HAL
open science

Effet de l'accélération artificielle du signal de parole sur la perception des dysarthries cérébelleuses: à propos de deux cas

Virginie Woisard-Bassols, Robert Espesser, Alain Ghio, Danielle Duez

To cite this version:

Virginie Woisard-Bassols, Robert Espesser, Alain Ghio, Danielle Duez. Effet de l'accélération artificielle du signal de parole sur la perception des dysarthries cérébelleuses: à propos de deux cas. Folia Phoniatrica et Logopaedica, 2010, 62 (4), pp.185 - 194. 10.1159/000314263 . hal-01616009

HAL Id: hal-01616009

https://hal.science/hal-01616009

Submitted on 23 Apr 2018

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.

Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2010;62:185–194
DOI: 10.1159/000314263

Published online: May 11, 2010

Effet de l'accélération artificielle du signal de parole sur la perception des dysarthries cérébelleuses: à propos de deux cas

Virginie Woisard-Bassols[a, b] Robert Espesser[a] Alain Ghio[a] Noël Nguyen[a] Danielle Duez[a]

[a]Laboratoire parole et langage, UMR6057 CNRS, Université de Provence, Aix-en-Provence, et
[b]Unité de la voix et de la déglutition, Service d'otorhinolaryngologie et de chirurgie cervicofaciale, Centre hospitalo-universitaire de Toulouse, Toulouse, France

Key Words

Compressed speech · Speech modification · Ataxic dysarthria · Intelligibility · Speech disorders

Résumé

Objectif: Dans les dysarthries cérébelleuses, la perte de la précision des mouvements articulatoires de la parole induit une dysrégulation temporelle avec un allongement irrégulier des durées contribuant globalement au ralentissement du débit de la parole. Ainsi, nous supposons que les distorsions du signal de parole sont largement influencées par le ralentissement et que la correction de ce ralentissement, par accélération du signal de parole, peut améliorer l'intelligibilité. ***Patientes et méthode:*** Deux patientes avec une dysarthrie ataxique ont été étudiées. L'intelligibilité de la parole a été évaluée subjectivement avec une échelle visuelle analogique et objectivement par la méthode des stimuli constants associée à une tâche d'identification d'un phonème cible en temps réel. L'accélération temporelle des stimuli est effectuée par l'algorithme SOLA. Un total de 144 stimuli a été créé à 3 vitesses différentes: 48 phrases à «vitesse naturelle» (sans accélération, VN), 48 phrases à «vitesse intermédiaire» (50% de l'accélération maximale, VI) et 48 phrases à «vitesse maximale» (vitesse moyenne des locuteurs témoins, VM). Les 144 phrases, dans les 3 présentations expérimentales, ont été réparties en 3 listes, chaque auditeur entendant une seule fois chaque phrase. Chaque liste est écoutée par 8 auditeurs. Les phrases sont présentées dans un ordre randomisé. Pour estimer l'effet de l'accélération, nous avons utilisé un modèle logit mixte pour le score de l'épreuve subjectif et les temps de réaction, un test de χ^2 pour le nombre des erreurs. ***Résultats:*** Pour les 2 locutrices, la probabilité d'une meilleure intelligibilité à l'épreuve subjective est significativement plus grande pour la condition VI ou VM que pour la condition VN. Le passage de la condition VN à la condition VI augmente le taux de satisfaction des auditeurs quant à l'intelligibilité de 78 à 87% pour Mme B et de 46 à 63% pour Mme Z. Les erreurs sont moins fréquentes avec la compression pour Mme Z. L'effet de l'accélération n'est jamais significatif sur les temps de réaction ($F_{2, 1024} = 2{,}14$, $p = 0{,}12$). ***Conclusions:*** Cette étude est la première à analyser l'effet d'une accélération temporelle uniforme sur de la parole pathologique. Elle montre que chez 2 patientes présentant une dysarthrie ataxique, le niveau perçu de l'intelligibilité est amélioré.

The Effect of Time-Compressed Speech on the Perception of Ataxic Dysarthria: A Report of Two Cases

Aim: In cerebellar dysarthria, the lack of precision in the articulatory motions leads to a temporal dysregulation with a decreased rate of speech. That is, our hypothesis is that the signal distortions are linked to the slowness of speech and that the acceleration of the rate improves the intelligibility of speech. ***Patients and Method:*** Two patients with a cerebellar pathology and an ataxic dysarthria participated in the present study. Speech intelligibility was assessed subjectively by a visual analog scale, and objectively by the constant-stimulus method associated with an identification task of a target phoneme in real time. The SOLA algorithm was used to accelerate the speech stimuli. 144 test sentences were created at 3 speed rates: 48 sentences at 'natural speed' (without compression, VN), 48 sentences at 'intermediate speed' (50% of the maximum rate, VI) and 48 sentences at 'maximum speed' (mean rate of a control group of speakers, VM). The 144 test sentences, in 3 experimental conditions, were distributed over 3 lists, each listener hearing each sentence only once; 8 subjects were assigned to each list. The sentences were presented in random order. A subjective assessment was performed

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2010 S. Karger AG, Basel
1021–7762/10/0624–0185$26.00/0

Accessible online at:
www.karger.com/fpl

Virginie Woisard-Bassols
Voice and Deglutition Unit – Ent Department
TSA 30030 Rangueil Larrey University Hospital
FR–31059 Toulouse Cedex 09 (France)
E-Mail woisard.v@chu-toulouse.fr

for the whole corpus. The statistical analysis was performed with a mixed logit model for the subjective scores and the reaction time and with a χ^2 test for the errors. ***Results:*** For each patient, the odds of an improved intellegibility were significantly greater for the VI or VM condition than for the VN condition. The effect tended to be stronger for patient Z. From the VN to the VI condition, the satisfaction rate regarding the intelligibility increased from 78 to 87% for patient B and from 46 to 63% for patient Z ($p = 0.03$). There were fewer errors for the accelerated speech for patient Z. The acceleration had no effect on the reaction times in the phoneme identification task ($F_{2, 1,024} = 2.14$, $p = 0.12$). ***Conclusions:*** In 2 cases of ataxic dysarthria, the perceived intelligibility was improved by a temporal acceleration of the speech disorders.

Copyright © 2010 S. Karger AG, Basel

La dysarthrie ataxique est un trouble de l'articulation de la parole secondaire à une lésion du cervelet ou des voies cérébelleuses.

Sur le plan physiopathologique, la perte de coordination et de la précision des mouvements liée à l'atteinte est à l'origine d'une dysrégulation avant tout temporelle de l'articulation de la parole [1, 2]. Pour Ackermann et al. [3], la limitation des gestes articulatoires ne peut pas compenser complètement l'exécution ralentie des mouvements dans les cas les plus marqués de bradykinésie. Récemment Spencer et Rogers [4] ont évoqué une rupture dans la programmation motrice de la parole associée au trouble de l'exécution motrice.

La littérature est riche en descriptions de la dysarthrie ataxique, mais les travaux établis à partir de locuteurs de langue française sont assez limités [5–8]. De manière générale, la parole est brusque, explosive, scandée, irrégulière et ralentie.

Les variations exagérées de l'intensité vocale et de la fréquence sont mises en rapport avec les irrégularités des mouvements articulatoires. Elles sont associées à une imprécision des consonnes et une distorsion des voyelles.

Au niveau des phonèmes [5, 8, 9], les principales anomalies concernent les consonnes occlusives sur le trait du voisement et la qualité de l'explosion. Les phonèmes [t, d] sont les plus souvent cités, suivis des phonèmes [p, b, k, g]. Les plosives non voisées en position initiale semblent les plus sensibles. La diminution du contraste entre les occlusives voisées et non voisées est objectivée par une étude mesurant le «voice onset time» en langue allemande [9].

La compétence prosodique est préservée en langue française du Québec, malgré une perte significative dans la réalisation prosodique [7]. Il persiste une nette différence d'intonation entre les énoncés interrogatifs et négatifs, même si les différences d'intonation sont significativement moindres que chez les sujets normaux. La durée des phrases, des syllabes et des pauses est augmentée. La réduction de l'accentuation prosodique est soulignée en langue anglaise [1] et suédoise [10].

Les caractéristiques intrinsèques et co-intrinsèques des voyelles ont été étudiées par Baudelle et al. [8] en langue française. Les contrastes temporels intrinsèques disparaissent en raison du trouble de la coordination alors que les contrastes fréquentiels intrinsèques sont préservés grâce au respect global de l'articulation. Les différences co-intrinsèques liées au voisement, de part l'inertie de ce paramètre, sont exacerbées pour ce qui est de la durée, mais s'estompent pour la fréquence. Les différences liées au mode et au lieu semblent, elles aussi, exacerbées par l'inertie des mouvements articulatoires.

L'ensemble de ces données met au premier plan le facteur temporel avec le ralentissement de la parole. Or, dans notre pratique clinique, l'accélération des enregistrements de parole des patients (2 fois la vitesse normale) améliore l'intelligibilité de la parole. Aucune information sur les effets de l'accélération du signal de parole dans les dysarthries n'a été retrouvée dans la littérature. Deux études décrivent une manipulation sur le débit de la parole, mais produite par le locuteur lui-même, une dans le cadre d'une sclérose latérale amyotrophique [11] et l'autre dans le cadre de la maladie de Parkinson et de la sclérose en plaques [12]. Or, les travaux de Janse [13], après comparaison entre la parole naturellement accélérée et différentes modalités d'accélération de la parole, démontrent que l'intelligibilité est meilleure avec une accélération linéaire. Dans le domaine médical, les études sur la parole compressée et la parole accélérée sont principalement concentrées sur les troubles de la perception auditive [14], l'intelligibilité dans le bruit ou chez le sujet âgé [15] et sur les capacités des cérébrolésés hémisphériques droits [16].

Pour vérifier cette hypothèse, notre étude repose sur 2 locutrices souffrant d'une dysarthrie cérébelleuse produisant une série de stimuli qui a été accélérée linéairement selon 2 vitesses différentes. Une évaluation subjective et une procédure en temps réel avec une tâche d'identification et la mesure du temps de réaction de l'auditeur ont permis d'analyser l'effet de l'accélération sur l'intelligibilité.

Les premières expériences d'accélération de la parole datent des années 50 [17] avec un champ d'application important dans le domaine de la cécité où cette technique est utilisée pour réduire le temps nécessaire à l'écoute d'un message [18]. L'altération de la compréhension du message est très lente jusqu'à 275 mots/min, puis se dégrade très rapidement lorsque les débits sont supérieurs [19]. Enfin, Lass et Prater [20] testent une accélération

variant de 100 à 300 mots/min dans une expérience évaluant le débit préférentiel pour l'audition d'une lecture de texte ou d'une conversation. Les résultats sont par ordre décroissant: 175, 200, 300 et 100 mots/min pour une population témoin. Par contre, il est à 200 mots/min pour les étudiants aveugles [19].

L'évolution des techniques de traitement numérique du signal a ensuite permis l'accélération ou le ralentissement d'un signal sonore sans modification de la fréquence. Catégorisés parmi les effets audionumériques adaptatifs [21], ces traitements ont une large place dans la création musicale numérique. Pour ce faire, plusieurs méthodes existent. Les traitements segment-temporels (type SOLA, PSOLA) répètent ou suppriment des segments d'échantillons, de la taille d'environ une période fondamentale, afin de modifier la durée. Ceci fonctionne particulièrement bien pour des signaux harmoniques au fondamental bien défini, d'où les applications en parole.

Les applications dans le domaine de la perception auditive sont nombreuses. Versfeld et Dreschler [15] ont établi pour un travail consacré à la compréhension de la parole dans le bruit des données normatives pour les phrases ayant fait l'objet d'une compression temporelle. Ainsi, en utilisant la méthode des limites, la proportion de réponse correcte en fonction du débit de la parole a été déterminée pour une population de sujets jeunes avec une audition normale. Le seuil d'intelligibilité de 50% des phrases est à 12,5 syllabes/s.

Gordon-Salant et Fitzgibbons [22] ont étudié les difficultés de reconnaissance de la parole accélérée en fonction de l'âge en vue d'applications dans le domaine de la presbyacousie. Ils ont comparé les performances en fonction de l'âge et de la présence d'une surdité, pour la reconnaissance de 4 modalités de compression temporelle de la parole: une compression linéaire uniforme à 50%, une compression sélective des pauses, une compression sélective des voyelles, une compression sélective des consonnes. Les performances des auditeurs ne sont pas modifiées par la compression sélective des pauses. Une diminution significative des performances est observée dans les 3 autres modalités de compression avec un moindre effet pour la compression sélective des voyelles, particulièrement pour les sujets âgés.

Dans les dysarthries cérébelleuses, la perte de la précision des mouvements articulatoires de la parole induit une dysrégulation temporelle avec un allongement irrégulier des durées contribuant globalement au ralentissement du débit de la parole. Celui-ci est associé à une exagération (amplification) des mouvements d'articulation; les lieux d'articulation sont relativement bien conservés.

Ainsi, nous supposons que les distorsions du signal de parole sont largement influencées par le ralentissement et que la correction de ce ralentissement peut améliorer l'intelligibilité.

Ce travail repose sur l'accélération de l'enregistrement audio de 2 patientes présentant une dysarthrie cérébelleuse de degré modéré à sévère pour:

- confirmer l'amélioration de l'intelligibilité par cette manipulation,
- déterminer le degré d'accélération nécessaire en fonction du ralentissement initial pour obtenir la meilleure intelligibilité.

Nous attendons une amélioration de l'intelligibilité en accélérant la vitesse du signal jusqu'à un débit normal (c'est-à-dire un débit produit par un locuteur non atteint d'un trouble pathologique de la parole).

Cela pourrait déboucher ultérieurement sur la mise au point d'une aide technique restituant le plus rapidement possible le message du locuteur avec un degré d'accélération adapté.

Matériel

Les locutrices

Deux patientes (Z et B) présentant une dysarthrie dans le cadre d'un syndrome cérébelleux ont fait l'objet de l'étude.

Mme Z, âgée de 34 ans, garde comme séquelle d'un accident vasculaire hémorragique du cervelet une dysarthrie sévère depuis 2 ans (grade C de l'adaptation française [23] du Frenchay Dysarthria Assessment [24] à l'épreuve de conversation: «La parole est sévèrement altérée, pouvant être comprise la moitié du temps. Il doit très fréquemment répéter»). L'examen clinique ne révèle aucune anomalie morphologique du carrefour aérodigestif. Les réalisations motrices de la sphère oro-bucco-faciale sont ralenties avec une diminution de l'amplitude et un trouble de la coordination motrice. Il existe un mouvement anormal à type de myoclonies bilatérales intéressant le vélopharyngolarynx. Les mouvements mandibulaires d'ouverture/fermeture sont saccadés. Le syndrome cérébelleux reste également invalidant au niveau de la marche avec des troubles de l'équilibre. La dysarthrie est de type ataxique, sévère avec un impact important sur la communication. Lors des épreuves de répétition, on note des erreurs inconstantes de sonorisation ou de désonorisation, de substitution de mode articulatoire, d'ajout de phonèmes ou de simplification de groupes complexes, une tendance à l'aperture vocalique. Le débit naturel est de 2,5 syllabes/s. La voix est instable tout au long de l'examen avec un caractère tremblé. Les possibilités de modulation de la voix sont réduites, le chant étant possible.

Mme B, âgée de 55 ans, souffre d'une dégénérescence spinocérébelleuse de type SCA1. Une dysarthrie ataxique s'aggrave progressivement depuis 5 ans. Le trouble de la parole reste de sévérité modérée (grade B: «La parole est anormale, mais intelligible. Le patient doit occasionnellement répéter»). La dysarthrie est associée à un trouble de la déglutition, des troubles de la marche

et des troubles sphinctériens. La force motrice et le tonus de la sphère oro-bucco-faciale sont diminués avec une lenteur et un asynchronisme des mouvements. Sur le plan de la parole, les erreurs relevées sont un allongement de certaines syllabes, quelques simplifications faisant suite à ces allongements. Le débit naturel est de 3,08 syllabes/s. Les reproductions prosodiques sont altérées, mais la prosodie reste subnormale en parole spontanée.

Un troisième patient a fourni les stimuli nécessaires à la session d'entraînement. Agé de 37 ans, ses troubles de l'élocution sont séquellaires d'un accident vasculaire du tronc cérébelleux datant de 6 mois. La dysarthrie cérébelleuse est modérée, classée B selon la classification par l'adaptation française du Frenchay Dysarthria Assessment [24].

Le matériel phonétique

Dans la mesure où nous voulons réaliser un test d'intelligibilité avec une modification de débit, l'utilisation de phrases est plus adaptée. Le risque d'erreur de transcription et la faible reproductibilité des tests d'intelligibilité sur de la parole pathologique justifie le choix de deux méthodes conduisant à la réalisation de deux expériences. Dans l'expérience 1, une méthode subjective repose sur l'utilisation d'une échelle visuelle analogique caractérisant globalement la production sans tâche de reconnaissance. L'expérience 2 est basée sur la méthode des stimuli constants [25]. Dans ce type d'expérience, le sujet doit à chaque fois répondre si oui ou non le phonème cible placé dans une phrase porteuse est présent. Ceci a été déterminant pour la constitution du corpus. Le choix des phonèmes vedettes – la constrictive [s] et l'occlusive [t] – résulte de l'analyse de la littérature décrivant les troubles phonétiques présents dans la dysarthrie ataxique. [t] a été retenue dans la mesure où les consonnes occlusives sont altérées sur le trait du voisement et la qualité de l'explosion [5]. Les occlusives non voisées en position initiale [9] semblent les plus sensibles. La constrictive [s] a été choisie pour pouvoir opposer une seconde cible d'une catégorie différente en conservant une sensibilité à un excès de voisement [8, 9]. Un test de clôture a permis le contrôle du degré de prédictibilité sémantique en éliminant de la liste des stimuli toutes les cibles prédictibles.

Un corpus de 48 stimuli a été constitué: 24 pour la détection du phonème [t] et 24 pour la détection du phonème [s]. Dans chaque groupe de 24 phrases, 12 contiennent le phonème cible et 12 ne le contiennent pas. La composition des stimuli peut donc être résumée comme suit: 12 stimuli avec le phonème cible [t], 12 stimuli distracteurs sans le phonème [t], 12 stimuli avec le phonème cible [s], 12 stimuli distracteurs sans le phonème [s]. Toutes les phrases ont été segmentées manuellement avec repérage du phonème cible.

Manipulation du signal

Le signal de parole a été accéléré par l'algorithme SOLA [26, 27] selon 2 niveaux de vitesse.

Le coefficient d'accélération a été déterminé après mesure de la durée moyenne de chaque phrase prononcée par 4 témoins de même sexe que les sujets de l'étude.

Soit dR_i la durée de référence de la phrase i et dP_{ij} la durée de la même phrase prononcée par le patient j, on définit:

$$kVM_{ij} = dP_{ij}/dR_i.$$

On a donc un coefficient d'accélération par phrase et par patient. KVM_{ij} est ici le coefficient de retour à la vitesse de référence, correspondant à une accélération maximale. Il correspondra à la modalité VM.

Une accélération à la vitesse intermédiaire (modalité VI) a aussi été effectuée avec comme le coefficient défini par:

$$kVI_{ij} = (1 + kVM_{ij})/2.$$

La vitesse originale (naturelle, non modifiée) correspond à la modalité VN.

Les données concernant les coefficients d'accélération, la durée des phrases à VN sont données dans le tableau 1. Nous avons ainsi obtenu un corpus de 48 phrases pour chacune des 3 vitesses VN, VI et VM.

Afin de contrôler la précision de la modification du débit, l'écart entre la position de chaque phonème vedette sur le signal modifié par rapport à sa position théorique a été mesuré sur un tiers du corpus. Cet écart étant toujours inférieur à 5 ms, nous avons pu utiliser le début du phonème comme temps de référence.

Pour la phase d'entraînement, un corpus de 8 phrases a été élaboré pour le phonème [p], c'est-à-dire 4 phrases avec le phonème cible [p] et 4 phrases sans [p]. Elles ont été accélérées suivant la même méthode.

Plan expérimental commun aux deux expériences

Les 144 phrases, dans les 3 conditions expérimentales (VN, VI, VM) ont été distribuées sur 3 listes construites de manière à ce que chaque auditeur ne puisse entendre 2 fois la même phrase. Tous les auditeurs ont ainsi écouté les 48 phrases avec un tiers des phrases en VN, un tiers en VI et un tiers en VM.

L'écoute était réalisée au casque, les auditeurs assis face à l'écran de l'ordinateur. La station automatisée de tests de perception et d'évaluation auditive et visuelle PERCEVAL [28] a été utilisée pour la passation.

Le traitement statistique a été réalisé avec le logiciel R.

Expérience 1

Méthode

Nous avons choisi, pour vérifier notre hypothèse, de quantifier l'intelligibilité par un score subjectif obtenu à l'aide d'une échelle visuelle analogique.

Un groupe de 24 auditeurs a coté manuellement sur une échelle visuelle analogique graduée de 10 cm «le degré de facilité ressenti pour comprendre le sens de la phrase qu'il vient d'entendre» (0 correspondant à très peu facile, 10 à très facile). Toutes les phrases, y compris celles ne contenant pas les phonèmes vedettes, ont été évaluées. Chaque série de phrases a été écoutée 4 fois pour chaque patiente. Les auditeurs, bénévoles, étaient des étudiants de l'école d'orthophonie de Toulouse de 1re et 2e année et des membres du personnel du service d'otorhinolaryngologie de l'hôpital Larrey à Toulouse.

Traitement statistique. Nous avons étudié l'effet de 2 facteurs – accélération et patient – sur le score subjectif.

Etant donnée la grande asymétrie de la distribution des scores (de 0 à 10), nous avons transformé cette variable en un facteur à 2 niveaux: score inférieur à 7 (faible intelligibilité, notée L) ou supérieur ou égal à 7 (bonne intelligibilité, codée H). La variable dépendante est désormais une variable binaire, appelé «score subjectif» prenant 2 valeurs: L ou H.

Tableau 1. Les stimuli

	Liste cible (t ou s)	Phonème (présent ou absent)	Patiente Z durée VN	Patiente B durée VN	Témoins secondes durée moyenne	Coefficient d'accélération, k_{VM} patiente Z	Coefficient d'accélération, k_{VM} patiente B
Il est très tard, il faut te lever pour partir	s	absent	4,245	3,823	2,393	0,56	0,63
J'ai accueilli l'agent qui vient d'entrer	s	absent	4,196	3,237	2,076	0,49	0,64
Mon frère n'a pas voulu venir avec moi	s	absent	4,562	2,775	1,975	0,43	0,71
Ma mère m'accompagne toute la journée	s	absent	4,155	2,728	2,112	0,51	0,77
La promenade dans la forêt est agréable	s	absent	4,949	2,89	2,302	0,47	0,80
Le chien a ramené des cailles et des perdrix	s	absent	4,77	3,251	2,356	0,49	0,72
La plongée a été longue et fatigante	s	absent	4,192	2,93	2,411	0,58	0,82
Nous cherchons des volontaires pour une étude	s	absent	5,278	3,031	2,386	0,45	0,79
L'eau est un bien naturel qu'il faut préserver	s	absent	5,24	3,359	2,740	0,52	0,82
Les travaux ont été terminés avant l'hiver	s	absent	5,377	2,824	2,345	0,44	0,83
J'ai cueilli du mimosa au bord du chemin	s	absent	5,068	3,413	2,360	0,47	0,69
Les enfants ont obtenu de bons résultats	s	absent	5,404	3,328	2,399	0,44	0,72
J'ai découvert un lézard au pied du mur	t	absent	4,157	2,723	2,013	0,48	0,74
Nous épions les flamands sur le lac	t	absent	4,027	2,366	2,041	0,51	0,86
Il y a encore des braises dans le foyer	t	absent	4,391	2,725	2,036	0,46	0,75
On a parcouru un réseau de galeries	t	absent	3,989	2,867	1,974	0,49	0,69
Ce club est composé de dix membres	t	absent	4,586	2,816	2,101	0,46	0,75
Le cheval a rué à la fin de la reprise	t	absent	4,9	3,303	2,582	0,53	0,78
Il a déménagé sur la région marseillaise	t	absent	5,626	3,369	2,405	0,43	0,71
Ce document est affiché à la mairie	t	absent	4,325	3,106	1,918	0,44	0,62
Les chaises sont empilées dans la cave	t	absent	4,608	2,674	2,126	0,46	0,80
On ramassera des mûres sur le chemin	t	absent	4,509	3,103	2,318	0,51	0,75
Le débat aura lieu au sénat en juillet	t	absent	4,625	3,046	2,411	0,52	0,79
J'ai vu des biches dans le parc	t	absent	3,585	2,35	1,735	0,48	0,74
Le tailleur en laine est sûrement fragile	s	présent	5,129	2,721	2,655	0,52	0,98
Elle a préparé la serviette pour lui	s	présent	3,951	2,528	1,956	0,50	0,77
La rivière souterraine fait un bruit terrifiant	s	présent	5,714	3,416	2,794	0,49	0,82
Le pauvre a une situation particulière	s	présent	5,133	2,619	2,380	0,46	0,91
La rencontre aura lieu au sénat en juillet	s	présent	5,336	3,517	2,418	0,45	0,69
La fête a lieu au stade à la fin du mois	s	présent	4,932	2,736	2,216	0,45	0,81
II faut savonner le perron de l'entrée	s	présent	4,435	2,754	2,144	0,48	0,78
Une fille songeuse regarde les étoiles	s	présent	5,778	3,646	2,569	0,44	0,70
Nous prendrons la sortie numéro deux	s	présent	4,424	2,876	2,291	0,52	0,80
Il a enfin une solution à présenter	s	présent	5,872	2,962	2,171	0,37	0,73
Ils ont vu un renard à l'entrée du sentier	s	présent	5,396	3,098	2,120	0,39	0,68
Le garçon trouve un beau livre	t	présent	3,878	2,553	1,900	0,49	0,74
Ces enfants sont tous agiles	t	présent	3,937	2,929	1,923	0,49	0,66
Ils admirent la toile dans le salon	t	présent	4,087	2,78	2,029	0,50	0,73
Ils sont revenus	t	présent	5,282	3,353	2,469	0,47	0,74
Aux beaux jours les tulipes sont belles	t	présent	3,964	2,682	1,900	0,48	0,71
J'ai préparé une tisane pour Léa	t	présent	4,398	2,614	1,980	0,45	0,76
J'ai mis la ceinture de ma fille	s	présent	4,269	2,651	1,757	0,41	0,66
Il a mangé une terrine de foie	t	présent	4,169	2,189	1,640	0,39	0,75
Elle aura tant de choses à nous dire	t	présent	4,605	2,457	1,763	0,38	0,72
Sous une tonnelle, se reposer est agréable	t	présent	7,607	4,335	2,803	0,37	0,65
Le gant est tombé sur la chaise	t	présent	3,549	2,195	1,921	0,54	0,87
Le président a terminé son exposé	t	présent	5,149	2,767	2,206	0,43	0,80
Le chat joue avec ton foulard	t	présent	4,059	2,794	2,179	0,54	0,78

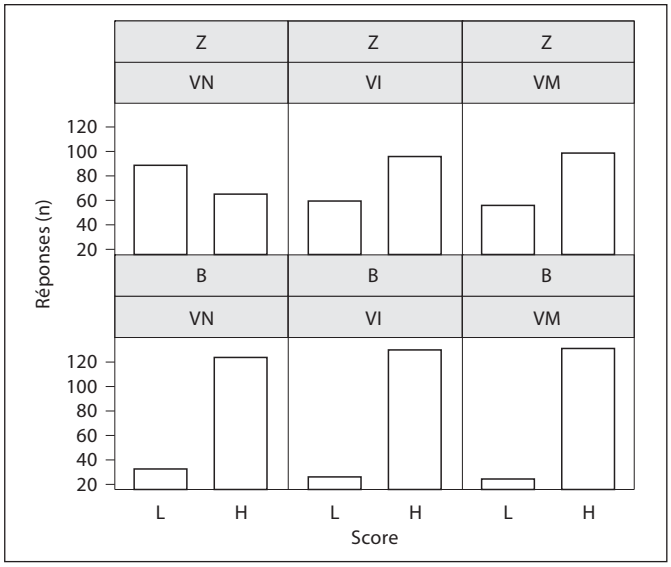

Fig. 1. Histogramme des scores de l'épreuve subjective pour les 2 locutrices; variable transformée en un facteur à 2 niveaux: score <7 (faible intelligibilité, notée L) ou ≥7 (bonne intelligibilité, codée H).

Les prédicteurs sont:

le facteur patient à 2 niveaux: Z, B;

le facteur accélération à 3 niveaux: VN, VI, VM.

Nous avons utilisé un modèle logit mixte [29, 30]. Un modèle logique prédit la probabilité d'une classification binaire, ici le taux d'appartenance à H. Un même auditeur fournissant plusieurs réponses, le facteur de groupement du modèle mixte est l'auditeur. Le terme aléatoire associe un simple offset.

Nous avons obtenu 931 mesures, et les distributions des données sont schématisées sur la figure 1.

Résultats

L'ajustement entre le modèle et les données est correcte. L'index de concordance C entre probabilité observée et probabilité estimée vaut 0,79.

Pour les 2 locutrices, la probabilité d'une «réponse H» (score >7) en débit VN est significativement plus faible que pour les débits VI et VM ($\beta = 0{,}21$, $z = 3{,}57$, $p < 0{,}001$). En revanche, la probabilité d'une réponse H n'est pas significativement différente en débit VI ou VM ($z = 0{,}34$, $p = 0{,}77$).

La probabilité d'une réponse H est significativement plus faible ($\beta = -0{,}7$, $z = -3{,}5$, $p < 0{,}001$) pour la patiente Z.

L'interaction locuteur:accélération atteint le seuil de significativité ($\beta = 0{,}12$, $z = 2$, $p < 0{,}05$). L'écart VN contre (VI + VM)/2 tend à être plus élevé pour la patiente Z.

Expérience 2

Méthode

Dans cette 2e expérience effectuée par un autre groupe d'auditeurs, apparaît sur l'écran la consigne «appuyer le plus rapidement possible sur le bouton du boîtier dès que le son cible est entendu». Pendant l'écoute de chaque phrase, le phonème cible à détecter est affiché à l'écran.

Après une phase d'entraînement de 2 phrases à VN et de 2 phrases à vitesse accélérée extraites de la liste des stimuli d'un 3eme sujet dysarthrique, l'auditeur réalisait le test de détection des phonèmes cibles. Dans chaque groupe la moitié des auditeurs commençait par la [t] et l'autre moitié par la liste [s]. Les phrases étaient présentées aléatoirement pour chaque liste.

Cette technique permet:

la mesure du temps de réaction sur les phrases présentant la cible;

la quantification des erreurs quand le sujet n'a pas appuyé alors que le phonème cible était présent;

la quantification des fausses alarmes quand le sujet a appuyé alors que le phonème cible était absent.

Les temps de réaction ont été mesurés à partir du début du signal sonore. Le temps de réaction par rapport au phonème cible a été calculé en soustrayant, au temps mesuré, l'intervalle de temps entre le début du signal et le début du phonème cible.

Huit auditeurs sont nécessaires pour l'écoute d'un stimulus. En considérant les 3 paliers de vitesse pour chaque enregistrement, 24 auditeurs sont recrutés pour chaque patiente, soit 48 auditeurs pour l'étude des 2 patientes.

La validité de cette expérience repose sur l'analyse des taux d'erreurs et de fausses alarmes.

La sensibilité est en moyenne de 94% avec un écart type de 5 et un minimum de 83%. Nous sommes face à un effet plafond. La spécificité est en moyenne de 95% avec un écart type de 6 et un minimum de 79%. Le nombre maximum de fausses alarmes est de 5 par auditeur. Neuf auditeurs ont réalisé au moins 3 fausses alarmes et la répartition par phrase est la suivante: aucune fausse alarme pour 7 phrases, 1 ou 2 fausses alarmes pour 9 phrases, au moins 3 fausses alarmes (entre 3 et 10) pour 8 phrases.

En fin d'étude est apparu que 2 phrases contenaient le phonème vedette [t], alors qu'elles étaient classées dans le groupe des phrases sans [t]. Cependant, ces phrases n'ont pas particulièrement donné lieu à des fausses alarmes (respectivement 2 et 3 fausses alarmes). Une des explications possibles est qu'il s'agit de liaisons. Pour les phrases contenant 2 fois le phonème cible, celui-ci était en position initiale et aucun auditeur n'a signalé le deuxième phonème. Nous n'avons donc pas modifié le plan de l'étude.

Il s'agit de:

la présence du phonème [t] dans le cadre de liaisons pour 2 phrases classées dans le groupe sans phonème vedette: «Ce document est affiché à la mairie» et «Les chaises sont empilées dans la cave»;

deux [s] successifs dans la phrase avec la vedette [s], le premier étant la cible: «Il a enfin une solution à présenter»;

deux [t], le premier étant la cible, le deuxième une liaison: «Sous une tonnelle, se reposer est agréable».

Traitement statistique. Pour les erreurs de détection, le taux d'erreur est faible: 69 erreurs. Leur distribution en fonction de l'accélération, par patiente et par phonème cible, a été étudié à l'aide de tableaux de contingence et un test du χ^2.

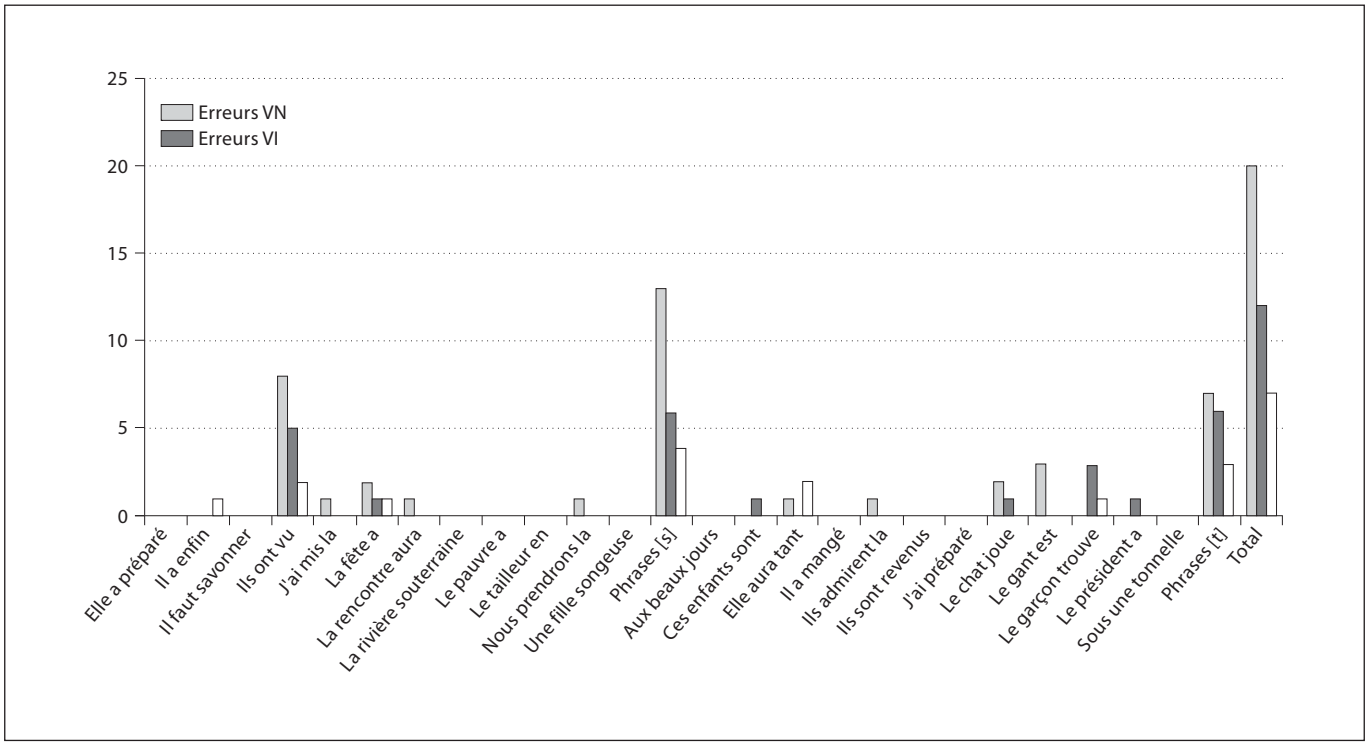

Fig. 2. Histogramme des erreurs pour Mme Z par phrase.

En ce qui concerne les temps de réaction, nous avons étudié l'effet de 3 facteurs: accélération, patient et consonne vedette.

La variable dépendante est le logarithme du temps de réaction mesuré à partir du phonème vedette.

Les prédicteurs sont:

le facteur patient à 2 niveaux: Z, B;

le facteur accélération à 3 niveaux: VN, VI, VM;

le facteur consonne à 2 niveaux: [s] et [t].

Nous avons utilisé un modèle linéaire mixte [29, 30].

Comme pour le traitement des données de l'épreuve subjective, le facteur auditeur a été pris comme facteur de groupement, le terme aléatoire associé étant un simple offset.

1117 mesures de temps de réaction sur 1152 sont comprises entre 250 et 2000 ms, soit une élimination de 3% des réponses, ce qui correspond à un pourcentage habituel pour ce type de procédure. La moyenne est de 806 ms avec un écart type de 349 ms.

Résultats

Pour les deux patientes, 27 erreurs ont été réalisées à l'écoute en VN, 22 en VI et 20 en VM (tab. 2). Pour Mme Z, une tendance est observée avec 20 erreurs en VN, 12 en VI et 7 en VM particulièrement sur le phonème [s] (fig. 2). Le test du χ^2 ne met pas en évidence une différence significative ($p > 0,5$ avec une correction de Bonferroni).

Tableau 2. Résultat du test du χ^2 pour les erreurs

Patiente	Série de phrase	Nombre d'erreurs VN	Nombre d'erreurs VI	Nombre d'erreurs VM	Valeur de p
Z	Phrases [s]	13	6	4	0,038
	Phrases [t]	7	6	3	0,42
	Phrases [s] + [t]	20	12	7	0,03
B	Phrases [s]	3	5	5	0,72
	Phrases [t]	4	5	8	0,44
	Phrases [s] + [t]	7	10	13	0,38

Les temps de réaction sont présentés sur la figure 3.

Pour les 2 locutrices, l'effet de l'accélération n'est jamais significatif ($F_{2, 1024} = 2,14$, $p = 0,12$), de même pour les interactions accélération:phonème ($F_{2, 1024} = 0,35$, $p = 0,7$) et accélération:locuteur ($F_{2, 1024} = 1,54$, $p = 0,2$). Les seuls effets significatifs concernent le locuteur et l'interaction phonème:locuteur. Les temps de réaction pour le phonème /s/ sont plus longs pour la patiente Z ($\beta = 0,28$, $t = 2,88$, $p < 0,01$) que pour la patiente B. Cet effet est moindre pour le phonème /t/ ($\beta = -0,19$, $t = -3,8$, $p = 0,0001$).

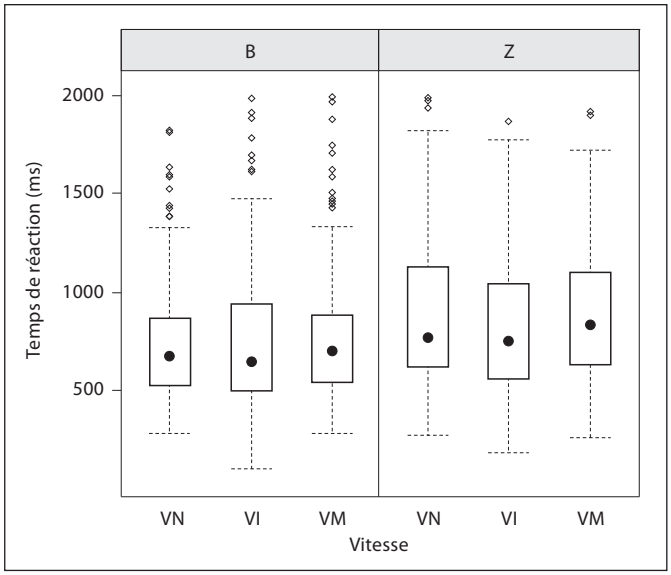

Fig. 3. Représentation en boîte des temps de réaction en fonction de la vitesse pour les 2 locutrices.

Les auditeurs ont été surpris par les altérations de la parole. Ils ont noté que certaines phrases étaient accélérées de manière «trop» rapide, mais aucun n'a repéré qu'il y avait 3 vitesses.

Discussion

Selon la mesure subjective de l'intelligibilité, une amélioration est obtenue par une accélération du signal de parole à vitesse intermédiaire ou maximale. Cet effet n'est pas retrouvé pour l'ensemble des 2 patientes à l'épreuve des stimuli constants. La nature différente des 2 tests peut expliquer cette dissociation. L'utilisation d'une échelle visuelle analogique appliquée à l'écoute de phrases a permis d'obtenir un jugement global incluant tous les paramètres participant à l'intelligibilité: le débit, la prosodie, des indices syntaxiques et sémantiques sans discrimination. Cependant, il ne s'agit pas d'une mesure de l'intelligibilité, mais de la perception subjective d'un niveau d'intelligibilité. Utiliser une tâche d'identification d'une cible phonétique sensibilisée par la mesure du temps de réaction est une méthode plus objective. Elle quantifie l'intelligibilité à partir de l'impact sur la distorsion de phonèmes. Rarement utilisée dans les dysarthries [31], cette méthode soulève la question de la pertinence des tâches d'identification et/ou de discrimination face à des distorsions (voire des élisions ou des substitutions) et de la sélection d'un nombre limité de cibles. Notre étude ne permet d'argumenter dans la mesure où la faible fréquence des distorsions phonétiques est à l'origine d'un effet plafond.

Un effet plus important paraît présent pour Mme Z. Subjectivement, sur une échelle linéaire, le passage à une vitesse intermédiaire augmente la probabilité estimée de percevoir l'intelligibilité comme meilleure de 46 à 63% (alors qu'elle passe de 78 à 87% pour Mme B) et la fréquence des erreurs diminue avec l'accélération. Ces deux locutrices présentent le même type de dysarthrie, mais avec des différences remarquables en terme de sévérité, de caractéristiques sémiologiques et en terme d'étiologie. D'ailleurs les temps de réaction sont globalement plus élevés pour Mme Z ($t = 2{,}88$, $p < 0{,}01$).

La sévérité de la dysarthrie s'accompagne d'une détérioration plus importante des structures prosodiques et phonétiques. Pour Mme Z le débit est beaucoup plus lent et la parole devient totalement incompréhensible dès qu'elle veut parler un peu plus vite. Nous pouvons émettre l'hypothèse que l'amélioration est plus sensible chez cette patiente en raison d'un degré de sévérité plus important comprenant une altération prépondérante des paramètres temporels.

Pour Mme Z, les phrases comportant le phonème vedette [s] sont associées à un taux d'erreurs plus élevé par rapport aux phrases comportant le phonème [t] et les temps de réaction sont plus faibles pour le phonème [t] ($t = 3{,}8$, $p = 0{,}0001$). Cette différence n'est pas expliquée pas les données du bilan orthophonique concernant les productions phonétiques. Mme Z, comme dans la description des dysarthries ataxiques, réalise surtout des erreurs de sonorisations et de rares substitutions de mode articulatoire. Mme B réalise très peu d'erreurs phonétiques. La différence naturelle de durée entre les deux phonèmes est la première raison à invoquer (le phonème [s] est plus long que le [t]). La seconde hypothèse est en rapport avec les effets du débit de la parole sur les distinctions segmentales. En effet, les compensations nécessaires ne sont pas univoques en fonction de la classe des phonèmes [32–34]. Par exemple, pour la distinction fricative/affricative, l'augmentation de la durée relative du silence précédent une affricative par rapport à la durée du bruit est discriminant lors de l'accélération du débit par un locuteur. Par contre, pour un contraste de voisement, lors de l'augmentation du débit de parole, la diminution de la durée du silence est nécessaire pour convertir une occlusive voisée en une occlusive non voisée. Le contexte global de la phrase a également une influence sur le jugement phonétique. Les variations de débit articulatoire ont un

impact beaucoup plus important que les variations de débit des pauses. Il ressort que les locuteurs utilisent différents indices pour différents tempi.

Bien que ces observations découlent de l'exploration de la parole naturellement accélérée, la compression linéaire conduit à des tempi différents. L'auditeur fait probablement appel à des stratégies perceptives différentes. De plus, dans l'expérience de Lass et Prater [20] évaluant le débit préférentiel pour l'audition d'une lecture de texte ou d'une conversation, l'accélération du débit à 200 et même 300 mots/min est préférée à un ralentissement à 100 mots/min.

Ainsi, l'impact de l'accélération sur les stratégies d'utilisation des indices acoustiques dans la segmentation peut contribuer à la différence des résultats entre certaines classes de phonèmes. Pour expliquer la meilleure perception de la parole accélérée des 2 patientes de cette étude, l'hypothèse de la restitution d'un débit préférentiel peut être proposée.

En l'absence de données pré-existantes dans la littérature et devant la difficulté du nombre d'auditeurs nécessaires à la multiplication des coefficients d'accélération, nous avons choisi arbitrairement 2 vitesses d'accélération correspondant:

- à la «normalisation» du débit de parole, pour la VM;
- à la moitié du coefficient d'accélération de la VM, pour la VI.

La comparaison des résultats entre la VI et la VM est en faveur d'un effet plafond notamment dans l'étude subjective.

Si l'on se réfère aux travaux réalisés avec des locuteurs et des auditeurs sains, le seuil d'intelligibilité diminue très lentement jusqu'à 10 syllabes/s, soit un seuil d'intelligibilité de 80% pour Versfeld et Dreschler [15], jusqu'à 175 mots/min pour Lass et Prater [20]. Ces seuils n'ont pourtant pas été atteints dans notre étude. Par contre, Gordon-Salant et Fitzgibbons [22] ont démontré une diminution significative des performances lors d'une compression linéaire à 50% en rapport avec la compression sélective des consonnes.

Il est difficile dans ces conditions de déterminer le meilleur coefficient d'accélération. Une méthode adaptative pourrait être proposée, comme l'ont réalisée Versfeld et Dreschler [15] avec éventuellement une compression sélective respectant de manière préférentielle les consonnes.

Le ralentissement du débit de parole observé chez les patients dysarthriques est parfois interprété comme un mécanisme compensatoire pour maintenir l'intelligibilité. Les arguments reposent sur des études mesurant l'effet d'un travail de rééducation ralentissant le débit avec des résultats intéressants et des études sur l'accélération par le locuteur du débit de parole sans effet positif [11, 12]. Dans la dysarthrie cérébelleuse, ceci est étayée par l'hypothèse d'un ralentissement compensatoire permettant aux segments concernés d'atteindre la cible. Cependant, les données de ce travail associées au fait que les épreuves utilisant les mouvements alternatifs rapides sont altérés avec une impossibilité d'accélérer la cadence convergent vers un ralentissement primitif du débit de parole.

Conclusions

L'effet de l'accélération temporelle linéaire du signal de parole sur une parole de dysarthrie cérébelleuse n'avait pas encore été évalué. Les résultats sont en faveur d'une amélioration de l'intelligibilité mesurée par une épreuve subjective. L'évaluation par la méthode des stimuli constants avec détection d'un phonème cible et mesure du temps de réaction est limitée par un effet plafond des réponses. Bien que le choix des tests d'intelligibilité et des modalités d'accélération du signal les plus adaptés reste à déterminer, ces résultats ouvrent des perspectives sur de nouvelles modalités de prise en charge des troubles de l'articulation de la parole pathologique.

Bibliographie

1 Kent RD, Kent JF, Rosenbek JC, Vorperian HK, Weismer G: A speaking task analysis of the dysarthria in cerebellar disease. Folia Phoniatr Logop 1997;49:63–82.
2 Kent RD, Kent JF, Duffy JR, Thomas JE, Weismer G, Stuntebeck S: Ataxic dysarthria. J Speech Lang Hear Res 2000;43:1275–1289.
3 Ackermann H, Hertrich I, Scharf G: Kinematic analysis of lower lip movements in ataxic dysarthria. J Speech Hear Res 1995;38: 1252–1259.
4 Spencer KA, Rogers MA: Speech motor programming in hypokinetic and ataxic dysarthria. Brain Lang 2005;94:347–366.
5 Gremy F, Chevrie Muller C, Garde E: Etude phoniatrique clinique et instrumentale des dysarthries. Technique I: résultats chez les malades présentant un syndrome cérébelleux. Rev Neurol 1967;116:401–442.
6 Gentil M: Acoustic characteristics of speech in Friedreich disease. Folia Phoniatr Logop 1990;42:125–134.
7 Le Dorze G, Ryalls J, Brassard C, Boulanger N, Ratté D: Caractéristiques prosodiques du langage dans la maladie de Parkinson et l'ataxie de Friedreich: comparaison avec des sujets normaux du point de vue neurologique. Folia Phoniatr Logop 1998;50:1–9.

8 Baudelle E, Vaissière J, Renard JL, Roubeau B, Chevrie-Müller C: Caractéristiques vocaliques intrinsèques et co-intrinsèques dans les dysarthries cérébelleuses et parkinsoniennes. Folia Phoniatr Logop 2003;55: 137–146.
9 Ackermann H, Hertrich I: Voice onset time in ataxic dysarthria. Brain Lang 1997;56: 321–333.
10 Schalling E, Hartelius L: Acoustic analysis of speech tasks performed by three individuals with spinocerebellar ataxia. Folia Phoniatr Logop 2004;56:367–380.
11 Weismer G, Laures JS, Jeng JY, Kent RD, Kent JF: Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. Folia Phoniatr Logop 2000;52:201–219.
12 Tjaden K, Wilding GE: Rate and loudness manipulations in dysarthria: acoustic and perceptual findings. J Speech Lang Hear Res 2004;47:766–783.
13 Janse E: Word perception in natural-fast and artificialy time-compressed speech. 15th Int Congr Phonet Sci, Barcelona, 2003, pp 3001–3004.
14 Gray BB: The intelligibility of time compressed words as a function of age and hearing loss. J Speech Hear Res 1969;12:443–448.
15 Versfeld NJ, Dreschler WA: The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. J Acoust Soc Am 2002; 111:401–408.
16 Leonard CL, Baum SR, Pell MD: The effect of compressed speech on the ability of right-hemisphere-damaged patients to use context. Cortex 2001;37:327–344.
17 Fairbanks G, Kodman F: Word intelligibility as a function of time compression. J Acoust Soc Am 1957;29:636–644.
18 Foulke E: The retention of information presented at an accelerated word rate. Int J Educ Blind 1966;14:11–15.
19 Foulke E, Sticht TG: Review of research on the intelligibility and comprehension of accelerated speech. Psychol Bull 1969;72:50–62.
20 Lass NJ, Prater E: A comparative sudy of listening rate preferences for oral reading and impromptu speaking tasks. J Commun 1973; 23:95–102.
21 Verfaille V: Effets audionumériques adaptatifs: théorie, mise en oeuvre et usage en création musicale numérique; thèse, Université Aix-Marseille II (spécialité ATIAM), 2003.
22 Gordon-Salant S, Fitzgibbons PJ: Source of age related recognition difficulty for compressed speech. J Speech Lang Hear Res 2001; 44:701–719.
23 Auzou P, Rolland-Monnoury V: Batterie d'évaluation clinique de la dysarthrie. Isbergues, Ortho-Edition, 2006.
24 Enderby PM: Frenchay Dysarthria Assessment. San Diego, College Hill Press, 1983.
25 Sawusch JR: Instrumentation and methodology for the study of speech perception; in Lass N (ed): Principles in Experimental Phonetics. Saint Louis, Mosby Yearbook, 1996, pp 525–550.
26 Roucos S, Wilgus AM: High quality time scale modification for speech. Proc Int Conf Acoust Speech Signal Processing, IEEE, New York, 1985, pp 493–496.
27 Hirst D, Espesser R: Automatic modelling of fundamental frequency using a quadratic spline function. Trav Inst Phonét Aix 2001; 15:75–85.
28 Ghio A, André C, Teston B, Cavé C: PERCEVAL, une station automatisée de tests de perception et d'évaluation auditive et visuelle. Trav Interdiscip Lab Parole Lang Aix 2003;22:115–133.
29 Pinheiro JC, Bates DM: Mixed-Effects Models in S and S-Plus: Statistics and Computing. New York, Springer, 2000, p 521.
30 Bates DM, Sarkar D: The LME4: linear mixed-effects models using 54 classes. R package version 0.99875–1. 2007. http://lib.stat.cmu.edu/R/CRAN/.
31 Gatignol P, Tanhere F, Clero D, Lobryeau C, Soudant J, Lamas G: Comparative evaluation of speech disorders and verbal and nonverbal communication within two groups of patients with facial paralysis (FP) and those who had undergone hypoglosso-facial anastomosis (HFA). Rev Laryngol Otol Rhinol 2003;124:291–297.
32 Miller JL, Grosjean F: How the components of speaking rate influence perception of phonetic segments. J Exp Psychol Hum Percept Perform 1981;7:208–215.
33 Miller JL, Dexter ER: Effects of speaking rate and lexical status on phonetic perception. J Exp Psychol 1988;14:369–378.
34 Kato H, Tsuzaki M: Effects of phoneme class and duration on the acceptability of temporal modifications in speech. J Acoust Soc Am 2002;111:387–399.