

Corrélats auditifs et cognitifs à la capacité de restauration de la parole accélérée

Caroline Jacquier et Fanny Meunier

Laboratoire Dynamique du Langage (UMR 5596 – Université Lumière Lyon 2)
14, avenue Berthelot – 69363 Lyon Cedex 07, France
Mél: jacquier@isc.cnrs.fr, fanny.meunier@univ-lyon2.fr

ABSTRACT

We explore the relationship between auditory measures, reading capacities and the ability to reconstruct time-compressed speech for individuals without language trouble. We focused on two short attributes of speech: Voice Onset Time (VOT) and second formant transition. Normal hearing subjects had to identify disyllabic CVCV non-words that have been time-compressed on both acoustic cues simultaneously. The time compression experience showed a large inter-individual variance and allowed us to contrast a good performer and a bad performer groups for speech perception. Complementary studies (audiometric test and reading skills evaluation) showed that there is no correlation between auditory measures and cognitive mechanisms of degraded speech reconstruction whereas there are specific correlations between reading capacities and performances in cognitive reconstruction.

1. INTRODUCTION

La compréhension du langage parlé est une tâche complexe menée quotidiennement qui représente un haut degré d'implication des fonctions cognitives. L'étude des enfants ayants des troubles du langage et de l'apprentissage -comme les dyslexiques qui ont un problème d'apprentissage de la lecture sans déficit intellectuel ou troubles neurologiques- révèle des difficultés à percevoir des segments brefs du signal de parole [1]. Selon l'hypothèse auditive de la dyslexie, il existerait une corrélation entre la perception auditive et les capacités de lecture. Tallal a montré, par exemple, qu'il existait une grande variabilité dans la capacité à lire des non-mots chez des sujets dyslexiques. Ce type de résultats suggère que le dysfonctionnement perceptif auditif affecterait les capacités à utiliser les compétences phoniques indispensables à la lecture. Mais, d'autres études ont montré des résultats contradictoires. Les déficits des mauvais lecteurs seraient imputables à leur perception spécifique de la parole c'est-à-dire à sa nature phonologique et non auditive. Le but de notre étude est d'établir pour des sujets sans trouble du langage les relations entre performances auditives, performances phonologiques et performances de lecture.

La variabilité intrinsèque du signal acoustique de la parole pose un important problème pour la modélisation de la compréhension de la parole. En effet, le signal de parole est constitué de nombreux segments acoustiques

modulables qui ont des degrés d'importance différents dans la perception de la parole [2-4]. Les recherches sur l'hypothèse phonologique de la dyslexie mettent en évidence des problèmes d'intégration, de traitement des sons rapides et brefs de la parole. Des déficits cognitifs dans la représentation phonologique des mots induiraient une faible conscience phonémique et des problèmes de traitement segmental des stimuli de parole.

Dans notre étude nous nous sommes intéressés à deux indices acoustiques rapides et brefs : le Délai d'Etablissement du Voisement (DEV) et la Transition du Formant 2 (TF2).

1.1. Les indices acoustiques

Le Délai d'Etablissement du Voisement (DEV)

Selon Lisker et Abramson [5], le DEV est défini comme l'intervalle de temps entre l'explosion de l'occlusive et le début du voisement. Le DEV peut-être négatif si le voisement débute avant la fin de l'explosion, nul si la synchronisation est parfaite et positif si le voisement commence un certain temps après la fin de l'explosion. L'aspiration correspond à un souffle s'échappant des poumons après l'explosion, ce phénomène est mis en évidence par un DEV positif plus long. Ainsi, les valeurs de DEV rendent compte du degré de voisement des consonnes.

La Transition Formantique (TF)

La TF correspond à un changement rapide de fréquence au moment de l'explosion de la consonne occlusive. Les changements rapides de fréquence sont primordiaux pour l'identification des segments acoustiques. La transition du second formant est un indice pour déterminer le lieu d'articulation des occlusives.

1.2. La reconstruction de la parole

La compréhension de la parole chez les sujets normo-entendants est une faculté cognitive très robuste qui résiste aux variabilités acoustiques du signal de parole. Certains mécanismes cognitifs semblent compenser et permettre la reconstruction du signal de parole altérée [6]. Ces capacités de reconstruction dépendent à la fois de la nature et du degré de distorsion appliqués au signal mais aussi de chaque individu : les capacités de perception et de compréhension de la parole dégradée sont propres à chacun, et une grande variabilité inter-individuelle est parfois observée dans des tâches d'intelligibilité.

Le but de notre étude est, d'une part, d'observer cette variabilité de reconstruction cognitive inter-individuelle à

partir d'une dégradation spécifique du signal de parole, et d'autre part, de tenter de comprendre et de circonscrire cette variabilité inter-individuelle chez des sujets normo-entendants et sans trouble du langage. Pour dégrader le signal, nous nous sommes intéressés à la dimension temporelle du signal acoustique en accélérant certains segments du signal. Ce paramètre semblant être critique pour la population dyslexique, nous avons voulu savoir ce qu'il en était chez les normo-entendants et si une grande variabilité était observable pour cette population. Nous avons manipulé deux indices acoustiques : le Délai d'Etablissement du Voisement (DEV) et la Transition du Formant 2 (TF2). Nous avons donc mesuré les effets de la compression temporelle d'indices acoustiques sur l'intelligibilité de la parole sur des sujets sans trouble du langage. Une expérience a été effectuée dans laquelle nous avons accéléré les deux indices simultanément. Nous avons ensuite conduit des études complémentaires afin de voir si les variabilités de performances observées étaient liées à des caractéristiques audiolinguistiques et/ou cognitives de lecture propres à chacun des participants.

2. EXPÉRIENCE de COMPRESSION TEMPORELLE

Trente deux participants, âgés de 18 à 32 ans, de langue maternelle française, et n'ayant jamais connu aucun trouble auditif, du langage, ou neurologique, ont participé à cette expérience.

2.1. Matériel et Méthode

Les stimuli étaient composés de 64 non-mots bisyllabiques de forme CVCV et de 16 fillers de forme VCV. Quatre consonnes occlusives (/b/, /d/, /p/, /t/) et deux voyelles (/a/ et /i/) ont été combinées pour construire chaque stimulus. Chaque consonne apparaît avec chacune des autres consonnes dans les deux syllabes et avec les deux voyelles ($4C_1 \times 4C_2 \times 2V_1 \times 2V_2 = 64$ CVCV). Les stimuli ont été produits par un locuteur français et enregistrés dans une chambre sourde. Les fichiers sons étaient sauvegardés sous le format wav et échantillonnés à 22 kHz (stéréo, 16 bits). La durée de chaque indice acoustique (le DEV et la TF2) a été mesurée manuellement pour chacun des items à l'aide du logiciel Praat. De même, le DEV a été segmenté manuellement à partir du début des pulsations périodiques régulières et de la détente de l'occlusion : le DEV est positif pour les occlusives non voisées et négatif pour les voisées. La TF2 a été délimitée à partir du changement brusque et rapide de la fréquence du F2, lors de la transition entre la consonne et la voyelle, jusqu'à la partie stable de la voyelle mesurée sur une représentation spectrographique en délimitant la zone où les formants sont parallèles à l'axe du temps. Pour les deux indices de chaque syllabe, la durée a été accélérée selon quatre conditions expérimentales de compression : une condition contrôle correspondant à la durée initiale, une condition 50% correspondant à 50% de la durée initiale, de même, une condition 25% et une condition 0% où les deux indices sont entièrement supprimés. La compression temporelle des indices acoustiques se fait par la méthode PSOLA (Pitch-Synchronous Overlap Add). Le but de l'analyse est d'effectuer un fenêtrage exactement centré

sur les périodes fondamentales du signal. Le signal de synthèse est alors reconstitué par superposition-addition (overlap-add) de ces formes d'onde élémentaires. De plus, les parties segmentées du signal acoustique (le DEV et la TF2) peuvent être accélérées alors que le reste du signal reste intact.

Les participants étaient assis dans une pièce silencieuse face à un écran d'ordinateur. Les stimuli étaient émis en modalité auditive binaurale à l'aide d'un casque (Beyerdynamic DT 48, 200Ω) et présentés dans un ordre aléatoire différent pour tous les participants. Les participants étaient informés qu'un signal de parole, pas nécessairement un mot, allait être présenté dans le casque et qu'ils devaient taper sur le clavier ce qu'ils avaient entendu. Un entraînement leur était auparavant proposé.

2.2. Résultats

Nous avons calculé les pourcentages d'identification des participants pour les items, les consonnes et les voyelles. Globalement, les voyelles étaient mieux identifiées que les consonnes. On observe pour les consonnes un effet de toutes les conditions de compression [$F(3,93)=652.32$, $p<.05$]. Les performances d'identification des consonnes dépendent aussi de la position de celles-ci : la première consonne était moins bien identifiée (70,4%) que la seconde (77%) [$F(1,31)=28.97$, $p<.05$]. On observe également une interaction entre ces deux facteurs [$F(3,93)=5.75$, $p<.05$] : selon la position (première ou deuxième syllabe), l'intelligibilité de la consonne était différemment modulée par la compression. Par ailleurs, nous avons observé une grande variabilité inter-individuelle des performances. Cette variabilité entre les 32 participants est plus importante pour la condition 25% pour les deux consonnes ($SD=0.13$ et $SD=0.15$). Par exemple, pour C1, les performances vont de 94% pour le meilleur sujet à 25% d'identification pour le moins bon. Cette variabilité reflète une différence dans les capacités à reconstruire la parole dégradée.

3. ETUDES COMPLEMENTAIRES

Suite aux résultats de l'expérience, qui montrent une grande variabilité des performances entre les participants, nous avons formé deux groupes de 12 sujets correspondant aux sujets les plus extrêmes dans leur performance à reconstruire la parole dégradée (groupes : Haute Performance, HP et Basse Performance, BP). Chaque sujet a passé un audiogramme et un test d'évaluation des capacités cognitives de lecture. Ces deux tests ont pour objectifs, d'une part, de valider la normalité des capacités auditives et de lecture des sujets et d'autre part, de mettre en évidence d'éventuelles corrélations entre les performances comportementales observées chez nos sujets et leurs capacités auditives et leurs capacités de lecture. En effet, les différences entre les sujets pourraient être liées autant à des déficits de langage (de type 'dyslexie') qu'à des troubles auditifs (perte de certaines fréquences acoustiques). Existe-t-il un lien entre l'audition et les performances cognitives de restauration de la parole ou/et un lien entre la lecture et les performances cognitives de reconstruction de la parole ?

3.1. Test d'audiométrie tonale

L'audiométrie tonale consiste à déterminer, pour plusieurs fréquences connues (125 à 8000 Hz), les seuils d'audition subjectifs d'un sujet. Le spectre de la voix s'étendant de 500 Hz à 2 kHz (voyelles : basses fréquences, consonnes : hautes fréquences). Les tests ont été pratiqués au laboratoire dans un caisson d'enregistrement insonorisé afin de s'isoler des bruits ambiants. L'audiomètre génère des fréquences à différentes intensités. Les sons purs sont présentés séparément pour chacune des oreilles à l'aide d'un casque. Leur intensité diminue progressivement jusqu'à ce que le sujet ne puisse plus l'entendre : c'est le seuil d'audition.

3.2. Test de lecture

Pour l'évaluation des capacités de lecture, nous avons utilisé le logiciel ECCLA (Evaluation diagnostic des Capacités Cognitives du Lecteur Adulte) conçu par Zagar, Jourdain et Lété, en 1995 [7]. Ce logiciel permet de déterminer la ou les sources de difficultés d'un lecteur. En effet, la lecture est conçue comme une activité cognitive modélisée par une suite d'étapes de traitement : visuel, phonologique et lexical. Le test est composé de 15 épreuves, divisé en 4 passations durant lesquelles nous enregistrons les temps de réaction des sujets. Différentes épreuves permettent l'étude de chaque étape de traitement. Pour l'étape de traitement visuel, les épreuves portent sur le codage des lettres, le jugement d'identité de lettres et le jugement de similitude de deux ensembles de lettres, prononçables ou non. Pour l'étude de l'étape de traitement phonologique, les épreuves sont une tâche de décision phonologique à choix forcé et une tâche de décision lexicale phonologique. De nombreuses épreuves évaluent l'étape de traitement lexicale : une tâche d'amorçage orthographique, des tâches de décision lexicale avec des effets de fréquence, de régularité et de la classe grammaticale, une tâche de décision phonologique à choix forcé et une tâche de catégorisation sémantique. De plus, les caractéristiques de la lecture (vitesse, temps de lecture, effet de fréquence, effet du contexte...) sont étudiées lors d'une tâche de lecture de textes mot à mot et une tâche de lecture de mots hors contexte. Ce test étant relativement long (environ 1h45), nous l'avons coupé en deux parties, de deux passations chacune, séparées par le test d'audiométrie qui durait 15 minutes.

3.3. Résultats

Audiométrie Tonale

Nous observons une diminution du seuil auditif pour les hautes fréquences (à partir de 4000 Hz) en particulier pour 6000 Hz. En effet, à 4000 et à 6000 Hz pour l'oreille droite, le groupe BP a un seuil auditif significativement plus bas (4000 Hz=5 dB, 6000 Hz=10 dB) que le groupe HP (4000 Hz=1.67 dB, 6000 Hz=4.58 dB) ainsi qu'à 6000 Hz pour l'oreille gauche (BP=10.83 dB, HP=4.58 dB).

Capacités de lecture

Le test ECCLA a fourni une importante masse de données qui ont été analysées de manière exhaustive. Nous ne

présenterons ici que les résultats les plus pertinents à notre étude. Le test de Student, qui compare les moyennes des groupes pour chaque indicateur, a révélé des différences significatives pour deux indicateurs seulement, un pour l'étape visuelle et un pour l'étape de lecture. Le premier indicateur « visuel » montre que la reconnaissance des mots semblent être plus efficace pour le groupe HP (6.67% d'erreurs) que pour le groupe BP (13.33%) lorsque nous leur demandons de juger de la similitude de deux groupes de lettres non prononçables qui diffèrent uniquement par une seule lettre (Ex. BCDTF BCDTG). Ce résultat est intéressant car cette différence entre les deux groupes n'est pas observée dans le cas où les suites de lettres sont prononçables. Le second indicateur « lecture » met en évidence un effet de fréquence lors de la lecture de mots hors contexte, c'est-à-dire que les temps de lecture des sujets BP sont plus longs que les sujets HP pour les mots de basse fréquence. Il semble donc que globalement le groupe HP soit de très bons lecteurs.

Corrélations : Reconstruction/Lecture pour HP et BP

Les résultats du groupe HP montrent uniquement des corrélations entre les performances et des indicateurs de l'étape de lecture. On observe pour ce groupe une corrélation négative entre les performances de reconstruction et le style de lecture ($r=-0.58$, $p<.05$), c'est-à-dire que plus les sujets HP lisent vite meilleurs ils sont à reconstruire la parole dégradée. Cette corrélation négative se retrouve entre les performances de reconstruction et le temps de lecture de mots fréquents ($r=-0.62$, $p<.05$).

Les résultats du groupe BP montrent de nombreuses corrélations entre leurs performances et des indicateurs des étapes de reconnaissance des mots. Une corrélation est observée pour une tâche de jugement de similitude de 2 suites de lettres prononçables, en effet, la diminution des performances de reconstruction est liée à une augmentation des temps de réponse ($r=-0.82$, $p<.01$). De même, dans une tâche de décision phonologique, la diminution des performances de reconstruction est corrélée à une augmentation des temps de réponse ($r=-0.60$, $p<.05$). Enfin, un résultat très intéressant est la corrélation négative observée dans une tâche de décision lexicale : la diminution des performances de reconstruction est liée à une augmentation des temps de réponse pour les mots irréguliers ($r=-0.67$, $p<.05$). Certains indicateurs portant sur les capacités propres à la lecture corréleront également significativement avec les résultats de reconstruction de la parole dégradée des sujets. Ainsi les temps recueillis dans une tâche de lecture de mots hors contexte corréleront négativement avec les performances de reconstruction ($r=-0.58$, $p<.05$). On observe également une corrélation entre la sensibilité à la longueur des mots dans une tâche de lecture de textes mot à mot et les performances de reconstruction ($r=-0.63$, $p<.05$) : plus les participants sont sensibles à la longueur des mots, moins ils sont performants dans la reconstruction du signal de parole dégradée. Enfin on observe une corrélation entre l'effet de contexte phrasique sur les temps de lecture et les performances de reconstruction ($r=0.69$, $p<.05$).

4. DISCUSSION

Dans cette étude, nous avons observé, dans une première partie, les effets de la compression temporelle de deux indices acoustiques brefs (le DEV et la TF2) sur l'intelligibilité de la parole par des sujets normo-entendants. Nous avons retrouvé une variabilité inter-individuelle des performances pour la perception auditive de non-mots, surtout à la condition de compression 25%. Cette variabilité pourrait être liée à un déficit auditif ou à un trouble d'apprentissage de la lecture comme Tallal l'a observé chez les dyslexiques. Parmi les sujets, nous avons distingué deux groupes : un groupe de personnes avec de hautes performances (HP) et un groupe avec de basses performances (BP). Pour tenter de comprendre et de circonscrire cette variabilité nous avons testé la perception auditive et les capacités de lecture des deux groupes.

Les résultats des tests audiométriques rendent compte d'une bonne audition de l'ensemble des sujets même si nous avons observé une augmentation des seuils auditifs pour les hautes fréquences, à partir de 4000 Hz. Ce résultat correspond aux travaux établissant une perte à 6000 Hz comme le premier indice de la perte auditive induite par le bruit. Dans notre étude, nous avons observé une légère asymétrie et une prédominance à droite de cette perte des hautes fréquences chez le groupe BP. L'asymétrie à droite pourrait expliquer la diminution des performances chez le groupe BP. En effet, les voies auditives afférentes se croisent au niveau de l'olive supérieure donc les informations auditives captées par l'oreille droite vont être projetées sur les aires auditives primaires du cortex gauche. Les aires du langage se trouvent également principalement dans l'hémisphère gauche donc la perte d'information au niveau de l'oreille interne droite va se répercuter sur le traitement du langage. Par ailleurs, au niveau acoustique, la bande de fréquence d'une consonne se situe entre 1500 et 6000 Hz alors que celle d'une voyelle est entre 250 à 1000 Hz. Donc, l'augmentation des seuils auditifs à partir de 4000 Hz pourrait expliquer, de manière générale, le déficit d'identification des consonnes par rapport aux voyelles. Par conséquent, il sera plus difficile de percevoir ces sons consonantiques et donc de reconnaître les stimuli de forme CVCV.

Les résultats du test d'évaluation des capacités de lecture nous conduisent à de multiples discussions. Tout d'abord, suite aux analyses statistiques nous avons observé des effets de prononçabilité et de longueur des non-mots sur les temps de réponse des sujets. Une suite de lettres non prononçable est très rapidement traitée par le groupe HP alors que les non-mots de deux syllabes sont traités plus lentement par le groupe BP. Pris ensemble, ces résultats nous permettent de suggérer que les difficultés du groupe BP pourraient être dues à une mauvaise activation des codes graphémiques et donc une mauvaise représentation des formes graphiques des lettres. Cette représentation graphémique entre en jeu, principalement, dans la voie phonologique de reconnaissance des mots. La voie phonologique permet de traiter les nouveaux mots ou les non-mots comme ceux utilisés dans notre expérience de compression temporelle. Donc, si la représentation graphémique de cette voie est déficiente, l'étape de

transformation des codes graphémiques en codes phonologiques serait perturbée et l'activation du lexique par le biais des codes phonologiques ou directement par les codes graphémiques serait plus faible. La reconnaissance des mots serait ainsi altérée. Ensuite, lors de l'analyse des corrélations, nous avons remarqué que les deux groupes n'évoluaient pas dans le même sens selon les étapes de traitement. Il apparaît que le groupe HP soit composé de très bons lecteurs alors que le groupe BP regroupe des sujets ayant des problèmes dans la reconnaissance des mots. Les résultats des corrélations permettent d'appuyer l'hypothèse émise précédemment sur les causes probables des déficits cognitifs des sujets BP. Nous avons montré, chez les sujets BP, que l'activation des chaînes de caractères et de codage de la position des lettres était déficiente dans l'étape visuelle. De même, nous avons mis en évidence des difficultés dans la transformation des codes graphémiques en codes phonologiques, ainsi que dans l'accès au lexique, c'est-à-dire lors de l'appariement du pattern graphémique avec une entrée du lexique mental du sujet. La reconnaissance des mots irréguliers nécessite un accès au lexique par la voie directe, l'utilisation par le système de correspondance graphème-phonème conduisant à une réponse erronée.

5. CONCLUSION

Une faible relation entre la perception auditive et la reconstruction de la parole dégradée a été mise en évidence pour l'atteinte des hautes fréquences. Par contre, il semblerait que les étapes cognitives utilisées lors de la lecture soient déficientes chez les personnes qui ont des difficultés à reconstruire de la parole accélérée.

BIBLIOGRAPHIE

- [1] P. Tallal. Auditory temporal perception, phonics, and reading disabilities in children, *Brain and language*, 9:182-198, 1980.
- [2] W. Serniclaes. Etude expérimentale de la perception du trait de voisement des occlusives du Français. *Ph. D. Dissertation, Université Libre de Bruxelles, 1987.*
- [3] R. D. Kent and K. L. Moll. Vocal-tract characteristics of the stop cognates. *Journal of Acoustical Society of America*, 46:1549-1555, 1969.
- [4] L. Lisker and A. S. Abramson. Some effects of context on voice onset time in English stops. *Language and Speech*, 10:1-28, 1967.
- [5] L. Lisker and A. S. Abramson. A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20:384-422, 1964.
- [6] F. Meunier, T. Cenier, M. Barkat, and I. Magrin-Chagnolleau. Mesure d'intelligibilité de segments de parole à l'envers en français. In *proc. of XXIVèmes Journées d'Etude sur la Parole*, pages 117-120, 2002.
- [7] D. Zagar, C. Jourdain and B. Lété. Le diagnostic cognitif des capacités de lecture : le logiciel ECCLA (Evaluation diagnostic des Capacités Cognitives du Lecteur Adulte). *Revue Française de Pédagogie*, 113:19-29, 1995.