



HAL
open science

Production de la parole en réponse à de multiples perturbations du feedback auditif

Jinyu Li, Leonardo Lancia

► **To cite this version:**

Jinyu Li, Leonardo Lancia. Production de la parole en réponse à de multiples perturbations du feedback auditif. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition), 2020, Nancy, France. pp.370-378. hal-02798560v2

HAL Id: hal-02798560

<https://hal.science/hal-02798560v2>

Submitted on 18 Jun 2020 (v2), last revised 23 Jun 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Production de la parole en réponse à de multiples perturbations du feedback auditif

Jinyu Li, Leonardo Lancia

Laboratoire de Phonétique et Phonologie, 19 rue des Bernardins, Paris, France
jinyu.li@sorbonne-nouvelle.fr, leonardo.lancia@sorbonne-nouvelle.fr

RESUME

Des études antérieures ont montré que la production de la parole dépend des conditions du feedback auditif. Cette étude vise à investiguer les interactions entre les effets de trois facteurs différents sur la production de la parole : 1) le retard du feedback auditif (DAF), 2) le décalage de la f0 du feedback auditif et 3) la complexité des syllabes qui composent les énoncés. Nous avons manipulé le feedback auditif de 20 locutrices francophones pendant la répétition de trois phrases. Nous avons pu observer que plus de la moitié des participantes ont tendance à suivre la direction du décalage de la f0 du feedback auditif en recevant en continu cette perturbation. La position syllabique des voyelles est un facteur important affectant l'effet des perturbations du feedback auditif. Cependant les analyses décrites dans cette étude ne montrent pas un effet interactif du DAF et du décalage de la f0 sur la production de la parole.

ABSTRACT

Speech production in response to multiple perturbations of auditory feedback

Previous studies have shown that speech production depends on the conditions of auditory feedback. In the present study we investigate the interactions between the effects of three different factors: 1) the delay of the auditory feedback, 2) the f0 shift of auditory feedback, and 3) the complexity of syllables composing utterances. We analysed a corpus of French sentences, repeated several times by 20 French-speaking female speakers while their auditory feedback was manipulated. We could show that more than half of the participants tend to follow the direction of the f0 shift of auditory feedback. The vowels' syllabic position is an important factor affecting the effect of perturbations of auditory feedback. However, the analyses reported in this study do not support the idea that DAF and f0 shift have an interactive effect on speech production.

MOTS-CLÉS : multiples perturbations du feedback auditif, retard du feedback auditif, décalage de la f0 du signal du feedback auditif, agentivité, position syllabique

KEYWORDS: multiple perturbations of auditory feedback, delayed auditory feedback, pitch-shifted auditory feedback, agency, syllabic position

1 État de l'art

Il est bien connu que les locuteurs réagissent rapidement au décalage de la fréquence fondamentale (f_0) de leur feedback auditif. En général, ils produisent des réponses compensatoires. C'est-à-dire qu'ils changent leur f_0 dans la direction opposée à celle du décalage, lors de la production des voyelles continues (ex : Burnett *et al.*, 1998), des mots monosyllabiques ou multi-syllabiques (Natke & Kalveram, 2001 ; Donath *et al.*, 2002), des phrases (Chen *et al.*, 2007), et même en présence d'un retard du feedback auditif (DAF, ici pour *delayed auditory feedback*, Kalveram & Jäncke, 1989) ou d'une perturbation du volume du feedback auditif (Larson *et al.*, 2000). De nombreux facteurs ont été signalés comme affectant la latence ou l'amplitude des réponses des locuteurs au décalage de la f_0 . Tout d'abord, la durée syllabique doit être suffisamment longue pour que les locuteurs puissent répondre au décalage (Natke & Kalveram, 2001). La vitesse de la réponse au décalage est proportionnelle à celle du déclenchement du décalage (Larson *et al.*, 2000). De plus, les réponses à cette perturbation du feedback auditif dépendent du contrôle dynamique de la f_0 par les locuteurs : dans les tâches qui nécessitent un contrôle précis et dynamique de la f_0 , l'amplitude des réponses est plus importante, la latence et le temps pour arriver au pic des réponses sont plus courts (Xu *et al.*, 2004).

Il est également important de noter que certains locuteurs suivent parfois la direction du décalage de la f_0 . Cela se produit davantage lorsque le décalage de la f_0 est prévisible (Behroozmand *et al.*, 2012) ou lorsque la plage de décalage est importante (Burnett *et al.*, 1998 ; Behroozmand *et al.*, 2012). De plus, il a été montré que la direction de la réponse des locuteurs était dépendante de la tâche. Par exemple, il a été prouvé que les mécanismes contrôlant les réponses au décalage de la f_0 du feedback auditif étaient sensibles à l'inflexion prévue de la f_0 (Chen *et al.*, 2007). La tendance à suivre ou contrer la perturbation de la f_0 est également liée aux fluctuations du comportement du système de production (Franken *et al.*, 2018) et à l'application de la perturbation à tous les essais de l'expérience ou seulement à certains essais (Franken *et al.*, 2019).

Si les locutrices utilisent un modèle interne (Pickering & Garrod, 2013) pour contrôler leur production de la parole, lorsque les locuteurs suivent la direction de la perturbation de la f_0 , il semble que l'écart entre les valeurs prévues et perçues de la f_0 augmente même. Pourquoi alors les locuteurs suivent-ils parfois la direction de la perturbation ? Ce comportement par ailleurs ressemble à celui observé quand les locuteurs produisent leurs énoncés de façon simultanée à ceux produits par une autre personne, car dans ce cas, ils tendent à imiter les caractéristiques de la voix de cette personne (Zheng *et al.*, 2011). Le comportement du système sensorimoteur lors de la production de la parole pourrait donc dépendre à la fois des conditions du feedback sensoriel et de l'attribution de l'agentivité de la voix entendue. Il manque toutefois à la littérature une étude systématique de la manière dont les facteurs définissant les conditions du feedback sensoriel interagissent. Dans cette étude, nous étudions les interactions entre les effets du DAF, de la perturbation de la f_0 du feedback auditif et de la complexité des syllabes composant les énoncés produits sur la production de la parole.

2 Expérimentation

Les données de cette étude ont été obtenues auprès de 20 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. Les participantes ont été divisées en deux groupes de dix selon la direction du décalage de la f_0 dans leur feedback auditif (voir plus de détails ci-dessous). Trois phrases de cinq syllabes, se différenciant en complexité syllabique, 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 : 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. Les participantes percevaient leur voix, soit manipulée, soit normale, à travers le casque. L'expérience a commencé par une phase de familiarisation au DAF. Il a été demandé aux participantes de lire le court texte français « La bise et le soleil » avec un DAF de 120 ms. 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 (Burnett *et al.*, 1998 ; Liu *et al.*, 2012). Ensuite, dans la phase de test, nous avons demandé aux participantes de répéter les trois phrases en leur rythme 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 se compose de 16 blocs. Les essais composant le premier bloc ont été considérés comme des essais de contrôle, puisqu'il n'y avait aucune altération du feedback auditif. Pendant tous les essais de chaque bloc suivant, la f_0 du signal du feedback auditif a été décalée d'une valeur constante tout le long du bloc (zéro, un ou deux demi-tons). Chaque degré de perturbation de la f_0 a été appliqué à cinq blocs choisis au hasard. Pour les 10 participantes composant le groupe 1, le décalage de la f_0 était toujours positif, tandis que pour les 10 participantes composantes le groupe 2, le décalage de la f_0 était toujours négatif. Au cours de chaque essai, la valeur de DAF a été choisie au hasard parmi 0, 60 et 120 ms. Chaque valeur de DAF a donc été appliquée deux fois dans chaque bloc.

3 Analyse et résultats

Nous avons caractérisé les énoncés produits par nos participantes en analysant la durée des voyelles et la valeur médiane de la f_0 de chaque voyelle. Nous avons d'abord étiqueté à la main une répétition de chaque phrase produite dans les essais de contrôle de chaque locutrice. Les frontières des phrases et des voyelles placées à la main ont été projetées sur les autres répétitions d'une même locutrice au moyen d'une procédure basée sur le *Dynamic Time Warping* (Sakoe & Chiba, 1978) s'appuyant sur une représentation cepstrale des énoncés (incluant les coefficients de 2 à 13 et leur deltas). En raison de la présence fréquente du dévoisement à la fin de chaque essai (comprenant une répétition des trois phrases), nous avons écarté les valeurs de f_0 extraites des voyelles finales des essais.

3.1 Effet du DAF sur la durée des voyelles

La première analyse a été menée pour estimer les effets du DAF sur la durée des voyelles (voir la figure 1), afin de vérifier si les effets de cette perturbation étaient cohérents avec ceux observés dans la littérature. Cette analyse a été menée sur l'ensemble des participantes. Nous avons testé l'interaction entre l'effet du DAF, l'effet de la présence d'un accent et l'effet de la phrase par le biais d'une régression par modèles mixtes, séparément pour chaque groupe de locutrices. Aussi bien dans ce modèle, que dans les modèles décrits dans les sections suivantes, les interactions non significatives qui ne contribuaient pas de manière significative à l'adéquation du modèle aux données ont été exclues. La contribution d'une interaction à l'adéquation du modèle a été évaluée en comparant, au moyen d'un test de χ^2 , les résidus du modèle obtenus avec et sans l'interaction. Chaque modèle comprenait un intercept aléatoire par locutrice et une pente spécifique à la locutrice pour l'effet de chaque prédicteur. Dans tous les modèles qui suivent, l'effet du DAF a été codé avec un contraste par différences successives, tandis que l'effet des phrases a été codé avec un contraste par écart à la moyenne, ce qui a donné lieu à un intercept correspondant au comportement moyen observé sur l'ensemble des phrases.

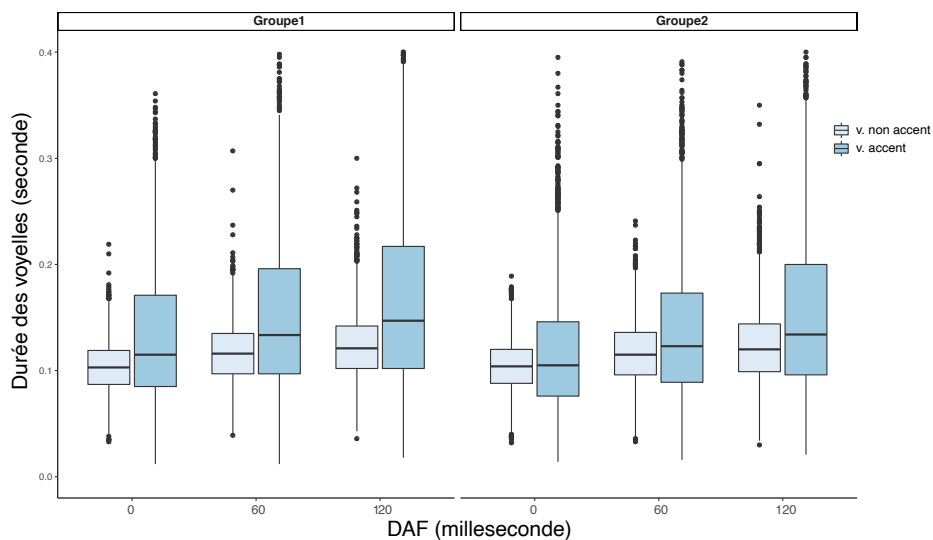


FIGURE 1 : Durée des voyelles accentuées et des voyelles non accentuées dans les conditions sans DAF (0 ms), avec le DAF de 60 ms ou de 120 ms du groupe 1 (qui est perturbé par le décalage positif de la f_0 du feedback auditif) et groupe 2 (qui est perturbé par le décalage négatif de la f_0 du feedback auditif)

Les voyelles accentuées sont généralement plus longues que les voyelles non accentuées pour le groupe 1 (estimée : 0,566, écart-type : 0,079, t-val. : 7,203, p-val. < 0,001) et le groupe 2 (estimée : 0,381, écart-type : 0,068, t-val. : 5,587, p-val. < 0,001). Les voyelles non accentuées sont allongées par le DAF de 60 ms par rapport à celles dans la condition sans DAF pour le groupe 1 (estimée : 0,202, écart-type : 0,046, t-val. : 4,444, p-val. < 0,001) et le groupe 2 (estimée : 0,239, écart-type : 0,063, t-val. : 3,818, p-val. : 0,001). Elles sont encore plus allongées par le DAF de 120 ms pour le groupe 1 (estimée : 0,113, écart-type : 0,046, t-val. : 2,468, p-val. : 0,019), mais non pas pour le groupe 2.

L'effet d'allongement est plus fort pour les voyelles accentuées avec le DAF de 60 ms à la fois pour le groupe 1 (estimée : 0,143, écart-type : 0,037, t-val. : 3.819, p-val. < 0,001) et pour le groupe 2 (estimée : 0,143, écart-type : 0,039, t-val. : 3.635, p-val. < 0,001). Ce même effet est aussi présent avec le DAF de 120 ms pour le groupe 1 (estimée : 0,159, écart-type : 0,038, t-val. : 4.218, p-val. < 0,001) et le groupe 2 (estimée : 0,191, écart-type : 0,040, t-val. : 4.802, p-val. < 0,001). Les voyelles dans la phrase qui est principalement composée des syllabes CV sont plus longues que celles dans les autres deux phrases, quelles qu'elles soient, non accentuées (estimée : 0,154, écart-type : 0,039, t-val. : 3.946, p-val. < 0,001) et cette différence augmente dans le cas des syllabes accentuées (estimée : 0,060, écart-type : 0,022, t-val. : 2.762, p-val. : 0,006) pour le groupe 1. L'effet de la structure syllabique est présent aussi dans les productions du groupe 2, mais n'augmente pas dans le cas des voyelles accentuées (estimée : 0,131, écart-type : 0,042, t-val. : 3.087, p-val. : 0,005). En revanche, les voyelles non accentuées dans la phrase qui est principalement composée des syllabes CCV sont plus courtes pour le groupe 1 (estimée : -0,247, écart-type : 0,039, t-val. : -6.337, p-val. < 0,001) et le groupe 2 (estimée : -0,261, écart-type : 0,042, t-val. : -6.160, p-val. < 0,001). Mais cet effet est réduit pour les voyelles accentuées aussi bien pour le groupe 1 (estimée : 0,263, écart-type : 0,022, t-val. : 12.165, p-val. < 0,001) que pour le groupe 2 (estimée : 0,293, écart-type : 0,023, t-val. : 12.857, p-val. < 0,001). En général, pour le groupe 1, la structure syllabique n'as pas d'impact sur l'effet du DAF et cela indépendamment de la présence de l'accent. Cependant pour le groupe 2, l'effet de DAF de 60 ms est moins important sur les voyelles non accentuées de la phrase qui est principalement composée des syllabes CCV (estimée : -0,060, écart-type : 0,027, t-val. : -2.190, p-val. < 0,029).

3.2 Effet du décalage de la f0 sur la f0 observée pour chaque locutrice

Afin d'évaluer les réponses des locutrices à la perturbation de la f0, nous avons effectué une régression linéaire pour chaque locutrice. Dans ces modèles, la variable dépendante était la valeur médiane de la f0 de chaque voyelle. Nous avons comparé les valeurs observées dans la condition de contrôle (enregistrées avant l'exposition aux perturbations de la f0) aux valeurs observées dans les blocs avec les perturbations d'un décalage positif ou négatif de la f0. Seules les voyelles produites au cours des essais sans DAF ont été prises en compte. En plus de l'effet du décalage de la f0, nous avons testé l'effet de l'accent et de la phrase, ainsi que les effets des interactions doubles et triples entre les prédicteurs. Dans le groupe 1, quatre locutrices montrent un effet significativement positif du décalage de la f0 quelle que soit la plage du décalage et à la fois dans les voyelles accentuées et non accentuées, dont une montre un effet plus fort dans les voyelles accentuées ; Une locutrice montre un effet significativement positif seulement lorsque la plage de la perturbation de la f0 était plus importante. Deux locutrices ont significativement modifié leur f0 en direction opposée de la perturbation. Dans le groupe 2, trois locutrices montrent un effet significativement négatif du décalage de la f0 quelle que soit la plage de décalage et à la fois dans les voyelles accentuées et non accentuées ; Une locutrice montre un effet significativement négatif seulement lorsque la plage de la perturbation était moins importante ; Une locutrice montre le même effet seulement lorsque la plage était plus importante ; Une locutrice montrent un effet significativement négatif quelle que soit la plage de décalage seulement pendant la production des voyelles accentuées. Pendant la production des voyelles non accentuées, cette locutrice montre ce même effet seulement lorsque la plage de la perturbation était moins importante. Une locutrice a significativement augmenté sa f0 en direction opposée de la

perturbation. En général, 55% des locutrices ont suivi la direction du décalage de la f0 au moins pendant la production des voyelles accentuée (voir dans la figure 2 la valeur médiane de la f0 observée dans les essais avec le décalage de la f0 mais sans DAF par rapport à celle observée dans les essais de contrôle de toutes les locutrices de tous les deux groupes). 30% des locutrices ne sont pas significativement affectées par le décalage de la f0. Seulement 15% des locutrices ont compensé le décalage de la f0.

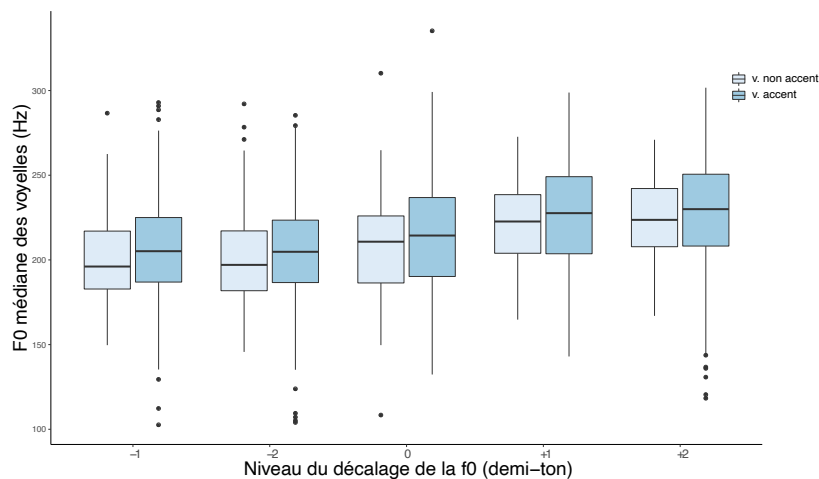


FIGURE 2 : La f0 médiane des voyelles accentuées et non accentuées des essais de contrôle (0) et des essais avec le décalage de la f0 négatif (-1, -2) ou avec le décalage de la f0 positif (+1, +2) du feedback auditif mais sans le retard du feedback auditif

3.3 Interactions entre les facteurs manipulés sur la durée des voyelles

Afin d'analyser les interactions potentielles entre les effets du décalage de la f0 et l'effet du DAF chez les locutrices qui ont suivi la perturbation de la f0, nous avons analysé comment la durée des voyelles était affectée par le décalage de la f0, la quantité du DAF, la phrase et par leurs interactions par le biais d'une régression par modèles mixtes, séparément pour chaque groupe de locutrices et pour les deux positions syllabiques (voyelles accentuées et non accentuées). Nous avons exclu les essais de contrôle. Chaque modèle (un modèle par groupe de locutrices et par position syllabique) comprenait un intercept aléatoire par locutrice et une pente aléatoire spécifique à la locutrice pour chaque prédicteur considéré. Les résultats des deux groupes sont assez homogènes. Les voyelles non accentuées de la phrase qui est principalement composée des syllabes CCV sont plus courtes à la fois pour le groupe 1 (estimée : -0,531, écart-type : 0,106, t-val. : -5.033, p-val. : 0,001) et pour le groupe 2 (estimée : -0,473, écart-type : 0,099, t-val. : -4.754, p-val. < 0,001). Les voyelles non accentuées sont allongées par le DAF de 60 ms par rapport à celles dans la condition sans DAF aussi bien pour le groupe 1 (estimée : 0,420, écart-type : 0,124, t-val. : 3.387, p-val. : 0,010) que pour le groupe 2 (estimée : 0,417, écart-type : 0,103, t-val. : 4.048, p-val. : 0,002). Cependant, pour tous les deux groupes, l'effet du DAF de 120 ms sur les voyelles non accentuées n'est pas significativement différent que celui de 60 ms. L'effet du DAF de 120 ms est réduit pour les voyelles non accentuées de la phrase CCV pour le groupe 1 (estimée : -0,132, écart-type : 0,054, t-val. : -2.454, p-val. : 0,014)

et aussi pour le groupe 2 (estimée : -0,139, écart-type : 0,049, t-val. : -2.832, p-val. : 0,005). Les voyelles accentuées sont allongées par le DAF de 60 ms par rapport à celles dans la condition sans DAF pour le groupe 1 (estimée : 0,287, écart-type : 0,073, t-val. : 3.949, p-val. : 0,004) et le groupe 2 (estimée : 0,353, écart-type : 0,079, t-val. : 4.486, p-val. < 0,001). Ces voyelles sont encore plus allongées par le DAF de 120 ms pour le groupe 1 (estimée : 0,231, écart-type : 0,073, t-val. : 3.171, p-val. : 0,013) et le groupe 2 (estimée : 0,239, écart-type : 0,079, t-val. : 3.023, p-val. : 0,006). Les voyelles accentuées de la phrase qui est principalement composée des syllabes CV sont plus longues que celles dans les autres deux types de phrase (estimée : 0,217, écart-type : 0,086, t-val. : 2.521, p-val. : 0,030). Pour les deux groupes, le décalage de la f_0 et le DAF n'ont pas un effet interactif sur la durée des voyelles des trois types de phrase, indépendamment de la présence de l'accent.

4 Discussion

Contrairement à la plupart des études précédentes indiquant que les locuteurs produisaient surtout des commandes motrices compensatoires en réponse à des perturbations inattendues de la f_0 du feedback auditif, dans cette étude, plus de la moitié des participantes ont tendance à suivre la direction du décalage de la f_0 , autrement dit, à adapter leurs commandes motrices au feedback auditif afin de stabiliser leur f_0 lorsque le feedback auditif est perturbé de façon continue pendant la production de phrases. En plus, cette adaptation se produit rapidement. Plusieurs études montrent que dans la parole synchronisée (lorsque deux locuteurs produisent les mêmes énoncés de façon simultanée), un locuteur tend à imiter la voix qu'il entend. Cela se produit même lorsque la voix entendue masque la voix du locuteur (Zheng *et al.*, 2011). De plus il a été démontré que dans les tâches de parole synchronisée, la suppression des activations du cortex auditif, qui est interprétée comme un index de la perception de notre propre voix, était absente (Jasmin *et al.*, 2016). En combinant le retard du feedback auditif avec la perturbation de la f_0 , nous avons produits des conditions d'énonciation qui se rapprochent de la parole synchronisée (ou les locuteurs ne sont jamais parfaitement synchronisés). Il est donc raisonnable d'en déduire que nos participantes n'ont pas traité le signal acoustique perçu comme le produit de leur propre activité phonatoire, mais comme celle d'une autre locutrice. Il est donc possible que les locutrices aient intégré le signal entendu lors de la production des phrases dans leur modèle prédictif interne pour anticiper le déroulement au fil du temps du signal entendu (Pickering & Garrod, 2013). Notre recherche fournit donc un nouveau paradigme pour explorer le sens de l'agentivité en parole et les résultats obtenus nous montrent que l'attribution de l'agentivité dans la production de la parole est flexible et dépendante des conditions d'énonciation.

Conformément à ce qui a été observé dans les études précédentes (ex : Kalveram & Jäncke, 1989), en réponse au retard du feedback auditif, les locuteurs allongent leurs syllabes, surtout les syllabes accentuées. Cette réponse a pour but de réduire le décalage entre les conséquences attendues des commandes motrices et les caractéristiques du signal acoustique. L'étude de Kalveram et Jäncke (1989) a cependant analysé la production de logatomes isolés composés uniquement de syllabes CV. Par conséquent, la sensibilité accrue des syllabes accentuées au DAF, par rapport aux syllabes non accentuées, pourrait être due aussi bien au fait que les voyelles accentuées sont par nature plus longues que les non accentuées, qu'à leur position dans la structure métrique. En demandant à nos participants

de prononcer des phrases dans lesquelles la complexité syllabique a été systématiquement manipulée, nous avons pu montrer que la sensibilité accrue des voyelles accentuées au DAF est principalement due à leur position dans la structure métrique. Car autrement on aurait dû observer un effet de la structure syllabique, étant donné que les voyelles contenues dans les syllabes CV sont plus longues que celles contenues dans des syllabes dont la structure est plus complexe. De plus, en demandant aux locuteurs de prononcer des phrases contenant plusieurs syntagmes accentuels, nous avons montré que la réponse au DAF est modulée par la présence d'un accent au niveau du syntagme accentuel. Ces résultats ont des implications importantes sur le rôle de la structure métrique et du phrasé dans le contrôle moteur de la parole qui méritent des recherches supplémentaires. Pour les locutrices qui ont suivi la direction du décalage de la f0 du feedback auditif, le décalage de la f0 et le DAF n'ont pas d'effets interactifs sur la durée des voyelles des trois types de phrase. Cela pourrait signifier que le fait d'intégrer le feedback auditif dans leur système de prédiction pendant la production de la parole ne suffit pas à les aider à mieux résister au DAF. Cependant, nous observons une absence de l'effet de renforcement du DAF de 120 ms par rapport celui de 60 ms sur les voyelles non accentuées pour ces locutrices. Il sera donc intéressant de comparer les résultats de ces deux groupes de locutrices aux groupes de locutrices qui produisent ces mêmes phrases avec un seul type de perturbation du feedback auditif (soit le DAF, soit le décalage de la f0).

Remerciements

Cette étude a été financée par le projet « MoSpeeDi. Motor Speech Disorders : characterizing phonetic speech planning and motor speech programming/execution and their impairments », subside CRSII5_173711/1 Sinergia du Fond National Suisse de la Recherche Scientifique et par le programme "Investissements d'Avenir" ANR-10-LABX-0083 (Labex EFL).

Références

- BEHROOZMAND, R., KORZYUKOV, O., SATTler, L., & LARSON, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *JASA*, 132(4), 2468–2477. DOI : doi.org/10.1121/1.4746984
- BURNETT, T. A., FREEDLAND, M. B., LARSON, C. R., & HAIN, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *JASA*, 103(6), 3153–3161. DOI : doi.org/10.1121/1.423073
- CHEN, S. H., LIU, H., XU, Y., & LARSON, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *JASA*, 121(2), 1157–1163. DOI : doi.org/10.1121/1.2404624
- DONATH, T. M., NATKE, U., & KALVERAM, K. T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *JASA*, 111(1), 357–366. DOI : doi.org/10.1121/1.1424870
- FRANKEN, M. K., ACHESON, D. J., MCQUEEN, J. M., HAGOORT, P., & EISNER, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic bulletin & review*, 25(4), 1458–1467. DOI : doi.org/10.3758/s13423-018-1494-x
- FRANKEN, M. K., ACHESON, D. J., MCQUEEN, J. M., HAGOORT, P., & EISNER, F. (2019). Consistency influences altered auditory feedback processing. *Quarterly Journal of Experimental Psychology*, 72(10), 2371–2379. DOI : doi.org/10.1177/1747021819838939

- HAIN, T. C., BURNETT, T. A., LARSON, C. R., & KIRAN, S. (2001). Effects of delayed auditory feedback (DAF) on the pitch-shift reflex. *JASA*, 109(5), 2146–2152. DOI : doi.org/10.1121/1.1366319
- JASMIN, K. M., MCGETTIGAN, C., AGNEW, Z. K., LAVAN, N., JOSEPHS, O., CUMMINS, F., & SCOTT, S. K. (2016). Cohesion and joint speech: Right hemisphere contributions to synchronized vocal production. *Journal of Neuroscience*, 36(17), 4669–4680. DOI : doi.org/10.1523/JNEUROSCI.4075-15.2016
- 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 Phoniatica*, 41(1), 30–42. DOI : doi.org/10.1159/000265930
- LARSON, C. R., BURNETT, T. A., KIRAN, S., & HAIN, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *JASA*, 107(1), 559–564. DOI : doi.org/10.1121/1.428323
- LARSON, C. R., SUN, J., & HAIN, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *JASA*, 121(5), 2862–2872. DOI : doi.org/10.1121/1.2715657
- LIU, H., WANG, E. Q., METMAN, L. V., & LARSON, C. R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PloS one*, 7(3). DOI : doi.org/10.1371/journal.pone.0033629
- MAX, L., & MAFFETT, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience letters*, 591, 25–29. DOI : doi.org/10.1016/j.neulet.2015.02.012
- NATKE, U., & KALVERAM, K. T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research*. DOI : [doi.org/10.1044/1092-4388\(2001\)045](https://doi.org/10.1044/1092-4388(2001)045)
- PICKERING, M. J., & GARROD, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347. DOI : doi.org/10.1017/S0140525X12001495
- SAKOE, H., & CHIBA, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43–49. DOI : [10.1109/TASSP.1978.1163055](https://doi.org/10.1109/TASSP.1978.1163055)
- XU, Y., LARSON, C. R., BAUER, J. J., & HAIN, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *JASA*, 116(2), 1168–1178. DOI : doi.org/10.1121/1.1763952
- ZHENG, Z. Z., MACDONALD, E. N., MUNHALL, K. G., & JOHNSRUDE, I. S. (2011). Perceiving a stranger's voice as being one's own: A 'rubber voice' illusion? *PloS one*, 6(4). DOI : doi.org/10.1371/journal.pone.0018655