- (ii) $f > 0$ sur $B(A) - A$,
- (iii) f est strictement décroissante sur les trajectoires dans $B(A) - A$.

Autrement dit, la dynamique minimise f . Si l'on se focalise sur les fonctions de Lyapounov en supposant que les attracteurs sont ponctuels, on retrouve les modèles dynamiques à la Thom-Zeeman proposés dès la fin des années 60. Ils se trouvent réalisés dans des réseaux de neurones formels.

Les dynamiques que l'on peut obtenir ainsi sont en général d'une redoutable complexité. Par exemple, dans le cas (totalement irréaliste sur le plan neurobiologique) où les connexions sont *symétriques*, Hopfield a remarqué que, pour $S = \{-1, +1\}$ et $g =$ fonction signe, les équations du réseau sont celles d'un système de spins en interaction. L'énergie minimisée par la dynamique est alors donnée par:

$$E = -\frac{1}{2} \sum_{i \neq j} w_{ij} y_i y_j + \sum_i T_i y_i .$$

Dans la mesure où les poids synaptiques w_{ij} fonctionnent comme l'analogue de constantes de couplage et où ils sont, de façon très intriquée, à la fois >0 et <0 , ces systèmes — qui exemplifient le cas *le plus simple* de réseaux de neurones formels — correspondent au cas *le plus complexe* de systèmes de spins, celui des verres de spins. Leur énergie présente un nombre considérable de minima relatifs locaux (états métastables) et pour accéder aux minima absolus globaux (états stables) les méthodes classiques du genre descente de gradient sont inopérantes. Il faut utiliser des algorithmes sophistiqués de physique statistique comme celui du *recuit simulé* (simulated annealing). On introduit un bruit dans le système, i.e. une "température" computationnelle T . On part d'une configuration initiale \mathbf{y}_0 à haute température (de façon à avoir accès à tous les bassins). On tire au hasard une configuration voisine \mathbf{y}_1 . Si $\Delta E < 0$ on effectue la transition. Si $\Delta E > 0$, on effectue la transition avec la probabilité

$$\frac{1}{1 + \exp\left(\frac{\Delta E}{T}\right)}$$

(cela permet de remonter les seuils). On itère jusqu'à un minimum local. On baisse alors T et on recommence le processus. On peut montrer que si T décroît suffisamment lentement (par exemple si $\lim_{n \rightarrow \infty} T_n \text{Log} n = c \geq D$ où D est la profondeur maximale des minima locaux non globaux) alors ce processus stochastique converge vers un minimum absolu lorsque $T \rightarrow 0$. Les minima absolus sont les configurations \mathbf{y} sur lesquelles se concentre lorsque $T \rightarrow 0$ la distribution de Gibbs:

$$G_T(\mathbf{y}) = \frac{1}{Z_T} \exp\left(\frac{-E(\mathbf{y})}{T}\right),$$

où Z_T est la fonction de partition:

$$Z_T = \sum_{\mathbf{y} \in M} \exp\left(\frac{-E(\mathbf{y})}{T}\right)$$

Dans les modèles connexionnistes la capture asymptotique par un attracteur A d'un stimulus s mettant le système dans un état initial σ_0 :

$$s \text{ (stimulus)} \rightarrow \sigma_0 \text{ (état initial)} \rightarrow A \text{ (attracteur)}$$

modélise le fait que l'attracteur A est le contenu de la *représentation* de s par le réseau. Les phénomènes de catégorisation sont pensés comme résultant de la partition de l'espace interne M en bassins d'attraction $B(A)$ d'attracteurs A qui fonctionnent comme autant de prototypes. Les "gradients de typicalité" s'interprètent alors comme des fonctions de Lyapounov sur les $B(A)-A$. Quant à l'opposition typique/non typique (i.e. générique/spécial), elle se trouve géométrisée au moyen de la stratification de M définie par les séparatrices des bassins.

- Toutefois, cette interprétation connexionniste en termes de seules dynamiques internes reste largement insuffisante pour modéliser les phénomènes de perception catégorielle. En effet, lorsqu'on fait varier des indices acoustiques ils opèrent comme des paramètres *de contrôle* et cela n'a rien à voir avec la capture presque instantanée d'un état initial par un attracteur (reconnaissance $\sigma_0 \rightarrow A$). Le phénomène de perception catégorielle porte *sur les attracteurs eux-mêmes* et non pas sur les états transients.

Il faut donc disposer de *deux* échelles de temps, c'est-à-dire de ce que l'on appelle une dynamique interne "rapide" et d'une dynamique externe "lente". La perception catégorielle s'explique alors au moyen de *bifurcations* d'attracteurs. Dans l'analogie bien connue : Réseaux de neurone \equiv Systèmes thermodynamiques la perception catégorielle devient l'analogie de phénomènes de transitions de phases.

4. Le modèle de Stevens

Kenneth Stevens est sans doute l'un des phonéticiens à avoir le mieux compris ce problème. Selon lui, la situation générale de la CP est la suivante. Il existe un système S (black box perceptive) dont les états internes \mathbb{A} (exprimés par des qualités perceptives) sont contrôlés par les paramètres d'un espace de contrôle W (indices acoustiques et/ou contrôles articulatoires) au moyen d'un champ de W dans \mathbb{A} : $\sigma : W \rightarrow \mathbb{A}$. Le point essentiel est que le contrôle est "non linéaire" (K. Stevens, 1972). Cela signifie que l'état interne P est stable sur de larges domaines de W et devient instable à la traversée de valeurs critiques K . Pour comprendre cette situation, Kenneth Stevens a dû développer (sans les connaître mathématiquement) les concepts essentiels d'un point de vue morphodynamique:

- (i) Stabilité/Instabilité (à la traversée de K l'état interne P devient instable),
- (ii) Quantitatif/Qualitatif (une variation quantitative implique ou non de variation qualitative),
- (iii) Variation/Invariance (une variation continue s'identifie à une invariance qualitative).

Il en a conclut à juste titre que la perception catégorielle est un phénomène global et qualitatif.

5. Modèles catastrophistes

La mathématisation du modèle de Stevens conduit immédiatement aux modèles catastrophistes où les indices acoustiques et les contrôles articulatoires constituent l'espace externe W et où le percept P (état interne de la black box) est défini par une dynamique interne X_W paramétrée par W .

Le champ $\sigma : W \rightarrow \mathbb{A}$ devient ainsi tout naturellement interprété comme un champ

$$\begin{array}{ccc} s : & W & \rightarrow & X \\ & w & \rightarrow & X_w \end{array}$$

de W dans l'espace X des dynamiques. On aura par exemple $X_W = \text{grad}f_w$ (avec $f_w : M \rightarrow \mathbb{R}$ une famille de fonctions potentiel sur M) dans le cas des modèles de gradient.

L'intérêt de cette approche est de permettre de définir des ensembles catastrophiques K (i.e. des catégorisations) *canoniquement associées aux dynamiques internes*. On définit d'abord la notion clef de *type qualitatif* des fonctions potentiel f . Cela permet de définir les *f structurellement stables* comme celles dont le type qualitatif résiste aux petites variations du contrôle w . Cela permet immédiatement de définir dans l'espace F des *f un ensemble catastrophique intrinsèque et global* K_F . Il est composé des $f \in F$ structurellement *instables*. Si $\sigma : W \rightarrow X$ est un champ, on peut alors définir un ensemble catastrophique dans W par image inverse en posant que K_W est dérivable de $\sigma^{-1}(K_F \cap \sigma(W))$.

La conséquence de ce point de vue est que des *modèles universels* pour les K_W seront fournis par *la stratification des déploiements universels des f de codimension finie*.

Les chercheurs du groupe du *Center for Complex Systems* de la *Florida Atlantic University* (Betty Tuller, J. A. Scott Kelso) ont accompli des travaux remarquables sur ce type de modèles dynamiques. Ils ont fait varier de façon continue un indice acoustique en faisant l'hypothèse que la perception est un processus auto-organisé de formation de patterns descriptible par des dynamiques non linéaires. Ils ont analysé la formation de seuils caractéristique d'une dynamique perceptuelle (avec multistabilité, perte de stabilité, switches catastrophiques). Par exemple, sur le continuum "say \rightarrow stay" engendré en

faisant varier l'onset du premier formant F_1 , ils ont observé les phénomènes catastrophistes caractéristiques :

- (i) de formation d'un seuil i.e. d'une valeur critique du paramètre de contrôle,
- (ii) d'hystérésis (lorsque la présentation des stimuli est séquentielle) :

“the presence of hysteresis may be regarded as strong evidence of non linearity and multistability in speech”,

- (iii) d'enhanced contrast :

“the boundary between patterns shifts to enhance differences between percepts.”

6. Les modèles a deux niveaux

Enfin il existe des modèles à deux niveaux, dont les premiers ont été proposés par Zeeman au début des années 80. Ils comportent :

- (i) Un niveau *dynamique* constitué de dynamiques internes X et d'attracteurs A . Les A sont interprétés comme des *hypothèses* sur l'identification des stimuli s .

- (ii) Un niveau *cognitif* opérant de la façon suivante :

— On analyse les stimuli s en traits distinctifs $t \in T$.

— Soit $w(A, t)$ le poids donné au trait t par l'hypothèse A . On a évidemment $w \in I = [0, 1]$ et $\sum_{t \in T} w(A, t) = 1$ (normalisation).

— Soit $r(A, t)$ la réalité de t sous l'hypothèse A .

— La vraisemblance $V(A)$ de l'hypothèse A pour s est la somme pondérée :

$$V(A) = \sum_{t \in T} r(A, t) w(A, t).$$

— On suppose alors que le système choisit l'hypothèse maximisant V .

V. LA CATÉGORISATION DES VOYELLES

Pour montrer que les modèles morphodynamiques s'introduisent tout naturellement dès que l'on prend en considération les bases acoustiques de la phonétique, nous allons considérer le cas le plus simple de catégorisation, celui des voyelles. Les voyelles sont les états stationnaires du résonateur vocal.

1. La fonction de transfert

Partons de la physique acoustique standard (cf. le classique Gunnar Fant :

Acoustic Theory of Speech Production, 1960, Mouton). Comme l'affirmait Fant :

"The speech wave is the response of the vocal tract filter systems to one or more sound sources".

L'élément essentiel est la *fonction de transfert* du tractus vocal. Elle se définit de la façon suivante. Le problème acoustique est un problème de propagation de vibrations. Dans un milieu fluide seules les vibrations longitudinales sont transmises.

Soit $u(t)$ le débit (la vitesse volumique) à la source en cm^3/s . On a par définition $u(t) = v(t)A$ où $v(t)$ est la vitesse des particules et A l'aire sectionnelle perpendiculaire au flux. Soit $p(t)$ la pression à la sortie en dynes/ cm^2 . On lui associe un débit $u_0(t)$. Le quotient p/u est l'impédance acoustique. $p(t)$, $u_0(t)$ et $u(t)$ sont des fonctions qui varient rapidement avec t . Pour les analyser, on considère (c'est une méthode standard) leurs *transformées de Laplace*.

Rappel: si $f(t)$ est une fonction du temps sa transformée de Laplace est la fonction d'une variable complexe s donnée par l'intégrale :

$$g(s) = \int_0^{\infty} f(t) e^{-st} dt$$

Ici, la variable complexe $s = \sigma + i\omega$ est une fréquence complexe, σ étant un facteur d'amortissement et ω une pulsation ($\omega/2\pi$ est une fréquence).

La fonction de transfert est définie comme le quotient des transformées de Laplace des débits à la sortie et à la source :

$$H(s) = \frac{U_0(s)}{U(s)}$$

(U est la fonction d'excitation de la source (glotte) et U_0 la fonction d'excitation de la sortie (bouche)). La restriction de H à l'axe imaginaire des $i\omega$ fournit le *spectre continu* de la voyelle.

Le principal intérêt de H est de permettre de définir rigoureusement les formants. Ils correspondent aux *pôles* de H . σ étant un facteur d'amortissement, plus ceux-ci sont proches de l'axe $i\omega$, plus les formants sont étroits et intenses.

2. L'analogie électrique

Rappelons qu'il existe une analogie électrique bien connue entre les résonateurs acoustiques et les circuits oscillants. Elle est donnée par le tableau suivant :

Voltage	→	pression sonore
Courant	→	vitesse volumique

$L =$ inductance	\rightarrow	inertie acoustique (la masse de l'air s'oppose à l'accélération)
$C =$ capacité	\rightarrow	compliance due à la compressibilité de l'air
$R =$ résistance	\rightarrow	résistance acoustique (puissance dissipée en friction sur les parois à cause de la viscosité de l'air)
$G =$ shunt-conductance	\rightarrow	puissance dissipée par conduction thermique sur les parois

3. Les formules pour les formants

Si les dimensions des cavités sont faibles et s'il y a peu de pertes, alors on a une propagation d'ondes planes et on obtient par exemple pour 1 formant:

$$H(s) = \frac{1}{LC s^2 + (RC + GL) s + GR + 1} = \frac{\omega_0^2}{(s - s_1)(s - s_2)}$$

avec $\omega_0^2 = \frac{1}{LC}$ jouant le rôle d'une fréquence propre. $H(s)$ est donc l'inverse d'un polynôme du second degré dont les coefficients sont des fonctions (algébriques) des paramètres "électriques", paramètres dépendant eux mêmes de façon plus complexe des paramètres acoustiques. Ses pôles sont les racines s_1 et s_2 de ce polynôme. Elles sont complexes conjuguées (car les coefficients sont réels) et seule la racine de partie imaginaire $\omega \geq 0$ est pertinente (et donne l'unique formant) car une fréquence est toujours ≥ 0 .

Pour le résonateur à un formant de la figure 1,

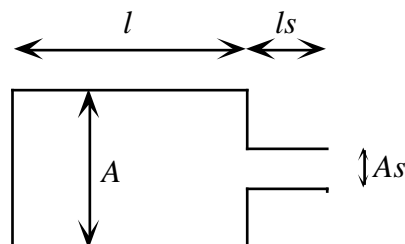


Figure 1

les paramètres sont donnés par les formules suivantes où As = section du tube ouvert, S = circonférence de la section de diamètre A , ρ = densité de l'air, c = vitesse du son, μ = coefficient de viscosité, λ = coefficient de conduction de la chaleur, η = constante adiabatique, c_p = chaleur spécifique de l'air à pression constante :

$$L = \rho l s / A$$

$$C = l A / \rho c^2$$

$$R = \frac{S}{A^2} \sqrt{\frac{\omega \rho \mu}{2}} (l + l_s)$$

$$G = S \frac{\eta - 1}{\rho c^2} \sqrt{\frac{\lambda \omega}{2 c_p \rho}} (l + l_s)$$

En appliquant la formule de résolution d'une équation du second degré, on voit que les pôles de H sont donnés par la formule :

$$\begin{aligned} s_1 &= -\frac{1}{2} \left(\frac{R}{L} + \frac{G}{C} \right) + i \sqrt{\frac{GR+1}{LC} - \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2} \\ s_1 &= \sigma_1 + i\omega_1 \\ s_2 &= \bar{s}_1 = \sigma_1 - i\omega_1 \end{aligned}$$

avec $\omega_1 = \sqrt{\omega_{01}^2 - \sigma_1^2}$ et $\omega_{01}^2 = \frac{GR+1}{LC}$. ω_1 est une fréquence propre perturbée.

Si les pertes sont faibles on a $GR \ll 1$, et

$$\omega_{01}^2 = \frac{GR+1}{LC} \cong \omega_0^2 = \frac{1}{LC} \gg \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2$$

On a alors $s_1 \cong i\omega_1$ et les formules approchées suivantes pour la fréquence F_1 du formant et pour sa largeur ΔF_1 :

$$F_1 = \frac{\omega_1}{2\pi} \cong \frac{1}{2\pi\sqrt{LC}}$$

$$\Delta F_1 = -\frac{\sigma_1}{\pi} = \frac{R}{2\pi L} + \frac{G}{2\pi C}$$

Pour la fonction de transfert le long de l'axe des fréquences, on a :

$$|H(f = i\omega)| = \frac{F_1^2}{(f + F_1) \left[(f - F_1)^2 + \frac{1}{4}(\Delta F_1)^2 \right]^{\frac{1}{2}}}$$

ce qui implique que F_1 et ΔF_1 définissent la fonction de transfert H .

Dans le cas de plusieurs formants, les formules sont beaucoup plus complexes mais donnent néanmoins des formules algébriques exactes. Par exemple, dans le cas à 2 formants (deux résonateurs, dont l'un comme ci-dessus, reliés par un tube), qui est absolument minimal pour les voyelles, on a :

$$H(s) = \frac{\omega_1 \omega_2}{s^4 + as^3 + bs^2 + cs + d}$$

avec

$$\omega_1 = \frac{1}{\sqrt{L_1 C_1}}, \quad \omega_2 = \frac{1}{\sqrt{L_2 C_2}}$$

$$a = \frac{R_1}{L_1} + \frac{R_2}{L_2} + \frac{G_1}{C_1} + \frac{G_2}{C_2}$$

$$b = \frac{1 + R_1 G_1}{L_1 C_1} + \frac{1 + R_2 G_2}{L_2 C_2} + \frac{R_1 G_2}{L_1 C_2} + \frac{R_2 G_1}{L_2 C_1} + \frac{G_1 G_2}{C_1 C_2} + \frac{R_1 R_2}{L_1 L_2} + \frac{1}{L_2 C_1}$$

$$c = \frac{(R_1(1 + R_2 G_2))}{L_1 L_2 C_2} + \frac{(G_2(1 + R_1 G_1))}{L_1 C_1 C_2} + \frac{R_1 + R_2 + R_1 R_2 G_1}{L_1 L_2 C_1} + \frac{G_1 + G_2 + G_1 G_2 R_2}{L_2 C_1 C_2}$$

$$d = \frac{(1 + R_1 G_1 + R_2 G_2 + R_1 G_2 + R_1 R_2 G_1 G_2)}{L_1 C_1 L_2 C_2}$$

les paramètres "électriques" étant donnés relativement à la géométrie des résonateurs par des formules analogues à celles du cas précédent.

Or on sait que l'on a besoin d'au moins trois formants pour une catégorisation correcte. Il existe donc une difficulté intrinsèque du problème.

En fait la situation réelle est encore plus compliquée. Il y a en effet une infinité de formants et la fonction de transfert $H(s)$ est un donc produit infini :

$$H(s) = \frac{\prod_{i=1}^{i=\infty} s_i \bar{s}_i}{\prod_{i=1}^{i=\infty} (s - s_i)(s - \bar{s}_i)} = \prod_{i=1}^{i=\infty} H_i(s)$$

$$|H_i(f = i\omega)| = \frac{s_i \bar{s}_i}{|s - s_i| |s - \bar{s}_i|_{(s=i\omega)}}$$

$$|H_i(f = i\omega)| = \frac{F_i^2 + \left(\frac{\Delta F_i}{2}\right)^2}{\sqrt{(f - F_i)^2 + \left(\frac{\Delta F_i}{2}\right)^2} \sqrt{(f + F_i)^2 + \left(\frac{\Delta F_i}{2}\right)^2}}$$

Les paires de pôles conjugués correspondent aux formants $F_1, F_2, \dots, F_\infty$ d'une voyelle donnée et leur position $s_i = \sigma_i + i\omega_i$ dans le plan complexe déterminent univoquement la réponse du tractus vocal car $F_i = \frac{\omega_i}{2\pi}$ et $\Delta F_i = -\frac{\sigma_i}{\pi}$.

Heureusement, on peut se restreindre à un nombre fini k de formants en introduisant un terme correcteur $K(s)$ tenant compte de la queue formantielle que l'on néglige.

$$H(s) = \frac{K(s) \prod_{i=1}^{i=k} s_i \bar{s}_i}{\prod_{i=1}^{i=k} (s - s_i)(s - \bar{s}_i)}$$

$K(s)$ est essentiel pour la synthèse mais il est très difficile à évaluer.

4. Trois points clés

Il faut retenir essentiellement trois choses de ces bases d'acoustique physique :

- (i) Pour n formants on doit considérer des *inverses de cuspoïdes* $\frac{1}{A^{2n}(s)}$

avec

$$A^{2n}(s) = s^{2n} + a_{2n-1} s^{2n-1} + \dots + a_1 s + a_0$$

c'est-à-dire les *déploiements universels* des singularités s^{2n} .

- (ii) Les coefficients de ces déploiements sont des fonctions algébriques *compliquées*

des paramètres de contrôle.

(iii) La conséquence essentielle en est que *la géométrie des bifurcations et des conflits possibles entre formants contraint donc de façon à priori et universelle la classification des voyelles.*

Les modèles les plus simples exigent les déploiements universels du cusp s^4 (2 formants) et du papillon s^6 (3 formants). On sait qu'il existe une complexité déjà notable de s^6 . Elle est pourtant réalisée physiquement par *les plus simples* des simulateurs physiques corrects de l'acoustique vocalique.

5. Modèles

La figure 2 montre comment la fonction de transfert, ou plutôt le logarithme de son module $\text{Log}(\text{Abs}(H(s)))$ varie en fonction des paramètres géométriques ls (figure (a)) et As (figure (b)) de la figure 1.

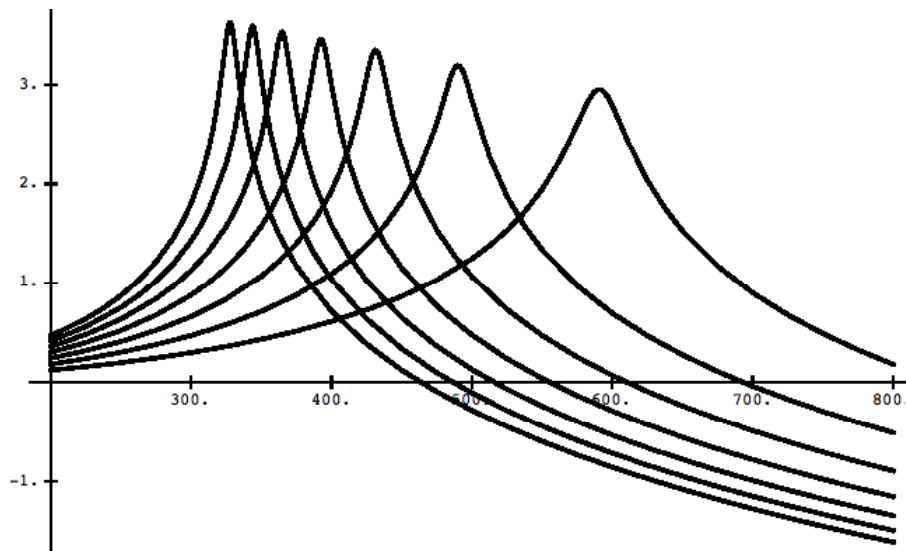


Figure 2 (a)

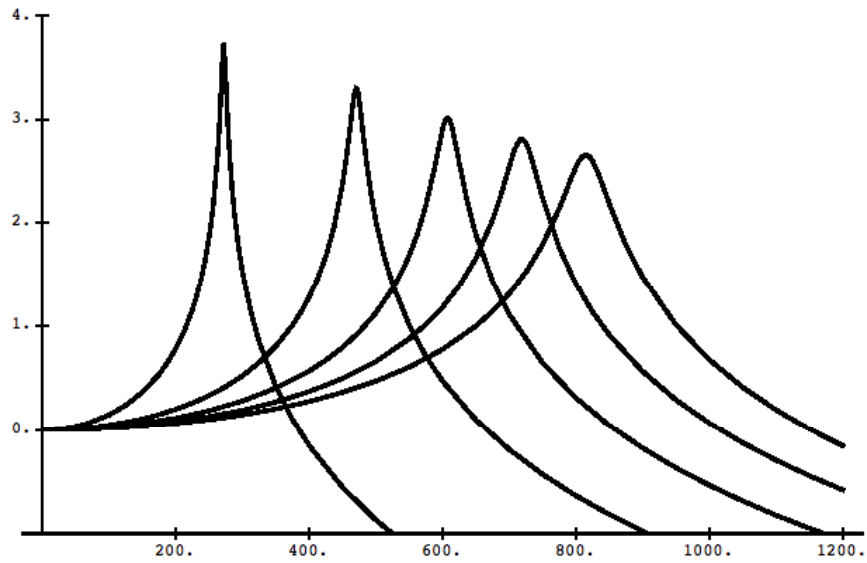


Figure 2 (b)

Dans le cas de 2 formants il faut un modèle à 4 tubes :

- cavité arrière : paramètres A_1 et l_1 ,
- zone de rétrécissement maximal : paramètres As_1 et ls_1 ,
- cavité avant : paramètres A_2 et l_2 ,
- lèvres : paramètres As_2 et ls_2 .

La figure 3 montre la variation de $\text{Log}(\text{Abs}(H(s)))$ en fonction de ls_1 (figure (a)), As_1 (figure (b)), ls_2 (figure (c)) et As_2 (figure (d)). Les valeurs numériques sont : longueur totale = $l = 17$, $l_1 = 6$, $A_1 = 5$, $l_2 = 6$, $A_2 = 5$, $ls_1 = 3$, $As_1 = 0.9$, $ls_2 = 2$, $As_2 = 0.9$.

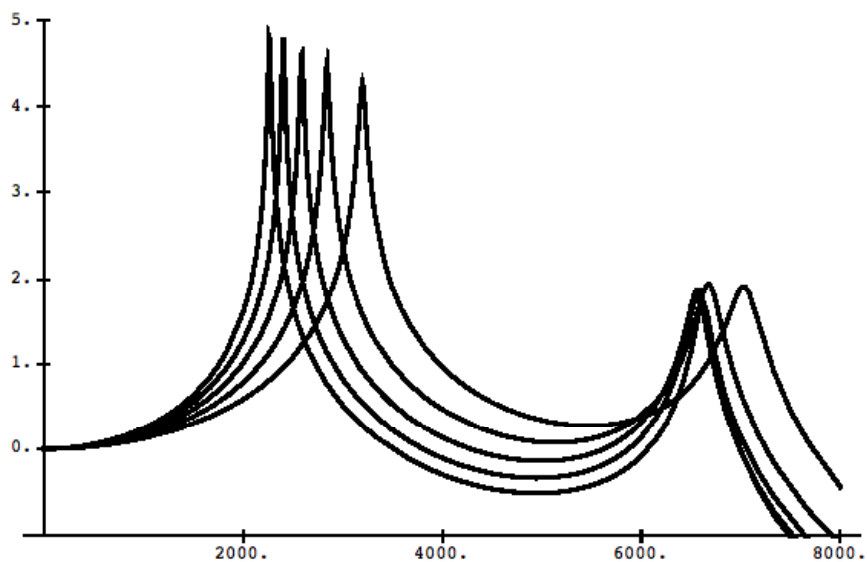


Figure 3 (a)

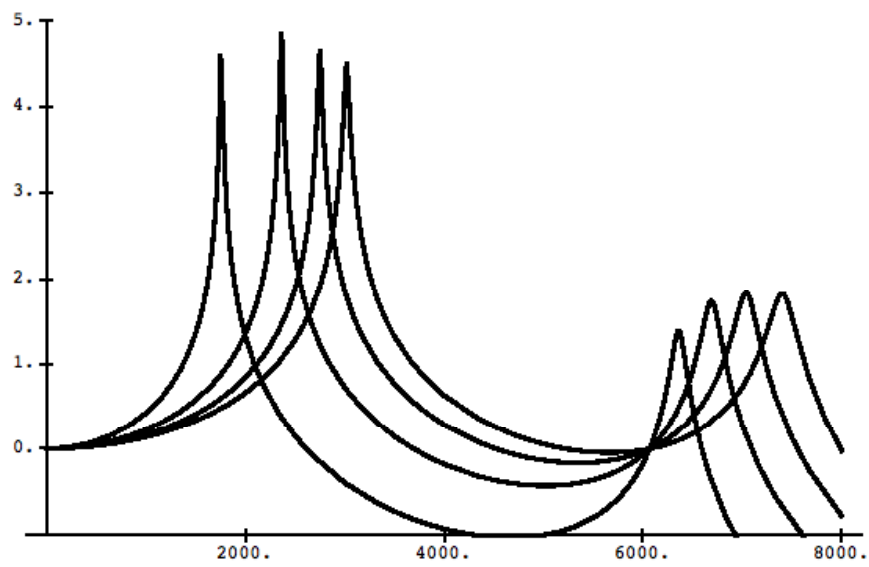


Figure 3 (b)

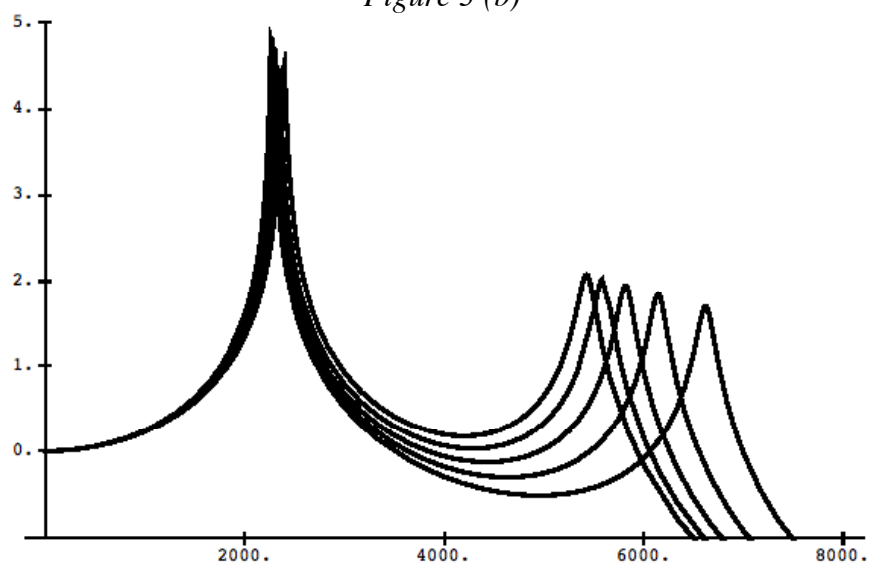


Figure 3 (c)

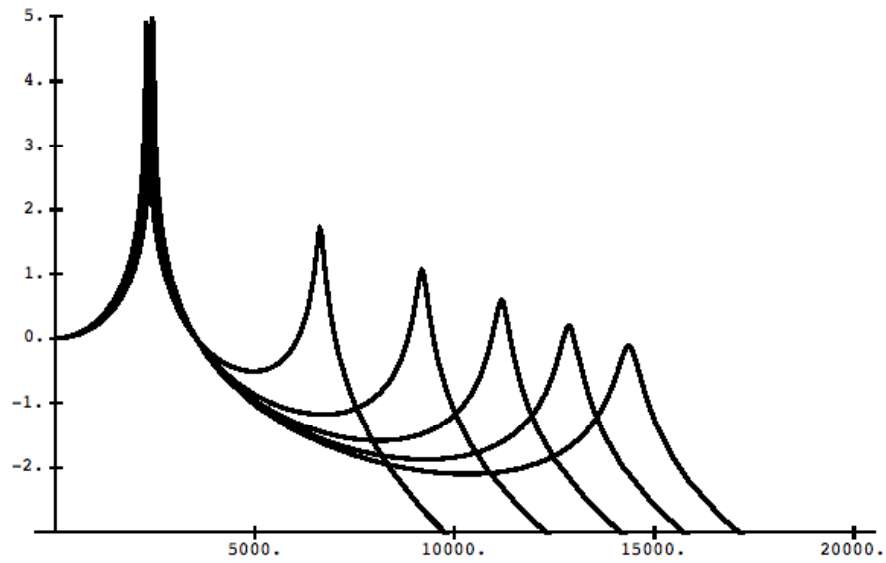


Figure 3 (d)

A la figure 4 (a) nous représentons le dénominateur de $H(s)$ lorsque le facteur d'amortissement σ augmente. Les figures 4 (b) et 4 (c) représentent $H(s)$.

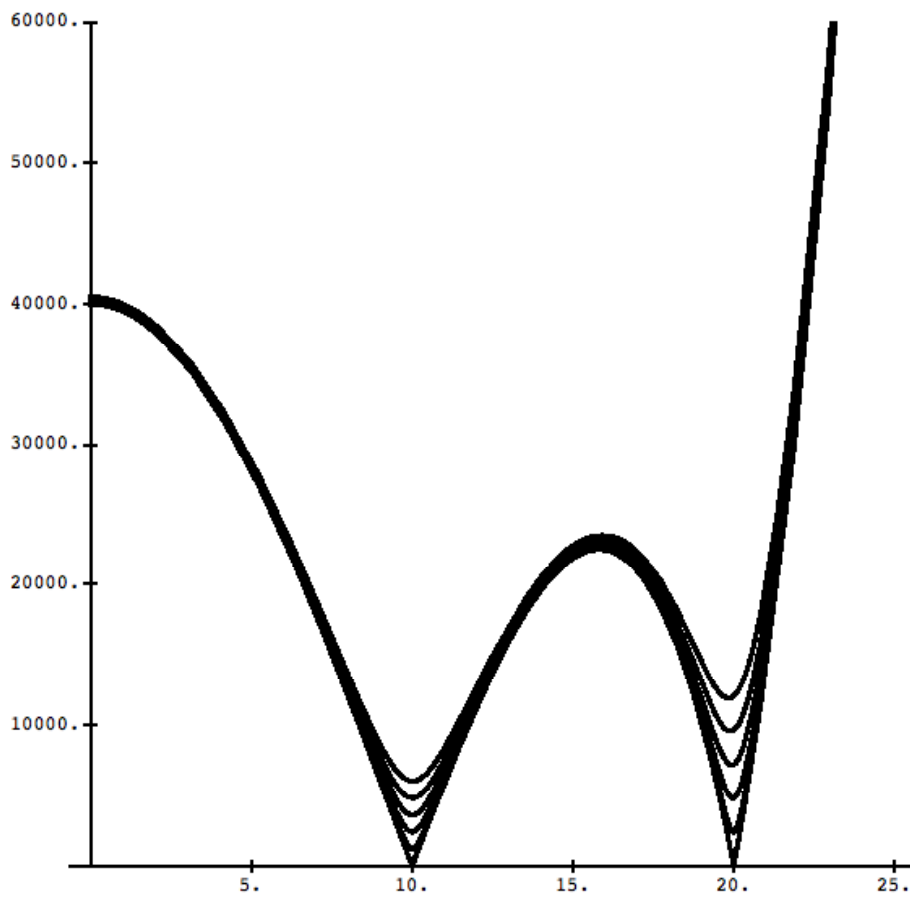


Figure 4 (a)

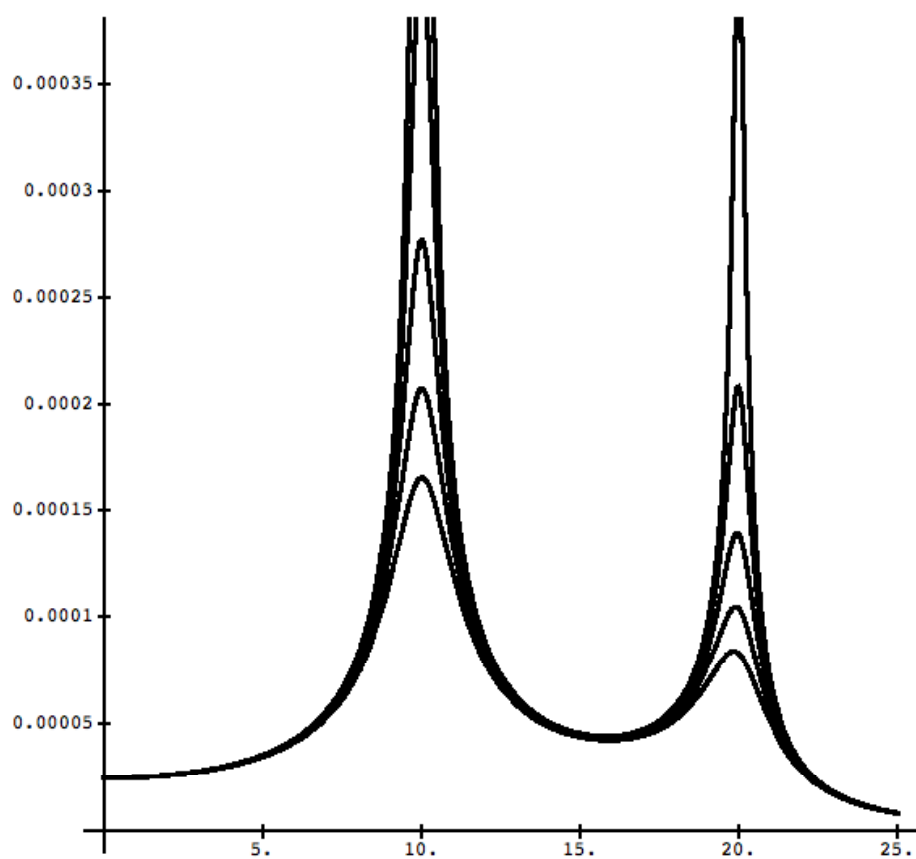


Figure 4 (b)

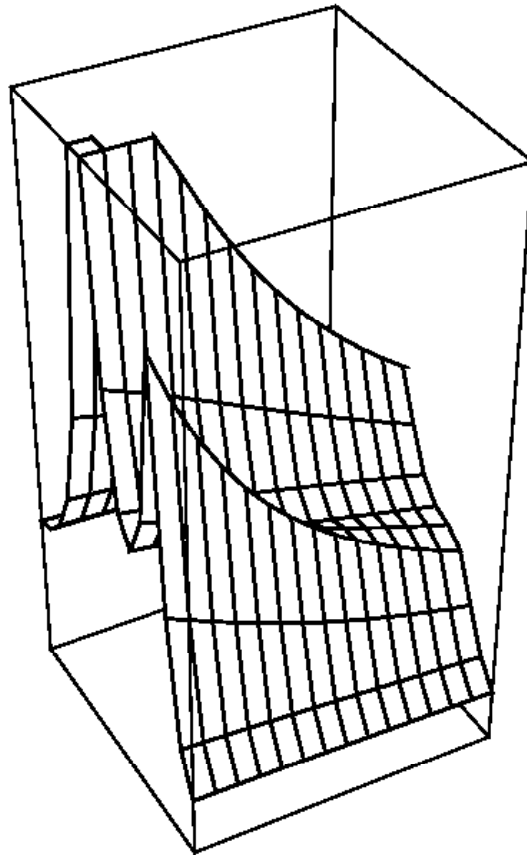


Figure 4 (c)

A la figure 5 nous représentons le dénominateur de $H(s)$ (figure (a)) et l'amortissement de $H(s)$ (figure (b)) dans le cas à 3 formants, c'est-à-dire pour la singularité s^8 .

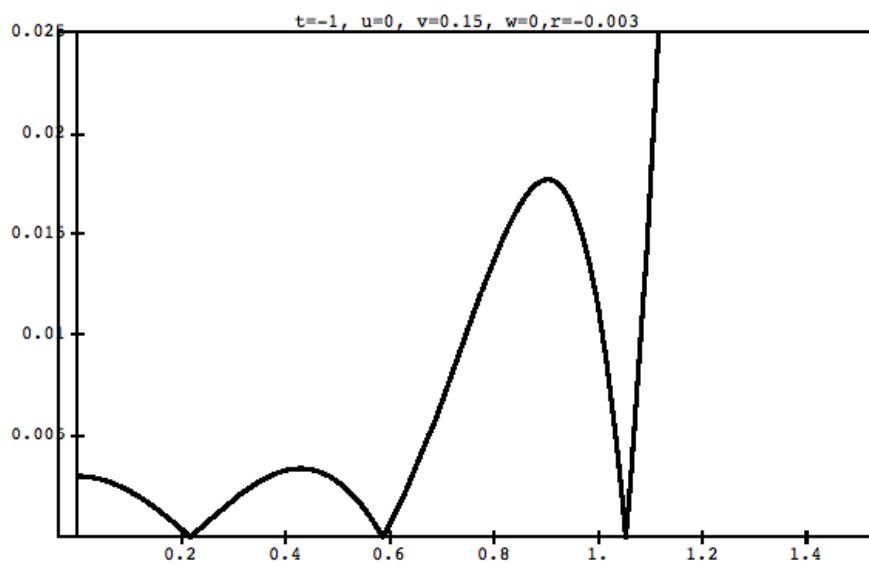


Figure 5 (a)

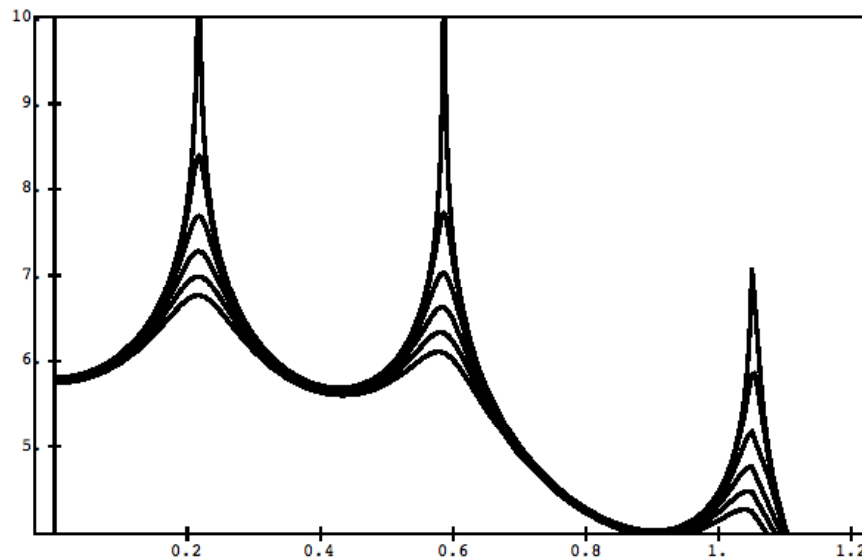


Figure 5 (b)

VI. LE PASSAGE A LA PHONOLOGIE

1. La transformation auditive

Une fois que l'on dispose de modèles acoustiques corrects, le problème est de savoir comment retrouver la classification *phonologique* des voyelles à partir de la catégorisation des espaces W (ou des espaces des formants puisque ceux-ci caractérisent la fonction de transfert). Pour ce faire, il faut d'abord passer du quantitatif au qualitatif comme nous l'avons expliqué dans *Les Catastrophes de la Parole*, et cela exige une *transformée auditive*.

Par exemple, dans le cas simple du triangle vocalique fondamental opposant */a/* au couple */i/-/u/*, la catégorisation phonologique, exprimée en termes des deux premiers formants F_1 et F_2 , repose sur une première opposition *Compact/Diffus* et sur une deuxième opposition *Grave/Aigu* qui vient se greffer sur le terme *Diffus* de la première. *Compact* signifie que les deux formants sont proches l'un de l'autre dans la zone centrale du spectre tandis que *Diffus* signifie au contraire qu'ils sont bien séparés, l'opposition *Grave/Aigu* explicitant alors quel formant, bas ou haut, domine l'autre.

L'on voit que pour passer à une description qualitative, il faudrait pouvoir considérer que, dans le cas *Compact*, les deux formants *fusionnent* en un seul. Le triangle vocalique fondamental serait alors interprétable au moyen du déploiement universel de la singularité *cusp*. Telle est la fonction de la transformée auditive.

Cette transformation auditive théoriquement nécessaire a été découverte expérimentalement. Nous pensons en particulier aux travaux de Chistovich (Institut

Pavlov de Saint Petersburg) et à ceux du groupe de l'ISC de Grenoble (notamment Jean-Luc Schwartz). L'idée est qu'il existe plusieurs représentations internes (RI) d'un même spectre vocalique.

(i) Une RI périphérique (cochléaire : il s'agit de la transduction par les cellules ciliées et du codage nerveux du signal acoustique). Elle est tonotopique et définit l'échelle physiologique des fréquences (transformation des Hertz en Barks). Elle est *pré-classificatoire*.

(ii) Une RI phonétique formantielle *post-classificatoire* (phonologique). Le point essentiel est que le rapport de distance entre ces deux représentations est fortement *non-linéaire*, ce qui est une façon d'exprimer le fait que le traitement phonétique est qualitatif.

(iii) Mais entre la RI cochléaire périphérique et la RI formantielle phonétique s'insère une RI *intermédiaire* comme l'ont montré les expériences sur les "centres de gravité" formantiels et le mécanisme ILB d'intégration spectrale large bande. Cette RI convolue la RI périphérique au moyen d'un filtre (disons gaussien), ce qui a pour effet de fusionner les formants dont la distance est inférieure à un seuil critique (d'environ 5 Barks).

Le mécanisme ILB et l'existence d'une distance critique d_c sont fondamentaux car ils permettent de passer du quantitatif au qualitatif dans la classification des voyelles (cf. les travaux de Fant, Stevens, Syrdal). On pourra par exemple caractériser les voyelles postérieures /u/ et /a/ par la condition $d(F_1-F_2) < d_c$. De même, pour les voyelles antérieures, on pourra distinguer le cas de /y/ correspondant à $d(F_1-F_2) > d_c$ du cas de /i/ correspondant à $d(F_2-F_3) < d_c$.

On passe ainsi des formants proprement dits de nature acoustique à des masses formantielles de nature perceptuelle qui sont plus grossières et présentent des accidents morphodynamiques de fission, de fusion et de compétition (conflit). A partir de là on peut reconstruire la catégorisation phonétique.

J.-L. Schwartz semble avoir prouvé la réalité de ce mécanisme comme mécanisme dynamique et ascendant (et non pas cognitif de haut niveau). Les discontinuités qualitatives confèrent des propriétés morphologiques à la perception et (Thèse p. 340) :

“les formes sont des données, elles préexistent à l'action de la mémoire et de l'intelligence (...) leur existence ne dépend ni d'un apprentissage ni de fonctions cognitives.”

2. Les voyelles focales et le papillon

Toutefois, la description du triangle vocalique de base par une catégorisation de type cusp opérant sur F_1/F_2 reste très insuffisante. On peut la raffiner à partir de la remarque suivante.

Il semble que les voyelles /i/ et /u/ soient *focales* au sens de Christian Abry, Louis-Jean Boë et Jean-Luc Schwartz, ce qui signifie qu'elles correspondent à des points où il y a convergence de formants et croisement par échange des cavités auxquelles les résonances sont affiliées. En effet (cf. les nomogrammes de Fant), la variation des formants vocaliques en fonction des paramètres articulatoires est, comme nous l'avons vu, fortement non-linéaire du fait que les coefficients des cuspoïdes sont des fonctions compliquées de ces paramètres.

Dans un modèle à 3 formants (encore trop simple mais déjà plus précis) correspondant à une catégorisation de type "papillon", cela signifie que /i/ et /u/ sont au voisinage des cusps F_1/F_2 et F_2/F_3 dans le déploiement universel de s^6 . On a donc une *hiérarchie* des catégorisations vocaliques correspondant aux phénomènes phonologiques de marquage. Au niveau le plus fondamental opère le cusp tandis qu'à un niveau supérieur /i/ et /u/ deviennent des centres organisateurs secondaires (voyelles focales) se déployant eux-mêmes en engendrant des sous-catégorisations subordonnées.

CONCLUSION

Le problème d'une classification phonético-phonologique qui soit à la fois quantitative et qualitative est mathématiquement non trivial pour la raison suivante. Pour disposer d'une classification qualitative, on doit considérer des déploiements universels, c'est-à-dire des formes normales tandis que pour effectuer des fits expérimentaux, on doit disposer de spectres continus numériquement précis. Ces deux exigences sont évidemment en conflit.

En effet, pour disposer de déploiements universels maniables, on doit tronquer à 3 ou 4 formants (singularités s^6 ou s^8) et on doit donc introduire une correction globale pour compenser l'effet de ce troncage (cf. les travaux d'Unto Laine). On doit ensuite bien positionner les valeurs des paramètres dans l'espace de contrôle. Or, les déploiements universels ont des espaces externes W de *dimension canonique*, liée à la nature de la singularité génératrice. Mais cette dimension est *inférieure* à celle de l'espace T des contrôles articulatoires A_i, l_i, As_i, ls_i . On a donc une application algébrique $T \rightarrow W$ avec des fibres compliquées, ce qui explique la non univocité bien connue du problème inverse consistant à remonter des formants aux contrôles articulatoires.

En conclusion, la géométrie des ensembles discriminants est riche. Elle catégorise (*i.e.* stratifie) l'espace de contrôle et définit donc des zones centrales dans les catégories (*i.e.* les strates ouvertes de la stratification). Cela permet de faire le lien avec des analyses statistiques, comme l'analyse factorielle, permettant de trouver ces valeurs centrales à partir de l'expérience.

REMERCIEMENTS

Je remercie beaucoup Per Aage Brandt de m'avoir invité à ce remarquable Symposium du Centenaire de la naissance de Roman Jakobson. Ces dernières années le structuralisme est devenu une science « dure » grâce à ses liens avec les sciences cognitives et il est donc particulièrement intéressant de le remettre en perspective à l'occasion d'une manifestation à la fois aussi solennelle et aussi amicale.

La version complète de ce texte (avec ses notes et sa bibliographie) est parue dans *Acta Linguistica Hafniensia*, 29, (1997), 239-269.