De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues

Alain Cazade

Apprentissage des Langues et Systèmes d'Information et de Communication

http://alsic.u-strasbg.fr
Vol. 2, Numéro 2, décembre 1999
pp 3 - 32
Recherche

De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues

Alain CAZADE
Université Paris IX Dauphine, France

Résumé : *Les logiciels de langues ont abondamment recours à l'affichage de courbes sonores autant pour rehausser leur crédibilité que pour tenter d'aider l'apprenant à fabriquer un lien entre sa production vocale et l'exemple qu'on lui demande d'imiter, alors qu'il ne sait seulement pas, le plus souvent, comment commencer à améliorer l'une pour mieux reproduire l'autre. Ces courbes sonores, de simples oscillogrammes la plupart du temps, peuvent-elles jouer un rôle quelconque dans ce sens ? Au moment où à tous les niveaux l'accent semble être enfin mis sur l'importance de la phonétique, de même que sur les spécificités de l'entraînement à la pratique de l'oral dans l'enseignement des langues vivantes, il semble intéressant de poser cette question même si nombreux sont ceux qui considèrent que de telles courbes sonores n'apportent rien, essentiellement pour l'insuffisance de l'information utile qu'elles contiennent. Si tel est le cas, on peut chercher à définir quels affichages aideraient les apprenants à mieux appréhender les mécanismes de la phonation et à faire progresser la qualité de leur prononciation. Une évolution semble déjà se dessiner qui vise à dépasser les faiblesses constatées. Il sera intéressant de voir si cette évolution va dans le bon sens.*

Introduction

Quand on s'intéresse de près ou de loin aux logiciels de langues actuellement disponibles, on ne peut s'empêcher de remarquer l'affichage abondant de courbes sonores accompagnant la diffusion des modèles proposés et les enregistrements de productions apprenantes. Ces courbes ont-elles un usage quelconque ? Ne sont-elles là que pour "faire joli", ou plutôt "faire sérieux", grâce à l'aspect scientifique, apparemment

inattaquable et définitif, qu'un joli graphe semble pouvoir donner, pour un utilisateur non averti, à toute présentation de résultats, dans un magazine scientifique ou non, à une présentation multimédia de courtier d'assurance ou même parfois à un exposé d'étudiant à court de démonstration ?

Nombreux sont ceux qui considèrent que leur affichage n'apporte rien et qu'il est en tout état de cause insuffisant. Il semble que le point mérite attention, tout spécialement au moment où à tous les niveaux, y compris institutionnels, et même si cela procède d'une attitude quelque peu volontariste, l'accent semble être enfin mis sur l'importance de la phonétique et la particularité de l'entraînement à la pratique de l'oral dans l'enseignement des langues vivantes. Pour replacer la question dans un cadre de réflexion un peu plus large, je me propose de faire également allusion dans cet article à d'autres représentations graphiques ainsi qu'à diverses fonctionnalités qu'il serait peut-être bon de voir proposées dans nos logiciels multimédias de langues à l'avenir.

1. Que trouve-t-on dans les logiciels actuellement ?

De nombreux logiciels proposent à l'apprenant de visualiser une courbe comme celle qui est proposée en figure 1. Il s'agit d'un oscillogramme (en anglais *waveform*), qui traduit visuellement les variations d'amplitude des vibrations causées par une source sonore (celui qui parle) dans l'air (ce pourrait être dans l'eau ou le long des parois d'une caisse de résonance d'un instrument de musique) avant qu'elles n'atteignent un récepteur (un micro, et au bout du compte une oreille, etc.). La perturbation causée par les mouvements des molécules qui vont et viennent dans l'air (sans toutefois le déplacer), plus ou moins vite et avec plus ou moins d'ampleur suivant les sons, est perçue par l'oreille ou par le microphone comme l'oeil voit un bouchon monter et descendre sur les vagues d'une mer plus ou moins agitée, sans que le bouchon se déplace, d'ailleurs (s'il n'y a pas de courant marin ou si le vent ne souffle pas). Les ondes produites sont, dans le cas de la parole, de natures multiples, certains sons étant assez réguliers, périodiques (reproduits de façon similaire sur une période donnée - cf. les sonorités simples produites par un instrument électronique ou même, éventuellement, les voyelles), certains autres pouvant être apparentés à du bruit, apériodiques donc, comme lorsqu'on prononce la lettre "f". Les oscillogrammes permettent de traduire ces suites de sonorités, périodiques ou non, sinusoïdales (avec leurs composantes harmoniques) ou non, sur un même graphe[1].

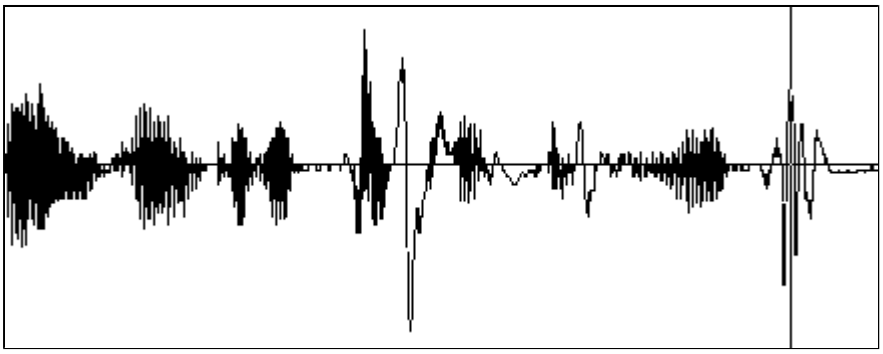


Figure 1 : Affichage d'un oscillogramme simple correspondant à la phrase : *"I'm going to the pictures tonight."* Copie d'écran d'une courbe affichée sous Wave Studio (1997)

Il n'est évidemment pas nécessaire de savoir et de comprendre en détail ce qui précède pour être capable d'établir certaines correspondances entre un oscillogramme et ce qu'on aura entendu, mais sera-t-on vraiment capable, sans une aide experte et des explications claires, de déchiffrer de telles correspondances avec précision, et surtout d'en tirer profit pour améliorer sa prononciation, comme disent pourtant les notices de la plupart des logiciels de langues ? Le graphique précédent correspond à la simple phrase en anglais: *I'm going to the pictures tonight*. Nous reviendrons sur cet exemple plus loin.

Certains logiciels, rares, essayent de permettre à l'apprenant de tirer un meilleur parti d'un affichage graphique. Les premières versions (jusqu'à la version 3.0) de Speaker (1997) développé par la société Neuroconcept proposaient pour chaque item modèle et pour la production apprenante correspondante trois affichages possibles. On notera au passage que le produit est canadien au départ et qu'il a été conçu par

des enseignants de langues.

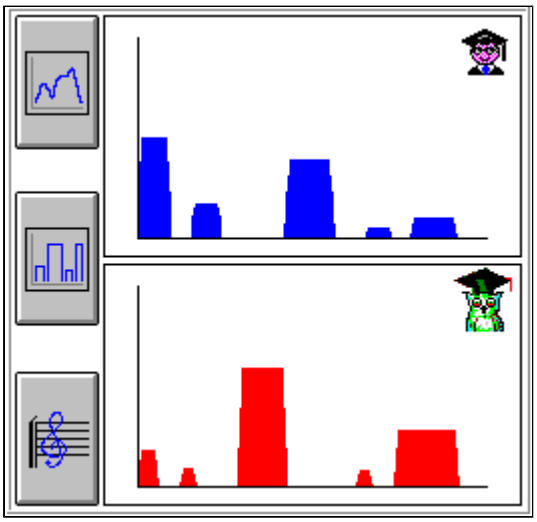


Figure 2 : Passages prononcés par l'apprenant et le modèle avec le volume sonore le plus élevé pour la phrase : *"This is a car, of course."*. Affichage supérieur = production de l'apprenant, affichage inférieur = modèle. Affichage proposé par la version 3.0 de Speaker (1997).

Le premier affichage, non reproduit pour économiser l'espace, est tout simplement la courbe sonore classique. Le deuxième,dont un exemple est reproduit en figure 2, permettait de ne proposer au regard de l'apprenant que certains éléments de la courbe sonore, correspondant à des passages prononcés avec le volume sonore le plus élevé. Un travail de filtrage repérait les pics de volumes (*peaks*) les plus distincts par rapport à un contexte proche, puis les situait sur l'axe horizontal (temps) en leur accordant une épaisseur et une hauteur correspondant à leur importance relative dans la totalité de l'item sonore traité.

L'affichage correspond à:*This is a car, of course*. Il a été récupéré durant une expérimentation du logiciel avec un apprenant d'un niveau faible mais attentif. Ce qui est prononcé par l'apprenant figure en haut et le modèle en bas. On reconnaît, dans ce dernier, les deux petites pointes correspondant au démonstratif puis au verbe un peu plus faible. L'article *a* n'a pas réussi à figurer sur le graphique, l'accent étant essentiellement porté sur le mot *car*. La voiture mentionnée ne doit pas être représentative de son espèce et le ton élevé (*high fall*, dynamique) du locuteur pour prononcer le mot paraît traduire son incompréhension devant l'étonnement de son interlocuteur. Avec un peu d'aide, l'apprenant pourra comprendre que, dans l'affichage du modèle, le marquage fort et un peu long du mot *course* peut souligner le sentiment légèrement scandalisé du locuteur qui ne se prive pas de donner à la voyelle centrale toute la longueur que sa valeur normale lui permet d'avoir. Il saisira peut-être également, si on attire son attention sur le fait, que dans sa propre production, l'accentuation n'est pas distribuée de la même manière, que le mot *car* n'est pas assez nettement mis en relief et que l'indignation n'est probablement pas suffisamment exprimée dans la finale : *of course*.

Les développeurs de chez Neuroconcept que j'ai rencontrés m'ont affirmé que bien d'autres variables que la seule variation d'amplitude étaient prises en compte pour établir ce type d'affichage mais ils ont préféré garder la recette de leur sélection pour eux. Gageons qu'une progression de la courbe d'intensité ou qu'une mesure de la rapidité des temps de montée du son devait en faire partie (une transitoire - un son se rapprochant de la percussion instrumentale, une plosive ou une occlusive, cf. les consonnes |p|, |k| - monte plus vite en intensité qu'une latérale ou qu'une vibrante - cf. les sons |l|, |r|).

Le mérite d'un affichage simplifié de la sorte est d'attirer l'attention directement sur des points "forts", dans différents sens du terme, de l'item travaillé, les variations du volume correspondant souvent, dans des productions orales courtes et surtout de structure simple, à des éléments clés de l'articulation sonore. Il n'en va certainement pas de même dans le cas de phrases plus complexes, avec une simple subordonnée et quelques variations d'intention expressive. Précisons que, le plus souvent dans les modules que j'ai pu tester, le niveau de travail proposé est assez peu élevé et que, même en cas de travail à un niveau

supérieur, les phrases proposées restent très courtes et relativement reconnaissables.

Le troisième affichage proposé (figure 3) essayait, quant à lui, de suivre la progression de la courbe mélodique et suivait la variation de la hauteur des sons. Toutefois, pour aider l'apprenant, le même principe de filtrage des données affichées ne faisait apparaître que quelques points clés permettant de suivre les mouvements particulièrement marquants de l'item sonore.

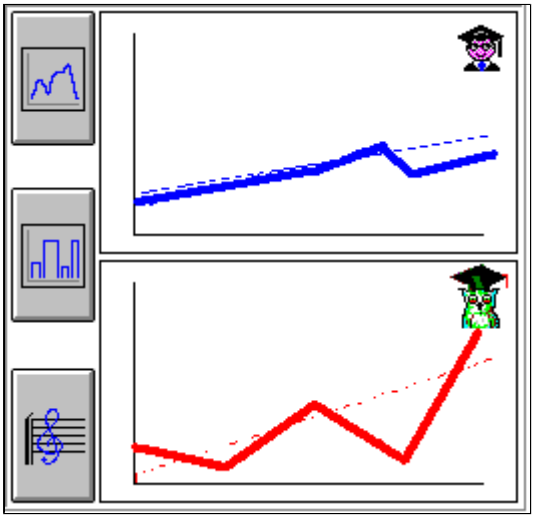


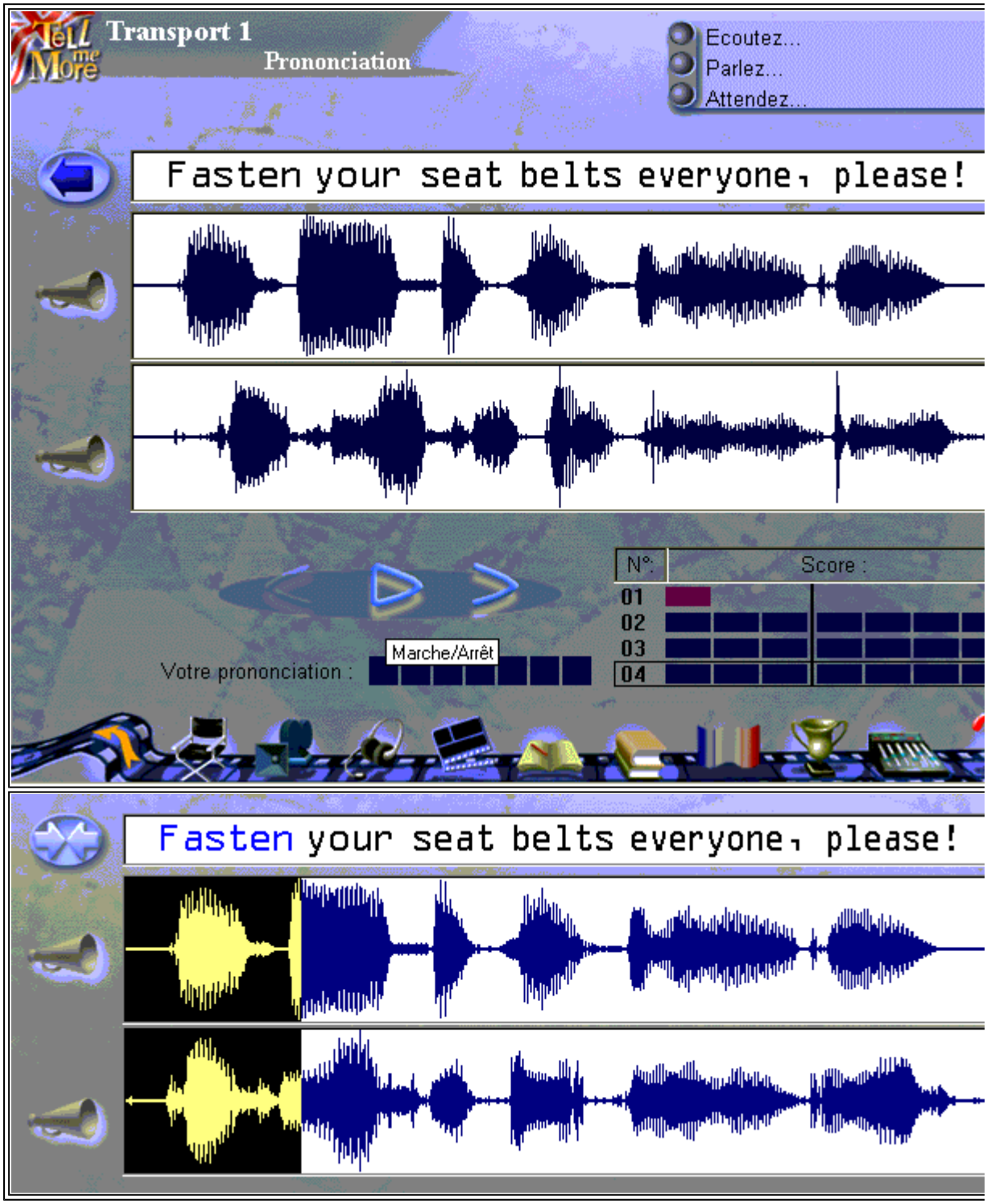
Figure 3 : Progression de la courbe mélodique et variation de la hauteur de son pour la phrase : *"This is a car, of course."*. Affichage supérieur = production de l'apprenant, affichage inférieur = modèle. Affichage proposé par la version 3.0 de Speaker (1997).

On remarquera une ligne en pointillé qui cherche à traduire la direction moyenne de la pente intonative. Le schéma affiché est en fait un peu étonnant, il donne une hauteur démesurée à l'attaque du mot *course* - nettement au-dessus du mot *car* - alors que le volume ne permettait pas de deviner cette hauteur. On regrettera que le schéma ne montre pas le début de chute qui succède à ce pic. Une véritable courbe affichant les variations mélodiques, une courbe F0 ou encore un mingogramme (voir infra) aurait probablement montré cela plus clairement. La production de l'apprenant est typique de ce qu'un jeune Français peut faire, n'osant pas jouer avec les variations de hauteurs de son qu'affectionnent les Anglo-saxons dans certaines prononciations très expressives.

L'intention qui présidait aux choix pédagogiques des concepteurs initialement était plus que louable, et on regrette fortement que cette triple possibilité ne soit plus offerte. L'explication, selon un développeur français que j'ai rencontré, tient au fait que les calculs avaient été faits, au départ, à partir du fonctionnement d'une carte vocale particulière, en rapport avec les caractéristiques physiques de la carte, ceci afin d'accélérer au maximum les délais d'affichage des données dans des environnements multimédias qui étaient souvent un peu limités à l'époque où le système a été développé (sous DOS dans sa première phase), une mémoire vive de 4 mégaoctets étant une rareté. Lorsque le nombre de cartes sonores a augmenté de façon spectaculaire, il n'a pas été possible, pour des raisons commerciales, d'obliger les utilisateurs à ne travailler qu'avec un seul type de carte. Les affichages ne donnaient pas, à partir de la plupart des nouvelles cartes disponibles, des résultats suffisamment satisfaisants, leur programmation logicielle n'étant pas assez indépendante du support physique; certains étaient même tout à fait fantaisistes. L'idée a donc été abandonnée et la seule courbe disponible désormais est la première, la moins intéressante de toutes, hélas.

D'autres produits comme Voicebook (1998) ou encore "Tell Me More" (1998) utilisent les affichages classiques non pas tant pour les possibilités d'analyse d'une production vocale qu'ils offrent aux

apprenants - le point est pourtant largement mis en valeur dans la documentation - que pour faciliter la sélection partielle en vue de la ré-audition d'une portion limitée d'un item. Il est vrai que, pour sélectionner ceci plutôt que cela, il faut pouvoir reconnaître la partie de l'affichage qui correspond à l'élément sonore qu'on veut isoler.



Figures 4 et 5 : Deux affichages proposés par "Tell Me More" (1998). Affichage supérieur = modèle, affichage inférieur = production de l'apprenant.

Dans "Tell Me More" (1998), par exemple, dans le cas où on veut effectuer une sélection partielle, il suffit de faire glisser le pointeur de la souris sur une des courbes sonores affichées. En fait, instantanément, les parties équivalentes des courbes des deux fichiers seront sélectionnées et en cliquant sur les haut-parleurs (figures 4 et 5, à gauche) correspondant à chaque piste (modèle en haut, apprenant en bas) on n'entendra que la partie sélectionnée.

Un point est intéressant à noter : il montre le soin avec lequel le logiciel a été conçu mais également les limites contre lesquelles un travail semblable ne peut que buter. Un élément aidant beaucoup au confort et à la clarté du travail est le fait que le texte correspondant au script de l'élément étudié apparaisse en haut de l'écran. Le passage dont il est question ici est extrait d'un module faisant partie de la partie *Advanced level* de *Tell Me More - English* (1998). Le fait de cliquer sur un mot (et non plus sur une ou l'autre courbe) sélectionnera l'équivalent (c'est du moins ce qui est visé, même si le résultat n'est pas toujours satisfaisant) dans la piste modèle et dans la piste apprenante. Pour les cas simples, comme à la fin avec le mot *please*, cela ne pose pas vraiment de problème, le mot étant clairement détacé du reste de la phrase. En revanche pour le mot *fasten*, l'expérience est plus délicate puisqu'un phénomène de liaison empêche de séparer les deux mots *fasten* et *your* autant visuellement, dans la courbe, que sur le plan sonore. La difficulté apparaît dans l'affichage représenté en figure 5 et est confirmée par l'audition qui ne permet pas d'entendre la fin du mot *fasten*. La finale n'est pas audible car trop intimement liée au début du mot suivant. Le développeur et l'expert linguistique qui a dû aider au montage de cette séquence a bien essayé de déborder un peu sur la partie de la courbe qui commence sous le mot *your*, mais il a sous-estimé la longueur de la finale nasale |n| du verbe précédent.

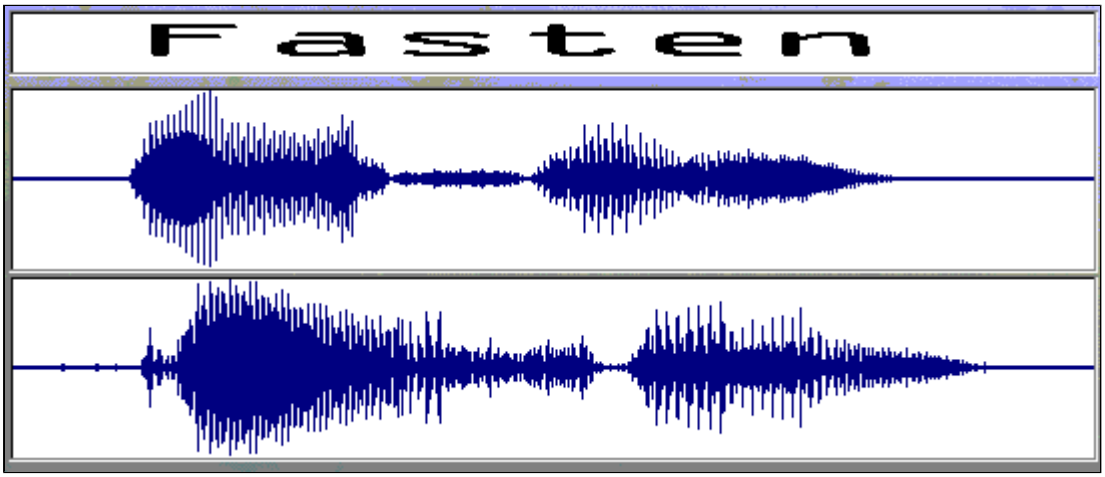


Figure 6 : Détail d'affichage proposé par "Tell Me More" (1998). Affichage supérieur = modèle, affichage inférieur = production de l'apprenant.

On appréciera malgré tout que, pour permettre un travail aussi fin que possible sans qu'il devienne trop spécialisé, il suffise de cliquer sur les deux flèches situées à gauche de l'affichage textuel (cf. figure 5 : la simple flèche s'est transformée en deux flèches opposant leurs pointes, suggérant habilement qu'une sélection plus précise est proposée) pour pouvoir travailler sur le mot sélectionné (ici, figure 6, le verbe *fasten*). Le mot est alors prononcé seul, sans contexte (il a été enregistré séparément) et est accompagné de l'affichage correspondant.

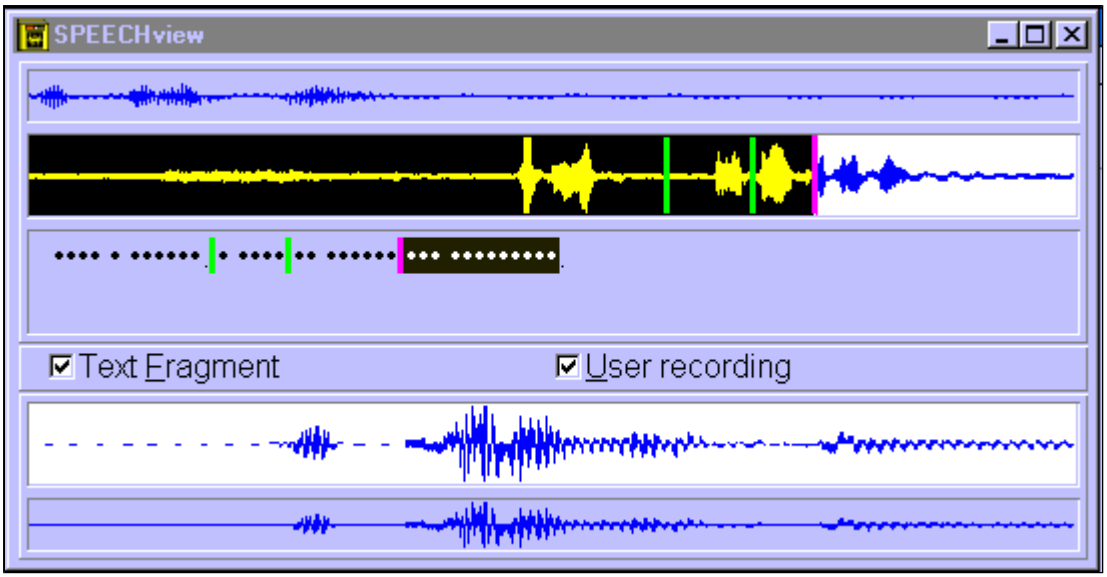


Figure 7 : Affichages proposés par VOICEbook (1998). Deux affichages supérieurs = modèle (entier et zoom); affichages inférieurs = production de l'apprenant.

Un autre logiciel, que l'on rencontre très souvent dans les centres de langues, utilise sobrement les formes sonores pour aider le travail de l'apprenant. Une partie du succès de VOICEbook (1998) vient du fait qu'il a le mérite de proposer, avec une présentation assez soignée, un travail simple et clair, tout utilisateur étant capable de comprendre dès le départ les visées du produit et les moyens utilisés pour y parvenir. Une interface d'enregistrement spécifique vient compléter le travail de restitution qui est l'activité centrale du produit. Cette interface (figure 7) donne accès, d'une part à la visualisation de la forme sonore du modèle (en haut), elle-même séparée en deux parties : la forme dans son entier (juste au-dessus du masquage de texte à compléter) et une sélection de cette forme, reconnaissable en inversion-vidéo claire dans la partie finale de la forme entière et reproduite de façon plus élargie dans la partie la plus haute de la fenêtre. On appréciera la possibilité, pour l'apprenant, de faire apparaître dans la forme entière, comme cela a été fait ici, des marqueurs séparant les divers éléments constitutifs de l'item sonore. Ce démarquage est doublé avec un même jeu de couleurs dans l'affichage du texte à démasquer. L'item travaillé correspond à la phrase : "*Just a minute.* | *I have* | *to answer* | *the telephone"* (les séparateurs - | - ont été placés aux emplacements correspondants). Si l'apprenant a sélectionné cette option, il lui suffit de cliquer sur telle ou telle partie, soit du masquage, soit de la forme sonore, pour que ce soit cette sélection uniquement que l'on puisse entendre. La forme sonore affichée ne sert donc que très indirectement au travail de compréhension et de restitution.

L'objet "courbe sonore" est davantage utile ici pour sa fonction de sélection qu'en tant qu'objet facilitant l'analyse d'un item sonore. Son rôle est malgré tout très intéressant de ce point de vue déjà. Les concepteurs du produit sont conscients du rôle délicat que peuvent jouer les courbes sonores et commentent le point avec honnêteté:

> *Ces formes d'ondes, dans l'état actuel de la technologie économiquement accessible, ne permettent pas une profonde analyse typologique des sons individuels qui composent un mot (comme "t", "w" ou "r"), mais elles permettent néanmoins d'identifier et séparer dans le temps les sons individuels comme entités et de montrer leur emphase relative ainsi que le développement rythmique de la phrase. Ces derniers aspects sont particulièrement important dans la langue anglaise.* (extrait du livret d'aide de VOICEbook, 1998)

Il est vrai que les *glides*, ("glissées", |w|, |j|), ou autres "sonantes" ("approximantes" comme |r|), ou même le |θ| et |ð| anglais, seront très difficilement reconnaissables sur de telles représentations. Tentons de voir s'il est malgré tout possible d'y repérer quelques détails.

2. À quoi ces courbes peuvent-elles éventuellement

servir ?

Qu'en tirer ? Pour un apprenant seul devant son ordinateur, sans apprentissage sérieux : peu de choses, vraiment. En revanche, si un enseignant un peu habitué à lire ces courbes se tient à proximité prêt à aider l'apprenant à déchiffrer de tels affichages, les choses peuvent être différentes et peut-être même plus riches qu'on ne le pense. Il n'est pas inutile de souligner, ici encore, le rôle irremplaçable de l'enseignant dans un environnement multimédia d'apprentissage des langues. Prenons quelques exemples de courbes réalisées à partir de phrases simples et courtes.

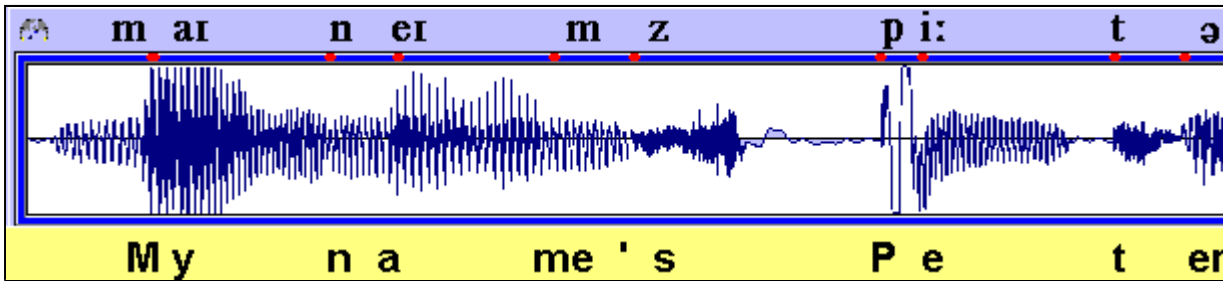
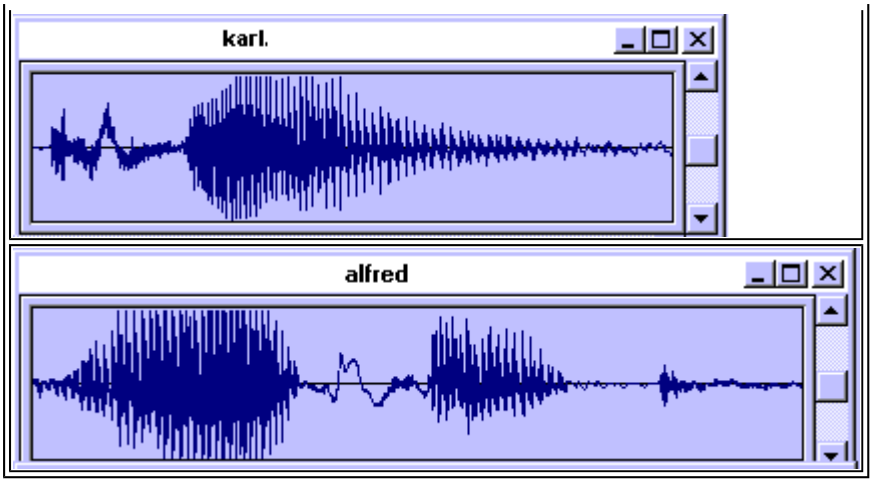


Figure 8 : Réalisée à partir d'une copie d'écran de "Wave Studio" (1997) retravaillée avec un logiciel d'édition graphique.

La petite phrase : "*My name's Peter.*" permet de mettre en valeur quelques formes typiques. Dans cette courte phrase, il est possible de reconnaître l'amorce du |m|. Les phonéticiens appellent cette articulation : "occlusive bilabiale", puisque le souffle est à un moment coupé par les deux lèvres servant à le former. Son émission nasale ne permet pas de tracer une courbe dont l'amplitude peut se mesurer avec celle de la diphtongue qui suit (correspondant au *y* dans *my* - voir la transcription phonétique au dessus de la figure 8), composée, elle, d'articulations voisées (émises avec l'aide des cordes vocales) et visiblement plus sonores. La première des deux parties de cette diphtongue se révèle plus sonore encore, sur le graphique, que la deuxième, appelée "fermante". L'articulation nasale qui suit, |n|, a, de nouveau, moins d'amplitude que la diphtongue qu'elle précède, construite par l'enchaînement de deux voyelles assez proches dans la partie de la cavité buccale où elles sont formées. La deuxième apparition du |m| correspond à une partie de la courbe comparable en partie à la première, elle se distingue nettement de la section correspondant au sifflement du |s|. Cette articulation ("fricative sifflante non voisée") est dessinée sur la courbe avec des zébrures très serrées, qui ne sont corrélables avec aucune vibration périodique, ce qui est typique du bruit fricatif.

L'attaque du |p| suit un moment de rupture qui correspond à la préparation d'une sorte d'explosion nécessaire pour que cette consonne ("occlusive bilabiale non voisée") soit produite : les deux lèvres, après s'être rejointes en bloquant momentanément le souffle et toute sonorité, laissent soudainement passer l'air. Le |i:| allongé est suivi d'un autre court moment durant lequel le son est interrompu : ce nouveau blocage correspond à l'occlusion qui précède le |t|. Un spécialiste ne reconnaîtra pas ici une prononciation britannique typique. La forme associée à l'émission du |t| ne montrant pas ce que cette "occlusive dentale" ("alvéolaire sourde") aurait produit chez un britannique, qui aurait partiellement rappelé la forme associée au |p| précédent. Il se trouve que le locuteur, dans ce cas, est un Américain d'un État de l'est. Son accent, même s'il se rapproche fortement de l'accent britannique par nombre de ses composantes, laisse ici apparaître sa différence. Le "schwa" final se prolonge en s'amoindrissant progressivement d'une manière qui, de nouveau, paraîtrait étrange chez un locuteur britannique type. Il est, en fait, associé à un discret |r| rétroflexe qui marque, lui aussi, l'origine réelle du locuteur.

Dans la courbe reproduite en figure 8, les différentes phases restent assez facilement reconnaissables.



Figures 9 et 10 : *"My name's Karl"* et *"My name's Alfred"*. Affichages réalisés à partir de copie d'écran de "Wave Studio" (1997) retravaillées avec un logiciel d'édition graphique.

Il en ira de même avec la suivante, "*My name's Karl*", dont seule la deuxième partie, qui correspond à la prononciation du nom *Karl*, est proposée ici (en figure 9). La prononciation de ce nom commence par un autre type d'occlusive ("occlusive vélaire" : la langue se colle au voile du palais et bloque le passage de l'air). Le |k| est non voisé (les cordes vocales ne vibrent pas) et sa forme est très irrégulière, très différente de celle qui représente la sonorité suivante, longue et ample. Il n'est pratiquement pas possible de repérer finement ici la trace du "*glide"* centralisant, ni de son enchaînement avec la consonne latérale sombre : |ɬ|.

La courbe correspondant au nom *Alfred* (figure 10) montre une amplitude forte pour la première syllabe (accentuée), à laquelle succède une forme de nouveau très irrégulière, sans périodicité, sans voisement. Ce passage correspond à la fricative sourde |f|, immédiatement suivie d'un |r| peu distinct du son |e| antérieur qu'elle introduit (et non pas |ɪ| ou |ə|, comme on aurait pu s'y attendre). La petite coupure suivante précède, comme dans les courbes précédentes, une nouvelle occlusive : |d|, alvéolaire, celle-ci; son amplitude est très faible ici puisqu'elle est située à la fin du nom *Alfred*, sans l'appui d'une syllabe sonore (à la différence de ce qu'on obtiendrait avec le nom *Daniel*, par exemple). Il est encore pratiquement impossible de situer la consonne "latérale" |ɬ| sombre, après la diphtongue initiale.

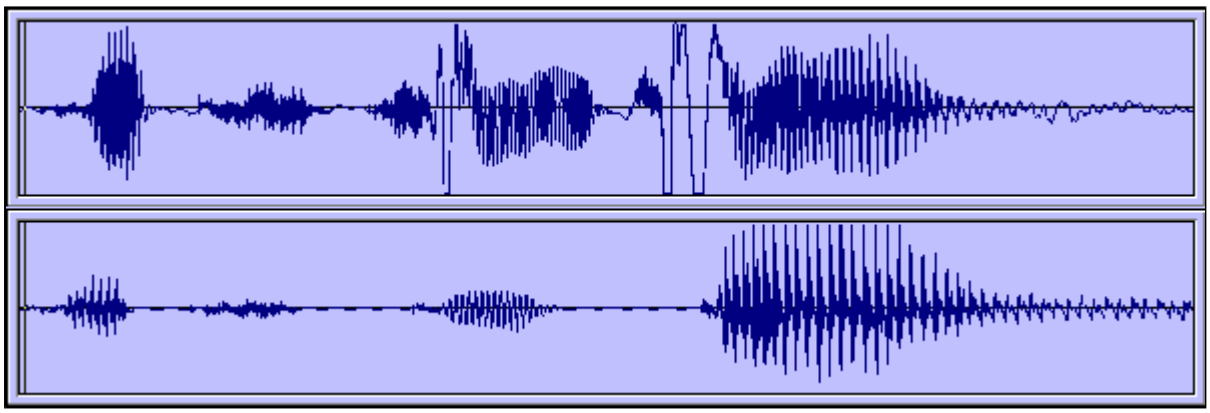


Figure 11 : *It's tea time.* Affichage partie supérieure par un anglophone, partie inférieure par un apprenant francophone. Affichages réalisés à partir d'une copie d'écran sous "Wave Studio" (1997) retravaillées avec un logiciel d'édition graphique.

L'exemple proposé en figure 11 correspond à la petite phrase : *It's tea time.* prononcée dans sa partie supérieure par un anglophone et dans sa partie inférieure par un apprenant francophone. Il sera peut-être intéressant de souligner certaines différences témoignant de difficultés que rencontrent souvent les francophones pour aborder la prononciation de l'anglais. Un détail qui saute aux yeux est probablement le

fait que les formes correspondant à l'émission de la consonne dentale |t| à la fois pour le mot *tea* et pour le mot *time* sont beaucoup moins visibles dans la courbe du francophone que chez l'anglophone. On sait que les anglophones font davantage retentir cette alvéolaire, en la faisant suivre d'un |h| expiré. Les deux courbes viennent de deux locuteurs différents, et les volumes de production ne sont clairement pas semblables. Il est donc difficile d'en tirer des conclusions précises sans l'aide de l'écoute. Pourtant on est tenté de déceler dans la forme très ample de la finale du francophone l'habitude classique qui consiste à peser sur les finales, alors que la forme supérieure ne montre pas le même déséquilibre. La position de l'accent, placé sur le premier mot du composé binominal *tea-time* pourrait faire apparaître un volume plus important sur le mot *tea* que sur le mot *time*, mais il est probable que ceci apparaîtra plus nettement sur une courbe d'intensité ou sur une courbe mélodique (montrant la hauteur des sons).

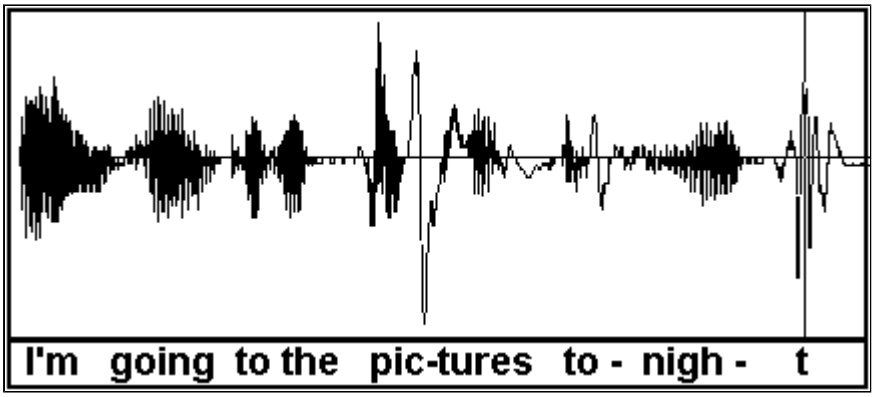


Figure 12 : Oscillogramme correspondant à la phrase : *"I'm going to the pictures tonight"*. Copie d'écran d'une courbe affichée sous Wave Studio (1997).

Revenons quelques instants sur la première courbe (reproduite figure 12). Quelques éléments deviennent, par analogie avec ce qui vient d'être dit, un peu plus clairs : l'ampleur imposante de la première diphtongue (*I'm*) qui rappelle l'entrée plus haut de *My name's* (figure 8), la position rapprochée de la préposition *to* précédant l'article *the*, la position des deux occlusives du mot *pictures* (bilabiale |p| et vélaire : |k|, comparables au |p| de Peter et au |k| de Karl, précédemment), la dentale finale de *tonight*, un peu forte pourtant (position du micro trop près de la bouche ?).

Il faut préciser ici que l'affichage des courbes - et donc toute possibilité d'interprétation - dépend fortement du type et de la qualité du micro ou/et de la carte vocale, de même que des capacités phonatoires de l'apprenant. Par ailleurs, notons qu'un même mot, une même suite de mots, voire exactement la même phrase pourront être prononcés par la même personne "exactement de la même manière", aux dires du locuteur, et pourtant produire des formes sonores réellement différentes, où les constantes ne seront pas toujours simples à repérer. Insistons sur le fait que les exemples choisis ici sont très simples et ne peuvent faire oublier que d'autres phrases, un peu plus complexes ou/et longues, contenant des liaisons ou des valeurs expressives particulières, poseront de réelles difficultés, même pour l'oeil averti.

Un pas vers l'inaccessible

Les remarques rapides qui précèdent montrent que si on habitue l'apprenant à repérer quelques détails caractéristiques ici et là, la présence de telles courbes peut se révéler utile. Surtout si celles-ci sont accompagnées, comme dans certains logiciels, du texte auquel elles correspondent. D'aucuns se demanderont probablement, dès lors, si un habillage dynamique du texte seul -cliquer sur un mot ou une sélection permettant de faire entendre la portion sonore correspondante- ne serait pas préférable à ces courbes. Il est possible de répondre que, pour bon nombre d'apprenants, les courbes représentent, telles qu'elles sont proposées déjà, un pas vers l'indicible, vers l'insaisissable. Il faut reconnaître que l'affichage de courbes sonores permet de présenter de façon visible et non subjective (avec la supériorité d'impact que le visuel a sur le sonore pour beaucoup) une transcription du problème que l'apprenant rencontre dans son travail d'approche de la production orale, processus mal maîtrisé dans notre société, pourtant très axée sur la communication parlée, processus aussi essentiel pour une bonne insertion dans le macrocosme sociétal

que celui de la respiration pour assurer la vie de l'individu. Cet affichage lui offre la possibilité, d'une certaine manière, de transformer la production orale en un objet-symbole permettant à sa sensibilité autant sensorielle qu'intellectuelle de commencer à avoir prise sur lui : le premier pas pour pouvoir envisager de progresser.

En revanche, il semble difficile de prétendre, comme le font pourtant de nombreux logiciels dont certains ont été cités ci-dessus, que cela peut aider à améliorer la prononciation. À moins de considérer que, en essayant vainement de reproduire une forme sonore équivalente à celle du modèle proposé, et donc en écoutant l'exemple et en le répétant un nombre considérable de fois, l'apprenant arrive à s'en imprégner si profondément qu'il finit par maîtriser une bonne partie de ses composantes sonores. Il est vrai que, sans ces affichages comparatifs, les exercices de répétition seraient certainement nettement raccourcis.

3. D'autres possibilités d'affichage

3.1 Courbes d'amplitude

D'autres courbes pourraient être également utilisées. Certaines seraient vraisemblablement moins utiles que d'autres. La courbe d'amplitude (figure 13, partie haute) n'aiderait pas beaucoup plus l'apprenant à se repérer ou à comprendre ce qu'il doit faire que le classique oscillogramme (en bas). Cette courbe d'amplitude est définie par les concepteurs du logiciel Wincecil (1997) (cf. décrit infra) comme la moyenne calculée à court terme de l'amplitude - ou magnitude - absolue d'un oscillogramme acoustique.

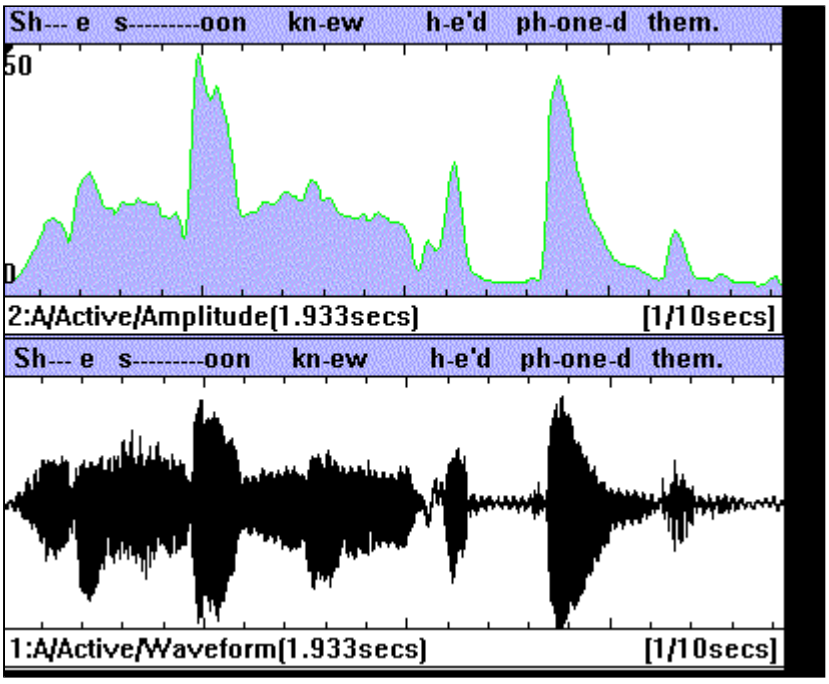


Figure 13 : Courbe d'amplitude et oscillogramme qu'on peut afficher en parallèle dans Wincecil (1997. Wincecil peut fonctionner sur un PC doté de moyens limités, et même sous Windows3x. Il reste préférable d'enregistrer les sons par un autre logiciel. Limites d'enregistrement : 11 ou 22 Khz.

3.2 Courbes mélodiques

Une courbe mélodique, en revanche, est beaucoup plus prometteuse. À la différence de l'oscillogramme vu précédemment, elle affichera la progression dans le temps de la fréquence fondamentale, FO, fréquence[2] des cordes vocales pendant un énoncé. Cela montre qu'à un instant *t* (une fraction de

seconde), le message sonore entendu est plus ou moins aigu ou grave.

Un logiciel proposé sur le site Internet du SIL (nd), intitulé Wincecil (1997), relativement simple d'usage et offrant déjà un certain nombre de possibilités d'affichages intéressantes (longueur maximum d'un son analysé: 3 secondes!), permet de visualiser une courbe mélodique. L'affichage reproduit en figure 14 correspond à la prononciation de la phrase : "*He probably won't remember anything by then, anyway*." (il ne se souviendra probablement de rien, à ce moment-là, de toutes façons).

Il montre la "courbe fondamentale", aussi appelée *F0*, de cette courte phrase. Certains points de la courbe ne sont pas affichés, ils correspondent à des passages non voisés (sans résonance des cordes vocales) de l'item sonore. Il est possible de fixer soi-même le seuil d'affichage ou de rejet des sons non voisés. Le logiciel permet d'écarter certaines fréquences parasites éventuellement mal interprétées, et c'est cette option qui a été choisie ici. On voit que la courbe varie entre 100 et 300 Hz, ce qui correspond à une voix féminine, dont la variation moyenne se situe généralement entre 150 et 300 Hz, alors qu'une voix masculine variera plutôt entre 80 et 200 Hz. Je dois préciser que j'ai retravaillé ce graphe sous un logiciel graphique simple, de façon à pouvoir faire figurer en haut de la courbe l'affichage des mots de la phrase (plutôt que leur transcription phonétique) et à matérialiser sur la courbe les secteurs correspondants. Il n'est pas en effet possible, dans ce logiciel, de faire afficher plus de 20 caractères (une police de l'alphabet phonétique international est fournie avec le logiciel).

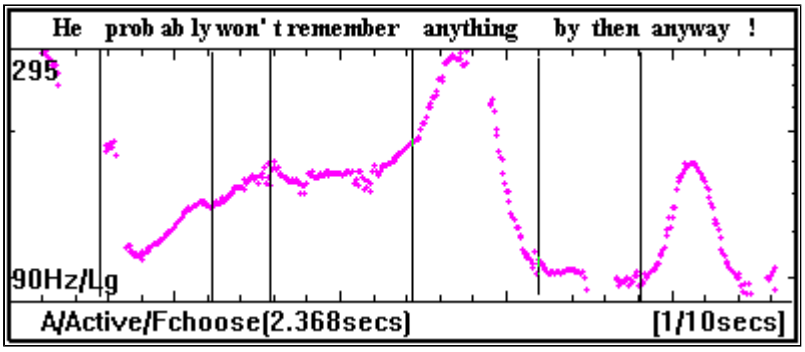


Figure 14 : Courbe F0 (Fondamentale) affichée par Wincecil (1997). Affichage *Fchoose* (certains signaux parasites ont été supprimés).

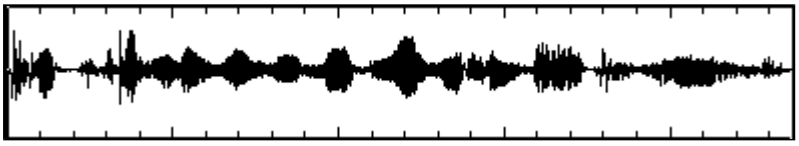


Figure 15 : Oscillogramme de la phrase : *He probably won't remember anything by then, anyway.* affiché par Wincecil (1997).

Comparativement à ce qu'un apprenant pouvait déceler en analysant l'oscillogramme classique correspondant (figure 15), beaucoup d'apprenants pourront mieux prendre conscience, avec la courbe mélodique, des intentions expressives du locuteur. La montée centrale sur le mot *anything*, par exemple, et l'écho qu'on peut en trouver dans le mot *anyway*, sera probablement mieux interprétée comme traduisant un sentiment de résignation devant le fait que le message dont il est question dans la phrase ne sera "même pas" mémorisé par le personnage concerné (*he*). Ce n'est pas sur la donnée temporelle, *by then* (à ce moment-là), qu'on insiste ici. La référence a dû être déjà introduite dans ce qui précède, ou est implicite dans le contexte. Il est probable que même à des fins de sélection partielle dans un item sonore (cf. supra), la courbe de la figure 14 serait plus pertinente et efficace que la courbe de la figure 15.

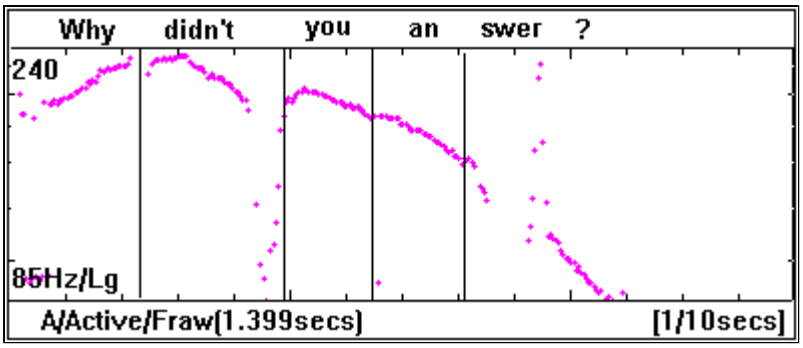
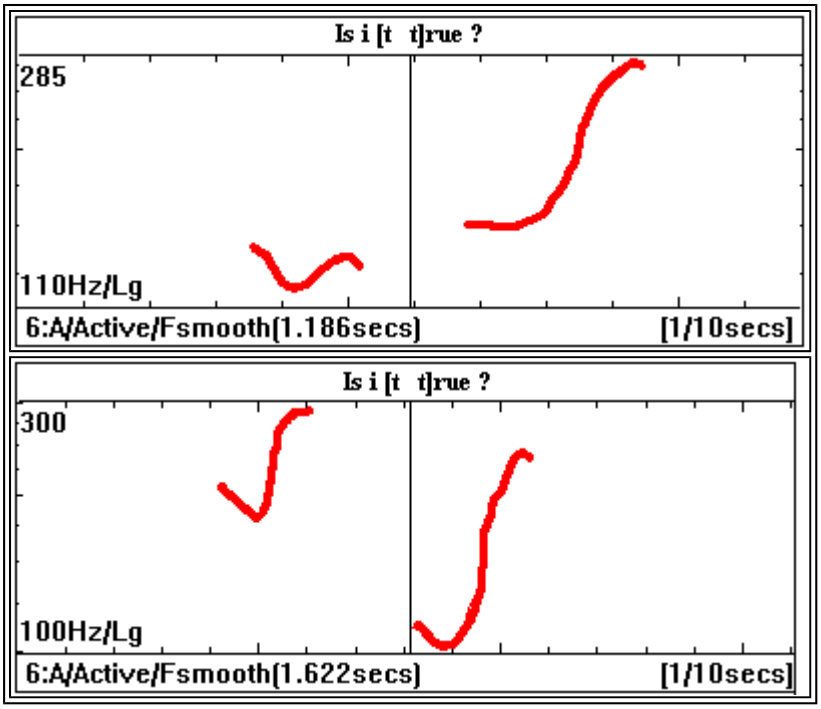


Figure 16 : Courbe Fondamentale *Fraw*, (non filtrée) de : *Why didn't you answer ?* affichée par Wincecil (1997). À la différence de la courbe reproduite en figure 14, celle-ci est affichée avec l'option : format "brut" (*raw*). On aperçoit quelques parasites ici et là qui peuvent malgré tout être utiles dans certains cas limites.

L'expérience montre que, même si pour certains cela peut paraître inimaginable, de nombreux apprenants ont de réelles difficultés à comprendre ce que veut dire : son élevé, aigu ou bas, grave. La courbe reproduite en figure 16, qui correspond à la question : "*Why didn't you answer ?"*, permettrait peut-être aux étudiants ayant du mal à s'approprier de telles références mélodiques, qu'une question ouverte (celles qui commencent, par exemple, par des adverbes interrogatifs en "w": *where*, *why*, *when*, etc.) commence souvent dans l'aigu pour pointer vers le grave à la fin. La comparaison avec une ou plusieurs transpositions visuelles d'une question fermée du style : "*Is it true ?*" (la réponse attendue est de type binaire : oui ou non), dont la courbe mélodique sera, elle, souvent ascendante, compléterait peut-être la démonstration. Dans les figures 17 et 18, l'affichage de gauche correspond à une simple question informative, qui indique que le locuteur n'émet aucune opinion sur la réponse qu'il attend, alors que l'affichage de droite, dont la deuxième partie repart du bas et se termine à un niveau inférieur à celui de la première, laisse passer une nuance d'étonnement dans la question, le locuteur semblant émettre quelque doute sur le fait que "cela puisse être vrai!".



Figures 17 et 18 : Deux versions de *Is it true ?*, affichages sous Wincecil (1997).

On comprend bien que l'affichage des courbes mélodiques pourrait certainement aider l'apprenant à progresser de façon déterminante, en lui proposant une autre approche de l'item à reproduire, en lui donnant la possibilité de suivre et de comprendre les intentions expressives qu'il contient, notamment

lorsqu'il lui sera demandé, dans le cas d'une simulation de conversation, de fournir une réplique adéquate. La reconnaissance d'enchaînements *fall-rise*, *rise-fall*, *fall-rise-fall* etc. (descente-montée, montée-descente, etc. de tonalité) ne sont pas seulement utiles pour le phonéticien mais aussi pour le simple ingénieur francophone envoyé, par exemple, aux États-Unis afin de préparer un contrat ou toute autre opération. Non seulement il pourra exprimer ses idées d'une façon plus explicite pour son auditoire, mais il saura aussi distinguer rapidement si son interlocuteur est favorable à ce qu'il propose ou s'il reste, au contraire, dubitatif, avec les changements radicaux sur le cours des choses et sur sa propre carrière, dans un sens ou dans l'autre, qu'on peut imaginer.

3.3 Spectrogramme - sonagramme

Dans un spectrogramme, l'axe horizontal est celui du temps et l'axe vertical est réservé aux fréquences. Une troisième dimension est ajoutée, opacifiant quelque peu l'affichage lorsque le spectrographe n'est pas assez finement réglé, qui correspond à l'amplitude : plus l'amplitude est élevée, plus le graphe s'assombrira à la hauteur correspondant à une fréquence donnée.

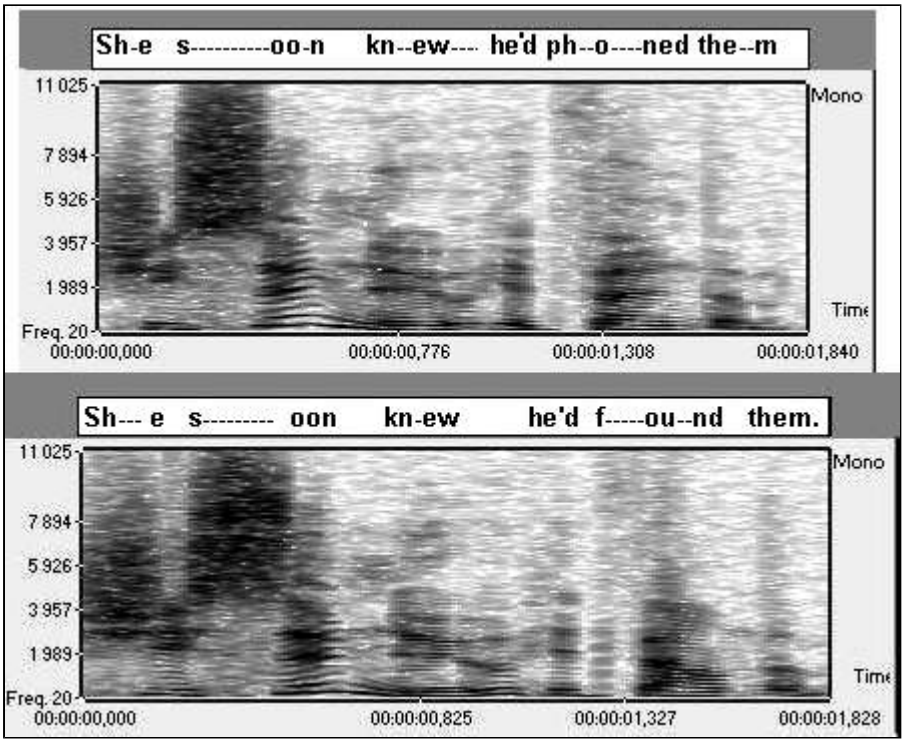


Figure 19 : Deux sonagrammes réalisés avec Sound Forge (1998), qui présente l'avantage de pouvoir choisir entre une représentation colorée ou non, pour séparer plus finement les différents niveaux d'amplitude. On peut également isoler les plages de fréquences. On peut, par ailleurs, ralentir le débit sans jouer sur la hauteur des sons, appliquer différents filtres pour atténuer ou supprimer les plosives et autres fricatives trop fortes, etc., et insérer des lignes de texte ou de commentaires.

Les deux spectrogrammes acoustiques, ou sonagrammes reproduits dans la figure 19 sont le résultat de la prononciation de deux courtes phrases dont le texte, inscrit en haut de chacun d'eux, comporte une unique variation jouant sur deux diphtongues différentes. Ce type d'opposition vocalique est souvent utilisé pour initier les étudiants aux secrets de la prononciation. La résolution du graphisme n'est pas assez fine pour laisser apparaître assez clairement ce qui fait l'intérêt particulier des sonagrammes : la possibilité de visualiser les harmoniques dont sont formés tous les éléments composant un item sonore complexe, comme l'avait conçu Fourier, dès le 19e siècle. On parviendra, avec une certaine habitude, à distinguer tous les phonèmes (unités minimales de son) de la phrase. On constate notamment que le |s| du mot *soon* est audible dans les régions hautes du spectre visible (entre 4000 et 12000Hz ici), et que sa traduction visuelle ne permet pas de discerner quelque couche d'harmoniques que ce soit. Il relève en cela d'un simple "bruit". Si la qualité du graphe était meilleure, les "plosives", cf. le |p| du nom *Peter*, laisseraient apparaître juste avant un vide (blanc), une zone verticale nette suivrait, annonçant une zone couvrant un

nombre important de fréquences. Les articulations centrales du même mot *soon* principalement, mais aussi, si on regarde bien, de *knew*, de *he'd*, de *phoned* et même de *them* paraissent, elles, plus organisées. Ces articulations laissent deviner un étagement de couches horizontales, empilées de façon plus ou moins serrée et qu'on pourrait compter si, de nouveau, la qualité du graphe était meilleure. Le nombre des harmoniques ainsi trouvé, et les hauteurs des plus caractéristiques parmi celles-ci permettent aux habitués de reconnaître les phonèmes qui leur sont associés. Chaque couche correspond en fait à ce qu'on appelle des "formants". On les compte de bas en haut et on les nomme F1, F2, F3 etc.. Naturellement, avant de pouvoir reconnaître les formants des phonèmes contenus dans un sonagramme représentant un texte de plusieurs mots, sur une échelle de fréquences montant à 12 ou 15 Khz, il sera nécessaire d'apprendre à les reconnaître isolément, sur différents sonagrammes présentant l'alphabet des phonèmes, avec un effet de zoom qui en détaillera les parties les plus reconnaissables.

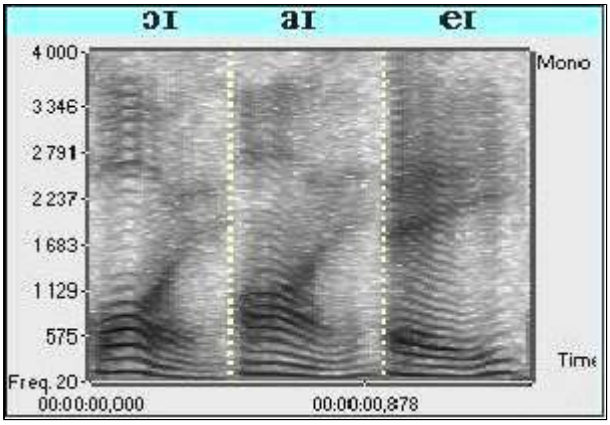


Figure 20 : Trois spectrogrammes de diphtongues, plages limitées à 4000Hz, sous Sound Forge (1998).

Pour illustrer ce point, la figure 20 ne montre que trois diphtongues de la langue anglaise et se limite volontairement à 4000Hz. On notera certaines similitudes. Le dessin du |ɪ| final de chaque diphtongue est assez semblable et permet de constater un éclaircissement parallèle entre environ 600 et 2000Hz. Il est intéressant de constater que les trois progressions vers le |ɪ| final se font de façon graduelle à partir des différents formants constituant la partie initiale de chacune des trois diphtongues.

En l'état actuel des choses, autant un tel affichage intéressera un spécialiste, autant il semble que leur déchiffrage demandera au néophyte trop de temps pour qu'il puisse en tirer un profit appréciable dans le cadre d'études de langues "ordinaires". Je doute que, même si on lui donne les outils pour le faire (zooms, recadrage, réglage des contrastes, etc.), l'apprenant moyen prenne le temps d'effectuer ce travail. Pourtant, si dans un avenir plus ou moins proche, ces réglages et effets de cadrages sont produits automatiquement (tout en restant modifiables) par un didacticiel de qualité, et si, en changeant la position de sa bouche, l'apprenant peut voir un affichage lui présenter, en temps réel, le sonagramme de tel ou tel phonème qu'il prononce, il n'est pas interdit de penser que, sentant un lien possible avec ce que j'appelais plus haut le côté insaisissable de la production orale, voyant qu'il peut avoir prise dessus, il y prenne goût et même qu'il puisse réellement améliorer sa prononciation.

3.4 Mingogrammes

Le recours aux "mingogrammes" permettrait probablement de bénéficier d'atouts semblables à ceux qui avaient été envisagés plus haut à propos des courbes mélodiques. L'exemple reproduit en figure 21, emprunté à un ouvrage d'initiation à la phonétique de Richard Lilly et Michel Viel (1993), correspond à l'enchaînement : "*I sent him a letter. - You didn't phone him ?*". Les trous entre les différentes formes visibles correspondent aux émissions non voisées.

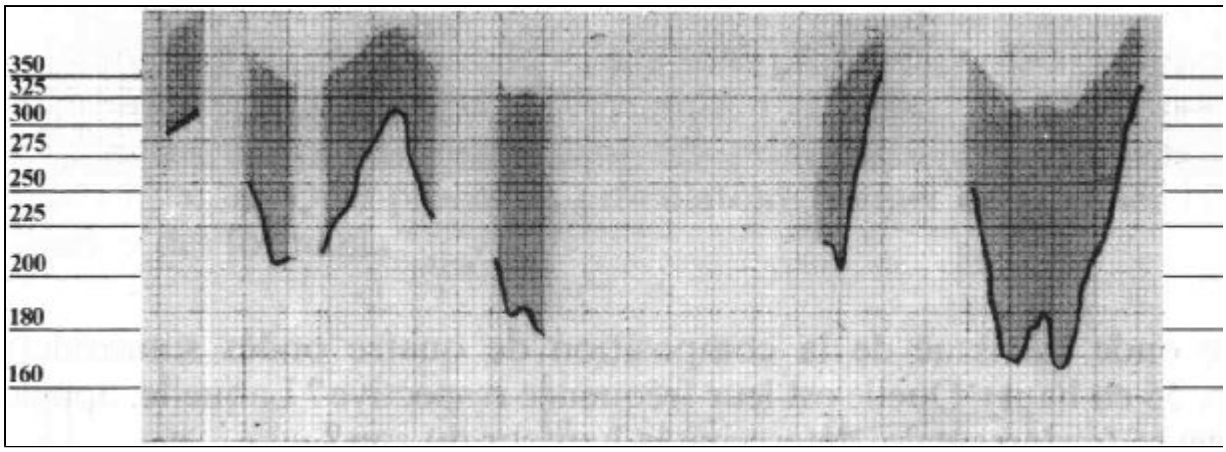


Figure 21 : Un exemple de mingogramme, reproduit de Lillly R. & Viel M. (1993, p. 53), correspondant à la phrase : *I sent him a letter. - You didn't phone him ?*

Par rapport à la courbe mélodique vue précédemment, on constate que le mingogramme ne se contente pas de schématiser la variation mélodique (fréquence fondamentale) d'un énoncé oral, il montre également l'amplitude des variations de la fréquence fondamentale. Si les données figurant sur ce type d'affichage sont plus riches, il faudra, en conséquence, que l'apprenant accepte de passer par une initiation un peu plus longue pour qu'il puisse en tirer parti. Mais ce qui a été dit du spectrogramme est probablement valable pour le présent affichage également : une automatisation des affichages en temps réel, avec comparaison constante entre le modèle et la production apprenante et possibilité d'interaction immédiate entre l'apprenant et le contenu de l'affichage, lui donnerait une possibilité de déborder les canaux habituels de la conscience et du raisonnement conceptuel pour parvenir à contrôler simultanément la position de tout son système phonatoire, pharynx, cavité buccale, palais, langue, cartilages, cordes vocales etc.

4. Ce qu'on aimerait pouvoir faire et faire faire

À la suite de ce que je viens de suggérer et au vu des différentes possibilités envisagées précédemment, pourquoi ne pas se demander ce dont il serait souhaitable de pouvoir disposer dans un bon logiciel de langues, même si cela demande un certain effort d'imagination et se rapproche plus de la projection que permet le rêve que d'une réalité prochainement accessible. L'idée précède souvent l'acte.

4.1 Analyse et guidage automatisés des productions orales

Des outils sont en voie de développement pour traiter les fichiers sonores de la même manière qu'on analyse les mots d'un texte et leurs assemblages, en leur attribuant une description fonctionnelle, sémantique et structurelle. Des procédures bénéficiant certainement des systèmes d'analyse automatisée du discours mais portant sur les sons et leur enchaînement, utilisant des bases de règles et de données rappelant un peu les modes de fonctionnement de l'I.A. (Intelligence Artificielle), permettront probablement un jour de faire qu'une production orale apprenante soit étalonnée avec suffisamment de sécurité pour que le système puisse lui proposer des améliorations possibles, des schémas préférentiels à imiter. Ces schémas, adaptés à sa voix, pourront lui être présentés visuellement sous forme de graphiques à recouvrir au plus près par les affichages provoqués par sa propre voix.

4.2 Affichage simultané de texte grâce à la reconnaissance vocale

La présence du texte, référent commun majeur à tous les types d'apprenants, peut fortement aider à étudier et à déchiffrer les courbes proposées. Or, la possibilité d'un affichage automatique du texte, auquel se rapporte chaque section d'une courbe au-dessus de celle-ci, n'est pas si inaccessible que cela et même très réalisable dès aujourd'hui. Les progrès actuels de la reconnaissance vocale permettent de l'envisager dès maintenant.

Il est vrai que la reconnaissance vocale est aujourd'hui encore trop axée sur un type de langue beaucoup

plus proche de l'écrit que de l'oral. Les systèmes actuellement disponibles dans le commerce servent essentiellement à récupérer des dictées de lettres, d'articles de journaux, des comptes-rendus d'analyses médicales ou des constats de dégâts dans le monde des assurances. L'ouverture sur l'oral fait partie des développements actuels dans ce domaine. Le traitement d'un énoncé oral inclut les nécessités propres à l'écrit mais doit en plus prendre en compte celles qui caractérisent la langue orale : les hésitations, les faux départs, les syntaxes hésitantes ou qui se modifient en cours de prise de parole, les différences de débit, les variations d'amplitude, la diversité des accents, etc.. La recherche sur la modélisation des comportements vocaux est en cours, elle vise à permettre un étalonnage plus rapide des profils vocaux et comportementaux, et donc à faire l'économie des procédures actuelles d'apprentissage, longues et fastidieuses. Lorsque de telles recherches auront abouti, il sera possible de traiter automatiquement et instantanément les éléments sonores d'un didacticiel et d'en proposer la transcription écrite dans des "boîtes" ou "bulles d'information" automatiquement mises à jour et accessibles via des zones sensibilisées sur l'écran de travail. Il en ira de même de l'accompagnement textuel des courbes sonores.

Il sera alors possible d'avoir une ligne spéciale se rapportant au modèle proposé par le didacticiel et une autre, séparée, alimentée par la production apprenante et affichée au dessus de sa transposition graphique. L'intérêt, pour l'apprenant, sera de pouvoir se faire une idée de la manière dont le système aura reconnu sa production et de pouvoir la modifier en conséquence. Certains logiciels commencent à oser présenter de telles fonctionnalités mais il est encore trop tôt pour juger leurs résultats suffisamment probants. Si on ajoute à cette indication les affichages de schémas-guides suggérés plus haut, on peut penser que les logiciels permettant l'apprentissage de la prononciation d'une langue étrangère et même sa remédiation ne sont pas irréalisables.

La transcription simultanée à l'écrit de la réponse orale apprenante permettra une analyse du contenu des productions orales apprenantes avec les mêmes possibilités que celles dont on dispose aujourd'hui pour traiter des réponses tapées au clavier, même s'il est vrai qu'il reste beaucoup à faire dans ce sens et que bon nombre de didacticiels ne se donnent pas la peine de mettre à profit les possibilités existantes. Ayant testé moi-même l'effet de l'intégration d'une telle procédure de reconnaissance vocale dans un environnement didacticiel que j'ai conçu, je peux dire, d'une part, que la chose est plus aisée qu'on ne le croit, d'autre part, que son effet en situation pédagogique est impressionnant et renouvelle fortement la motivation de l'apprenant.

4.3 Visualisation iconographique contextuelle

Diverses fonctionnalités d'aide visuelle animée et contextuelle, que l'on trouve disséminées et plus ou moins bien réussies dans certains produits existant déjà, doivent être généralisées et améliorées. Il est vrai que, tant que la qualité d'analyse des productions sonores apprenantes n'aura pas fait de progrès sensibles, leur utilité et surtout leur opportunité risque de n'être pas convaincante, mais dès qu'on pourra reconnaître dans une réponse orale un assemblage de formants typiques de tel phonème à un emplacement où on en attendait un autre, par exemple, ou quand il sera possible d'y distinguer une courbe intonative différente de celle qu'on prévoyait, il pourra être d'un grand secours à l'apprenant de se voir proposer un travail plus fin appuyé sur la visualisation d'une vidéo ou d'une séquence animée. L'animation graphique pourrait montrer l'évolution des positions de la langue, des lèvres, du flux de l'air expiré, etc. dans une courbe de la bouche au moment où on prononce tel phonème, en opposition à celui qui aura été émis à tort.

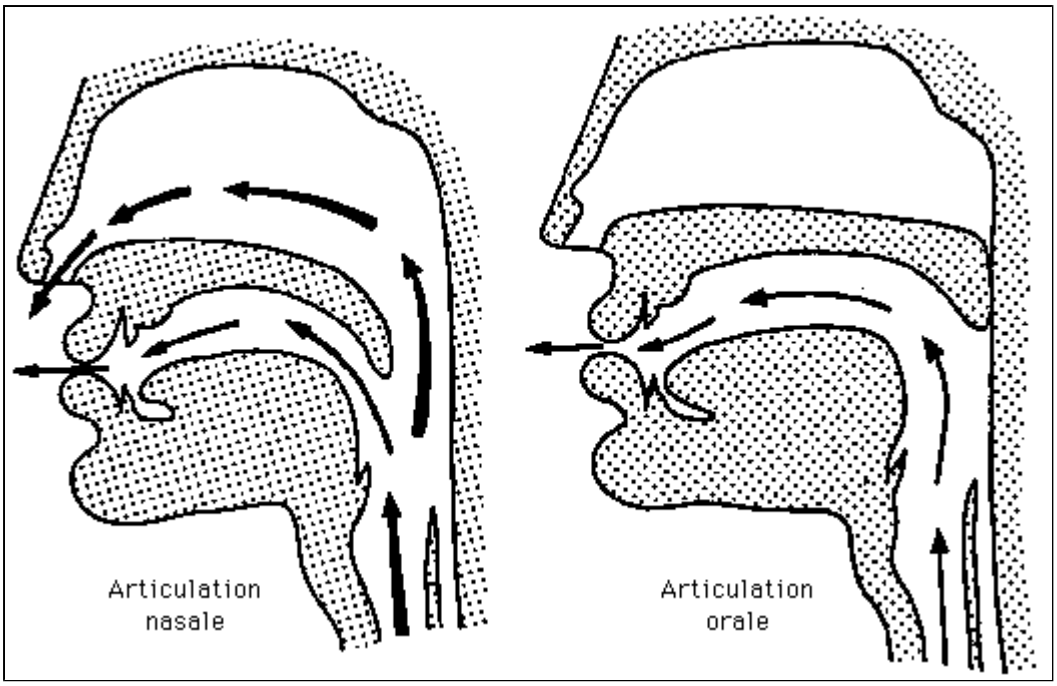


Figure 22 : Ce schéma est extrait d'une "Introduction à la phonétique" en cours d'élaboration sur le site Internet de l'Université de Lausanne (1999). Les illustrations sonores sont déjà accessibles sur ce site mais les graphiques ne sont pas animés.

Les schémas apparaissant en figure 22 pourraient être animés et accompagnés de plusieurs boutons sonores permettant d'apprécier de façon plus "parlante" les différentes conditions d'émission des sons. Même sur l'Internet, où les vitesses de transfert de données sont encore très limitées dans des conditions ordinaires de fonctionnement, les possibilités d'activation de fichiers sonores aux formats compressés MP3, MP4, AU ou autres sont désormais possibles. Les même possibilités existent pour exécuter en direct des fichiers graphiques animés peu gourmands comme les "GIF animés" (GIF = *Graphic Interface Format*)[3].

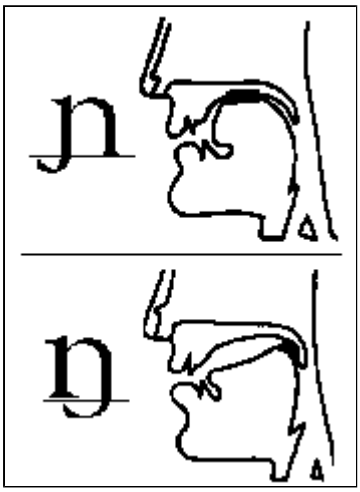


Figure 23 : Mêmes références pour ces graphiques que pour la figure 22, au site de l'Université de Lausanne (1999).

Les enseignants d'anglais savent que certains apprenants ont des difficultés sérieuses à contrôler l'usage du

voile du palais : des graphiques animés et sonores devraient aider à faire passer, pour ne prendre qu'un exemple parmi tant d'autres possibles, les différences de prononciation qui existent entre la consonne centrale du mot français "oignon" (voir haut de la figure 23) et celle du mot anglais "*singing"*, et la nuance qu'il y a par rapport à un autre mot comme *conglomerate*. Les différentes prononciations du son |r|, battu ou roulé, à l'espagnole, à la française, à l'anglaise, à l'écossaise etc. devraient en être améliorées d'autant.

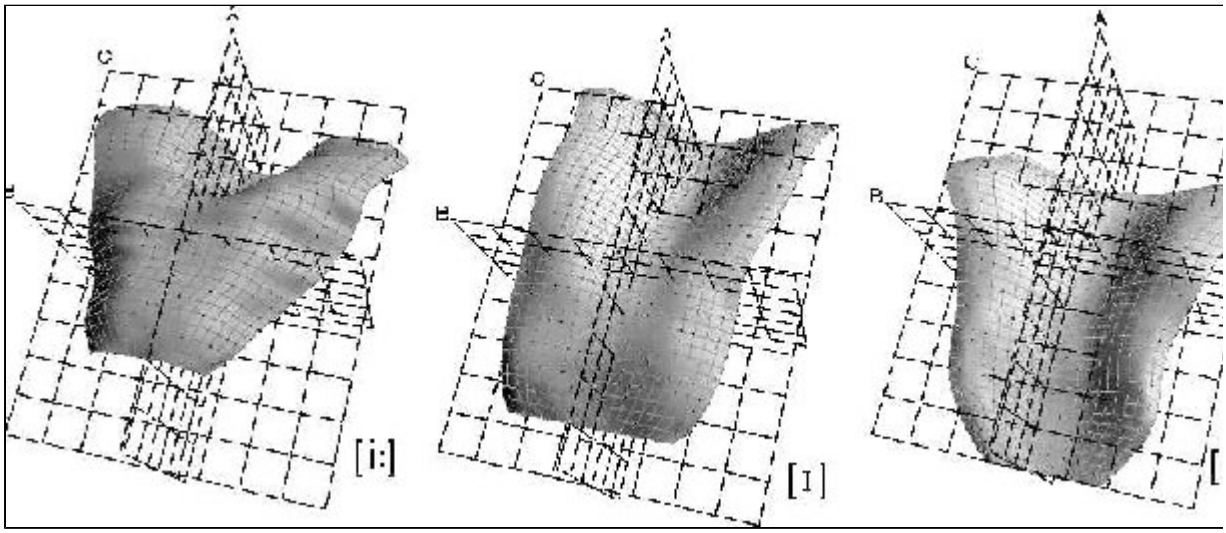


Figure 24 : Les graphiques ci-contre sont extraits du site du *Vocal Track Visualization Laboratory* (nd). On y trouvera quelques vidéos aux Rayons X annoncées dans le texte. Elles sont téléchargeables.

Une autre animation ciblant plus précisément le rôle de la langue, pourrait compléter ce qui précède et mieux mettre en valeur les différences flagrantes entre plusieurs prononciations de ce que bien des Français appellent "le" *i* ou "le" *a* en anglais. Des travaux sont réalisés dans ce sens et les graphiques reproduits ci-dessus donnent une idée de l'intérêt que peut être celui de visualiser de façon précise et claire les variations de position que la langue peut prendre suivant les sonorités (figure 24) même si, bien évidemment, il conviendrait de montrer les positions de la langue d'une manière différente, plus attrayante, "en contexte", dans une représentation plus globale d'un locuteur en train de parler. On peut, désormais, voir tel temple égyptien, aujourd'hui en ruine, reconstitué en 3D et l'observer sous tous les angles possibles, par un simple mouvement de la souris de bas en haut ou de gauche à droite. De même avec un corps humain en écorché, avec le fonctionnement du coeur, etc. dans une encyclopédie médicale. Le même type de présentation portant sur le fonctionnement du système phonatoire pourra certainement favoriser la perception de l'insaisissable dont il était question plus haut. Des vidéos réalisées aux rayons X durant la prononciation de phrases ou de mots isolés (cf. légende la figure 24) permettent de savoir précisément comment tout cela fonctionne, mais il semble préférable, pour des raisons de taille de fichiers et surtout de lisibilité, de présenter de tels mouvements sous formes d'animations graphiques.

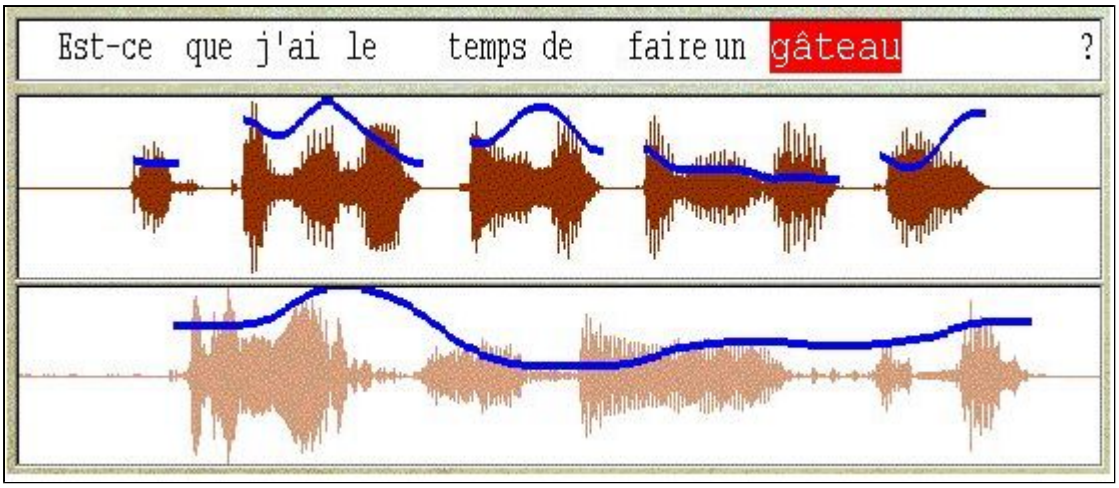


Figure 25 : Un des affichages possibles dans la dernière version de Talk To Me (1998).

J'ai inclu des réflexions voisines de celles que je viens d'évoquer dans plusieurs articles consacrés à certains logiciels axés sur la reconnaissance vocale, notamment à propos de la sortie de "Tell Me More" (Cazade, 1999a), et durant un séminaire que j'ai organisé à Paris Dauphine en avril 1999, centré sur ce sujet et j'ai eu la bonne surprise de constater que l'idée avait fait son chemin. Je ne peux véritablement penser qu'il y ait un lien de cause à effet entre ceci et cela et que mes suggestions aient été responsables de l'évolution qu'on peut constater dans les deux figures 25 et 26. Je pense plutôt que le *Zeit Geist*, l'esprit du temps, a dû faire son travail. Les deux graphiques joints sont issus d'un logiciel qui sort précisément au moment où je dois finir cet article[4] et montrent, pour le premier (figure 25) comment un affichage multiple (ici oscillogramme et courbe fondamentale) peut aider l'apprenant, beaucoup plus efficacement que le seul oscillogramme, à se repérer visuellement, spécialement dans la comparaison de l'intonation à imiter et de la sienne propre. Les deux courbes peuvent être affichées séparément ou en mode cumulé, au choix de l'apprenant.

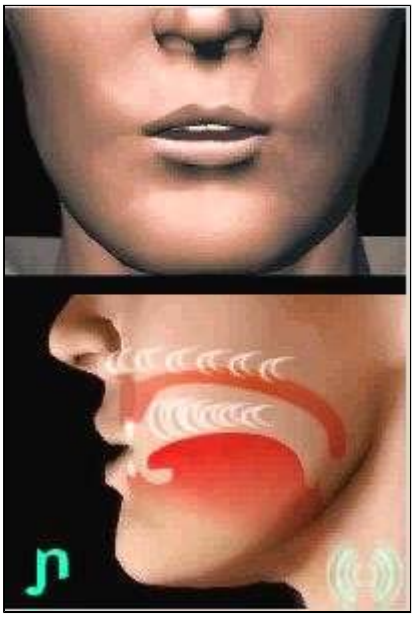


Figure 26 : Un autre affichage prévu dans la dernière version de Talk To Me (1998).

Le deuxième (figure 26) est emprunté à une des nombreuses animations graphiques qui accompagnent le logiciel et qui présentent, en 3D, les mouvements du bas du visage, simultanément de profil et de face, au moment de la prononciation des différents phonèmes. Il est intéressant de noter que l'animation montre :

1. le symbole phonétique concerné (cf. oignon),
2. la cavité buccale (sans la fosse nasale, pour des raisons de simplification) et la bouche en mouvement,
3. les mouvements de la langue,
4. le moment précis où les cordes vocales fonctionnent et
5. la progression de la propagation des ondes dans les cavités buccale et nasale.

L'effet d'ensemble est attirant et stimulant, même si certaines animations ou choix de sonorités exemples (le phonème central |A| de *but*, par exemple) pourraient être discutées.

Conclusion et suggestions

Il semble que les moyens mis en oeuvre pour favoriser la compréhension par les apprenants des mécanismes de la parole, et tout particulièrement de la production des sons en langues étrangères, soient en train de connaître une évolution sérieuse, qui reléguera probablement les pauvres oscillogrammes actuellement disponibles au rayon des antiquités. Beaucoup de choses sont possibles à partir de l'existant, pour peu qu'on l'améliore quelque peu, comme les éditeurs semblent le comprendre.

Il ne faut pas nier l'intérêt de certaines schématisations dont il a été question dans les lignes qui précèdent. Les choix proposés par tel logiciel de simplifier une partie des données à présenter à l'apprenant restent, semble-t-il, toujours intéressants. Bien repensés et recalculés afin de pouvoir prendre en compte la multiplicité des cartes vocales et autres interfaces sonores du type "puce intégrée au microprocesseur" (cf. les ordinateurs portables) présentes sur le marché aujourd'hui, ils fourniraient une aide simple et précieuse à bien des apprenants en langues dès maintenant. Le même principe de filtrage des données devrait permettre à d'autres courbes graphiques (spectrogrammes, mingogrammes, courbes mélodiques etc.) d'être un peu plus compréhensibles que ce que l'on trouve à l'heure actuelle dans les publications réservées aux spécialistes de la phonologie et de la phonétique.

Plutôt que de présenter toutes ces courbes exclusivement les unes des autres, pourquoi ne pas envisager de les proposer de façon cumulée, l'une au dessus ou à côté de l'autre, par exemple, ou même l'une par-dessus l'autre comme le fait un certain éditeur. Il devrait être possible de partager l'espace d'affichage d'une seule fenêtre en gardant un axe horizontal commun, afin de montrer des données différentes sur les parties haute et basse de la fenêtre d'affichage. La juxtaposition des données variables permettrait certainement de faire pressentir le rôle de certaines convergences ou divergences. Il serait peut-être même plus parlant d'envisager une superposition en différentes couleurs de ces données, avec des effets de transparence appropriés : une courbe montrant une progression des dynamiques, tenant compte des temps de montée et de fuite des sons, une autre courbe indiquant la progression des fréquences, par exemple. Une autre paire significative pourrait joindre une courbe mélodique à une courbe dynamique.

De tels affichages pourraient atteindre les apprenants moins spontanément aptes à déceler les fluctuations de la voix. Le choix des types de données pouvant être affichées simultanément devrait être laissé à l'apprenant, afin qu'il puisse estimer lui-même ce qui est le plus pertinent ou utile par rapport à son propre profil cognitif. L'étude des préférences d'affichage choisies par les apprenants serait très fructueuse tant pour faire avancer le produit que pour déterminer des attitudes, des profils d'écoute dominants. Les adeptes des sciences de l'éducation, les oralistes, les méta-cogniticiens pourraient trouver des pistes de recherche aussi passionnantes qu'utiles au développement de nos outils d'enseignement de langues.

Un point doit être mis en avant : la nécessité pour l'apprenant de pouvoir manipuler, influer au maximum sur les différents modes de restitution possibles de tous les items sonores (ceux qu'il aura produits et ceux qu'on lui proposera comme modèles à étudier). Parmi les possibilités de restitution à offrir pour mieux comprendre les affichages de courbes, la réécoute en bénéficiant d'un effet de ralenti ne modifiant pas la hauteur des sons (comme se contentent pourtant de le faire quelques didacticiels) devrait jouer un rôle de premier plan. Les tests que j'ai pu faire avec quelques produits[5], font penser que la chose est réalisable. La segmentation automatique, prenant en compte les baisses de volume ou d'intensité dans le flux sonore pour marquer l'endroit où il est possible d'insérer un repère ou un silence plus ou moins grand, propre à faciliter la compréhension ou la répétition partielle ultérieurement, est également une piste à creuser et une fonctionnalité qui devrait être généralisée. Le didacticiel LAVAC (1999) permet déjà ce travail aussi bien

pour les productions sonores que pour l'étude de documents vidéos.

Pour pouvoir tirer quelque parti que ce soit, aussi bien des courbes sonores de natures diverses qui ont été envisagées que des animations graphiques auxquelles il vient d'être fait allusion, il semble bien qu'une initiation aux bases élémentaires de la phonétique soit indispensable, en préalable ou/et en accès libre au moment où l'apprenant le jugera utile. S'il est donné à l'apprenant de comprendre avec les yeux autant qu'avec les oreilles les différences qui existent entre un bruit et un son, entre les sonorités simples et complexes, d'aborder les principes de la phonation, le rôle des échanges excitateurs / résonateurs (buccal, labial, nasal), l'alphabet des phonèmes mentionné supra, les différences entre sonorités postérieures, centrales ou antérieures, entre voyelles arrondies ou non, entre les différents degrés d'aperture possibles, entre les consonnes sifflantes ou fricatives, plosives et occlusives, etc., et si tous ces éléments lui sont proposés avec la possibilité de choisir, ensuite de revoir ce qu'il aura choisi, d'en piloter lui-même la présentation, de voir la transcription de sa propre production juxtaposée ou projetée par-dessus le modèle étudié, alors les explications phonétiques deviendront aussi indissociables des logiciels de langues que le sont les cartes vocales ou les oscillogrammes aujourd'hui.

Références

Bibliograhie

Cazade, A., (1999a). "Pour intégrer des outils pédagogiques multimédias dans l'enseignement de l'anglais : *Tell Me More (Auralog)*", *Les Cahiers de l'Apliut*, vol.18,4, juin 1999.

Lilly, R. & Viel, M. (1993). *Initiation raisonnée à la phonétique de l'anglais*. Paris : Hachette.

Université de Lausanne (1999). *Cours d'initiation à la phonétique*. Lausanne, Suisse : Université de Lausanne. Consulté en août 1999 : http://www.unil.ch/ling/phonetique/api.html

Logiciels de traitement des sons

Cubase Audio (version 3.5, 1997). Consulté en novembre 1999 : http://www.steinberg.net

Sound Forge (1991, version 4.5: 1998) - Outil quasi professionnel d'acquisition et de manipulations de sons de toutes natures. Sonicfoundry et RealNetworks. Consulté en novembre 1999 : http://www.Sonicfoundry.com

Wave Studio (version utilisée : 3.12, 1997, dernière version : 4.06, 1999). Logiciel d'acquisition et de manipulations simples de fichiers sonores. Il est livré avec toute carte "Sound Blaster", Creative Labs, qui est le "standard" des cartes sonores. La dernière version a des facilités de sélection intéressantes. Consulté en novembre 1999 : http://www.soundblaster.com

Wincecil (version 2.2 : 1994, dernière version : 1997). Logiciel sous PC, téléchargeable et gratuit. Ne traite que 3 secondes maximum. Aussi accessible depuis le site du SIL (nd). Consulté en novembre 1999 : http://www.jaars.org/icts/software/cecil/wincecil/wc22.zip

Didacticiels

LAVAC (version 4.03.i-1999). LAVAC (Laboratoire Vidéo Actif Comparatif) a été conçu par Tony Toma. Éditeur C3 : Montpellier.Consulté en novembre 1999 : http://www.alizes.fr/cp3i.

Speaker (version 3.0 et 3.1 1997, version 4.0 1999). Neuroconcept : Suresnes. Consulté en août 1999 : http://www.neuroconcept.com

Talk To Me (1998). Auralog : Voisins-le-Bretonneux. Consulté en août 1999 : http://www.auralog.fr

Tell Me More (version testée : English, advanced level, 1998). Auralog : Voisins-le-Bretonneux. Consulté en août 1999 : http://www.auralog.fr

VOICEbook (1998). Englishear System : Paris. Consulté en août 1999 : http://www.voicebook.com

Sites Internet

SIL (nd). Le site de SIL International (anciennement Summer Institute of Linguistics) propose de nombreux outils, polices phonétiques, et logiciels téléchargeables et gratuits (dont Wincecil) ou simplement référencés. Dallas, TX, États-Unis : SIL International. Consulté en novembre 1999 : http://www.sil.org/computing/catalog/

Vocal Track Visualization Laboratory (nd). Site du *Vocal Track Visualization Laboratory*. University of Maryland : Baltimore, États-Unis. Consulté en août 1999 : http://som1.ab.umd.edu/~mstone/lab.html ou http://iacl.ece.jhu.edu/agg/projects/vtv.html

Références complémentaires non référencées dans le texte de l'article

Bibliograhie

Cazade, A. (1998). "Souplesse et contrainte du tout numérique". In *Les laboratoires multimédias*. Revue *Les Dossiers de l'ingénierie éducative*, 27, septembre 1998. Paris : Centre National de Documentation Pédagogique (CNDP). Consulté en novembre 1999 : http://www.cndp.fr/DOSSIERSIE/27/P2_1.pdf

Duchet, J.L. (1981). *La Phonologie*, Collection "Que Sais-je ?" Paris : Presses Universitaires de France.

Keller, E. (dir.) (1994). *Fundamentals of Speech Synthesis and Speech Recognition*. John Wiley and Sons.

Laver, J. (1994). *Principles of phonetics*. Cambridge : Cambridge University Press.

Lieberman, P. & Blumstein, S. (1988). *Speech physiology, speech perception and acoustic phonetics*. Cambridge : Cambridge University Press

Pierrehumbert, J.B. & Hirshberg, J. (1990). "The meaning of intonational contours in the interpretation of discourse." In *Intentions in Communication*, Cohen, Morgan & Pollack (dirs.). Cambridge, Mass : MIT Press.

Roac, P. (dir.)(1992). *Computing in linguistics and phonetics; introductory readings*. Londres : Academic Press Ltd.

Roac, P. (dir.) (1991). *English Phonetics and Phonology - A practical course*. (+ student's book + cassettes). Cambridge : Cambridge University Press.

Selkirk, E.O. (1984). *Phonology and syntax : The relation between sound and structure*. Cambridge, MA : MIT Press.

Shisler, B.K. (1997). *Dictionnaire des phonestèmes anglais - qui jouent un rôle important dans la formation des mots en anglais*. Consulté en août 1999 : http://www.geocities.com/SoHo/Studios/9783/phond1.html

Titze I.R. (1994).*Principles of Voice Production*. Prentice Hall.

Wells, J.C. (1990). *Longman Pronunciation Dictionary*. Londres : Longman.

Logiciels de traitement des sons

BitScope (nd). Un oscilloscope digital, relativement bon marché, également analyseur logique et outil d'acquisition. Consulté en novembre 1999 : http://www.bitscope.com/about/applications.html

IPOX (nd). Synthétiseur de parole, développé à l'université d'Oxford. Système utilisant Yorktalk. Consulté en août 1999 : http://www.phon.ox.ac.uk/~public/IPOX/ipox.htm

KPE (nd). Synthétiseur de voyelles. Consulté en août 1999 : http://svr-

www.eng.cam.ac.uk/comp.speech/Section5/Synth/klatt.kpe80.html

Phthong (nd). Logiciel d'initation à la phonétique de l'anglais, téléchargeable. Consulté en novembre 1999 : http://www.chass.utoronto.ca/~stairs

Rsynth (version 2.0, nd). Synthèse vocale, graticiel. . Consulté en novembre 1999 : ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/synthesis/rsynth-2.0.tar.gz

SFS (nd). Logiciel sous Unix et PC. University College London. Consulté en novembre 1999 : http://www.phon.ucl.ac.uk/

Signalize (nd). Logiciel sous Mac avec démonstrations téléchargeables. Un des logiciels les plus utilisés sur ce support. Consulté en novembre 1999 : http://agoralang.com/signalyze.html

SoundScope (nd). Logiciel sous Mac. Utilisé en phonologie, enseignement et recherche. Consulté en novembre 1999 : http://www.speech.cs.cmu.edu/comp.speech/Section1/Labs/soundscope.html et http://www.gwinst.com/web-pages/SoS.html

Speech Lab (nd). Logiciel sur la phonétique de l'anglais, fabriqué en Allemagne par Media enterprise - Trier. Consulté en août 1999 : http://www.Mediaenterprise.de

Wavesmith (nd). Analyseur de formes sonores. Version démo téléchargeable. Consulté en novembre 1999 : http://www.wavebuilder.com

Winpitch (nd). Logiciel sous PC. Démos téléchargeables. Consulté en août 1999 : http://www.winpitch.com/

Winspeech (nd). Logiciel de synthèse vocale, graticiel. Consulté en novembre 1999 : http://www.pcww.com

Sites Internet

CSLU (1997). Analyse de spectrogrammes. Center for spoken language understanding. États-Unis : . University of Oregon. Consulté en août 1999 : http://cslu.cse.ogi.edu/tutordemos/SpectrogramReading/spectrogram_reading.html

Dillon, G.L. (1999). Ressources diverses en phonétique. Washington, DC, États-Unis : Université de Washington. Consulté en décembre 1999 : http://faculty.washington.edu/dillon/PhonResources/PhonResources.html

ILG (nd). Images synthétisées de visages en train de parler et mouvements des lèvres. Institut du Langage de Grenoble. Grenoble, France : Université de Grenoble. Consultés en août 1999 : http://ophale.icp.grenet.fr/2.1.html et http://ophale.icp.grenet.fr/2.6.html

Ladefoged, P. (nd). Pages personnelles. Los-Angeles, CA, États-Unis : UCLA. Consulté en août 1999 : http://www.humnet.ucla.edu/humnet/linguistics/people/ladefoge/ladefoge.html

MAS (1998). Un modèle d'analyse des formes sonores complexes de voyelles par Michael A. Stokes. Indianapolis, IN, États-Unis : MAS Enterprises. Consulté en août 1999 : http://www.indy.net/~masmodel

Sentence Processing Lab (nd). Quelques articles en ligne et données bibliographiquesLawrence, KS, Etats-Unis : University of Kansas. Consulté en août 1999 : http://ondas.splh.ukans.edu

Université de Leeds (nd). *Speech Visualisation Tutorial*. Document, en anglais, d'initiation à la phonétique. Leeds, GB : Leeds University. Consulté en août 1999 : http://www.psyc.leeds.ac.uk/research/cogn/speech/tutorial/index.htm

Université de Lund (nd). Tutoriel. Lund, Suède : Lund University. Consulté en août 1999 : http://www.ling.lu.se/research/speechtutorial/tutorial.html

Notes

[1] En recourant essentiellement à un mode de calcul mathématique qu'on appelle la "transformée de Fourier", ou une de ses variantes, du nom du mathématicien français du 19o siècle qui en est l'auteur.

[2] Une fréquence correspond au nombre de cycles qu'effectue une vibration pendant un durée donnée. L'unité de fréquence habituellement choisie est le Hertz (Hz), et correspond à un nombre de cycles par seconde.

[3] La technique Shockwave, la programmation en langage Java permettent même un assemblage multimédia très sophistiqué. Certains didacticiels déjà présents sur le marché mettent en avant quelques vidéos ou animations graphiques qui, hélas, ne sont pas bien convaincantes. "Parlons anglais" (Nathan) est dans ce cas.

[4] "Talk to me" (1998) La version en question est sortie en septembre 99. J'avais envoyé un des articles visés à cet éditeur en mai 99.

[5]Quelques essais avec Wincecil (1997) ont été assez intéressants à des vitesses ralenties par un facteur 1,3 ou même 1,5. Certaines phrases particulièrement rapides sont restées assez reconnaissables avec un facteur 1,7. Sound Forge (1998) ou même "Cubase audio" (1997) sont également très intéressants.

[Note de la rédaction] Le lecteur désireux de poursuivre ses lectures dans le domaine de l'application des méthodes de traitement automatique de la parole à l'aide à l'apprentissage d'une langue pourra aussi se reporter aux articles suivants parus dans la revue américaine Language Learning & Technology (LLTJ), avec laquelle nous collaborons :

- Chun, D.M. (1998). "Signal Analysis Software For Teaching Discourse Intonation". *Language Learning & Technology (LLTJ),* vol.2, 1, juillet 1998. pp 61-77. Consulté en décembre 1999 : http://polyglot.cal.msu.edu/llt/vol2num1/article4/index.html
- Ehsani, F & Knodt, E. (1998). "Speech Technology In Computer-Aided Language Learning: Strengths And Limitations Of A New Call Paradigm". *Language Learning & Technology (LLTJ),* vol.2, 1, juillet 1998. pp. 45-60. Consulté en décembre 1999 : http://polyglot.cal.msu.edu/llt/vol2num1/article3/index.html
- Eskenazi, M. (1999). "Using Automatic Speech Processing For Foreign Language Pronunciation Tutoring: Some Issues And A Prototype". *Language Learning & Technology (LLTJ),* vol. 2, 2, janvier 1999. pp. 62-76. Consulté en décembre 1999 : http://polyglot.cal.msu.edu/llt/vol2num2/article3/index.html

Signalons, enfin la constitution en 1999 du groupe de travail INSTIL (SIG Integrating Speech Technology in (Language) Learning) sur cette même tématique, groupe commun à l'assocation nord-américaine CALICO et l'association européenne EUROCALL.

- INSTIL (1999). Site du groupe de travail INSTIL. Dundee, GB : Université d'Abertee. Consulté en décembre 1999 : http://dbs.tay.ac.uk/instil/

A propos de l'auteur

Alain CAZADE est co-responsable du C.R.L. (Centre de Ressources en Langues) multimédia de L'Université Paris IX Dauphine et de l'"Observatoire des Technologies nouvelles" de la SAES (Socié té des Anglicistes de l'Enseignement Supérieur). Anciennement responsable pédagogique, à Paris XIII, du CFIPE (Centre de formation à l'informatique pédagogique) de l'Académie de Créteil. Auteur d'un logiciel auteur hypermédia de recherches sur l'apprentissage des langues: "Help Yourself " (sous Windows 3x & 9x - non commercialisé).
Courriel : cazade@dauphine.fr
Adresse: Université Paris IX Dauphine, Pl. Ml Delattre de Tassigny, 75775 Paris Cedex 16, France.