

GÉNÉRATION ET GESTION DE MODÈLES GRAPHIQUES EN PHONÉTIQUE ACOUSTIQUE

Roussi Nikolov
Université de Plovdiv Paissii Hilendarski

COMPUTER-AIDED GENERATION AND MANAGEMENT OF ACOUSTIC MODELS

Russi Nikolov
Paisii Hilendarski University of Plovdiv

The new phonetic module of TREFL – a portable, versatile and efficient computer tool – was designed in view of the frequent need of language researchers and students for accessing, collecting, manipulating and representing acoustic data. It allows, in particular, successful management of multilingual collections of numerical speech models and graphical representations for direct visual comparison between them or with real-time acoustic analysis results. The latter type of comparison is particularly interesting when it concerns the speech of language learners.

Key words: experimental phonetics, acoustic data visualization and comparison

1. La modélisation acoustique: son rôle et les types de représentations graphiques

L'objet de cet article est de présenter une réflexion méthodologique portant sur l'utilisation d'un nouveau logiciel de génération et de gestion de graphiques acoustiques, avec une ergonomie qui maximise la facilité de leur utilisation comme modèles, au double sens du terme : d'une part, en tant que représentations simplifiées et linguistiquement pertinentes de la réalité acoustique et, d'autre part, en tant qu'objets de référence, notamment de comparaison et d'imitation en matière d'évaluation et de correction phonétique.

Quant aux critères de pertinence d'un modèle acoustique, il y a lieu de distinguer ici entre deux niveaux de modélisation, celui de la langue et celui du locuteur. Cette distinction est justifiée en raison de la nature

stochastique des données relatives aux traits phoniques linguistiquement pertinents. Toutes les caractéristiques phoniques des messages oraux produits par un locuteur lui sont propres et participent de son identification vocale, certaines d'entre elles étant linguistiquement importantes de par leur fonction d'identification des unités phoniques et, par là, des unités de sens. La question se pose de savoir si la modélisation linguistique sous forme d'une bibliothèque de modèles individuels linguistiquement pertinents se justifie didactiquement comme une alternative au modèle plus abstrait encore qu'est la norme linguistique. La réponse est positive à condition d'admettre 1) le rôle des bases de données très volumineuses comme une forme de modélisation en soi et 2) le principe didactique d'individualisation des parcours d'apprentissage, principe d'autant plus valable ici qu'il s'agit de didactique de l'oral: *“There’s a very interesting theory that if you want to improve your pronunciation, you should choose one person you want to sound like, and you basically copy that person.”*

Les représentations graphiques des caractéristiques acoustiques de la parole que ce logiciel appelé *TREFL* peut produire se divisent, en premier lieu, en deux types : des graphiques (fig. 1) et des grilles de référence (fig. 2). Le critère de cette distinction est purement formel : chaque ligne (horizontale ou verticale) d'une grille de référence correspond à une quantité instantanée, par exemple la valeur, mesurée à un moment donné, du premier formant d'une voyelle ; une telle valeur donne lieu à une ligne horizontale dans le système de coordonnées du spectrogramme ou à une ligne verticale dans le système de coordonnées du spectre vocalique. Par opposition, un graphique acoustique peut avoir la forme d'une représentation 3D (le spectrogramme), d'un nuage de points (valeurs formantiques dans l'espace F2/F1), d'une courbe plane (celles par exemple représentant l'évolution de la fréquence fondamentale ou de l'intensité en fonction du temps) ou de plusieurs courbes dans un même espace plan (les traces formantiques). Les objets acoustiques (graphiques et grilles) générés, stockés et gérés par *TREFL*, visuellement et numériquement, sont issus de données soit directes, soit traitées statistiquement ; dans les deux cas, ces objets sont considérés comme des modèles pour deux raisons: 1) ils peuvent servir de références dans des situations diverses de recherche ou d'apprentissage et 2) les caractéristiques acoustiques de la parole extraites et utilisées sont soit linguistiquement pertinentes, soit identificatrices du locuteur, ce dernier cas étant intéressant en matière d'imitation artistique ou de phonétique légale (Olsson 2008: 11 – 17).

Afin d'avoir un accès facile et rapide aux objets graphiques, ils sont automatiquement structurés (lors de leur enregistrement) et filtrés (lors de

leur utilisation ultérieure) selon quatre critères : langue (français, anglais, bulgare...), type d'objet (graphique ou grille d'écran), graphique et fenêtre actifs ; les deux derniers critères déterminent respectivement la nature du graphique de premier plan (spectrogramme, traces des formants, spectre...) et le logiciel d'analyse acoustique utilisé (*Speech Analyzer, Real-Time Spectrum, Real-Time Spectrogramme*, etc.). L'association automatique par défaut de ces méta-informations aux objets graphiques n'empêchera pas l'utilisateur, au besoin, de réorganiser/filtrer manuellement les données.

La génération et la gestion de modèles graphiques de référence de nature acoustique comme méthode de recherche et didactique est un vaste domaine dont les limites sont ici établies en vue de l'utilisation du logiciel *TREFL*. Par comparaison avec son prédécesseur *SARP*, il ne s'agit plus de se limiter aux seules voyelles ni aux représentations traditionnelles (parfois simplistes) F1/F2 ; tous les graphiques/grilles de référence sont gérables et superposables. La méthodologie correspondante implique l'examen d'une autre question connexe et plus large, à savoir la constitution et la gestion de ressources linguistiques orales avec optimisation des collections de modèles de prononciation suivant deux critères majeurs : l'exhaustivité (relativement à des critères établis de norme de prononciation) et la non-redondance.

2. L'analyse acoustique de la parole : étapes et outils

L'analyse acoustique de la parole se décompose en trois étapes. Elle commence par une transformation mathématique qui consiste à représenter la fonction du signal acoustique (présente sous forme d'une suite numérique obtenue par la digitalisation de l'onde sonore) comme la somme de fonctions trigonométriques simples (cf. *analyse harmonique*, ou *analyse de Fourier*). Cette analyse mathématique s'interprète physiquement comme la décomposition du phénomène vibratoire complexe en vibrations simples. À la fin de cette étape, on obtient le spectrogramme (ou, ce qui revient au même, les spectres successifs). Tout comme la forme d'onde, le spectrogramme est une représentation graphique intégrale du son, en ce sens qu'elle contient par principe des informations suffisantes pour convertir inversement le graphique en son.

La deuxième étape de l'analyse acoustique de la parole est celle de la modélisation acoustico-phonétique, au sens à la fois conceptuel et matériel du terme. Modéliser acoustiquement un objet ou un phénomène de parole signifie, d'abord, de concevoir un système de traits acoustiques qui décrit le mieux possible cet objet ou phénomène (c'est-à-dire d'une façon à la fois économique et pertinente), et de donner ensuite une forme matérielle,

optimale elle aussi, à ce système, en l'occurrence par représentation visuelle. Par exemple, la mélodie de la parole est modélisée par la fréquence fondamentale et représentée par le graphique du F0 en fonction du temps; les voyelles sont modélisées à partir de leur structure formantique et souvent représentées dans l'espace (F1, F2) ou (F1, F2, F3).

La troisième et dernière étape de l'analyse acoustique de la parole, celle qui précède immédiatement l'interprétation des résultats, est l'évaluation des données analytiques par comparaison avec des données de références correspondant au modèle utilisé. Cette évaluation peut être numérique, basée parfois sur des méthodes statistiques sophistiquées et rigoureuses, mais aussi visuelle – très pratique et utile, surtout en phonétique corrective –, par superposition de graphiques du même type.

Pour pallier le manque de fonctionnalités des logiciels d'analyse acoustique les plus complets et les plus utilisés tels que *Praat*, *Speech Analyzer*, *Speech Filing System*, *WinPitch* etc., notamment aux étapes 2 et surtout 3 ci-dessus de l'analyse acoustique de la parole, les spécialistes du langage oral au laboratoire *LAPS* ont entrepris de développer, sous la direction de J.-Y. Dommergues, un outil informatique multifonctionnel (intégré à *TREFL*) qui crée un environnement de travail plus efficace et plus cohérent, grâce respectivement aux nouvelles fonctions et aux nouvelles interactions suivantes.

Les logiciels mentionnés peuvent superposer des graphiques de type différents correspondant à un même énoncé oral, mais il est tout aussi intéressant, sinon plus, de superposer des graphiques du même type et correspondant à des réalisations différentes d'un même énoncé, par exemple, les traces des formants pour le locuteur A dessinés sur le spectrogramme pour le locuteur B. Il en va de même pour les spectres, les F0, les intensités, les F1/F2, etc., avec la possibilité de jouer avec les couleurs et les transparences.

TREFL permet aussi d'enregistrer un spectre ou un spectrogramme obtenu sous *Speech Analyzer* et de les utiliser ensuite comme références sous *Real-time Spectrum* et *Real-Time Spectrogram*, respectivement, qui sont des logiciels spécialisés d'analyse acoustique en temps réel. Les échelles et les dimensions des graphiques sont normalisées automatiquement.

Une troisième possibilité nouvelle consiste à dessiner librement sur l'écran de l'ordinateur (électroniquement, bien sûr, et plus facilement qu'avec un feutre) des grilles utilisateurs, c'est-à-dire des lignes horizontales et/ou verticales, de couleurs différentes, globalement ou associées à un type de graphique particulier, à les enregistrer, ouvrir et modifier. Des lignes horizontales sur les spectrogrammes et verticales sur

les spectres, correspondant à des formants vocaliques de références sont des moyens tout aussi simples qu'efficaces en matière de techniques d'entraînement articulatoire par feedback visuel des traits acoustiques linguistiquement pertinents.

Les fonctions du logiciel *SARP* aussi ont été intégrées à *TREFL*, avec plusieurs optimisations au niveau de l'ergonomie, ce dernier logiciel étant initialement conçu comme un outil de gestion de bases de données. *TREFL* est ainsi devenu, dans sa version actuelle, une extension fonctionnelle importante à *Speech Analyzer*, *Real-time Spectrum* et *Real-time Spectrogram*, ainsi que, dans une moindre mesure, à tout autre logiciel d'analyse acoustique.

3. Le module phonétique de *TREFL* – traitement de données brutes et analytiques

Le nouveau module phonétique de *TREFL* est une sorte de gestionnaire de ressources phonétiques et de logiciels apportant chacun des fonctions complémentaires dans un système cohérent d'analyse acoustique de la parole. Or l'originalité de *TREFL*, comme outil de recherche en phonétique expérimentale, mais aussi d'aide à l'acquisition autonome de la prononciation, consiste dans l'automatisme, la souplesse et l'ergonomie des opérations de superpositions de toutes sortes de graphiques et grilles de référence.

3.1. Gestion de graphiques à l'aide de *TREFL* – enregistrement et accès, superpositions et réglage de la transparence

Des logiciels tels que *Speech Analyzer*, *Praat*, *Speech Filing System*, *Real-time Spectrum* et *Real-time Spectrogram* sont des instruments haut de gamme en matière d'analyse acoustique pointue et polyvalente, en temps réel ou différé. *TREFL* offre une fonction inédite, à savoir la superposition de graphiques et de grilles de référence. Il ne s'agit pas, comme c'est le cas pour *Speech Analyzer* et *Praat* par exemple, de combiner des systèmes de coordonnées ayant en commun l'axe du temps pour mixer des graphiques de type différent ; il est autrement plus intéressant de superposer des graphiques d'un seul et même type et d'un seul et même segment de parole qui a été produit soit par un seul et même locuteur dans des situations de communication différentes, soit par des locuteurs différents, par exemple un apprenant et un locuteur natif.

La copie d'écran reproduite à la figure 1 illustre le contraste entre les traces formantiques du mot bulgare *napu* /pa'rɪ / (signifiant *argent*) et du mot français *Paris* /pa'Ri/. Le fait est qu'un francophone associera très

facilement la forme phonique de ce mot bulgare au mot français de constitution phonémique identique (au point de se demander, lors d'un premier séjour en Bulgarie, comment se fait-t-il qu'on y parle si souvent de la belle ville de Paris). La comparaison des deux productions est cependant très instructive en matière de prononciation (avec ou sans accent étranger).

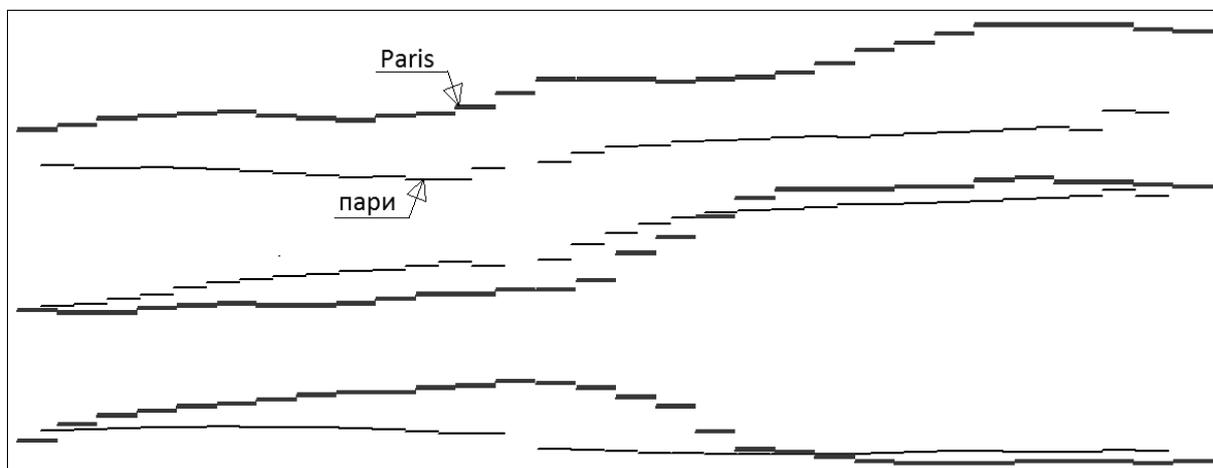


Fig. 1. Traces formantiques du mot bulgare пари /pa'ri/ et du mot français Paris /pa'ri/

3.2. Les grilles de référence

Cette fonction de *TREFL* permet de dessiner librement sur l'écran de l'ordinateur des lignes verticales et/ou horizontales de couleurs différentes, de les effacer ou de les déplacer, ainsi que de sauver les grilles ainsi obtenues en les organisant dans des bases de données de façon à y avoir un accès facile et quasi automatique en fonction du contexte (logiciel et type de graphique), de la langue, du locuteur, de l'unité phonétique, etc. La figure 2 contient une grille de lignes verticales dont chacune correspond à une valeur formantique, ainsi qu'un graphique de référence – le spectre de la voyelle /i/ – et le spectre de la même voyelle affiché en temps réel dans le logiciel *Real-time Spectrum* et représenté ici à un moment donné de la production de la voyelle par un apprenant du français – très avancé à en juger par la reproduction assez fidèle des trois (voire quatre) premiers formants vocaliques. Pour évaluer en temps réel le spectre produit par l'apprenant, ce dernier peut choisir comme modèle la grille ou le spectre de référence, ou afficher les deux à la fois (comme c'est le cas pour la figure 2).

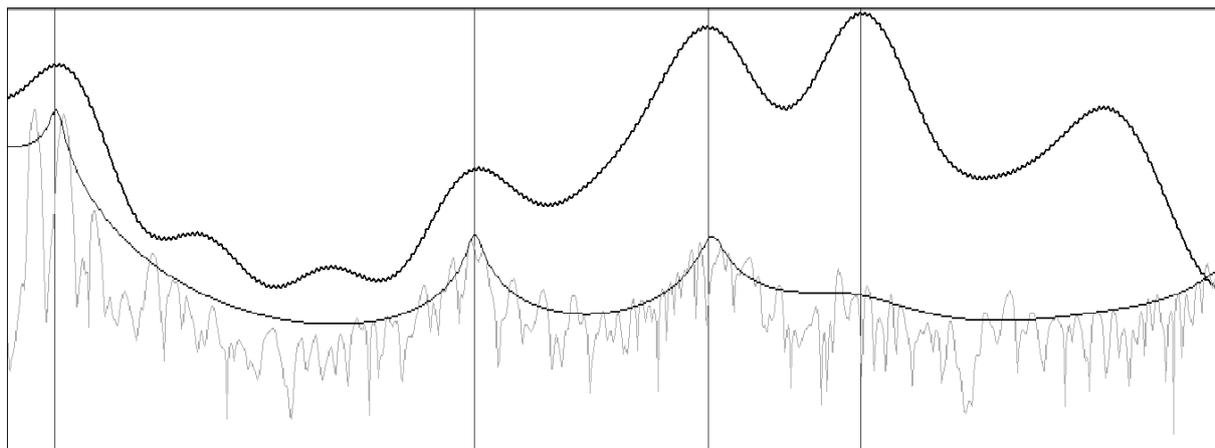


Fig. 2. Une grille de référence comme représentation simplifiée et pertinente du spectre

L'avantage de la grille par rapport au spectre est que l'évaluation visuelle est plus immédiate, car la grille élimine le facteur de l'intensité sonore globale, celle-ci n'étant pas pertinente quand on examine un son isolé ; en revanche, négliger l'intensité relative des pics correspondant aux formants vocaliques peut parfois (rarement) induire une évaluation erronée.

4. URL des principaux logiciels

Real-time Spectrogram, Web. 14/05/2013. <<http://www.phon.ucl.ac.uk/resource/sfs/rtgram/>>

Real-time Spectrum, Web. 14/05/2013. <<http://www.phon.ucl.ac.uk/resource/sfs/rtspect/>>

SARP – SA Rapid Plot, Web. 14/05/2013. <<http://web.uni-plovdiv.bg/rousni/sarp/>>

Speech Analyzer, Web. 14/05/2013. <<http://www-01.sil.org/computing/sa/index.htm>>

TREFL, Web. 14/05/2013. <<http://web.uni-plovdiv.bg/rousni/>>

RÉFÉRENCES BIBLIOGRAPHIQUES

Ditcheva et al. 2009: Ditcheva, M., Dommergues, J.-Y., Ryst, E. *Modélisation et acquisition des voyelles à l'aide du logiciel SARP*. Plovdiv University “Paissii Hilendarski” – Bulgaria, Scientific Works, Vol. 47, Book 1 Part B, 2009, 258–273.

Olsson 2008: Olsson, J. *Forensic Linguistics*. New York: Continuum International Publishing Group, 2008.