# Restructuration rythmique de la parole produite sous feedback auditif retardé

Jinyu LI[1] Leonardo Lancia[1]
(1) Laboratoire de Phonétique et Phonologie, 19 rue des bernardins, Paris, France
jinyu.li@sorbonne-nouvelle.fr, leonardo.lancia@sorbonne-nouvelle.fr

## RÉSUMÉ

Sous *delayed auditory feedback* (DAF) les durées des gestes de parole deviennent moins uniformes, alors que les cibles atteintes deviennent plus uniformes. Nous proposons que ces effets ne soient pas simplement le résultat des ralentissements locaux, permettant d'atteindre plus précisément les cibles acoustiques, mais qu'ils soient dus à une réorganisation des structures coordinatrices sous-jacentes au rythme de la parole. 10 femmes françaises ont répété trois phrases avec trois degrés de DAF. Nous avons analysé comment le DAF affecte le degré de coordination temporelle entre les modulations d'amplitude observées à l'échelle temporelle des syllabes et à l'échelle de la proéminence supra-syllabique dans cinq bandes spectrales séparées. Conformément à notre hypothèse, cette coordination est positivement corrélée au retard du retour auditif dans la quatrième bande (1719 ~ 3913Hz), liée principalement aux gestes vocaliques. En d'autres termes, la coordination temporelle entre patrons rythmiques syllabiques et supra-syllabiques dépend du retard auditif.

## ABSTRACT

**Rhythmic restructuration of speech produced under delayed auditory feedback.**
Under delayed auditory feedback (DAF) the duration of speech gestures becomes less uniform, while the acoustic configurations achieved themselves become more uniform. We hypothesize that these changes are not simply the consequence of local slowing down, permitting to reach the targets more accurately, but are associated with the reorganization of the coordinative structures underlying speech rhythm. We recorded 10 French females repeating three French sentences with three degrees of DAF and we analyzed the effects of DAF on the degree of temporal coordination between amplitude modulations observed at the time scales of syllable and at that of supra-syllabic prominence production in five separated spectral bands. Consistently with our hypothesis, this coordination gets stronger with increasing DAF in the fourth bands (1719 ~ 3913Hz), mainly related to vocalic gestures. In other words, the temporal coordination between syllabic and supra-syllabic rhythmic patterns depends on the auditory feedback delay.

MOTS-CLÉS : rythme de la parole, feedback auditif retardé, modulation d'amplitude, bandes spectrales, structure temporelle, contrôle sensorimoteur

KEYWORDS: speech rhythm, delayed auditory feedback, amplitude modulation, spectral bands, temporal structure, sensorimotor control

# 1 Introduction

Grâce à la flexibilité du contrôle moteur de la parole, les locuteurs sont capables d'adapter leur comportement à toutes sortes de facteurs contextuels, y compris ceux qui dépendent de leurs interlocuteurs ou des conditions environnementales. La flexibilité observée dans la dimension temporelle, c'est-à-dire dans les durées des gestes de parole, est particulièrement intéressante. En effet, cette flexibilité doit coexister avec les contraintes qui règlent très précisément la coordination entre les différents paramètres impliqués dans la production de séquences complexes de gestes de parole. Une méthode efficace pour étudier le mécanisme de contrôle temporel dans le traitement de la parole est le paradigme du *Delayed Auditory Feedback* (DAF) qui entraîne les locuteurs à entendre leur voix plus tard que prévu. Les locuteurs peuvent être perturbés de différentes manières par le DAF (voir Sasisekaran, 2012 pour une revue). L'une des stratégies les plus courantes utilisées par les locuteurs en réponse à cette perturbation consiste à allonger les syllabes, en particulier dans leurs parties vocaliques (e.g., Kalveram et Jäncke, 1989). Cependant, le degré de flexibilité temporelle dans la production de la parole sous DAF est modulé par des facteurs prosodiques tels que la position des syllabes (ou segments) dans la structure prosodique (Kalveram et Jäncke, 1989 ; Li et Lancia, 2022). Autrement dit, la flexibilité temporelle de la production de la parole dépend de caractéristiques qui varient sur des échelles temporelles différentes, comme le sont la nature consonantique ou vocalique des sons ou le degré d'accentuation de la syllabe qui les contient. Ces échelles temporelles, par ailleurs, définissent les deux domaines rythmiques plus saillants dans la perception de la parole, c'est-à-dire le rythme syllabique et le rythme constituant les patrons de proéminence supra-syllabique (i.e., ces évènements prosodiques, qui dans une langue donnée, rendent une syllabe plus saillante que celles qui l'entourent). Il a été montré que les différences rythmiques entre les énoncés d'une même langue ou des langues différentes peuvent être expliquées par le degré de coordination temporelle entre les évènements syllabiques et supra-syllabiques (Leong et Goswami, 2015 ; Lancia, et al., 2019). Selon plusieurs auteurs (e.g., Cummins et Port, 1998 ; Lancia et al., 2019 et références citées), la mise en place des patrons rythmiques hiérarchiquement organisés par des relations de coordination permet de structurer les processus phonétiques se déroulant sur les échelles temporelles différentes et joue donc un rôle central dans le contrôle moteur de la parole.

En utilisant le paradigme du DAF, nous avons aussi montré (Lancia et al., 2020) que lorsque le retard du retour auditif augmentait, les durées des gestes de parole devenaient moins uniformes, alors que les cibles atteintes devenaient plus uniformes. D'un côté, ces résultats ne sont pas surprenants. Nous nous attendons à ce que les ralentissements induits par le DAF permettent d'atteindre plus précisément les cibles phonétiques, car la durée des gestes articulatoires augmente. Cependant il est aussi possible que, pour faire face aux perturbations du retour auditif, les locuteurs réorganisent les structures coordinatrices sous-jacentes au rythme de la parole de manière à simplifier les patrons moteurs. Si c'est le cas, nous nous attendrions que dans un paradigme de DAF, l'augmentation du retour auditif soit corrélé à l'augmentation du degré de coordination entre le rythme syllabique et le rythme supra-syllabique. De plus, du moment que la réponse au DAF varie selon la nature des sons et leur position dans la structure prosodique, il devrait s'en suivre que les différentes parties de l'énoncé ne participent pas de la même manière à la réorganisation rythmique sous DAF. Afin de tester nos hypothèses, nous avons conduit une expérience de production sous DAF (voir section 2), et nous avons analysé le degré de coordination temporelle entre les modulations d'amplitude observées à l'échelle temporelle des syllabes et à l'échelle de la proéminence supra-syllabique dans cinq bandes spectrales séparées (voir section 3).

# 2 Expérimentation

Les données de cette étude ont été obtenues auprès de 10 locutrices francophones. Toutes les participantes étaient étudiantes universitaires ayant le français comme langue maternelle, et n'ayant pas de problèmes d'audition et de parole connus. Trois phrases de cinq syllabes, se différenciant par rapport à leur complexité, ont été créées. Les trois phrases se composent respectivement et principalement de syllabes dont la structure est 1) CV : Vivien vit le vin ; 2) CVC : Jacqueline gère le jour ; 3) CCV(C) : Bradley brise le bras. L'expérience s'est effectuée dans une chambre sourde. Les participantes, assises devant l'ordinateur donnant les consignes, portaient un micro et un casque à travers lequel elles percevaient leur voix, soit manipulée, soit sans modifications. La manipulation a été faite en MATLAB en utilisant le programme AUDAPTER (Cai et al, 2008). L'expérience a commencé par une phase de familiarisation au DAF, dans laquelle nous avons demandé aux participantes de lire le court texte français « La bise et le soleil » avec un DAF de 120ms. Pendant cette phase de familiarisation, le volume du feedback auditif dans le casque a été ajusté afin de minimiser la perception par les participantes de leur propre voix en dehors du casque. Ensuite, dans la phase de test, nous avons demandé aux participantes de répéter les trois phrases avec une vitesse d'élocution confortable. Les essais expérimentaux (chacun consistant en une répétition des trois phrases dans un ordre aléatoire différent) ont été organisés en blocs de six. Au total, le test s'est composé de 16 blocs et donc chaque phrase a été répétée 96 fois par chaque locutrice. Les essais composant le premier bloc ont été considérés comme des essais de contrôle, puisqu'il n'y avait aucune altération du feedback auditif. Au cours de chaque essai de chaque bloc suivant, la valeur de DAF a été choisie au hasard parmi 0, 60 et 120ms. Chaque valeur de DAF était donc appliquée deux fois dans chaque bloc.

# 3 Méthodes d'analyse

Nous avons d'abord identifié manuellement les frontières des phrases dans PRAAT. Pour analyser la coordination entre la production des syllabes et celle de la proéminence supra-syllabiques nous avons procédé en deux étapes. D'abord nous avons extrait de chaque enregistrement et de chaque bande spectrale considéré deux signaux qui dépendent des deux rythmes (syllabique et supra-syllabique). Ensuite nous avons appliqué une mesure de la coordination temporelle entre les signaux extraits.

## 3.1 Extraction des rythmes syllabiques et supra-syllabiques

Dans plusieurs études, les relations entre les rythmes syllabiques et supra-syllabiques ont été étudiées en analysant leurs contributions à la modulation d'amplitude du signal acoustique. Plus précisément, l'amplitude de la forme d'onde est filtrée par des filtres passe-bande afin de capturer les changements relativement lents induits par la production de proéminences supra-syllabiques et les changements plus rapides dus à la production de syllabes. Dans cette approche, tous les sons sont considérés comme contribuant de manière égale à la définition des patrons rythmiques, quelles que soient leurs caractéristiques spectrales. Comme l'a noté Leong (2013), cela peut être problématique car différents sons contribuent différemment aux caractéristiques rythmiques de la production de la parole (Cummins et Port, 1998 ; Patel et al., 1999). Pour faire face à ce problème, Leong et ses collaborateurs proposent de partitionner le spectre en un petit nombre de sous-bandes, chacune capturant les canaux spectraux adjacents dont les modulations d'énergie sont corrélées (ex : Leong, 2013 ; Leong et Goswami, 2015), et d'analyser séparément les modulations d'amplitude dans chaque sous bande. Ensuite, le signal acoustique est filtré par plusieurs filtres passe-bande pour capturer l'énergie présente dans chaque bande. La modulation de l'amplitude de chaque signal filtré est calculée au moyen de la

transformation de Hilbert. Enfin, les patrons rythmiques syllabiques et supra-syllabiques (ci-après syllAM et stressAM) sont extraits de chaque signal de modulation d'amplitude au moyen de filtres passe-bande (frontières : 2,5Hz et 12Hz pour syllAM et 0,9Hz et 2,5Hz pour stressAM).

Afin de déterminer le nombre et les frontières des sous-bandes contenant les canaux spectraux adjacents et corrélés, chaque signal a été d'abord soumis à un banc de filtres ERB. Chacun des 96 signaux analysés dans notre étude correspond à un enregistrement contenant une répétition des trois phrases, produite par une des locutrices. Ensuite, les représentations spectrales obtenues subissaient une procédure de réduction de dimensions basée sur l'Analyse en Composantes Principales (ACP). Plus précisément, en suivant l'approche de Leong (2013), chaque signal a été d'abord sous-échantillonné à 16000 Hz, pré-emphatisé et soumis à un banc de filtres composé de 28 canaux ERB répartis entre 100Hz et 7250Hz. La modulation d'énergie au fil du temps de chaque canal a été obtenue en calculant l'amplitude de la transformation de Hilbert. Les 28 séries temporelles correspondant aux modulations d'amplitude ont été échantillonnées à 1000 Hz, filtrées en passe-bas (fréquence de coupure : 40 Hz) et soumises à l'ACP. Cela implique que chaque analyse a été conduite sur quelques milliers de configurations des amplitudes spectrales par composantes principales (CPs). Après avoir ordonné les CPs par la quantité de variance expliquée, nous avons retenu les coordonnées des CPs (représentant les directions de chaque CP dans l'espace des 28 canaux ERB) et les ratios de variance expliquée (les ratios entre la variance expliquée par chaque CP et la variance totale observée) indexées par rapport à l'ordre des CPs. En calculant le ratio moyen de variance expliquée pour chaque CP sur l'ensemble des enregistrements, il était possible de sélectionner le nombre minimum de CPs qui permettait en moyenne d'expliquer plus de 65% de la variance spectrale totale observée (ce qui dans notre cas a permis de sélectionner les quatre premiers CPs). Les valeurs absolues des cordonnées ont été calculées pour chaque CP sélectionnée, et la moyenne des cordonnées ainsi rectifiées a été calculée sur les différents enregistrements pour chaque CP (comme indiqué en Figure1).

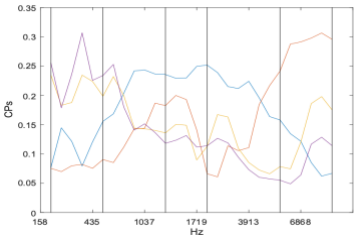


FIGURE 1 - Coordonnées rectifiées moyennes des 4 premières CPs. Les lignes verticales indiquent les limites des bandes spectrales individuées (voir texte).

L'algorithme proposé par Leong (2013) pour identifier les régions des sous-bandes spectrales est basé sur l'hypothèse qu'à l'intérieur de chacune de ces régions, les valeurs absolues des coordonnées des CPs ont tendance à augmenter, tandis qu'à ses frontières, des creux devraient être observés. La procédure est ainsi basée sur l'identification de pics et de creux dans les coordonnées rectifiées moyennes des CPs en Figure1. Afin de réduire les effets des fluctuations aléatoires dans la forme des coordonnées des CPs sur l'identification des sous-bandes spectrales, nous avons ajouté à l'algorithme original un test de significativité visant à identifier les creux parasites dans les coordonnées des CPs. Si $C_i$ est l'emplacement d'un creux observé dans la courbe des coordonnées moyennes d'un des CPs et si $P_{i,1}$ et $P_{i,2}$ sont les emplacements des pics qui le précèdent et qui le suivent sur la même courbe, ce creux sera retenu seulement si les valeurs absolues des coordonnées des CPs obtenues à partir des différents enregistrements correspondant à $C_i$ sont simultanément plus petites que les valeurs

correspondant à $P_{i,1}$ et $P_{i,2}$. Les comparaisons ont été effectuées au moyen de t-tests unilatéraux appariés, avec $\alpha$=0,05. L'élimination des creux parasites a créé des intervalles délimités par deux creux et contenant deux pics dont seul le plus élevé a été retenu. Selon la procédure proposée par Leong (2013), les frontières des bandes spectrales sont déterminées par l'inspection visuelle des pics et des creux restants dans les coordonnées moyennes des CPs. Plus précisément, une sous-bande spectrale doit contenir au moins deux pics et deux bandes spectrales doivent être séparées par au moins deux creux. Dans notre travail, ces deux critères ont été formalisés en une procédure automatique : D'abord nous avons identifié tous les intervalles qui contenaient au moins deux pics. Ensuite, pour chaque frontière séparant deux intervalles consécutifs, nous avons identifié le dernier pic du premier intervalle et le premier pic du deuxième et nous avons vérifié s'il y avait au moins deux creux entre deux pics. Nous avons calculé la distance entre le pic de chaque intervalle qui ne satisfaisait pas ces deux conditions et le pic de chaque intervalle qui les satisfaisait. À la suite de cette étape, nous avons intégré chacun des intervalles défaillants à l'intervalle dont le pic est le plus proche. Enfin, afin de déterminer les frontières entre les bandes spectrales consécutives, nous avons calculé la valeur moyenne entre les bords adjacents. La procédure décrite a permis d'identifier cinq bandes spectrales dont les frontières sont 158, 435, 1037, 1719, 3913, 7250 Hz (voir Figure2).

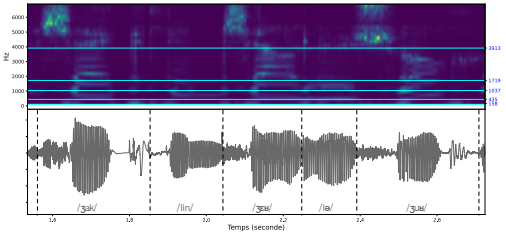


FIGURE 2 - Spectrogramme d'une phrase entière (« *Jacqueline gère le jour* »). Les lignes horizontales séparent les cinq bandes spectrales. Les valeurs des frontières des bandes (en *Hz*) sont indiquées à droite (158, 435, 1037, 1719, 3913, 7250).

## 3.2 Calcul de l'indice de coordination

Comme proposé par Leong, la coordination entre $syllAM$ et $stressAM$ dans chaque sous-bande est basée sur les valeurs de phase instantanées des signaux $syllAM$ et $stressAM$ et sur le calcul du *Phase Locking Value* (PLV) dans différentes fenêtres temporelles. Plus précisément, nous avons calculé la phase instantanée de chaque signal à chaque échantillon, ce qui permettait d'obtenir la phase relative généralisée entre les deux signaux (i.e., la différence entre les phases instantanées des deux signaux, normalisée par rapport aux différences de fréquence entre les deux signaux). L'index PLV est une mesure de la variabilité de la relation temporelle représentée par la phase relative généralisée dans un intervalle de temps considéré. La phase instantanée a été calculée comme l'angle de la transformation de Hilbert de chaque modulation d'amplitude. Cependant, afin de réduire les artefacts habituellement observés dans l'application de cette approche, les signaux oscillatoires ont été soumis à une procédure de normalisation de l'amplitude inspirée de celle décrite par Huang et al. (2009) et forçant les signaux à osciller entre -1 et 1 à chaque cycle. L'indice de *phase locking* a été obtenu en calculant le premier moment de la transformation de Fourier de la distribution des valeurs de la différence de phase généralisée observées entre $syllAM$ et $stressAM$ dans chaque fenêtre temporelle : $PLV = < cos\Delta\phi(t) >^2 + < sin\Delta\phi(t) >^2$, où les crochets représentent des moyennes ; $\Delta\phi(t) = n \times \phi_{syllAM}(t) - m \times \phi_{stressAM}(t)$ ; $\phi_{syllAM(t)}$ et $\phi_{stressAM}(t)$ sont la phase à l'instant $t$ de $syllAM$ et $stressAM$ respectivement ; et $m$ et $n$ sont deux entiers tels que, si $\omega_{syllAM}$ et $\omega_{stressAM}$ sont les

fréquences des syllAM et stressAM, alors $n \times \omega_{syllAM} = m \times \omega_{stressAM}$. Alors que dans l'approche originale $m$ et $n$ étaient respectivement fixés à 1 et 2, nous avons calculé ces valeurs une fois pour chaque répétition de chaque phrase. Dans ce but, nous avons considéré toutes les valeurs de phase observées dans l'intervalle de temps correspondant à la phrase et avons adopté l'approche proposée par Rosenblum et al. (2001). En pratique, nous avons calculé les 90 PLVs qui peuvent être obtenues avec toutes les combinaisons possibles de $m$ et de $n$ allant de 1 à 10 (sans inclure ceux où $m = n$ avec $m \neq 1$) et nous avons choisi la combinaison donnant la PLV la plus élevée. Avec les valeurs de $m$ et $n$ ainsi obtenues, nous avons calculé les PLVs au fil du temps en considérant les valeurs de phase contenues dans des fenêtres temporelles longues de 200ms, décalées avec un pas de 5ms.

# 4 Analyses statistiques et résultats

Les modèles mixtes ont été effectués avec RStudio (2019) en utilisant le paquet lmerTest (Kuznetsova et al., 2017) pour estimer l'effet des différents niveaux du DAF, interagissant avec l'effet de la position de lecture des phrases dans chaque essai et l'effet du type des phrases, sur les *Phase Locking Values* (PLVs), séparément dans chaque bande spectrale. Environ 2700 points de données ont été fournis à chaque modèle. Les prédicteurs des modèles étaient donc les niveaux du DAF, la position des phrases (réf : position1- les phrases lues en premier dans chaque essai) et le type des phrases. L'effet du DAF a été codé avec un contraste par différences successives (0-60-120ms), tandis que l'effet du type des phrases a été codé avec un contraste par écart de la moyenne, ce qui a donné lieu à un intercept correspondant au comportement moyen observé sur l'ensemble des phrases. Les effets aléatoires comprenaient des intercepts par participant et des pentes aléatoires spécifiques au participant pour chaque prédicteur. Afin d'obtenir des modèles parcimonieux, nous avons comparé, au moyen d'un test du $\chi 2$ (en utilisant le paquet MASS – Ripley et al., 2002), les résidus du modèle obtenus avec ou sans chacune des interactions entre les prédicteurs. Les interactions qui ne contribuaient pas au pouvoir explicatif de nos modèles ont été ensuite éliminées.

L'effet du DAF varie en fonction des bandes spectrales. Dans la **bande4** (1719 ~ 3913Hz, voir Figure3), les PLVs sont plus élevées sous DAF de 120ms que 60ms (estimée : 0,030, écart-type : 0,011, t-val. : 2,725, p-val. : 0,007). Pourtant, les PLVs ne se distinguent pas significativement entre la condition sans DAF et la condition sous DAF de 60ms et les PLVs ont même une légère tendance à être réduites par le DAF de 60ms par rapport à celles dans la condition sans DAF. Mais l'interaction positive entre le DAF (60ms-0ms) et le type des phrases principalement composées des syllabes CV (ci-après phrases CV) montre que pour les phrases CV en position initiale, les PLVs sont augmentées par le DAF de 60ms (estimée : 0,038, écart-type : 0,015, t-val. : 2,511, p-val. : 0,012), tandis qu'en positions médiane et finale, les PLVs ont également tendance à être réduites par le DAF de 60ms, comme montré par l'interaction triple significativement négative entre le DAF (60ms-0ms), le type des phrases CV et la position des phrases (position médiane : estimée : -0,051, écart-type : 0,022, t-val. : -2,335, p-val. : 0,020 ; position finale : estimée : -0,068, écart-type : 0,022, t-val. : -3,148, p-val. : 0,002). Les PLVs dans les phrases CCV(C) sont généralement plus élevées par rapport aux PLVs moyennes des trois phrases (estimée : 0,038, écart-type : 0,015, t-val. : 2,511, p-val. : 0,012), indépendamment de la condition du DAF et de la position des phrases. Dans les autres bandes, l'effet général du DAF est absent. Dans la **bande1** (158~435Hz), l'interaction négative entre le type des phrases CCV(C) et la position finale (estimée : -0,039, écart-type : 0,010, t-val. : -3,761, p-val. < 0,001) montre que les phrases CVC(C) sont liées aux PLVs moins élevées en position finale. Contrairement à ce qui est observé dans la bande4, les PLVs sous DAF de 120ms ont tendance à être moins élevées que celles sous DAF de 60ms, ainsi que cette tendance est encore plus forte pour les phrases CV, comme montré par l'interaction négative entre le DAF (120ms-60ms) et le type des phrases CV (estimée : -0,036, écart-type : 0,018, t-val. : -1,975, p-val. : 0,048). Pourtant, l'interaction

triple entre le DAF (120ms-60ms), le type des phrases CV et la position médiane est significativement positive (estimée : 0,060, écart-type : 0,026, t-val. : 2,342, p-val. : 0,019), ce qui veut dire que l'effet négatif du DAF de 120ms est réduit en position médiane. Dans la **bande2** (435~1037Hz), les phrases CCV(C) présentent les PLVs moins élevées par rapport aux PLVs moyennes des trois phrases (estimée : -0,045, écart-type : 0,009, t-val. : -4,855, p-val. < 0,001), mais en position médiane, les PLVs sont plus élevé comme montré par l'interaction positive entre le type des phrases CCV(C) et la position médiane (estimée : 0,020, écart-type : 0,010, t-val. : 2,024, p-val. : 0,043). Les phrases produites en positon finale sont généralement liées aux PLVs moins élevées (estimée : -0,015, écart-type : 0,007, t-val. : -2,104, p-val. : 0,041). L'interaction triple significativement négative entre le DAF (60ms-0ms), le type des phrases CCV(C) et la position finale (estimée : -0,051, écart-type : 0,024, t-val. : -2,113, p-val. : 0,035) montre que les PLVs des phrases CCV(C) en position finale sont encore plus réduites par le DAF de 60ms. Dans la **bande3** (1037~1719Hz), les phrases CV présentent les PLVs plus élevées (estimée : 0,070, écart-type : 0,006, t-val. : 11,831, p-val. < 0,001) mais les phrases CCV(C) moins élevées (estimée : -0,053, écart-type : 0,009, t-val. : -6,213, p-val. < 0,001) par rapport aux PLVs moyennes des trois phrases. Ces deux effets sont encore plus forts lorsque les phrases ont été lues en position finale comme montré par l'interaction positive entre le type des phrases CV et la position finale (estimée : 0,030, écart-type : 0,007, t-val. : 4,244, p-val. < 0,001) et l'interaction négative entre le type des phrases CCV(C) et la position finale (estimée : -0,014, écart-type : 0,007, t-val. : -2,007, p-val. : 0,045). Pourtant, l'interaction triple significativement négative entre le DAF (120ms-60ms), le type des phrases CV et la position finale (estimée : -0,035, écart-type : 0,017, t-val. : -2,001, p-val. : 0,046) montre que les PLVs des phrases CV en position finale sont moins élevées sous DAF de 120ms que sous 60ms. Dans la **bande5** (3913~7250Hz), les PLVs sont moins élevées dans les phrases CV que les PLVs moyennes des trois phrases (estimée : -0,021, écart-type : 0,008, t-val. : -2,679, p-val. : 0,015), indépendamment de la condition du DAF et de la position des phrases.

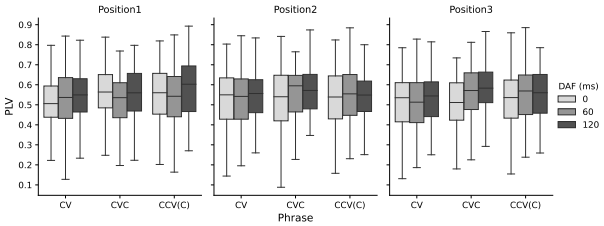


FIGURE 3 - Effets de différents niveaux du DAF sur les *Phase Locking Values* (PLVs) dans la quatrième bande (1719~3913Hz), séparémment pour les phrases contenant les syllables de différentes structures et en trois positions de lecture dans les essais expérimenaux

# 5 Discussion

Dans ce travail, nous cherchons à comprendre la nature des changements provoqué par le DAF à la fois sur la durée des gestes de parole et sur les configurations acoustiques qui sont propres à ces gestes. Nous avons proposé que ces effets du DAF ne soient pas simplement dus aux ralentissements locaux au cours de la production des énoncés mais qu'ils soient le résultat d'une restructuration des structures coordinatrices sous-jacentes au rythme de la parole. L'analyse de la coordination entre les composantes rythmiques de la parole en fonction des bandes spectrales nous permet de mieux capturer les effets du DAF sur différents types de sons, puisque les modulations d'énergie dans les bandes inférieures pourraient correspondre principalement aux gestes vocaliques et celles dans les bandes

supérieures aux gestes consonantiques (Leong et Goswami, 2015). Les résultats de notre étude montrent que les effets du DAF (à un degré élevé : 120ms) sur les PLVs « se localisent » principalement dans la quatrième bande (1719 ~ 3913Hz), dont la zone est très proche de celle de Leong et Goswami (2015) (1750~3900Hz). Dans les données de ces auteurs, l'énergie acoustique dans cette bande dépend principalement de la production des consonnes, tandis que dans nos données, cette bande correspond surtout aux hauts formants de certaines voyelles (voir la Figure2). En général, on peut en conclure que le DAF induit une réorganisation des patrons rythmiques de la parole, mais qui ne concerne qu'une sous-partie du spectre de la parole, Cette conclusion est conforme à notre hypothèse. Le fait que l'association entre la nature consonantique ou vocalique des sons et les sous-bandes spectrales dans cette étude diverge de celle obtenue par Leong et Goswami (2015) peut dépendre de différences entre les langues ciblées (Français vs. Anglais), entre les sons composant les énoncés utilisés, entre les locuteurs ou entre les conditions d'énonciation des deux études (i.e., parole produite sous DAF ou sans DAF). Nous constatons également que la structure syllabique affecte l'organisation rythmique, ce qui est suggéré par une différente coordination entre les rythmes syllabique et supra-syllabique dans les phrases contenant principalement les syllabes plus complexes (CCV(C)). Dans la quatrième bande ces phrases montrent des niveaux de coordination entre les deux rythmes plus élevés que ceux que l'on observe dans les autres phrases. Tandis que dans les trois premières bandes, ces phrases montrent des valeurs de coordination moins élevés que les phrases contenant surtout des syllabes CV or CCV. Cependant, l'effet du DAF n'interagit pas avec celui de la structure syllabique. Autrement dit, la structure syllabique en général ne conditionne pas la réorganisation rythmique sous DAF.

Les résultats de cette étude sont cohérents avec résultats d'une étude précédente (Li et Lancia, 2022), dans laquelle nous montrons que l'effet d'allongement induit par le DAF interagit avec la structure prosodique de façon que les voyelles accentuées en français situées à la frontière droite des syntagmes accentuels sont plus allongées par le DAF. En plus, dans les deux études on trouve que cet effet est indépendant de la structure (et de la durée intrinsèque) des syllabes. Les deux études montrent ensemble que la structure temporelle peut être hiérarchiquement réorganisée sous l'effet du DAF, ainsi que la présence du DAF a tendance à mettre plus en évidence certains constituants (ex. les voyelles accentuées) des énoncés en fonction de leur rôle dans la hiérarchie temporelle ou prosodique. Les études sont ainsi complémentaires : l'une a exploré la relation entre la structure prosodique et l'allongement des gestes vocaliques sous DAF, tandis que l'autre a mis en évidence la relation entre l'allongement des gestes vocaliques et la réorganisation rythmique (ex. réorganisation de la coordination entre les rythmes sous-jacents).

# 6 Conclusion

La présence du retard dans le feedback auditif des locuteurs peut provoquer une réorganisation des structures coordinatrices sous-jacentes au rythme de la parole qui se manifeste surtout dans la production des sons vocaliques. Cependant, l'association entre la nature des unités acoustiques (voyelles vs. consonnes) et les bandes spectrales pourrait dépendre de différents facteurs, et aussi varier entre locuteurs. Nous planifions donc de conduire des analyses séparées par locuteur et par niveau de DAF, afin de pouvoir prendre en compte ces types de variation dans l'analyse de l'organisation rythmique. Par ailleurs, comme les effets du DAF sur une sous-partie du spectre de la parole sont constatés dans ce travail, il serait intéressant de comparer les données acoustiques aux données articulatoires dans les futures recherches. Ce qui nous permettrait de déterminer comment la réorganisation rythmique affecte la production des gestes articulatoires et la coordination entre les articulateurs.

# Références


CAI S., BOUCEK M, GHOSH SS, GUENTHER FH, PERKELL JS. (2008). A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. In *Proceedings of the 8th Intl. Seminar on Speech Production, Strasbourg*, 65-68.

CUMMINS, F., PORT, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26*(2), 145–171.

HUANG, N. E., WU, Z., LONG, S. R., ARNOLD, K. C., CHEN, X., BLANK, K. (2009). On instantaneous frequency. *Advances in Adaptive Data Analysis, 1*(02), 177–229.

KALVERAM, K. T., JÄNCKE, L. (1989). Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition. *Folia Phoniatrica, 41*(1), 30–42.

KUZNETSOVA, A., BROCKHOFF, P. B., CHRISTENSEN, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(1), 1–26.

LANCIA, L., KRASOVITSKY, G., & STUNTEBECK, F. (2019). Coordinative patterns underlying cross-linguistic rhythmic differences. *Journal of Phonetics, 72*, 66-80.

LANCIA, L., LI, J., GOLDSTEIN, L. (2020). Complexity patterns underlying speech production activity. *International Seminar on Speech Production,* 2020 (online).

LEONG, V. (2013). Prosodic Rhythm in the Speech Amplitude Envelope: Amplitude Modulation Phase Hierarchies (AMPHs) and AMPH Models. University of Cambridge.

LEONG, V., GOSWAMI, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PloS One, 10*(12), e0144411.

LI, J., LANCIA, L. (2022). Effects of delayed auditory feedback interacting with prosodic structure. *Speech Prosody 2022* (accepté)

PATEL, A. D., LÖFQVIST, A., NAITO, W. (1999). The acoustics and kinematics of regularly timed speech: A database and method for the study of the p-center problem. *Proceedings of the 14th International Congress of Phonetic Sciences, 1*, 405–408.

RSTUDIO TEAM. (2019). RStudio: Integrated Development Environment for R. RStudio, Inc. http://www.rstudio.com/.

RIPLEY, B., VENABLES, B., BATES, D. M., HORNIK, K., GEBHARDT, A., FIRTH, D., RIPLEY, M. B. (n.d.). Package 'mass'.

ROSENBLUM, M., PIKOVSKY, A., KURTHS, J., SCHÄFER, C., TASS, P. A. (2001). Phase synchronization: From theory to data analysis. In *Handbook of biological physics, Vol. 4*, pp. 279–321. Elsevier.

SASISEKARAN, J. (2012). Effects of delayed auditory feedback on speech kinematics in fluent speakers. *Perceptual and Motor Skills*, 115(3), 845–864.