

Orthophonie: la musique du langage

Par Louis Rougnon Glasson (Ortograf-fr) novembre 2008

Du point de vue phonétique, la langue française compte une vingtaine de voyelles et une vingtaine de consonnes, d'où une succession de phonèmes à la fois commode à prononcer et harmonieuse à entendre, une syllabe étant très souvent constituée par l'association d'une seule consonne et d'une seule voyelle.

Simple et évident comme b-a-ba, comme ch-ou-chou, ou comme b-on-bon.

Le mot "chtroumpf" a obtenu un grand succès pour la bonne raison qu'il comporte pas moins de six consonnes pour une seule voyelle. Difficile de faire plus dans notre douce France. Il se trouve qu'il est emprunté à l'allemand. Quant au mot "tract", avec quatre consonnes pour une seule voyelle, sa prononciation est encore plus problématique en raison de la nature des consonnes qui interviennent, et il a déjà besoin d'être manié avec précaution pour ménager la fluidité de la phrase.

Dans la conversation courante, on prononce environ six syllabes par seconde. La durée de prononciation d'une voyelle est plus longue que celle d'une consonne, elle est supérieure à un dixième de seconde.

Les voyelles

La propriété caractéristique des voyelles, c'est qu'on peut les chanter durablement sur une note donnée, c'est à dire à une fréquence imposée.

Le signal correspondant est périodique, la fréquence est d'environ 80 hertz pour une voix d'homme, et d'environ 120 hertz pour une voix de femme.

Le motif correspondant à une période se reproduit alors quelque 12 fois pour une voix d'homme et 18 fois pour une voix de femme, pour le cas d'une voyelle située dans une phrase prononcée normalement.

Remarque: En raison de la brièveté de sa prononciation, la fréquence d'une voyelle ne peut pas être définie rigoureusement.

Si le son correspondant à une voyelle est émis par exemple pendant 0,15 seconde, on considère qu'il correspond à une superposition d'une infinité de sons de très faibles amplitudes et de fréquences voisines, qui sont en phase et s'additionnent pendant cet intervalle de temps seulement, et qui se détruisent par interférences le reste du temps.

Une relation dite d'incertitude (bien connue par ailleurs en mécanique quantique) donne un ordre de grandeur de l'incertitude sur la fréquence en fonction de la durée du signal. Plus le signal est bref, plus l'incertitude sur sa fréquence est grande.

Pour l'exemple abordé ici, un ordre de grandeur de l'écart entre les fréquences extrêmes sera de 6 hertz.

Un signal périodique est considéré comme étant une superposition de signaux simples sinusoïdaux, de fréquences bien définies.

La composante sinusoïdale dont la fréquence est la plus basse est appelé le **fondamental**. On a vu que sa fréquence est typiquement de 80 Hz ou de 120 Hz.

Les autres composantes sinusoïdales ont des fréquences **multiples de celle du fondamental**, on les appelle les **harmoniques**.

Un son donné d'une voyelle donnée est donc équivalent à une somme de sons qui seraient produits avec une série de diapasons, avec des fréquences et des puissances judicieusement choisies.

C'est l'ensemble des harmoniques d'une voyelle qui caractérise son **timbre**. Pour le son "ou", les amplitudes des différents harmoniques

sont toutes faibles, et on peut donc considérer que le signal est pratiquement sinusoïdal.

"Sinusoïdal", ça veut dire que la membrane d'un haut-parleur produisant ce son aura le mouvement d'aller-retour le plus répandu que l'on puisse trouver dans la nature, et qui est analogue à un mouvement de balancier.

La voyelle "i" est caractérisée par un timbre aigu, ce qui signifie que des harmoniques de fréquences élevées ont des amplitudes importantes. Mais ceci n'est pas forcément évident à première vue si l'on considère la courbe observée à l'oscillo ou à l'ordinateur, tout simplement parce que la sensibilité de l'oreille est beaucoup plus grande aux alentours de 1000 Hz qu'aux alentours de 100 Hz. Les "amplitudes importantes" des fréquences élevées le sont beaucoup plus pour l'oreille que pour l'oscillo.

Pour chaque voyelle, il existe deux bandes de fréquences caractéristiques, dans lesquelles les harmoniques ont des amplitudes importantes. Ces bandes de fréquences sont appelées "**formants**"

Problème de l'inventaire des sons

La diversité des accents régionaux pose le problème de l'inventaire des sons que l'on peut considérer comme étant les **sons constitutifs normaux** de la langue. Un tel inventaire est nécessaire pour proposer une norme claire et bien définie par exemple à des étrangers qui apprennent notre langue.

Pour les voyelles, cette question de la définition de la prononciation normale concerne d'une part le timbre, d'autre part la durée: quel timbre peut-on considérer comme étant le timbre normal de telle voyelle ? et quelle longueur ? Par exemple "il sut" ne se prononce pas exactement comme "il

sue", ou bien "nous" ne se prononce pas exactement comme "noue"

Dans la pratique, pour les sons réellement observés, le nombre de combinaisons du facteur timbre et du facteur durée que l'on trouve réellement est assez restreint. Par exemple, pour les sons de la famille du "o", les Grecs anciens connaissaient d'une part le "o grand" (oméga) et le "o petit" (omicron). Une analyse fine de la manière dont ces sons sont produits montrerait que le timbre propre du oméga s'accommoderait mal d'une énonciation trop brève, et que, à l'inverse, le timbre propre du "omicron" rendrait malcommode le fait de le faire durer.

Les caractéristiques d'un son peuvent être fortement influencées par les sons qui le précèdent ou qui le suivent immédiatement, d'où l'existence des **semi-voyelles ou semi-consonnes**: ce sont:

- le "ou bref" de "oui", "watt", "ouate", "loin",

- le "u bref" de "lui",

- et le "i bref", ou "i mouillé", de "lion", de "yéti".

Ces voyelles étant prononcées brièvement, leur timbre est moins précis. Par exemple, l'écriture de "loin" montre que le "ou bref" est ici l'aboutissement d'un "o bref". On peut dire que ces voyelles sont consonnisées, tout simplement parce qu'elles précèdent une autre voyelle et qu'on a tendance à rechercher l'alternance voyelle-consonne.

Cette tendance naturelle à une alternance voyelle-consonne a même fait apparaître un y, avec valeur de i consonnisé, dans le mot turc "tiyatro" (théâtre). On a la même chose lorsque de jeunes enfants prononcent "dëyor", pour "dehors".

Les sons de la famille du "o" apparaissent dans des expressions telles que "aux Commores". Le premier se retrouve dans "peau", "paume", le deuxième se retrouve dans "pot", "pomme", et le troisième se retrouve dans "sort", "cor", "encore": ce dernier est toujours suivi d'un r.

Les sons de la famille du "e" sont fournis par une expression telle que "mes deux belles-soeurs de gênet". On en trouve ici six, mais les inventaires habituels se limitent à cinq. On estime couramment que le "et" de "belle" est identique au "ê" de "gênet".

Un autre problème est au niveau des voyelles nasales: "on", "an", "in", "un". Dans bien des régions, la prononciation du "un" a dérivé jusqu'à être confondue avec celle du "in". Cette évolution pourrait bien être irréversible, et donc "brun" pourrait bien être définitivement condamné à être prononcé "brin".

Si l'on voulait revenir en arrière, il faudrait tenir compte d'un facteur particulier assez curieux. La différenciation un/in subsiste dans la **mémoire** des gens, en fonction de l'apprentissage qu'ils ont reçu. Mais les déficiences auditives qui apparaissent avec l'âge font que beaucoup de personnes auraient bien du mal de distinguer les deux sons à l'oreille, même dans le cas où ces sons seraient prononcés de façon soignée.

Les consonnes

Dans le langage courant, leur durée est plus courte que celle des voyelles. La nature du son bref correspondant, c'est un "bruit coloré", c'est à dire qu'on n'y trouve pas de fréquences particulières assez bien définies, mais un peu toutes les fréquences, avec des amplitudes plus importantes dans certaines bandes.

Pour certaines consonnes, on peut **faire durer** les bruits colorés qui les caractérisent. C'est le cas pour v, f, j, ch, z, s, et enfin pour r.

Les différenciations v/f, j/ch, et z/s font en outre intervenir des **différences de niveaux sonores**.

Pour les consonnes d, t, g, k, d, t, l, m, n, il n'est pas possible de faire durer les bruits colorés qui les constituent, parce qu'elles ont également pour fonctions caractéristiques de **provoquer une variation d'intensité sonore, dont l'ampleur et la durée sont peu modulables**.

La reconnaissance vocale

Il est évident que la machine écrivant ce qu'on lui dicte, moyennant un minimum d'interventions manuelles, n'en est pour l'instant qu'à ses premiers balbutiements, si l'on peut dire.

Avec certains logiciels, un utilisateur donné doit d'abord faire identifier par la machine chacun des différents mots qu'il compte utiliser.

Cette façon de faire est avantageuse pour les pays de langue anglaise, parce qu'une prononciation donnée ne donne en général qu'une possibilité d'écriture, les règles d'accord y étant réduites à leur plus simple expression. Elle y est judicieuse parce que les différences d'accents rendent difficile une reconnaissance universelle de chacun des phonèmes.

Avec d'autres logiciels, les caractéristiques de chaque phonème ont préalablement été identifiées de la manière la plus générale possible. L'ordinateur cherche à identifier chaque phonème à mesure qu'il est produit.

Ce système est avantageux pour des langues qui ont une prononciation bien définie et qui s'écrivent phonétiquement, telles que l'espagnol ou l'italien, parce que les difficultés du passage à l'écriture en orthographe normale ne doivent pas être insurmontables.

Pour la langue française, étant donné que nos correcteurs d'orthographe détectent uniquement, pour les rejeter, les graphies qui ne peuvent pas exister, la reconnaissance vocale n'a que peu d'intérêt en dehors de cas bien particuliers. Avec un type de logiciel comme avec l'autre, le texte donné par la machine doit être repris en détail pour en corriger l'orthographe.

Il est vrai que la frappe d'un texte est en elle-même non dénuée d'intérêt.