HAL
open science

Perception audio-visuelle de séquences VCV produites par des personnes porteuses de Trisomie 21 : une étude préliminaire

Alexandre Hennequin, Amélie Rochet-Capellan, Marion Dohen

To cite this version:

Alexandre Hennequin, Amélie Rochet-Capellan, Marion Dohen. Perception audio-visuelle de séquences VCV produites par des personnes porteuses de Trisomie 21 : une étude préliminaire. JEP-TALN-RECITAL 2016 - conférence conjointe 31e Journées d'Études sur la Parole, 23e Traitement Automatique des Langues Naturelles, 18e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, Jul 2016, Paris, France. hal-01348019

HAL Id: hal-01348019

https://hal.science/hal-01348019v1

Submitted on 22 Jul 2016

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.

Perception audio-visuelle de séquences VCV produites par des personnes porteuses de Trisomie 21 : une étude préliminaire

Alexandre Hennequin[1], Amélie Rochet-Capellan[2] & Marion Dohen[1]
(1) Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France
(2) CNRS, GIPSA-Lab, F-38000 Grenoble, France
{alexandre.hennequin, amelie.rochet-capellan, marion.dohen}@gipsa-lab.grenoble-inp.fr

RESUME

La parole des personnes avec trisomie 21 (T21) présente une altération systématique de l'intelligibilité qui n'a été quantifiée qu'auditivement. Or la modalité visuelle pourrait améliorer l'intelligibilité comme c'est le cas pour les personnes « ordinaires ». Cette étude compare la manière dont 24 participants ordinaires perçoivent des séquences VCV voyelle-consonne-voyelle) produites par quatre adultes (2 avec T21 et 2 ordinaires) et présentées dans le bruit en modalités auditive, visuelle et audiovisuelle. Les résultats confirment la perte d'intelligibilité en modalité auditive dans le cas de locuteurs porteurs de T21. Pour les deux locuteurs impliqués, l'intelligibilité visuelle est néanmoins équivalente à celle des deux locuteurs ordinaires et compensent le déficit d'intelligibilité auditive. Ces résultats suggèrent l'apport de la modalité visuelle vers une meilleure intelligibilité des personnes porteuses de T21.

ABSTRACT

Auditory-visual Perception of VCVs Produced by People with Down Syndrome: a Preliminary Study

The speech of people with Down Syndrome (DS) is systematically altered resulting in an intelligibility loss. This was quantified only auditorily. The visual modality could actually improve intelligibility, as is the case for "ordinary" people. The present study compares the way 24 ordinary participants perceive VCV sequences (vowel-consonant-vowel) produced by four adults (2 with DS and 2 ordinary) and presented in noise in three modalities: auditory, auditory-visual and visual. The results confirm an intelligibility loss in the auditory modality for speakers with DS. However, for the two speakers involved in this study, visual intelligibility is equivalent to that of the ordinary speakers and compensates for the auditory intelligibility loss. These results put forward the importance of integrating multimodality to improve the intelligibility of people with DS.

MOTS-CLES : Parole, Multimodalité, Perception, Trisomie 21, Apport visuel.

KEYWORDS: Speech, Multimodality, Perception, Down Syndrome, Visual input.

1 Introduction

La Trisomie 21 (T21) est une anomalie génétique très fréquente, présente dans toutes les sociétés et causée par la présence d'un chromosome 21 surnuméraire dans le génotype. Cette anomalie induit des troubles anatomiques, physiologiques et cognitifs. Elle est la première cause génétique de déficience intellectuelle (Katz & Lazcano-Ponce, 2008). Le trouble de production de la parole est systématique et n'est pas seulement imputable à la déficience intellectuelle (Kumin, 2006 ; Bunton *et al*., 2007 ; Kent & Vorperian, 2013). Les compétences intellectuelles d'un individu étant souvent inférées de sa capacité à s'exprimer, le trouble de la parole des personnes avec T21 est un enjeu de prise en charge central pour améliorer leur intégration sociale. Or peu d'études ont quantifié l'intelligibilité des personnes avec T21 et la manière dont les personnes « ordinaires »[1] non familiarisées perçoivent leur parole. De plus, les études sur la perception de la parole produite par des locuteurs tout-venant montre que la vision du visage du locuteur améliore la perception de sa parole, notamment en milieu bruité (e.g. Grant & Seitz, 2000). Dans ce cadre, notre objectif est d'évaluer si la vision aide à la perception de la parole produite par des locuteurs avec T21.

Kent et Vorperian (2013) ont publié une revue des travaux de recherche sur la production de la parole par les locuteurs avec T21 depuis les années 1950 selon 4 axes : la voix, l'articulation, la fluence/prosodie et l'intelligibilité. Cette revue montre d'abord que l'intérêt pour la recherche sur la parole des personnes avec T21 a récemment augmenté surtout concernant les aspects articulatoires. Les résultats des études menées sont globalement mitigés et parfois contradictoires du fait du nombre limité de participants et de l'utilisation de méthodologies très variées. Par exemple, bien que la fréquence fondamentale (F0) soit perçue comme étant plus faible chez les personnes avec T21, les résultats d'études acoustiques suggèrent plutôt des valeurs de F0 plus importantes chez ces personnes. La qualité vocale des personnes avec T21 est souvent décrite comme rauque mais cette impression n'est pas quantifiée. Les personnes avec T21 font beaucoup d'erreurs articulatoires et/ou phonologiques dans la production de mots qui rappellent la parole dysarthrique (Bunton *et al*., 2007 ; Kumin, 2006). La littérature rend aussi compte de dysfluences et de différences prosodiques, le bégaiement est notamment fréquent. Plusieurs études, se basant principalement sur des jugements perceptifs et des questionnaires aux familles, rapportent des problèmes d'intelligibilité (Kumin, 2006, 2012 ; Bunton *et al*., 2007). On note de plus la présence de fortes idiosyncrasies. Par exemple, des scores similaires à un test d'intelligibilité de mots peuvent être associés à des profils d'erreurs différents (Bunton *et al*., 2007). Enfin, le trouble de la parole des personnes avec T21 s'observe dès un très jeune âge, avec des différences de développement tels qu'un retard observable sur le babillage canonique ou la production plus fréquente de sons n'étant pas de la parole.

Toujours selon Kent et Vorperian (2013, voir aussi Kumin, 2012), les difficultés de parole décrites ci-dessus ont des origines très variées : problèmes de contrôle moteur, d'audition, de retours somatosensoriels dans la cavité orale, déficience cognitive (avec notamment un déficit du traitement de l'information auditive sérielle), anomalies physiologiques et anatomiques du conduit vocal. La cavité orale des personnes avec T21 est notamment plus petite donnant l'impression d'une macroglossie (*i.e*. une langue anormalement volumineuse) alors que le pharynx possède des caractéristiques de volume et de taille usuelles. La dentition et le palais sont aussi affectés dans la majorité des cas. Ces anomalies ont des conséquences sur la précision de positionnement des

[1] Le terme « ordinaires » est utilisé ici pour « tout-venant » en accord avec la terminologie préconisée par un des organismes financeurs du project : la FIRAH.

articulateurs dans la production de la parole. De plus, la pression intra orale et l'activation musculaire lors de la production de la parole sont supérieures à celles des personnes « ordinaires ». Ces observations suggèrent que la parole demande un effort moteur particulièrement important aux personnes avec T21. Elles s'accordent avec les travaux sur les mouvements des membres qui suggèrent des seuils d'activation musculaire plus hauts que chez les personnes tout-venant, liés à l'observation d'une hypotonie générale au repos (Latash *et al*., 2008).

Concernant les aspects perceptifs, en regard des spécificités physiologiques et anatomiques décrites ci-dessus, le déficit d'intelligibilité de la parole des personnes avec T21 a été essentiellement décrit dans la modalité auditive. On peut dès lors s'interroger sur le rôle de la modalité visuelle dans la perception de la parole des personnes avec T21. Cette modalité est-elle moins touchée que la modalité auditive ? Peut-elle rendre la parole plus intelligible qu'en modalité auditive seule ? Voir son interlocuteur aide à mieux percevoir et détecter sa parole notamment lorsque celle-ci est perturbée comme en milieu bruité (pour une revue, voir Dohen, 2009). Les informations auditives et visuelles sont de plus complémentaires et non redondantes. Summerfield (1987) a comparé les confusions auditive et visuelle des consonnes en anglais et montre que le mode d'articulation est plus robuste en auditif alors que c'est le lieu d'articulation en visuel. L'apport visuel aide ainsi à la perception de la parole produite par des locuteurs « ordinaires », mais qu'en est-il de celle produite par des locuteurs avec T21 ?

Hustad et Cahill (2003) se sont intéressés à l'apport de la modalité visuelle dans la perception de phrases ayant une faible prédictibilité sémantique produites par 5 locuteurs avec des dysarthries moyennes à sévères. La modalité visuelle s'est révélée avoir un apport pour un locuteur seulement, souffrant d'une dysarthrie sévère. Keintz *et al*. (2007) ont réalisé une étude similaire avec 8 patients souffrant de la maladie de Parkinson associée à une dysarthrie et des auditeurs expérimentés et non-expérimentés. Leurs résultats montrent une amélioration significative de l'intelligibilité en modalité audiovisuelle par rapport à en audio seul pour les 3 locuteurs ayant les scores d'intelligibilité les plus réduits, équivalente pour les deux groupes d'auditeurs.

Ce travail s'intéresse à l'apport de la modalité visuelle pour la perception de la parole de deux jeunes adultes avec T21 (avec une bonne intelligibilité) par des personnes tout-venant (n'ayant jamais ou qu'occasionnellement interagi avec des personnes avec T21) en comparaison de celle de deux adultes tout-venant de même genre et sexe.

2 Méthodologie

2.1 Participants au test perceptif

24 personnes ont participé à cette étude, toutes de langue maternelle française (12 femmes ; âge : 25,1 (moy) ±3 (e.t.)). Aucune n'a rapporté de problèmes de vision non corrigé, de trouble phonologique ou de la parole. Toutes ont passé un test audiométrique et aucun déficit auditif n'a été constaté. Ils ont été dédommagés pour leur participation par une carte cadeau de 15€.

2.2 Locuteurs et stimuli

Quatre locuteurs de langue maternelle française ont été sélectionnés parmi les locuteurs enregistrés pour une étude précédente (voir Rochet-Capellan & Dohen, 2015) : 2 personnes « ordinaires » (1 homme, 22 ans – 1 femme, 21 ans) et 2 personnes avec T21 (1 h, 21 ans – 1 f, 19 ans). Les locuteurs avec T21 ont été choisis pour leur relativement bonne intelligibilité d'après un

pré-test de perception et les locuteurs « ordinaires » par appariement en âge et en genre. Le pré-test d'intelligibilité a été réalisé en modalité auditive et sans bruit auprès de participants experts sur un ensemble de 7 locuteurs avec T21.

Les stimuli audio-visuels correspondent à 16 séquences de type Voyelle-Consonne-Voyelle (VCV) avec V=[a] et C={[b], [d], [g], [p], [t], [k], [f], [s], [ʃ], [v], [z], [ʒ], [l], [ʁ], [m], [n]}. Chaque VCV était produit 3 fois et nous avons choisi comme stimulus pour le test perceptif la plus claire de ces trois répétitions aux niveaux auditif et visuel. L'enregistrement des stimuli a été réalisé en chambre sourde. Les participants étaient assis, portaient un micro serre-tête (Sennheiser HP4) et étaient filmés avec une caméra numérique HD (Panasonic HC-X920). Ils devaient répéter des séquences VCV qu'ils entendaient via un haut-parleur. L'audio a été échantillonné à 44100 Hz (carte son externe Focusrite Scarlett 6i6). Chaque fichier audio a été normalisé en intensité à 70dB avec Praat puis ajouté à un bruit de type « cocktail party » (BDBRUIT, Zeiliger *et al*., 1994) avec un rapport signal sur bruit de -4dB. Les fichiers audio ont ensuite été resynchronisés aux vidéos (logiciel FFmpeg : https://www.ffmpeg.org, résolution 960x540 pixels) et créés en 3 versions : audio seul (A, avec image d'un haut parleur), vidéo seul (V) et audio vidéo (AV) soit un total de 192 stimuli : 4 locuteurs x 3 conditions x 16 VCV.

2.3 Procédure pour le test perceptif

Les participants au test perceptif étaient assis devant un bureau, à 60 cm d'un écran de 24 pouces, et portaient un micro casque (Audio Technica BPHS1). Le signal acoustique a été numérisé à 48 kHz (carte son externe Focusrite Scarlett 6i6). L'expérience a été programmée en utilisant la Psychophysics Toolbox (http://psychtoolbox.org/, sous Matlab). Le test perceptif était divisé en trois blocs correspondant aux trois modalités (A, V et AV) de 64 stimuli chacun (16 VCVs x 4 locuteurs). L'ordre des blocs était contrebalancé d'un participant à l'autre. L'ordre des stimuli à l'intérieur de chacun des blocs était aléatoire d'un bloc à l'autre et d'un participant à l'autre.

Les participants étaient informés qu'ils allaient entendre, voir, ou voir et entendre une personne prononcer un son de parole deux fois de suite et qu'ils devraient répéter ce qu'ils avaient perçu. La répétition a été utilisée pour s'affranchir d'ambiguïtés de transcription orthographique. Il leur était précisé d'interpréter les séquences comme des sons n'ayant aucun sens mais aucune information n'était fournie sur la structure du son. Après un entraînement avec des stimuli non bruités, un extrait du bruit leur était présenté. Chaque essai avait la structure suivante : la vidéo intégrant deux répétitions du stimulus était jouée au centre de l'écran. Après avoir vu et/ou entendu les deux répétitions, le participant donnait sa réponse orale puis appuyait sur une touche d'un clavier pour passer au stimulus suivant.

2.4 Transcription des réponses et analyses

Les réponses audio fournies par les participants ont été retranscrites selon le code suivant : avantV1-V1-C-V2-après V2. Chaque partie a été transcrite phonétiquement ou codée comme vide (ex : « brata » pour « ata », avantV1='br' – V1='a' – C='t' – V2='a' – après V2=''). Une consonne non perçue était codée par 'h'. Si la réponse était une voyelle unique (ex : 'a' pour 'ata'), elle était codée en V2 (V1='' – C='h' – V2='a'). Une réponse impossible à retranscrire était codée : V1=' ?' – C=' ?' – V2=' ?'. Une réponse correcte désigne le cas où **V1**, **C** et **V2** correspondent à la stimulation et où **avantV1**='' et **après V2**=''. Notre mesure principale est le nombre de réponses correctes. Une analyse plus détaillée des erreurs sur les consonnes et les voyelles a aussi été réalisée. L'analyse des résultats a été faite avec le logiciel R (https://www.r-

project.org/) avec des analyses de la variance (ANOVA). Les comparaisons post-hoc ont été réalisées avec des tests de Student avec correction de Bonferroni.

3 Résultats

L'analyse montre que 44,2% des réponses fournies par les participants étaient correctes et 54,4% comportaient au moins une erreur. Seules 1,4 % des réponses n'ont pas pu être transcrites. Le groupe de locuteurs et la modalité n'ont pas d'effet significatif sur ce pourcentage ($p > 0,1$). On rappelle que pour les résultats obtenus, le niveau de chance est à 6.25% (16 séquences VCV possibles).

3.1 Réponses correctes

La figure 1 présente les pourcentages de réponses correctes en fonction du groupe de locuteurs (Ord vs T21) et de la modalité de présentation (AV vs A vs V). L'ANOVA dont les résultats sont reportés ci-dessous comportait deux facteurs intra-sujets (groupe de locuteurs et modalité) et un facteur inter-sujets (ordre de présentation des modalités).

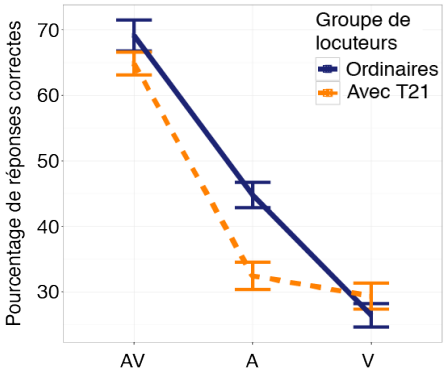

FIGURE 1 : Pourcentage de réponses correctes en fonction du groupe de locuteurs (Ord, T21) et de la modalité

La modalité a un effet significatif sur ce pourcentage ($F(2,36)=263.5$ – $p < 0.001$) : c'est en modalité AV qu'on obtient le plus de réponses correctes suivie des modalités A puis V (A vs AV : $t(23)=-12.8$; $p < 0.001$ – A vs V : $t(23)=5.4$ $p < 0.001$). Les résultats sont globalement meilleurs pour les locuteurs ordinaires que pour les locuteurs porteurs de T21 ($F(1,18)=14.6$ – $p = 0.001$). Ceci dépend cependant de la modalité (groupe de locuteurs * modalité : $F(2,36)=13.6$ – $p < 0.001$) : en modalité A, les locuteurs ordinaires sont significativement mieux perçus que ceux avec T21 ($t(23)=6.7$ – $p < 0.001$) mais ça n'est pas le cas en modalité V ($t(23) = -0.94054$ – $p > 0.9$). En AV, on observe seulement une tendance (non significative après correction pour les comparaisons multiples) à ce que les locuteurs ordinaires soient mieux perçus que ceux porteurs de T21 ($t(23)=2,1$ – $p > 0.1$). L'ordre de passage des modalités a également un effet significatif ($F(5,18)=3,2$ – $p < 0.05$) et interagit avec la modalité ($F(10,36)=3,2$ – $p < 0.01$). L'effet d'ordre s'observe seulement dans la modalité V : les résultats en modalité V sont meilleurs quand elle est passée après la modalité AV plutôt qu'avant.

3.2 Erreurs

Les erreurs ont été classées en trois catégories : insertions avant/après la séquence VCV (avantV1≠'' et/ou aprèsV2≠''), erreurs sur la première/seconde voyelle et erreurs sur la consonne. On notera que ces erreurs ne sont pas exclusives entre elles : il est par exemple possible d'avoir une erreur sur la consonne et une autre sur la ou les voyelles. Les erreurs les plus fréquentes sont celles sur la consonne (70,1% du nombre total d'erreurs) et il y a relativement peu d'erreurs sur V1 et V2 (respectivement, 7,1 et 7%) et d'insertions avant ou après la séquence VCV (7,9% pour les deux).

Réponses sur C – Les réponses sur C ont été classées en catégories : correcte, confusion (avec une autre consonne) et autre (*e.g.*, ajout d'une ou plusieurs consonnes). La figure 2 présente la

répartition des réponses sur C en fonction de la modalité, du groupe de locuteurs et du type de réponse. Les réponses correctes et les confusions représentent 94,5% du total des réponses sur la consonne. L'ANOVA a été réalisée sur les pourcentages d'erreurs avec trois facteurs intra-sujets : groupe de locuteurs, modalité et type d'erreur.

Les confusions avec une autre consonne sont les erreurs les plus fréquentes (F(1,23)=228,3 – p<0,001) et il y en a significativement plus en modalités A et V qu'en AV (F(2,46)= 206,4 – p <0,001). On constate qu'il n'y a globalement pas de différence entre les groupes (F(1,23)=2,4 p= 0,1) bien qu'en modalité A, il y ait plus d'erreurs pour les locuteurs avec T21 que pour les ordinaires (groupe de locuteur * modalité : F(2,46)=3,6 – p<0,05).

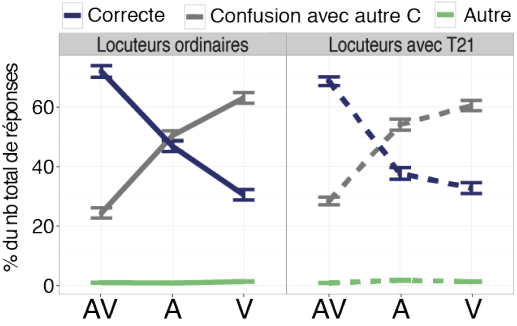

FIGURE 2 : Répartition des réponses sur la consonne en fonction du type de réponse, du groupe de locuteurs (Ord vs T21) et de la modalité (AV, A, V).

Erreurs sur V1 et V2 – Elles ont été classées par catégorie : confusion avec une autre voyelle et voyelle non perçue. La figure 3 fournit les pourcentages d'erreur sur V1 (haut) et sur V2 (bas) en fonction de la modalité, du groupe de locuteurs et du type d'erreur. Les ANOVA réalisées comportent trois facteurs intra-sujets : groupe de locuteurs, modalité et type de réponse.

Erreurs sur V1 – Il y a significativement moins d'erreurs sur V1 en modalité AV qu'en A et V (F(2,46)=4 – p<0,05). Les erreurs sont significativement plus fréquentes pour les locuteurs avec T21 que chez les ordinaires (F(1,23)=22,1 – p<0,001) mais seulement en modalité A (groupe de locuteur * modalité : F(2,46)=11,8 – p<0,001). Il n'y a pas de différence entre les types d'erreurs (F(1,23)=0,001 – p=0,98). En modalité A, alors que pour les locuteurs ordinaires on observe plus de confusions que de non-perceptions, c'est l'inverse pour les locuteurs avec T21 (groupe de locuteur * modalité * type de réponse : F(2,46)=6,8 – p<0,001).

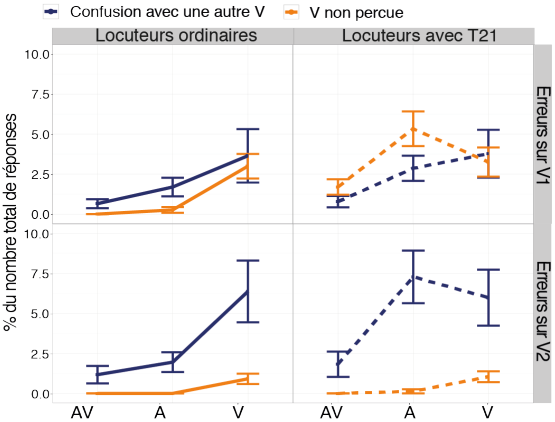

FIGURE 3 : Pourcentage des erreurs sur V1 (haut) et V2 (bas) en fonction du type d'insertion, du groupe de locuteurs (Ord vs T21) et de la modalité (AV, A, V).

Erreurs sur V2 – Il y a significativement moins d'erreurs sur V2 en modalité AV qu'en A et V (F(2,46)=6,2 – p<0,005). Les erreurs sont significativement plus fréquentes pour les locuteurs avec T21 que pour les ordinaires (F(1,23)=8 – p<0,01) mais seulement en modalité A (groupe de locuteur * modalité : F(2,46)=6,6 – p<0,005) et seulement pour les confusions avec une autre voyelle (groupe de locuteur * modalité * type d'erreur : F(2,46)=5,3 – p<0,01). Les confusions avec une autre voyelle sont significativement plus fréquentes que les non-perceptions (F(1,23)=8,3 – p<0,01) mais surtout en modalités A et V (modalité * type d'erreur : F(2,46)=4,1 – p=0,03).

Insertions avant V1 ou après V2 – Elles ont été classées par catégorie : insertion d'une consonne, insertion d'une voyelle et autre (*e.g.*, insertions multiples). La figure 4 fournit les pourcentages d'insertions avant V1 (haut) et après V2 (bas) en fonction de la modalité, du groupe de locuteurs et du type d'insertion. Les ANOVA réalisées comportent trois facteurs intra-sujets : groupe de locuteurs, modalité et type d'insertion.

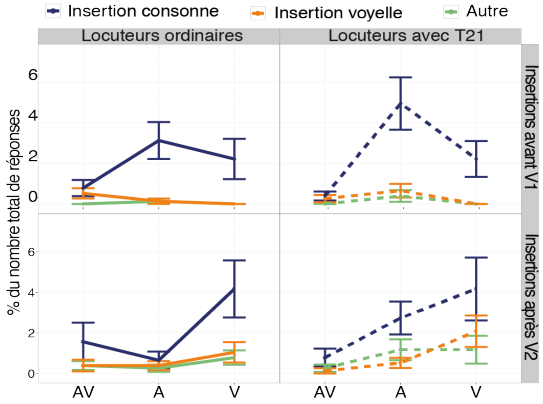

FIGURE 4 : Pourcentage des insertions avant V1 (haut) et après V2 (bas) en fonction du type d'insertion, du groupe de locuteurs (Ord vs T21) et de la modalité (AV, A, V).

Insertions avant V1 – Les insertions avant V1 sont plus fréquentes en A et V qu'en AV, ce type d'erreur étant quasi inexistant en modalité AV (modalité : $F(2,46)=5.8$ – $p < 0.01$). La catégorie d'erreurs la plus fréquente est l'insertion d'une consonne, les autres types d'insertion ne se produisant quasiment pas ($F(2,46)=213.8$ – $p = 0.001$). Il n'y a pas de différence entre les groupes de locuteurs ($F(1,23)=1.8$ – $p = 0.2$). Cependant, les insertions d'une consonne sont plus fréquentes en A qu'en V pour les locuteurs porteurs de T21 mais pas pour les locuteurs ordinaires (groupe de locuteurs * modalité : $F(2,46)=5.8$ – $p < 0.01$).

Insertions après V2 – Il y a plus d'insertions après V2 en modalité V qu'en A et AV ($F(2,46)=4.1$ – $p < 0.05$) et pour les locuteurs porteurs de T21 que pour les locuteurs ordinaires ($F(1,23)=4.8$ – $p < 0.05$). On ne constate pas d'effet du type d'insertion ($F(2,46)=2.9$ – $p > 0.05$). Pour les locuteurs ordinaires, le pourcentage d'insertions après V2 est équivalent en modalités A et AV mais plus important en V alors que pour les locuteurs porteurs de T21, le nombre d'insertions suis un ordre V>A>AV (modalité * groupe de locuteurs : $F(2,46)=3.8$ – $p < 0.05$).

4 Conclusions et discussion

Cette étude compare la perception multimodale, par des participants tout-venant, de séquences Voyelle-Consonne-Voyelle (VCV) produites par des locuteurs avec T21 à celles produites par des locuteurs « ordinaires ». L'objectif était d'évaluer si le fait de voir le locuteur avec T21 parler en plus de l'entendre aide à mieux percevoir sa parole. Comme il a été largement rapporté dans la littérature (cf. Dohen, 2009 pour une revue), les résultats de cette étude montrent que, quel que soit le locuteur (avec T21 ou « ordinaire »), la parole mélangée à du bruit est mieux perçue en modalité AV qu'en modalité A puis V (AV>A>V). Tous groupes de locuteurs confondus, les erreurs de loin les plus fréquentes sont celles sur la consonne qui impliquent une confusion de celle-ci avec une autre consonne : il s'agissait en effet du seul phonème qui variait au cours du test (Voyelle=[a]). Globalement, les pourcentages de bonnes réponses sont faibles en modalités A et V (moins de 50% de bonnes réponses dans les deux cas) semblant suggérer une très faible intelligibilité pour les deux groupes de locuteurs. Ces pourcentages sont cependant bien supérieurs au hasard (6,25%) et sont liés au nombre important de réponses possibles. Nous nous intéressons dans la suite aux résultats dans chacune des modalités.

Modalité A – Les locuteurs « ordinaires » sont significativement mieux perçus que ceux avec T21, ce qui va dans le sens des études rapportant un déficit de l'intelligibilité auditive chez les

personnes avec T21 (Kumin, 2006 ; Bunton *et al*., 2007; Kent & Vorperian, 2013). En dehors du fait que tous les types d'erreurs soient globalement plus fréquents chez les locuteurs avec T21, il y a plusieurs cas où les différences sont plus importantes que la moyenne. Les insertions d'une consonne avant ou après la séquence VCV sont ainsi particulièrement plus fréquentes pour les locuteurs avec T21 que pour les tout-venants. La deuxième voyelle est de plus beaucoup plus souvent confondue avec une autre voyelle pour les locuteurs avec T21. Cela pourrait s'interpréter par une difficulté accrue à séparer la parole des locuteurs avec T21 du bruit ambiant. On rappelle que le bruit ambiant est de type « cocktail party » c'est-à-dire composé d'une multitude d'autres signaux de parole qui peuvent être confondus avec ou assimilés au signal cible à identifier. Notons que cette séparation des sources est d'autant plus complexe en modalité A. On observe de plus que la première voyelle est beaucoup plus souvent non perçue que confondue chez les locuteurs avec T21 alors que c'est l'inverse chez les tout-venant. Ceci suggère que les locuteurs avec T21 ont des difficultés à produire une voyelle intelligible auditivement à l'initialisation d'une séquence, le manque d'intelligibilité étant ici accentué par la présence de bruit.

Modalité AV – Le fait que l'écart d'intelligibilité entre les groupes soit beaucoup plus faible qu'en modalité A (non significatif après corrections pour les comparaisons multiples) suggère que la modalité visuelle permet de compenser au moins en partie, et pour les deux locuteurs avec T21 de cette étude, le déficit d'intelligibilité auditive. En modalité AV, les types d'erreurs ne dépendent pas du groupe de locuteurs sauf pour la première voyelle pour laquelle on observe significativement plus de non-perceptions pour les locuteurs avec T21.

Modalité V – Dans le cadre de cette étude, utilisant des séquences VCV et testant seulement deux locuteurs avec T21, on observe que les locuteurs avec T21 sont aussi intelligibles visuellement que les tout-venants. Notons que lorsque la modalité V est présentée avant AV, les pourcentages de réponses correctes sont moins bons quel que soit le groupe de locuteur. Cet effet d'ordre est probablement lié au fait que les stimuli sont mieux perçus en AV : les participants ont pu mémoriser l'association audio-visuelle d'un stimulus donné, la confrontation au stimulus visuel peut ensuite servir d'amorce à la réponse pour produire la bonne réponse. Cet effet de l'avantage de l'ordre AV-V pour la modalité V n'a, à notre connaissance, pas été clairement mis en évidence par les études antérieures sur les personnes « ordinaires ». Ce résultat, observé pour les deux groupes de locuteurs, suggère que l'inclusion de personnes avec T21 pourrait augmenter l'attention des participants pour l'information visuelle en condition AV.

Cette étude préliminaire suggère un apport de la modalité visuelle pour améliorer l'intelligibilité des personnes avec T21. Les difficultés sur la première voyelle observées en modalité A et qui se maintiennent en AV font écho aux résultats en contrôle moteur suggérant un seuilplus haut d'activation musculaire chez les personnes avec T21 (Latash *et al*., 2008). Cette difficulté à initialiser le mouvement pourrait diminuer l'énergie sur la première voyelle et rendre son exécution plus difficile que la deuxième voyelle. L'inertie de retour au repos pourrait aussi être plus importante, ce qui pourrait contribuer à expliquer les insertions en fin de VCV. Les erreurs sur la consonne sont probablement liées à l'anatomie du conduit vocal et au manque de précision des points d'articulation portés par la langue. La dépendance des résultats aux lieux et modes d'articulation de la consonne reste à évaluer. Il faudra également corroborer ces résultats pour d'autres voyelles. Des travaux antérieurs ayant montré un apport plus important de la modalité visuelle dans la parole dysarthrique seulement pour les dysarthries sévères (Hustad & Cahill, 2003 ; Keintz *et al*., 2007), on peut aussi s'interroger sur l'effet du choix des locuteurs sur nos résultats : est-ce que l'apport du visuel serait encore plus important pour des personnes avec T21 avec une moins bonne intelligibilité ?

Remerciements

Ce travail a reçu le soutien du European Research Council dans le cadre du 7ème Programme de la Communauté Européenne (FP7/2007-2013 Grant Agreement no.339152- “Speech Unit(e)s”). Il s’inscrit de plus dans le cadre du projet « Communiquons Ensemble » subventionné par la Fondation Internationale de la Recherche Appliquée sur le Handicap (FIRAH). Les auteurs remercient les participants, l’Association de Recherche et d’Insertion Sociale des Trisomiques (ARIST) et les professionnels de l’ESAT-SAJ de l’ARIST.

Références

BUNTON, K., LEDDY, M., & MILLER, J. (2007). Phonetic intelligibility testing in adults with Down syndrome. *Down Syndrome Research and Practice*, 12(1), 1–4.

DOHEN, M. (2009). Speech through the ear, the eye, the mouth and the hand. Lecture Notes in *Computer Science* (including Subseries Lecture Notes in *Artificial Intelligence* and Lecture Notes in *Bioinformatics*), 5398 LNAI, 24–39.

GRANT, K. W., & SEITZ, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, 108(3), 1197-1208.

HUSTAD, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language & Hearing Research*, 51(3), 562–573.

HUSTAD, K. C., & CAHILL, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. American Journal of Speech-Language Pathology.

KATZ, G., & LAZCANO-PONCE, E. (2008). Intellectual disability: definition, etiological factors, classification, diagnosis, treatment and prognosis. *salud pública de méxico*, 50, s132-s141.

KEINTZ, C. K., BUNTON, K., & HOIT, J. D. (2007). Influence of visual information on the intelligibility of dysarthric speech. American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association, 16(3), 222–34.

KENT, R. D., VORPERIAN, H. K., KREIMAN, J., & MAASSEN, B. A. M. (2013). Speech Impairment in Down Syndrome: A Review, *Journal of Speech, Language & Hearing Research*, 56(1), 178–210.

KUMIN, L. (2006). Speech intelligibility and childhood verbal apraxia in children with Down syndrome. *Down’s Syndrome, Research and Practice*, 10(1), 10–22.

KUMIN, L. (2012). *Early communication skills for children with Down syndrome: A guide for parents and professionals*. États-Unis : Woodbinehouse

LATASH, M., WOOD, L., & ULRICH, D. (2008) *What is currently known about hypotonia, motor skill development, and physical activity in Down syndrome*. Down Syndrome Education Online (http://www.down-syndrome.org/reviews/2074/).

ROCHET-CAPELLAN, A., & DOHEN, M. (2015). Acoustic characterisation of vowel production by young adults with Down syndrome. Actes de *The International Congress of Phonetic Scences*.

SUMMERFIELD, Q. (1987). Comprehensive Account of Audio-Visual Speech Perception. In: Dodd, B., Campbell, R. (eds.), *Hearing by Eye: The Psychology of Lip-reading*, pp. 3--51. Lawrence Erlbaum Associates, Hillsdale, NJ.

ZEILIGER J., S. J.-F. (1994). BDBRUIT, une base de données parole de locuteurs soumis à du bruit. Dans Actes des *10ièmes journées d'Étude sur la Parole* (pp. 287-290).