déterminisme, nous pouvons émettre l'hypothèse de l'existence d'un attracteur. Le signal change radicalement d'un phonème à l'autre, ce qui garantit, en théorie, que deux phonèmes se situent sur des hypersurfaces qui ne se coupent pas.

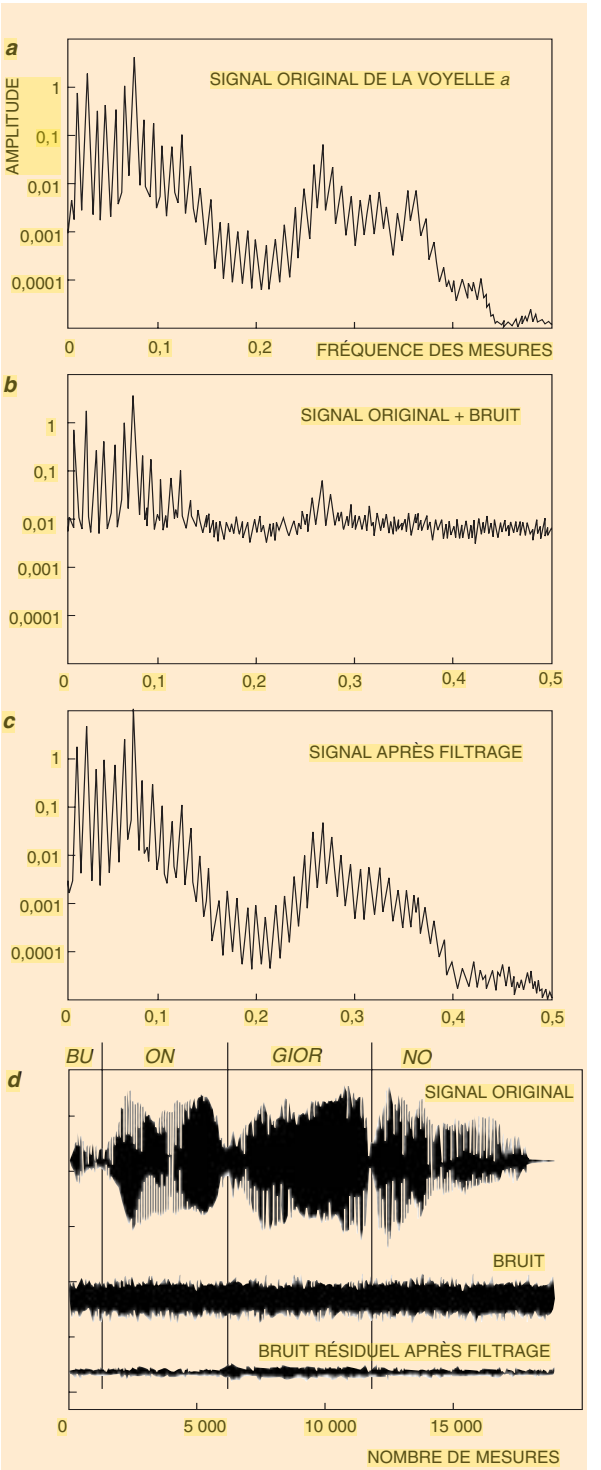
En pratique, les hypersurfaces caractéristiques des phonèmes ne se croisent pas, à condition que la dimension de l'espace des phases reconstruit soit suffisamment élevée. Quand on l'augmente, on «déplie» les hypersurfaces et elles ne se recoupent pas avec elles-mêmes. Ici, on augmente cette dimension de façon à éviter que deux phonèmes différents soient représentés par une même structure dans l'espace des phases reconstruit. Cette condition est importante pour retrouver le signal vocal. En effet, quand la trajectoire présente des intersections, le système ne peut être déterministe, car, si le point d'intersection constituait par exemple une condition initiale, le système pourrait évoluer aussi bien dans une direction que dans une autre, ce qui est contraire à l'idée de déterminisme.

Simple comme bonjour

Pour reconstruire les attracteurs caractéristiques de chaque phonème, nous avons observé que chaque vecteur devait couvrir une fenêtre temporelle de longueur équivalente au moins à un sous-phonème. En pratique, on enregistre un point de mesure s_k, environ toutes les 0,045 milliseconde (échantillonnage à 22,05 kilohertz). La longueur d'un sous-phonème varie entre 80 et 150 points, soit environ 3,5 et 6,7 millisecondes. Il suffit de choisir le délai temporel τ de telle sorte que $m\tau$ soit de l'ordre de grandeur de la longueur d'un phonème. D'après de récentes études empiriques, la dimension m associée à la phonation prolongée d'une voyelle est de trois pour une voix normale et de cinq pour un sujet affecté d'une pathologie des cordes vocales. En revanche, une phrase entière est trop complexe pour être analysée facilement. Nous supposons cependant que la dimension impliquée dans un sous-phonème ne dépasse pas dix. D'après le théorème de F. Takens, le choix d'une dimension m égale à deux fois la dimension obtenue pour un phonème augmentée de quelques unités, devrait permettre de disposer d'un espace des phases reconstruit dans lequel les phonèmes pourront être distingués sans ambiguïté.

En analysant plusieurs phrases, et en comparant les résultats du filtre réalisé avec différents délais temporels et différentes dimensions, nous avons établi que l'espace des phases reconstruit optimal est obtenu en prenant un délai temporel égal à l'intervalle séparant cinq points consécutifs de la série temporelle, soit 0,18 milliseconde. Nous avons choisi une dimension m égale à 25. Tous nos résultats ont été obtenus avec ces deux valeurs pour les paramètres τ et m.

Pour vérifier les performances du filtre, nous avons effectué une série d'expériences, de plus en plus complexes. Nous avons commencé par l'étude de la phonation prolongée de la voyelle *a*. Nous avons d'abord enregistré un signal en l'absence de bruit *(voir la figure 4)*, puis nous l'avons perturbé au point de détruire beaucoup des structures initialement présentes. Nous avons ainsi obtenu un signal aplati, caractéristique d'un bruit «blanc», c'est-à-dire pour lequel la densité spectrale de puissance est constante quelle que soit la fréquence. En appliquant le filtre, nous avons identifié presque toutes les structures de départ, bien qu'elles se trouvent au-dessous du niveau du bruit (leur amplitude est inférieure à celle du bruit). Peu importe à ce stade de comprendre



4. LE FILTRE est d'abord validé sur une voyelle soutenue : le signal d'un *a* émis pendant trois secondes est enregistré : il forme une série temporelle *(a)*. Un bruit numérique est ajouté et perturbe le signal *(b)*. Le signal est restauré après le filtrage *(c)* : la quasi-totalité des structures présentes dans le signal original réapparaissent. Puis le filtre est mis à l'épreuve avec le signal *bonjour (buon giorno) (d)*, auquel est ajouté un signal de bruit *(au centre)* ; on obtient ainsi le signal d'entrée du filtre. Le signal obtenu à la sortie du filtre est comparé au signal vocal original : la différence est un bruit faible, considéré comme un bruit résiduel *(en bas)*.

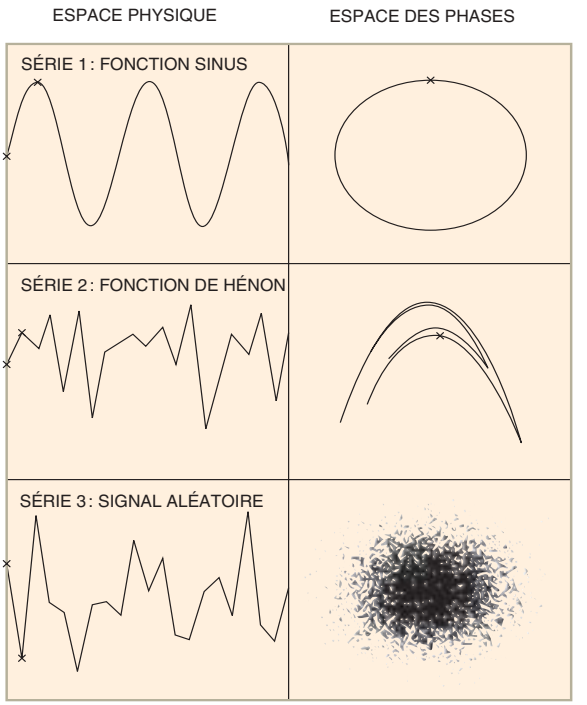
Reconstruction de l'espace des phases

L'étude de la dynamique des systèmes est plus facile dans un espace mathématique nommé espace des phases. On y représente l'évolution du système sous la forme d'une trajectoire, dont la structure renseigne directement sur la nature du système étudié : erratique ou chaotique, périodique ou non, etc.

La reconstruction de l'espace des phases peut être illustrée par de simples cas bidimensionnels. Considérons trois séries temporelles. Les coordonnées sont le temps t en abscisse et la valeur de la série à l'instant t, $f(t)$ en ordonnée *(à gauche)*. À ces trois séries, on associe les portraits de phase correspondants *(à droite)* : leurs coordonnées sont $(f(t), f(t + d))$, c'est-à-dire les valeurs de la série au temps t en abscisse et les valeurs de la série au temps $(t + d)$ en ordonnée, où d est l'intervalle qui sépare deux mesures. Dans l'espace des phases reconstruit, à chaque couple de points des figures de gauche correspond un seul point des figures de droite, après avoir éliminé la dépendance explicite de la variable temporelle.

La première série temporelle est une sinusoïde $y(t) = \sin \omega t$, un signal déterministe simple ; le portrait de phase correspondant est un cercle. Représenté par deux coordonnées $(x(t) = \cos \omega t$ et $y(t) = \sin \omega t)$, il est obtenu ici avec $y(t) = x(t + d)$ où $d = \pi/2\omega$. La deuxième série temporelle représente la fonction de Hénon, un système simple qui engendre du chaos. Son portrait de phase a une forme de banane. Enfin, un algorithme de génération de nombres aléatoires a produit la troisième série. Dans l'espace des phases reconstruit correspondant, aucune structure n'apparaît : le signal a une origine stochastique.

Bien que les deuxième et troisième séries semblent aléatoires, les espaces des phases reconstruits sont distincts : l'origine déterministe du deuxième signal contraint les variables de l'espace des phases reconstruit à former une trajectoire nommée attracteur étrange ou chaotique. En revanche, dans le cas de la troisième série, aucune structure ne ressort de cet espace des phases reconstruit.



ce que ces structures représentent, l'essentiel est qu'elles apparaissent de nouveau après le traitement du signal.

Nous avons reproduit l'expérience avec le mot *bonjour (buon giorno)*, d'une durée égale à une seconde environ *(voir la figure 4)*. À l'entrée du filtre, nous ajoutons un bruit numérique au signal original. Puis, nous comparons le signal obtenu à la sortie au *bonjour* initial : la différence est faible et peut être considérée comme le bruit résiduel. Ce bruit résiduel n'est pas constant, mais il varie, en particulier dans la transition entre les syllabes *bon* et *jour* (*buon* et *giorno*), où un pic apparaît. Un silence imperceptible se glisse entre les deux mots et le signal ne comporte plus de structures périodiques. À ce niveau, l'estimation de l'hypersurface sur laquelle effectuer la projection présente un plus grand degré d'incertitude. Les états situés à proximité des transitions entre les phonèmes occupent des régions de l'espace reconstruit qui n'appartiennent ni au phonème qui se termine ni à celui qui commence. Dans ce contexte, le filtrage se fait avec une qualité inférieure. Sinon, les phonèmes occupent des régions distinctes de l'espace des phases reconstruit, et on identifie leur point de séparation dans l'espace des phases et dans le temps, donc dans la série temporelle. Les phonèmes sont ainsi automatiquement reconnus.

Pour vérifier le fonctionnement du filtre, nous avons aussi fait varier le niveau du bruit introduit, de 10 à 500 pour cent de la variation du signal, c'est-à-dire que le signal superposé va d'un bruissement à peine perceptible à un bruit qui entraîne la perte complète de l'intelligibilité. Nous avons comparé les performances de notre filtre avec les résultats obtenus en utilisant l'algorithme qui sert de référence dans le filtrage de signaux vocaux, dit algorithme d'Ephraim et Malah. Notre filtre améliore les performances du filtrage, notamment pour des niveaux élevés de bruit.

Le filtre que nous avons conçu consiste à créer un espace des phases reconstruit à partir des données scalaires, ce qui, d'un point de vue strictement mathématique, a un sens seulement pour des signaux déterministes et stationnaires. Nous avons montré que la représentation dans un espace des états et le confinement dans une hypersurface sont valables au niveau de phonèmes distincts ; il faut seulement considérer un espace de dimension suffisamment élevée avec les paramètres de reconstruction qui conviennent. Le nombre élevé de dimensions de cet espace vectoriel aide à résoudre la non-stationnarité du signal vocal. Notre algorithme utilise des outils de la théorie du chaos sans toutefois que l'on ait à se préoccuper de savoir si la voix humaine est assimilable à un système chaotique ou non.

Lorenzo MATASSINI travaille dans la Société *Daimler-Chrysler*, en Allemagne.

J. GLEICK, *La théorie du chaos. Vers une nouvelle science*, Albin Michel, 1989.

A. DAHAN-DALMEDICO, *Chaos et déterminisme*, Points Sciences Le Seuil, 1992.

I. EKELAND, *Le chaos : un exposé pour comprendre*, Dominos Flammarion, 1996.

Holger KANTZ et Thomas SCHREIBER, *Nonlinear Time Series Analysis*, Cambridge University Press, 1997.