TELECOM
PARIS
Département SIGNAL

Mémoire de Soutenance de Thèse

présenté pour obtenir le titre de
Docteur de l'Ecole Nationale
Supérieure des Télécommunications.

Spécialité : Signal et Image

Soumya BOUABANA

Modélisation des Mouvements Articulatoires de la Langue par la Méthode de la LPC Multi-impulsionnelle.

A SOUTENIR EN 06/11 1995 DEVANT LE JURY COMPOSÉ DE

Jean-Pierre	TUBACH	Président
Christophe	d'ALESSANDRO	Rapporteurs
Paul	JOSPA	
Bjorn	LINDBLOM	Examinateurs
Shinji	MAEDA	
Alain	MARCHAL	

Ecole Nationale Supérieure des Télécommunications
Enseignement Supérieur des Télécommunications

à Toussaint...

Remerciements

J'exprime ma profonde gratitude à monsieur S. Maeda qui, en m'accueillant dans le laboratoire du groupe parole du département Signal, m'a donné l'occasion de conduire des travaux particulièrement enrichssants. Qu'il trouve ici l'expression de ma haute considération pour ses qualités de scientifique et surtout pour ses critiques émises avec une grande clairvoyance.

Je remercie très respectueusement monsieur le Professeur J. P. Tubach d'avoir accepté de présider ce jury.

Je tiens à remercier vivement messieurs C. d'Alessandro et P. Jospa de m'avoir fait l'honneur d'être les rapporteurs de cette thèse, et monsieur A. Marchal d'avoir accepté de participer à mon jury.

Je remercie très chaleureusement monsieur B. Lindblom pour l'intérêt qu'il a porté à ce travail et d'avoir accepté de participer au comité du jury.

Que soient ici remerciés tous ceux qui, en m'accordant leur amitié, leur symtpathie ou leurs conseils, ont permis que ce travail soit mené à bien, en particulier messieurs R. Carré et G. Blanchet. Qu'ils trouvent ici le témoignage de mon estime pour leur sympathie à mon égard.

Je voudrais également réserver la dernière phrase à mes parents. *Merci.*

RÉSUMÉ

La description physique d'un son, qu'elle soit acoustique en terme de spectrogramme ou articulatoire dans une séquence cinéradiographique, dépend de plusieurs facteurs : le contexte phonétique, la vitesse d'élocution, l'accent syllabique .... Ce sont autant de facteurs qui affectent le patron des sons.

L'objectif principal de cette thèse est de tenter d'éclairer certains aspects du contrôle moteur lors de la production de la parole en utilisant des techniques de traitement du signal. Plus précisément, le problème que nous nous sommes posé est le suivant : disposant d'enregistrements de films radiologiques de la coupe sagittale de la langue lors de la production de la parole, est-il possible grâce à l'analyse des variations trame par trame des séquences cinéradiographiques de détecter des régularités qui existent dans la parole et de les caractériser ? À priori le problème est double. En premier lieu, il faut modéliser les variations temporelles ou "mouvements" de la langue. Ensuite, il faut analyser les paramètres de contrôle du modèle, en l'occurrence le signal d'excitation du filtre, en vue de comprendre l'organisation spatio-temporelle des commandes des différentes composantes articulatoires de la langue.

La première partie est consacrée aux mécanismes de contrôle des mouvements musculaires. L'étude comparative des mécanismes les plus répandus chez l'homme, nous a conduit à retenir le modèle de type source-filtre pour l'analyse des mouvements des différents paramètres articulatoires. Deux filtres invariants placés en cascade constituent le modèle de chaque paramètre articulatoire. Le premier est un filtre intégrateur. Il est associé au système mécanique de la chaîne de production de la parole. Le deuxième est d'ordre 2. Il correspond au système physiologique. La source, quant à elle, est représentée par une séquence d'impulsions. L'énergie de l'erreur entre le mouvement synthétisé et celui de référence est inférieure à 30% lorsque le nombre d'impulsions est proche du nombre de syllabes dans la phrase. Ce résultat s'explique par le fait que la durée effective de la réponse impulsionnelle du modèle est comparable à celle de la durée moyenne de la syllabe française, $180ms$.

La deuxième partie s'ouvre sur la détermination du nombre minimum d'impulsions de commande $\mathbf{m}^*$ nécessaire pour la synthèse des patrons des fréquences de formants des sons produits avec une précision raisonnable au sens perceptif. Ce nombre est inférieur ou égal au nombre de syllabes dans la phrase. Nos résultats suggèrent que l'organisation spatio-temporelle de la commande des articulateurs de la langue entraîne la production d'un mouvement élémentaire dont la durée est comparable à celle de la syllabe française. De plus, le nombre effectif des impulsions d'un articulateur semble être étroitement lié à celui des traits phonétiques qui lui sont inhérents.

Abstract

The temporal variation of the tongue profiles, derived from X-ray film data, are described in terms of those of four articulatory parameters. A filter model of movement for each tongue parameter is developed in which the impulse response is described by a damped, linear second-order system. The filter describing the system are assumed to be time invariant, and the excitation model is represented by a train of pulses. This model, excited by a limited but large number of pulses, produced a faithful synthesized movement.

Our objective is not the faithful reproduction of the movement. Rather, our interest is to find out something about the nature of inter-articulator coordination during speech by analyzing the organization of pulses. Because of this objective, we determine the minimum number of pulses necessary to describe movements using the difference limen of formant frequencies as an acoustic criterion.

The results show that the number of pulses necessary for each movement to synthesized formants frequencies of vowels is always less or equal to the number of syllables in the sentence. This is explained by the fact that the effective duration of the filter's impulse response is comparable to that of the average syllable duration in the 10 sentences, $180ms$. It means that articulatory controls involve a unit of production whose length is about the size of syllable. Moreover, it suggests that the number of pulses depends on the phonetics features in the utterance, of which their realization is related to particular articulatory parameters.

Table des matières

Liste des figures

Liste des tableaux

Introduction

La description physique d'un son, qu'elle soit acoustique en terme de spectrogramme ou articulatoire dans une séquence cinéradiographique, dépend de plusieurs facteurs : le contexte phonétique, la vitesse d'élocution, l'accent syllabique .... Ce sont autant de facteurs qui affectent le patron des sons. Aussi, des variabilités inter-locuteurs mais également intra-locuteur, sont observées pour un phonème donné.

Au cours des dernières décennies, des efforts considérables ont été développés pour étudier ces variabilités. Cependant, il n'existe toujours pas de descriptions suffisantes des informations acoustiques ou articulatoires pour spécifier le son dans des contextes variés. De ce problème découle une question fondamentale sur la définition de l'entité phonétique qui entraîne une structure de contrôle invariante dans le temps.

L'objectif principal de cette thèse est de tenter d'éclairer certains aspects du contrôle moteur lors de la production de la parole en utilisant des techniques de traitement du signal. Plus précisément, le problème que nous nous sommes posé est le suivant : disposant d'enregistrements de films radiologiques de la coupe sagittale de la langue lors de la production de la parole, est-il possible grâce à l'analyse des variations trame par trame des séquences cinéradiographiques de détecter des régularités qui existent dans la parole et de les caractériser ? À priori le problème est double. En premier lieu, il faut modéliser les variations temporelles ou “mouvements” de la langue. Ensuite, il faut analyser les paramètres de contrôle du modèle, en l'occurrence le signal d'excitation du filtre, en vue de comprendre l'organisation spatio-temporelle des commandes des différentes composantes articulatoires de la langue.

Enregistrement et analyse des données

Dans notre travail, nous avons utilisé le modèle articulatoire de Maeda pour décrire l'évolution temporelle du profil de la langue. Celui ci est décrit par quatre paramètres articulatoires ; un paramètre extrinsèque, la position de la mâchoire **jw** et trois paramètres intrinsèques, la position de la langue **tp**, la forme de la langue **ts** et la position de l'apex **tt**.

L'analyse des mouvements de ces paramètres dans le cas des séquences $CVCV$ a montré que la variabilité observée dans les positions d'un articulateur donné est linéairement coordonnée à celle d'un autre articulateur. Notre analyse a mis en évidence la prépondérance de l'activité de certains articulateurs pour la production d'un phonème donné.

Modélisation des mouvements articulatoires de la langue

Les mouvements des articulateurs de la langue qui correspondent à des phrases prononcées par deux locutrices sont modélisés par des modèles source-filtre. La source est spécifiée par un train d'impulsions générées de manière à minimiser l'énergie de l'erreur entre le mouvement observé et celui synthétisé. Le filtre est conçu en se basant sur les connaissances et les résultats des travaux menés dans le domaine des mécanismes de contrôle moteur et de la structure musculaire. Notre modèle est composé de deux filtres qui décrivent les caractéristiques des systèmes physiologique et mécanique de la chaîne de communication chez l'homme. Notre analyse a montré que le système mécanique est du premier ordre, donc le rôle de la masse est négligé. Quant au système physiologique, il est du second ordre. Il permet de générer la force d'excitation. Ces filtres sont choisis auto-régressifs et invariants dans le temps. Leur contrôle s'effectue par les impulsions de commande.

Détermination de l'excitation

La technique utilisée pour déterminer la séquence d'excitation est celle proposée par Atal et Remde (1982) [?] pour le codage de la parole. Elle fournit une solution suboptimale mais efficace. Les amplitudes et positions des impulsions sont déterminées de manière itérative. Les différentes séquences d'impulsions montrent des différences inter-articulateurs mais également inter-locuteurs. Toutefois, elles montrent certaines régularités par rapport aux caractéristiques acoustiques des unités phonétiques.

Analyse acoustique

Les résultats de la modélisation source-filtre ont ensuite été analysés dans le domaine acoustique. À partir des mouvements synthétisés nous avons calculé les fréquences des quatre premiers formants sur les segments vocaliques. Nous avons montré que les articulateurs de la langue ne sont pas tous contrôlés de la même manière. Selon les caractéristiques acoustiques de la voyelle, l'activation d'un articulateur peut présenter une importance primaire ou secondaire.

Organisation du document

Le premier et deuxième chapitres sont consacrés à la présentation et au traitement des données radiofilms. Nous donnons un aperçu des différentes techniques utilisées en traitement de la parole pour la modélisation du conduit vocal. Nous décrivons en détail le modèle utilisé, celui de Maeda.

Le troisième chapitre présente une analyse exhaustive des mouvements observés des articulateurs de langue.

Le quatrième chapitre est quant à lui consacré aux mécanismes de contrôle des mouvements musculaires. Notre premier souci étant de trouver un modèle de contrôle, il nous semble naturel de prospecter d'abord dans cette direction. L'étude comparative des mécanismes les plus répandus, nous a conduit à retenir le modèle de type source-filtre pour notre analyse. Une ébauche sur les différents critères de choix des paramètres de contrôle du modèle parachèvera ce chapitre.

L'analyse des mouvements des composantes articulatoires de la langue par le modèle LPC multi-impulsionnel est abordée au cinquième chapitre. Le problème de l'invariance des modèles des articulateurs est discuté. Des interprétations physiques (biomécanique, phonétique) des résultats obtenus sont suggérées.

Le chapitre six tente d'illustrer l'organisation des impulsions de commande en fonction des unités phonétiques telles que la syllabe et phonème. Les résultats obtenus dans le cas des deux locutrices sont comparés.

Enfin, le dernier chapitre s'ouvre sur la détermination du nombre d'impulsions de commande minimum m^* nécessaire pour la synthèse des patrons des fréquences de formants avec une précision raisonnable au sens perceptif. Dans ce chapitre, nous examinons d'abord l'effet de la modélisation de chaque composante articulatoire sur les corrélats acoustiques. Ensuite, nous montrons la prépondérance de certaines composantes dans la production des traits phonétiques. Enfin, nous utilisons les seuils de perception des voyelles établis par Flanagan pour déterminer les séquences d'impulsions S^*. Des exemples sont présentés et commentés.

Chapitre 1

Représentation et mesure du conduit vocal

1.1 Pourquoi des données articulatoires ?

La parole est un moyen de communication qui est basé sur la perception et la compréhension par un auditeur du signal sonore produit par un locuteur. Ce signal doit donc être adapté à la perception et contenir des traits invariants par rapport aux caractéristiques phonétiques malgré les aspects variables des gestes articulatoires pour la même information segmentale. Toutefois, une telle prévision de l'invariance des patrons auditifs pour les phonèmes est largement discutable. L'extraction de la partie la plus stable d'une voyelle d'un signal acoustique donne un résultat perceptif inattendu. Souvent, cette déviation ne peut être complètement expliquée par les effets de coarticulation des phonèmes adjacents. Or, le signal acoustique défini dans un espace physique monodimensionnel, n'est autre que le résultat du travail multidimensionnel des articulateurs du conduit vocal. Les observations articulatoires constituent un apport considérable d'informations qui ne peuvent être facilement dérivés des signaux acoustiques.

Aussi, il y a eu un intérêt croissant à savoir si les caractéristiques invariantes correspondant aux phonèmes sont représentées à des niveaux du processus de production plus élevés que le niveau acoustique. Les deux niveaux qui ont fait l'objet du plus grand nombre de travaux sont la forme du conduit vocal et les mouvements articulatoires d'une part, et les patrons des contractions musculaires d'autre part. Les données observées à ces deux niveaux sont obtenues respectivement par des mesures radiographiques et électromyographiques (*cf. sections* **??**, **??**) . Cette dernière technique consiste à observer directement l'activité musculaire. Quant à la première, elle permet de suivre les variations temporelles de la coupe sagittale du conduit vocal durant la parole. En partant des données articulatoires, les chercheurs espèrent arriver à une meilleure compréhension de ce qu'est la parole.

1.2 Techniques d'enregistrement des mouvements du conduit vocal

Le développement des algorithmes de traitement automatique de la parole nécessite de plus en plus de données susceptibles de décrire les signaux de la parole. Afin d'optimiser l'exploitation de ces algorithmes, des efforts intenses ont été fournis pour mettre en oeuvre des techniques d'enregistrement des signaux de parole aussi bien au niveau acoustique qu'articulatoire. Il est à noter que les données articulatoires sont plus difficiles à obtenir que les données acoustiques, en particulier les données sur le contour des organes qui ne sont pas dirtectement accessibles. Dans la section suivante, nous décrivons brièvement les techniques de mesure des mouvements articulatoires du conduit vocal les plus utilisées.

1.2.1 Enregistrements ultrasons

Les mesures ultrasons fournissent une relative bonne qualité des observations bidimensionnelle des organes inaccessibles de parole, tels que la langue et le larynx, sans risque de radiation dans le corps du locuteur. Elles ont été recommandées pour détecter aussi bien les contractions musculaires que la forme de la surface de la langue (Sonies et *al* 1981 [?]). L'inconvénient principal de cette technique est lié aux effets de la charge mécanique à l'extérieur de la peau. Les signaux ultrasons sont facilement réfléchis entre les organes solides ou liquides et l'air.

1.2.2 Enregistrements radiofilms

1.2.2.1 Faisceau de rayons X

Les rayons X ont été pendant longtemps la seule source d'information pour l'étude des mouvements de la langue. Le faisceau des rayons X émerge uniformément d'un générateur dans un large angle solide. Le problème de cette méthode est d'exposer le locuteur à une dose de radiation importante pendant une longue durée. Aussi même les mouvements correspondant aux voyelles stationnaires sont difficiles à enregistrer. En raison de l'effet dangereux du rayonnement la quantité de donnée des mouvements enregistrés est réduite. Ceci représente un véritable obstacle pour la validation des résultats d'analyse.

1.2.2.2 Micro-faisceau de rayons X

Cette méthode est différente de la précédente dans la mesure où le faisceau des rayons X est extrêmement fin. Un calculateur contrôle d'une manière adaptative la direction du rayon X sur la position cible du capteur collée sur la surface des organes articulatoires, par exemple, la langue. La position du capteur est identifiée et calculée en temps réel. L'analyse trame par trame des données utilisées dans la méthode traditionnelle est limitée à celle de la position du capteur. Pendant l'acquisition des données, l'émission automatique des rayons X n'est effectuée que lorsque l'identification de la position du capteur est nécessaire. La dose de rayonnement est donc extrêmement faible en comparaison à celle produite par les autres systèmes à rayons X. Ce système a été implémenté par Fujimura, Kiritani et Ishida en 1968 [?].

1.2.3 Enregistrements magnétiques

Cette technique est utilisée pour suivre l'évolution des mouvements articulatoires lors de la production de la parole, au moyen d'une petite bobine de $4 \times 4 \times 2.5mm$ fixée sur l'articulateur (Perkell 1992 [?]). Celle ci joue le rôle de capteur ou récepteur. Deux ou plusieurs émetteurs magnétiques, qui correspondent à de grandes bobines (la fréquence de fonctionnement est égale à $60kHz$) sont placés au-dessus et derrière la tête du locuteur, de telle manière que leurs axes principaux soient parallèles entre eux et perpendiculaires au plan sagittal médian du conduit vocal. Ils déterminent la position des capteurs.

La difficulté majeur de cette technique est de placer les capteurs sur des axes parallèles aux émetteurs de manière à obtenir des mesures précises. Cette contrainte est difficile à réaliser dans la cavité buccale où les déformations de la langue placent les récepteurs sur des axes non parallèles à ceux des émetteurs. Des mesures de haute précision ont été obtenues dans le cas de mouvements correspondant à des articulateurs de structure anatomique rigide tel que la mandibule (Nguyen et Marchal 1992 [?]).

1.2.4 Enregistrements électromyographiques

Les mesures de l'activité musculaire représentent la meilleure technique pour l'observation directe de l'évolution physiologique du conduit vocal lors de la production de la parole. L'utilisation d'électrodes prédomine dans les études électromyographiques (EMG). L'EMG est considéré comme le moyen le plus puissant pour établir les principes de contrôle de parole *via* des mesures directes. Aussi, de nombreux travaux ont été effectués, notamment, par l'équipe “physiologie de la parole” aux laboratoires de Haskins pour étudier le contrôle articulatoire (Bell-Berti et Harris 1982 [?], Baer et *al.* 1988 [?]).

L'intérêt de cette méthode réside dans la nature des signaux enregistrés qui représente un niveau plus élevé dans le processus de production. Les signaux EMG sont, en effet, en relation directe avec les forces musculaires. Toutefois, l'interprétation de ces signaux, qui représentent les différentes contributions des multiples fibres du muscle, est difficile pour des analyses quantitatives.

Notre travail est s'effectué sur des données radiofilms obtenues par l'émission d'un faisceau de rayons X selon le système d'enregistrement présenté ci-dessous.

1.3 Données

1.3.1 Réalisation technique

D'après Bothorel et *al.* (1986) [?] de l'Institut de Phonétique de Strasbourg (IPS), le matériel radiologique utilisé se compose d'un générateur de rayons X, d'un amplificateur de brillance, d'un bloc optique, d'une caméra avec film photosensible (caméra 1) et d'un moniteur vidéo (*fig.*??). L'émission des rayons X est centrée sur les organes articulatoires du conduit vocal. Les rayons X ne sont pas émis d'une façon continue mais par impulsions synchrones avec l'ouverture du diaphragme de la caméra. La tête du sujet, maintenue fixe

Figure 1.1 *Salle de radiocinématographie vue de dessus, d'après Bothorel et al* 1986 [7].

grâce à un crâniostat, est placée à $50cm$ du générateur. Un amplificateur de brillance capte les variations du champ de rayons X dont l'intensité est modifiée selon la région organique traversée et transforme ces variations de champs en variations lumineuses proportionnelles. Celles-ci vont impressionner le film photosensible de la caméra 1. Afin de rendre plus visible le contour des lèvres sur les clichés radiologiques, un filtre de plomb composé de plusieurs plaques a été placé entre le sujet et le générateur de rayons. Les labiofilms sont réalisés à l'aide des caméras 2 et 3 qui filment respectivement les lèvres de face et à 45 degrés. Les trois caméras sont synchronisées électroniquement à la vitesse de 50 images par seconde.

Ces données articulatoires sont accompagnées de documents acoustiques qui sont des enregistrements sur magnétophone et les spectrogrammes correspondants.

1.3.2 Corpus

Le corpus correspond à des articulations de courtes phrases françaises. Celles-ci sont constituées de quatre à six syllabes dont la dernière syllabe est du type consonne-voyelle (CV). On dispose au total de dix phrases prononcées par deux locutrices de langue maternelle française, âgées de 23 à 26 ans :

Sujet 1. *P.B.*, née en Allemagne (Fribourg-en-Brisgau), a séjourné en différents points de la métropole et outre-mer.

Sujet 2. *D.F.*, née dans le Haut-Rhin (Mulhouse) a tendance à surarticuler par rapport à PB.

Nous disposons, en outre, d'un ensemble de croquis de logatomes, séquences $CVCV$, où les consonnes (C) sont de type occlusives $/p, d, g/$ et les voyelles (V) sont les trois voyelles extrêmes $/a, i, u/$. La première consonne correspond à la bilabiale /p/ qui n'implique pas les articulateurs de la mâchoire et de la langue. Ces séquences sont prononcées dans la phrase "Répéter...trois fois" par une troisième locutrice française.

Sujet 3. *M.I*, locutrice de langue maternelle française.

L'examen des croquis des logatomes présentés en dynamique, devrait aider à comprendre plus aisément l'organisation des stratégies de la commande articulatoire. Les phrases, contenant plus d'éléments phonétiques et linguistiques, sont plus difficiles à analyser. En revanche, elles devraient contenir plus d'informations concernant la production de la parole naturelle. La liste des phrases et des logatomes est présentée dans l'annexe A.

Annexe A : Corpus

Liste des 10 phrases françaises.

Phrases		*Transcription phonétique*
01	*Ma chemise est roussie*	/m a ʃ m i z ε r u s i/
02	*Voilà des bougies*	/v w a l a d ε b u ʒ i /
03	*Donne un petit coup*	/d ɔ n œ̃ p t i k u/
08	*Une réponse ambiguë*	/y n r e p õ s ã m b i g y/
09	*Louis pense à ça*	/l w i p ã s a s ɑ/
15	*Mets tes beaux habits*	/m ε t e b o z a b i/
17	*Une pâte à choux*	/y n ə p a t a ʃ u/
18	*Prête-lui seize écus*	/p r ε t l w i s e z e k /
24	*Chevalier du gué*	/ʃ ə v a l j e d y g e/
28	*Il fume son tabac*	/i l f y m s ɔ̃ t ɑ b ɑ/

Liste des logatomes

Logatomes		*Transcription phonétique*
01	*pada*	/p a d a/
02	*padou*	/p a d u/
03	*pagou*	/p a g u/
04	*pida*	/p i d a/
05	*pigou*	/p i g u/
06	*pouga*	/p u g a/
07	*pougou*	/p u g u/
08	*poudi*	/p u d i/
09	*piga*	/p i g a/
10	*pouda*	/p u d a/
11	*padi*	/p a d i/
12	*pidou*	/p i d u/
13	*pigi*	/p i g i/
14	*paga*	/p a g a/
15	*pougi*	/p u g i/
16	*pipi*	/p i p i/
17	*pidi*	/p i d i/
18	*pagi*	/p a g i/
19	*pinpin*	/p œ̃ p œ̃/
20	*toutou*	/t u t u/
21	*tintin*	/t œ̃ t œ̃/
22	*peup*	/p œp/
23	*keke*	/k ε k ε/

Chapitre 2

Paramètres articulatoires du conduit vocal

2.1 Système de production

Pour générer un message vocal, le locuteur procède à la production d'un ensemble de commandes qui, du système nerveux central jusqu'aux muscles via le système nerveux périphérique, vont permettre de piloter les évolutions de la forme du conduit vocal. Au niveau de la réalisation articulatoire, de nombreux acteurs sont mis en jeu. Ils sont regroupés le plus souvent selon leur localisation en deçà ou au delà du larynx en deux sous-systèmes, *sub* et *supraglottique* (*fig.* **??**, d'après Flanagan 1972 [**?**]).

2.1.1 Système subglottique

Il englobe les *poumons* et la *trachée*. Cet ensemble, appelé soufflerie, se comporte comme un générateur de débit d'air qui alimente le *larynx*.

2.1.2 Larynx

Le *larynx* contribue à la génération de la source vocale. Alimenté en pression relativement constante par le système subglottique, il génère, au cours de la phonation et par vibration des cordes vocales, l'onde qui véhiculera le signal sonore. Le réglage de la tension et de l'écart entre les cordes vocales permet la diminution ou l'augmentation de la fréquence et/ou de l'amplitude de la source.

Figure 2.1 *Conduit vocal, d'après Flanagan* 1972 [8].

2.1.3 Système supraglottique

Il se compose de deux parties, la *partie orale* et la *partie nasale*. La première comprend le *pharynx* et un ensemble d'articulateurs : la *langue*, la *mandibule* et les *lèvres* qui participent à la mise en forme de l'onde sonore issue de la glotte. La partie la plus mobile de l'ensemble est la *langue* qui ne comporte pas moins de 17 muscles différents ancrés sur les os ou les organes voisins ; elle peut modeler à volonté la forme du conduit oral. Les *lèvres*, composées également de plusieurs muscles, peuvent se déformer dans de larges proportions de manière continue. Enfin, la *mandibule* qui, contrairement au maxillaire fixe, dispose d'une grande mobilité dans les trois plans vertical, horizontal et transversal.

La *partie nasale* se compose de deux cavités fixes : les deux *fosses nasales*. Le degré du couplage nasal dans la phonation dépend de la position du voile du palais. Celui-ci constitue le prolongement mobile de la voûte du palais. Complètement abaissé, il aiguille la totalité de l'onde acoustique vers les fosses nasales. Relevé, il les met entièrement hors circuit et le flux d'air est dirigé vers les lèvres. Des positions intermédiaires sont possibles. Les données anatomiques du conduit nasal sont rares. Après quelques travaux préliminaires (House et Stevens 1956 [?]), ce n'est que très récemment que les problèmes théoriques liés à cette cavité et leur simulation ont été réactivés (Maeda 1982 [?], Lonchamp 1988 [?], Dang et Honda 1994 [?]).

2.2 Description des articulateurs du conduit vocal

2.2.1 La langue

Pièce maîtresse de la cavité buccale, la langue est une structure non rigide, très déformable, dotée d'une grande vitesse de déplacement, en particulier son apex. Elle est difficilement observable par les techniques de mesure.

La méthode classique de description de la position de la langue était d'en spécifier le point le plus haut. On a pensé que la reproduction approximative de la forme de la langue, à partir des deux paramètres représentant ce plus haut point sur l'axe horizontal et l'axe vertical, était suffisamment pertinente(Stevens et House 1955 [?], Lindblom et Sundberg 1971 [?]). Mais ce type de représentation n'a pas apporté d'informations suffisantes sur son fonctionnement. D'autres techniques d'analyse ont alors été utilisées pour permettre une reproduction plus précise de la forme de la langue, ainsi que nous le verrons dans la section ??.

2.2.2 Les lèvres

Extrémité visible et donc la plus accessible du conduit vocal, les lèvres ont été photographiées, cinématographiées et plus récemment filmées en vidéo (Abry et Boë 1986 [?], Fukuda et Hiki 1982 [?]). La section du tube labial est caractérisée par une impédance acoustique propre. Cette impédance acoustique est proportionnelle au rapport de la longueur l à la section A du tube labial, lorsque le rayonnement aux lèvres est négligé. La relation A/l, qui est une information articulatoire, a été largement utilisée pour spécifier des phénomènes acoustiques. Nous citons, en particulier, les travaux de Stevens et House (1955) [?] sur la description quantitative de l'articulation dans la production des voyelles.

2.2.3 La mandibule

Les mouvements complexes de cet organe (rotations et translations) ont longtemps été écartés des mesures et de la modélisation. En effet, les premiers modèles articulatoires se caractérisent par l'absence de paramètres mandibulaires (Öhman 1967 [?], Harshman et *al.* [?]). L'abaissement-rotation mandibulaire provoquant une augmentation de l'aperture buccale, dans la plupart des cas reliée aux mouvements de la langue, c'est sur cette dernière qu'ont été portés tous les efforts de description.

Toutefois, cette attitude a changé au début des années 70. Les nouveaux modèles articulatoires ont intégré la mâchoire comme paramètre à part entière (Lindblom et Sundberg 1971 [?], Mermelstein 1973 [?], Maeda 1992 [?]). La mâchoire est alors définie par la distance inter-incisive et un centre de rotation. Les modèles proposés ont permis d'étudier l'influence spécifique de cet articulateur dans la parole, tel que son rôle déterminant dans la durée des syllabes (Lindblom 1983 [?]).

2.3 Modèles de production

Dans des systèmes complexes comme celui de la production de la parole, il est difficile d'étudier le comportement des multiples degrés de liberté existant dans le système articulatoire à l'échelle microscopique. Néanmoins, ce système est organisé pour des actions correspondant à un petit nombre d'articulateurs fonctionnels pour la production de la parole. Le problème fondamental est donc d'identifier les articulateurs qui décrivent au mieux les fonctions accomplies par le système de production pour structurer les sons de la parole.

De nombreux modèles de production tentent de décrire et de modéliser la forme générale de la fonction d'aire ou de la coupe sagittale au moyen d'un nombre réduit de paramètres indirectement liés aux articulateurs réels. Parmi ces modèles, nous distinguons deux groupes selon les considérations à la base de leur définition : *les modèles* visant à décrire directement la *fonction d'aire* et *les modèles articulatoires* décrivant la coupe sagittale du conduit vocal.

2.3.1 Modélisation de la fonction d'aire

2.3.1.1 Modèles paramétriques

Ces modèles ont une caractéristique commune : ils définissent la position X_c et la section minimale A_c de la constriction principale dans la région du corps de la langue (Stevens et House 1955 [?], Fant 1980 [?]). L'évolution de l'aire à partir de X_c (à l'exception des sections correspondant aux lèvres et au larynx) est généralement représentée par une fonction continue (hyperbole, parabole, sinusoïde). Ces modèles de conduit vocal ont été utilisés avec profit dans des études sur la relation entre la forme du conduit vocal et les propriétés acoustiques pour des voyelles.

2.3.1.2 Modèles basés sur les caractéristiques acoustiques du tube

Ce groupe de modèles est basé sur l'étude des propriétés des tubes acoustiques. Les chercheurs tels que Heinz (1967) [?], Mermelstein (1967) [?] et Schroeder (1967) [?] sont à l'origine de cette démarche. Schroeder s'est intéressé à l'étude des fonctions de perturbation de l'ensemble du tube acoustique de section uniforme qui entraînent la variation d'un seul formant. La même année, Heinz a généralisé le travail de Schroeder au cas de fonctions d'aire quelconques. Quelques années plus tard, Fant et Pauli (1974) [?] ont proposé la fonction de sensibilité. Cette fonction représente la variation des fréquences de formants en fonction du lieu de perturbation de la fonction d'aire. Cette fonction s'est montrée intéressante pour étudier la relation fonction d'aire/paramètres acoustiques.

Le modèle à région distinctives (DRM) a été proposé, en 1988, par Mrayati et *al.* [?]. La définition du modèle est basée sur l'étude des relations qui lient de petites perturbations des fonctions d'aire d'un tube acoustique aux variations des fréquences des formants. À partir de ces relations, des régions sont définies permettant de contrôler transversalement et, plus tard longitudinalement (Chennoukh 1995 [?]), les dimensions du tube et de réaliser des constrictions à des endroits quantifiés le long de celui-ci.

2.3.2 Modélisation articulatoire

Plusieurs modèles basés sur une première approximation de la coupe sagittale du conduit vocal avec un nombre de degrés de liberté très limité ont été développés. Dans cette section nous présentons la description des principaux modèles que nous avons classé en quatre catégories : géométriques, statiques, statistiques et physiologiques.

2.3.2.1 Modèles géométriques

Ils ont tous en commun de simplifier le conduit vocal par une lecture géométrique de la coupe sagittale. Les articulateurs peuvent être représentés par un ensemble d'éléments géométriques simples, tels qu'un cercle (Coker 1966 [?], Mermelstein, 1973 [?] et Flanagan 1970 [?]). Mermelstein (1973) a utilisé ce type de modèle pour définir des règles de synthèse en dynamique pour produire des séquences de type VCV. En 1976, Coker [?] propose une autre version du modèle qu'il a étudié avec Fujimura en 1966, et l'intègre dans un système de synthèse articulatoire.

Toutefois, la simplicité d'approche de ce type de modélisation se heurte vite à des mauvais rendus physiologiques : la rigidité circulaire de la langue ne peut prendre en compte sa capacité d'aplatissement-abaissement ; les mouvements consonantiques de l'apex nécessitent aussi une plus grande complexité.

2.3.2.2 Modèle d'interpolation

D'autres auteurs ont cherché à caractériser les formes de la langue par des valeurs directement mesurées sur les images radiofilms. En 1971, Lindblom et Sundberg [?] ont proposé une procédure de reproduction de la forme de la coupe sagittale à partir de la hauteur et la largeur de l'ouverture des lèvres, le bout et le corps de la langue, et la hauteur du larynx. Ils ont adopté une représentation où ils montrent trois positions principales de la composante maxillaire pour lesquelles sont reproduites les voyelles. Ces position définissent une caractéristique relative au lieu de constriction qui est soit avant, soit central, soit arrière.

Ces modèles présentent certaines limitations dues essentiellement au choix intuitif des paramètres. Ils ne présentent, en effet, aucune garantie mathématique assurant l'optimalité des paramètres choisis.

2.3.2.3 Modèles mathématiques et statistiques

D'autres modèles ont été élaborés avec une approche mathématique basée sur le calcul de paramètres à partir de l'analyse des données observées, notamment radiographiques. Le premier modèle de ce type est celui de Liljencrants (1971) [?]. Celui ci a réussi à décrire le vecteur représentant la forme sagittale de la langue pour dix voyelles suédoises en terme d'une série de Fourier. Il a montré que deux paramètres suffisent à spécifier le vecteur avec une bonne précision. Ces deux paramètres correspondent à l'amplitude et à la phase de la première composante de la série. Le problème de ce type de modèle est de trouver une interprétation en termes articulatoires des composantes extraites.

Quelques années plus tard, ces résultats ont été confirmés par des méthodes d'analyse factorielle multi-variables. Cette technique est basée sur l'extraction des composantes principales de la coupe sagittale repérée sur une grille qui permettent par simple combinaison linéaire de décrire l'ensemble du conduit buccal. Ces modèles sont plus proches de la géométrie effective du conduit vocal. Par exemple, Harshman et ses collègues (1977) [?] ont montré que les formes de la langue peuvent être décrites par deux paramètres. L'un décrit un mouvement d'avancement de la racine de la langue accompagné du mouvement élévation de la partie antérieure du corps de la langue et l'autre paramètre décrit un mouvement d'élévation et de recul de la langue. Maeda (1978) [?] a proposé un modèle analytique de la forme du conduit vocal, basée sur l'étude statistique des mesures radiographiques (*cf.* section **??**).

Avec les mêmes objectifs, Hashimoto (1982) [?] et Sasaki(1982) [?] ont décrit les profils de la langue à l'aide d'une décomposition polynomiale quadratique. Ces auteurs ont proposé des relations polynomiales entre les formants et la forme de la langue correspondante et ils ont dérivé des relations de correspondance entre les espaces articulatoire et acoustique.

2.3.2.4 Modèles physiologiques

Les modèles que nous venons de décrire ne reflètent pas la structure biomécanique des articulateurs du conduit vocal. Et pourtant il est naturel de penser que les caractéristiques physiologiques de ces articulateurs imposent des contraintes à ses déformations pour produire les sons du langage. La simulation biomécanique permet de mettre en évidence ces contraintes et la première tentative a été menée par Perkell (1974) [?]. Son modèle est construit à partir d'éléments musculaires visco-élastiques dont l'organisation reflète la structure musculaire de la langue décrite dans le plan sagittal médian.

Pour intégrer la propriété fondamentale de la conservation du volume de la langue, un modèle 3D, basé sur la méthode des éléments finis, a été proposé par Kiritani et *al.* (1976) [?]. Cette première tentative montre déjà que la contraction du génioglosse postérieur a pour effet de déplacer vers l'avant la masse de la langue tout en élevant sa partie antérieure.

La poursuite de cette recherche de modélisation 3D en éléments finis a conduit à des versions améliorées du modèle 3D de la langue incluant les lèvres et le larynx (Kakita et Fujimura 1977 [?]). Les fonctions d'aire et de transfert ont pu ainsi être calculées. Ces modèles biomécaniques sont des modélisations statiques qui n'intègrent que l'élasticité (Wilhelms-Tricarico, 1995 [?]). Les contraintes dynamiques dues à l'inertie (ou la masse) ne sont pas intégrées dans ces modèles.

Toutes ces méthodes de modélisation ont permis le développement d'un nombre substantiel de travaux qui relèvent de questions qui demeurent essentielles pour la compréhension du processus de production, de la mesure et de la représentation. Dans le paragraphe suivant, nous décrivons un modèle basé sur la méthode de l'analyse factorielle multi-variables, modèle que nous avons utilisé pour étudier la commande spatio-temporelle de la langue.

2.4 Modèle de Maeda

Pour une description de la parole, le conduit vocal peut être considéré comme composé de sous-systèmes mécaniques indépendamment contrôlés. Maeda (1978) [?] a déterminé les composantes linéaires représentant de tels sous-systèmes par analyse factorielle comme décrit ci-dessous.

2.4.1 Mesure de la forme du conduit vocal

La forme de la langue et des parois de la cavité buccale sont mesurées en utilisant un système de coordonnées semi-polaires, comme le montre la figure ??. La position des coordonnées semi-polaires est fixée par rapport au maxillaire supérieur comme référence. Les points d'intersection entre les contours et les grilles des coordonnées sont repérés par une procédure automatique basée sur un algorithme de recherche binaire. A chaque instant, la forme de la langue est décrite par un vecteur de 27 variables, $\mathbf{x}_{tng}$ (Maeda 1992 [?]). Ces mesures ont été effectuées pour l'ensemble des images radiofilms du corpus. Le vecteur mesuré, $\mathbf{x}_{tng}$, est converti en un vecteur centré et réduit, noté $\mathbf{z}_{tng}$, comme suit:

$$\mathbf{z}_{tng} = (\mathbf{x}_{tng} - \bar{\mathbf{x}}_{tng})/\sigma_{tng}, \qquad (2.1)$$

où $\bar{\mathbf{x}}_{tng}$ et σ_{tng} sont respectivement le vecteur moyen et l'écart type de $\mathbf{x}_{tng}$ calculés sur l'ensembles des dix phrases.

Figure 2.2 *Mesure de la forme du conduit vocal en utilisant les coordonnées semi-polaires (les cercles indiquent les points mesurés), d'après Maeda 1992.*

La section labiale est modélisée par un tuyau de section elliptique uniforme. Trois variables décrivent la forme des lèvres :

- la *longueur* du pavillon labial ou protrusion mesurée à partir du profil sagittal ;
- la *hauteur* ;
- la *largeur* frontale de l'aperture des lèvres.

Le vecteur centré et réduit, $\mathbf{z}_{lip}$, est donc composé de trois variables.

La forme du larynx est décrite par deux points à ses extrémités qui définissent sa *position.*

Le mouvement de la mandibule, *ouvert/fermé*, est décrit par *la position de la mâchoire*, notée **jw**. Le paramètre **jw** est défini par la projection de la distance entre les incisives centrales supérieures et inférieures sur une ligne droite. L'angle entre cette ligne droite et la ligne horizontale est déterminé de façon à ce que la proportion de la variance extraite des données de la langue, qui correspond au paramètre **jw**, soit maximale.

Les variations temporelles de ce paramètre affectent celles des articulateurs des trois sections du conduit vocal. C'est pourquoi on a procédé à la soustraction de son effet des mesures des différentes variables définies.

La forme du conduit vocal correspondant à chaque trame est ainsi représentée par un vecteur de variables mesurées. Ne disposant pas des images labiofilms numérisées pour les sujets DF et MI, les variables des lèvres n'ont pas été mesurées pour ces deux locutrices.

2.4.2 Modèle par composantes linéaires

L'analyse factorielle consiste à représenter les formes observées de la langue ($\mathbf{z}_{tng}$) d'une part, et celles des lèvres ($\mathbf{z}_{lip}$) d'autre part (après la soustraction de la composante mâchoire) par la somme pondérée de composantes linéaires comme suit (Maeda 1979 [?]) :

$$\mathbf{z}_j = \mathbf{A}_j \mathbf{y}_j, \tag{2.2}$$

où j est l'indice spécifiant l'articulateur de la langue ou des lèvres, $\mathbf{y}_j$ est le vecteur des composantes (le vecteur des paramètres) et $\mathbf{A}_j$ est une matrice de coefficients de pondération des paramètres de $\mathbf{y}_j$, c'est à dire "factor patterns".

Les résultats de cette analyse appliqués au vecteur $\mathbf{z}_{tng}$, ont conduit à choisir les trois premières composantes principales pour décrire la forme du profil de la langue. Ces composantes sont orthogonales et donc statistiquement indépendantes.

L'analyse des effets de ces trois composantes sur la coupe sagittale a montré une forte correspondance avec les mouvements des principaux paramètres articulatoires intrinsèques de la langue :

- la première composante est associée à *la position du corps de la langue* (**tp**). Elle est décrite par deux états : *antérieur* ou *postérieur* ;
- la deuxième composante, associée à *la forme du corps de la langue* (**ts**), est soit *arquée* soit *plate* ;
- la troisième composante correspond à *la position de l'apex de la langue* (**tt**). Elle est également décrite par deux états : *haut* ou *bas*.

Le tableau **??** montre la variance extraite de chacune de ces composantes intrinsèques ainsi que celle de la composante extrinsèque **jw**. Ces paramètres décrivent plus de 90% de la variance du profil de la langue observée durant la parole pour les trois locutrices. Les

$articulateur$	$var_{jw}(\%)$	$var_{tp}(\%)$	$var_{ts}(\%)$	$var_{tt}(\%)$	$Total(\%)$
PB	15.6	45.8	20.4	9.2	91%
DF	24.0	32.4	23.4	11.0	90%
MI	33.9	31.2	25.0	5.0	95%

Tableau 2.1 Variance extraite des paramètres articulatoires de la langue.

variances extraites des paramètres dépendent plus ou moins de la locutrice. Elles présentent néanmoins certaines régularités. Notons, par exemple, que la somme des variances var_{jw} et var_{tp} est approximativement constante pour les trois locutrices. Ce résultat signifie qu'il existe une coordination entre le mouvement de la position de la langue et celui de la mâchoire. Ainsi, la variance de la mâchoire **jw** est plus faible pour la locutrice PB que DF. En revanche, la variance de la position de la langue **tp** de PB est plus élevée. Quant aux valeurs des variances des paramètres **ts** et **tt**, elles sont pratiquement identiques pour les deux premières locutrices.

Pour la partie labiale et le larynx, trois paramètres articulatoires ont été extraits. Deux paramètres intrinsèques aux lèvres correspondent à la protrusion du pavillon labial **lp** et à la hauteur frontale de l'aperture des lèvres **lh**. Ces deux paramètres plus le paramètre **jw** constituent le vecteur $\mathbf{y}_{lip}$. Le troisième paramètre **lx** décrit le mouvement du larynx.

Grâce à cette analyse, un modèle articulatoire formé de sept composantes a été élaboré. Il permet de reconstituer la forme du conduit vocal avec une précision suffisante pour calculer les caractéristiques acoustiques notamment pour les voyelles.

2.4.3 Patron temporel des paramètres articulatoires du conduit vocal

Les patrons temporels de chacun des paramètres articulatoires $\mathbf{y}_j$ sont calculés à partir des données observées $\mathbf{z}_j$ de la manière suivante :

$$\mathbf{y}_j = \mathbf{A}_j^{-1}\mathbf{z}_j. \tag{2.3}$$

Ils sont considérés comme les mouvements des articulateurs correspondants.

Les figures **??** à **??** illustrent des exemples de l'évolution trame par trame des paramètres articulatoires de la langue et des lèvres. L'intervalle entre deux trames est de $20ms$. Il correspond à une fréquence d'échantillonnage de $50Hz$. Les amplitudes de ces signaux sont des valeurs centrées et réduites. Les frontières qui délimitent les segments des phonèmes sont déterminées à partir de la lecture des spectrogrammes. Nous avons choisi de montrer deux phrases différentes "Une réponse ambiguë" et "Louis pense à ça" prononcées par les deux locutrices PB et DF afin d'observer les variabilités et les régularités inter et intra locuteurs.

Ces figures montrent des mouvements lents oscillants entre des valeurs extrêmes. Globalement, les mêmes articulateurs suivent des trajectoires comparables pour les deux locutrices. Des *régularités* apparaissent, d'une manière évidente, aussi bien au niveau des transitions que dans les segments stationnaires.

Les variations inter-locutrices observées sont liées essentiellement à l'application des facteurs prosodiques telles que la durée et l'accentuation. Un exemple est relevé dans la phrase 08 où le degré d'aperture de la mâchoire correspondant au segment de la voyelle ouverte /ã/ ("aN" dans la figure) est élevé chez la locutrice DF qui a choisi de bien articuler. Tandis que la même portion de la trajectoire obtenue pour la locutrice PB montre un mouvement de fermeture qui serait le résultat d'une faible articulation de la mâchoire masquée par l'effet du contexte consonantique. En contre partie, les mouvements des trois articulateurs de la langue sur le même segment vocalique sont plus distincts pour la locutrice PB que DF. Ce phénomène est appelé *compensation*.

Par ailleurs, des variations contextuelles sont observées chez la même locutrice. Elles reflètent l'influence mutuelle des unités linguistiques successives. Ces variations sont dues principalement à la *coarticulation*. Pour illustrer ces propos prenons le cas de la mâchoire. D'une manière générale, les gestes consonantiques possèdent un plus grand degré de constriction que les gestes vocaliques. Le degré d'aperture de la position de la mâchoire pour la même voyelle varie en fonction du contexte consonantique, ainsi que le montre l'exemple de la phrase 17. Cette phrase contient deux voyelles /a/ dans des contextes différents. La première est précédée par la consonne bilabiale /p/, elle correspond à un degré d'ouverture élevé. La deuxième est située dans un contexte post-alvéolaire/palato-alvéolaire (/t a ʃ/), elle est obtenue avec la position de la mâchoire neutre (valeur nulle). Un autre exemple de cet effet contextuel sur l'ouverture de la voyelle /a/, est montré dans les figures **??** et **??**. Notons la différence entre les degré d'aperture de la même voyelle /a/ dans un contexte post-alvéolaire/bilabial (/t a b/) et en fin de phrase.

Tous ces résultats sont repris d'une manière plus exhaustive dans le chapitre suivant.

Des descriptions qualitatives et quantitatives obtenues à partir de l'analyse des données CVCV prononcées par la locutrice MI sont présentées et commentées.

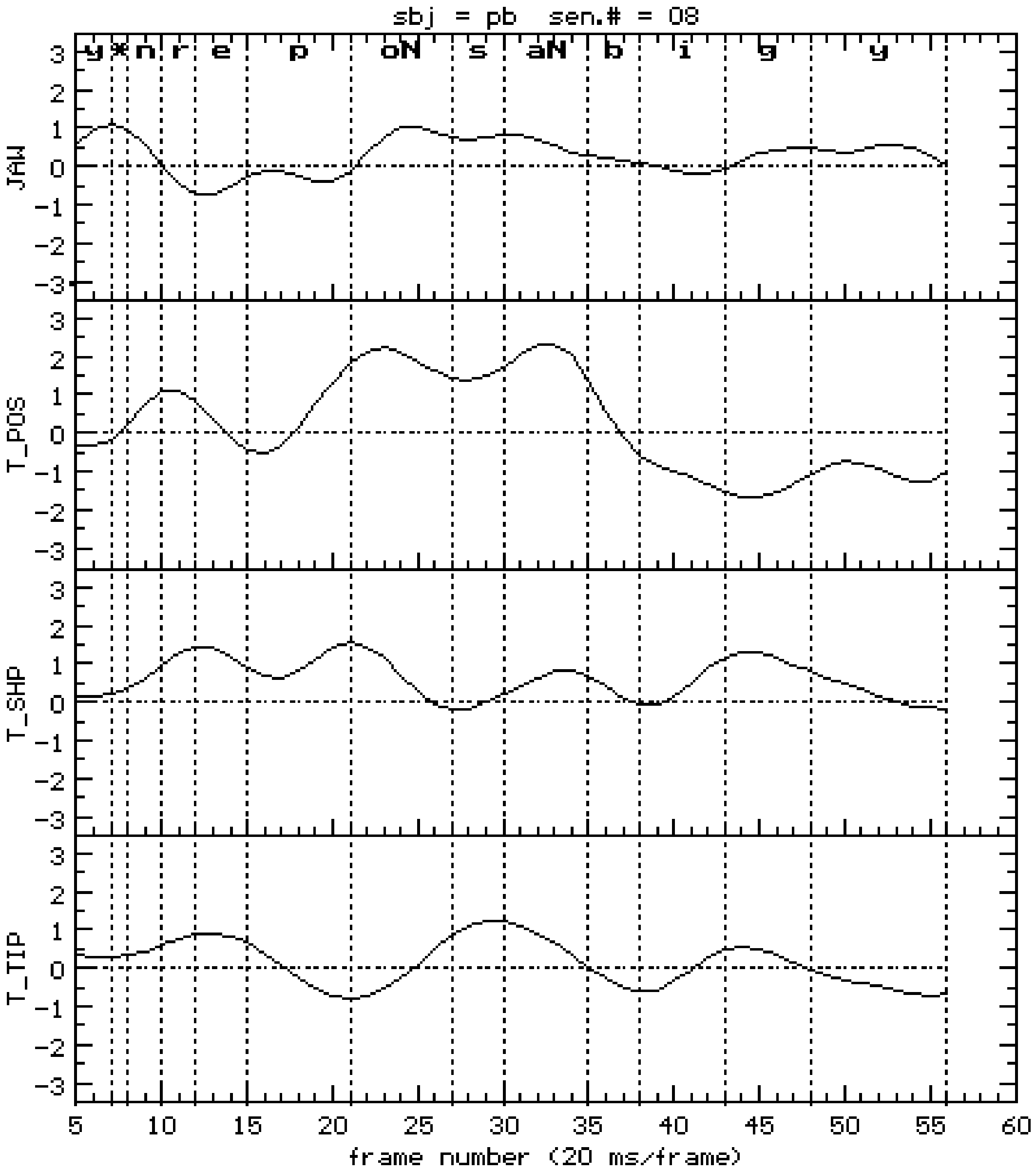


Figure 2.3 *Les patrons temporels des* 4 *paramètres articulatoires de la langue pour la locutrice PB et la phrase* 08*.*

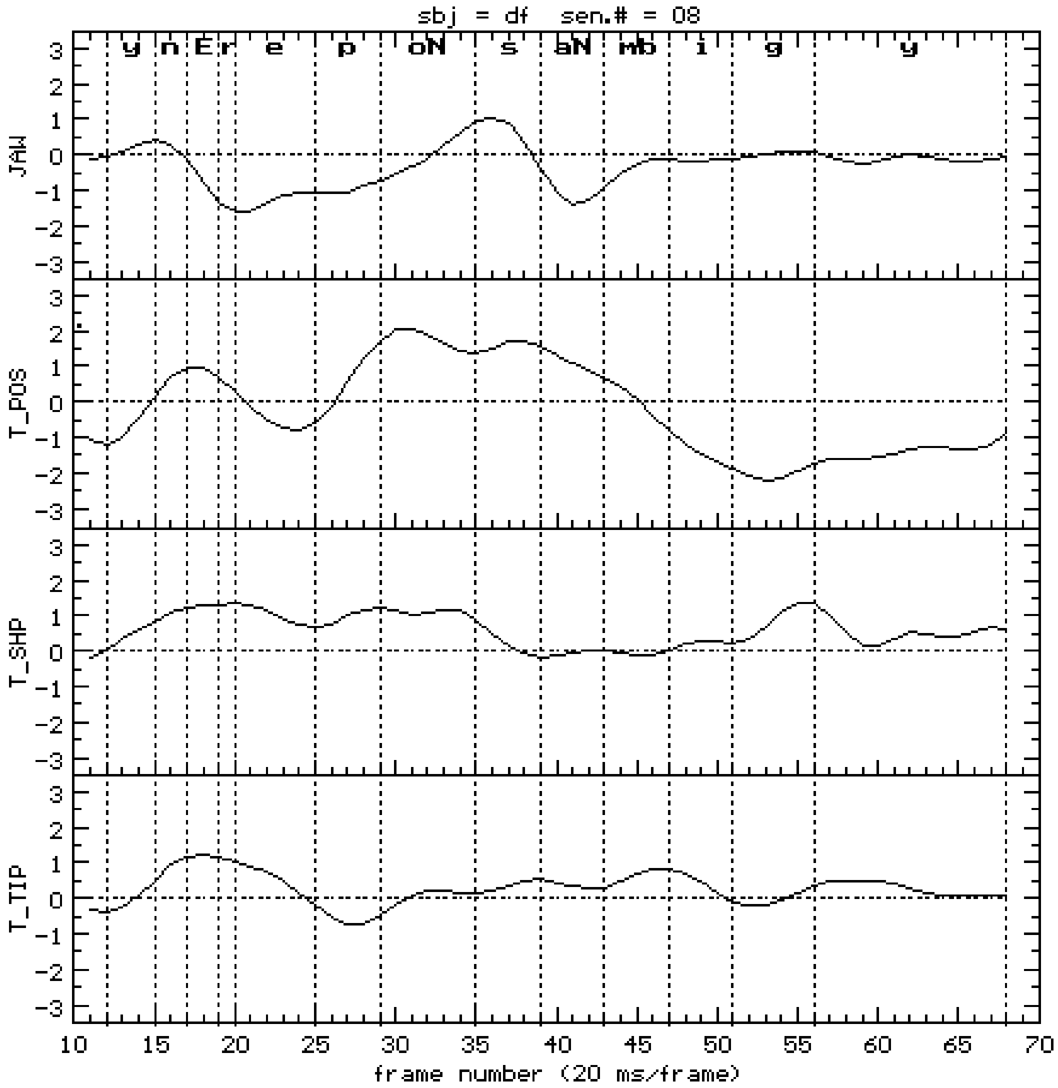


Figure 2.4 *Les patrons temporels des* 4 *paramètres articulatoires de la langue pour la locutrice DF et la phrase* 08.

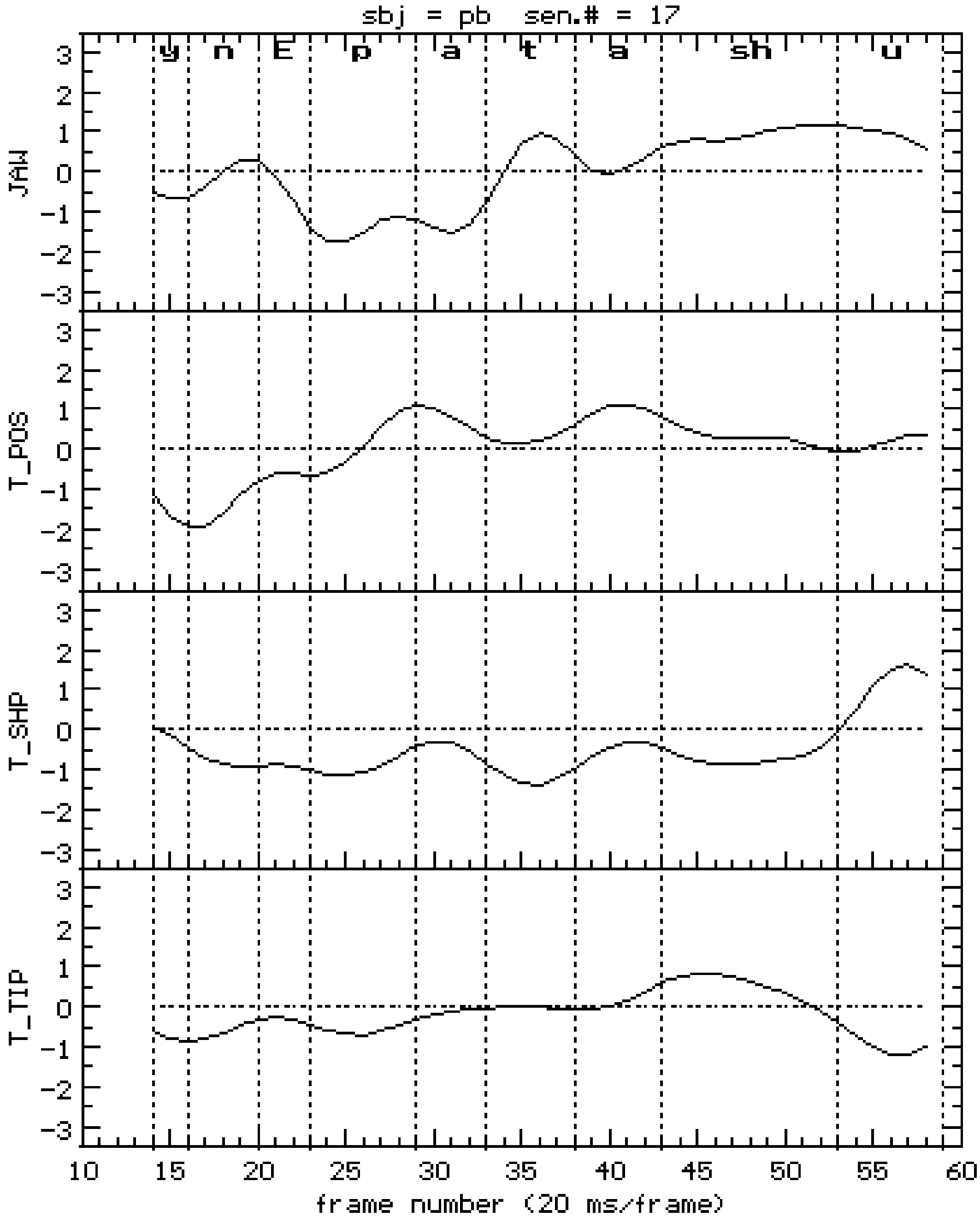


Figure 2.5 *Les patrons temporels des 4 paramètres articulatoires de la langue pour la locutrice PB et la phrase 17.*

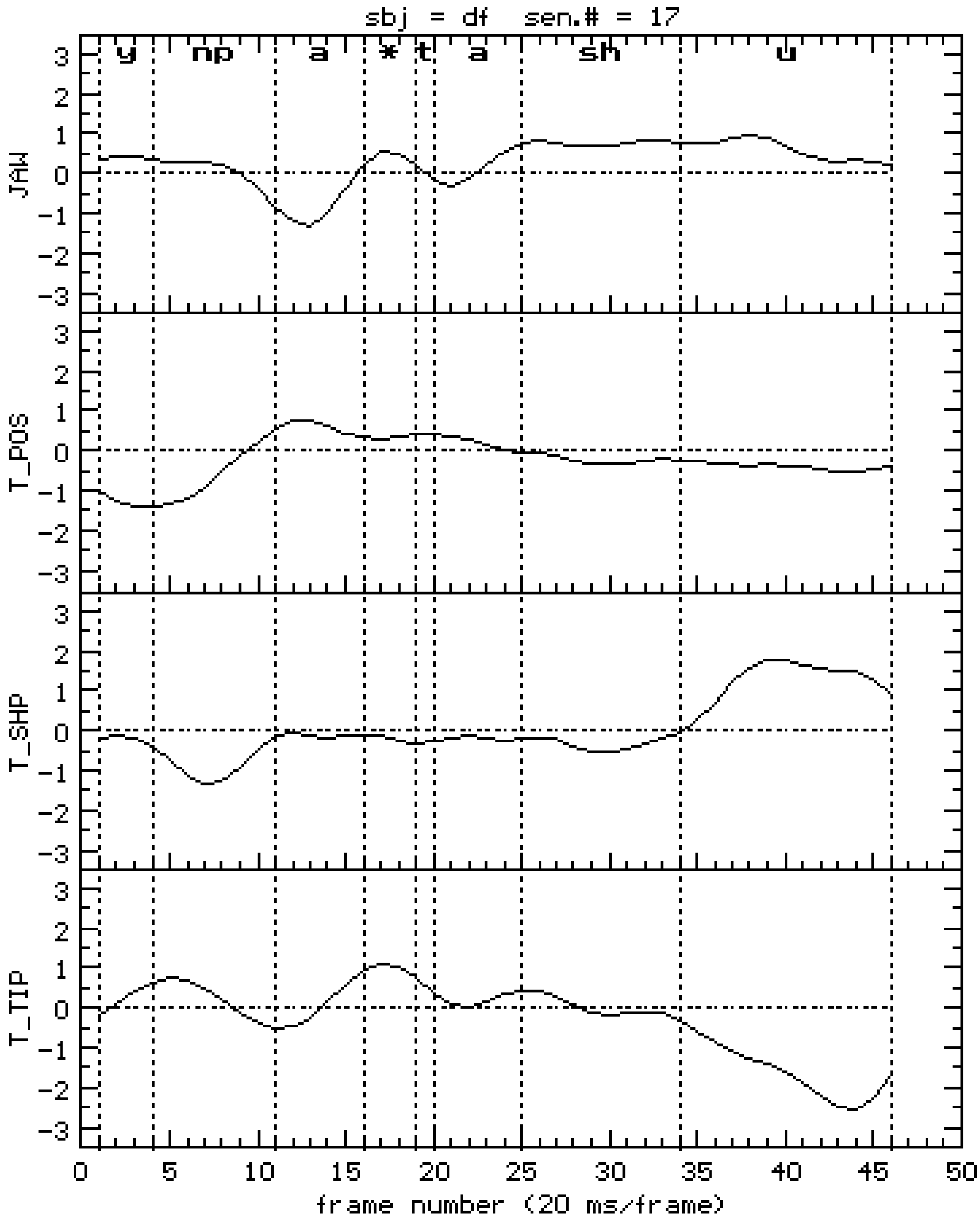


Figure 2.6 *Les patrons temporels des 4 paramètres articulatoires de la langue pour la locutrice DF et la phrase* 17.

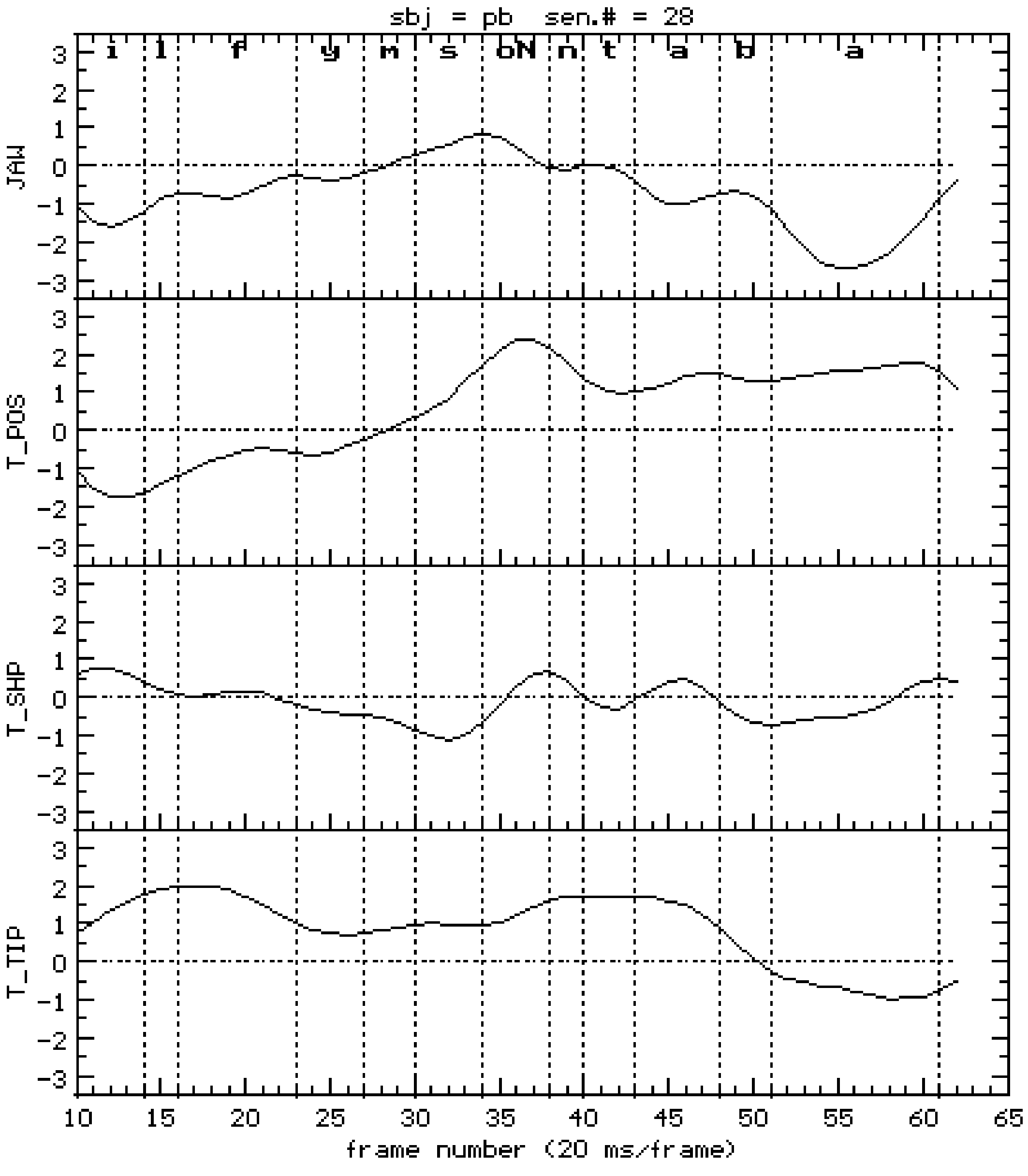


Figure 2.7 *Les patrons temporels des 4 paramètres articulatoires de la langue pour la locutrice PB et la phrase 28.*

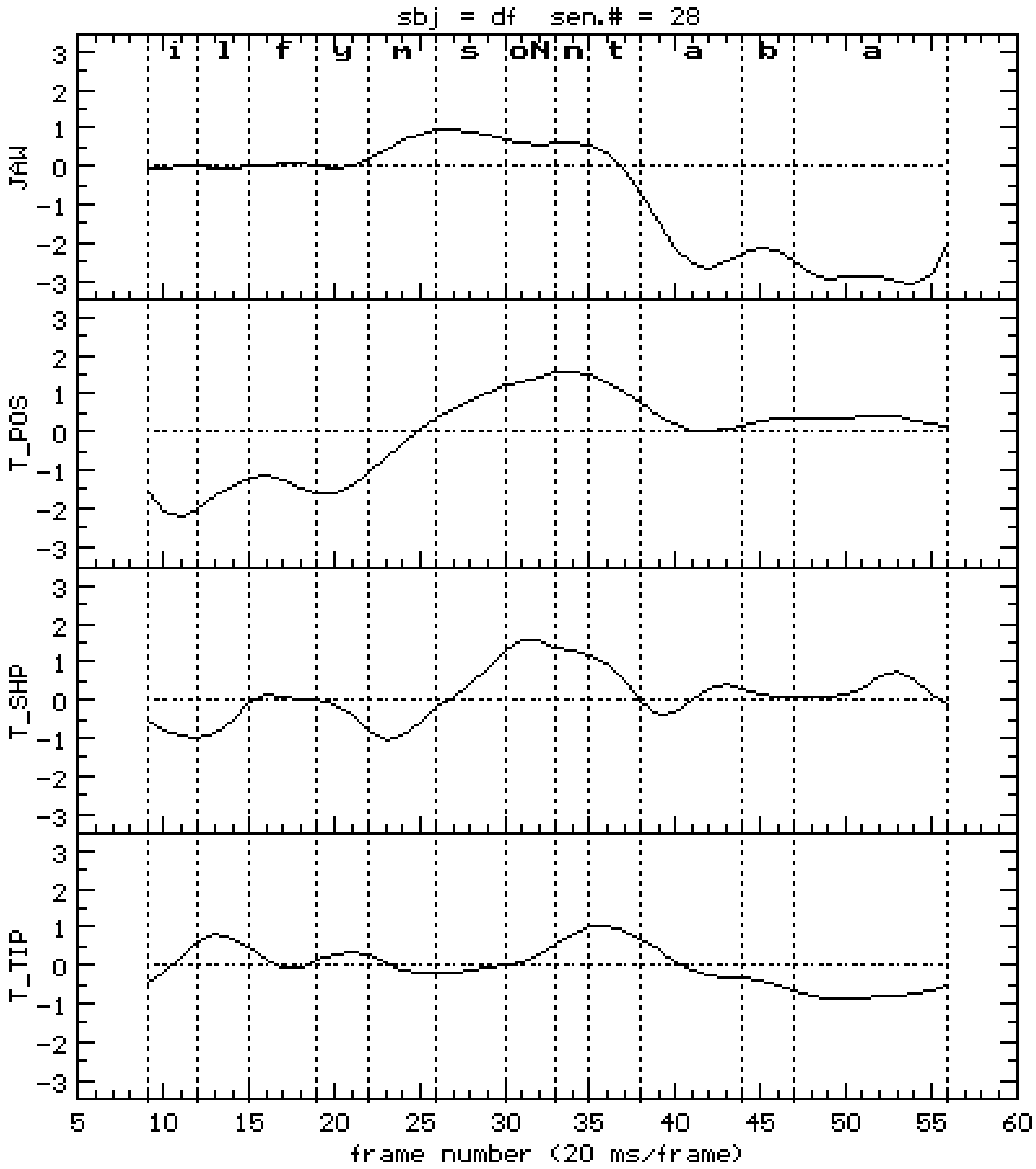


Figure 2.8 *Les patrons temporels des 4 paramètres articulatoires de la langue pour la locutrice DF et la phrase 28.*

Chapitre 3

Analyse des données

3.1 Introduction

La production de la parole requiert l'interaction des multiples articulateurs du conduit vocal dans une large structure fonctionnelle. Un des centres d'intérêt du contrôle moteur est la manière dans laquelle le système nerveux contrôle cette structure. Des études suggèrent que les multiples articulateurs sont modulés de façon interdépendante. Ainsi, la variabilité observée dans l'examen des données expérimentales n'est pas aléatoire mais plutôt systématique.

Dans ce chapitre, nous allons examiner comment les paramètres de la langue sont coordonnés dans les séquences voyelle-consonne occlusive-voyelle décrites dans la section **??**. Il apparaît que tous les paramètres n'ont pas le même degré d'activité dans la production d'un phonème donné. Un articulateur peut, en effet, avoir un rôle primaire ou secondaire, ainsi que nous le verrons ci-dessous.

3.2 Séquences $/pCV_1CV_2/$

Un exemple des mouvements observés est montré dans la figure **??** pour les séquences /pada/ et /paga/. Les mouvements examinés comprennent le déplacement de la mâchoire **jw** et ceux de la position **tp**, la forme **ts** et l'apex **tt** de la langue qui correspondent aux mouvements des quatre articulateurs décrits dans le chapitre précédant. Les deux consonnes sont sélectionnées par rapport au degré d'implication de la langue lors de leur production : l'apex de la langue est impliqué pour /d/, et la forme de la langue est l'articulateur primaire pour /g/. La figure **??** montre une similitude de l'allure des mouvements sur les segments vocaliques. Néanmoins, une différence de niveau d'activité est observée. Ainsi que nous le verrons ci-dessous, les mouvements de la langue sont soit activement impliqués soit passivement contraints selon le phonème produit.

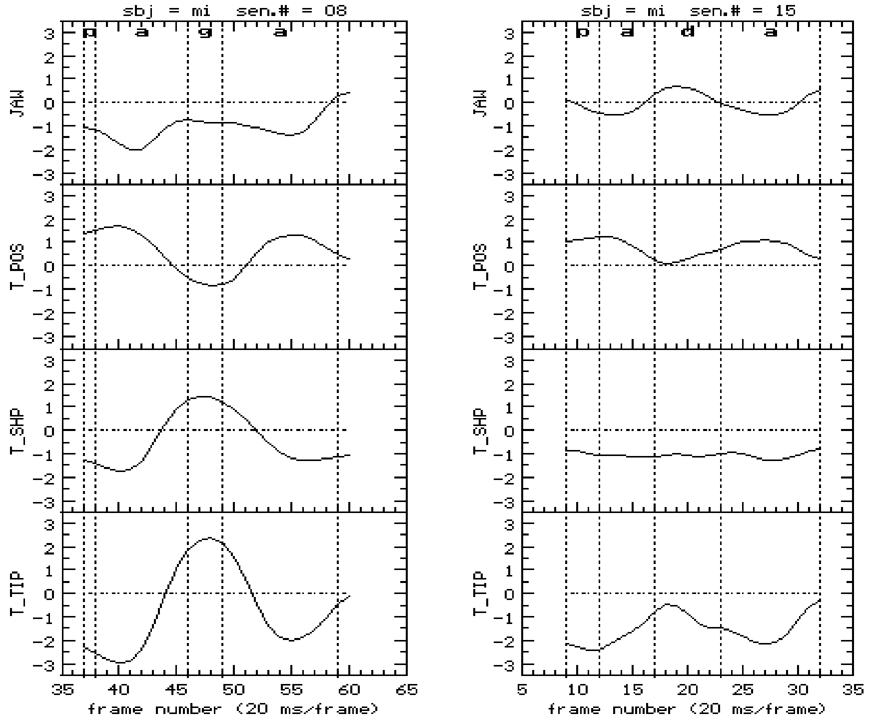


Figure 3.1 *Paramètres articulatoires de la langue pour les séquences /pada/ et /paga/.*

3.3 Relation entre les paramètres articulatoires de la langue

La parole consiste en une séquence de segments qui sont décrits par des traits distinctifs. Ces traits suggèrent comment les articulateurs bougent et donc comment les propriétés acoustiques sont produites par ces mouvements. Pour des types de segments, telles que les voyelles, la description des traits tels que haut/bas et postérieur/antérieur est comparable dans les espaces articulatoires et acoustiques. Les cibles acoustiques (F_1 et F_2, par exemple) indiquent, grosso-modo, quels articulateurs sollicités sont ajustés pour réaliser ces cibles. Pour les consonnes, certains traits, tels que la spécification des articulateurs qui réalisent les constrictions sont décrits en termes articulatoires. La génération des segments fait en général appel à la coordination des mouvements de plusieurs articulateurs. Dans le cas des consonnes, par exemple, la description segmentale spécifie l'articulateur (et sa position) dont le rôle est majeur pour la formation de la constriction dans le conduit vocal (tels que les lèvres, le dos de la langue et le corps de la langue). Les mouvements des articulateurs secondaires doivent être coordonnés temporellement avec le premier articulateur.

Les articulateurs sont assujettis à des sources de variabilité diverses : les variations phonémiques qui marquent les différences des réalisations acoustiques du même phonème, le contexte segmental ou coarticulation progressive et régressive, le contexte suprasegmental ou prosodique tel que l'accentuation et la durée des phonèmes, les caractéristiques du locuteur (sexe, âge, dialecte ...). Ces sources de variabilité peuvent interagir entre elles. Dans cette section, nous allons nous intéresser à ces caractéristiques de la parole.

3.3.1 Gestes vocaliques

Pour examiner les relations qui existent entre les différents paramètres articulatoires de la langue, nous avons représenté les valeurs (positions) des maxima et minima des mouvements sur les segments phonémiques.

Les voyelles /i/ et /u/

Ces deux voyelles sont caractérisées par des valeurs absolues élevées des positions de **tp** pour /i/ et de **ts** pour /u/. L'antériorité du paramètre **tp** est déterminante pour la production de la voyelle /i/. La valeur de ce paramètre est relativement invariante et indépendante du contexte. Ceci se traduit par la formation d'un nuage compact sur l'axe **tp** ainsi que le montre la figure **??**. L'activité des paramètres **jw** et **ts** est également faible (*fig.* **??**-a et **??**-b). Cette voyelle est relativement fermée. La position de **ts** varie légèrement autour de la valeur moyenne.

Le paramètre primaire pour la production de la voyelle /u/ est **ts**. La figure **??**-a) montre une valeur de **ts** pratiquement constante pour les différents contextes. La position de la langue est plutôt postérieure. Elle montre une faible variabilité qui est, toutefois, supérieure à celle de **ts**. Le paramètre le plus variable est **jw** (*fig.* **??**-c). La mâchoire est plus ouverte lorsqu'elle est précédée par une consonne /g/ que par une consonne /d/. Ceci est probablement dû au ballonnement de la langue pour /g/ qui requiert plus d'espace dans la bouche que /d/ et donc l'abaissement de la mâchoire. Cette abaissement de la position de la mâchoire se manifeste comme une coarticulation progressive lors de la voyelle qui suit.

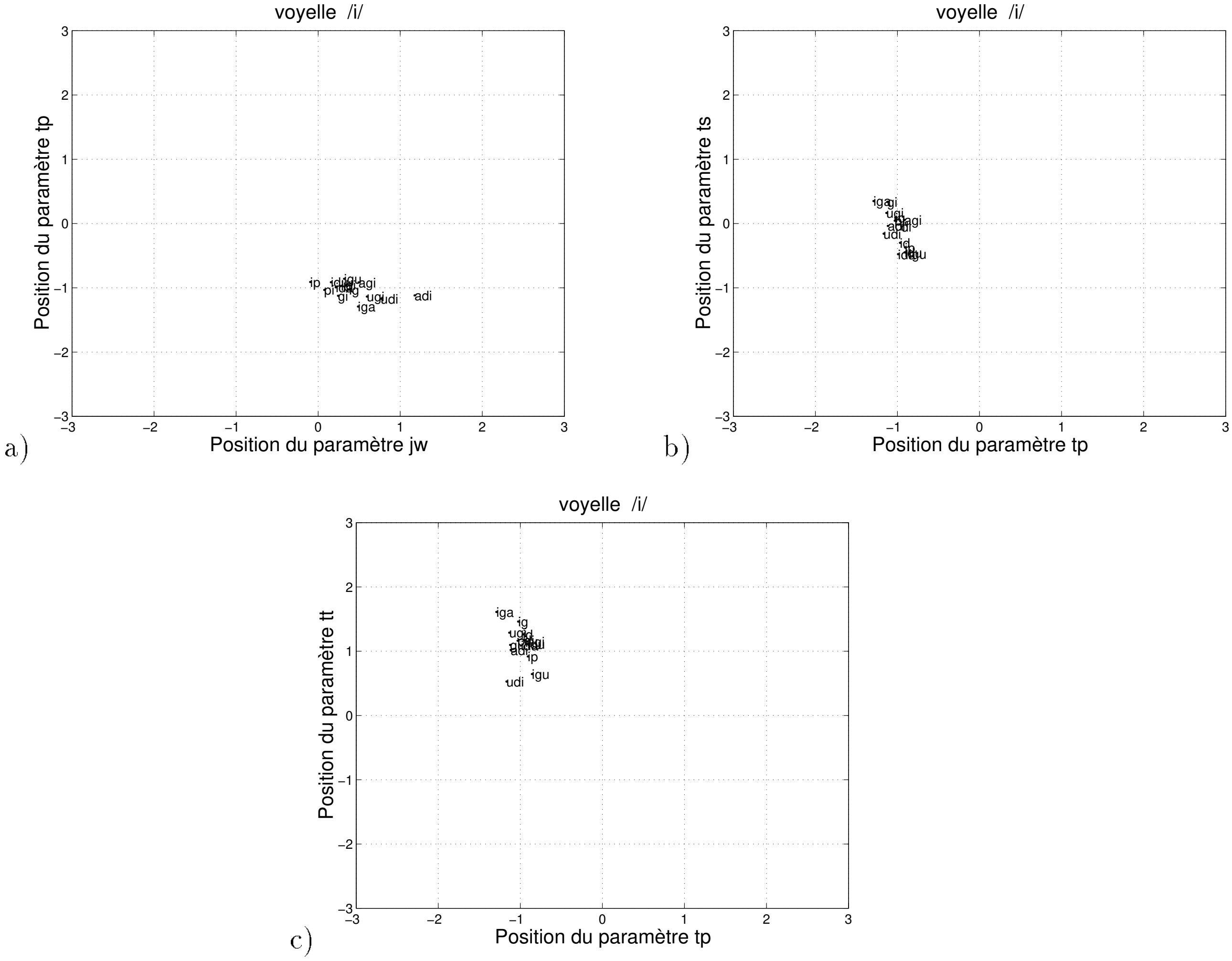


Figure 3.2 *Relation entre les positions des paramètres articulatoires de la langue pour la voyelle /i/ située au début et à la fin des séquences /pV_1CV_2/*

La voyelle /a/

La figure ?? montre une relation linéaire entre les deux articulateurs majeurs, **jw** et **tp**. Ce type de relation résulte de la compensation acoustique (Maeda, 1995 [?]). La déviation d'un paramètre peut être compensée par l'ajustement de la position d'un autre paramètre de manière à ce que le résultat acoustique reste relativement invariant. Cette variabilité articulatoire est liée au contexte. Pour chaque voyelle initiale /a/ précédée par une consonne bilabiale /p/, **jw** est plus ouverte et **tp** est plus postérieure que pour la même voyelle mais en position finale précédée par une consonne occlusive /d/ ou /g/. Des relations linéaires similaires sont observées entre **ts** et **jw** (*fig.* ??-b), et **ts** et **tp** (*fig.* ??-c). Les positions **jw**-**ts** montrent une corrélation positive : l'ouverture de la mâchoire implique l'aplatissement de la forme de la langue. Ces relations entre les paramètres articulatoires signifient que la

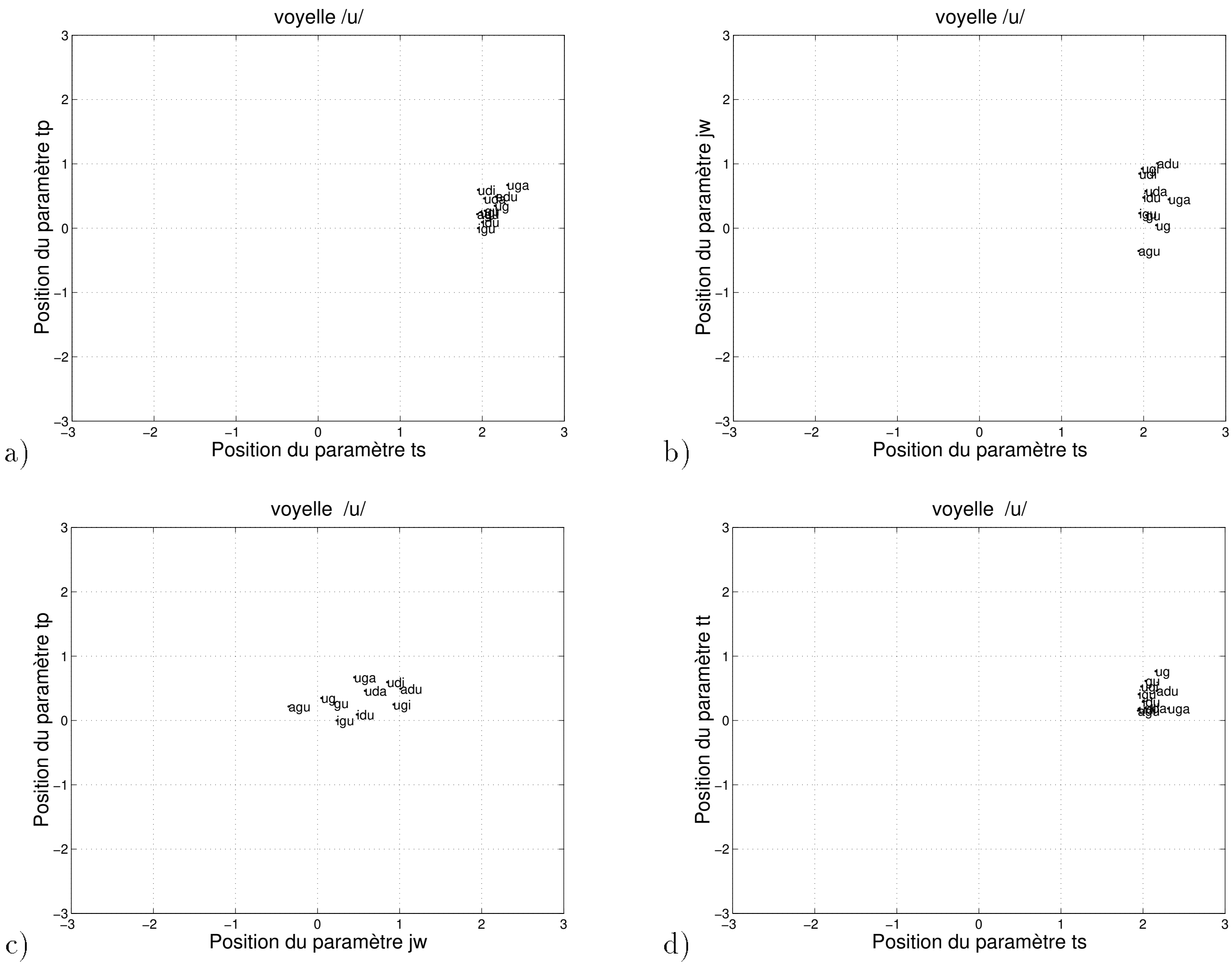


Figure 3.3 *Relation entre les positions des paramètres articulatoires de la langue pour la voyelle /u/ située au début et à la fin des séquences* $/pV_1CV_2/$.

connaissance des positions de certains paramètres sélectionnés permet de déterminer celles du reste des paramètres en exploitant leur structure de coordination.

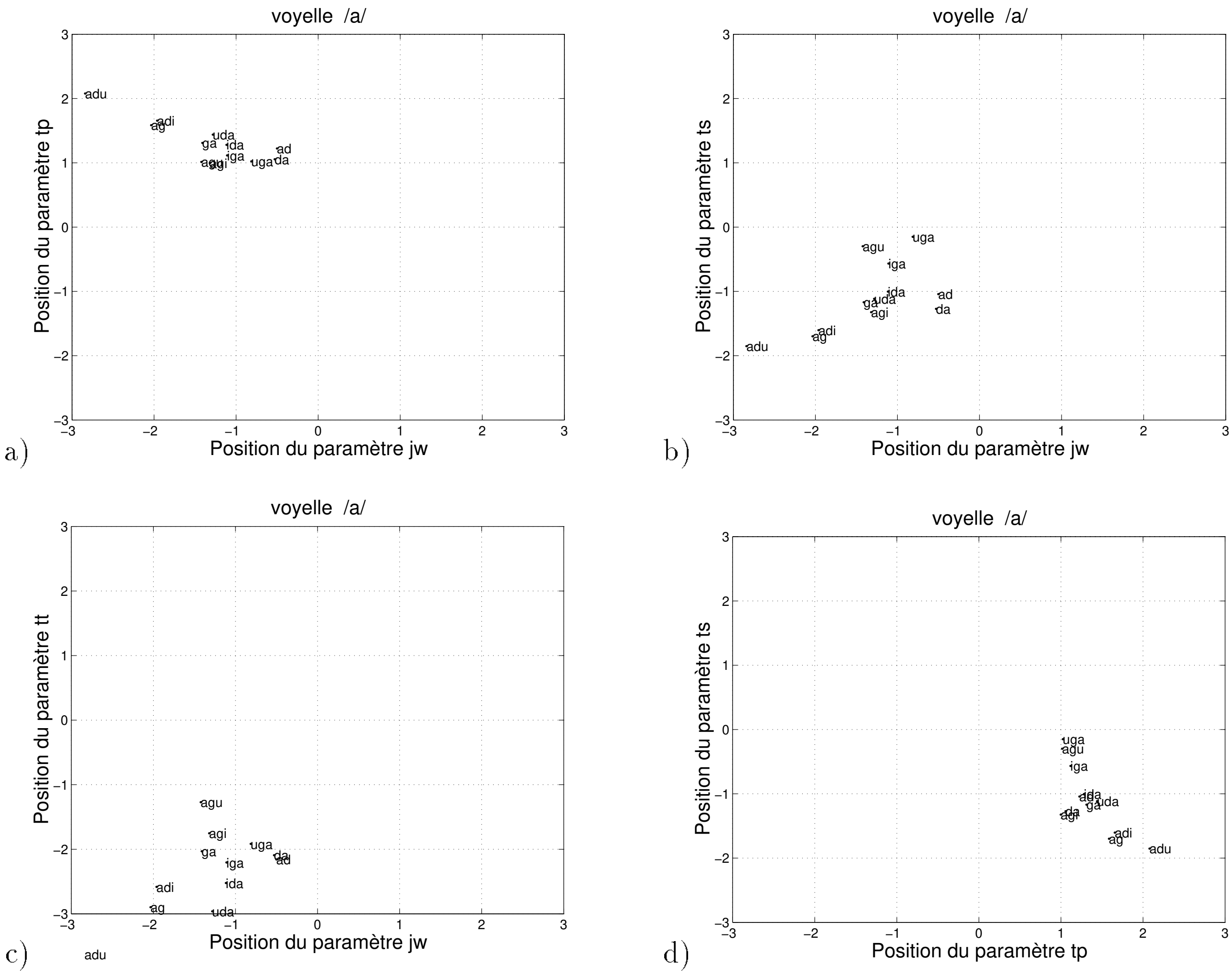
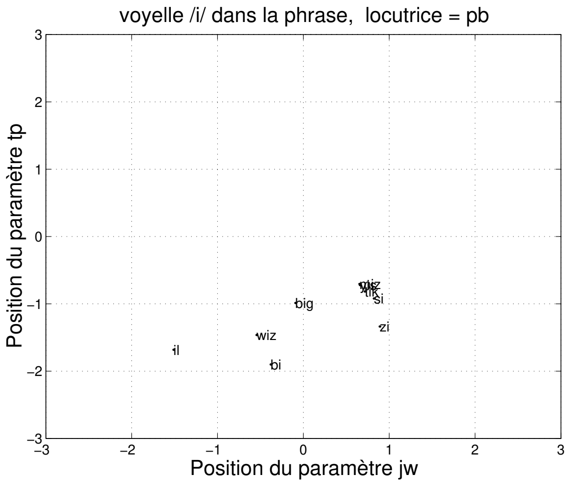


Figure 3.4 *Relation entre les positions des paramètres articulatoires de la langue pour la voyelle /a/ située au début et à la fin des séquences /pV_1CV_2/.*

Le degré de variabilité observé dans les positions des articulateurs pour les voyelles semble dépendre du degré d'aperture de la mâchoire. Les voyelle fermées /i/ et /u/ présentent une faible variabilité, tandis que la voyelle ouverte /a/ montre une importante variabilité qui manifeste une coordination inter-articulateurs. L'invariance relative des positions des articulateurs pour les voyelles fermées semble être particulière à ce type de données, c'est à dire des séquences /CV_1CV_2/ prononcées dans une phrase fixe. En effet, l'analyse des mouvements articulatoires des voyelles prononcées dans des phrases naturelles montre des résultats différents. La figure **??** présente les position de **tp** en fonction de celles de **jw** pour la voyelle fermée /i/ extraite des phrases prononcées par les deux locutrices. Les positions des deux articulateurs des deux locutrices montrent une large variabilité mais sont linéairement corrélés. Le degré de coarticulation est, probablement, plus important dans la production de phrases naturelles que dans celle du corpus de test.

a) 

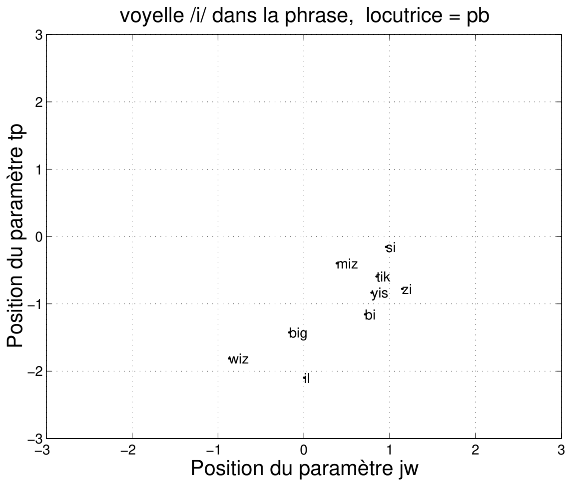
b) 

Figure 3.5 *Relation entre les positions des paramètres* **jw** *et* **tp** *pour la voyelle /i/ extraite des phrases naturelles.*

3.3.2 Gestes consonantiques

Nous allons maintenant examiner la nature et l'étendue de la variabilité dans la production des consonnes.

La consonne /g/

La figure ??-a) montre que les deux articulateurs **tp** et **ts** sont mutuellement dépendants lors de la production de la consonne /g/ ; les positions **tp**-**ts** montrent une relation linéaire. Cependant, la relation entre **jw** et **ts** semble plus complexe en raison de la dépendance contextuelle (*fig.* ??-b). Lorsque la consonne /g/ est précédée par la voyelle ouverte /a/, la position de la mâchoire est relativement basse. Pour réaliser la constriction consonantique, cette ouverture est compensée par un ballonnement accentué de la forme de la langue qui correspond à une valeur de **ts** positive et élevée. Ce phénomène apparaît clairement dans les séquences /paga/ et /pagu/. Lorsque le contexte vocalique est fermé (/pugi/ et /pigi/) la position de **jw** est également haute et la forme de la langue est moins ballonnée, correspondant à une faible valeur positive de **ts**. De plus, on observe une coarticulation régressive de la voyelle arquée /u/ (V_2) sur la forme de la langue qui est nettement plus élevée dans ce cas. Enfin le mouvement de la mâchoire a tendance à être d'autant plus fermé que la position de la langue est antérieure (*fig.* ??-c).

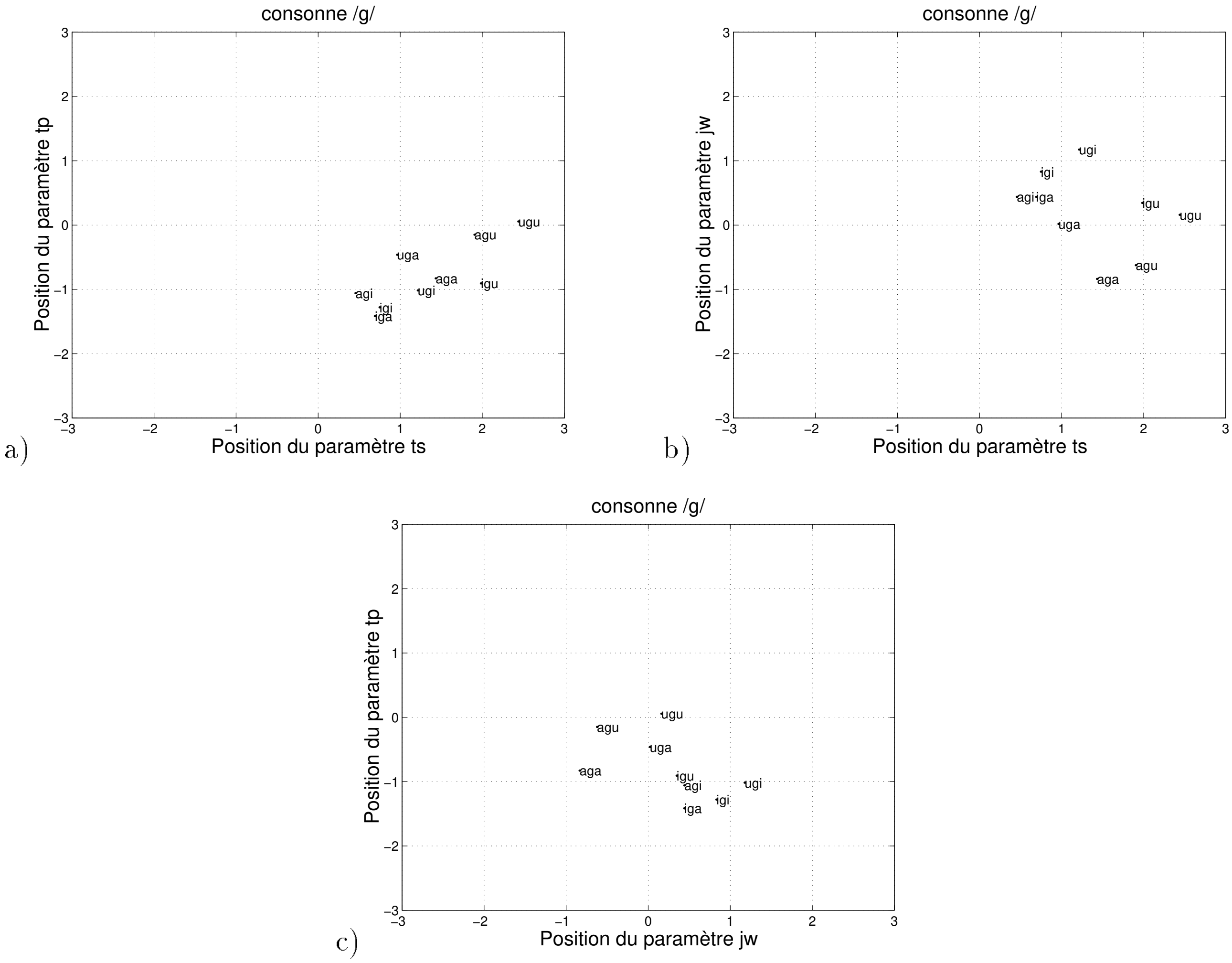


Figure 3.6 *Relation entre les positions des paramètres articulatoires de la langue pour la consonne /g/ dans les séquences /pV_1CV_2/.*

La consonne /d/

L'articulateur primaire pour la consonne /d/, l'apex de la langue **tt**, est linéairement dépendant de la position de la langue **tp**, ainsi que le montre la figure (*fig.* **??**-a). L'apex est d'autant plus haut que le contexte vocalique est antérieur. Dans le cas de /pada/ où la position de la langue est située dans un contexte vocalique fortement postérieur, la position de **tt** est relativement basse. Afin de réaliser la constriction consonantique, un aplatissement du paramètre **ts** est produit (*fig.* **??**-b). Par conséquent, la constriction se fait par le corps de la langue. La position de la mâchoire est plutôt fermée pour cette consonne, en comparaison à la vélaire /g/, ainsi que le montre la figure (**??**-c).

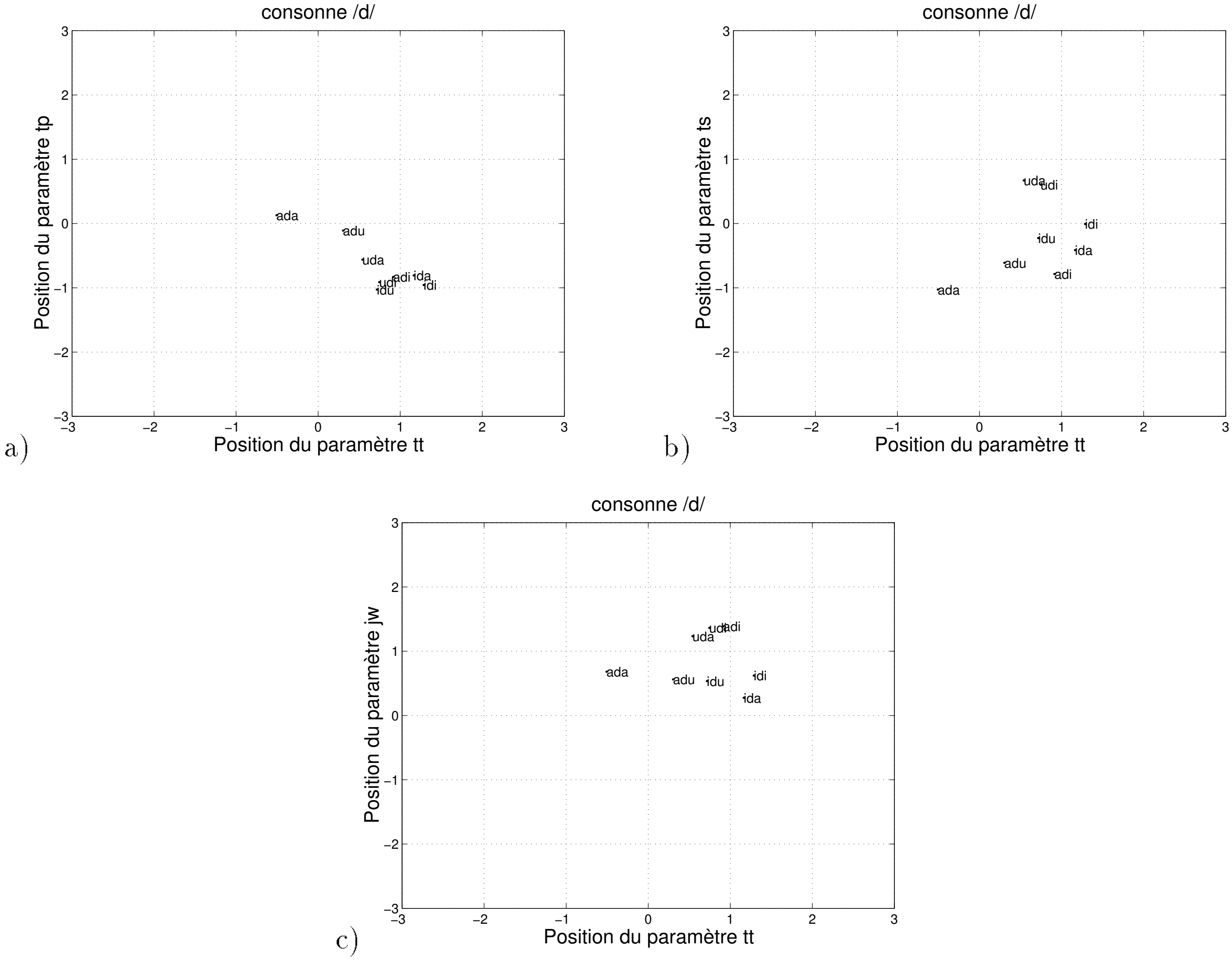


Figure 3.7 *Relation entre les positions des paramètres articulatoires de la langue pour la consonne /d/ dans les séquences /pV_1CV_2/.*

<u>Remarque</u>

En résumé, le tableau **??** présente les valeurs moyennes et les écarts types des positions des quatre articulateurs pour chacune des voyelles extrêmes et des deux consonnes intervocaliques calculées sur l'ensemble des séquences V_1CV_2. Le tableau montre que tous les paramètres ne varient pas de leur position moyenne de la même manière. Une valeur moyenne élevée signifie que l'articulateur participe activement dans la production du phonème. Selon le son produit, certains paramètres s'avèrent plus actifs que d'autres. Dans le cas de /u/, par exemple, l'articulateur **ts** est dominant. Il présente une valeur moyenne élevée et un écart type faible.

D'autres phonèmes nécessitent la participation active de plusieurs articulateurs. Pour la production de la voyelle /i/, les articulateurs activés sont **tp** et **tt**. La valeur moyenne de **tt** est aussi élevée que celle de **tp**. Mais la position de l'articulateur **tp** montre moins de variabilité contextuelle. Aussi, nous sommes tentés de penser que l'articulateur **tt** est secondaire et que sa contribution effective pour la réalisation de /i/ est faible. Dans le cas de la voyelle /a/ et des deux consonnes plusieurs articulateurs sont impliqués. Deux articulateurs, **ts** et **tt**, exhibe une importante activité pour la production de la consonne /g/. L'articulateur primaire **ts** est sensible aux effets contextuels (*cf.* section **??**), sa variance est élevée. L'articulateur **tt** ne contribue pas directement à la réalisation de la constriction vélaire, son activité est, probablement, une conséquence de la continuité élastique du corps de la langue. La voyelle /a/ semble être le produit de la coordination d'au moins trois paramètres à savoir **jw**, **tp** et **ts** qui ont des valeurs absolues élevées et proches. Tandis que la consonne /d/ montre de faibles valeurs moyennes pour l'ensemble des paramètres de la langue. Il est intéressant de remarquer que l'écart type des paramètres, dans le cas multi-articulateurs, sont souvent élevés. Ceci peut être interprété comme étant la manifestation de la coordination inter-articulateurs.

articulateur	*/a/*	*/i/*	*/u/*	*/d/*	*/g/*
jw	-1.36 (0.6)	0.38 (0.3)	0.44 (0.4)	0.69 (0.6)	0.21 (0.6)
tp	1.31 (0.3)	-1.03 (0.1)	0.34 (0.2)	-0.63 (0.4)	-0.78 (0.5)
ts	-1.09 (0.5)	-0.09 (0.3)	2.05 (0.1)	-0.22 (0.6)	1.31 (0.6)
tt	-2.33 (0.6)	1.11 (0.3)	0.37 (0.2)	0.64 (0.5)	1.75 (0.5)

Tableau 3.1 *Moyennes et écarts types (entre parenthèses) des paramètres articulatoires calculés sur l'ensemble des séquences* $/pV_1cV_2/$.

3.4 Conclusion

En conclusion, note analyse montre que la variabilité observée dans les mouvements des paramètres articulatoires de la langue ne s'effectue pas d'une manière aléatoire. Les mouvements des articulateurs sont coordonnés entre eux pour réaliser un son donné.

La suite de notre travail consiste à modéliser les patrons temporels, "mouvements articulatoires", de la langue en utilisant des techniques de traitement du signal. Le but est de contribuer à l'étude de la commande des différents articulateurs de la langue en vue d'appréhender l'organisation spatio-temporelle du contrôle de ces mouvements. Avant d'aborder le chapitre de la modélisation de ces mouvements, nous avons jugé utile de présenter une ébauche sur les mécanismes de contrôle des mouvements musculaires chez l'être humain. Ceci va nous permettre de fournir des interprétations physiques aux résultats d'analyse obtenus.

Chapitre 4

Contrôle des mouvements musculaires

4.1 Description des processus de contrôle des mouvements

La capacité de l'homme à produire un mouvement ne se limite pas à une simple commodité lui permettant de marcher, jouer ou manipuler des objets. Elle a un aspect critique pour son développement intellectuel et même émotionnel. Le mouvement prend différentes formes. Certains mouvements sont génétiques (contrôle des articulateurs, réflexes...) alors que d'autres sont acquis. Leur maîtrise dépend de la pratique et de l'expérience (apprendre à parler, conduire une voiture, écrire à la machine ...).

Depuis le début du siècle, les physiologistes et les anatomistes tentent d'étudier les fonctions des muscles par des stimulations artificielles. Aussi bien acquis que génétiques, les mouvements peuvent être simples ou très complexes. Leur étude comprend deux niveaux, le *niveau biomécanique* et le *niveau neurophysiologique*.

Le premier traite des aspects physiques et mécaniques du système biologique. Afin de comprendre le mécanisme de production des mouvements, il est nécessaire de comprendre le fonctionnement du système musculaire avec l'ensemble des articulations et leurs caractéristiques mécaniques (*cf.* section **??**).

Le second traite du fonctionnement du système nerveux central pour le contrôle des contractions musculaires. Les performances du contrôle de mouvements ou du contrôle moteur sont fonctions de mécanismes variés qui semblent opérer d'une manière dépendante pour produire des mouvements continus avec une économie d'énergie.

Le problème posé dans l'étude des systèmes moteurs est de modéliser le mécanisme biomécanique de production des mouvements, et ensuite de comprendre comment est organisé le système nerveux central pour contrôler les différents muscles et articulations d'une manière parfaitement coordonnée.

4.2 Niveau neurophysiologique

Le système moteur humain est très complexe de par l'interaction des différents processus et mécanismes qui le constituent. En effet, il est probablement inexact de penser que le système moteur humain n'est représenté que par un seul mode de contrôle. Présentement, le seul moyen de comprendre les comportements complexes des mouvements est de considérer l'humain comme un système complexe composé de plusieurs mécanismes séparés. Ainsi, les scientifiques ont identifié des processus de contrôle qui correspondent à des classes de fonctionnement spécifiques de l'ensemble du système moteur. Chaque classe est étudiée d'une manière isolée et indépendante.

Plusieurs facteurs, notamment les informations sur le milieu, peuvent influencer le comportement moteur de différentes manières, que l'on peut classer en deux catégories : le *contrôle à boucle ouverte* et le *contrôle à boucle fermée*. Ces deux processus interviennent dans le fonctionnement du contrôle moteur pour des mouvements différents et sous des conditions différentes. Cette classification est l'une des plus répandue dans le contrôle moteur.

Dans ce chapitre, nous décrivons brièvement les processus à boucle fermée et à boucle ouverte. Nous nous attardons sur ce dernier qui est utilisé pour l'étude de la production de la parole.

4.2.1 Contrôle à boucle fermée

Le mouvement à boucle fermée correspond à un environnement prévisible parce que parfaitement stable comme dans le tir à l'arc, bowling, ... etc. C'est un système servo-mécanique qui implique le traitement de la boucle de réaction par la détermination de l'erreur de correction. Cette erreur est obtenue en comparant le mouvement mesuré et celui désiré.

La théorie de contrôle à boucle fermée est bien adaptée à l'étude des mouvements lents et des mouvements précis. Elle permet également d'expliquer les réflexes.

4.2.2 Contrôle à boucle ouverte

Ce type de mouvement demande une adaptation rapide aux changements imprévisibles de l'environnement. La réponse est préprogrammée. Un bon exemple de mouvement à boucle ouverte est sans doute celui correspondant aux mouvements produits dans un duel où les comportements de l'adversaire sont imprévisibles tels que l'escrime et le catch. Nous présentons ici deux mécanismes qui sont connus dans la production de la parole à savoir le modèle source-filtre ou "Impulse-Timing IT" (Schmidt 1988 [?]) et le modèle masse-ressort ou "Equilibrium-Point Hypotheses EPH" (Asatryan et Feldman 1965 [?]).

4.2.2.1 Modèle source-filtre

Dans ce type de modèle le contrôle de la trajectoire des mouvements est assuré par l'activation d'un muscle. Pour cela, on applique une force excitatrice $F(t)$ dont la durée t_0, l'amplitude A_0 et l'instant d'excitation sont les paramètres à contrôler par le programme moteur. La trajectoire du mouvement musculaire $x(t)$ peut être ainsi entièrement spécifiée par les trois paramètres de la force d'excitation et par les caractéristiques musculaires (*fig.* ??).

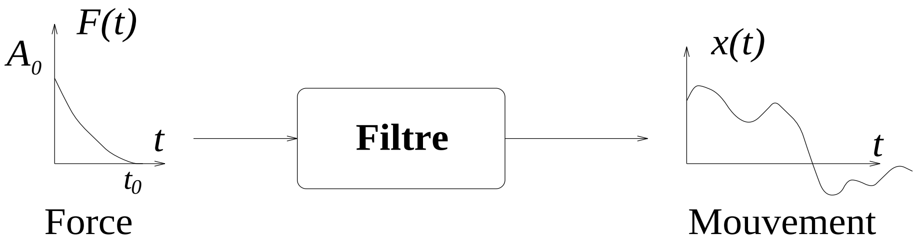


Figure 4.1 *Modèle source-filtre.*

4.2.2.2 Modèle masse-ressort

Les muscles et les articulations sont caractérisés par une élasticité. Cette propriété signifie qu'on peut assimiler le comportement musculaire à celui d'un système de ressort complexe. Dans la figure ??, nous présentons un exemple simplifié de cette modélisation. Les deux ressorts de raideurs $k1$ et $k2$ schématisent les fonctions des muscles agonistes et antagonistes. Les cibles correspondent au point d'équilibre mécanique, spécifié par l'égalité entre les force agonistes et antagonistes. Lorsque les ressorts sont étirés par rapport à leur longueur au repos, les forces en présence sont exclusivement des forces de rappel. Avec ce modèle, la durée et le rythme des impulsions excitatrices n'interviennent pas dans le mouvement de changement de positions.

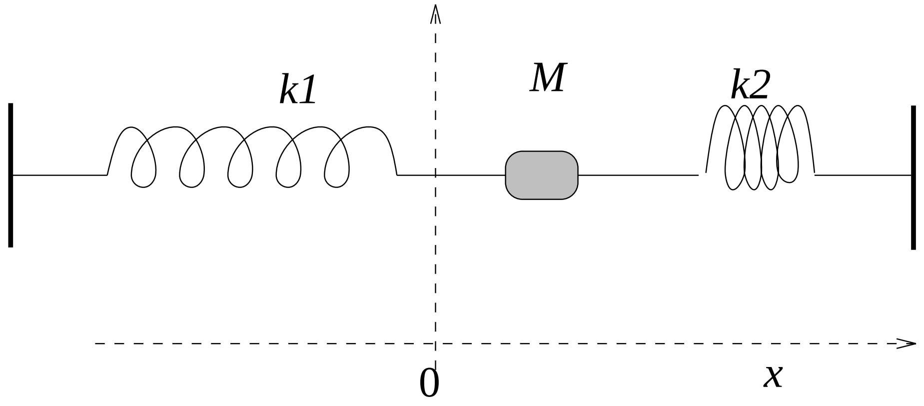


Figure 4.2 *Modèle masse-ressort.*

Le modèle masse-ressort est plus simple que le modèle source-filtre. En effet, la production des mouvements nécessite seulement la connaissance des deux niveaux d'activation ou cibles du muscle spécifiés par les constantes de raideur $k1$ et $k2$.

Cependant, les expériences qui sont en faveur du modèle masse-ressort sont effectuées sur une catégorie de mouvements simples et unidirectionnels. En effet, ce mécanisme semble capable d'expliquer avec précision le mouvement musculaire dans une articulation (le coude, par exemple), mais n'est pas capable de coordonner les mouvements entre plusieurs articulations. Les mouvements qui invoquent plus d'une direction ou les mouvements rapides semblent être mieux expliqués par le modèle source-filtre (Schmidt 1988 [?]).

Les mouvements des articulateurs du conduit vocal durant la production de la parole montrent une grande complexité spatiale ; le modèle source-filtre est donc le mieux adapté à leur modélisation.

4.3 Niveau biomécanique

La complexité des systèmes biologiques et physiologiques réside fondamentalement dans l'ordre élevé et la non linéarité de ces systèmes au niveau moléculaire. La méthode traditionnelle pour étudier de tels systèmes est de commencer avec des modèles macroscopiques simples, ensuite d'augmenter graduellement la complexité du modèle en fonction des nouvelles connaissances acquises et des questions posées.

Deux types de modèles de structures différentes ont été développés pour décrire les systèmes musculaires :

- le premier est basé sur la structure musculaire classique du modèle de Hill (*cf.* section ??). Il consiste en un modèle non linéaire d'ordre élevé qui est décrit par une équation différentielle d'ordre supérieur ;
- le second est basé sur l'analyse des signaux d'entrée/sortie pour une tâche donnée. Il découle de la description de l'équation différentielle du second ordre qui s'avère suffisante pour une variété de mouvements.

Chaque modèle présente des avantages et des inconvénients (Winters 1987 [?]). Le choix doit être défini par les limites et les objectifs de l'étude. Très simple, le modèle peut présenter des carences qui empêchent une compréhension approfondie des propriétés physiologiques importantes du système. Très complexe, le modèle peut conduire à une impossibilité d'exploiter les informations suffisantes liées au comportement du système. On peut ainsi très vite se perdre dans la résolution de nombreux problèmes liés à une analyse détaillée du modèle, tels que l'identification des paramètres ou l'excès de variables inconnues.

4.3.1 Modèle des contractions musculaires

Le muscle présente une flexibilité lui permettant de produire des contractions qui paraissent sous forme de mouvements. L'étude de la visco-élasticité des forces de contractions internes au muscle est indispensable pour clarifier le mécanisme des contractions musculaires. Nous présentons ici le modèle de simulation des phénomènes de contractions musculaires avec des composantes visco-élastiques.

4.3.1.1 Structure et mécanisme de contraction des muscles

La structure anatomique du muscle est montrée dans la figure **??**. Ses contractions sont de deux natures (Akazawa et *al.* 1969 [?]) :

1. *les contractions isométriques*, pour lesquelles la longueur du muscle demeure constante,
2. *les contractions isotoniques* qui entraînent une diminution de la longueur du muscle sans un changement de tension.

Le muscle est composé d'un grand nombre de fibres organisées en unités motrices (*fig.* ??*(b)*). Le nombre de fibres dans une unité motrice varie d'un muscle à l'autre et

Figure 4.3 *Structure typique d'un muscle.*

même dans le même muscle. L'ensemble des fibres d'une unité motrice se contractent simultanément lorsque cette dernière reçoit un stimulus du nerf moteur. Chaque fibre comprend plusieurs sarcomères segmentés par des lignes en zigzag connues sous le nom de *z-lines*. La structure microscopique des sarcomères est illustrée dans les figures **??**(d) et **??**(e) où on note, en particulier, les deux filaments actine et myosine. La contraction musculaire est considérée comme le résultat des réactions entre ces deux filaments. Dans ces réactions, les filaments coulissent avec un faible changement dans leur longueur (the sliding theory) (Huxley 1969 [?]). La contraction du muscle dépend de deux paramètres : le nombre des unité motrices activées et la fréquence d'excitation de ces unités (MacNeilage 1973 [?]), de manière à ce que le nombre d'unités sollicitées par le mouvement soit d'autant plus grand que la force appliquée est grande.

4.3.1.2 Modèle de Hill

Sur la base de la structure des tissus musculaires et de la théorie de glissement des filaments ou "sliding theory", Hill (1938) [?] a élaboré un modèle de représentation simple de la structure musculaire. Chaque élément du modèle, qui est présenté dans la figure **??**, correspond à une partie anatomique du muscle. Ainsi, la constante de raideur de la composante élastique en série K_s inclut les tendons, *z-line* et les filaments actine et myosine. La constante de raideur de la composante élastique en parallèle K_p représente le sarcolemma et les tissus de connexions. Le coefficient de viscosité résistive B correspond à la viscosité due aux mouvements des filaments. La force interne de contraction C est générée par les réactions entre les filaments. Ainsi, le muscle est simulé par un modèle visco-élastique.

Ce modèle a été soutenu par de nombreux groupes de chercheurs, aussi bien pour la modélisation des muscles isolés que pour celle des muscles à plusieurs articulations. Des équations différentielles ordinaires ont été utilisées pour résoudre ce type de modèle (Hill 1938 [?]).

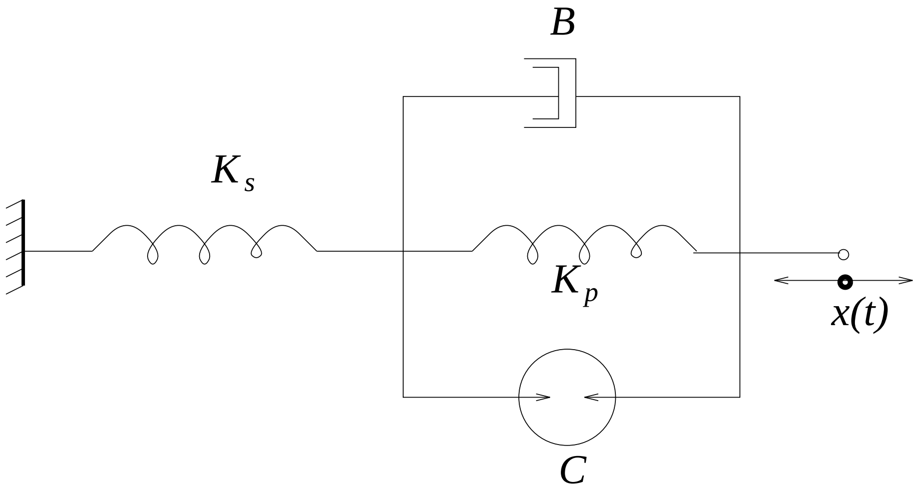


Figure 4.4 *Modèle de la structure musculaire de Hill,* 1938 [48].

4.3.2 Modèle du second ordre

Les mouvements musculaires dynamiques sont souvent modélisés par de simples systèmes linéaires du second ordre, masse-ressort, voire des systèmes du premier ordre, visco-élastique. Ces systèmes sont décrits par l'équation de Newton :

$$M\ddot{x}(t) + R\dot{x}(t) + Kx(t) = F(t), \tag{4.1}$$

qui est caractérisée par une masse M, une résistance mécanique R et un coefficient de raideur K. Le signal de sortie $x(t)$ représente le déplacement de la masse M en fonction du temps. Le signal d'entrée $F(t)$ représente la force d'excitation externe ou la charge.

La figure **??** montre le circuit mécanique équivalent d'un tel système. Le modèle du second ordre traite le système musculaire comme une boite noire à une entrée et une sortie dont le contenu est approximé par les trois paramètres macroscopiques M, R et K.

L'approximation du second ordre a été appliquée dans de nombreuses études des mouvements humains (Yasuhara 1983 [?], Stein et *al.* 1988 [?], Baratta et Solomonow 1990 [?],

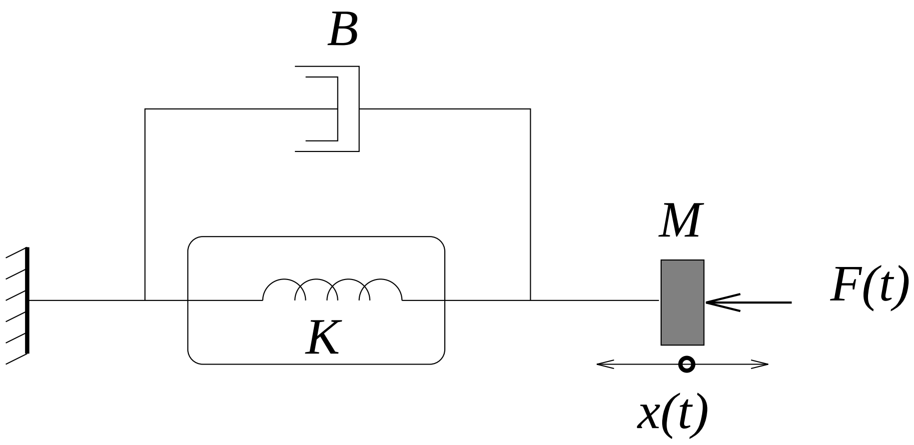


Figure 4.5 *Système mécanique d'ordre 2.*

Bobet et *al.* 1993 [?]). Son utilisation est motivée par le fait que l'extension du modèle de Hill à un muscle entier crée immédiatement un ordre élevé du modèle. Cette approche entraîne donc une complexité qui est indésirable comparée à la simplicité théorique du modèle du second ordre. En revanche, la validité de cette approximation dépend de la nature du mouvement. En effet, les non linéarités dues aux composantes élastiques et aux contractions peuvent exclure l'utilisation des modèles linéaires, notamment lorsque l'activité musculaire est très rapide (Chapman et Harrower 1977 [?]).

Dans ce type de modèle, le problème central est l'identification des paramètres de contrôle du système nerveux pour effectuer les changements de position des articulateurs. Nulle part, ce problème n'est plus apparent que dans l'étude de la production de la parole. Le choix d'un paramètre de contrôle comme la force, la longueur au repos, la raideur, la viscosité ou la vitesse fait l'objet d'un débat toujours ouvert.

4.3.3 Choix des paramètres de contrôle

Plusieurs chercheurs ont utilisé différents paramètres pour contrôler les modèles des mouvements articulatoires. Pour analyser le mouvement d'un point de la langue Sonoda et Kiritani (1976) [?] ont approximé le mouvement dynamique par la sortie d'un système linéaire du second ordre. La raideur et l'amortissement du modèle sont estimés de manière à approximer le mieux possible le mouvement observé de la langue pour les différentes transitions CVC.

Plus récemment, des études sur les effets de l'accentuation et la vitesse d'élocution durant la parole se sont basées sur le contrôle de la raideur. Dans ce type de problème, le changement de stress est directement exprimé par la variation de la raideur de l'articulateur et la relation entre le déplacement et la vélocité du mouvement est modélisée par un système masse-ressort où la raideur du ressort est la variable à contrôler.

Nous citons, en particulier, les travaux de Browman et Goldstein (1985) [?] qui ont montré comment la fréquence du mouvement vertical de la lèvre inférieure, et donc le coefficient de raideur d'une équation du second ordre non amortie, varie en fonction des différentes unités linguistiques telles que le type de la voyelle et la position du stress dans une trisyllabe.

Plus tard, ces travaux ont été repris par Smith et Browman (1993) [?] mais en examinant, cette fois-ci, les conséquences de la variation du coefficient d'amortissement. Ils ont montré la robustesse de l'effet des facteurs linguistiques sur la fréquence naturelle (coefficient de raideur) suivant différentes valeurs du coefficient d'amortissement.

Par ailleurs, Ostry et Munhall (1985) [?] ont montré que la croissance du rapport vitesse maximale à l'amplitude du mouvement du dos de la langue, qui diminue la durée du mouvement, peut être produite par un système du second ordre dans lequel le coefficient de raideur croît.

Enfin, pour modéliser la trajectoire du point le plus haut de la langue, Perrier et *al.* (1989) [?] ont proposé une modélisation distribuée à deux ressorts où les paramètres de contrôle sont la raideur et le temps (le temps de transition entre cibles et la tenue de la cible). Ce modèle repose sur le principe du modèle masse-ressort ou "equilibrium point hypotheses" (*cf.* section **??**).

Tous ces travaux sont, en partie, basés sur les résultats obtenus par Cooke (1980) [?] dans son étude du mouvement du coude où il influe directement sur la raideur pour passer d'une cible à l'autre.

D'autres auteurs se sont intéressés à d'autres paramètres de contrôle. Asatryan et Fel'dman (1965) [?] ont montré que les mouvements peuvent être produits en agissant sur la longueur au repos des muscles agoniste et antagoniste sans contrôle de la raideur des muscle du bras.

Lindblom (1967) [?] a proposé un modèle articulatoire pour l'étude de la durée des voyelles en représentant les articulateurs des lèvres et de la mandibule par un système amorti masse-ressort invariant dans le temps où les forces appliquées ont la même amplitude mais des durées différentes.

Pour modéliser les mouvements articulatoires de la langue, nous avons choisi un filtre invariant. Ce choix est motivé par la simplicité du modèle qui s'exprime en terme d'économie d'énergie. En effet, Nelson (1983) [?] a constaté que dans la plus part des cas, les mouvements précis résultant d'un apprentissage satisfaisaient à des principes physiques d'économie. Ces principes sont imposés par les limites de l'effort musculaire susceptible d'être produit ; il y aura ainsi minimisation de la durée, ou de la force, ou de la vitesse maximale ou de l'énergie ou encore des variations de l'accélération. Il a ensuite mis en évidence que les profils de vitesse pour les deux cas correspondant à l'énergie et aux variations de l'accélération étaient très proches de celui d'un système du second ordre.

Enfin, à partir de mesures faites avec le système "microbeam" sur deux locuteurs, il a montré que les déplacements de la mâchoire suivaient les minimisations de l'énergie et des variations de l'accélération. Or pour un modèle du second ordre à raideur constante, le patron de vitesse $V(t)$ est identique à celui obtenu en minimisant les variations de l'accélération ; lui-même très proche de celui obtenu à énergie minimale. Ces résultats sont illustrés dans la figure ??.

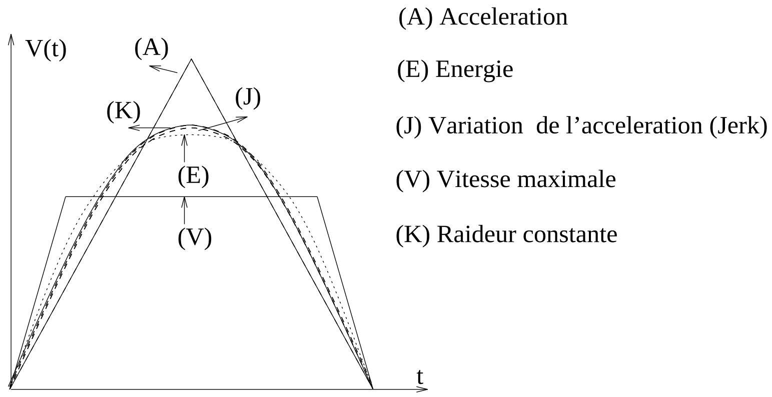


Figure 4.6 *Comparaison des profils de vitesse pour des mouvements de même durée et de même amplitude correspondant à la minimisation de différents coûts.*

Ceci suggère que le modèle masse-ressort avec une constante de raideur invariante est un modèle optimum pour la production de mouvements lisses avec une économie d'énergie.

Chapitre 5

Modélisation des patrons temporels des paramètres articulatoires

Nous avons vu dans le chapitre précédant que le modèle source-filtre est un bon compromis entre la modélisation et l'identification des mouvements articulatoires d'une part, et l'analyse de la coordination du contrôle moteur d'autre part. Dans le processus de production de la parole, on ne dispose que de la sortie du système, en l'occurrence les variations temporelles de la forme du conduit vocal. Dans ce cas, la modélisation source-filtre consiste à séparer la source (commande musculaire) du filtre (la réponse impulsionnelle) en partant des seules données d'observations. Une solution consiste à identifier d'abord le filtre et ensuite, à déterminer l'excitation. Dans ce chapitre, nous nous intéressons à l'identification du modèle du filtre par la méthode de la LPC (Linear Predictive Coding).

L'efficacité de cette méthode dépend de la nature du bruit dans le signal observé. Lorsque le bruit est quelconque, il est difficile d'identifier le modèle. Or le signal observé qui correspond aux mouvements articulatoires est composé d'un signal utile et d'un bruit additif de mesure dont les caractéristiques sont inconnues. Aussi, on se propose, dans un premier temps, d'exposer deux méthodes de débruitage : la méthode par filtrage et la méthode par projection du signal dans l'espace des composantes principales. Les résultats d'analyse par la méthode source-filtre des données débruitées sont présentés et discutés dans un second temps.

5.1 Méthodes de débruitage

Les mouvements obtenus à partir des images radiofilms sont entachés de bruit. Ce bruit est généré par deux sources différentes :

1. *Un bruit de mesure* dû au tracé manuel, trame par trame, des images radiofilms qui donnent les patrons temporels des profils du conduit vocal.

2. *Les effets de bord* dus aux discontinuités présentes au début et à la fin de chaque phrase. En effet, seules les images correspondant à la parole ont été numérisées.

Dans la phase de prétraitement de l'analyse des données par la méthode LPC, nous avons testé deux méthodes de débruitage : la méthode de *débruitage par filtrage* (Duvaut 1991 [?]) et la méthode de *débruitage par projection du signal dans l'espace des composantes principales* (Cooley et Lohnes 1971 [?]).

5.1.1 Débruitage par filtrage

Les variations relativement rapides, dues aux erreurs de mesure, observées sur les mouvements $y(n)$ sont traduites par des fréquences élevées. Un filtrage passe-bas permet de lisser les mouvements. Il consiste en un simple moyennage temporel des données sur une fenêtre de longueur L échantillons

$$s(n) = 1/L \sum_{i=-L/2}^{L/2} y(n-i), \qquad n = 0, ..., N, \qquad L = 3. \tag{5.1}$$

Pour atténuer les effets de bord, la façon la plus immédiate de tronquer un signal $s(n)$ sur un intervalle de temps $\{0, ..., N-1\}$ est de le multiplier par une fenêtre $w(n)$ de durée finie, comme suit :

$$s(n) \triangleq s(n)w(n) \tag{5.2}$$

Un nombre considérable de fenêtres existe. Les plus courantes sont : hamming, hannig, regtangulaire, triangulaire et blackman (Duvaut 1991 [?]). La fenêtre de Hamming a été retenue pour notre application. Elle est définie par :

$$w(n) = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1}), \qquad 0 \leq n \leq N-1 \tag{5.3}$$

Le résultat de cette méthode de débruitage est présenté dans la figure **??** où le signal débruité $s(n)$ *(trait plein)* est superposé au signal observé $s_{obs}(n)$ *(pointillé)*. Les variations rapides présentes dans le mouvement $s_{obs}(n)$ sont complètement éliminées.

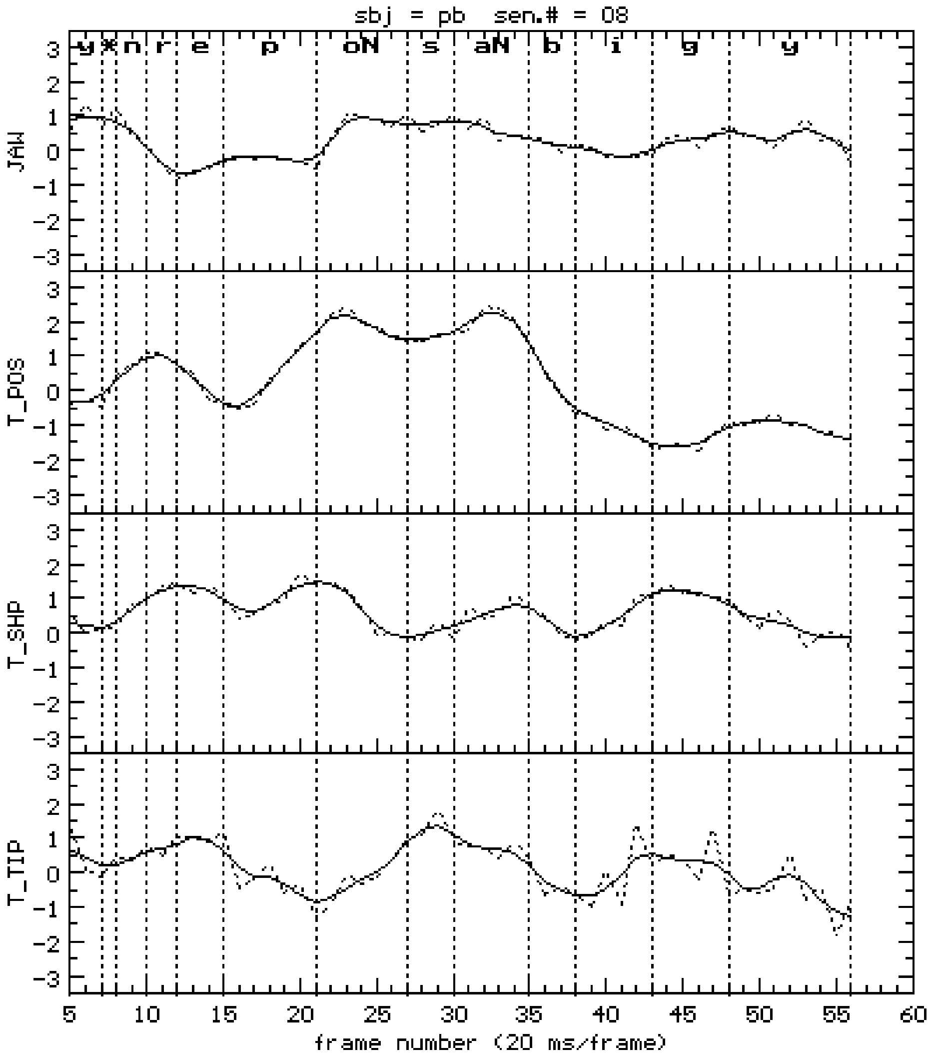


Figure 5.1 *Débruitage des mouvements observés par la méthode de filtrage.*

5.1.2 Débruitage par projection du signal dans l'Espace des Composantes Principales (ECP)

L'objet de cette méthode est de trouver un espace de représentation du signal observé de manière à ne garder que les informations les plus pertinentes.

On dispose de n vecteurs dans un espace de dimension d, c'est à dire d'une matrice d'observations $\mathbf{X}$ $(n \times d)$:

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) \tag{5.4}$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nd} \end{bmatrix} \tag{5.5}$$

Il s'agit de trouver un sous-espace W de dimension $d' < d$ qui représente au mieux les points de l'ensemble des observations. Pour cela, il suffit d'introduire la matrice de variance Σ définie par :

$$\Sigma = 1/n \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^t \tag{5.6}$$

Cette matrice est un indicateur de dispersion de l'ensemble des observations. Le sous-espace W est engendré par un sous-ensemble de vecteurs propres $\mathbf{v}_i$ de la matrice Σ qui forment une base orthogonale.

Pour le choix de la dimension d' du sous-espace de projection, on représente sur une courbe (D) le pourcentage cumulé d'inertie expliquée par les valeurs propres λ_i. On arrête d'ajouter des axes lorsque cette courbe présente un palier (*fig.* **??**). Le sous-espace W est engendré par les vecteurs propres correspondant aux premières valeurs propres situées avant la rupture de la courbe. Dans la figure **??**, nous avons tracé en pointillé la courbe dérivée de (D) qui permet de lire directement le point de rupture. Celui-ci correspond à l'abscisse du premier point pratiquement nul de (D').

Le tableau **??** montre les variances exprimées obtenues pour les différents paramètres articulatoires. Les paramètres **jw** et **tp** semblent être robustes au bruit. Ils comportent plus de 90% de la variance totale. Cependant, les valeurs de l'inertie expliquée du paramètre **tt** de la locutrice PB et des paramètres **ts** et **tt** de la locutrice DF sont relativement faibles. Rappelons que ces paramètres correspondent aux dernières composantes du modèle articulatoire de Maeda (*cf.* section **??**). Ceci signifie que les paramètres articulatoires sont d'autant plus sensibles au bruit de mesure que leur contribution à la reconstitution de la coupe sagittale de la langue est faible.

Inertie expliquée (%)	jw	tp	ts	tt
PB	91 (7)	94 (6)	90 (6)	65 (5)
DF	92 (6)	94 (6)	78 (6)	80 (4)

Tableau 5.1 *Inertie expliquée des paramètres articulatoires de la langue correspondant à la dimension de l'espace de projection définie (entre parenthèse).*

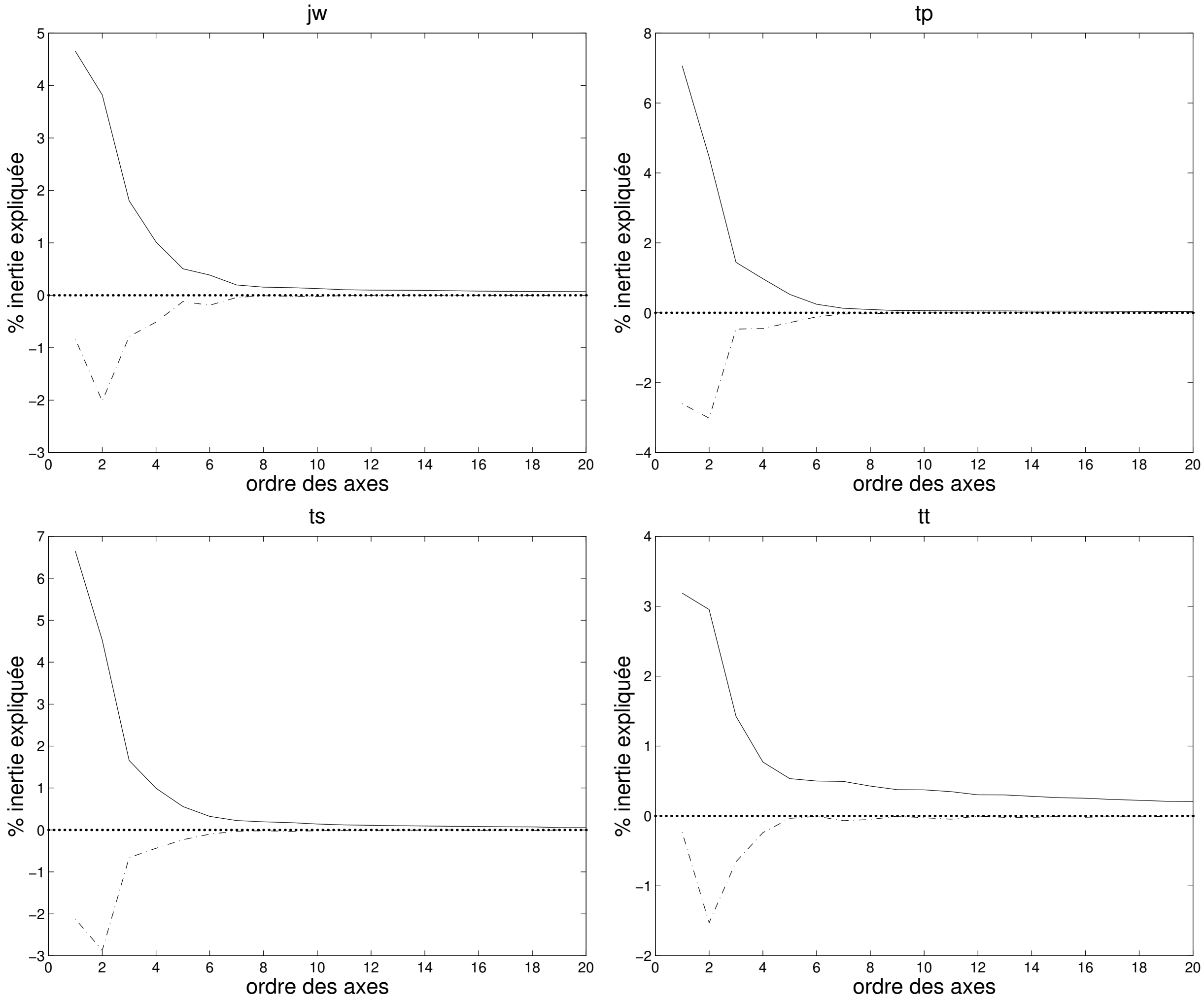


Figure 5.2 *Choix de la dimension de l'espace de projection.*

Le vecteur dont les composantes sont les projections des points $\mathbf{x}_i$ sur le k^{eme} vecteur $\mathbf{v}_k$ est appelé k^{eme} composante principale $\mathbf{s}_k$: ce vecteur a donc n composantes. La matrice $\mathbf{S}$ de reconstitution du signal est de dimension $(n \times d')$. Elle s'écrit :

$$\mathbf{S} = \mathbf{X}\mathbf{V}' \tag{5.7}$$

avec $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_{d'})$ et $\mathbf{V}' = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{d'})$.

La figure **??** montre un exemple de mouvements débruités par cette méthode.

Qualitativement, les résultats obtenus par les deux méthodes sont proches, en particulier pour les trois premiers paramètres. Le dernier paramètre **tt** obtenu par la deuxième méthode est plus lisse. Ces observations ne permettent pas de choisir la méthode à utiliser. C'est pourquoi, nous avons décidé d'appliquer l'analyse LPC sur les signaux obtenus par les deux méthodes de débruitage. Notre choix sera guidé par les résultats de l'analyse.

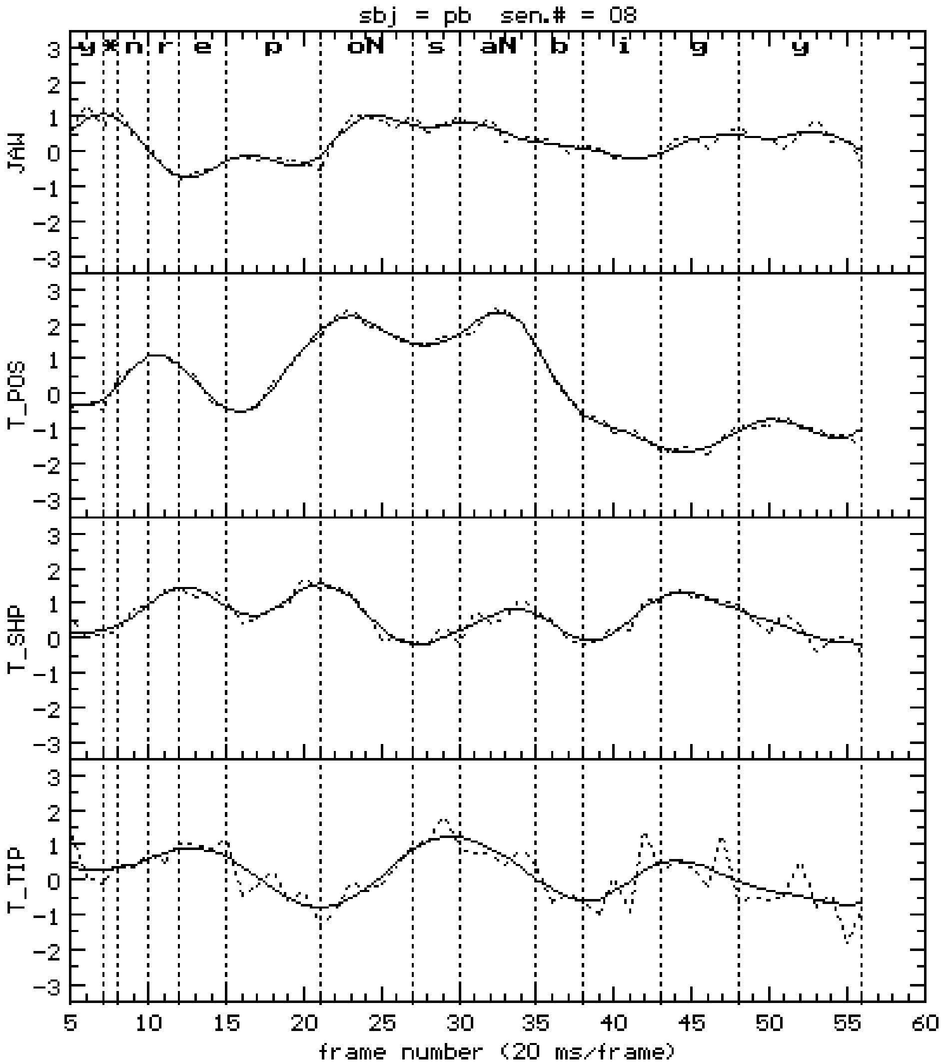


Figure 5.3 *Débruitage des mouvements observés par la méthode de projection du signal dans l'espace des composantes principales.*

5.2 Modèle des paramètres articulatoires de la langue

5.2.1 Description

La langue peut être considérée comme étant composée de plusieurs sous-systèmes mécaniques contrôlés indépendamment les uns des autres (Öhman, 1967 [?]). Cette hypothèse est bien adaptée au cas des paramètres du modèle de Maeda qui vérifient l'orthogonalité statistique (*cf.* section ??). On suppose donc qu'on peut décrire la variation temporelle de chaque paramètre articulatoire (mouvement) séparément. La figure ?? donne le schéma bloc du principe de la modélisation des paramètres de la langue.

Chaque sous-système est basé sur le principe du modèle source-filtre où la commande dépend entièrement de la force d'excitation (*cf.* section ??). Le signal de sortie qui correspond au mouvement $s(n)$ est produit par un filtre invariant. Ce dernier représente la partie neuro-musculaire. La source, elle, correspond à la commande musculaire.

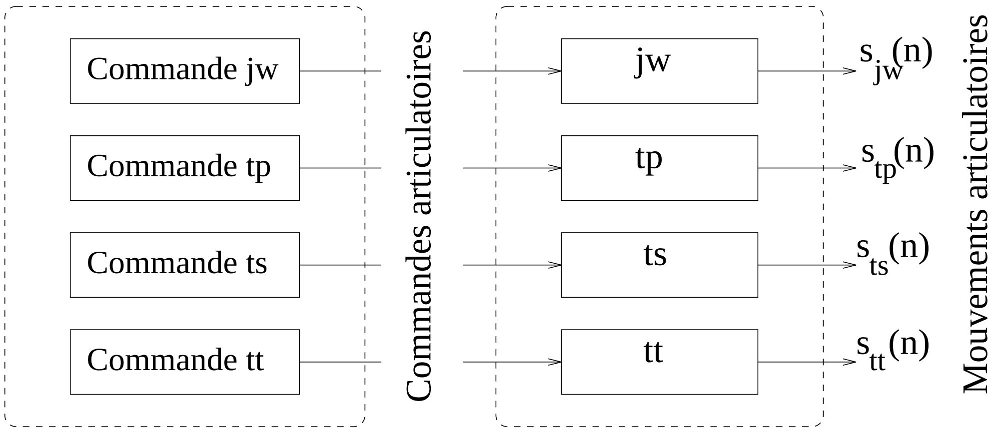


Figure 5.4 *Schéma bloc de la commande des différents articulateurs.*

On distingue deux systèmes dans la partie neuro-musculaire. Un système physiologique (entre le niveau neuro-moteur et la force musculaire) se comporte comme un générateur de la force musculaire. Un système mécanique (entre la force et le mouvement articulatoire) qui réalise les caractéristiques de transfert de l'articulateur. Ces deux systèmes sont placés en cascade ainsi que le montre la figure ??. Ainsi, la commande $v(n)$ est transformée en un signal $F(n)$ qui va exciter le système mécanique pour produire le mouvement $s(n)$.

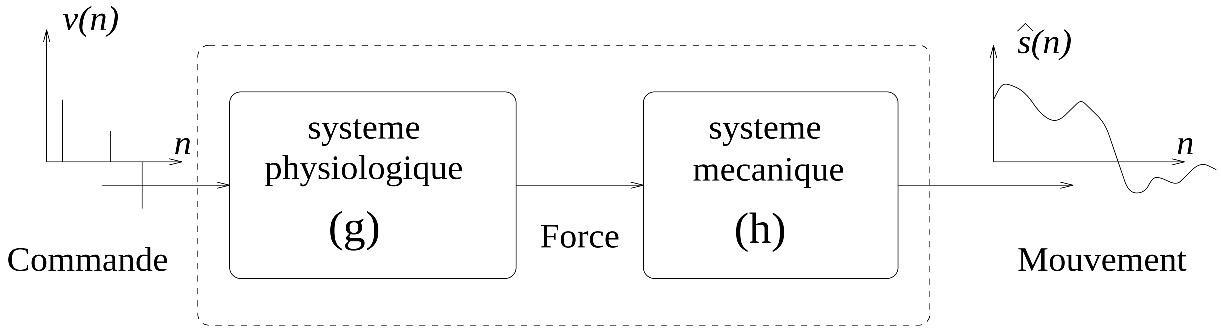


Figure 5.5 *Représentation de la chaîne de production de la parole.*

L'analyse des mouvements articulatoires de chaque paramètre va donc consister à identifier deux filtres g et h invariants dans le temps qui correspondent aux systèmes de la chaîne de production de la parole. Ils représentent le modèle de synthèse.

Le mouvement synthétisé $\hat{s}(n)$ est obtenu par la convolution de l'entrée $v(n)$ avec le modèle $g * h$ comme suit :

$$\hat{s}(n) = v(n) * (g(n) * h(n)) \tag{5.8}$$

À l'issue de cette analyse, nous établirons une analogie entre notre modèle de synthèse et la chaîne de production (*cf.* section **??**).

5.3 Analyse du modèle

5.3.1 Filtre de dé-accentuation, g

On suppose que le sous-système g est un filtre de dé-accentuation. Il est obtenu par un filtre intégrateur du premier ordre qui dépend d'un coefficient α ainsi que le montre l'équation de la fonction de transfert :

$$\mathbf{G}(z) \quad = \quad \frac{1}{1-\alpha z^{-1}}. \tag{5.9}$$

La réponse impulsionnelle de ce filtre est donnée par l'équation suivante (*fig.* **??**) :

$$g(n) \quad = \exp(-n/\tau) \tag{5.10}$$

où $\tau = \log \alpha$.

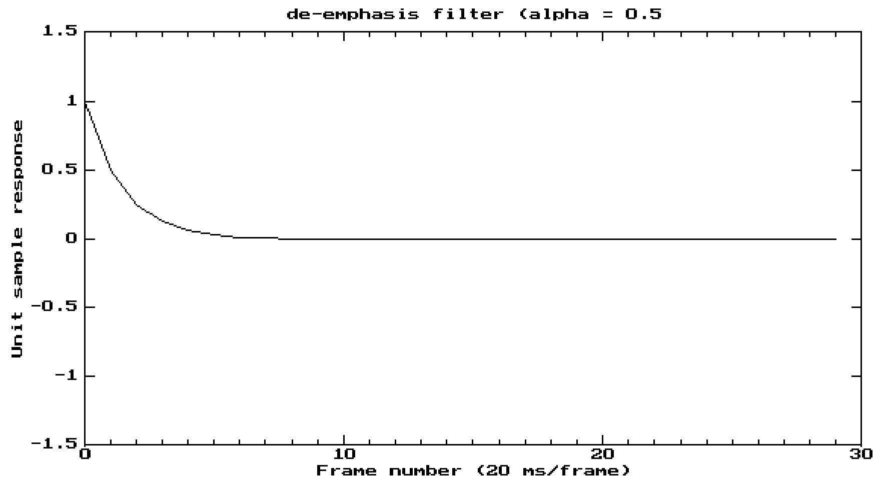


Figure 5.6 *Réponse impulsionnelle du filtre intégrateur d'ordre* 1.

Le temps de réponse T_r d'un système du premier ordre est sensiblement égal à 3 fois sa constante de temps τ, c'est-à-dire, T_r est proportionnel à α. Ainsi, la réponse impulsionnelle du filtre g est-elle d'autant plus brève que le coefficient α est petit. Nous verrons en **??** l'importance du choix du coefficient α sur la détermination du signal d'excitation en synthèse.

Remarquons, par ailleurs, que la réponse impulsionnelle de ce système est fortement amortie.

En phase d'analyse, le filtre inverse g^{-1}, appelé filtre de pré-accentuation, permet d'aplatir la pente spectrale du signal d'observation, afin de faciliter la détection des pôles du modèle h que nous présentons ci-dessous.

5.3.2 Modèle auto-régressif AR, h

L'analyse spectrale par la modélisation auto-régressive est une technique de haute résolution d'analyse spectrale paramétrique, en traitement du signal. Elle doit son succès surtout à sa facilité de mise en oeuvre. L'estimation des valeurs de paramètres du modèle auto-régressif est relativement facile, car elle se fait par la résolution d'un système d'équations linéaires. Par ailleurs, l'analyse spectrale d'enregistrements courts fournit d'excellents résultats.

Le sous-système h est un filtre auto-régressif AR (tout pôles) pour lequel le signal de sortie $s(n)$ à l'instant n est estimé par la combinaison linéaire de p échantillons précédants du signal $\{s_{n-1}, ..., s_{n-p}\}$ pondérés par les coefficients $\{a_i, i = 1, ..., p\}$:

$$\hat{s}(n) \quad = \quad -\sum_{i=1}^{p} a_i s_{n-i} \quad n = 0, ..., N \tag{5.11}$$

les coefficients $a_i, i = 1, ..., p$, sont choisis de telle sorte qu'on minimise l'énergie de l'erreur de prédiction ϵ_n : $\mathrm{E}\,[\epsilon_n^2] = \mathrm{E}\,[(s(n) - \hat{s}(n))^2]$.

Le processus décrit par s_n étant supposé stationnaire, les coefficients a_i sont indépendants du temps et sont identifiés par la méthode d'autocorrélation qui consiste à résoudre le système d'équations linéaires suivant [?] :

$$\begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}, \tag{5.12}$$

avec $\rho_i = \mathrm{E}\,[s_n s_{n-i}]$.

La matrice d'autocorrélation $[[\rho_i]]$ d'ordre p est de Toeplitz (matrice symétrique dont les éléments de la diagonale sont identiques).

La résolution de ce type d'équations, appelées équations de Yule-Walker, est obtenue par l'algorithme récursif de Levinson [?]. Dans le cas d'un modèle AR d'ordre 2, le système

?? se réduit à un système de deux équations linéaires à deux inconnues. Les coefficients a_1 et a_2 sont aisément calculés :

$$a_1 = \frac{\rho_0\rho_1 - \rho_1\rho_2}{\rho_1^2 - \rho_0^2} \tag{5.13}$$

$$a_2 = \frac{\rho_0\rho_2 - \rho_1^2}{\rho_1^2 - \rho_0^2} \tag{5.14}$$

Par ailleurs, la modélisation AR de la séquence s_n peut aussi être vue comme étant le filtrage d'un bruit blanc ϵ_n (erreur de prédiction) par un filtre tout pôle. La sortie à l'instant n du filtre ainsi défini, ne serait que la prédiction $\hat{s}(n)$ de $s(n)$. Il vient donc que la transformée en z de la réponse impulsionnelle du filtre h est :

$$H(z) = \frac{1}{1 + \sum_{i=1}^{p} a_i z^{-i}}, \tag{5.15}$$

Dans la phase d'analyse, nous avons été confronté au choix de l'ordre (p) du filtre auto-régressif h et à la détermination de la valeur du coefficient (α) du filtre g. Dans les sections ?? et ?? nous expliquons comment ont été déterminés ces coefficients.

5.3.3 Détermination de l'ordre du filtre h

Dans notre étude des contractions musculaires (*cf.* section ??), nous avons vu que le modèle le plus répandu pour la modélisation des mouvements articulatoires des différentes parties du corps humain est, sans conteste, le modèle du second ordre. Pour vérifier cette hypothèse, nous avons effectué deux calculs ; le spectre du signal et le critère informationnel d'Akaïke.

5.3.3.1 Analyse spectrale

Le calcul du spectre d'un signal par la FFT (Fast Fourier Transform) permet d'avoir une première estimation du nombre de pôles existant dans le signal. La précision de la méthode FFT est suffisante pour l'analyse d'une large classe de signaux, notamment les signaux oscillants de structure simple. Le calcul de la FFT a été effectué sur l'ensemble des données observées correspondants aux dix phrases concaténées, comme suit :

$$S_n = \frac{1}{N}\sum_k s_k W_N^{kn}. \tag{5.16}$$

où $W_N = \exp -j2\pi/N$

La figure ?? montre le spectre des mouvements des quatre paramètres de langue. Chacun des quatre spectres montre un seul pic dans les basses fréquences qui correspond à une paire de pôles complexes conjugués. Ceci suggère que les mouvements de la langue peuvent être modélisés par des modèles du second ordre.

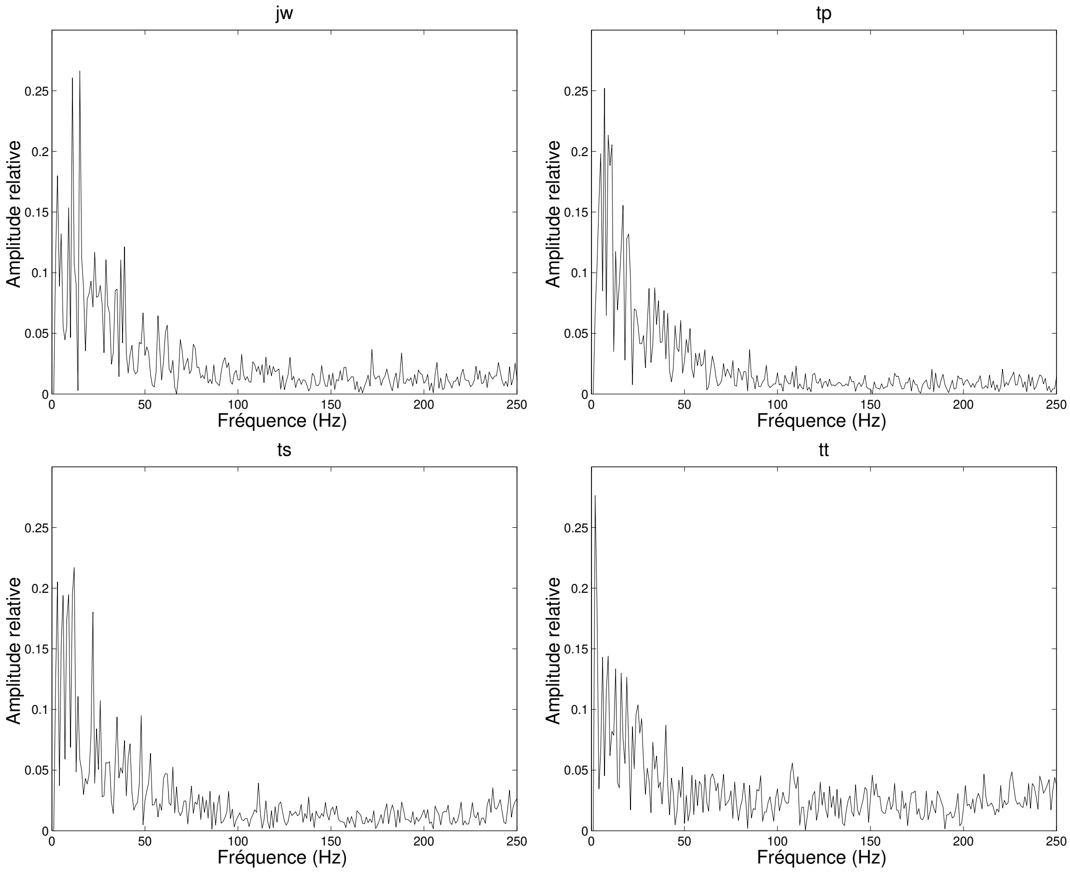


Figure 5.7 *La FFT des mouvements des différents paramètres.*

5.3.3.2 Critère informationnel d'Akaïke

Dans le cas d'un modèle AR où l'ordre p du filtre est à priori inconnu, Akaïke a proposé un critère pour déterminer la valeur de p : Akaïke Information Criterion [?][?]. Le critère $AIC(p)$ est défini par

$$AIC(p) \triangleq \log(\nabla_p^*/R_0) + 2p/kN, \tag{5.17}$$

où N, ∇_p^*, R_0 et k sont respectivement le nombre d'échantillons du signal, l'énergie de l'erreur de prédiction pour p pôles, la variance du signal et le coefficient de compensation par fenêtrage ($k = 0.4$ dans le cas de Hamming).

Le premier terme de l'équation **??** qui représente l'énergie de prédiction normalisée décroît en fonction de p. Tandis que le second terme croît proportionnellement à p. Il exprime le coût de calcul qui est d'autant plus élevé que la valeur de p est grande. La fonction $AIC(p)$ présente donc un minimum donné par p^* qui correspond à un rapport "précision/complexité" optimal. La figure **??** montre un exemple de calcul de la fonction $AIC(p)$ pour la phrase "*une réponse ambigüe*".

Dans cette figure, l'ordre optimal de prédiction est $p^* = 2$. En effet, pour $p > 2$ la pente de la courbe croît progressivement. Ce résultat est obtenu pour les quatre paramètres dans le cas des dix phrases. Ceci signifie, que du point de vue du rapport "précision/complexité de calcul" le modèle linéaire du second ordre est recommandé.

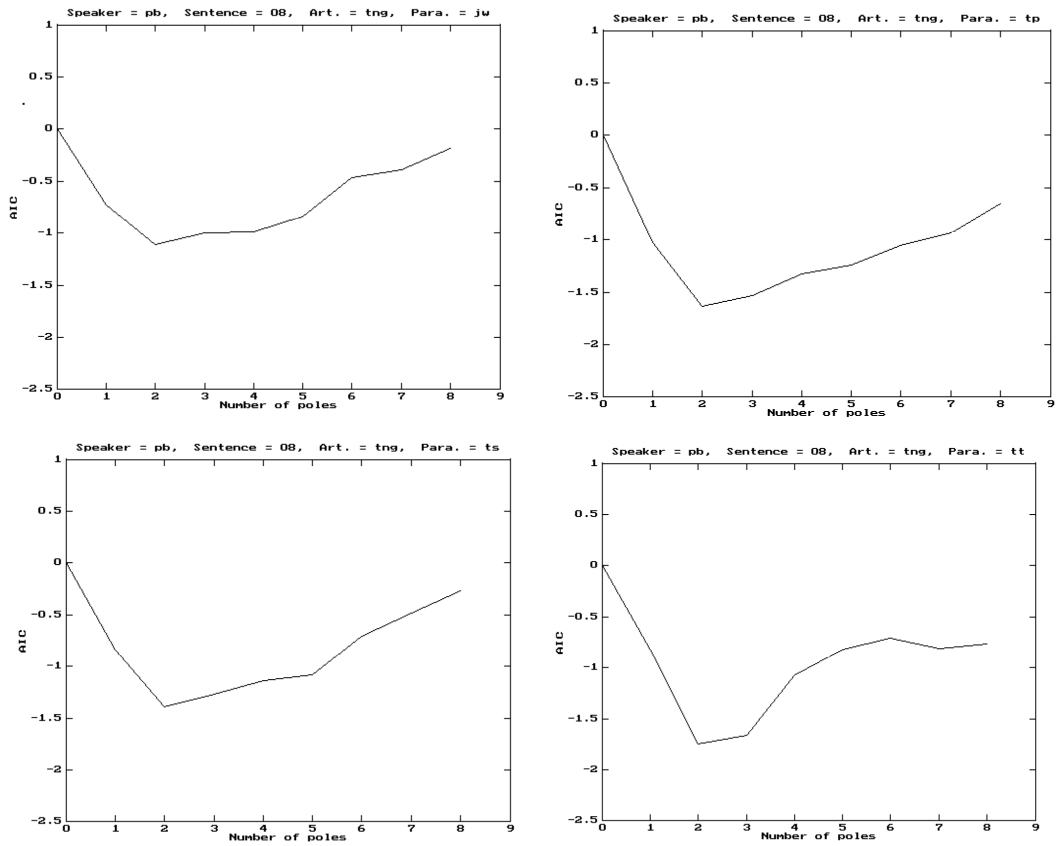


Figure 5.8 *Courbes du critère informationnel d'Akaïke (AIC) en fonction de l'ordre de prédiction p.*

5.3.4 Détermination du coefficient du filtre g

Le coefficient α du filtre de pré-accentuation a été déterminé empiriquement de façon que l'énergie de l'erreur entre le signal synthétisé, $\hat{s}(n)$, et le signal utile, $s(n)$, soit faible. Cette erreur est obtenue pour α égal à 0.5 et ce pour les quatre paramètres **jw**, **tp**, **ts** et **tt**. Notons que pour les valeurs de α comprises entre 0.7 et 0.4 la synthèse des mouvements articulatoires reste de bonne qualité. Cependant, en dehors de cet intervalle et pour le même type du signal d'excitation, la synthèse des mouvements est mauvaise.

La valeur du coefficient α détermine, en partie, le temps de réponse du filtre g et par conséquent agit directement sur la durée de la réponse impulsionnelle du modèle global. Par ailleurs, les variations moyennes des mouvements articulatoires correspondent approximativement à la durée de la syllabe. La synthèse des mouvements est donc d'autant meilleure que la durée effective D de la réponse impulsionnelle du modèle de synthèse $h * g$ est comparable à celle de la syllabe. Cette valeur est obtenue pour α égal à 0.5.

Dans le cas où le coefficient α est supérieur à 0.7, la durée D dépasse celle des transitions des mouvements. Ceci implique une complexité dans la synthèse des patrons des mouvements ; la courbe d'erreur ne décroît pas d'une manière monotone. Lorsque α est inférieur à 0.4, la durée D est petite. La synthèse nécessite un plus grand nombre d'impulsions pour atteindre les même performances que celles obtenues pour α égal à 0.5.

5.4 Invariance du modèle des paramètres articulatoires

Une fois les paramètres des filtres g et h déterminés, nous nous sommes proposés de vérifier l'hypothèse d'invariance du filtre AR. Pour chaque phrase l de notre corpus, les valeurs des coefficients $\{a_i, i = 1, 2\}$ du filtre auto-régressif sont identifiées par les équations **??** et **??**. Les coefficients sont, ensuite, convertis en parties réelle $\Re z_{jl}$ et imaginaire $\Im z_{jl}$, des pôles correspondants. L'indice j désigne le type de paramètre traité (**jw**, **tp**, **ts** ou **tt**). De plus, les fréquences naturelles F_{jl}, les bandes passantes B_{jl} et les coefficients de qualités Q_j sont calculés pour chaque phrase :

$$F_{jl} = (f_e/2\pi)\tan^{-1}(\Im z_{jl}/\Re z_{jl}), \tag{5.18a}$$

$$B_{jl} = -(f_e/\pi)\ln|z_{jl}|, \tag{5.18b}$$

$$Q_j = 1/l\sum_l (F_{jl}/B_{jl}). \tag{5.18c}$$

où f_e est la fréquence d'échantillonnage égale à $50Hz$.

Cette analyse a d'abord été appliquée aux données débruitées par la méthode de filtrage, ensuite aux données obtenues par la projection du signal observé dans l'espace des composantes principales.

5.4.1 Cas des données débruitées par la méthode de filtrage

Les dix paires de pôles, calculés à partir des signaux débruités par la méthode de filtrage, sont reportés dans le cercle unité. Sur la figure ?? apparaissent deux nuages de points dispersés de part et d'autre de l'axe réel positif, pour chaque paramètre j. Le tableau ?? montre des valeurs élevées des écarts types σ_j pour les paramètres F $(0.9 \leq \sigma_j \leq 4.3Hz)$ et B $(1.4 \leq \sigma_j \leq 3.3Hz)$ et ce pour les deux locutrices PB et DF.

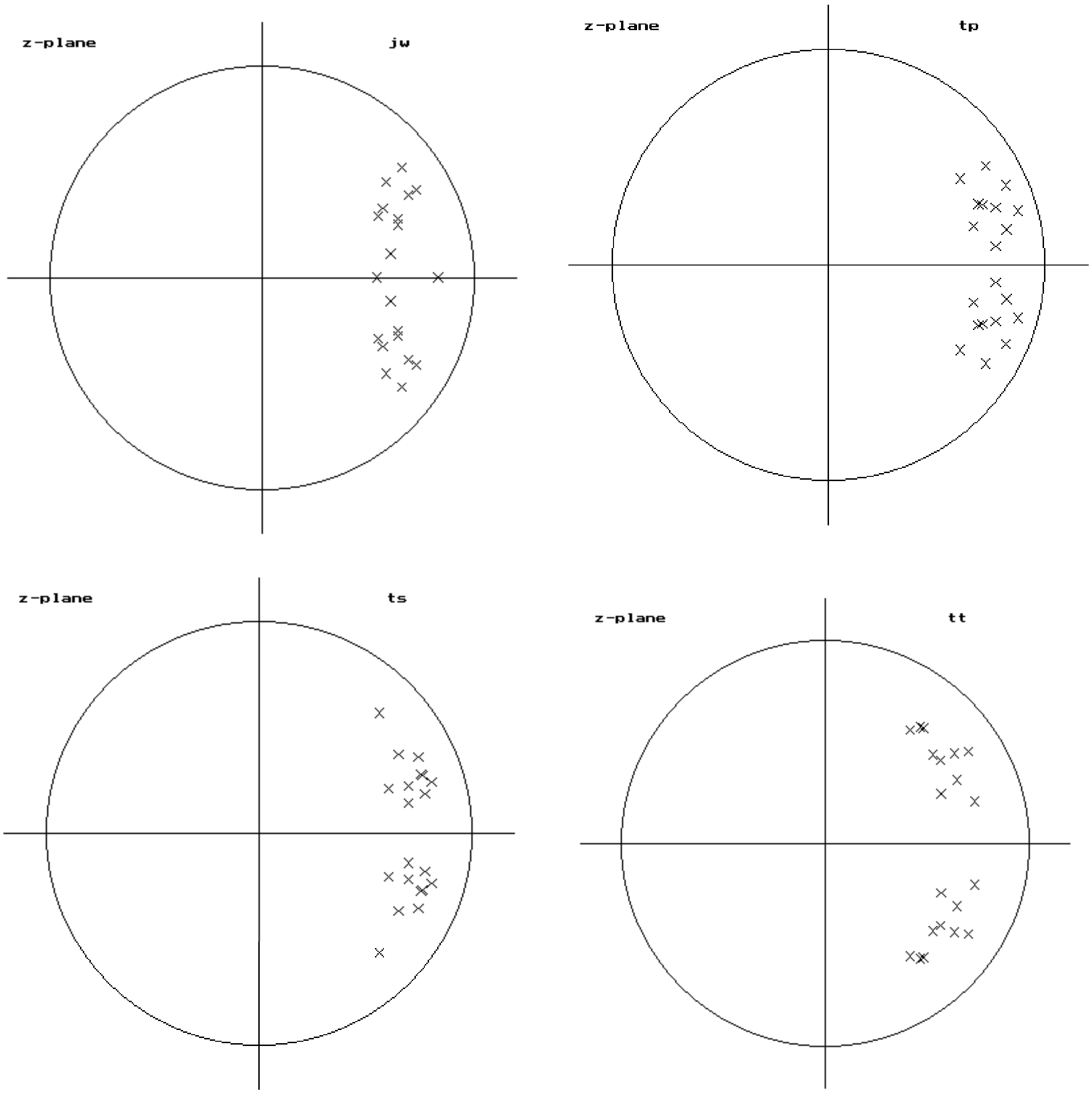


Figure 5.9 *Position des paires de pôles conjugués dans le cas des données débruitées par la méthode de filtrage.*

De ce résultat découle une question fondamentale sur la validité de l'hypothèse de stationnarité des processus modélisés, que nous avons posée dans les termes suivants : Quelle

	Locutrice PB			*Locutrice DF*		
articulateurs	$F_{(Hz)}$	$B_{(Hz)}$	Q	$F_{(Hz)}$	$B_{(Hz)}$	Q
jw	3.5	5.0	0.8	3.3	3.9	1.0
σ_{jw}	2.6	3.3		0.9	1.7	
tp	2.8	3.4	1.0	2.8	4.4	0.7
σ_{tp}	1.4	2.0		3.0	1.4	
ts	3.0	4.1	0.8	4.1	5.0	1.0
σ_{ts}	1.8	1.6		4.3	1.9	
tt	5.0	5.2	1.0	2.6	4.7	0.6
σ_{tt}	2.7	1.7		4.0	1.9	

Tableau 5.2 *Valeurs moyennes et écarts types des paramètres F_{jl}, B_{jl} et Q_j. Cas du débruitage par filtrage.*

est l'origine de cette dispersion du nuage de pôles ? Celle ci est due soit à un mauvais débruitage qui altère la position des pôles, soit à l'absence de caractéristiques invariantes dans les paramètres articulatoires de la langue. Dans ce dernier cas, le choix de filtre invariant est inadapté à la modélisation de ce type de signaux. Pour répondre à cette question, la première étape était d'appliquer l'analyse LPC aux données obtenues par la deuxième méthode d'extraction du bruit.

5.4.2 Cas des données lissées par projection du signal dans l'espace des composantes principales

Les pôles z_{jl}, les fréquences naturelles F_{jl} et les bandes passantes B_{jl} sont recalculés avec les données débruitées par la méthode de la projection du signal dans l'espace des composantes principales. Les résultats obtenus sont présentés dans la figure **??** et le tableau **??**.

On observe une compacité des deux nuages de points sur le cercle unité qui est traduite par une nette diminution des écarts types σ_j par rapport aux données débruitées par la méthode de filtrage. Ces derniers sont inférieurs à 0.9 pour les trois articulateurs **jw**, **tp** et **ts**. Les modèles de ces trois paramètres présentent donc une certaine invariance par rapport aux différentes phrases. Ces résultats sont favorables à l'hypothèse de l'invariance temporelle de nos modèles articulatoires. La variance σ_{tt} correspondant à l'apex est relativement élevée. Le mouvement de cet articulateur est complètement noyé dans le bruit de mesure. Il correspond au dernier paramètre du modèle de Maeda qui ne représente que 9.2% de la variance du profil de la langue observée sur les données radiofilms. Il est donc difficile de séparer le signal utile du bruit.

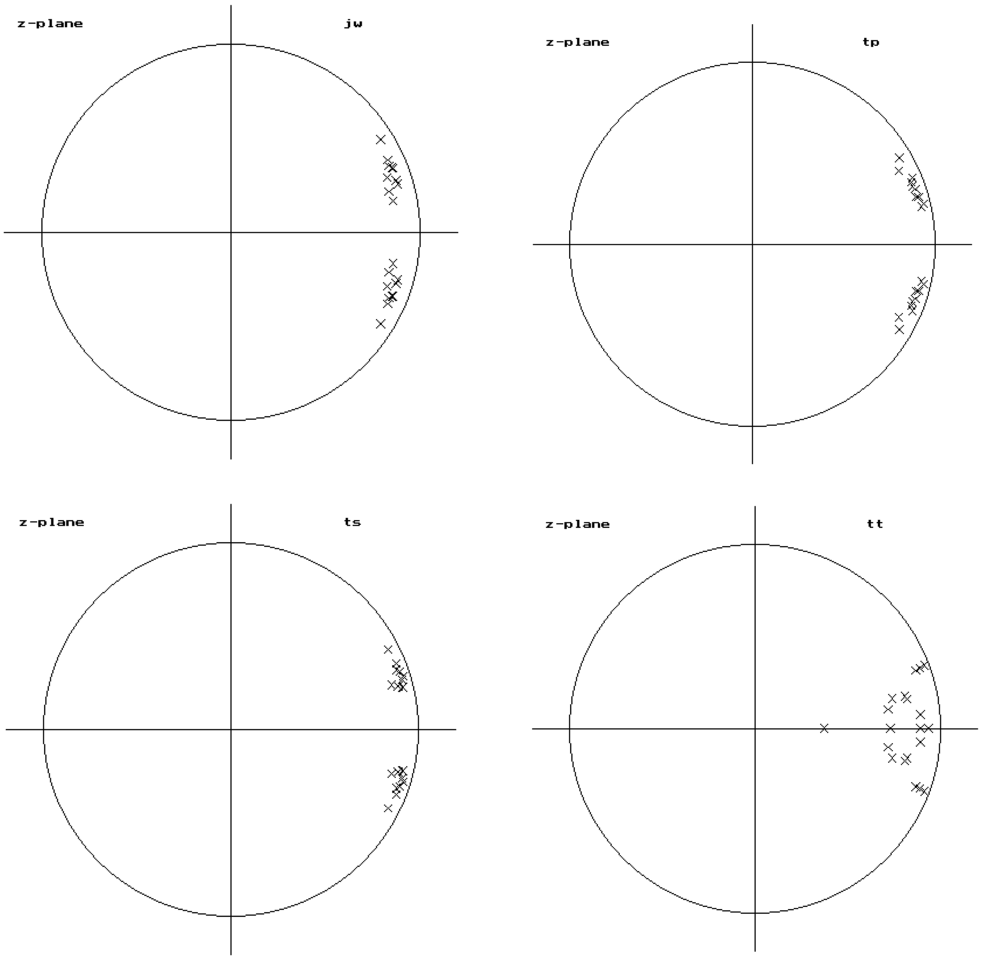


Figure 5.10 *Position des paires de pôles conjugués dans le cas des données débruitées par la méthode de projection du signal dans l'ECP.*

	Locutrice PB			*Locutrice DF*		
articulateurs	$F_{(Hz)}$	$B_{(Hz)}$	Q	$F_{(Hz)}$	$B_{(Hz)}$	Q
jw	2.8	1.7	1.7	3.0	1.0	2.9
σ_{jw}	0.6	0.2		0.4	0.1	
tp	2.8	1.0	2.7	2.7	1.1	2.5
σ_{tp}	0.7	0.1		0.6	0.1	
ts	2.5	1.0	2.4	2.6	2.3	1.1
σ_{ts}	0.3	0.1		1.1	0.3	
tt	1.5	2.3	0.7	2.6	1.2	2.1
σ_{tt}	1.2	2.6		1.1	0.2	

Tableau 5.3 *Valeurs moyennes et écarts types des paramètres F_{jl}, B_{jl} et Q_j. Cas du débruitage par ECP.*

Il est intéressant de remarquer que la locutrice DF présente des coefficients $Q_{\mathbf{jw}}$ et $Q_{\mathbf{tt}}$ plus élevés que la locutrice PB. Ceci signifie que le mouvement de la mâchoire est moins amorti et donc plus actif pour la première locutrice. En effet, durant l'expérience on a constaté que DF articule plus pour produire les séquences de phonèmes constituant les phrases. Les fréquences naturelles $F_{\mathbf{jw}}$, $F_{\mathbf{tp}}$ et $F_{\mathbf{ts}}$ ont des valeurs comprises entre $2.5Hz$ et $3.0Hz$ et demeurent pratiquement inchangées d'une locutrice à l'autre. Néanmoins, on observe une différence inter-locutrice relativement élevée pour le paramètre **tt**. La fréquence $F_{\mathbf{tt}}$ de la locutrice PB est particulièrement faible. Elle est égale à $1.5Hz$. Visiblement, le mouvement propre (haut/bas) de l'articulateur de l'apex dépend du mouvement de la mâchoire. Lorsque la mâchoire est peu mobile, l'espace de déplacement de l'apex est réduit. De faibles mouvements de **tt** suffisent pour réaliser les constrictions invoquant cet articulateur. Ceci implique une économie de mouvement de l'articulateur apical et donc une fréquence naturelle correspondante plus faible pour PB.

Par ailleurs, pour le débruitage des logatomes qui sont de durée relativement courte, nous avons procédé à une concaténation des données en effectuant une interpolation linéaire sur un segment que nous avons choisi égal à 5 échantillons, soit $100ms$. Cette durée correspond à la moitié de la durée moyenne de la syllabe française. Les fréquences naturelles, les bandes passantes et les coefficients de qualité, présentés dans le tableau **??**, sont comparables a ceux obtenus pour les phrases. Néanmoins, les mouvements sont plus amortis et plus lents notamment pour les paramètres **jw**, **tp** et **ts**. Ceci se traduit par des coefficients de qualité et des fréquences naturelles plus faibles. Ces différences sont en rapport avec la vitesse d'élocution qui est plus élevée pour la production des phrases (constituées de 5 à 6 syllabes) que celle des séquences bi-syllabiques isolées.

articulateur	$F_{(Hz)}$	$B_{(Hz)}$	Q
jw	1.8	2.1	0.9
tp	2.0	1.4	1.4
ts	1.4	1.4	1.0
tt	2.3	1.4	1.7

Tableau 5.4 Valeurs de F_j, B_j et Q_j obtenues pour les logatomes.

L'analyse de ces résultats suggère que pour un style de parole donné (effort articulatoire, vitesse d'élocution), les paramètres articulatoires de la langue possèdent des caractéristiques invariantes dans le temps.

5.5 Calcul des réponses impulsionnelles

Les quatre filtres de synthèse correspondant aux quatre paramètres articulatoires ont été construits à partir des centres de gravité des nuages de points respectifs, présentés dans la figure **??**. L'ensemble des pôles calculés est situé à l'intérieur du cercle unité aussi les fonctions de transfert qui leurs sont associées correspondent à des systèmes stables. Les réponses impulsionnelles des systèmes de synthèse incluant le filtre de dé-accentuation et le filtre AR d'ordre 2 sont montrées dans la figure **??**.

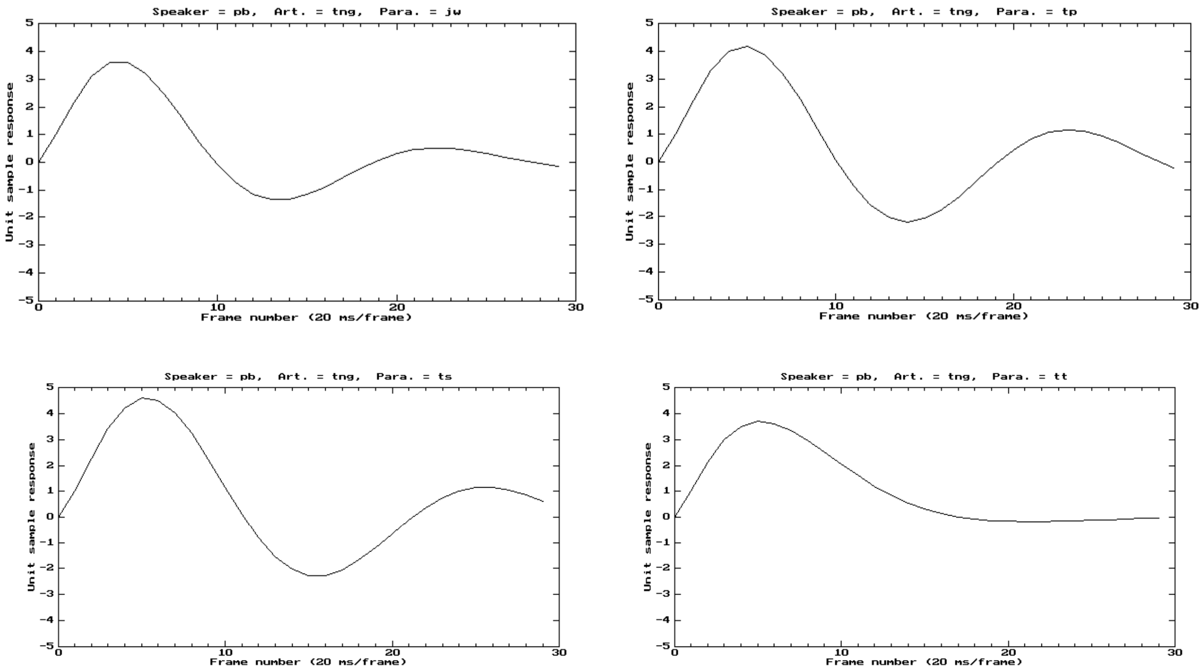


Figure 5.11 *Réponses impulsionnelles.*

Les quatre courbes montrent une oscillation plus ou moins amortie. Le signal de commande du système linéaire $h * g$ étant une séquence d'impulsions, ces réponses impulsionnelles sont considérées comme le geste élémentaire qui, par superposition, donne naissance aux mouvements articulatoires. Nous appelons la durée effective de la réponse impulsionnelle l'intervalle de temps entre l'instant d'activation de l'articulateur vers une position cible et l'instant de retour à l'état neutre. Il est important de remarquer que la durée effective ainsi définie de chacune de ces réponses impulsionnelles (**jw**, **tp**, **ts** et **tt**) est comprise entre 180 et 200ms. L'effet d'une impulsion de commande s'étend donc sur une longueur égale à la longueur moyenne de la syllabe française. En termes articulatoires, lorsque le système est excité par une impulsion, le mouvement effectué par l'articulateur atteint sa cible et revient à son état de repos après une faible oscillation. Par exemple, pour l'articulateur **tp**, une impulsion de commande produit un mouvement antérieur (ou postérieur) suivi d'un mouvement postérieur (ou antérieur) avant de retrouver l'état au repos.

5.6 Interprétation physique du modèle

Dans notre étude nous avons décrit le modèle de synthèse à l'aide de deux filtres. Dans la chaîne de production de la parole chez l'homme, on distingue deux systèmes : un système physiologique et un système dynamique (mécanique). Le problème est d'établir une analogie entre notre modèle et la chaîne de production.

Les nombreux travaux menés ces dernières années en vue d'examiner et de synthétiser les mouvements des différentes parties du corps tels que le bras, la main, le genou, les lèvres... etc, utilisent le modèle linéaire du second ordre. En effet, en dépit de la complexité des systèmes musculaires, la trajectoire du mouvement articulatoire est souvent reproduite à l'aide d'une équation différentielle caractérisée par une masse, une résistance mécanique, un coefficient de raideur et une force d'excitation ainsi que nous l'avons expliqué dans la section **??**. Dans ce cas, le filtre de dé-accentuation g correspond au système physiologique. Ce filtre, fortement amorti, consiste à transformer la séquence d'impulsions en une forme d'onde exponentielle qui représente la force d'excitation du système dynamique.

Toutefois, pour modéliser le mouvement de la main durant l'écriture, Yasuhara (1983) [?] a identifié deux systèmes g_m et h_m. Son analyse l'a conduit à inverser les rôles des filtres. Il a associé le système oscillant h_m (composé par la paire de pôles complexes conjugués) au système physiologique, et le système fortement amorti g_m au mouvement dynamique de la main. Son choix est basé sur le calcul de la valeur du coefficient de qualité Q ; lorsque les pôles sont réels, ce coefficient est nul. Par conséquent, la réponse impulsionnelle engendrée est très amortie. Yasuhara associe l'amortissement aux caractéristiques mécaniques des systèmes.

Cette interprétation nous semble plus réaliste. Mais notre constation est basée sur le calcul de la fréquence naturelle. En effet, la fréquence des pôles du système du second ordre identifié est environ $3Hz$ (*tab.* **??** et **??**). Or, la fréquence de résonance mécanique des tissus musculaires humains est estimée supérieure à $30Hz$. Par exemple, la fréquence d'oscillation des lèvres durant le relâchement des consonnes bilabiales est approximativement égale à $33Hz$ (Fujimura 1961 [?]). La résonance mécanique des tissus des joues, mesurée dans les cas tendues et relâchées, est comprise entre 30 et $60Hz$ (Ishizaka et *al.* 1975 [?]). Ces expériences suggèrent que la fréquence naturelle des articulateurs est supérieure à $3Hz$. Par conséquent, le filtre h du second ordre ne peut correspondre au système mécanique. Il est associé au système physiologique, et est considéré comme un générateur de force.

Par ailleurs, pour détecter une fréquence de l'ordre de la fréquence de résonance mécanique, soit $30Hz$, il faut échantillonner le signal avec une fréquence au moins égale à $60Hz$ de manière à vérifier le théorème de Shannon (fréquence d'échantillonnage supérieure ou égale à deux fois la fréquence de coupure du signal). Or, dans le cas des mouvement des articulateurs de la langue les signaux observés sont échantillonnés à $50Hz$. La fréquence naturelle du système musculaire des articulateurs de la langue n'est donc pas modélisée. De plus, il est facile de montrer que la partie de la courbe de la fonction de transfert, qui correspond aux fréquences inférieures à la fréquence de résonance d'un système du second ordre, est donnée essentiellement par la raideur du système. Tandis que la partie de la courbe dont les fréquences sont supérieures à la fréquence de résonance est engendrée par l'effet de la masse. Si on coupe le spectre des mouvements à une fréquence inférieure à la fréquence de résonance, le mouvement musculaire se comportera comme un filtre d'ordre 1. Ceci suggère que, sur le support spectral observé qui est inférieure à $25Hz$, les articulateurs de la langue peuvent être considérés essentiellement comme des systèmes visco-élastiques où l'effet de la masse est négligé. Cette approche est cohérente avec la nature du modèle des contractions musculaires décrit par Hill (*cf.* section **??**).

Remarque

Les caractéristiques du filtre h et le coefficient α du filtre g du mouvement de la main sont comparables à ceux des articulateurs de la langue, en particulier, pour la locutrice DF, ainsi que le montre le tableau **??**. Les valeurs du coefficient α obtenues pour les articulateurs de la langue sont comprises entre 0.5 et 0.7. Pour ces valeurs l'erreur de synthèse demeure inchangée.

articulateur	$F_{(Hz)}$	$B_{(Hz)}$	Q	α
main	3.4	1.2	2.7	0.7
jw	3.0	1.0	2.9	[0.5 0.7]
tp	2.7	1.1	2.5	[0.5 0.7]
ts	2.6	2.3	1.1	[0.5 0.7]
tt	2.6	1.2	2.1	[0.5 0.7]

Tableau 5.5 *Comparaison des caractéristiques des filtres h et g obtenues pour le mouvement de la main durant l'écriture et celui de la langue durant la parole (locutrice DF).*

5.7 Conclusion

Dans ce chapitre nous avons modélisé les mouvements des articulateurs de la langue par des modèles à caractéristiques invariantes. Deux filtres constituent chaque modèle. Un filtre visco-élastique d'ordre 1 et un filtre du second ordre. La fréquence naturelle des différents articulateurs dépend plus du style de parole produite que du locuteur. Le problème, maintenant, est de déterminer le modèle d'excitation et d'observer comment s'effectue la commande des articulateurs lors de la production des phrases.

Chapitre 6

Détermination de l'excitation

Dans cette section nous présentons le modèle multi-impulsionnel utilisé pour l'excitation du filtre de synthèse LPC (Atal et Remde 1982 [?]). Dans ce modèle, le signal d'excitation consiste en une séquence d'impulsions dont les positions et les amplitudes sont à déterminer. Il est clair que pour un nombre d'impulsions m assez grand, en particulier lorsque $m = N$ (N égal au nombre d'échantillons du signal), le signal synthétisé est identiquement égal au signal original. Dans le domaine de la commande articulatoire, cette méthode nous a paru commode pour l'étude du rapport qui existe entre la commande et les unités phonétiques d'une part, et celle de la stratégie de commande entre les différents articulateurs de la langue d'autre part.

6.1 Position et amplitude des impulsions

La procédure d'analyse par synthèse pour déterminer les positions et amplitudes des impulsions de la séquence d'excitation est illustrée par le diagramme bloc de la figure **??**.

Le synthétiseur LPC produit les échantillons $\hat{s}(n)$ du mouvement synthétisé en réponse à l'excitation $v(n)$ de telle sorte que :

$$\hat{s}(n) = \sum_{i=1}^{N} f_{n-i} v_i, \qquad 1 \leq n \leq N. \tag{6.1}$$

où $f = g * h$. Dans le cas où le signal d'excitation est une séquence d'impulsions, c'est-à-dire,

$$v(n) = \sum_{k=1}^{m} \gamma_k c_n^{(k)}. \tag{6.2}$$

où $\gamma_k c_n^{(k)}$ représente une impulsion d'excitation à l'instant k et d'amplitude γ_k, l'équation **??** peut s'écrire sous forme matricielle comme suit :

$$\hat{\mathbf{s}}_k = \gamma_k \mathbf{F} \mathbf{c}_k, \qquad 1 \leq k \leq m. \tag{6.3}$$

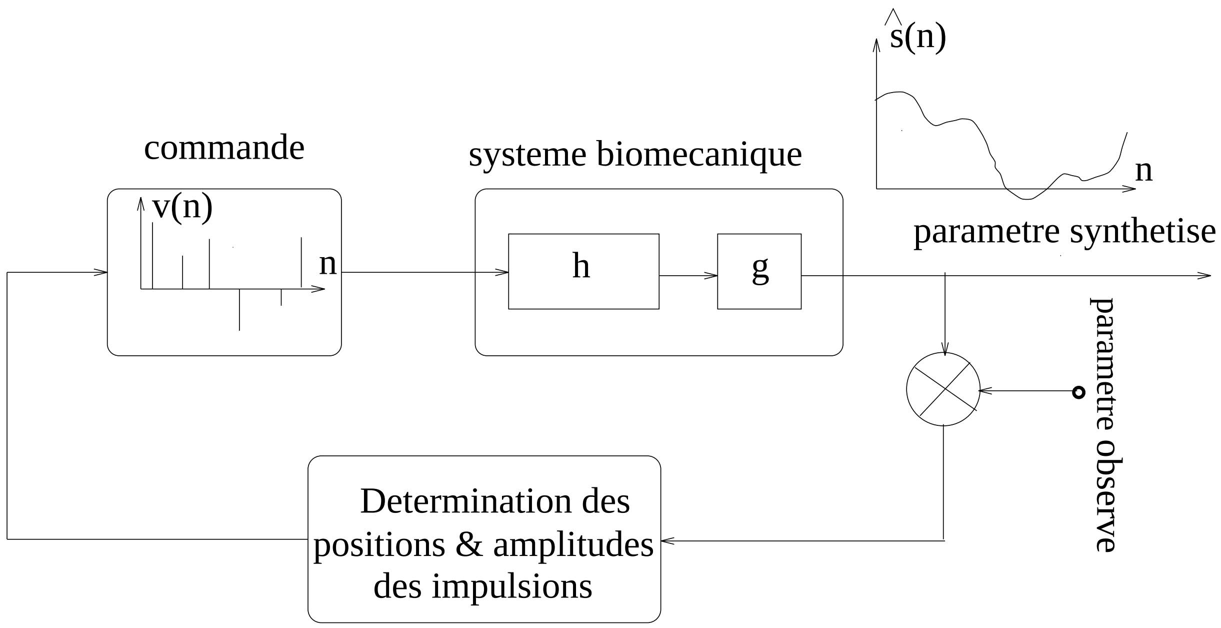


Figure 6.1 *Diagramme bloc de la procédure de l'analyse par synthèse pour la détermination des positions et amplitudes des impulsions de l'entrée d'excitation.*

où $\mathbf{F}$ est une matrice $N \times N$ dont l'élément de la n^{eme} ligne et la i^{eme} colonne est donné par f_{n-i}.

Les échantillons $\hat{s}(n)$ sont ensuite comparés aux échantillons observés $s(n)$ pour produire le signal d'erreur $\mathbf{E}_k$ qui s'écrit

$$\mathbf{E}_k = \sum_{n=1}^{N} (|s(n) - \hat{s}(n)^{(k)}|^2) \tag{6.4}$$

ou encore, en remplaçant $\hat{s}(n)^{(k)}$ par l'expression **??**,

$$\mathbf{E}_k = (\|\mathbf{s} - \gamma_k \mathbf{F} \mathbf{c}_k\|^2), \tag{6.5}$$

où $\| \quad \|^2$ est la somme au carré des composantes des vecteurs.

Les positions et les amplitudes des impulsions sont calculées de façon à minimiser cette erreur. En effet, la valeur optimale de l'amplitude γ_k qui minimise $\mathbf{E}_k$ peut être obtenue aisément en calculant $\partial \mathbf{E}_k / \partial \gamma_k = 0$ qui conduit au résultat suivant :

$$\gamma_k = \frac{\sum_{i=1}^{N} s_i f_{i-k}}{\sum_{i=1}^{N} f_i^2} \tag{6.6}$$

et

$$\mathbf{E}_k = \sum_{i=1}^{N} s_i^2 - \frac{\left[\sum_{i=1}^{N} x_i f_{i-k}\right]^2}{\sum_{i=1}^{N} f_i^2} \tag{6.7}$$

La position optimale c_k, qui est associée à l'indice k, minimise l'erreur $\mathbf{E}_k$. Cette position peut être déterminée en calculant le maximum du deuxième terme du membre de droite de l'équation ci-dessus. Le numérateur de ce terme n'est autre que l'expression de l'intercorrélation entre la réponse impulsionnelle f et le signal original $\mathbf{s}$ qui ne dépend que de la position c_k. Ainsi, la position optimale c_k est donnée directement par le maximum de la fonction d'intercorrélation $\sum_{i=1}^{N} s_i f_{i-k}$.

L'amplitude des impulsions est ensuite calculée par l'équation **??** à l'itération k (Atal 1986 [**?**]). De cette manière, les positions et amplitudes des impulsions de l'excitation sont obtenues séquentiellement. Les amplitudes γ_k sont recalculées à chaque nouvelle itération.

6.2 Calcul de l'erreur de synthèse

Dans cette section, nous allons observer comment évolue l'énergie de l'erreur entre les deux mouvements observé et synthétisé lorsque le nombre d'impulsions m augmente. Cette erreur est calculée pour chaque phrase. Elle est normalisée par l'énergie de l'erreur totale obtenue pour $m = 0$ de la façon suivante :

$$\varepsilon(m) \triangleq \left[\sum_{n=1}^{N} \mathbf{E}_m(n)^2\right] / \left[\sum_{n=1}^{N} s(n)^2\right]. \tag{6.8}$$

Ainsi que le montre l'exemple des figures **??** et **??**, l'erreur de synthèse décroît d'une manière monotone. Chacune des courbes diminue rapidement jusqu'au nombre d'impulsions égal au nombre de syllabes dans la phrase, en l'occurrence $m = 4$. Si bien que la pente de cette portion de la courbe coupe l'axe des abscisses en un point situé au voisinage du point m égal à 4. Pour cette valeur, l'énergie de l'erreur est déjà inférieure à 30%.

Au delà de cette valeur et jusqu'à m égal au nombre de phonèmes ($m = 9$), la pente de l'erreur devient relativement faible. Dans cette partie du graphe l'énergie de l'erreur est comprise entre 10% et 30% de l'énergie du signal observé.

Une saturation est observée lorsque m est supérieur au nombre de phonèmes. Dans ce cas, les amplitudes des dernières impulsions déterminées deviennent extrêmement faibles et leur effet est peu significatif. Remarquons que ces résultats sont obtenus dans 90% des cas. La synthèse peut requérir un nombre d'impulsions plus élevé pour atteindre les mêmes performances, ainsi que le montre la figure **??**-b).

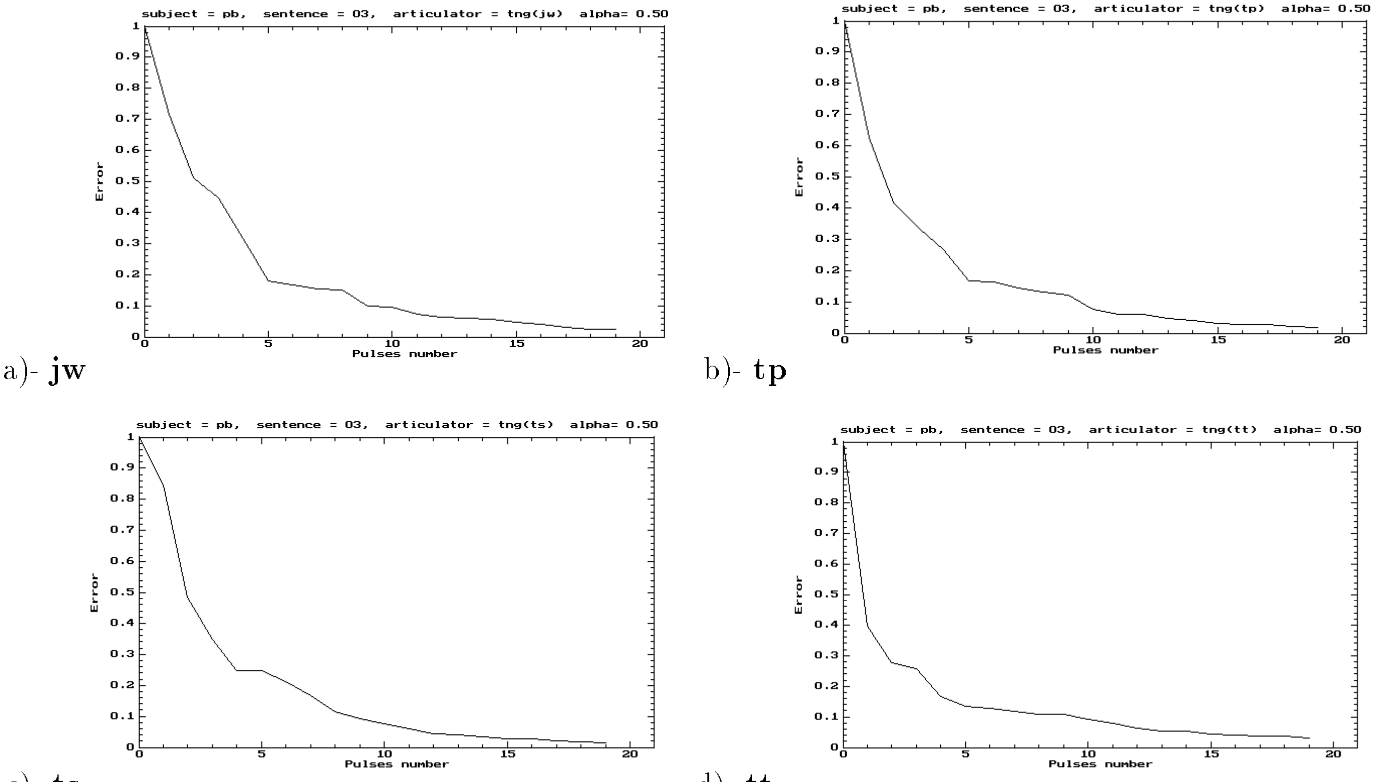


Figure 6.2 *Énergie de l'erreur de synthèse pour la phrase* 03 *et la locutrice PB.*

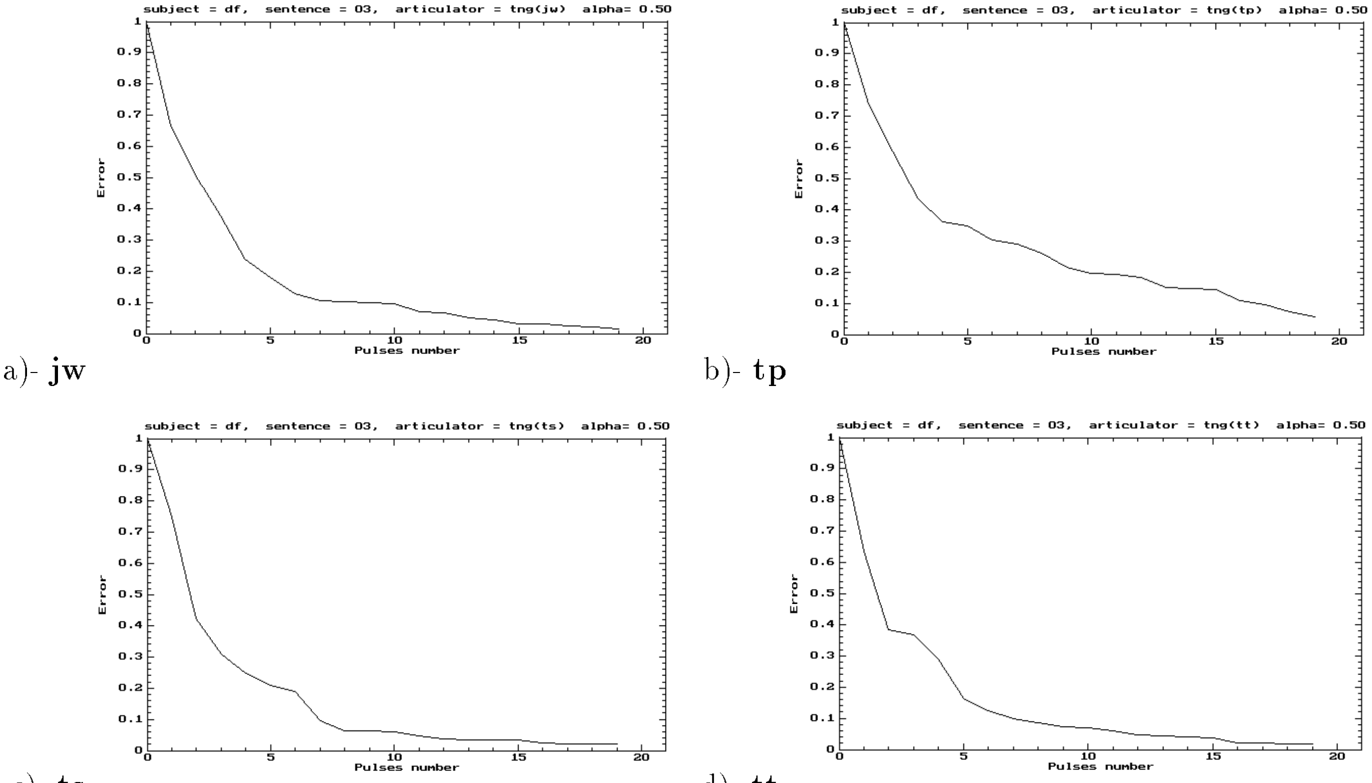


Figure 6.3 *Énergie de l'erreur de synthèse pour la phrase* 03 *et la locutrice DF.*

6.3 Description du mouvement synthétisé

Afin d'élucider la reconstitution des mouvements observés par la méthode multi-impulsionnelle, nous avons présenté dans les figures **??** et **??** le mouvement synthétisé (en pointillé) lorsque le nombre d'impulsions m est égal à 1, que nous avons comparé au mouvement observé (en trait plein).

Les mouvements des paramètres articulatoires de la langue évoluent entre deux positions extrêmes qui correspondent à l'activité maximale de chaque articulateur. Dans les figures présentées, les valeurs positives de la position **jw** correspondent à la fermeture de la mâchoire et les valeurs négatives à l'aperture. La position de la langue **tp** est postérieure pour des valeurs positives et antérieure pour des valeurs négatives. La forme de la langue **ts** est arquée lorsque les valeurs sont positives et plate lorsque les valeurs sont négatives. Enfin, l'apex **tt** est soit haut soit bas respectivement positif ou négatif. Le signal d'excitation de chaque paramètre est représenté en dessous des mouvements correspondants. Dans tous les cas de figures, une impulsion de commande génère un mouvement qui correspond à la réponse impulsionnelle du modèle identifié $f * g$. Après une légère oscillation, ce mouvement s'annule. On dit que l'articulateur est à sa position moyenne. La position et l'amplitude de l'impulsion sont déterminées de manière à réaliser un ajustement optimal du mouvement généré et d'une position cible du signal observé.

L'effet d'une impulsion s'étend sur un segment de durée égale à $200ms$, soit la durée de la première oscillation de la réponse impulsionnelle. Lorsque le mouvement observé évolue dans le sens de l'oscillation secondaire, le signal généré couvre deux syllabes (*fig.* **??**-b). Dans ce cas, il suffit d'une impulsion de faible amplitude pour produire le mouvement du segment qui suit l'oscillation principale (*fig.* **??**-b). Un deuxième cas peut se présenter (*fig.* **??**-b) : lorsque le mouvement évolue dans le sens opposé à celui de l'oscillation secondaire, il faut une impulsion d'amplitude élevée pour changer la trajectoire de l'articulateur (*fig.***??**-b). Le mouvement observé est reconstitué en superposant les réponses impulsionnelles produites par la séquence des commandes déterminées.

6.4 Organisation de la commande

Dans les sections suivantes, nous allons discuter le rapport entre la séquence des impulsions et les unités phonétiques, telles que la syllabe, le phonème ...etc. Notre but est de savoir si la synthèse des mouvements des articulateurs présente des caractéristiques phonétiques en fonction du nombre d'impulsions.

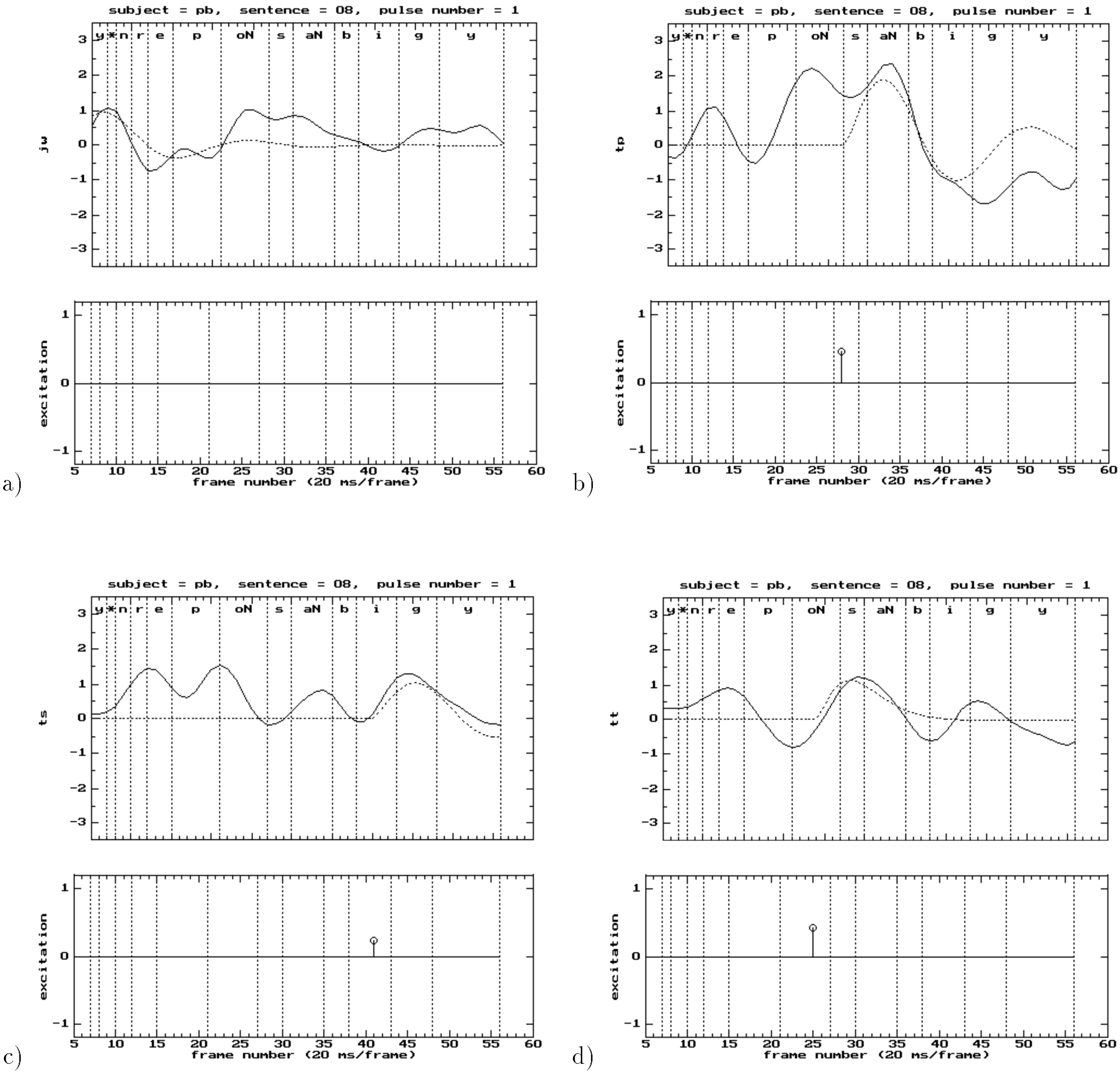


Figure 6.4 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase 08, nombre d'impulsions :* $m = 1$.

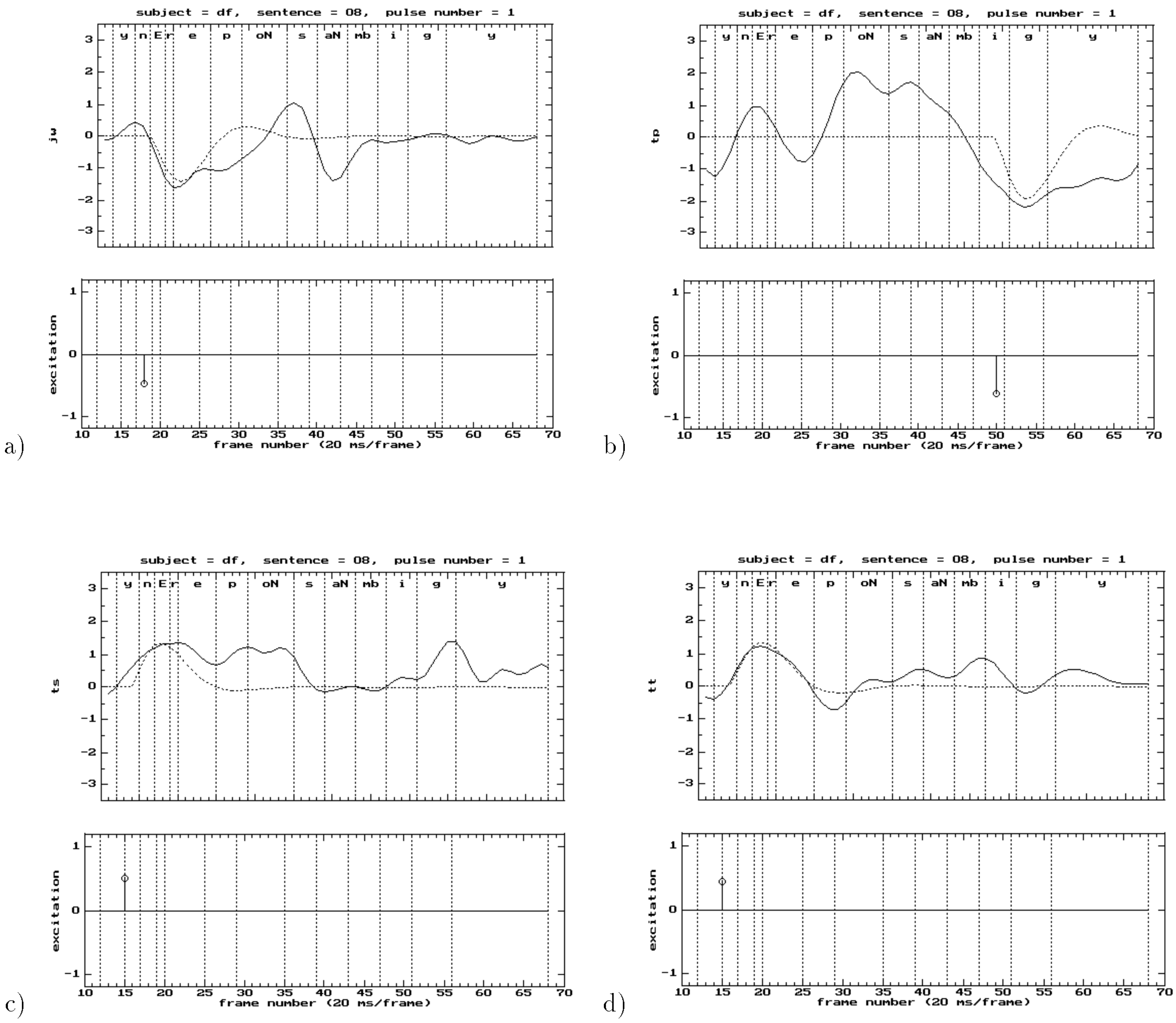


Figure 6.5 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase 08, nombre d'impulsions :* $m = 1$.

6.4.1 Organisation des impulsions par rapport à la syllabe

La durée effective de la réponse impulsionnelle étant comparable à celle de la durée moyenne de la syllabe française, nous nous sommes proposé d'observer l'évolution du signal synthétisé lorsque le nombre d'impulsions est égal au nombre de syllabes m_{syl}. Ce choix est également motivé par les résultats obtenus par le calcul de l'énergie de l'erreur de synthèse (*cf.* section **??**). La première question à laquelle nous allons essayer de répondre est la suivante : à partir du nombre réduit m_{syl} de commandes peut-on entrevoir des régularités de contrôle intra et inter-locutrices ?

Les mouvements synthétisés obtenus par la séquence S_{syl} composée de m_{syl} impulsions montrent des différences avec les mouvements observés. Ce nombre d'impulsions est insuffisant pour reproduire fidèlement le mouvement observé. Les différences sont indifféremment relevées sur les segments vocaliques et consonantiques.

La génération des segments de parole fait appel à la coordination des mouvements de plusieurs articulateurs, ainsi que nous l'avons vu dans le chapitre **??**. Cependant, tous les articulateurs ne montrent pas la même activité pour la production d'un son donné. À chaque son sont associés des articulateurs primaires et des articulateurs secondaires. Les figures **??** à **??** montrent que lorsque le paramètre est primaire pour la réalisation du trait distinctif d'un segment, les deux locutrices génèrent des impulsions d'amplitude élevée, marquées par des croix sur les figures. Ainsi sur le segment /i/, une impulsion est systématiquement générée pour l'articulateur **tp** (*fig.* **??**-b, **??**-b, **??**-b, **??**-b, **??**-b, **??**-b). De la même manière, on relève une impulsion sur les segments de la voyelle /u/ correspondants à **ts** (*fig.* **??**-c, **??**-c, **??**-c, **??**-c). La locutrice DF nécessite un nombre d'impulsions supérieur à m_{syl} pour générer l'impulsion nécessaire pour produire le ballonnement de la forme de la langue pour la voyelle /u/ **??**-c). Le patron du mouvement de **ts** qui correspond au segment /g/ est également reproduit pour les deux locutrices (*fig.* **??**-c, **??**-c). Ce résultat est également vérifié, lorsque le phonème est obtenu par les mouvements coordonnés de plusieurs articulateurs. Ces observations suggèrent que la commande des articulateurs primaires est indépendante de la locutrice . Lorsque l'articulateur est secondaire, les positions et amplitudes des impulsions générées ne sont pas toujours comparables pour les deux locutrices.

L'intervalle entre deux impulsions varie entre $80ms$ (la moitié de la durée moyenne de la syllabe) et $400ms$ (deux fois la durée moyenne de la syllabe). Les positions et amplitudes des impulsions dépendent de la nature des segments phonétiques qui se succèdent. Lorsque le paramètre est en position neutre, il n'est pas contrôlé (*fig.* **??**-a, **??**-d). En revanche, le maintien d'un mouvement d'amplitude élevée nécessite plusieurs impulsions rapprochées (*fig.* **??**-b, **??**-b, **??**-c). On note qu'il est parfois nécessaire de générer une impulsion pour inhiber l'effet des oscillations secondaires de la réponse impulsionnelle. Un exemple est montré dans la figure **??**-a où la position nulle de la mâchoire correspondant au segment /b i g y/ est obtenue par deux impulsions d'excitation.

D'une manière générale, les impulsions sont générées pour réaliser les transitions et atteindre les positions cibles des articulateurs. Elles sont émises avec une avance temporelle de 40 à $80ms$ par rapport à l'instant correspondant à la position cible. L'analyse de la séquence des premières impulsions déterminées, montre qu'il n'existe pas une relation biunivoque entre les m_{syl} impulsions et les segments syllabiques dans la phrase. Toutefois, lorsque l'articulateur est activé pour produire un son, son mouvement évolue sur une durée syllabique. Le problème est de définir un critère pour activer les articulateurs impliqués dans la production des segments de phonèmes.

Les séquences S_{syl}^{PB} et S_{syl}^{DF} sont différentes. Les différences observées peuvent être liées à l'erreur de synthèse qui est relativement élevée. Les deux séquences convergent-elles, lorsque l'erreur de synthèse devient faible?

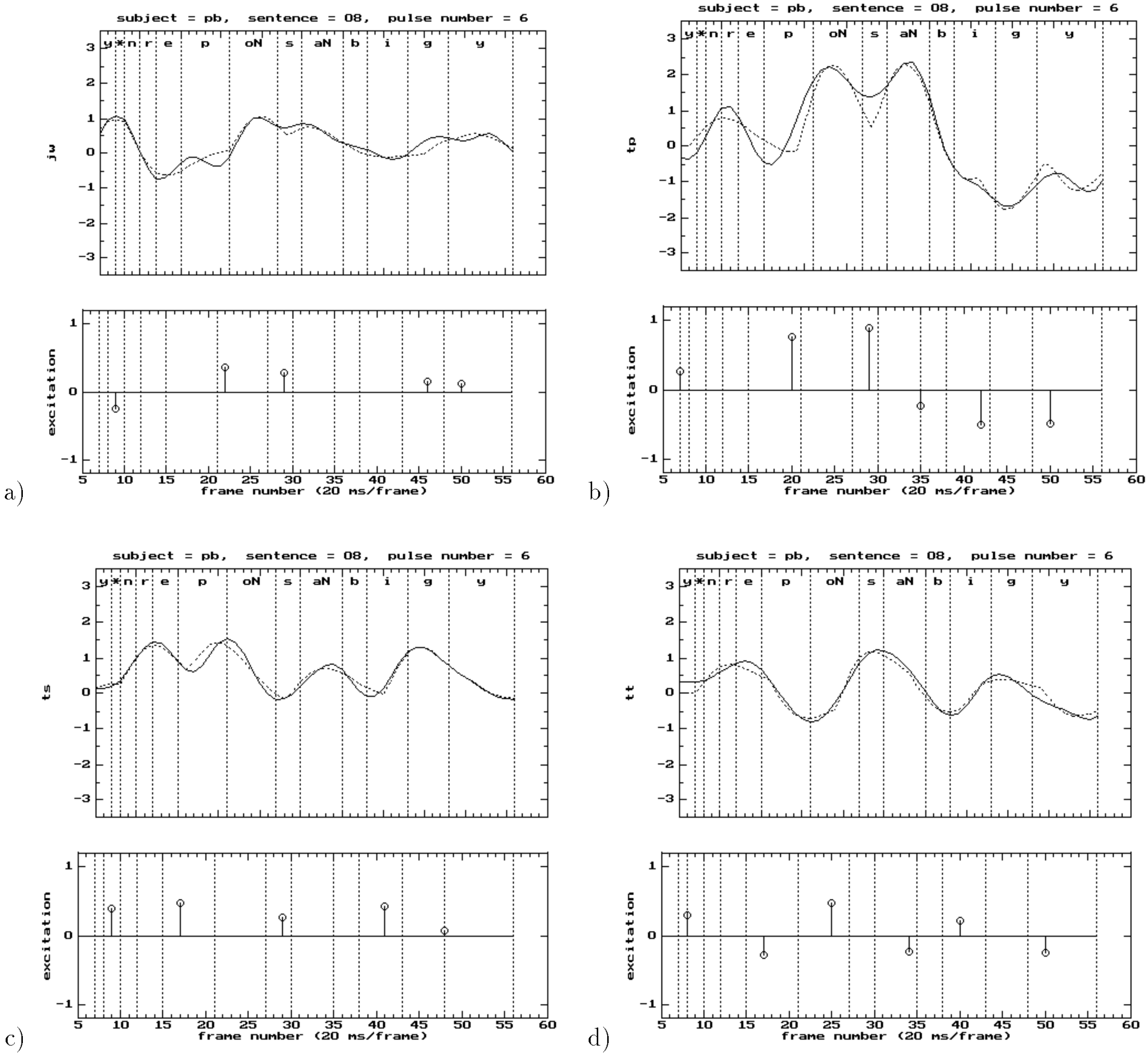


Figure 6.6 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase 08, nombre d'impulsions :* $m = 6$ *(nombre de syllabes).*

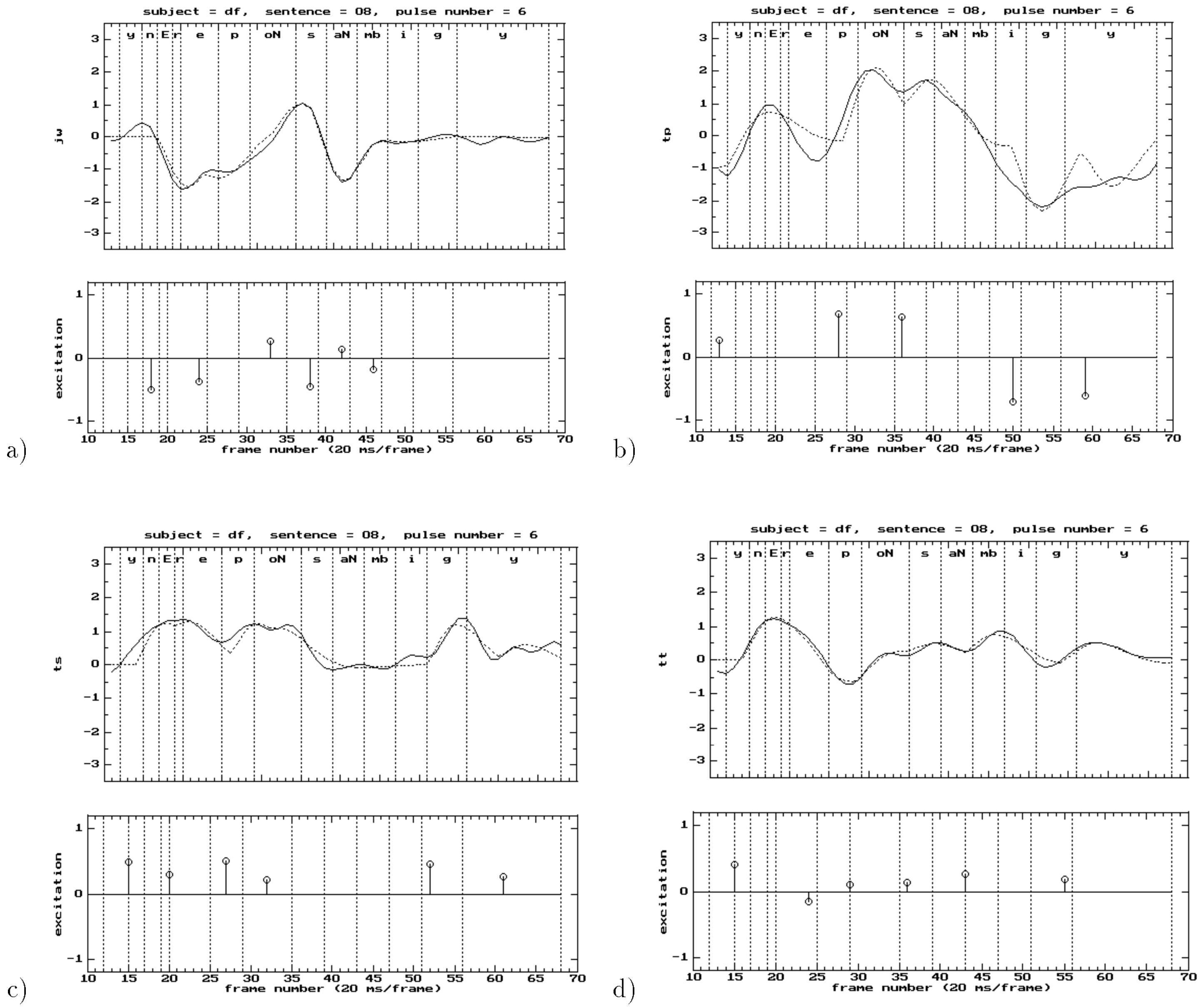


Figure 6.7 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase 08, nombre d'impulsions : $m = 6$ (nombre de syllabes).*

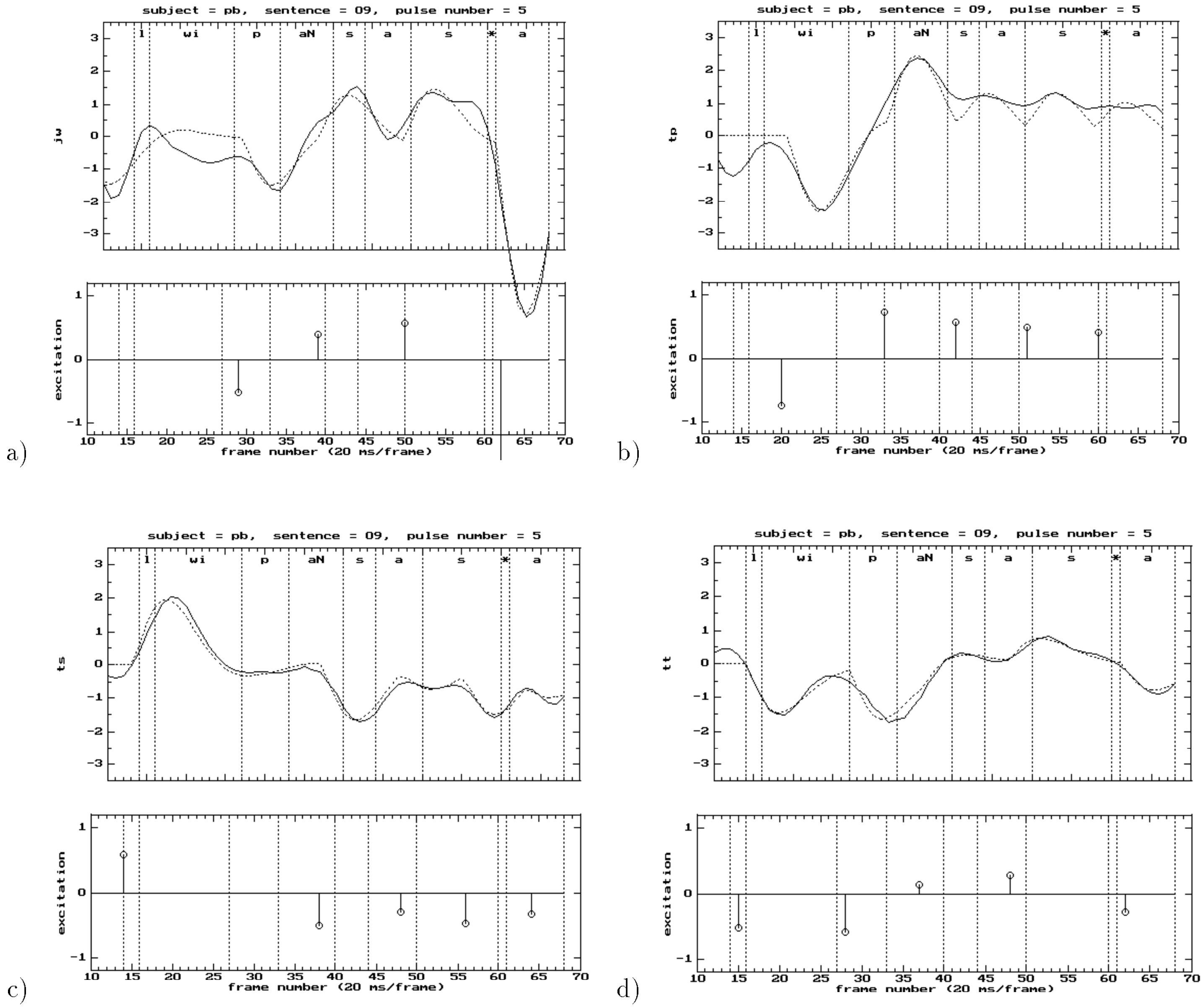
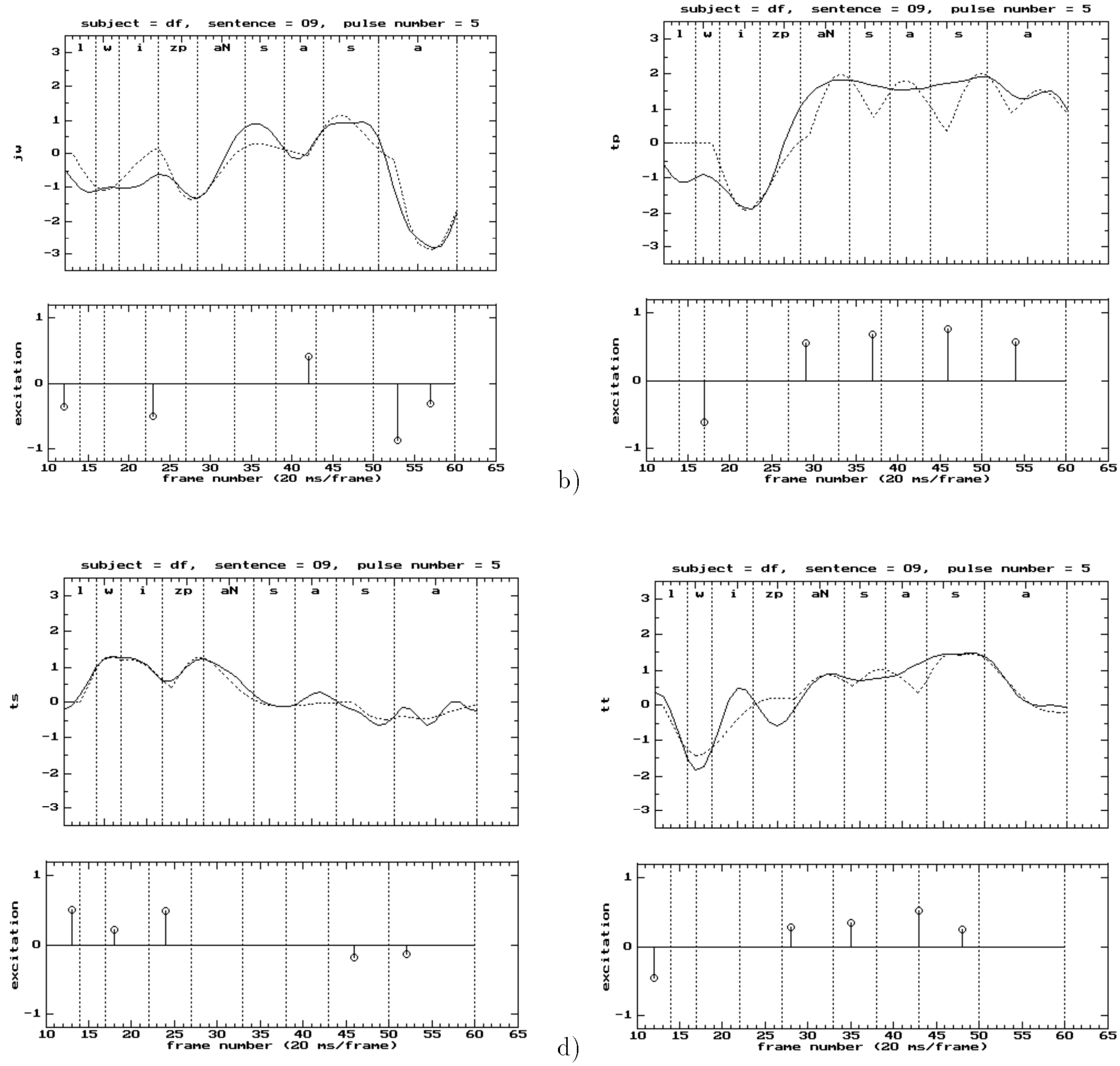


Figure 6.8 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase 09, nombre d'impulsions : $m = 5$ (nombre de syllabes).*



a) b) c) d)

Figure 6.9 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase 09, nombre d'impulsions :* $m = 5$ *(nombre de syllabes).*

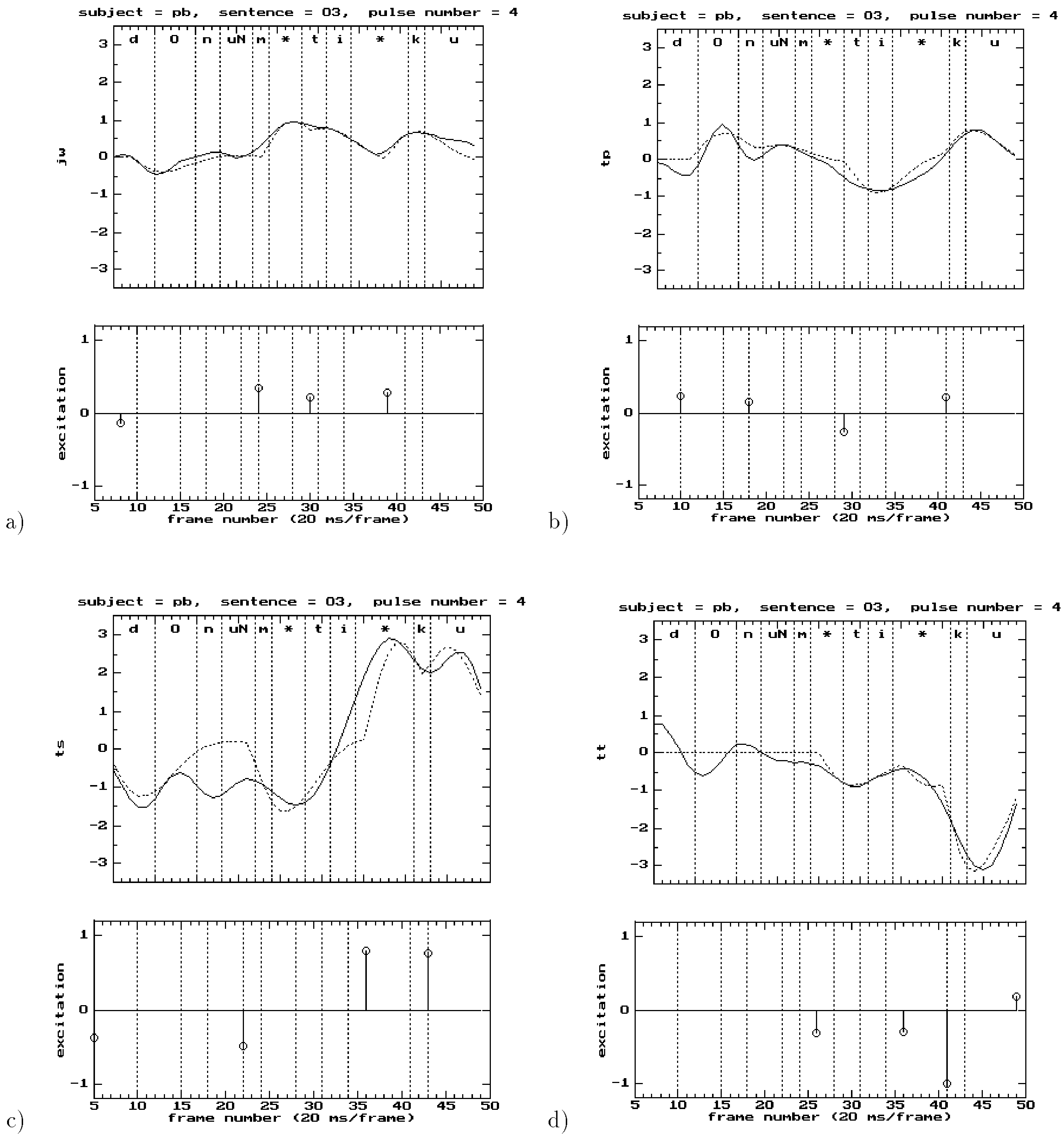


Figure 6.10 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase 03, nombre d'impulsions : $m = 4$ (nombre de syllabes).*

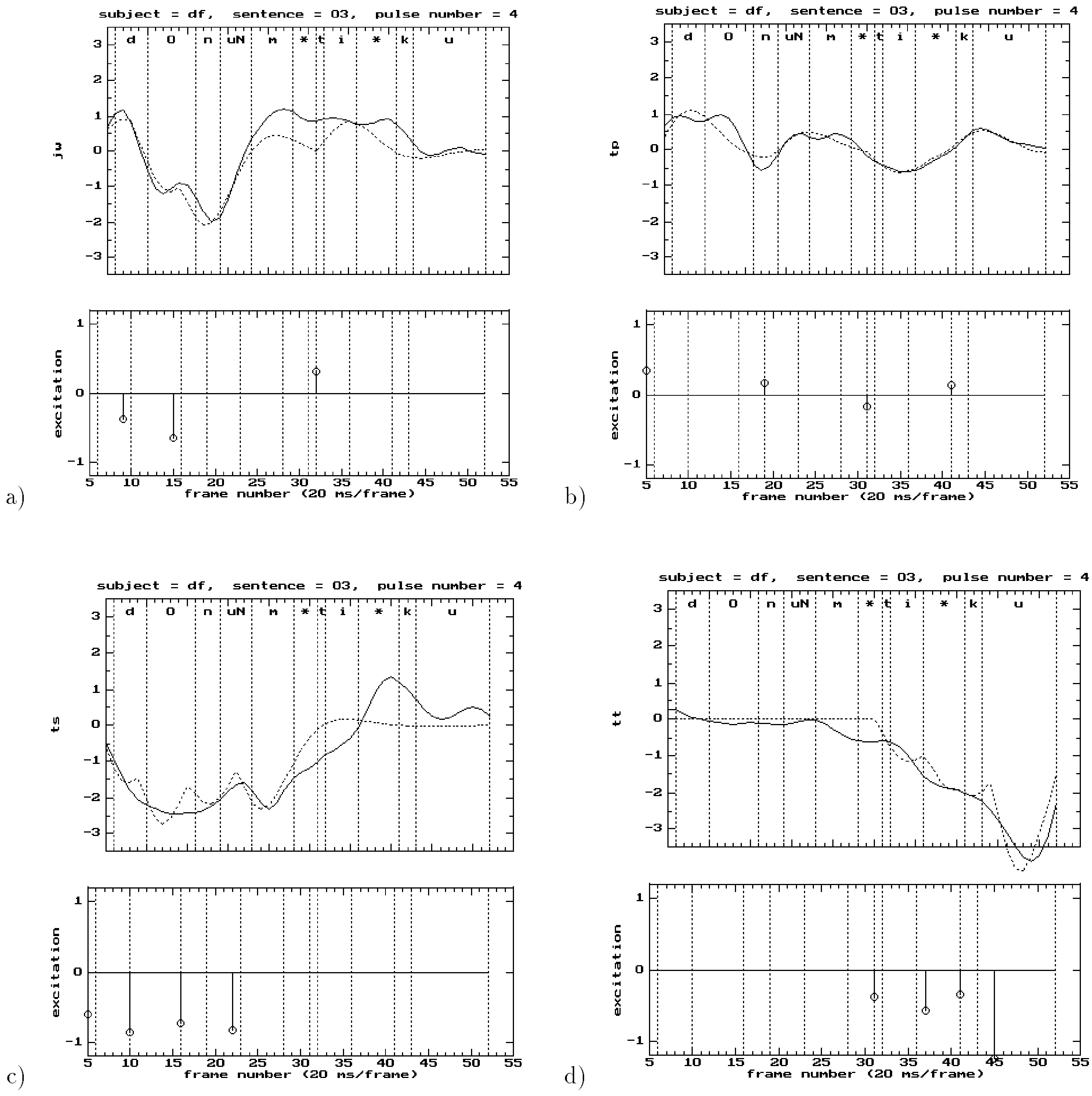


Figure 6.11 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase* 03*, nombre d'impulsions :* $m = 4$ *(nombre de syllabes).*

6.4.2 Organisation par rapport au phonème

Les mouvements des différents articulateurs sont maintenant synthétisés par la séquence S_{pho}. Celle ci correspond au nombre d'impulsions égal au nombre de phonèmes dans la phrase, $m = m_{pho}$. Les figures **??** à **??** montrent que les mouvements synthétisés et ceux observés sont pratiquement identiques. L'enveloppe formée par les amplitudes des impulsions suit globalement la trajectoire des mouvements générés. Les impulsions ne sont pas uniformément distribuées sur les segments phonémiques. On note l'absence d'impulsions sur certains segments, notamment lorsque l'articulateur est dans une position nulle (*fig.* **??**-a, **??**-d, **??**-c). En revanche, des segments peuvent avoir jusqu'à trois impulsions. Par exemple la voyelle finale /u/ pour l'articulateur **ts** (*fig.* **??**-c), et la voyelle finale /a/ pour **jw** (*fig.* **??**-a). Ceci signifie que les séquences S^{PB} et S^{DF} ne convergent pas lorsque l'erreur de synthèse devient faible. Les différences les plus souvent observées sont de deux types :

1. Un groupement d'impulsions, une impulsion d'amplitude élevée émise pour une locutrice peut être remplacée par un ensemble de deux à trois impulsions de plus faible amplitude pour l'autre locutrice ;

2. Les impulsions de faible amplitude qui sont souvent propres à la locutrice.

Ces différences sont relevées sur des parties des trajectoires des mouvements qui ont la même allure pour les deux locutrices. Seules l'amplitude et l'étendue du mouvement diffèrent. Nous avons, toutefois, relevé des différences qui agissent sur l'évolution de la trajectoire, ainsi que le montrent les figures **??**-a) et **??**-a) sur le segment /s aN/ de la phrase /y n E r p on **s aN** b i g y/. Dans le cas de la locutrice DF, une impulsion négative est générée pour produire l'ouverture de la mâchoire, tandis qu'une impulsion positive est émise pour la locutrice PB. Un autre exemple où la commande de l'articulateur **jw** s'opère différemment par les deux locutrices est montré dans les figures **??**-a) et **??**-a). Sur le segment /l wi/ de la phrase 09, on note une fermeture de la mâchoire correspondant à une impulsion positive et une aperture de l'articulateur pour DF générée par une impulsion d'amplitude négative. Ces différences ne semblent pas se répercuter sur le reste des articulateurs. Ceci suggère que les deux locutrices peuvent choisir des stratégies de commande différentes pour produire le même phonème.

En dépit de la détermination d'impulsions supplémentaires dans la séquence S_{pho}, les impulsions obtenues dans la sous séquence S_{syl} (marquées par des croix sur les figures) gardent des amplitudes relativement élevées. D'autres impulsions d'amplitude élevée peuvent, cependant, être générées. Elles contribuent activement à la mise en forme de la trajectoire des mouvements, ainsi que le montre la figure **??**-c). Dans cet exemple, la reconstitution du mouvement de l'articulateur primaire **ts** sur le segment /k u/ en fin de la phrase est obtenue avec des impulsions qui n'appartiennent pas à la séquence S_{syl}. Par ailleurs, des impulsions de la séquence S_{syl} peuvent avoir des amplitudes extrêmement petite qui n'ont probablement aucun effet acoustique (*fig.* **??**-c). Dans ce cas, il semble qu'un nombre d'impulsions m inférieur à m_{syl} suffit pour synthétiser le mouvement de l'articulateur. Aussi, nous sommes tentés de conclure que la reconstitution des mouvements des phrases n'est pas basée sur le nombre de syllabe.

L'analyse des impulsions d'amplitude très faible dans les modèles d'excitations S_{pho} montre qu'il est difficile de trouver des similitudes entre les deux locutrices. Nous pensons que contrairement aux impulsions d'amplitude élevée, les impulsions de faible amplitude sont propres à la locutrice. La question que nous nous posons est si notre analyse doit porter sur la séquence S_{pho}.

Remarque

Le tableau **??** présente le nombre d'impulsions d'amplitude négative m_- de l'articulateur **jw** généré par les deux locutrices. Ces impulsions contrôlent l'aperture de la mâchoire. Le nombre m_-^{DF} de la locutrice DF est plus grand que celui généré par la locutrice PB. Dans le cas de la phrase 03 où m_-^{DF} et m_-^{PB} sont égaux, on relève une différence appréciable entre les amplitudes des impulsions des deux locutrices qui sont plus élevées pour DF que pour PB. Nous pensons que cette différence est liée au mode de production des phrases lors de l'enregistrement ; la locutrice DF avait tendance à mieux articuler et donc l'activité de la mâchoire était plus accentuée. Cette différence ne semble pas se répercuter sur les séquences d'impulsions des paramètres **tp** et **ts**.

Phrase	08	09	03
PB	4	4	2
DF	9	7	2

Tableau 6.1 Nombre d'impulsions d'amplitude négative du paramètre **jw**.

Ces observations suggèrent que des régularités de commande existent entre les séquences d'impulsions générées par les deux locutrices . L'analyse de la commande entre les articulateurs s'est avérée complexe sur les séquences S_{pho} en raison du nombre élevé d'impulsions. Elle est estimée incomplète sur les séquences S_{syl} dont le nombre d'impulsions est, dans certains cas, insuffisant pour produire tous les segments de la trajectoire des mouvements qui sont prépondérants pour la production des phonèmes.

Les deux nombres m_{syl} et m_{pho} ne semblent pas être critiques pour le contrôle des articulateurs de la langue, lors de la production des phrases.

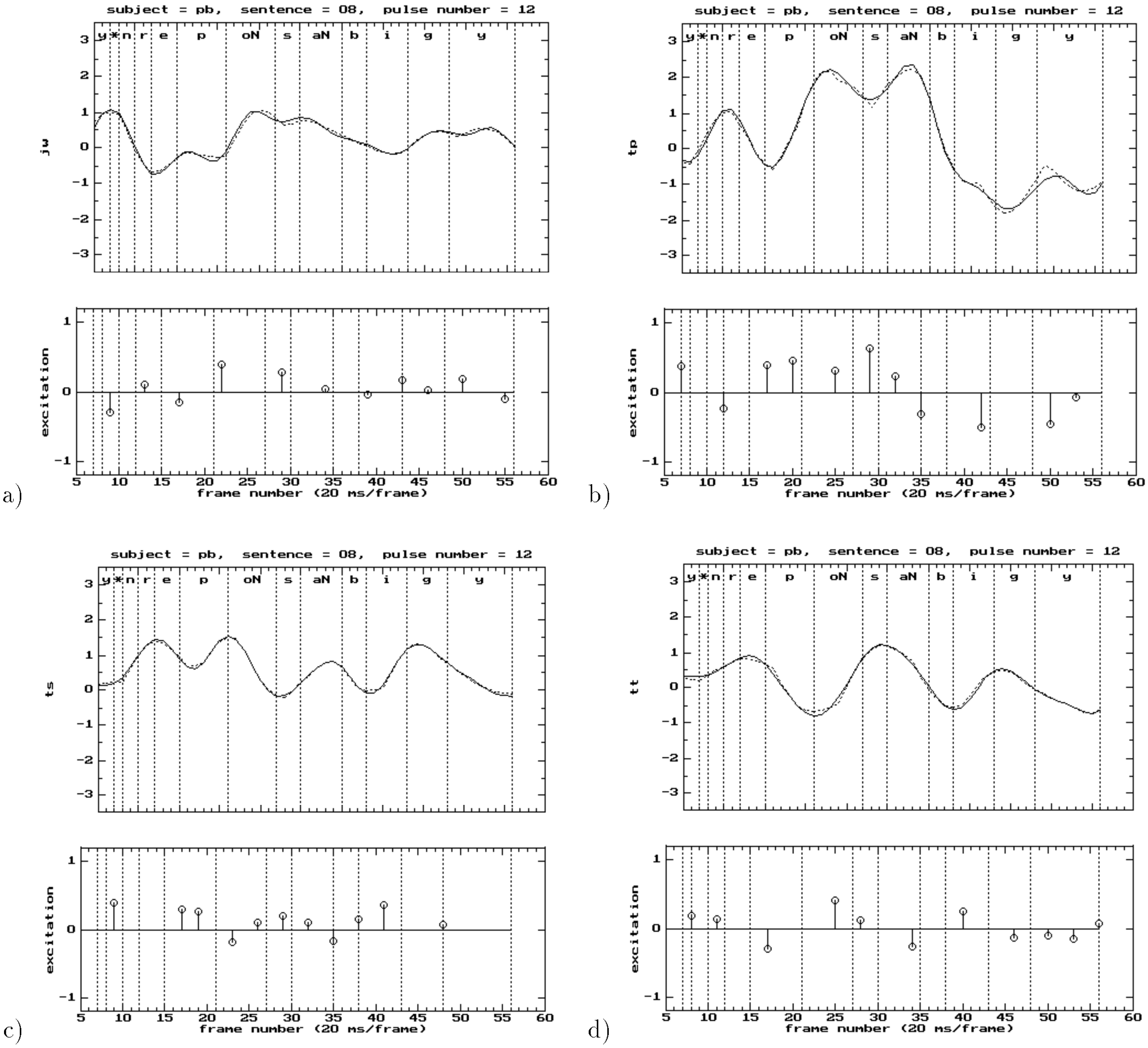


Figure 6.12 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase* 08*, nombre d'impulsions :* $m = 12$ *(nombre de phonèmes).*

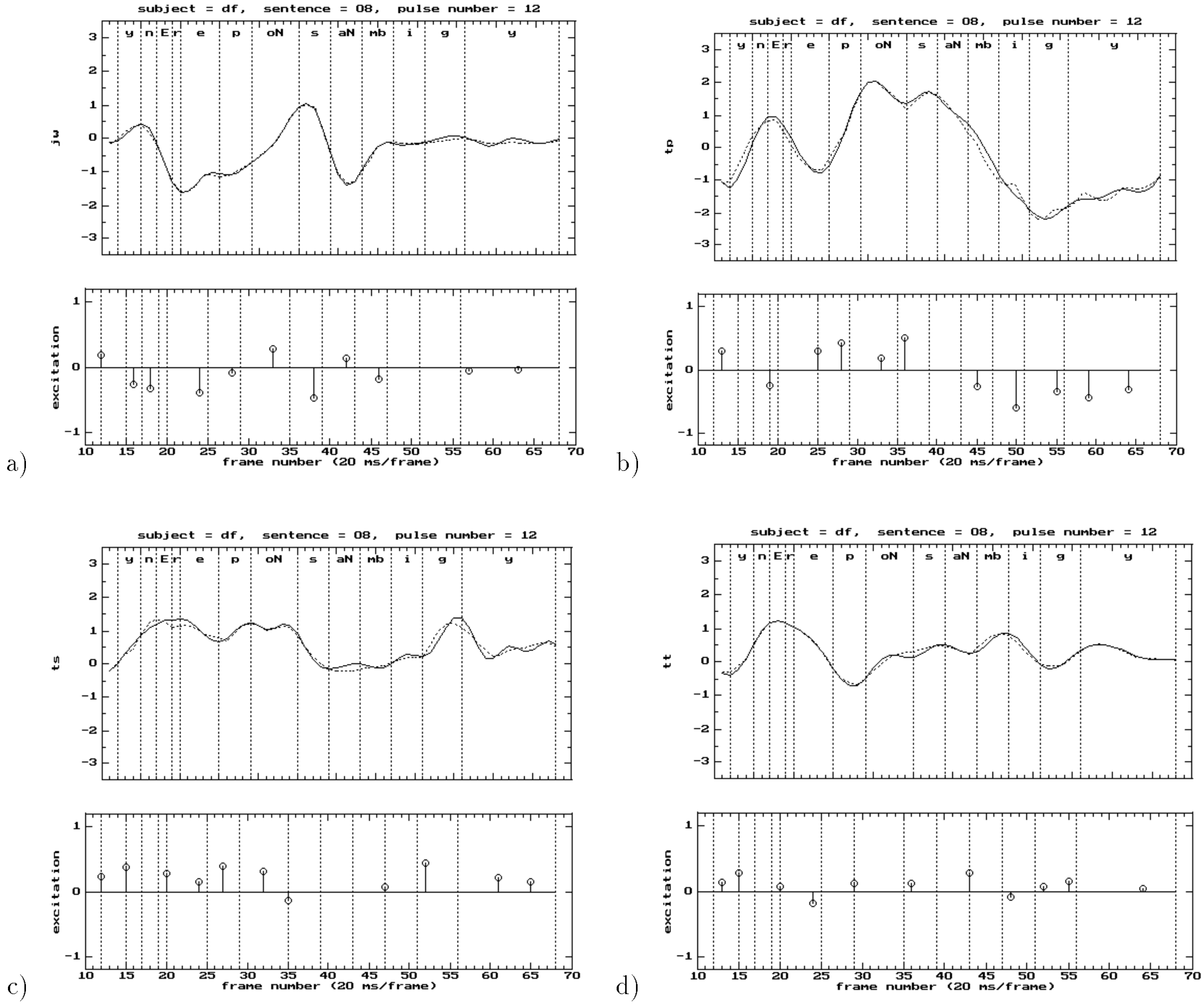


Figure 6.13 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase 08, nombre d'impulsions : m = 12 (nombre de phonèmes).*

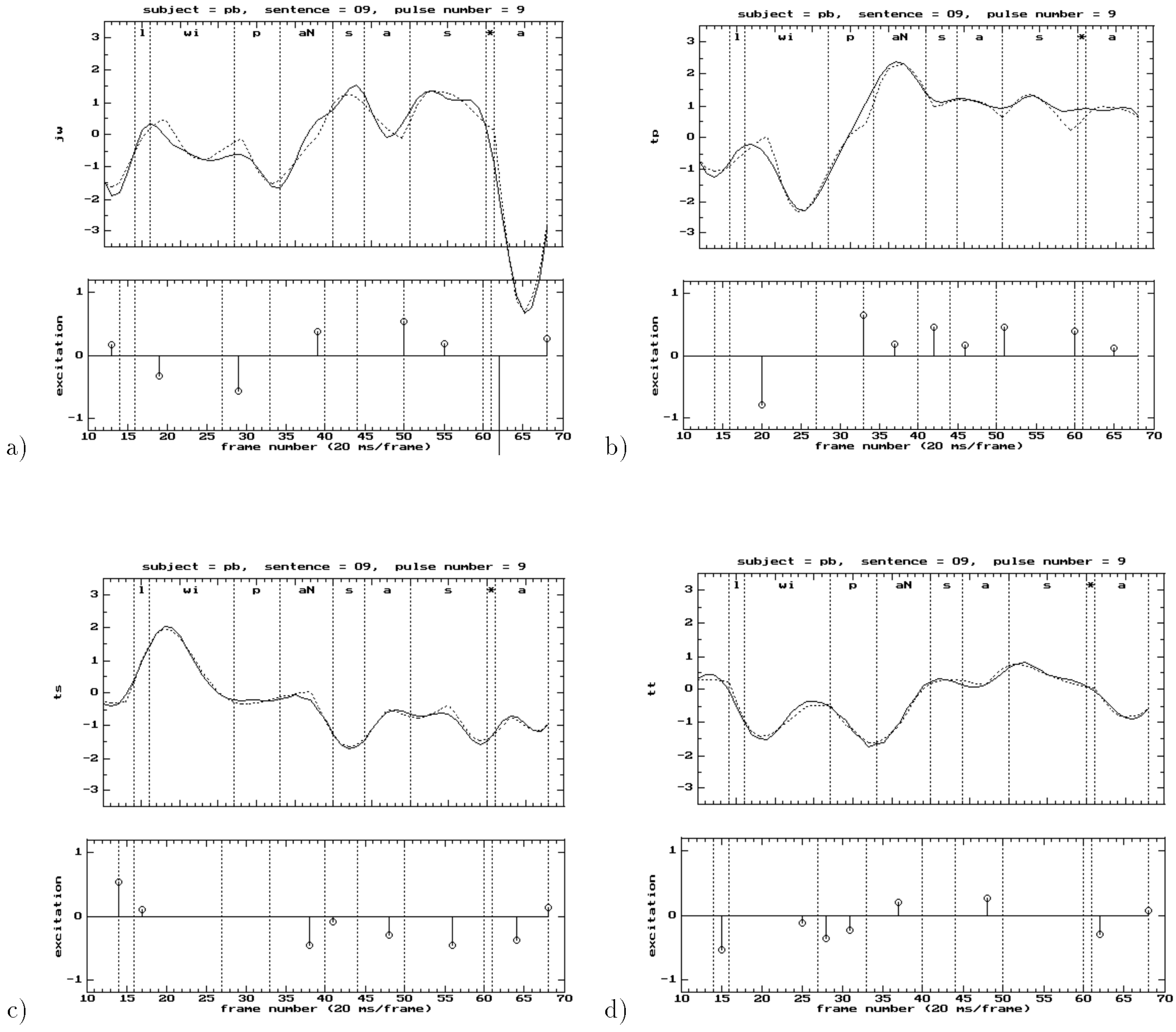


Figure 6.14 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase* 09*, nombre d'impulsions :* $m = 9$ *(nombre de phonèmes).*

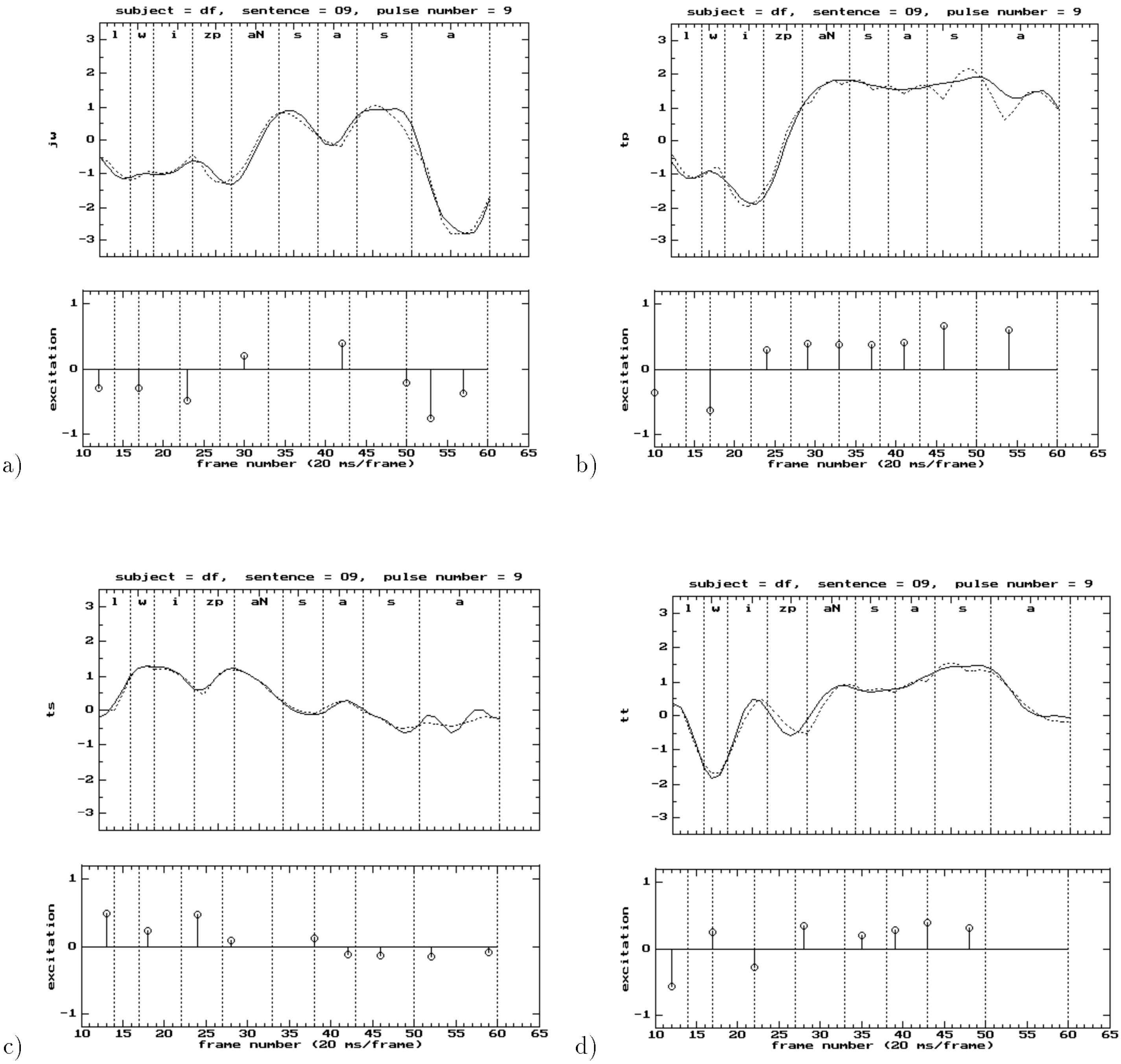


Figure 6.15 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase* 09*, nombre d'impulsions :* $m = 9$ *(nombre de phonèmes).*

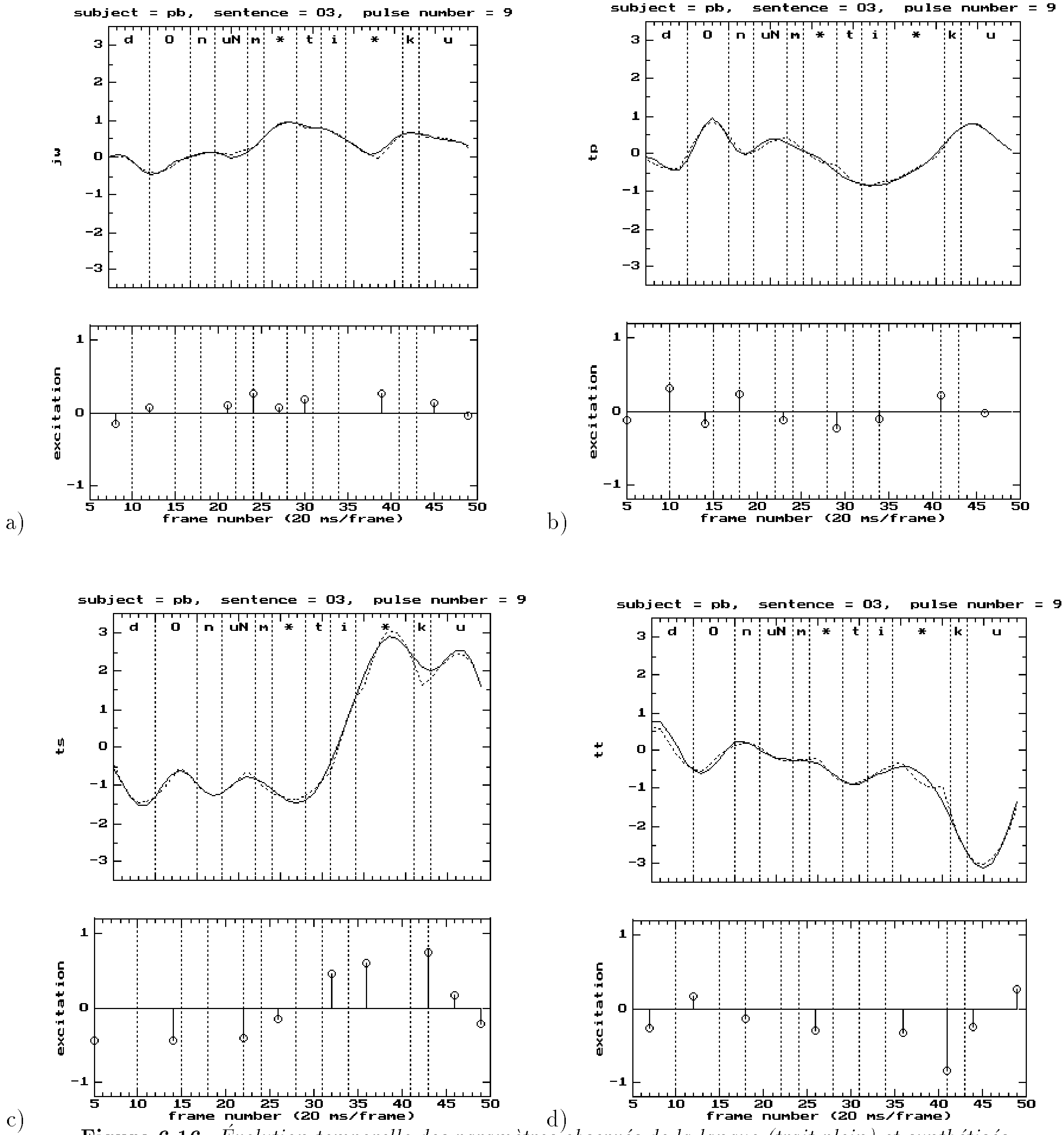


Figure 6.16 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice PB, phrase* 03*, nombre d'impulsions :* $m = 9$ *(nombre de syllabes).*

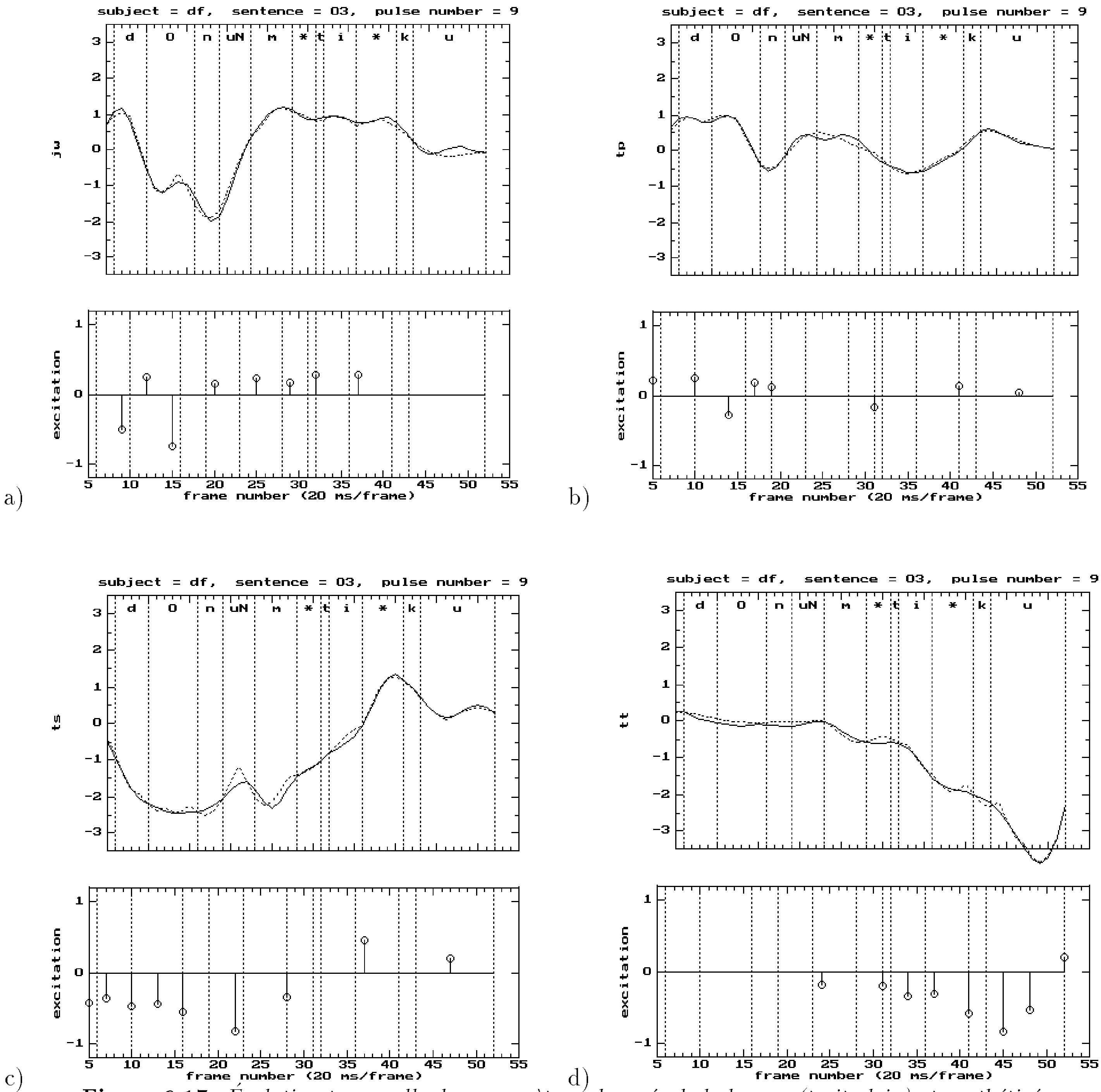


Figure 6.17 *Évolution temporelle des paramètres observés de la langue (trait plein) et synthétisés (pointillé). Locutrice DF, phrase 03, nombre d'impulsions : $m = 9$ (nombre de phonèmes).*

6.5 Conclusion

À l'issue de ce chapitre, une question s'impose : faut-il analyser les commandes sur le nombre d'impulsions correspondant à l'erreur de saturation ? C'est à dire sur le nombre m qui permet une synthèse fidèle du mouvement, sachant qu'en production on n'a pas besoin d'une grande précision de mouvement pour réaliser les sons, notamment les voyelles ? Ou bien suffit-il d'un nombre m^* plus réduit à déterminer ?

Dans la suite de notre travail, notre problème est de déterminer le nombre d'impulsions nécessaires pour produire les mouvements qui correspondent à une synthèse acoustiquement correcte des phrases. Dans le domaine articulatoire, il est difficile de trouver un critère de décision qui permet de déterminer les impulsions utilisées pour la synthèse. C'est pourquoi nous avons projeté notre analyse dans le domaine acoustique.

Chapitre 7

Analyse acoustique

L'analyse des résultats de la synthèse des paramètres articulatoires par la méthode de la LPC multi-impulsionnelle a conduit à des interrogations liées essentiellement à la détermination des impulsions de commande qui jouent un rôle substantiel dans la synthèse des mouvements. Dans le présent chapitre, nous cherchons à élucider cette question. Pour ce faire, nous décrivons, tout d'abord, la méthode de calcul des fréquences de formants. Ensuite, nous discutons l'importance du rôle que joue chaque paramètre articulatoire dans la production des traits phonétiques inhérents aux différents segments qui constituent une phrase. Enfin, nous déterminons le nombre d'impulsions minimum qui correspond aux fréquences de formants calculés avec une précision au dessous du seuil de perception de l'oreille humaine.

7.1 Détermination des patrons formantiques

Le passage des paramètres articulatoires aux fréquences des formants de la parole consiste en quatre étapes à savoir :

1. *le calcul de la forme sagittale* ;
2. *le calcul de la fonction d'aire* ;
3. *la simulation acoustique du conduit vocal* par un modèle électrique de ligne de transmission ;
4. *le calcul des fréquences des formants.*

Ces différentes étapes sont illustrées dans le schéma bloc de la figure **??** et expliquées dans les sections suivantes.

Figure 7.1 *Schéma bloc du passage des paramètres articulatoires aux fréquences des formants.*

7.1.1 Calcul de la forme sagittale

La forme vectorielle $\mathbf{z_{tng}}$ de la langue à chaque trame est reconstituée en appliquant une somme pondérée linéaire sur les quatre paramètres articulatoires de la langue de la manière suivante (*cf.* section **??**) :

$$\mathbf{z}_{tng} = \mathbf{A}\mathbf{y}_{tng}, \tag{7.1}$$

$\mathbf{A}$ est la matrice de pondération, $\mathbf{y}_{tng}$ est le vecteur des composantes articulatoires (**jw**, **tp**, **ts**, et **tt**). Pour aboutir à la forme sagittale correspondant à chaque trame, le vecteur $\mathbf{z}_{tng}$ est ensuite projeté sur un système de coordonnées semi-polaires dont la référence est fixée par rapport au maxillaire. Les contours de la langue sont ainsi reconstitués. Les sections labiale et du larynx, modélisées par deux tuyaux elliptiques uniformes, sont connectés aux deux extrémités du conduit vocal principal. Le premier, vers les lèvres, est calculé en fonction des composantes labiales **lp**, **lh** et **jw**. Le second, vers le larynx, est spécifié par sa position intrinsèque **lx** et **jw**. Le paramètre **lx** du larynx est supposé constant dans cette étude. Sa valeur est identiquement nulle. Elle correspond à la valeur moyenne des données observées.

7.1.2 Calcul de la fonction d'aire

Le passage "coupe sagittale-fonction d'aire", pour une configuration donnée du conduit vocal, est basée sur la formule [?] :

$$A = \alpha d^{\beta}, \tag{7.2}$$

où A et d sont respectivement l'aire de la section et la dimension sagittale. Les coefficients α et β dépendent de la position de la grille dans le système semi-polaire. La figure **??** illustre un exemple de calcul de fonction d'aire pour la voyelle /i/.

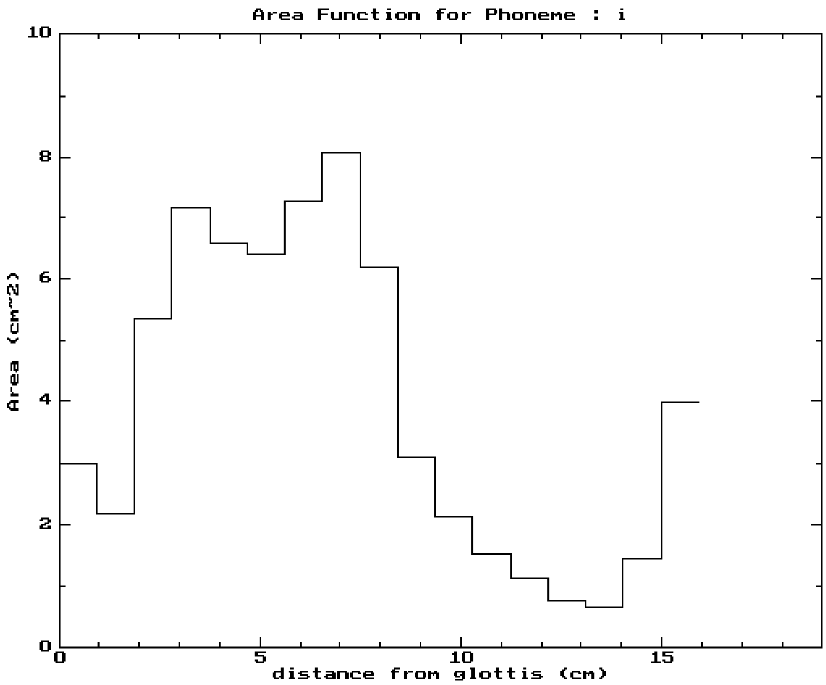


Figure 7.2 *Exemple de fonction d'aire obtenue par les six paramètres observés pour la voyelle /i/.*

7.1.3 Modèle acoustique de simulation du conduit vocal

Une méthode de simulation du conduit vocal dans le domaine temporel a été développée par Maeda (Maeda, 1982 [?]) en vue de produire une synthèse de la parole de bonne qualité à partir du modèle articulatoire. Cette méthode est basée sur la résolution des équations acoustiques de pression et du débit volumique des ondes acoustiques du son dans un tube.

Dans le conduit vocal, le flux d'air créé par la pression des poumons, est modulé par les vibrations des cordes vocales à l'orifice glottale. La source, correspondant à la modulation du flux et le filtre, représentant les caractéristiques acoustiques du conduit vocal, interagissent. La génération et la propagation des ondes acoustiques dans le conduit vocal sont pilotées par un système d'équations qui tient compte de cette interaction. Le modèle source filtre, couramment utilisé dans la synthèse, est une approximation de ce système d'équations acoustiques. Il consiste à supposer que l'impédance d'entrée du filtre est infinie. Par conséquent, la synthèse de la parole par simulation, dans laquelle les propriétés acoustiques du conduit vocal et l'interaction source filtre est prise en compte, produit des signaux de parole de meilleure qualité que le modèle source filtre. Le modèle de simulation comporte

- *une source de pression d'air constante* qui génère le flux et du bruit ;
- *une section étroite variable dans le temps* représentant la glotte ;
- *un tube* correspondant au conduit vocal couplé à la cavité nasale.

Le conduit vocal est discrétisé dans le temps et dans l'espace. Il comporte M tubes de longueur variable de la glotte aux lèvres. L'ensemble constitue une ligne de transmission. Chaque tube est modélisé par un circuit électrique qui exprime les lois acoustiques dans un tuyau non rigide. Il considère l'impédance mécanique due au mouvement des parois qui sont soumises à l'onde sonore. L'effet de cette impédance se traduit par un amortissement de résonances en basses fréquences de la fonction de transfert du conduit vocal et, par la modification de la valeur du premier formant (en basses fréquences). De plus, le modèle prend en compte l'effet de la charge acoustique due au rayonnement aux lèvres (*fig.* **??**).

Par ailleurs, les éléments acoustiques, tels que la viscosité due au frottement des particules de l'air le long du conduit vocal et la conduction thermique due à la proximité des particules de l'air le long de la paroi qui produisent des pertes relativement faibles par rapport aux deux précédants, sont négligés.

Le conduit vocal est correctement représenté par le réseau en ligne de transmission lorsque la fréquence d'échantillonnage f_s est égale à $40kHz$ (discrétisation temporelle) et la longueur maximale d'une section spatiale est de $1cm$ (discrétisation spatiale). Ce modèle

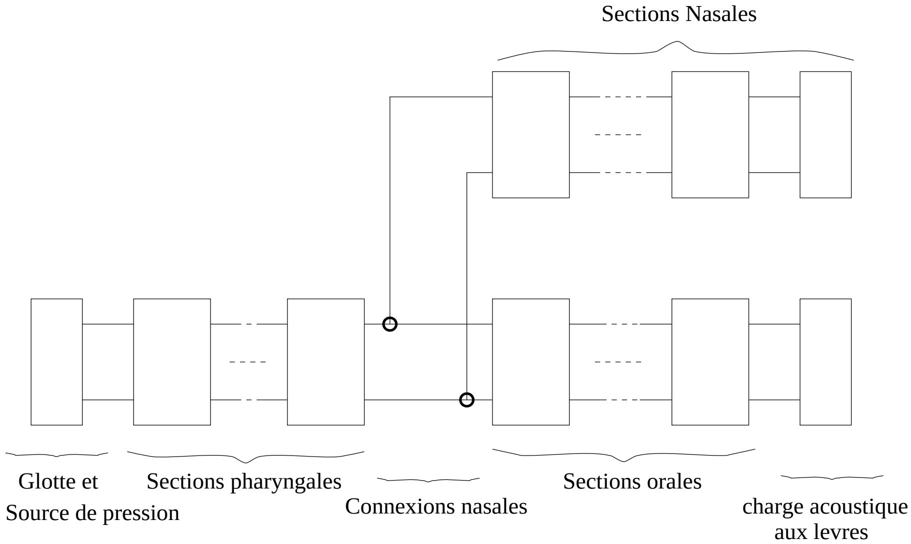


Figure 7.3 *Représentation du modèle en ligne de transmission électrique du conduit vocal.*

ne comprend pas les cordes vocales. La fonction temporelle de la variation de la section de la glotte est spécifiée directement comme une partie de la fonction d'aire du conduit vocal. Notons enfin, que ce modèle est implémenté pour le calcul des formants des voyelles orales et nasales. La génération de turbulence pour la production des consonnes n'a pas encore été mise au point.

7.1.4 Calcul des fréquences des formants

Les fréquences et amplitudes des formants $F_n, n = 1...4$, sont estimées par un algorithme de détection des pics de la fonction de transfert donnée par le modèle électrique de la ligne de transmission correspondant à chaque trame. On note que les fonctions de transfert sont obtenues par le calcul des équations linéaires dans le domaine fréquentiel. Cette méthode est plus rapide que celle du calcul des équations différentielles dans le domaine temporel (Maeda, 1982 [?]). Les patrons formantiques reconstitués à partir des sept paramètres articulatoires débruités constituent les patrons de références, notés $\mathcal{F}_{ref}$. Ils sont comparés aux spectrogrammes enregistrés au laboratoire lors de l'expérience. Les trajectoires des formants de $\mathcal{F}_{ref}$ sont correctement reproduites du point de vue qualitatif. La figure **??** montre la similitude du spectrogramme et du patron des fréquences des formants pour la phrase "*une réponse ambiguë*".

Cette méthode de calcul des fréquences des quatre premiers formants des voyelles a été appliquée aux différentes phrases.

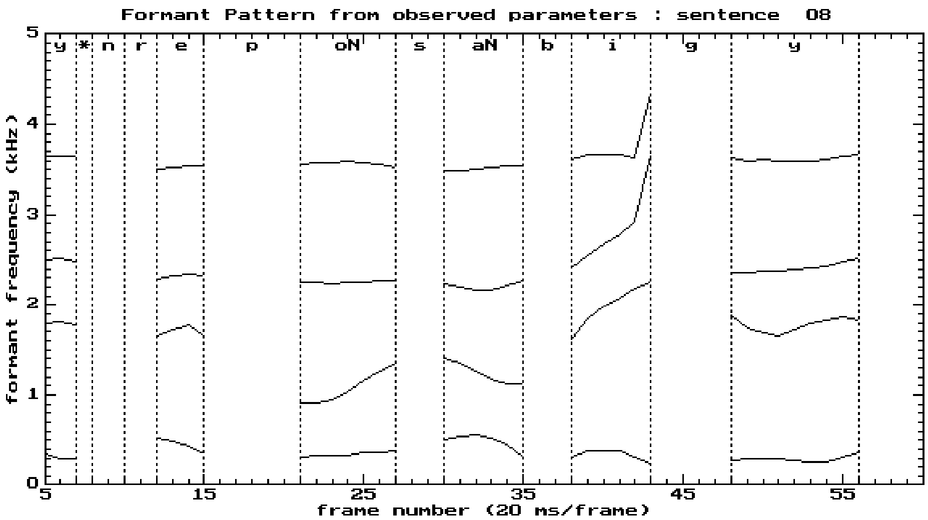


Figure 7.4 *Comparaison des fréquences des formants calculées $\mathcal{F}_{ref}$ et du spectrogramme.*

7.2 Effet du débruitage des données radiofilms sur les patrons des formants

Le premier traitement appliqué aux données des mouvements articulatoires correspond à un débruitage. Afin d'observer les conséquences de cette opération dans le domaine acoustique, les patrons des fréquences de formants ont été directement calculés à partir des mouvements observés. Ils sont notés $\mathcal{F}_{obs}$. La figure ?? montre les deux patrons $\mathcal{F}_{obs}$ en pointillé et $\mathcal{F}_{ref}$ en trait plein. On remarque que les trajectoires des formants subissent un effet de lissage de la même manière que les mouvements articulatoires. Les trajectoires lissées de $\mathcal{F}_{ref}$ suivent l'allure tracée par les trajectoires bruitées. Les informations acoustiques n'ont donc quasiment pas été affectées par l'opération de débruitage.

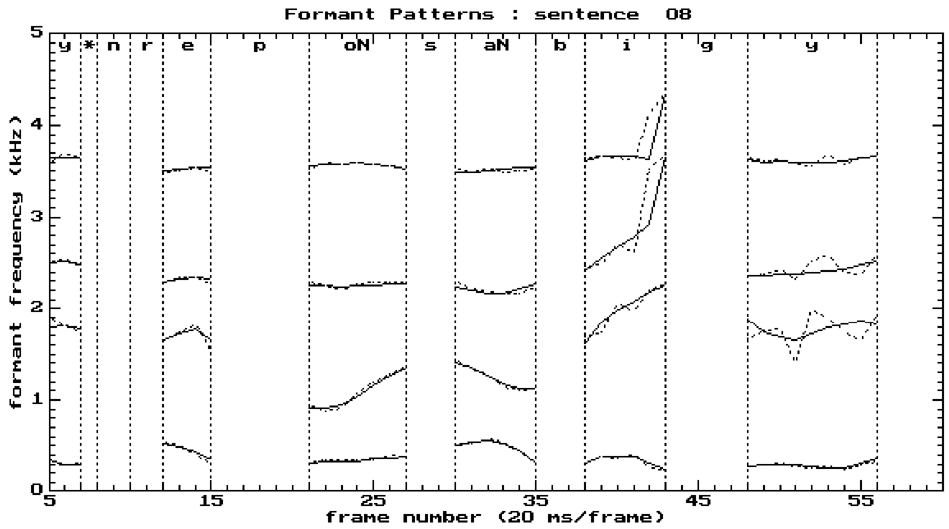
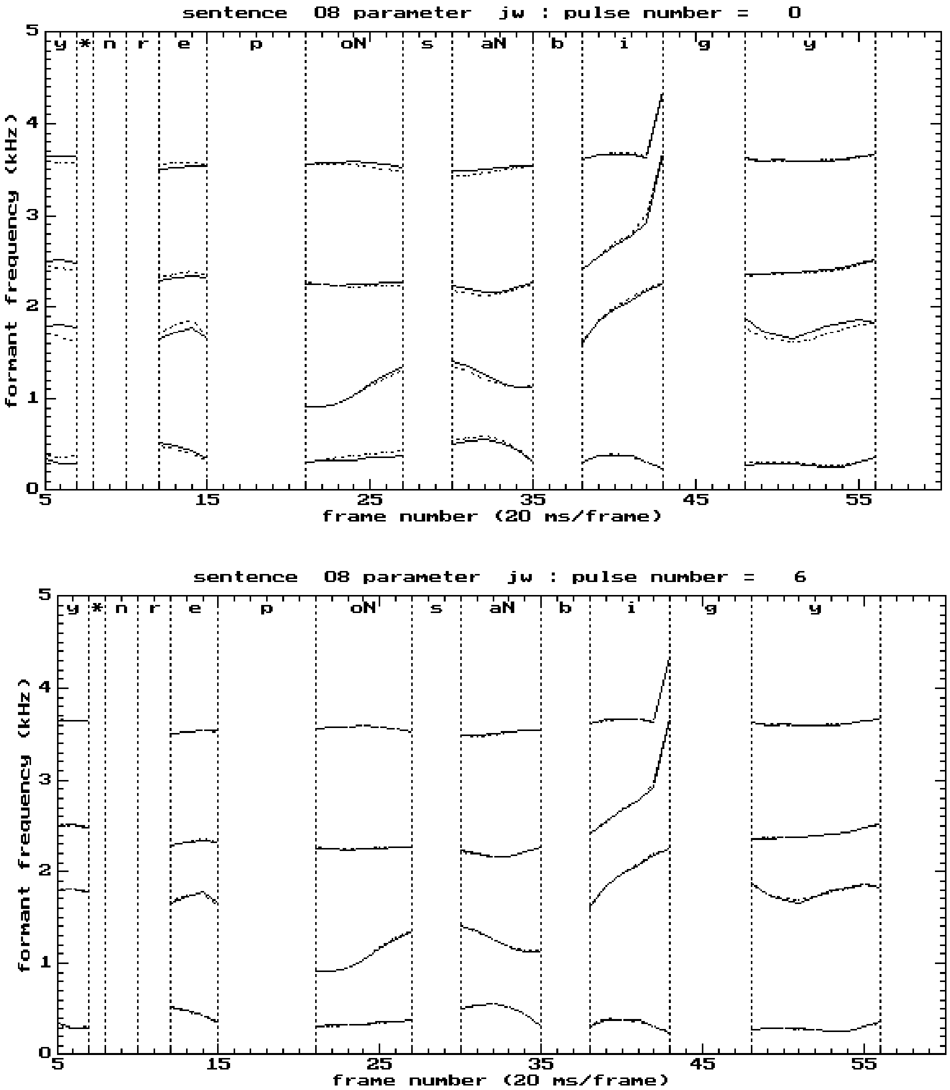


Figure 7.5 *Patrons des formants $\mathcal{F}_{obs}$ (trait plein) et $\mathcal{F}_{ref}$ (trait en pointillé).*

7.3 Effet acoustique des paramètres articulatoires

Dans cette section, nous décrivons qualitativement le rapport entre les phénomènes articulatoires et acoustiques. Nous discutons le rôle que joue chaque paramètre dans la réalisation des caractéristiques acoustiques des différents segments d'une phrase. Dans la suite, le patron $\mathcal{F}_{m,j}$ correspond aux fréquences des formants calculées à partir d'un paramètre articulatoire j ($j =$ **jw**, **tp**, **ts** ou **tt**), synthétisé en fonction du nombre d'impulsions m. Les six autres paramètres nécessaires pour la reconstitution de la forme du conduit vocal sont les paramètres de référence.

Les figures **??** à **??** comparent le patron des formants de référence $\mathcal{F}_{ref}$ (trait plein) et celui des formants calculés $\mathcal{F}_{m,j}$ (pointillé) pour les deux valeurs de m, $m = 0$ et $m = 6$. Lorsque $m = 0$, le paramètre articulatoire synthétisé j du patron $\mathcal{F}_{m,j}$ est inactif. Le patron des formants $\mathcal{F}_{0,j}$ obtenu pour $m = 0$ permet donc d'observer l'effet du paramètre j sur la production des séquences phonétiques composant la phrase.

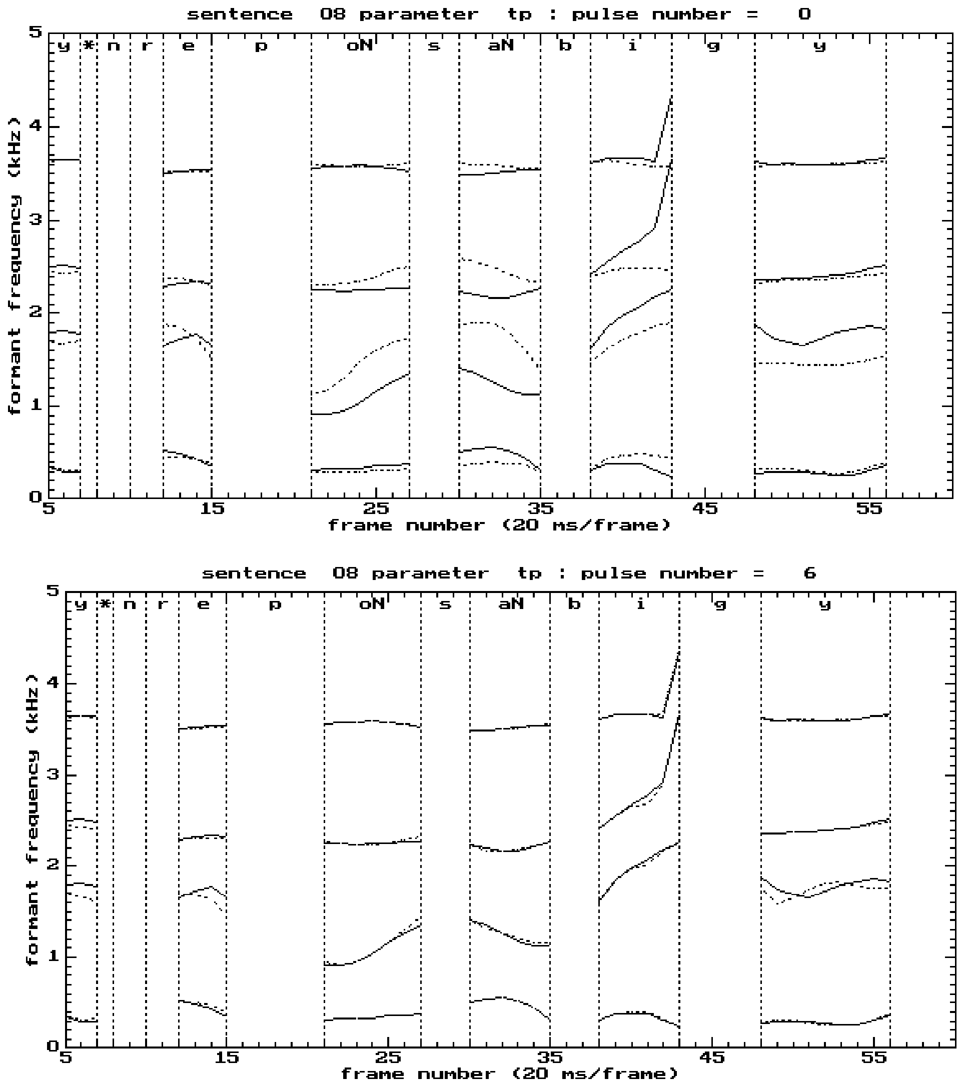


a)

b)

Figure 7.6 *Patrons de formants* $\mathcal{F}_{ref}$ *(trait plein) et* $\mathcal{F}_{m,\mathbf{jw}}$ *(pointillé). Nombre d'impulsions : a-* $m = 0$*. b-* $m = 6$*.*

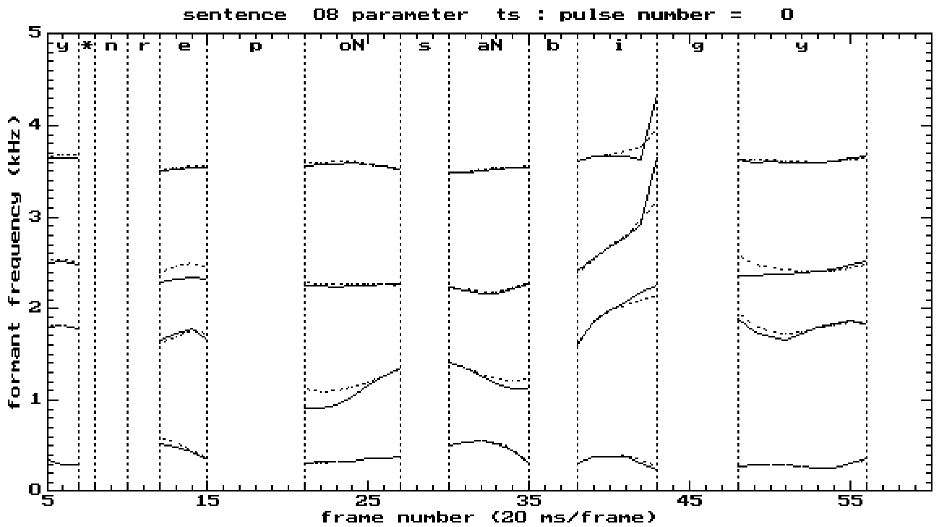
D'une manière générale, ces figures montrent que la distance entre les fréquences des formants de référence $\mathcal{F}_{ref}$ et celle des formants calculés $\mathcal{F}_{m,j}$ diminue lorsque le nombre d'impulsions m augmente. Toutefois, cette distance dépend clairement du paramètre synthétisé j et des propriétés phonétiques du segment acoustique considéré, ainsi que nous le verrons ci-dessous.



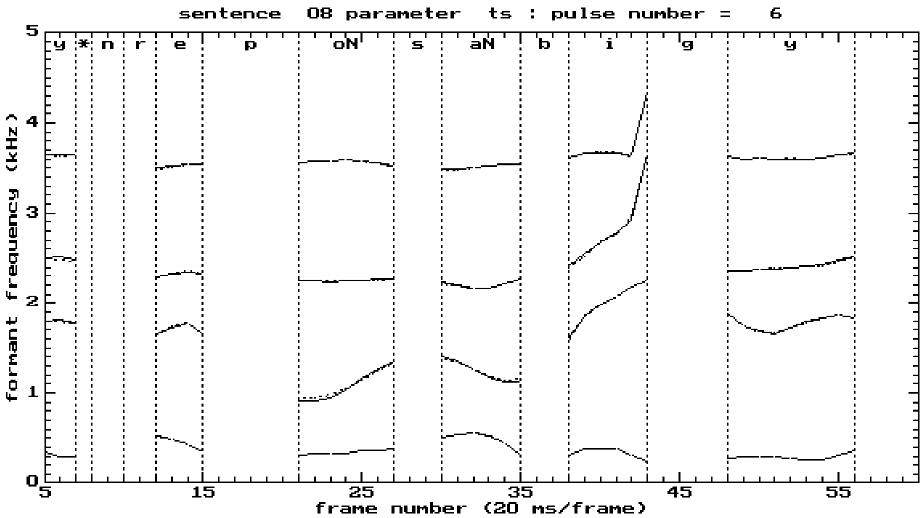
a)

b)

Figure 7.7 *Patrons de formants* $\mathcal{F}_{ref}$ *(trait plein) et* $\mathcal{F}_{m,\mathrm{tp}}$ *(pointillé). Nombre d'impulsions : a-* $m = 0$*. b-* $m = 6$*.*



a)



b)

Figure 7.8 *Patrons de formants $\mathcal{F}_{ref}$ (trait plein) et $\mathcal{F}_{m,\text{ts}}$ (pointillé). Nombre d'impulsions : a- $m = 0$. b- $m = 6$.*

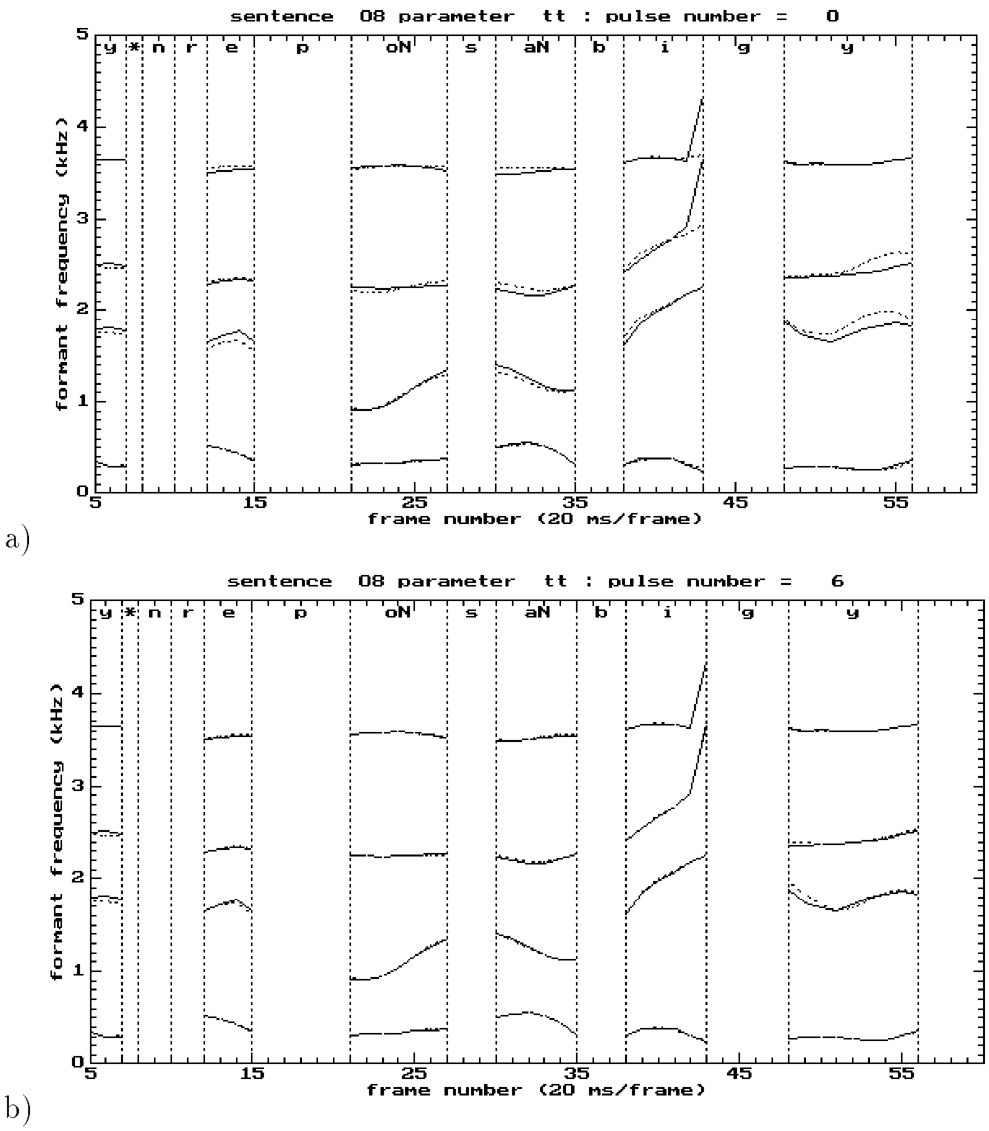


Figure 7.9 *Patrons de formants $\mathcal{F}_{ref}$ (trait plein) et $\mathcal{F}_{m,\texttt{tt}}$ (pointillé). Nombre d'impulsions : a- $m = 0$. b- $m = 6$.*

7.3.1 Effet de la position de la mâchoire

Lorsque la position de la mâchoire correspond à la valeur moyenne, les deux patrons des fréquences des formants de référence et calculées ne sont pas très différents. Ce résultat s'explique par le fait que la locutrice PB utilise très peu sa mâchoire quand elle parle ainsi que nous l'avons fait remarquer dans le chapitre ??. Néanmoins, il est intéressant de noter que cette différence est plus importante pour les voyelles fermées que pour les voyelles ouvertes. Ceci suggère que lorsque le mouvement de la mâchoire est de faible amplitude le phénomène de compensation par la langue, pour atteindre la configuration désirée de la fonction d'aire, est plus important pour les voyelles dont le degré d'aperture est élevé.

Lorsque une voyelle ouverte est située en fin de phrase, elle est accentuée. Dans ce cas, le trait distinctif correspondant à l'aperture est déterminant pour la synthèse correcte de cette voyelle. Ce phénomène est observé sur la voyelle finale /a/ dans les deux phrases "*Louis pense à ça*" et "*il fume son tabac*". La distance entre les patrons $\mathcal{F}_{0\mathbf{jw}}$ et $\mathcal{F}_{\mathbf{ref}}$ est plus élevée sur le segment vocalique final que sur le segment de la voyelle /*a*/ dans un contexte alvéolaire (/s a s/), ainsi que le montre la figure ??.

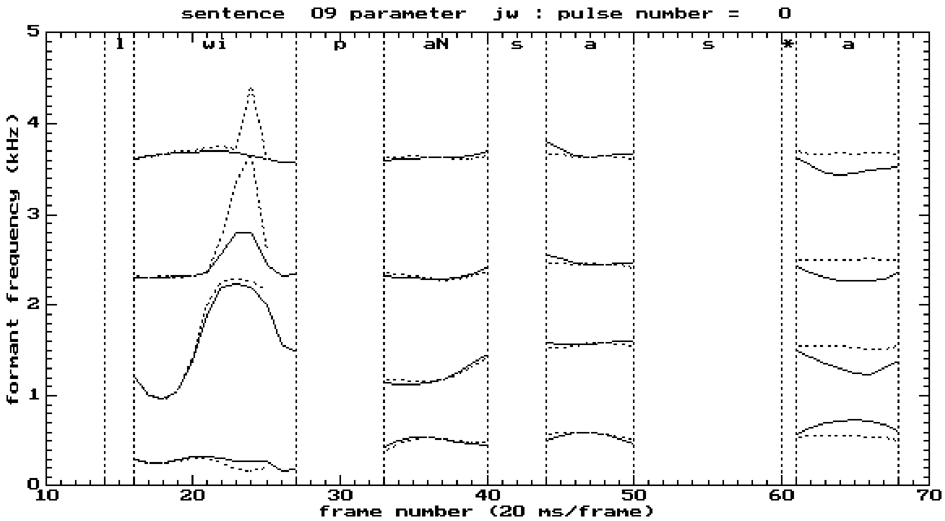


Figure 7.10 *Patrons de formants $\mathcal{F}_{ref}$ (trait plein) et $\mathcal{F}_{m,\mathbf{jw}}$ (pointillé). Nombre d'impulsions $m = 0$.*

<u>Remarque</u>

Le /wi/ sur la figure montre une distance importante entre les fréquences de références et celles calculées des deux derniers formants F_3 et F_4. Cette erreur est liée à un problème d'identification des pics du spectre. Lorsque la mâchoire est inhibée, les fréquences des formants sont erronées. Dans le cas présent, les deux formants F_2 et F_3 se sont confondus. Le détecteur de pics n'en a identifié qu'un seul. On remarque, en effet, que le troisième formant calculé est situé dans la zone des fréquences du formant F_4.

7.3.2 Effet de la position de la langue

La figure **??**-a montre l'effet de la position de la langue **tp** sur l'antériorité/postériorité des voyelles dont le corrélat acoustique est F_2, mais également sur le degré d'aperture et la labialité caractérisée respectivement par les paramètres acoustiques F_1 et F_3. Il apparaît clairement que la distance entre les deux formants F_1 et F_2 est fortement biaisée lorsque le paramètre **tp** est inhibé. Notons, en particulier, la diminution de l'écart $F_1 - F_2$ pour les voyelles antérieures $/i/$ et $/y/$ et l'augmentation de cet écart pour les nasales $/aN/$ et $/oN/$. Ceci signifie que lorsque $\mathbf{tp} = 0$, les voyelles s'éloignent de leur caractéristiques acoustiques. Elles ont tendance à se neutraliser.

Par ailleurs, les perturbations introduites sur la valeur du premier formant F_1 (une augmentation de F_1 pour les voyelles fermées, $/i/$ par exemple, et sa diminution pour les voyelles ouvertes, $/oN/$ par exemple) d'une part, et sur la valeur du troisième formant (F_3 augmente pour les voyelles arrondies $/y/$, $/aN/$ et $/oN/$ et diminue pour la voyelle étirée $/i/$) d'autre part suggèrent que l'inhibition de **tp** modifie fortement les caractéristiques de la fonction d'aire d'un phonème donné.

Le mouvement de la position de la langue paraît prépondérant pour la production des traits distinctifs des voyelles en général, et la réalisation du trait antériorité/postériorité en particulier.

7.3.3 Effet de la forme de la langue

Ce paramètre contribue à la réalisation des voyelles pour lesquelles la forme de la langue est arquée ($/u/$, par exemple) ou plate ($/a/$, par exemple). Ce paramètre est, toutefois, moins important pour la production des voyelles plates, notamment lorsque le degré d'aperture de la mâchoire est faible. Dans la figure **??**, la postériorité de la voyelle $/u/$ est largement détériorée. Ce paramètre est indispensable pour la production des phonèmes caractérisés par une forme ballonnée de la langue.

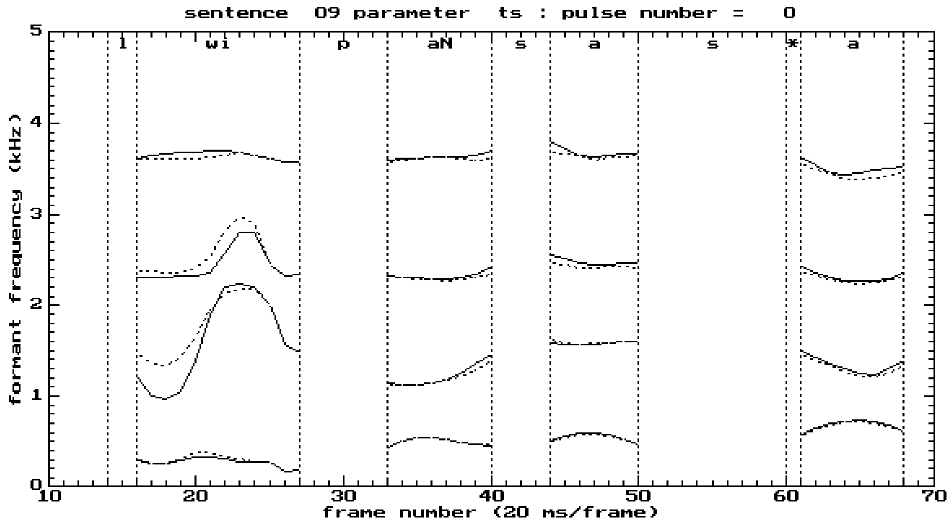


Figure 7.11 *Patrons de formants $\mathcal{F}_{ref}$ (trait plein) et $\mathcal{F}_{\mathrm{ts}}$ (pointillé). Nombre d'impulsions $m = 0$.*

7.3.4 Effet de la position de l'apex

Enfin, nous avons relevé une faible contribution du paramètre **tt** dans la production des voyelles et ce en dépit de l'importante activité de ce paramètre (*cf.* chapitre **??**). L'apex de la langue est plus sollicité pour la production des consonnes, en particulier les consonnes dentales.

Ces observations suggèrent que les articulateurs de la langue (**jw**, **tp** et **ts**) produisent un effet direct sur les corrélats acoustiques. Toutefois, les quatre paramètres de la langue ne sont pas toujours tous nécessaires pour obtenir les patrons acoustiques des voyelles. Pour une voyelle donnée, un paramètre peut jouer un rôle primaire ou secondaire. Par ailleurs, un trait distinctif peut être associé à plus d'un paramètre.

7.4 Calcul de l'"énergie" de l'erreur formantique

Dans cette section, nous examinons l'effet de la synthèse d'un seul paramètre articulatoire j de la langue sur la production des quatre premiers formants des voyelles pour les différentes phrases. Le but est d'en déduire la validité acoustique de la séquence d'impulsions de commande déterminée dans le domaine articulatoire. Pour cela, nous avons calculé l'"énergie" de l'erreur entre les fréquences de formants des patrons de référence $\mathcal{F}_{ref}$ et calculés $\mathcal{F}_{m,j}$ en fonction du nombre d'impulsions m.

7.4.1 Échelle de perception des patrons formantiques

La description acoustique et phonétique de la qualité des voyelles est traditionnellement exprimée en termes de patrons formantiques. La spécification de la fréquence du fondamental et des quatre premiers formants est considérée suffisante pour une synthèse correcte. Pour se rapprocher d'une résolution comparable à celle de l'oreille humaine, on est amené à convertir les fréquences du spectre harmonique en une grandeur z exprimée sur l'échelle non linéaire bark. La grandeur z correspond à une échelle naturelle qui est liée à l'anatomie de l'oreille interne (membrane basilaire). Les cellules ciliées, réparties régulièrement le long de la membrane basilaire, permettent à l'oreille d'apprécier la hauteur de l'onde sonore perçue appelée tonie. Ainsi, l'oreille divise toute la zone des fréquences audibles ($16Hz$ à $20kHz$) en 24 bandes critiques adjacentes dont les frontières ont été déterminées par des expériences psyco-acoustiques. Un décalage de fréquence d'une bande critique entraîne un déplacement de $1.3mm$ du lieu d'excitation principal de la membrane basilaire de l'oreille [?][?]. L'échelle que forme ces bandes critiques présente donc une relation très étroite avec la hauteur perçue. La relation entre f (Hz) et z ($Bark$) peut être approximée par la formule suivante [?] :

$$f = 650 \sinh(z/7) Hz, \tag{7.3}$$

ou bien par l'équation inverse

$$z = 7 \ln\{(f/650) + [(f/650)^2 + 1]^{1/2}\} Bark. \tag{7.4}$$

Notons qu'en dessous de $500Hz$, cette relation est linéaire, au-delà, elle est approximativement logarithmique.

7.4.2 Normalisation locale de l'erreur

Les fréquences des quatre premiers formants, $\{F_n, n = 1, ..., 4\}$, sont converties en $\{z(n), n = 1, ..., 4\}$, exprimées en bark à l'aide de la formule **??**. Notons

- $z_l(n)$ les fréquences en bark qui correspondent aux patrons $\mathcal{F}_{ref}$;
- $z_{l,j,m}(n)$ celles des patrons $\mathcal{F}_{j,m}$.

Les fréquences, notées $z_{l,j,m}(n)$, où l est la phrase considérée et j est le paramètre synthétisé, dépendent du nombre d'impulsions m.

L'énergie de l'erreur entre les fréquences des formants $z_{l,j,m}(n)$ et $z_l(n)$ est calculée de la manière suivante :

$$\mathbf{E}_{l,j,m} = \sum_{k=1}^{N} \sum_{n=1}^{4} (z_{l,j,m}(n,k) - z_l(n,k))^2, \tag{7.5}$$

où N est le nombre d'échantillons des segments vocaliques dans la phrase l.

Cette erreur est ensuite normalisée par l'erreur maximale $\mathbf{E}_{l,j,0}$ obtenue pour le nombre d'impulsions égal à zéro, comme suit :

$$\mathcal{E}_{l,j,m} = \frac{\mathbf{E}_{l,j,0}}{\mathbf{E}_{l,j,0}}, \tag{7.6}$$

La figure **??** montre un exemple de courbes d'erreur correspondant aux quatre paramètres. L'allure de ces courbes est proche de celle obtenue dans le domaine articulatoire (voir figures **??**, **??**). En effet, on observe une décroissance monotone dans les quatre cas. Néanmoins, ces courbes montrent des différences avec les courbes articulatoires. Notons en particulier, qu'elles ne présentent que deux paliers contrairement aux courbes d'erreur obtenues dans le domaine articulatoire qui en présentent trois. Le premier palier a une pente rapide qui coupe l'axe des abscisses en un point m_1 inférieur ou égal au nombre de syllabes dans une phrase. Le deuxième palier correspond à la saturation.

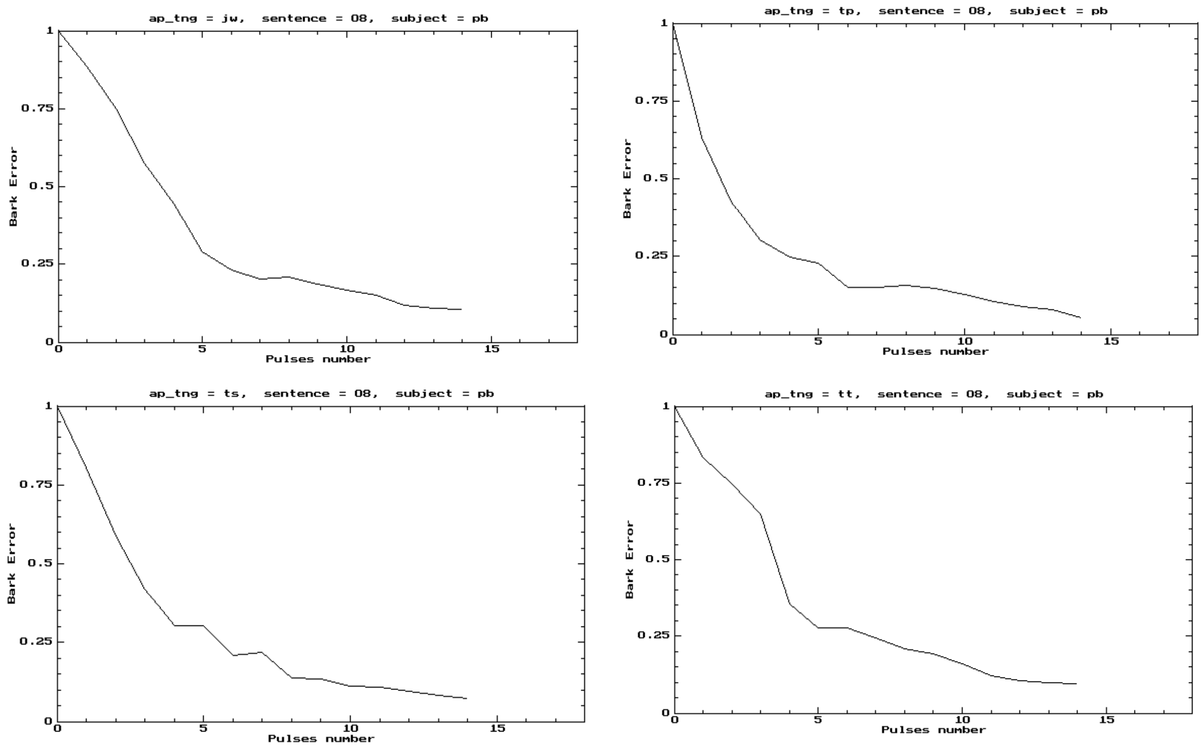


Figure 7.12 *Erreur des fréquences des formants : cas de la normalisation locale.*

La décroissance monotone de la courbe d'erreur, observée ci-dessus, signifie que les impulsions déterminées dans le domaine articulatoire par la minimisation de l'erreur entre les mouvements de référence et les mouvements synthétisés minimisent également l'erreur acoustique exprimée en terme de fréquences de formants. Ceci suggère une interaction directe entre les phénomènes articulatoires et acoustiques. Néanmoins, certaines courbes présentent des escaliers qui correspondent à des impulsions pour lesquelles l'erreur est maintenue constante. La question qu'on s'est posée est : ces impulsions peuvent-elles être omises sans dégrader la qualité de la synthèse ? Cette question est traitée plus loin.

7.4.3 Normalisation globale de l'erreur

Dans cette partie, nous cherchons à déterminer les paramètres prépondérants pour le calcul des patrons formantiques en estimant la contribution propre à chaque paramètre dans la synthèse acoustique. Pour cela, l'erreur $\mathbf{E}_{l,m,j}$ calculée dans la section **??** est normalisée par la même quantité D_l quelque soit le paramètre j,

$$D_l = \sum_{k=1}^{N} \sum_{n=1}^{4} (z_{l,0}(n,k) - z_l(n,k))^2. \tag{7.7}$$

Dans cette expression $z_{l,0}(n,k)$ est donné par les sept paramètres inhibés ($m = 0$). D_l est l'erreur commise lorsque l'ensemble des paramètres du conduit vocal sont nuls. Elle est indépendante du paramètre j considéré. Nous obtenons ainsi

$$\mathcal{E}_{l,m,j} = \frac{\mathbf{E}_{l,m,j}}{D_l}. \tag{7.8}$$

Le tableau **??** montre le pourcentage d'erreur de chaque paramètre lorsque $m = 0$. Notons que la somme des erreurs sur j pour une phrase donnée ($\sum_j \mathcal{E}_{l,0,j}$, $\{j =$ **jw**, **tp**, **ts** et **jw**$\}$) est souvent supérieure à 100%. Ce ci signifie que la reconstitution de la fonction d'aire ne se limite pas à une simple superposition des contributions partielles de chaque paramètre. Néanmoins, ces pourcentages d'erreur donnent une première estimation de la prépondérance de chaque paramètre.

Le paramètre **tp** présente les pourcentages les plus élevés. Il contribue activement à la production des corrélats acoustiques des voyelles. Le deuxième paramètre est **jw**. Son inhibition peut engendrer des erreurs importantes de l'ordre de 40%. En revanche, lorsque le paramètre **ts** ou **tt** est neutre, l'erreur maximale induite $\mathcal{E}_{l,0,j}$ est faible.

La figure **??** montre un exemple typique de l'allure des courbes de l'énergie des erreurs en fonction de m. Le nombre d'impulsions qui correspond à la saturation est nettement plus grand pour le paramètre **tp**. Dans le cas des paramètres **ts** et **tt**, trois impulsions suffisent pour atteindre la saturation. Par conséquent, le nombre minimum d'impulsions nécessaires m_j^* pour la synthèse des patrons acoustiques n'est pas le même pour tous les paramètres de

articulateur	**jw** *(%)*	**tp** *(%)*	**ts** (%)	**tt** (%)
phrase 01	22	50	22	6
02	24	47	22	16
03	13	27	33	14
08	14	79	16	15
09	35	65	19	15
15	22	73	25	16
17	26	49	21	12
18	38	58	16	13
24	40	84	29	33
28	35	81	09	23

Tableau 7.1 *Contribution de chaque paramètre dans la production des patrons des formants en pourcentage.*

la langue. Pour déterminer la valeur de m_j^*, nous allons choisir un critère de décision qui permet d'obtenir la sous séquence optimale d'impulsions S_j^*.

7.5 Détermination du nombre d'impulsions nécessaires pour chaque paramètre

7.5.1 Seuil de perception

Pour l'analyse et la synthèse des phonèmes, il n'est pas nécessaire d'avoir une précision des fréquences des formants excédant celle que l'oreille humaine peut détecter. Aussi, de nombreuses expériences ont été effectuées pour déterminer le seuil de perception des différences fréquencielles pour les formants des voyelles. Flanagan a déterminé le minimum de précision nécessaire pour l'analyse et la synthèse des formants F_1 et F_2 des voyelles (Flanagan 1955 [**?**], 1972 [**?**]). Il a entrepris des tests pour juger de la similitude de deux sons voisins pour différentes valeurs des fréquences de chaque formant séparément. Le seuil de la variation des fréquences des formants est atteint lorsque la différence

$$\Delta F_n = F_{ref,n} - F_n \tag{7.9}$$

est perçue. F_n est la fréquence du formant testé. Le seuil, noté DL_n (Difference Limen), est calculée comme suit :

$$DL_n = \frac{\Delta F_n}{F_{ref,n}} \tag{7.10}$$

Les résultats obtenus par les tests indiquent la nécessité d'une bonne résolution. En effet, le tableau **??** montre des valeurs de DL_n inférieures à 6% pour les différentes fréquences

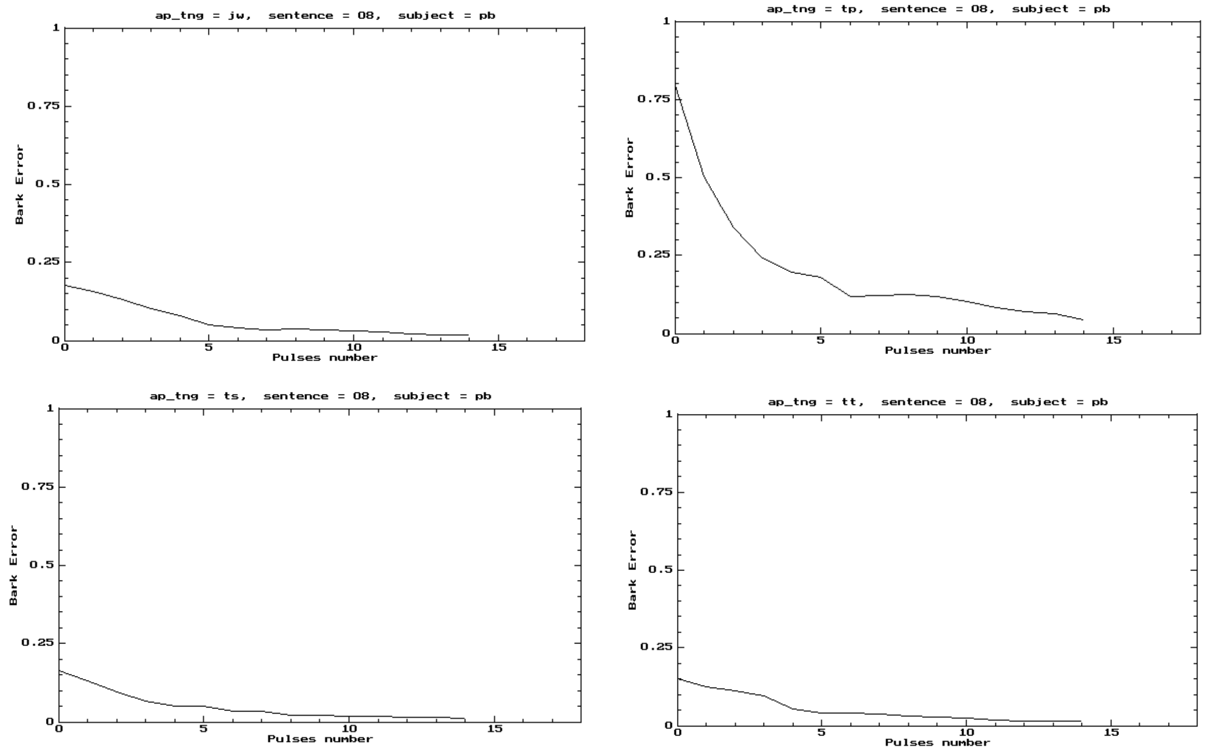


Figure 7.13 *Erreur des fréquences des formants : cas de la normalisation globale.*

du test. Ces valeurs dépendent de la distribution des formants dans le spectre. Elles sont d'autant plus faibles que les formants voisins de $F_{ref,n}$ sont proches.

Fréquence F_n *(Hz)*	300	500	700	1000	1500	2000
$-DL_n$ (%)	6	5	4	2	3	5
$+DL_n$ (%)	4	5	3	5	5	1

Tableau 7.2 *Seuils de perception pour différentes valeurs des fréquences des formants* F_1 *et* F_2 *d'après Flanagan,* 1955 [75].

7.5.2 Détermination du nombre d'impulsions minimal pour chaque paramètre

Pour une phrase l, le nombre minimum d'impulsions m_j^* est obtenu lorsque l'erreur de chacun des segments vocaliques est inférieure à DL_n, et ce pour les deux formants F_1 et F_2. En d'autres termes, l'erreur formantique doit vérifier l'inéquation suivante :

$$DL_n \geq \frac{\sum_{k=v_i}^{v_f} F_l(n,k) - \sum_{k=v_i}^{v_f} F_{l,j,m_j^*}(n,k)}{\sum_{k=v_i}^{v_f} F_l(n,k)} \qquad \forall n, \quad \forall V_i. \tag{7.11}$$

où V_i est la voyelle considérée. Elle est définie sur la durée $[v_i, v_f]$.

Le tableau **??** présentent les valeurs m_j^* obtenues pour les différentes phrases. D'une manière générale, $m_{\mathbf{jw}}^*$ et $m_{\mathbf{tp}}^*$ sont comparables au nombre de syllabes m_{syl} dans la phrase. Quant aux valeurs $m_{\mathbf{ts}}^*$ et $m_{\mathbf{tt}}^*$, elles sont nettement plus faibles que m_{syl}.

phrase	$m_{\mathbf{jw}}^*$	$m_{\mathbf{tp}}^*$	$m_{\mathbf{ts}}^*$	$m_{\mathbf{tt}}^*$	m_{syl}
01	5	5	1	1	5
02	3	4	1	3	5
03	3	3	2	2	4
08	5	7	3	1	6
09	5	5	1	2	5
15	5	6	2	2	5
17	5	3	3	2	5
18	5	6	0	0	5
24	4	6	2	3	5
28	6	8	1	3	5

Tableau 7.3 *Nombre minimum d'impulsions pour chaque paramètre articulatoire comparé au nombre de syllabe dans la phrase.*

La valeur m_j^* semble dépendre des traits distinctifs qui sont inhérents au paramètre. Puisque tous les paramètres ne contribuent pas de la même manière à la production des corrélats acoustiques, ainsi que nous l'avons vu plus haut, il est normal qu'ils ne soient pas contrôlés de la même manière. Les paramètres, qui ont un effet direct sur les corrélats acoustiques, requièrent nécessairement une plus grande précision pour la synthèse de leurs mouvements. Ils demandent donc un plus grand nombre d'impulsions.

Remarque

Des cas particuliers peuvent se présenter. Par exemple, lorsque l'amplitude du mouvement est relativement élevée pour des phonèmes successifs présentant le même état, le patron du mouvement montre un palier pour maintenir l'état. Dans ce cas, la synthèse nécessite un grand nombre d'impulsions. En effet, dans le cas de la phrase 28 ("*il fume son tabac*"), cinq impulsions sont nécessaires pour la synthèse du mouvement de **tp** correspondant à la séquence trisyllabiques "*son tabac*" ; les trois voyelles de cette séquence sont toutes postérieures.

7.5.3 Reconstitution des patrons formantiques à partir du nombre d'impulsions minimal

Les mouvements des quatre paramètres articulatoires, $s^*_{\mathbf{jw}}(n)$, $s^*_{\mathbf{tp}}(n)$, $s^*_{\mathbf{ts}}(n)$ et $s^*_{\mathbf{tt}}(n)$, sont synthétisés à partir des séquences d'impulsions S^*_j déterminées dans la section précédente. Les patrons des fréquences des formants des voyelles, $\mathcal{F}_{syn}$, sont ensuite reconstitués en utilisant les quatre paramètres synthétisés simultanément. Seuls les paramètres des lèvres correspondent aux paramètres de référence.

La variation des fréquences des formants calculée sur la durée des voyelles

$$\Delta F_{syn,n} = \frac{\sum_{k=v_i}^{v_f} F_l(n,k) - \sum_{k=v_i}^{v_f} F_{syn,l}(n,k)}{\sum_{k=v_i}^{v_f} F_l(n,k)}, \tag{7.12}$$

n'est pas toujours inférieure au seuil DL_n. C'est pourquoi, d'autres impulsions ont été rajoutées pour atteindre ce seuil. Le tableau **??** présente les nouvelles valeurs m^*_j nécessaires pour obtenir les fréquences de formants de $\mathcal{F}_{syn}$ avec une précision au dessous de DL_n.

phrase	$m^*_{\mathbf{jw}}$	$m^*_{\mathbf{tp}}$	$m^*_{\mathbf{ts}}$	$m^*_{\mathbf{tt}}$
01	5	6	3	1
02	4	5	3	4
03	3	5	4	3
08	5	7	3	2
09	5	6	1	2
15	5	6	4	3
17	5	3	3	2
18	5	7	3	3
24	4	7	2	3
28	6	8	1	4

Tableau 7.4 Nombre minimum d'impulsions pour la synthèse des patrons de formants.

La figure **??** montre les patrons temporels des quatre paramètres synthétisés par les nouvelles séquences d'impulsions S^*_j (*tab.* **??**) pour la phrase "*Louis pense à ça*". Le patron des formants correspondant est montré dans la figure **??**. En dépit des différences observées entre les mouvements synthétisés et ceux observés, notamment pour les paramètres **ts** et **tt**, le patron $\mathcal{F}_{syn}$ est correctement calculé. Les mêmes résultats sont observés pour la phrase 08, "*une phrase ambiguë*" (*fig.* **??**, **??**).

Comment sont organisées les commandes ? Tout d'abord, on remarque que les impulsions des articulateurs **jw** et **tp** sont générées de manière à reconstituer les mouvements accentués d'ouverture et de fermeture de la mâchoire d'une part et d'antériorité/postériorité de la langue d'autre part. Lorsque le mouvement de l'articulateur dépend du contexte, la commande va dépendre de l'influence contextuelle. Par exemple, la voyelle /i/ dans la phrase "louis pense à ça" nécessite une impulsion pour contrôler le mouvement de la mâchoire (*fig.* **??**-a). Or ce dernier n'est pas contrôlé lorsque /i/ est dans la phrase "une réponse ambiguë" (*fig.* **??**-a). Toutes les impulsions sont générées telles que les maxima et les minima des mouvements soient atteints. L'articulateur **tp** montre une importante activité. La postériorité/antériorité est un trait distinctif prépondérant pour la majorité des voyelles françaises. Aussi, les impulsions des séquences S^*_{tp} sont générées de manière à reconstituer les mouvements de l'articulateur **tp** avec une bonne précision . Il est souvent nécessaire de contrôler la consonne qui précède pour obtenir une meilleure précision sur les segments vocaliques. Ainsi, une impulsion est nécessaire sur le segment consonantique /s/ pour reproduire le mouvement de la voyelle /a/, ainsi que le montre la figure **??**-b). De la même façon, la réponse impulsionnelle de l'impulsion générée sur le segment /g/ est nécessaire pour la synthèse du mouvement de la voyelle finale /y/ (*fig.* **??**-b).

Par ailleurs, on note l'absence d'impulsions de contrôle sur les segments vocaliques des articulateurs **ts** et **tt**. Le ballonnement de la forme de la langue pour la voyelle /aN/ (*fig.* **??**-c) et son aplatissement pour la voyelle /a/ (*fig.* **??**-c) ne semblent pas nécessaires pour le calcul des fréquences des formants. Sur ces segments le mouvement de **ts** est identiquement nul. Deux impulsions suffisent pour avoir le mouvement effectif de l'apex pour les deux phrases. Ces deux impulsions ne permettent pas de reproduire l'intégralité des mouvements. En particulier, les mouvement des voyelles /i/, /y/ (*fig.* **??**-d) et /a/ (*fig.* **??**-d). Cependant, l'erreur $\Delta F_{syn,n}$ calculée sur ces segments est inférieure à DL_n. Ils semble que les mouvements observés ne sont pas tous contrôlés par le système moteur. Quelle est donc leur l'origine ? Nous pensons que ces mouvements ne sont probablement pas propres aux articulateurs. Ils sont générés soit par l'effet des contraintes physiologiques liées aux propriétés de la structure de la langue (continuité élastique) et résultent donc des mouvements des articulateurs primaires, soit par l'effet des contraintes neurologiques du système nerveux central qui contrôle la coordination de l'ensemble des articulateurs actifs comme une unité fonctionnelle. Dans les deux cas, ils jouent un rôle secondaire dans la production des segments phonétiques correspondants.

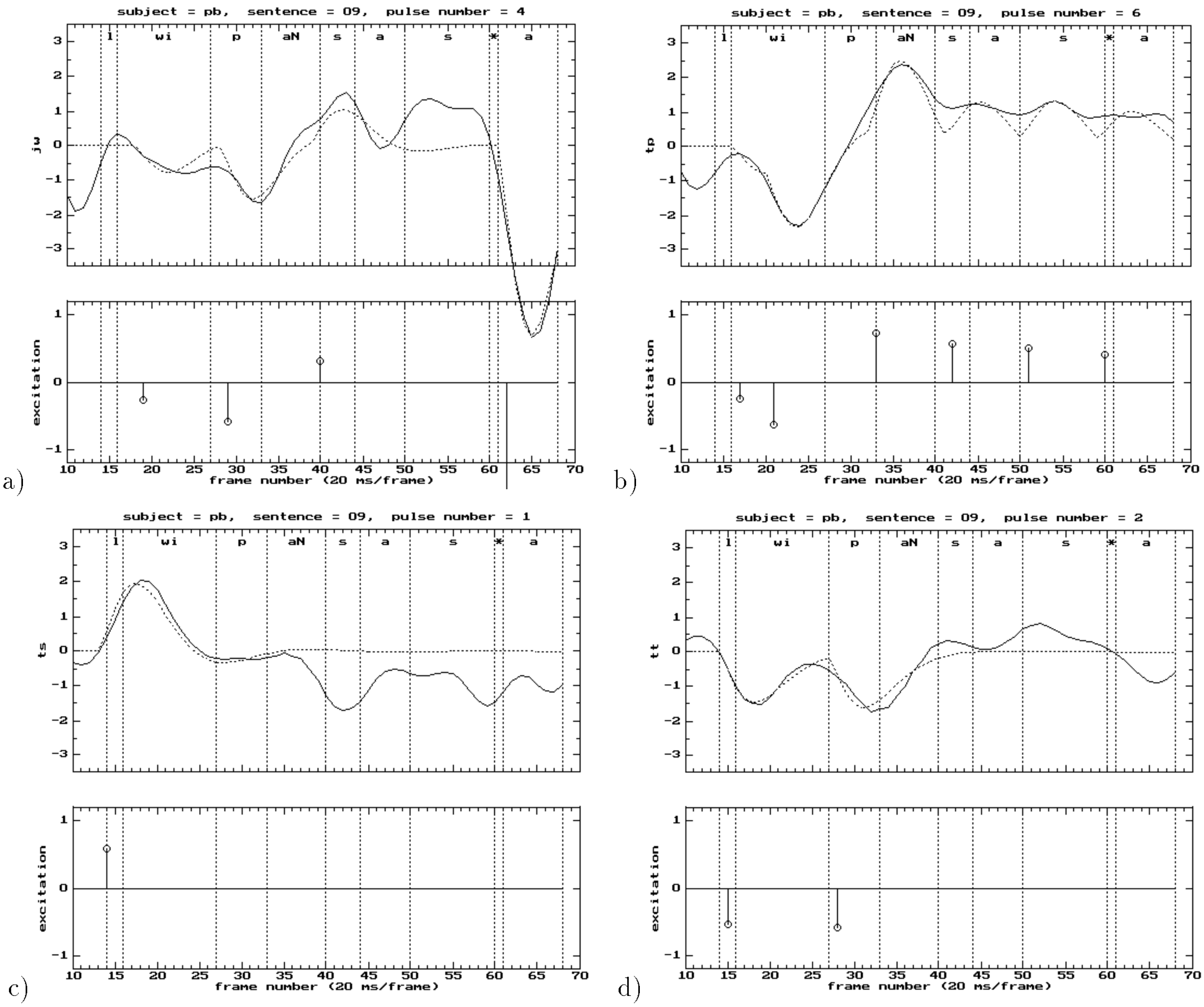


Figure 7.14 *Patrons temporels des paramètres* **jw**, **tp**, **ts** *et* **tt** *observés (trait plein) et synthétisés (pointillé) respectivement avec* $m^*_{\mathbf{jw}}$, $m^*_{\mathbf{tp}}$, $m^*_{\mathbf{ts}}$ *et* $m^*_{\mathbf{tt}}$ *pour la phrase* 09.

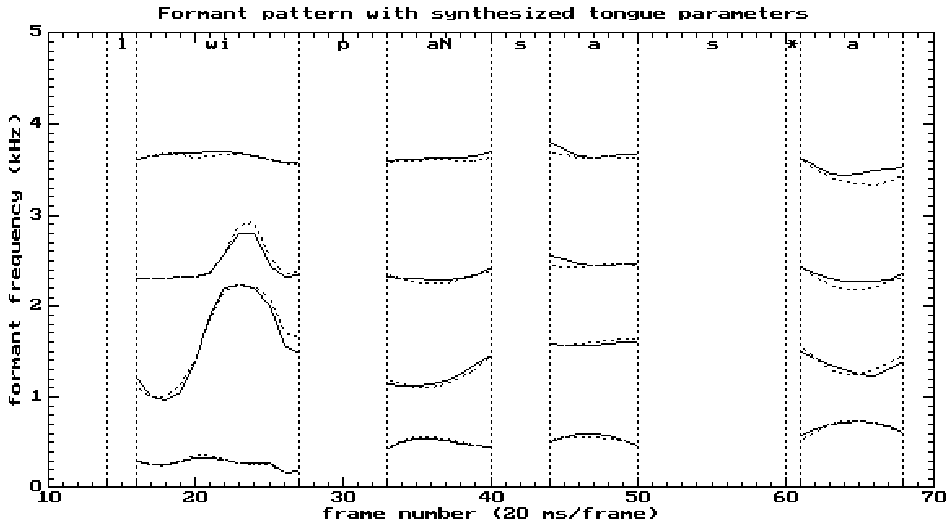


Figure 7.15 *Patron des fréquences des formants obtenu par les mouvements synthétisés* $s^*_{\mathrm{jw}}(n)$, $s^*_{\mathrm{tp}}(n)$, $s^*_{\mathrm{ts}}(n)$ *et* $s^*_{\mathrm{tt}}(n)$ *pour la phrase* 09.

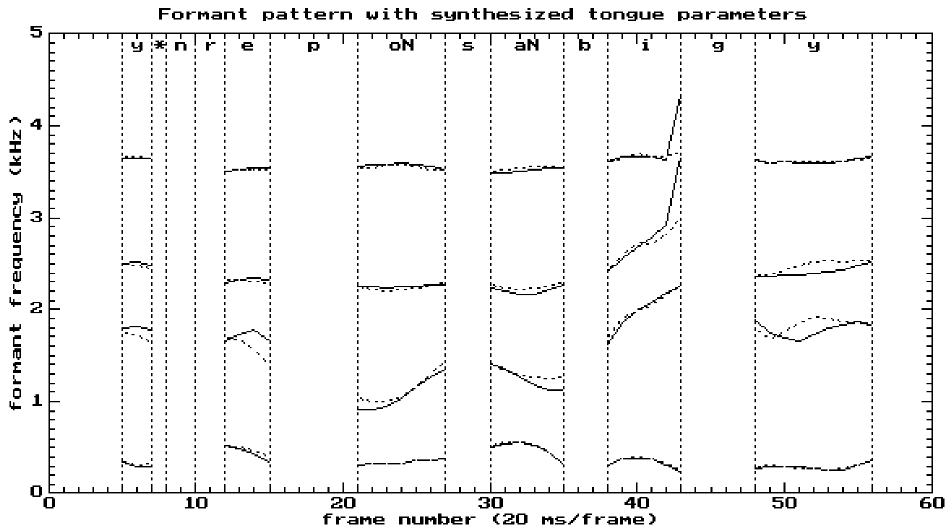


Figure 7.16 *Patron des fréquences des formants obtenu par les mouvements synthétisés* $s^*_{\mathrm{jw}}(n)$, $s^*_{\mathrm{tp}}(n)$, $s^*_{\mathrm{ts}}(n)$ *et* $s^*_{\mathrm{tt}}(n)$ *pour la phrase* 08.

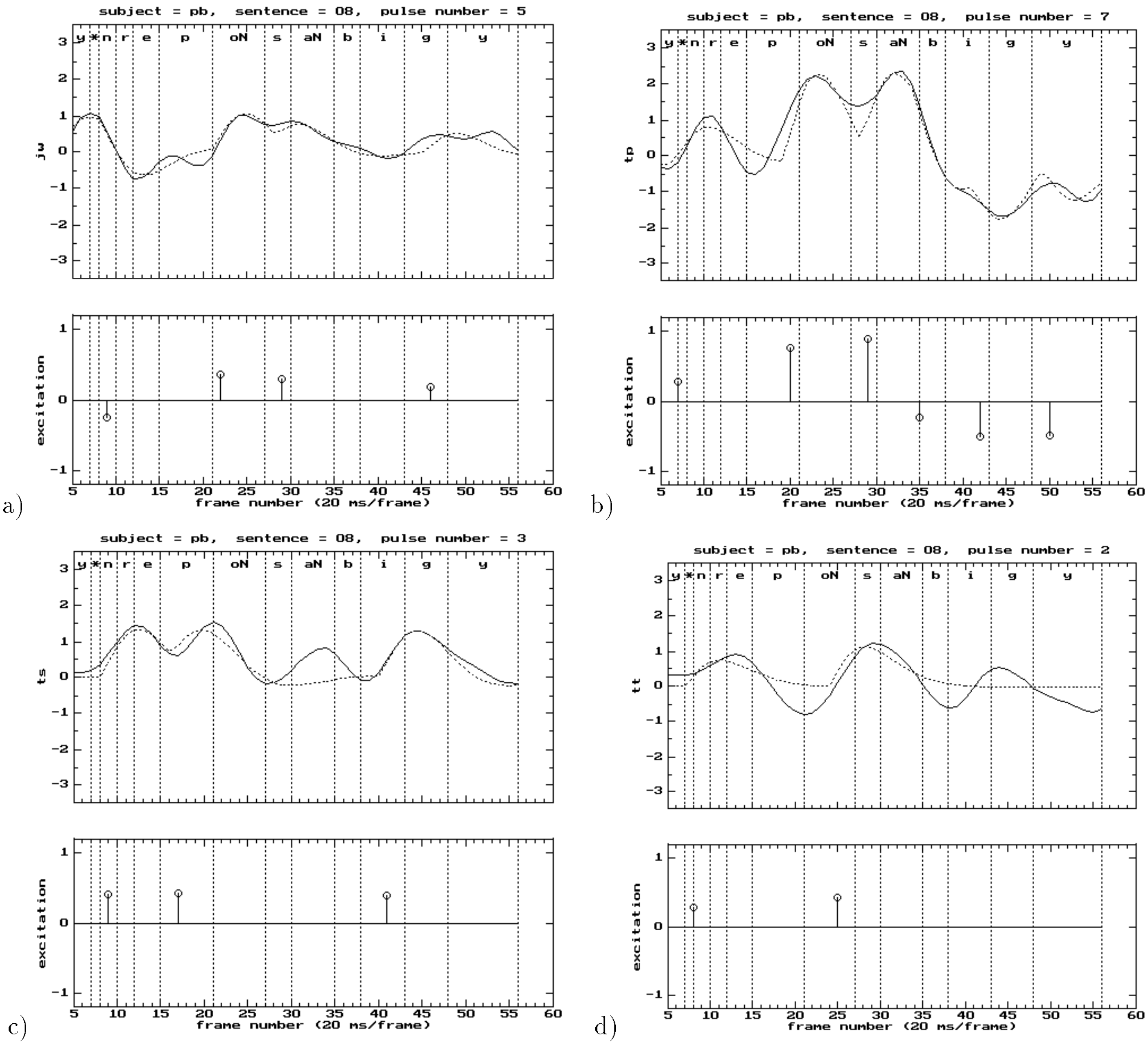


Figure 7.17 *Patrons temporels des paramètres* **jw**, **tp**, **ts** *et* **tt** *observés (trait plein) et synthétisés (pointillé) respectivement avec* $m^*_{\mathbf{jw}}$, $m^*_{\mathbf{tp}}$, $m^*_{\mathbf{ts}}$ *et* $m^*_{\mathbf{tt}}$ *pour la phrase* 08.

Chapitre 8

Conclusion

Au terme de ce travail qui s'intitule "Modélisation des mouvements articulatoires de la langue par la méthode de la LPC multi-impulsionnelle", nous voudrions situer les résultats auxquels nous sommes parvenus et en esquisser les perspectives.

Il convient toutefois de souligner que la quantité réduite des données articulatoires a imposé des limitations à nos résultats. En effet, notre travail a porté sur seulement 30 secondes de parole continue composée de courtes phrases françaises et de séquences CVCV.

Les signaux analysés correspondent à l'évolution temporelle (mouvement) des paramètres articulatoires de la langue extraits du modèle de Maeda. Ces paramètres sont la position de la mâchoire (**jw**), la position de la langue (**tp**), la forme de la langue (**ts**) et la position de l'apex (**tt**).

L'apport principal de notre travail se situe dans l'utilisation de techniques de traitement du signal pour l'analyse des articulateurs de la langue lors de la production de la parole. Pour ce faire, notre travail a pris pour référence de départ les travaux de Atal et Remde, 1986, [?] sur les modèles d'excitation pour le codage de la parole. En particulier, nous utilisons leur approche du modèle source-filtre. Dans ce modèle le problème consiste à séparer le signal de la commande musculaire (source) de la réponse impulsionnelle (filtre) à partir des observations du mouvement dynamique des différents articulateurs. En ce qui nous concerne le modèle source-filtre auquel nous sommes parvenus se présente comme suit :

- Les caractéristiques du filtre sont supposées invariantes. L'optimalité d'un tel filtre, exprimée en termes d'économie d'énergie, est la principale motivation de notre choix.

- Deux sous-systèmes placés en cascade constituent notre filtre. Le premier est d'ordre 1 donc visco-élastique. Le second d'ordre 2 présente une fréquence de résonance relativement basse, de l'ordre de $4Hz$. À l'encontre des interprétations jusqu'alors utilisées dans le cadre de l'étude des mouvements articulatoires chez l'homme, nous avons associé le sous-système du second ordre au système physiologique de la chaîne de production de la parole. Le sous-système d'ordre 1, quant à lui, est associé au système mécanique. Notre interprétation signifie que l'effet de la masse du système mécanique est négligée. Elle est soutenue par la nature visco-élastique des contractions musculaires ainsi que l'a décrit Hill, (1938 [?]), dans son modèle.

- La source correspond au paramètre de contrôle. Elle est représentée par un train d'impulsions. La réponse impulsionnelle du filtre correspond donc au mouvement élémentaire produit. Sa durée effective est comparable à la durée moyenne de la syllabe française. Ce résultat suggère que l'organisation articulatoire au niveau neuromusculaire implique la production d'un mouvement élémentaire dont la durée est comparable à celle de la syllabe, $180ms$.

- L'analyse par synthèse montre qu'un nombre fini d'impulsions est suffisant pour modéliser l'évolution temporelle des paramètres articulatoires. L'énergie de l'erreur entre le mouvement synthétisé et celui de référence décroît rapidement pour les premières impulsions. Si bien que sa valeur est inférieure à 30% lorsque le nombre des impusions est voisin du nombre de syllabes dans la phrase. Ce résultat s'explique par le fait que l'effet de la réponse impulsionnelle du modèle s'étend sur la durée moyenne de la syllabe française.

En entreprenant cette étude, notre intention était d'aboutir à une synthèse à partir du texte. C'est pourquoi nous nous sommes reportés, dans la deuxième partie de notre travail, au domaine acoustique pour l'analyse des résultats de la synthèse des mouvements articulatoires.

- L'erreur quadratique entre les fréquences des formants synthétisés et celles de références décroît rapidement pour les premières impulsions. D'une manière générale, l'importance de l'effet des premières impulsions est équivalente dans les deux domaines articulatoire et acoustique. Ceci suggère l'existance d'une relation directe entre les variations articulatoires et les conséquences acoustiques.

- Par ailleurs, nous avons montré que l'effet des différents articulateurs de la langue n'est pas similaire pour la production d'un son donné. Néanmoins, le nombre d'impulsions minimum, qui produit une différence inférieure au seuil de perception DL entre les fréquences des formants de références et synthétisés, est inférieur ou égal au nombre de syllabes dans la phrase. Il est, toutefois, plus élevé pour les paramètres **jw** et **tp** que pour **ts** et **tt**. Lorsque le nombre d'impulsions correspondant à un articulateur est élevé, comme dans le cas de **tp**, une importante activité de l'articulateur est observée. L'articulateur **tp** est, en effet, sollicité pour produire l'antériorité/postériorité observée dans la majorité des phonèmes. Notre analyse montre que les mouvements des paramètres **jw** et **tp** sont prépondérants pour la production des voyelles. Dans le cas du paramètre **ts**, les impulsions émises génèrent des mouvements qui correspondent à un ballonnement ou un aplatissement accentué de la langue. Seuls les phonèmes qui invoquent ces formes particulières de la langue nécessitent des impulsions. Le nombre des impulsions semble être étroitement lié au nombre de traits phonétiques inhérents au paramètre.

Pour conclure, nous retiendrons que la méthode source-filtre s'avère pertinente quant à la modélisation des mouvements articulatoires de la langue et à l'étude du contrôle moteur. C'est pourquoi, nous ne saurions terminer sans souligner les possibles perspectives que ce travail nous a permis de mettre en évidence, et qu'il serait intéressant de reprendre dans des travaux ultérieures :

- **Vérification des résultats obtenus** sur un ensemble de données plus large correspondants à des séquences CVCV constituées de l'ensemble des consonnes et voyelles françaises. Un tel apport de données permet de déterminer le degré d'activité de chacun des articularteurs de la langue pour la production des phonèmes, et de généraliser ainsi les résultats obtenus.

- **Synthèse des segments consonantiques des phrases,** les impulsions de commande sont déterminées de manière à ce que l'erreur de la synthèse des formants soit inférieure au seuil de perception DL sur les segments vocaliques des phrases. Les parties consonantiques ne sont pas traitées. Cette limitation présente une véritable entrave à l'étude globale de la commande déterminée par la modélisation MLPC. La synthèse des consonnes est donc une phase incontournable pour obtenir une analyse complète des impulsions de commande.

- **Modélisation vectorielle,** Dans ce travail, nous avons supposé que les paramètres articulatoires sont controlés indépendamment les uns des autres. Cette hypothèse est soutenue par l'indépendance statistique des paramètres articulatoires correspondant aux composantes principales de l'analyse factorielle appliquée aux contours de la langue. Or, l'analyse des séquences CVCV a montré une dépendance linéaire entre les positions extrèmes des articulateurs. Les paramètres sont donc globalement indépendants mais présentent une dépendance au niveau segmental. Un filtre multicanaux récursif permet de modéliser la dépendance observée.

Références

[1] S. Atal and J. R. Remde, "A new model of lpc excitation for producing natural-sounding speech at low bit rates," *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 614–617, 1982.

[2] B. C. Sonies, T. H. Shawker, T. E. Hall, L. H. Gerber, and S. B. Leighton, "Ultrasonic visualization of tongue motion during speech," *J. Acoust. Soc. Am.*, vol. 70, pp. 683–686, 1981.

[3] O. Fujimura, "Methods and goals of speech production research," *Language and Speech*, vol. 33, no. 3, pp. 195–258, 1990.

[4] J. S. Perkell, M. H. Cohen, M. A. Svirsky, and M. L. Matthies, "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *J. Acoust. Soc. Am.*, vol. 92, no. 6, pp. 3078–3096, 1992.

[5] N. Nguyen and A. Marchal, "Assessment of an electromagnetometric system for the investigation of articulatory movements in speech production," *J. Acoust. Soc. Am.*, vol. 91, no. 2, pp. 1152–1155, 1992.

[6] F. Bell-Berti and K. S. Harris, "Temporal patterns of coarticulation: Lip rounding," *J. Acoust. Soc. Am.*, vol. 71, pp. 449–454, 1982.

[7] T. Baer, P. J. Alfonso, and K. Honda, "Electromyography of the tongue muscles during vowels in /evp/ environment," *Annu. Bull. Res. Inst. Logop. Phoniatr. (Univ. Tokyo)*, vol. 22, pp. 7–19, 1988.

[8] A. Bothorel, P. Simon, F. Wioland, and J. Zerling, "Cinéradiographie des voyelles et consonnes du français," *Travaux de l'Institut de Phonétique de Strasbourg*, 1986.

[9] J. L. Flanagan, *Speech Analysis Synthesis and Perception.* Springer-Verlag Berlin. Heidelberg. New York. 2nd Edition, 1972.

[10] A. S. House and K. N. Stevens, "Analog study of the nazalization of vowels," *J. of Speech and Hearing Disorders*, vol. 21, pp. 218–232, 1956.

[11] S. Maeda, "The role of the sinus cavities in the production of nasal vowels," *ICASSP IEEE*, pp. 911–914, 1982.

[12] F. Lonchamp, *tudes sur la production de la parole. Les indices acoustiques de la nasalit vocalique. La modification du timbre par la fr quence fondamentale.* Nancy, Doctorat d' tat en Lettre et Sciences Humaines., 1988.

[13] J. Dang and K. Honda, "Morphological and acoustical analysis of the nasal and the paranasal cavities," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2088–2100, 1994.

[14] K. N. Stevens and A. S. House, "Development of a quantitative description of vowel articulation," *J. Acoust. Soc. Am.*, vol. 27, no. 3, pp. 484–493, 1955.

[15] B. Lindblom and J. Sundberg, "Acoustical consequences of lip, tongue, jaw, and larynx movement," *J. Acoust. Soc. Am.*, vol. 50, no. 4, pp. 1166–1179, 1971.

[16] C. Abry and L. J. Bo , "Laws for lips," *Speech Communication*, vol. 5, pp. 97–104, 1986.

[17] Y. Fukuda and S. Hiki, "Characteristics of the mouth shape in the production of japanese - stroboscopic observation," *J. Acoust. Soc. Am.*, vol. 3, no. 2, pp. 75–91, 1982.

[18] S. E. G. hman, "Numerical model of coarticulation," *J. Acoust. Soc. Am.*, vol. 41, no. 2, pp. 310–320, 1967.

[19] R. Harshman, P. Ladefoged, and L. Goldstein, "Factor analysis of tongue shapes," *J. Acoust. Soc. Am.*, vol. 62, no. 3, pp. 693–707, 1977.

[20] P. Mermelstein, "Articulatory model for the study of speech production," *J. Acoust. Soc. Am.*, vol. 53, no. 4, pp. 1070–1082, 1973.

[21] S. Maeda, "Mod lisation articulatoire du conduit vocal," *Journal de Physique*, vol. IV, no. 2, pp. 191–199, 1992.

[22] B. Lindblom, *Economy in Speech Gesture.* Springer-Verlag, New-York, In The Production of Speech, P.F. Mac Neilage ed., 217-245, 1983.

[23] G. Fant, "Relations between area functions and the acoustic signal," *Phonetica*, no. 37, pp. 55–86, 1980.

[24] J. M. Heinz, "Perturbation functions for the determination of vocal tract area eigenvalues," *STL-QPSR*, no. 1, pp. 1–14, 1967.

[25] P. Mermelstein, "Determination of the vocal tract shape from measured formant frequencies," *J. Acoust. Soc. Am.*, no. 41, pp. 1283–1294, 1967.

[26] M. R. Schroeder, "Determination of the geometry of the human vocal tract by acoustic measurements," *J. Acoust. Soc. Am.*, vol. 41, no. 4, pp. 1002–1010, 1967.

[27] G. Fant and S. Pauli, "Spatial characteristics of vocal tract wall impedance," *Proc. Speech Comm. Sem.*, vol. 2, no. 4, 1967.

[28] M. Mrayati, R. Carr , and B. Gu rin, "Distinctive regions and modes: A nw theory of speech production," *Speech Communication*, vol. 7, pp. 257–286, 1988.

[29] S. Chennoukh, *Mod lisation du conduit vocal en r gions distinctives. Synth se d'ensemble voyelle-voyelle et voyelle-consonne-voyelle.* cole Nationale Sup rieure des T l communications. Paris, Th se de Doctorat., 1995.

[30] C. Coker and O. Fujimura, "A model for specification of the vocal tract area function," *J. Acoust. Soc. Am.*, vol. 40, no. 5, p. 1271, 1966.

[31] J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schafer, and N. Umeda, "Synthetic voices for computers," *IEEE Spectrum*, vol. 7, pp. 22–45, 1970.

[32] C. Coker, "A model of articulatory dynamics and control," *Proc. of the IEEE*, vol. 64, pp. 452–460, 1966.

[33] J. Liljencrants, "Fourier series description of the tongue profile," *QPSR*, vol. 4, pp. 9–18, 1971.

[34] S. Maeda, "Une analyse statistique sur les positions de la langue : étude préliminaire sur les voyelles françaises.," *9ème Journée d'Etude sur la Parole.*, no. 2, pp. 307–314, 1978.

[35] K. Hashimoto and K. Sasaki, "On the relationship between the shape and position of tongue for vowels.," *J. of Phonetics*, vol. 10, pp. 141–149, 1982.

[36] K. Hashimoto and K. Sasaki, "Performance of articulatory parameters to express formant frequencies.," *J. Acoust. Soc. Japan.*, vol. 3, pp. 141–149, 1982.

[37] J. Perkell, "A physiologically-oriented model of tongue activity during speech production," *Ph. Diss., MIT, Cambridge*, 1974.

[38] S. Kiritani, K. Miyawaki, O. Fujimura, and J. E. Miller, "A computational model of the tongue," *Ann. Bull. RILP*, no. 10, pp. 243–251, 1976.

[39] Y. Kakita and O. Fujimura, "Computational model of the tongue : A revised version," *J. Acoust. Soc. Am.*, vol. 62, no. S1, S15 (A), pp. 310–320, 1977.

[40] R. Wilhelms-Tricarico, "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3085–3098, 1995.

[41] S. Maeda, "Un mod le articulatoire de la langue avec des composantes lin aires.," *10ème Journée d'Etude sur la Parole.*, pp. 1–9, 1979.

[42] S. Maeda, "Articulatory co-ordination and its neurobiological aspects: An essay," *Proceedings of the 13-th ICPhS*, vol. 2, pp. 76–83, 1995.

[43] R. A. Schmidt, *Motor Control and Learning : A Behavioral Emphasis.* Human Kinetics Publishers, Inc., 1988.

[44] D. G. Asatryan and A. G. Feldman, "Functional tuning of the nervous system with control of movement or maintenance of a steady posture, i. mechanographic analysis of the work of the limb on execution of a postural task," *Biophysics*, vol. 10, pp. 925–935, 1965.

[45] J. M. Winters and L. Stark, "Muscle models: What is gained and what is lost by varying model complexity," *Biol. Cybern.*, vol. 55, pp. 403–420, 1987.

[46] K. Akazawa, M. Fujii, and T. Kasai, "Analysis of muscular contraction mechanism by viscoelastic model," *Technology Reports of the Osaka University*, vol. 19, no. 902, pp. 577–595, 1969.

[47] H. E. Huxley, *Structural Organization and the Contraction Mechanism in Striated Muscle.* New York: Colombia University Press, In S. Devons (Ed), Biology and the physical sciences, 1969.

[48] P. F. MacNeilage, "Preliminaries to the study of single motor unit activity in speech musculature," *J. of Phonetics.*, vol. 1, pp. 55–71, 1973.

[49] A. V. Hill, "The heat of shortening and the dynamic constants of muscle," *Proc. Roy. Soc. Brit.*, vol. 125, pp. 136–136, 1938.

[50] M. Yasuhara, "Identification and decomposition of fast handwriting system," *IEEE Transaction on Circuits and Systems*, vol. 30, no. 11, pp. 828–832, 1983.

[51] R. B. Stein, F. W. J. Cody, and C. Capaday, "The trajectory of human wrist movements," *J. of Neurophysiology.*, vol. 59, pp. 1814–1829, 1988.

[52] R. Baratta and M. Solomonow, "The dynamic response model of nine different skeletal muscles," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 3, pp. 243–251, 1990.

[53] J. Bobet, R. B. Stein, and M. N. Oguztoreli, "A linear time-varying model of force generation in skeletal muscle," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 10, pp. 1000–1006, 1993.

[54] A. E. Chapman and P. T. Harrower, "Linear approximations of muscle mechanics in isometric contraction," *Biol. Cybernetics*, vol. 27, pp. 1–7, 1977.

[55] Y. Sonoda and S. Kiritani, "Analysis of tongue point movements by linear second-order system model," *Ann. Bull. RILP*, no. 10, pp. 29–35, 1976.

[56] C. P. Browman and L. Goldstein, "Dynamic modeling of phonetic structure," *Phonetic Linguistics, edited by V. A. Fromkin (Academic Press, New York)*, pp. 35–53, 1985.

[57] C. L. Smith, C. P. Browman, R. S. McGowan, and B. Kay, "Extracting dynamic parameters from speech movement data," *J. Acoust. Soc. Am.*, vol. 93, no. 3, pp. 1580–1588, 1993.

[58] D. J. Ostry and K. G. Munhall, "Control of rate and duration of speech movements," *J. Acoust. Soc. Am.*, vol. 77, pp. 640–648, 1985.

[59] P. Perrier, C. Abry, and E. Keller, "Vers une modélisation des mouvements du dos de la langue," *J. Acoustique*, vol. 2, pp. 69–77, 1989.

[60] J. D. Cooke, *The Organization of Simple, Skilled Movements, in Tutorials in Motor Behavior.* Eds. G. E. Stelmach and J. Requin (North-Holland Publishing Company), 1980.

[61] B. Lindblom, "Vowel duration and a model of lip mandibule coordination," *STL-QPSR*, no. 4, pp. 1–29, 1967.

[62] W. Nelson, "Physical principles for economies of skilled movements," *Biological Cybernetics*, vol. 46, pp. 135–147, 1983.

[63] P. Duvaut, *Traitement du signal : concepts et applications.* dition Herm s, 1991.

[64] W. W. Cooley and P. R. Lohnes, *Speech Analysis Synthesis and Perception.* John Wiley & Sons, Inc. Printed in United States of America, 1971.

[65] J. Makhoul, "Linear prediction : A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[66] S. Kay and L. Marple, "Spectrum analysis : A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1320–1419, 1981.

[67] F. Makhoul, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. AC-19, no. 6, pp. 716–723, 1974.

[68] O. Fujimura, "Bilabial stop and nasal consonants : A motion picture study and its acoustical implications," *Journal of Speech and Hearing Research*, vol. 4, no. 3, pp. 233–247, 1961.

[69] K. Ishizaka, J. C. French, and J. L. Flanagan, "Direct determination of vocal tract wall impedance," *IEEE. Transactions on Acoustique, Speech and Signal Processing*, vol. ASSP-23, no. 4, pp. 370–373, 1975.

[70] S. Atal, "High-quality speech at low bit rates: Multi-pulse and stochastically excited linear predictive coders," *Proc. ICASSP*, pp. 1681–1684, 1986.

[71] J. M. Heinz and K. N. Stevens, "On the derivation of area functions and acoustic spectra from cineradiographic films of speech," *J. Acoust. Soc. Am.*, vol. 36, p. 1037, 1955.

[72] S. Maeda, "A digital simulation methode of the vocal tract system," *Speech Communication*, vol. 1, pp. 199–229, 1982.

[73] Calliope, *La parole et son traitement automatique.* Masson. Collection Technique et Scientifique des Télécommunications CNET, 1989.

[74] E. Zwicker and R. Feldtkeller, *Psychoacoustique.* Masson. Collection Technique et Scientifique des T l communications CNET, 1981.

[75] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the humain ear," *J. Acoust. Soc. Am.*, vol. 66, no. 6, pp. 1647–1652, 1979.

[76] J. L. Flanagan, "A difference limen for vowel formant frequency," *J. Acoust. Soc. Am.*, vol. 27, no. 3, 1955.