



De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication

Lionel Fontan

► **To cite this version:**

Lionel Fontan. De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication. Linguistique. Université Toulouse le Mirail - Toulouse II, 2012. Français. <NNT : 2012TOU20113>. <tel-00797883>

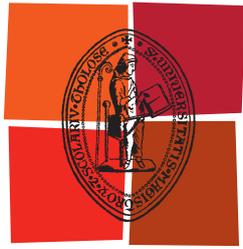
HAL Id: tel-00797883

<https://tel.archives-ouvertes.fr/tel-00797883>

Submitted on 7 Mar 2013

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.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 2 Le Mirail (UT2 Le Mirail)

Cotutelle internationale avec :

Présentée et soutenue par :
Lionel FONTAN

Le jeudi 8 novembre 2012

Titre :

De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole
pathologique en situation de communication

ED CLESCO : Sciences du langage

Unité de recherche :

Unité de recherche interdisciplinaire OCTOGONE (E.A. 4156)

Directeur(s) de Thèse :

Robert RUIZ, MCF HDR, Université de Toulouse II - Le Mirail

Pascal GAILLARD, MCF, Université de Toulouse II - Le Mirail

Virginie WOISARD, PH, CHU Toulouse Rangueil

Rapporteurs :

Lorraine BAQUE, PU, Université Autonome de Barcelone

Serge PINTO, CR, Laboratoire Parole et Langage, Aix-en-Provence

Autre(s) membre(s) du jury :

Renaud GARREL, PU, PH, Université Montpellier I, CHRU Montpellier

DE LA MESURE DE L'INTELLIGIBILITÉ À L'ÉVALUATION
DE LA COMPRÉHENSION DE LA PAROLE PATHOLOGIQUE
EN SITUATION DE COMMUNICATION

Lionel Fontan

Novembre 2012

Remerciements

Je tiens tout d'abord à adresser mes sincères remerciements à toutes les personnes qui ont accepté de diriger et d'encadrer mon travail : Claude Legros, Robert Ruiz, Élie Serrano, Pascal Gaillard et Virginie Woisard. Ils ont tous été présents quand il le fallait, et j'ai eu la chance de bénéficier tout au long de ma thèse d'un suivi attentif et rigoureux.

Je suis également très reconnaissant à Lorraine Baqué et à Serge Pinto d'avoir accepté d'être les rapporteurs de ce travail de thèse, et à Renaud Garrel d'avoir bien voulu participer à mon jury de soutenance.

Je remercie le personnel de l'Unité de la Voix et de la Déglutition du service d'Oto-Rhino-Laryngologie de l'hôpital Larrey à Toulouse. Leur gentillesse et leur aide m'ont été très précieuses. Merci également à Julie Bordenave et Astrid Ros, qui ont effectué les passations de l'étude de faisabilité et de validation du test EloKanz dans le cadre de leur mémoire d'orthophonie.

À toutes les autres personnes qui m'ont apporté leur aide et leur expertise pour ce travail : je pense notamment à Mireille Costes, Pascale Fichaux-Bourin, Stéphanie Grand, Saïd Jmel, Jean-Luc Nespoulous, Michèle Puech, Julien Tardieu...

Un grand merci aux 156 personnes qui ont accepté de participer aux expériences relatées dans ce volume...

Au laboratoire Octogone dans son ensemble, avec une pensée particulière pour mes collègues doctorants et jeunes chercheurs : Charlotte Alazard**¹, Karine Aura, Audrey Bathélémy, Aurore Berland, Marie-Mandarine Colle-Quesada, Claire Del Olmo, Carine De Martin, Sören Frappart, Mélanie Jucla, Yolibeth Machado Key, Cynthia Magnen**¹, Vanda Marijanovic, Émilie Massa, Margarita Muñoz Garcia, Marie Lacabanne, Martha Serrano, Charlotte Vallat, Nikki Van Leeuwen, Lucie Vialettes, Margarita Zubareva et tous les autres... Merci pour votre soutien, et pour les moments passés avec vous.

Merci aux collègues du laboratoire CLLE, dont, en particulier : Clémentine Adam, Caroline Atallah, Laurie Buscail, Nathalie Dehaut, Buddy Dirat, Fanny Lallemann, Aurélie Guerrero, Christel Le Bellec, Lydia-Mai Ho-Dac, Stéphanie Lopez, François Morlane-Hondère, Christelle Pêcher, Aurélie Picton, Marie-France Roquelauze, Caitlin Smith, Cécile Viollain et, *last but not least*, Marianne Vergez-Couret.

1. ces personnes ayant partagé mon bureau pendant des mois, voire des années pour le cas de C. Alazard, leur mérite augmente de manière hautement significative ($P < 0,01$).

Merci aux autres collègues de l'Université de Toulouse II et d'ailleurs – Vanessa Andréani, Lolla Choucavy, Estelle Delpech, Marjorie « Hennaff » Raufast... qui m'ont soutenu.

Je pense aussi au département de Sciences du Langage de l'Université de Toulouse II, qui m'a accueilli pendant trois années en tant que moniteur. Cette expérience d'enseignement, menée de pair avec le travail de thèse, a été très enrichissante.

À Chanida, qui est la personne qui m'a le plus supporté – dans tous les sens du terme – pendant ce travail... Merci pour tout !

À ma famille qui m'a soutenu tout au long de la thèse, avec une pensée particulière pour ma mère, Monique Montserret, et ma grand-mère Arlette Fontan.

À tous mes amis sans qui je n'aurais pas tenu le coup... Je pense à tous mes amis bigourdans : Christophe Gonzalez « eth Gonz », Nathalie Larré-Larrouy « era Nath », Jean-Louis Oribès « eth Luis », Vincent Sarrazin « eth Vinz » et tous les autres, ainsi que mes amis de Toulouse et d'ailleurs : Aurélie Baudé, Estelle Marie-Sainte, Alice Rambaud...

À Marcel Dalens

Sommaire

Introduction	1
I État de l'art et problématique	7
<hr/>	
1 Développement des tests d'intelligibilité de la parole dans les domaines de l'acoustique et des télécommunications	11
1.1 Introduction	12
1.2 De l'élaboration empirique à la standardisation des premières mesures subjectives de l'intelligibilité	14
1.3 Développement de nouvelles méthodes de mesure subjective de l'intelligibilité	18
1.4 Vers des mesures objectives et prédictives	23
1.5 Synthèse et discussion	28
2 Méthodes d'évaluation et de mesure de l'intelligibilité pour les troubles pathologiques de production de la parole (TPPP)	35
2.1 Introduction	36
2.2 Tâches d'identification de stimuli	38
2.3 Tâches d'évaluation sur des échelles subjectives	40
2.4 Facteurs influençant les scores d'intelligibilité	44
2.5 Synthèse et discussion	51
3 Problématique : intelligibilité et compréhension de la parole en situation de communication	55
3.1 Introduction	56
3.2 Perspectives psycholinguistiques sur la perception et la compréhension de la parole	59
3.3 Quelle perspective sur la compréhension de la parole dans notre travail ?	78
3.4 Relation entre les mesures subjectives d'intelligibilité et la compréhension de la parole : état de la question et hypothèses	80
4 Choix préliminaires et développement du test EloKanz	85
4.1 Choix d'une procédure pour l'évaluation de la compréhension de la parole	86
4.2 Création du logiciel EloKanz	96

II	Expériences	109
<hr/>		
5	Études préliminaires	111
5.1	Sélection du jeu d'images et études de l'accord sur le nom	112
5.2	Test de l'interface locuteur d'EloKanz	132
5.3	Conclusion et discussion	139
6	Intelligibilité et compréhension de la parole dans le bruit	143
6.1	Objectifs et hypothèses	145
6.2	Matériel	148
6.3	Population et procédure	157
6.4	Résultats	161
6.5	Discussion	168
6.6	Conclusion	175
7	Études de faisabilité clinique et de validation du test EloKanz	177
7.1	Introduction	179
7.2	Étude de la faisabilité côté <i>locuteur</i> : enregistrement des patients . . .	180
7.3	Étude de la faisabilité côté <i>auditeur</i> : écoute des patients	205
7.4	Éléments de validation du test EloKanz	220
	Discussion générale : des différences entre l'intelligibilité et la com- préhension de la parole en situation de communication	233
	Conclusion	241
	Bibliographie	245
	Index des citations	264
	Liste des figures	265
	Liste des tableaux	270
	Annexes	277
<hr/>		
A	Première expérience d'accord sur le nom	277
A.1	Sujets testés	277
A.2	Interface utilisée	279
A.3	Noms les plus fréquemment attribués pour chaque image	281
A.4	Noms correspondant à d'autres concepts autres que ceux visés par les images	284
A.5	Noms d'images désignées par des synonymes	286
A.6	Images donnant lieu à des noms/concepts super ordonnés	286
A.7	Images donnant lieu à des noms/concepts subordonnés	287

B	Pré-test pour la seconde expérience d'accord sur le nom	291
B.1	Consigne	291
B.2	Interface auditeur montrée aux sujets	292
B.3	Résultats	293
C	Seconde expérience d'accord sur le nom, et test de l'ergonomie de l'interface <i>locuteur</i> d'EloKanz	295
C.1	Sujets	295
C.2	Consignes utilisées	296
D	Expérience sur l'intelligibilité et la compréhension de la parole dans le bruit	299
D.1	Sujets	299
D.2	Phrases utilisées comme stimuli	301
D.3	Cabine Petra-2	302
D.4	Algorithme général pour la concaténation des fichiers audio	304
D.5	Consignes	305
E	Études de faisabilité et de validation du test Elokanz	307
E.1	Patients inclus dans l'étude	307
E.2	Grille d'observation pour l'enregistrement des patients	309
E.3	Questionnaire post-test (patients et auditeurs)	312
E.4	Lettre d'informations aux patients	313
E.5	Formulaire de consentement éclairé	314
E.6	Feuille de passation du <i>Mini-Mental State - MMSE</i>	315
E.7	Salle d'enregistrement et d'écoute	316
E.8	Consigne donnée aux patients	317
E.9	Questions et remarques faites par les patients durant la passation	319
E.10	Commentaires post-test des patients	321
E.11	Grille d'observation pour l'écoute des patients par les auditeurs	322
E.12	Consigne donnée aux auditeurs	326
E.13	Sujets auditeurs inclus dans l'étude	328
E.14	Questions et remarques des auditeurs liées aux consignes	329
E.15	Questions et remarques des auditeurs à propos de l'ergonomie	330

Introduction

Ce travail de recherche répond à un besoin exprimé par des médecins et des orthophonistes travaillant auprès de patients souffrant de *troubles pathologiques de production de la parole* (TPPP) : anomalies d'origine structurelle (anatomique) ou neurologique engendrant des difficultés d'élocution. La prise en charge et le suivi des patients nécessite de disposer de méthodes d'évaluation fiables et reproductibles permettant de quantifier leurs performances orales. Habituellement ces évaluations sont effectuées grâce à différents types de *tests d'intelligibilité*, qui impliquent la participation d'un jury d'auditeurs plus ou moins nombreux, et plus ou moins habitués à ce type de parole.

De nombreuses études ont porté sur les difficultés accompagnant ces méthodes d'évaluation, concernant :

- leur fiabilité : une variabilité importante peut être observée auprès de populations d'auditeurs différant par exemple de par leur familiarité avec les locuteurs, avec le type de pathologie concernée, avec la tâche qui leur est octroyée ou encore avec les stimuli employés. De même, les scores d'intelligibilité dépendent fortement de la méthode mise en œuvre et du matériel utilisé ;
- leur validité : la nature des liens entre les scores d'intelligibilité de la parole et l'aptitude du locuteur à se faire comprendre dans une situation de communication naturelle sont encore peu connus (Hustad, 2008) ;
- leur « praticabilité » en milieu clinique : les procédures d'évaluation requièrent généralement des moyens – en termes de temps et de personnel – qui ne sont pas nécessairement à disposition dans les services.

Notre travail s'est principalement concentré sur la question de la *validité* de ces méthodes, c'est-à-dire sur la capacité des procédures de test d'intelligibilité à « capturer » la performance de communication verbale du locuteur. En d'autres termes, nous nous sommes attaché à l'étude *de la relation existant entre l'intelligibilité et la compréhension de la parole en situation de communication*. Cette question est d'une importance cruciale pour la recherche comme pour la pratique clinique. En effet les scores d'intelligibilité sont non seulement utilisés pour suivre les progrès des patients dans le temps – et donc pour juger de l'efficacité de thérapies ou d'interventions, mais aussi pour prendre des décisions aussi importantes que la prise en charge thérapeutique d'une personne ou, au contraire, de l'arrêt d'un traitement (Ertmer, 2010; Skahan *et al.*, 2007). Il est donc particulièrement important de s'assurer de la capacité des scores d'intelligibilité à prédire la performance du locuteur en situation de communication.

Un premier élément qui peut nous amener à douter de ce pouvoir de prédiction est l'objectif pour lequel les tests d'intelligibilité ont été conçus au départ. Comme nous l'exposons dans le chapitre 1, historiquement, ce sont des ingénieurs en télécommunication qui ont été les premiers à développer des méthodes de mesure de l'intelligibilité de la parole. Leurs travaux répondaient à de nouveaux besoins liés au développement des lignes téléphoniques commerciales au début du XX^{ème} siècle, et à l'élaboration de systèmes de communication radio entre la Première et la Seconde Guerre Mondiale (Hawley, 1977c). L'augmentation des distances couvertes par les lignes téléphoniques causait des dégradations importantes du signal de parole, alors que dans le domaine militaire les difficultés de communication étaient principalement dues à l'environnement bruyant dans lequel se trouvaient les interlocuteurs (par ex. les cockpits d'avion¹). L'optimisation des systèmes de radio ou télé transmission était alors d'une importance cruciale, et avec elle le besoin de recourir à des procédures d'évaluation du succès de la communication verbale (Kent *et al.*, 1989; Schiavetti, 1992; Weismer et Martin, 1992).

Les problèmes auxquels étaient confrontés les ingénieurs étaient donc le fait de dégradations acoustiques intervenant dans les *canaux de communication*. Or, ce sont les mêmes méthodes qui ont été ensuite portées au domaine clinique pour évaluer cette fois la compétence de communication *de locuteurs* (chapitre 2). Ces méthodes consistent généralement pour le locuteur à lire des stimuli linguistiques (mots, phrases), et pour un ou des auditeur(s) à retranscrire les paroles entendues. Le score d'intelligibilité correspond alors au pourcentage d'unités (généralement de mots) dont tous les phonèmes ont été correctement retranscrits par le ou les auditeur(s).

Dans le chapitre 3 nous posons la question du lien existant entre ces mesures, fondées sur le taux de transfert de *code linguistique*, et la performance de communication verbale du locuteur en situation de communication. En effet, si les ingénieurs en télécommunication étaient intéressés par la capacité des canaux de communication à transmettre les unités signifiantes de la parole, la question centrale qui se pose dans le cas des TPPP est celle du transfert de *sens*. La question de la relation entre ces deux types de transferts est loin d'être évidente, comme nous l'illustrons à travers un bref état de l'art sur les études psycholinguistiques de la perception et de la compréhension de la parole. Ainsi la façon dont l'être humain apparie les signaux de parole perçus aux unités linguistiques suscite encore de nombreux débats, tout comme la façon dont il intègre les informations linguistiques et contextuelles lors de la compréhension d'énoncés verbaux.

De ce point de vue, la capacité de tâches orientées vers la perception du code linguistique, comme les tests d'intelligibilité, à rendre compte de l'aptitude du locuteur à *se faire comprendre* par des tiers en situation de communication est discutable. Plusieurs études ont ainsi tenté de rendre compte de la relation existant entre les scores obtenus par des tâches de retranscription orthographique, et les scores obte-

1. l'article « Speech Intelligibility in Naval Aviation » de Steer (1945) est un exemple typique des recherches menées à cette époque.

mus par des tâches visant à observer la compréhension d'énoncés produits par des patients (Beukelman et Yorkston, 1979; Hustad, 2008). Cependant ces études se sont heurtées à deux problèmes principaux :

- le premier concerne la représentativité de la tâche de compréhension utilisée. Dans les études existantes, les auditeurs n'étaient pas engagés dans une véritable situation de communication, dans laquelle ils auraient été amenés à répondre aux énoncés entendus par un comportement quelconque (Hustad, 2008) ;
- la deuxième difficulté provient du fait qu'il n'existe aucun moyen de contrôler de manière objective la sévérité des troubles dont souffrent les patients. Ceci cause un problème méthodologique majeur dans la mesure où cette variable influence à la fois les scores d'intelligibilité et de compréhension – masquant ainsi la véritable nature de leur relation.

L'apport principal de notre travail s'inscrit en réponse directe aux deux points que nous venons d'évoquer. Nous avons tout d'abord élaboré une méthode permettant d'évaluer la compréhension de l'auditeur en observant ses réactions à des énoncés verbaux (chapitre 4). Cette méthode s'appuie sur une conception pragmatique de la compréhension de la parole, et consiste à demander à l'auditeur de répondre à des consignes verbales en effectuant des actions sur des objets. La compréhension de l'auditeur est alors jugée en fonction de l'adéquation entre ses actions et les consignes données par le locuteur. Ce protocole est directement inspiré de tests utilisés en aphasiologie et en psychologie animale (*cf.* notamment Savage-Rumbaugh *et al.*, 1993). Nous avons implémenté cette procédure dans un logiciel informatique, que nous avons nommé « EloKanz ». Outre le travail de programmation informatique, ce développement nous a demandé d'effectuer plusieurs études préliminaires que nous relatons dans le chapitre 5.

Ensuite, pour dépasser le problème posé par le contrôle de la sévérité des TPPP, nous avons choisi d'utiliser des stimuli de parole non pathologique, que nous avons dégradés artificiellement et de manière contrôlée en les diffusant avec du bruit (chapitre 6). En effet puisque les hypothèses des différences entre l'intelligibilité et la compréhension de la parole en situation de communication concernent uniquement les processus cognitifs en jeu chez l'auditeur, rien ne nous empêchait de recourir à ce paradigme pour étudier la relation entre les deux types de tâches.

Nos résultats montrent que les scores de retranscription orthographique ne permettent pas de prédire les performances de compréhension de la parole en situation de communication. Les implications cliniques de ce résultat sont importantes, dans la mesure où dans la grande majorité des cas la performance de communication verbale des patients atteints de TPPP est aujourd'hui évaluée à l'aide de tests d'intelligibilité. Nos données constituent donc un argument fort en faveur du développement de tests spécialement destinés à évaluer cette dimension particulière.

Toutefois, le recours à des tests s'approchant de situations de communication « naturelles », tels que celui que nous avons conçu à travers le logiciel EloKanz, peut impliquer l'utilisation de matériel et de consignes qui ne sont pas forcément adaptés aux contraintes du domaine clinique. Pour étudier cette question, nous avons

mené dans un dernier temps une étude exploratoire de faisabilité et de validation du logiciel EloKanz en milieu clinique, dans une unité hospitalière destinée aux troubles de la parole (chapitre 7).

Première partie

État de l'art et problématique

*Nous ne pouvons qu’imaginer de quelle manière le domaine de l’intelligibilité de la parole serait aujourd’hui différent si la neurologie, la linguistique, la psychologie perceptive, la phonétique, ou l’informatique avaient été suffisamment avancées et instrumentées pour être à même de mener le champ avant 1940.*² (Hawley, 1977a)

Il est difficile d’appréhender ce que recouvre la notion d’*intelligibilité de la parole* à travers les différentes définitions – explicites ou implicites – proposées dans la littérature scientifique. En effet, si les définitions ont en commun d’avoir trait à l’évaluation de la réception de signaux de parole par un ou plusieurs auditeur(s), elles se distinguent de par la nature des *processus cognitifs* censés être en jeu chez les auditeurs :

- la *reconnaissance* d’unités linguistiques (ex. Steeneken, 2001; Doyle et Leeper, 1997) ;
- l’*identification* d’unités linguistiques (ex. Saunders et Cienkowski, 2002; Searl *et al.*, 2001) ;
- ou encore la *compréhension* d’unités linguistiques (ex. Özsancak, 2001; Yunusova *et al.*, 2005).

Également, les *unités* sur lesquelles portent ces processus sont de natures diverses :

- *phonèmes* (ex. Searl *et al.*, 2001; Doyle *et al.*, 1988) ;
- *mono-* ou *polysyllabes* dénués de sens (« logatomes » – ex. Jongmans *et al.*, 2006; Benoit *et al.*, 1994) ;
- *mono-* ou *polysyllabes* pourvus de significations (ex. Doyle et Leeper, 1997; Miralles et Cervera, 1995) ;
- *phrases* (ex. Cienkowski et Speaks, 2000; Garcia et Dagenais, 1998) ;
- ou encore *discours* (ex. De Bodt *et al.*, 2002; Cox et McDaniel, 1989).

Cependant cet ensemble hétéroclite de références et de pratiques reflète bien moins une diversité dans la définition de l’intelligibilité de la parole que la dépendance de toute *mesure* à une *méthode* particulière. En effet, pour pouvoir être définie

2. « One can only speculate how the field of speech intelligibility would be different today if the sciences of neurology, linguistics, perceptual psychology, phonetics, or computers had been sufficiently advanced and instrumented so that they might have led the field before 1940. ».

précisément, l'intelligibilité de la parole doit être appréhendée comme une *mesure*, c'est-à-dire comme une grandeur relative dont l'essence ne peut être saisie sans tenir compte de la méthode servant à l'évaluer (Kent *et al.*, 1989; Özsancak, 2001). Ceci est très clairement exprimé par D. Sadek-Khalil (1997, p. 52-53) qui, en vue de la préparation d'un colloque, avait demandé à plusieurs conférenciers de définir « en quelques lignes » l'intelligibilité de la parole :

[...] j'ai compris dès le début que j'avais fait une erreur (erreur qui s'est avérée fructueuse) : il est impossible, en effet, de définir *en quelques lignes* l'intelligibilité comme on donne couramment une définition. Les réponses auraient peut-être été conformes à mes prévisions si j'avais demandé une définition par la méthode de mesure de l'intelligibilité. Je m'explique. Comment peut être défini un mot tel que « *poids* » ? [...] Définir un mot consiste à dire tout ce qu'il implique de *déjà connu*, de *préalablement conçu*. Pour « *poids* », le plus court chemin et le plus direct eut été de décrire la méthode de mesure. Pour l'intelligibilité, nous aurons donc à chercher ce qu'apporterait la description des méthodes de mesure, s'il en est.

Dans cette partie, nous allons commencer par suivre le conseil de D. Sadek-Khalil et décrire les principaux tests d'intelligibilité de la parole existant. Comme nous le verrons, les premiers tests d'intelligibilité ont été conçus afin de répondre à des problématiques bien particulières dans le domaine de l'ingénierie des systèmes de télé ou de radio transmission. Nous décrivons ces premières mesures de l'intelligibilité et leurs évolutions dans notre premier chapitre (p. 11). Il nous paraît en effet important d'adopter un point de vue historique dans la mesure où les tests développés dans le domaine des télécommunications ont ensuite été portés dans le domaine clinique pour évaluer les performances de personnes souffrant de troubles pathologiques de production de la parole.

Chapitre 1

Développement des tests d'intelligibilité de la parole dans les domaines de l'acoustique et des télécommunications

Sommaire

1.1	Introduction	12
1.2	De l'élaboration empirique à la standardisation des premières mesures subjectives de l'intelligibilité	14
1.2.1	Le test de Campbell (1910)	14
1.2.2	Le test de Fletcher et Steinberg (1929)	16
1.3	Développement de nouvelles méthodes de mesure subjective de l'intelligibilité	18
1.3.1	Essais aux mots	18
1.3.2	Essais aux phrases	19
1.3.3	Tests de rime	19
1.4	Vers des mesures objectives et prédictives	23
1.4.1	L'Indice d'Articulation (<i>Articulation Index - AI</i>)	23
1.4.2	L'Indice d'Intelligibilité de la Parole (<i>Speech Intelligibility Index - SII</i>)	24
1.4.3	L'Indice de Transmission de la Parole (<i>Speech Transmission Index - STI</i>)	24
1.4.4	Le Niveau d'Interférence avec la Parole (<i>Speech Interference Level - SIL</i>)	25
1.4.5	La perte d'articulation aux consonnes	25
1.4.6	Le modèle de Faiget	26
1.5	Synthèse et discussion	28

1.1 Introduction

« *Mister Watson, come here, I want to see you*¹ ». Ce sont les paroles prononcées par Alexander Graham Bell, le 10 mars 1876, lors de l'essai de son premier téléphone. De l'autre côté de la ligne téléphonique, dans une salle distante, Mr Watson – l'assistant de Bell – et quelques autres personnalités assistent à l'événement et prennent note des paroles entendues.

FIGURE 1.1 – Transmission d'un message de Salem à Boston par A. G. Bell, en mars 1877 (<http://www.life.com> – dernière consultation le 13/10/2010)



A. G. Bell (1876) dira plus tard, à propos de l'évaluation des performances de ce premier système téléphonique :

De manière générale, la prononciation était inintelligible, excepté lors de l'emploi de phrases familières. [...] Les sons élémentaires de l'anglais étaient prononcés successivement dans l'un des combinés téléphoniques, et les effets produits dans l'autre combiné étaient notés.²

Bell créa ainsi un précédent, en proposant de tester l'intelligibilité de la parole par le biais de comparaisons entre la production et la réception des sons de parole – évaluée par la retranscription des unités perçues. Comme le souligne Hawley (1977c, p. 2) à ce propos, « *en téléphonie, un précédent devient vite une tradition*³ ». Nous verrons en effet que la plupart des méthodes subjectives de mesure de l'intelligibilité de la parole qui sont utilisées aujourd'hui s'inscrivent dans ce paradigme de prononciation/retranscription qui a été développé aux débuts de la téléphonie. De plus, ce

1. « Monsieur Watson, venez ici, je veux vous voir ».

2. « Indeed as a general rule, the articulation was unintelligible except when familiar sentences were employed. [...] The elementary sounds of the English language were uttered successively into one of the telephones and its effects noted at the other. ».

3. « in telephony, a precedent quicly becomes a tradition ».

paradigme est non seulement utilisé pour des problématiques touchant à la transmission de signaux de parole (par ex. Peng *et al.*, 2011), mais aussi à la production (par ex. Habib *et al.*, 2010; Haley et Martin, 2010), et à la réception de la parole (par ex. Gnansia *et al.*, 2010).

Dans la section suivante nous nous intéressons tout d'abord aux premières mesures subjectives de l'intelligibilité à travers l'exemple de deux tests souvent présentés comme fondateurs : le test de Campbell (1910, *cf.* p. 14) et le *Test d'Articulation* de Fletcher et Steinberg (1929, *cf.* p. 16). Ces tests étant assez lourds à mettre en place – en particulier de par le temps nécessaire à l'entraînement des sujets locuteurs et auditeurs – des tests subjectifs plus simples à mettre en œuvre ont été développés par la suite (*cf.* section 1.3, p. 18), avant la création de mesures objectives et prédictives (*cf.* section 1.4, p. 23). Enfin, nous proposons à la fin de ce chapitre de faire la synthèse des méthodes de mesure de l'intelligibilité de la parole, et d'en discuter la procédure au regard de différentes théories de la communication – en particulier celles de Shannon et Weaver (1949) et de Jakobson (1963).

1.2 De l'élaboration empirique à la standardisation des premières mesures subjectives de l'intelligi- bilité

En tant que domaine de recherche, l'intelligibilité de la parole n'a commencé à faire l'objet de véritables expérimentations scientifiques qu'à partir du début du 20^{ème} siècle. C'est à cette époque que les ingénieurs en télécommunication ont commencé à développer des tests subjectifs permettant de quantifier le taux de dégradation de la parole lors de sa transmission par un *medium* particulier (Kent *et al.*, 1989; Schiavetti, 1992; Weismer et Martin, 1992). Ces tests consistaient en la lecture de listes de mots ou d'énoncés par des locuteurs, et en leur retranscription orthographique ou phonétique par des auditeurs situés de l'autre côté du système ou de l'environnement à évaluer. Construits au départ de manière *ad hoc* par les équipes de recherche, les tests d'intelligibilité ont ensuite fait l'objet de procédures de standardisation, et ce dès les années 1910-20 ; c'est alors que le concept d'*intelligibilité de la parole* – *speech intelligibility* – est né (Weismer et Martin, 1992), et avec lui tout un champ de recherche.

Ainsi selon Gabor (1947), des tests d'intelligibilité étaient déjà menés dans les laboratoires du Bureau des Postes en Grande-Bretagne en 1905, mais ces expérimentations n'ont donné lieu à aucune publication. Ce n'est que quelques années plus tard que des articles scientifiques feront leur apparition, émanant pour la plupart des laboratoires *Bell*, le laboratoire le plus influent de l'époque (Hawley, 1977a). L'article *Telephonic Intelligibility* de Campbell (1910) est l'un des plus importants de cette période pionnière.

1.2.1 Le test de Campbell (1910)

L'étude que rapporte Campbell (1910) est un exemple typique des problématiques rencontrées par les ingénieurs en téléphonie au début du 20^{ème} siècle. Ces difficultés sont notamment liées à l'échelle de grandeur des réseaux téléphoniques que l'on souhaitait construire, d'une envergure bien supérieure aux réseaux qui avaient été réalisés jusqu'alors. De même, le développement de lignes commerciales impliquait la création de nouveaux appareils téléphoniques destinés à une plus large clientèle, appareils nécessitant par conséquent d'être testés. Dans ce contexte le travail de Campbell visait à observer dans quelle mesure la qualité de transmission de la parole était affectée par l'utilisation de tel ou tel appareil téléphonique, ou par la longueur de la ligne téléphonique.

Pour ce faire Campbell a conçu un test d'intelligibilité reposant sur la reconnaissance de 20 syllabes de type Consonne-Voyelle (CV), dont le noyau était toujours la voyelle longue /i:/ (cf. table 1.1). 10 listes de 100 syllabes ont été constituées, dans lesquelles chacune des 20 consonnes utilisées apparaissait 5 fois dans une position attribuée de manière aléatoire. Le test se déroulait de la manière suivante : les dix listes étaient lues par des locuteurs, le signal de parole était transmis *via* le système téléphonique dont on désirait évaluer la performance, et de l'autre côté

des observateurs notaient pour chaque stimulus la consonne qu'ils pensaient avoir entendue. Le score d'intelligibilité correspondait alors à la proportion de consonnes « correctement transmises » par le système téléphonique, autrement dit au pourcentage d'équivalence entre les consonnes contenues dans les listes et les consonnes notées par les auditeurs.

TABLE 1.1 – Exemples de stimuli du test de Campbell (1910)

Syllabe	Exemple(s)	Syllabe	Exemple(s)
/bi:/	<i>bee</i>	/ni:/	<i>knee</i>
/tʃi:/	<i>Chi(le)</i>	/pi:/	<i>pea</i>
/di:/	<i>de(pot), dee</i>	/ri:/	<i>re(bus), rei</i>
/fi:/	<i>fee</i>	/si:/	<i>see</i>
/gi:/	<i>Gi(zeh)</i>	/ʃi:/	<i>she</i>
/hi:/	<i>he</i>	/ti:/	<i>tea</i>
/dʒi:/	<i>gee</i>	/θi:/	<i>the(ory)</i>
/ki:/	<i>key</i>	/vi:/	<i>ve(nus), vee</i>
/li:/	<i>lee</i>	/ji:/	<i>ye</i>
/mi:/	<i>me</i>	/zi:/	<i>ze(bra), zee</i>

D'autres listes de syllabes seront plus tard constituées dans la continuité du test proposé par Campbell. Toutefois, ces listes n'ont pas fait l'objet de publications. Elles étaient construites de deux manières différentes (Fletcher et Steinberg, 1929) :

- soit les syllabes étaient créées en combinant les sons de la langue de façon aléatoire. Dans ce cas les listes étaient construites de manière à ce que tous les sons de la langue apparaissent à des fréquences relativement équivalentes. Le problème posé par ce type de liste était que des syllabes de forme très peu fréquente en anglais, et donc difficiles à prononcer pour les locuteurs, étaient construites ;
- soit les listes étaient construites en prenant soin que les syllabes et les sons de parole soient représentatifs de la langue. Dans ce cas le souci principal était que seul un nombre limité de syllabes pouvait être construit. Or un nombre limité d'items implique une plus grande difficulté à équilibrer les listes, et peut également conduire à la familiarisation des auditeurs avec les stimuli.

De nombreux travaux ultérieurs ont porté sur ces questions de la représentativité et de l'« utilisabilité » des listes de stimuli. Traitant de cette problématique, l'article *Articulation Testing Methods* de Fletcher et Steinberg (1929) est indubitablement la contribution scientifique la plus importante de cette époque. À ce titre il deviendra la référence la plus citée en psychoacoustique pendant le quart de siècle suivant sa publication (Hawley, 1977a).

1.2.2 Le test de Fletcher et Steinberg (1929)

Le principe des Tests d'Articulation proposés par Fletcher et Steinberg (1929) est le même que celui défini par Campbell (1910) : des sons de parole sont prononcés par un locuteur ; ces sons sont transmis *via* un système de transmission ou bien *via* le canal aérien jusqu'à l'oreille d'un observateur, dont le rôle est de noter les sons qu'il pense identifier. La figure 1.2 représente une feuille de passation du test :

- dans les colonnes intitulées *called* sont inscrits les sons de parole prononcés par le locuteur ;
- dans les colonnes intitulées *observed* sont notés les sons de parole identifiés par l'auditeur.

Chaque syllabe⁴, de forme Consonne-Voyelle-Consonne (CVC) est prononcée à la suite d'une phrase porteuse⁵ (par ex. « *The first group is [CVC]* »⁶), dont on peut voir la liste sur la gauche de la feuille de passation. Après chaque phrase, l'observateur (l'auditeur) note la dernière syllabe entendue. À la fin du test, la correspondance entre les syllabes prononcées et les syllabes retranscrites est vérifiée pour chaque paire d'items : si la *totalité* des phonèmes (ici transcrits à l'aide de l'alphabet tchèque) correspondent, alors le transfert de la syllabe est considéré comme effectif (score de 1), sinon il est considéré comme nul (score de 0). L'intelligibilité de la parole équivaut alors au rapport entre le nombre de syllabes correctement retranscrites et le nombre de syllabes énoncées, soit 51,5 % pour l'exemple donné en figure 1.2.

Les listes de logatomes présentent deux inconvénients majeurs. Tout d'abord, ils nécessitent un entraînement assez poussé des locuteurs et auditeurs (CALLIOPE, 1989) : les premiers doivent apprendre à énoncer de manière régulière et naturelle des listes de pseudo mots, et les seconds doivent être rompus à l'utilisation de conventions de retranscription orthographique⁷. Ensuite, les tests aux logatomes, tels qu'ils ont été conçus au départ, ne permettent pas d'identifier quels sont les contrastes phonétiques les plus perturbés par les conditions de transmission du signal de parole. Pour répondre à ces deux questions ont été créés a) d'une part des tests faisant intervenir de véritables mots de la langue (*essais aux mots* et *essais aux phrases*) et b) d'autre part des *tests de rime* qui mettent l'accent sur les contrastes phonétiques de la langue.

4. du point de vue sémantique il s'agit de *logatomes* – c'est-à-dire d'unités linguistiques dénuées de sens.

5. comme son nom l'indique le rôle d'une phrase porteuse est de « porter » un stimulus. L'ajout d'une phrase porteuse permet en premier lieu de préparer l'auditeur à l'arrivée du stimulus, et d'éviter ainsi un effet de surprise. La diffusion de la phrase porteuse permet aussi de faire en sorte que les dégradations causées par le canal de transmission se stabilisent avant la diffusion du stimulus. Par exemple, dans une salle réverbérante, la réverbération du signal de parole a tendance à augmenter jusqu'à une phase de stabilisation. La phrase porteuse permet de diffuser le stimulus après que la réverbération s'est stabilisée.

6. « Le premier groupe est [CVC] ».

7. on peut se rendre compte de ces difficultés en observant la feuille de passation donnée en exemple par Fletcher et Steinberg (1929, figure 1.2), sur laquelle les items sont écrits à l'aide de l'alphabet tchèque – une langue étrangère pour les auditeurs qui ont utilisé ces feuilles.

1.2. De l'élaboration empirique à la standardisation des premières mesures
subjectives de l'intelligibilité

FIGURE 1.2 – Feuille de passation du Test d'Articulation de Fletcher et Steinberg
(1929, p. 8)

DATE <u>3-16-29</u>		SYLLABLE ARTICULATION <u>51.5%</u>					
TITLE OF TEST <u>PRACTICE TESTS</u>		CONDITION TESTED <u>1500~ Low Pass FILTER</u>					
TEST NO. <u>10</u>		OBSERVER <u>W.H.S.</u>					
LIST NOS. <u>5-9-37</u>		CALLER <u>E.B.</u>					
NO.		OBSERVED	CALLED	OBSERVED	CALLED	OBSERVED	CALLED
1	THE FIRST GROUP IS	<i>māv</i>	<i>nāv</i>	<i>pōz</i>	<i>poth</i>	<i>kōb</i>	<i>kōb</i>
2	CAN YOU HEAR	<i>pōch</i> ✓	<i>pōch</i>	<i>nēz</i>	<i>nēzh</i>	<i>shēth</i>	<i>siz</i>
3	I WILL NOW SAY	<i>seng</i> ✓	<i>seng</i>	<i>jōch</i> ✓	<i>jōch</i>	<i>fūch</i> ✓	<i>fūch</i>
4	AS THE FOURTH WRITE	<i>chūd</i> ✓	<i>chūd</i>	<i>thām</i> ✓	<i>thām</i>	<i>thōl</i> ✓	<i>thōl</i>
5	WRITE DOWN	<i>run</i> ✓	<i>run</i>	<i>hab</i> ✓	<i>hab</i>	<i>poth</i> ✓	<i>poth</i>
6	DID YOU UNDERSTAND	<i>chiz</i>	<i>kiz</i>	<i>def</i>	<i>doth</i>	<i>wām</i> ✓	<i>wām</i>
7	I CONTINUE WITH	<i>foz</i>	<i>fozh</i>	<i>chech</i>	<i>chej</i>	<i>gūm</i>	<i>gūn</i>
8	THESE SOUNDS ARE	<i>lōl</i> ✓	<i>lōl</i>	<i>lun</i>	<i>lon</i>	<i>nāsh</i>	<i>nāth</i>
9	TRY THE COMBINATION	<i>jās</i>	<i>zhāth</i>	<i>shāl</i> ✓	<i>shāl</i>	<i>vōg</i> ✓	<i>vōg</i>
10	PLEASE RECORD	<i>thāth</i>	<i>thāsh</i>	<i>muz</i> ✓	<i>muz</i>	<i>lung</i>	<i>long</i>
11	WRITE THE FOLLOWING	<i>wūr</i> ✓	<i>wūr</i>	<i>lēd</i>	<i>bēd</i>	<i>diz</i>	<i>dizh</i>
12	NOW TRY	<i>yāp</i> ✓	<i>yāp</i>	<i>wif</i> ✓	<i>wif</i>	<i>kak</i>	<i>tak</i>
13	THIRTEEN WILL BE	<i>mād</i>	<i>māj</i>	<i>gōst</i> ✓	<i>gōst</i>	<i>thār</i>	<i>zhār</i>
14	YOU SHOULD OBSERVE	<i>bēch</i>	<i>bēk</i>	<i>thav</i>	<i>sāv</i>	<i>must</i> ✓	<i>must</i>
15	WRITE CLEARLY	<i>gēm</i>	<i>dēm</i>	<i>kōf</i> ✓	<i>kōf</i>	<i>yō'd</i> ✓	<i>yō'd</i>
16	NUMBER 18 IS	<i>thēb</i>	<i>vēb</i>	<i>rōg</i> ✓	<i>rōg</i>	<i>jet</i> ✓	<i>jet</i>
17	YOU MAY PERCEIVE	<i>jok</i>	<i>jost</i>	<i>thip</i> ✓	<i>thip</i>	<i>rēp</i>	<i>rēj</i>
18	I AM ABOUT TO SAY	<i>gaf</i> ✓	<i>gaf</i>	<i>yar</i> ✓	<i>yar</i>	<i>thēp</i>	<i>hēp</i>
19	TRY TO HEAR	<i>hus</i> ✓	<i>hus</i>	<i>zhāt</i> ✓	<i>zhāt</i>	—	<i>chuv</i>
20	PLEASE WRITE	<i>hiv</i>	<i>thith</i>	<i>kūk</i>	<i>tūk</i>	<i>thēf</i>	<i>thēsh</i>
21	LISTEN CAREFULLY TO	<i>tōg</i> ✓	<i>tōg</i>	<i>fung</i> ✓	<i>fung</i>	<i>bās</i> ✓	<i>bās</i>
22	THE LAST GROUP IS	<i>shōt</i> ✓	<i>shōt</i>	<i>thēv</i>	<i>vēsh</i>	<i>thōf</i>	<i>shaf</i>

1.3 Développement de nouvelles méthodes de mesure subjective de l'intelligibilité

1.3.1 Essais aux mots

Egan (1948) est un des premiers auteurs à avoir pointé du doigt les difficultés d'entraînement rencontrées avec les listes de logatomes. Son article « Articulation Testing Methods » est souvent présenté comme fondamental en ce qui concerne les tests de reconnaissance de mots. Egan propose dans cet article 20 listes contenant chacune 50 mots monosyllabiques de fréquence courante en anglais. Pour le français, il existe le matériel élaboré par Zuliani (1988) (*cf.* table 1.2). Ce matériel comprend 10 listes de 34 mots triphonémiques mono- ou dissyllabiques. Ces listes sont équilibrées phonétiquement : la fréquence d'occurrence des phonèmes correspond à leur fréquence en français courant.

TABLE 1.2 : Liste de mots de Zuliani (1988), d'après Faiget (1997)

N°	A	B	C	D	E	F	G	H	I	J
1	bile	dru	allée	oncle	neuve	buée	mur	béat	leur	heureux
2	dort	idée	acquis	tuant	dalle	ride	balle	veule	enfant	dette
3	meurt	seul	rase	leurre	faire	pâle	soude	range	inné	longe
4	gaine	orne	jute	somme	héros	suant	nef	gale	souc	gave
5	sage	épais	mille	épais	gaz	cri	change	néon	chœur	ami
6	honteux	urée	sonde	flot	caisse	ailé	gale	port	tôle	seul
7	mule	panne	douce	cure	foc	raide	trou	avis	élu	tasse
8	cale	hanté	ville	gris	mainte	mieux	malle	taille	prend	prêt
9	rive	rave	mare	pince	prix	natte	tonne	meurt	fade	front
10	neuf	soie	causse	rôle	haineux	colle	peur	poule	motte	sur
11	sol	teinte	appât	hibou	hisser	soupe	cale	dite	tri	crin
12	cran	vif	heurte	brun	rite	tonte	rampe	sœur	essaim	vol
13	phase	moule	cieux	sève	ému	vèle	corps	frais	meule	ruse
14	chatte	on-dit	fête	haineux	riche	nage	vite	suc	voûte	épi
15	pays	pâle	veule	ara	ondée	pompe	rance	époux	pile	dieu
16	pelle	code	ronde	seul	aveu	site	laide	mage	haché	ôter
17	dupe	athée	embout	aidé	pelle	heureux	dard	creux	héron	ride
18	ami	fange	foi	vide	rade	loi	seul	vote	doit	étang
19	suc	gaule	brin	meurt	honteux	rente	gîte	élan	ronde	gène
20	gel	ici	drap	fine	lieu	cave	cahot	sec	baisse	raide
21	doute	allée	aucun	coule	embout	tire	père	loi	grue	poids
22	aider	cran	raide	mêle	pic	écho	effet	mien	jeune	année
23	heurte	monte	puis	agit	lance	somme	mine	inné	ami	aucun
24	aveu	nasse	langue	pousse	ours	drain	élan	date	épais	roule
25	dense	houleux	pire	état	aucun	mène	tronc	coupe	lande	pile
26	ponte	prêt	colle	dote	latte	preux	dieu	errer	anneau	masse
27	brun	heureux	naine	néant	humant	agit	acquis	assaut	vieux	banque
28	sœur	sac	sotte	cœur	laine	saoûle	voie	élu	cette	sire
29	inné	bise	suée	zèle	patte	aisé	pèse	rente	vide	nef
30	auvent	encre	neuve	achat	range	chance	ponce	dîne	serre	veule
31	quinte	vide	penche	sente	selle	bœuf	sainte	roche	case	cuit
32	soie	orée	irait	toit	doit	appât	veule	hideux	puis	manche

Liste de mots de Zuliani (1988) (suite)

33	lire	seiche	humer	peur	hideux	rôle	puce	seize	lisse	selle
34	tire	meule	telle	édit	suie	lente	aidé	ponte	cure	doute

Afin d'illustrer les différences de temps d'entraînement pour des tests aux logatomes et des tests aux mots comme ceux proposés par Egan (1948) et Zuliani (1988), la norme ISO/TR 4870 (1991) donne les durées typiques d'apprentissage de locuteurs et d'auditeurs (table 1.3).

TABLE 1.3 – Durées typiques d'apprentissage pour des tests aux mots *vs* aux logatomes (ISO/TR 4870, 1991, p. 17)

	Essais sur des grands corpus pseudo-ouverts ^a		Essais sur des petits corpus fermés	
	Logatomes	Mots	Logatomes	Mots
Durée typique d'apprentissage des locuteurs	8 heures	1 heure	2 heures	5 minutes
Durée typique d'apprentissage de l'auditeur	24 heures	12 heures	2 heures	5 minutes

a. Pour environ 20 listes de 50 éléments de test

1.3.2 Essais aux phrases

Comme pour les essais aux mots, le principal avantage des tests aux phrases est de réduire l'entraînement nécessaire aux sujets locuteurs et auditeurs. Toutefois, une grande différence oppose les tests de reconnaissance de mots aux tests de reconnaissance de phrases : du fait de la redondance des informations présentes dans le signal de parole, l'écart entre des scores d'intelligibilité très bas et ceux correspondant à une intelligibilité très haute est beaucoup plus réduit dans le cas des tests aux phrases. Autrement dit, dans le cas des tests aux phrases les scores ont tendance à atteindre beaucoup plus rapidement le plafond correspondant à une intelligibilité parfaite que cela n'est le cas pour les scores obtenus par des tests aux mots.

1.3.3 Tests de rime

Les tests de rime se focalisent sur un aspect particulier de l'intelligibilité de la parole : la différenciation phonémique – c'est-à-dire la capacité de l'auditeur à distinguer entre les phonèmes de la langue. Ainsi chaque item de test ne porte que sur un son de parole particulier, que l'auditeur doit transcrire ou bien sélectionner

dans une liste. Cela permet d'identifier quels sont les phonèmes ou bien les contrastes phonétiques (ex. voisement, nasalité, etc.) qui posent un problème de reconnaissance à l'auditeur. Ces tests sont dits « explicatifs » puisqu'ils permettent ainsi d'établir une sorte de « diagnostic phonétique » mettant au jour les principales difficultés de l'auditeur en matière de perception phonémique.

Il existe différentes versions des tests de rime. Nous avons choisi de présenter les trois principales :

- le Test de Rime de Fairbanks (*Fairbanks Rhyme Test – FRT* ; Fairbanks, 1958) ;
- le Test de Rime Modifié (*Modified Rhyme Test – MRT* ; House *et al.*, 1965) ;
- le Test de Rime à visée diagnostique (*Diagnostic Rhyme Test – DRT*).

Le Test de Rime de Fairbanks (Fairbanks Rhyme Test – FRT ; Fairbanks, 1958)

Le matériel utilisé pour le Test de Rime de Fairbanks consiste en 5 listes de 50 mots. Les items de chaque liste ne se différencient que par leur consonne initiale. Deux items pris dans deux listes différentes constituent donc ce que l'on appelle en phonologie une *paire minimale*, soit deux morphèmes ne se distinguant que par un seul phonème. La table 1.4 présente quelques items⁸ à titre d'exemple.

TABLE 1.4 – Exemples d'items dans les 5 listes du Test de Rime de Fairbanks (1958)

Item	Liste 1	Liste 2	Liste 3	Liste 4	Liste 5
1	/hɑ:t/ (hot)	/gɑ:t/ (got)	/nɑ:t/ (not)	/p ^h ɑ:t/ (pot)	/lɑ:t/ (lot)
2	/p ^h ēi/ (pay)	/mēi/ (may)	/dēi/ (day)	/wēi/ (way)	/sēi/ (say)
3	/t ^h ɑ:p/ (top)	/hɑ:p/ (hop)	/p ^h ɑ:p/ (pop)	/mɑ:p/ (mop)	/k ^h ɑ:p/ (cop)
...
50	/lʌk/ (luck)	/t ^h ʌk/ (tuck)	/dʌk/ (duck)	/sʌk/ (suck)	/bʌk/ (buck)

Le Test de Rime de Fairbanks est un test de complétion orthographique. Les auditeurs disposent d'une feuille de passation sur laquelle sont notés les stimuli, dans l'ordre où ils vont être entendus. La première consonne de chaque item n'est pas écrite (ex. « _ot », « _ay »), charge à l'auditeur de compléter chaque mot en fonction du son qu'il perçoit.

8. nous avons donné dans la table 1.4 les transcriptions phonétiques des items, en nous appuyant sur leur prononciation en anglais américain standard. Cependant dans les listes utilisées par Fairbanks seuls les mots écrits entre parenthèses étaient présentés.

Le Test de Rime Modifié (Modified Rhyme Test – MRT ; House *et al.*, 1965)

House *et al.* (1965) ont utilisé sensiblement le même type de test que Fairbanks (1958) afin de procéder à l'évaluation de divers systèmes de communication vocale – en majorité destinés à des applications militaires. Leur test, appelé *Modified Rhyme Test* (MRT), est en réalité une version simplifiée du test de Fairbanks, les auteurs désirant réduire les coûts en temps et en équipement pour la passation des tests. House *et al.* (1965) cherchaient également à diminuer les temps d'entraînement des locuteurs et des auditeurs : leur objectif était de pouvoir, à terme, employer des sujets naïfs.

Pour ce faire les auteurs ont choisi de réduire les choix possibles pour l'auditeur. Le matériel de test comprend 50 items, dont chacun contient 6 mots de forme CVC ne se distinguant que par une consonne. 25 items présentent des mots dont la consonne initiale diffère, et les 25 autres items sont constitués de mots se distinguant par leur consonne finale. Cette réduction des choix proposés à l'auditeur (6 phonèmes par item) permet d'éliminer les problèmes d'apprentissage du vocabulaire, mais aussi d'annihiler les effets relatifs à la familiarité des mots (CALLIOPE, 1989).

Le Test de Rime à visée diagnostique (Diagnostic Rhyme Test - DRT)

Le DRT a été développé par W. D. Voiers et ses collègues, qui désiraient réduire au minimum l'influence du contexte informationnel dans les tests subjectifs utilisés pour l'évaluation des systèmes de traitement de la parole (Voiers, 1983). Les auteurs ont constitué une liste de 96 paires de mots de forme CVC dont les consonnes initiales ne se distinguent que par un des traits phonologiques suivants : *voisé/non voisé*, *grave/non grave*, *compact/diffus*, *interrompu/non interrompu*, *vocalique/non vocalique*, ou *nasal/non nasal*.

Le test de Voiers a ensuite été appliqué au français par Peckels et Rossi (1973), dans leur *test de diagnostic par paires minimales*. La table 1.5 présente quelques items de ce test, et les traits distinctifs qui leur sont associés.

TABLE 1.5 – Exemples d'items du test de diagnostic par paires minimales (d'après
 Cartier et Rossi, 1975), et les traits distinctifs associés

<i>vol</i>			<i>fol</i>	(contraste \pm voisé)
<i>vosne</i>			<i>zone</i>	(contraste \pm grave)
<i>corps</i>			<i>port</i>	(contraste \pm diffus)
<i>bol</i>			<i>vol</i>	(contraste \pm interrompu)
<i>note</i>			<i>lotte</i>	(contraste \pm nasal)
<i>gourd</i>			<i>court</i>	(contraste \pm voisé)
<i>bouse</i>			<i>douze</i>	(contraste \pm grave)
<i>coule</i>			<i>poule</i>	(contraste \pm diffus)

1.4 Vers des mesures objectives et prédictives

La majorité des modifications ou des développements qui ont été apportés aux tests subjectifs d'intelligibilité de la parole répondaient à leur difficulté de mise en œuvre, et à des soucis d'objectivité et de reproductibilité. Hawley (1977b, p. 120) a très bien résumé ces questions :

« Nous avons recours aux tests d'intelligibilité uniquement parce qu'ils représentent la méthode la plus valide et la plus fiable pour mesurer la compréhension de la parole. (Mais) ils sont onéreux et demandent du temps. Ils nécessitent plusieurs sujets, de l'équipement, et un expérimentateur compétent. Pire encore, ils sont difficiles à répliquer dans d'autres laboratoires ou bien à un moment différent. »⁹

Pour pallier ces difficultés des recherches ont été entreprises dès les années 1940 pour essayer de prédire les scores d'intelligibilité à partir de critères objectifs. L'objet de ces méthodes est donc de retrouver, par le calcul ou la mesure, les scores d'intelligibilité qui seraient obtenus dans les mêmes conditions grâce à des tests subjectifs.

1.4.1 L'Indice d'Articulation (*Articulation Index - AI*)

L'Indice d'Articulation permet de prédire les scores de Tests d'Articulation, tels que celui développé par Fletcher et Steinberg (1929, p. 16). Développé à l'origine dans les laboratoires *Bell* par French et Steinberg (1947), il est obtenu en mettant en rapport les niveaux de parole et de bruit dans certaines bandes de fréquences avec les scores d'intelligibilité habituellement obtenus – dans les mêmes conditions – par des mesures subjectives.

En d'autres termes l'Indice d'Articulation permet de retrouver la proportion de syllabes dont tous les phonèmes seraient correctement identifiés par des auditeurs, et ce en se basant uniquement sur un critère, un seul « facteur de dégradation de l'intelligibilité » : le bruit entrant en concurrence, à la hauteur de l'oreille de l'auditeur, avec le signal de parole¹⁰.

9. « Speech intelligibility tests are undertaken only because they offer the most convincing and trustworthy method of measurement of the ability to understand speech. They are expensive and time consuming; they require several subjects, some equipment, and a skilled experimenter; and, even worse, they are difficult to replicate in other laboratories or at other times. » (Hawley, 1977b, p. 120).

10. à ce propos, l'indice d'articulation n'a rien à voir avec la production de parole, comme sa dénomination semblerait l'indiquer; de nombreux auteurs – notamment dans le domaine de l'audition – ont en conséquence choisi de se référer à cette mesure sous le nom d'Indice d'Audibilité (*Audibility Index - AI*) (Hornsby, 2004).

1.4.2 L'Indice d'Intelligibilité de la Parole (*Speech Intelligibility Index - SII*)

Tout comme l'Indice d'Articulation, l'indice d'intelligibilité de la parole est une mesure comprise entre 0.0 et 1.0, et qui est « hautement corrélée avec l'intelligibilité de la parole » (ANSI S3.5, 1997, p. 1).

Néanmoins, le SII est une mesure plus élaborée : L'AI est calculé à partir de mesures ou d'estimations des niveaux de parole et de bruit dans plusieurs bandes de fréquences contiguës (généralement 20 bandes, parfois certaines bandes d'octaves ou encore des bandes de tiers d'octave), dont chacune contribue de façon égale à l'intelligibilité de la parole (Kryter, 1962). Le SII, quant à lui, va plus loin et introduit dans le calcul des fonctions d'importance de fréquences (*frequency importance functions - FIFs*). Le fait que toutes les zones fréquentielles ne contribuent pas de manière égale à l'intelligibilité de la parole est donc pris en compte dans ce calcul (Hornsby, 2004).

1.4.3 L'Indice de Transmission de la Parole (*Speech Transmission Index - STI*)

Le STI est une méthode permettant à la fois d'effectuer des mesures objectives de l'intelligibilité de la parole et de prédire cette dernière. Elle a été créée dans les années 1960-70 par Herman J.M. Steeneken et Tammo Houtgast, qui désiraient alors mettre en place un système de prédiction de l'intelligibilité pour pallier la lourdeur d'analyses reposant sur des tests subjectifs lors de l'évaluation de systèmes de transmission radio ou téléphonique (Houtgast et Steeneken, 2002).

Le principe de cette méthode est de considérer l'ensemble des éléments perturbateurs (bruit, réverbération, etc.) propres à un canal de communication (système de transmission radio ou téléphonique, salle, etc.) comme un « filtre » altérant certaines caractéristiques du signal acoustique de la parole et réduisant ainsi son intelligibilité. Steeneken et Houtgast ont effectué un certain nombre d'analyses portant sur des signaux artificiels (reproduisant certaines caractéristiques acoustiques de la parole continue – notamment le spectre moyen et l'enveloppe spectrale), avant et après leur passage dans divers canaux de transmission. Les auteurs ont démontré qu'il existait une corrélation significative entre la perte de précision des différences spectrales du signal initial (dus entre autres facteurs au bruit et à la durée de réverbération), et les résultats obtenus par des tests d'évaluation subjectifs réalisés sous les mêmes conditions. Cela, en plus de la prise en compte d'autres facteurs relatifs au canal de transmission (échos) et aux interlocuteurs (sexe, effort vocal du locuteur ; distance entre les interlocuteurs), a permis d'élaborer un système de prédiction du STI couramment utilisé aujourd'hui pour l'évaluation ergonomique de la communication parlée (NF EN ISO 9921, 2004).

Le terme de STI renvoie donc à la fois :

- à une évaluation *objective* de l'intelligibilité, obtenue par la comparaison entre les caractéristiques spectrales d'un signal artificiel aux points d'émission et de réception ;

- à une méthode de *prédiction* de l'intelligibilité, puisque l'on peut mettre en correspondance certains paramètres physiques des canaux de transmission et des interlocuteurs avec les scores STI obtenus pour les mêmes conditions.

Le STI, qu'il soit mesuré ou prédit, peut enfin être mis en rapport avec les scores d'évaluation subjective de l'intelligibilité.

La méthode « complète » et originale de mesure du STI – à laquelle on se réfère sous le sigle STI-14 – nécessite la prise en compte de 14 fréquences de modulation différentes, et ce pour 7 bandes d'octaves. Cette méthode, si elle est applicable pour une grande variété de systèmes de communication, était difficile à effectuer par le passé – notamment à cause de durées de calcul trop importantes. Pour cette raison, des versions « simplifiées » de la méthode de calcul du STI ont été élaborées en fonction des besoins dans divers domaines :

- le STI-3, qui ne prend en compte que trois bandes d'octaves ;
- le STITEL (STI for TELEcommunication channels), qui a été conçu pour évaluer les systèmes de télécommunication ;
- le STIPA (STI for Public Address systems), qui a été développé pour les systèmes de sonorisation ;
- le RASTI (Room Acoustical STI, ou RApid STI), qui est largement utilisé en acoustique des salles.

Enfin, plus récemment le STI a subi des modifications qui ont débouché sur le développement du STI_r (*Revised STI*, Steeneken et Houtgast, 2002).

1.4.4 Le Niveau d'Interférence avec la Parole (*Speech Interference Level - SIL*)

Le SIL est une méthode d'évaluation et de prédiction de l'intelligibilité de la parole. Beaucoup moins complexe que le STI, son champ d'application est restreint aux situations de communication directe en milieu bruyant. Le niveau SIL est défini comme la différence entre le niveau de parole et le niveau d'interférence du bruit avec la parole, tous deux étant déterminés dans la position de l'auditeur. Une intelligibilité correcte de la communication parlée est assurée lorsque la différence des niveaux est égale ou supérieure à 10 dB dans la position de l'auditeur.

Le niveau SIL peut être mesuré, ou bien prédit à partir des paramètres suivants : niveau de bruit ambiant, effort vocal du locuteur, utilisation ou non d'un protecteur individuel contre le bruit, distance entre les interlocuteurs et statut (langue maternelle ou non) chez l'auditeur de la langue utilisée par le locuteur (NF EN ISO 9921, 2004).

1.4.5 La perte d'articulation aux consonnes

Peutz (1971) a observé que les consonnes contribuaient davantage aux scores d'intelligibilité que les voyelles, notamment dans un environnement peu réverbérant et non bruyant (Faiget, 1997). En se fondant sur cette observation, Peutz a proposé

le modèle %AL_{cons} (*Articulation Loss of CONSonants*). À partir de données liées au canal de communication (dimensions de la salle, distance entre les interlocuteurs, etc.) ce modèle prédit la perte de définition des consonnes, cette dernière permettant à son tour de prédire les scores d'intelligibilité.

1.4.6 Le modèle de Faiget

Faiget *et al.* ont élaboré un modèle de prédiction de l'intelligibilité adapté aux conditions d'écoute difficiles dans des salles équipées d'un système de sonorisation (voir notamment Faiget, 1997; Faiget et Ruiz, 1999). Une des originalités de ce modèle est de séparer l'influence de paramètres propres à la salle (réverbération) et propres au système de sonorisation (directivité, réponse en fréquence), ce qui lui permet d'atteindre des seuils de prédiction de l'intelligibilité plus élevés que ceux du STI dans les conditions pour lesquelles il est prévu ($r = 0,96$).

En conclusion, les méthodes de prédiction de l'intelligibilité constituent des solutions pratiques pour estimer les scores subjectifs d'intelligibilité en prenant en compte les paramètres physiques pertinents (bruit, réverbération, etc.) en fonction des applications visées (évaluation de systèmes de sonorisation, de lieux, etc.). La table 1.6 indique la correspondance entre les descripteurs correspondant à une évaluation subjective de l'intelligibilité sur une échelle à 5 points et les scores obtenus par la plupart des tests subjectifs, objectifs et prédictifs de l'intelligibilité dont nous avons parlé dans cette partie.

TABLE 1.6 – Relations entre différents indices d'intelligibilité (NF EN ISO 9921, 2004)

Évaluation de l'intelligibilité ^a	Note par phrase ^b %	Note par type de mots phonétiquement équilibrés ayant un sens ^c %	Note par type de mots CVC _{EQB} dépourvus de sens %	STI	SIL ^d dB	SII ^{d, e}
Excellente	100	> 98	> 81	> 0,75	21	-
Bonne	100	93 à 98	70 à 81	0,60 à 0,75	15 à 21	> 0,75
Correcte	100	80 à 93	53 à 70	0,45 à 0,60	10 à 15	-
Faible	70 à 100	60 à 80	31 à 53	0,30 à 0,45	3 à 10	< 0,45
Mauvaise	< 70	< 60	< 31	< 0,30	< 3	-

a. qualification conformément à une échelle à cinq points.

b. la note par phrase fait référence à des phrases simples, aux mots CVC_{EQB} dépourvus de sens avec une distribution équitable des phonèmes, et à la note des mots phonétiquement équilibrés (relative à la liste de Harvard phonétiquement équilibrée).

c. selon Anderson et Kalb (1987).

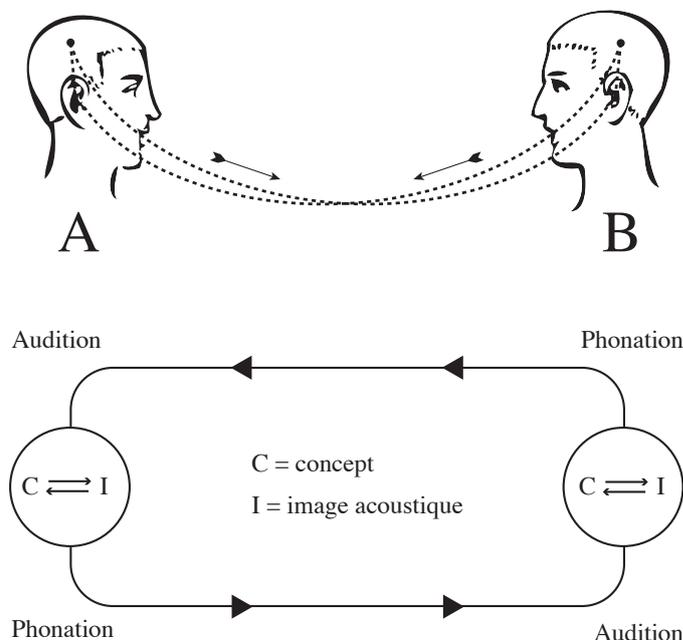
d. les méthodes SIL et SII ne font référence qu'aux conditions de bruit ambiant.

e. la procédure SII ne prévoit pas d'intervalles de référence accompagnés de descripteurs (*faible intelligibilité*, *intelligibilité moyenne*, etc.). La norme ANSI S3.5 (1997) donne deux valeurs de référence pour cet indice, à savoir *bonne* > 0,75 et *faible* < 0,45.

1.5 Synthèse et discussion

Dans cette partie nous nous intéressons à ce que recouvre la notion d'intelligibilité de la parole au vu des tests que nous avons décrits, et au regard de différentes théories de la communication. Nous pensons en particulier au circuit de la parole présenté par Ferdinand de Saussure (1995, 1916 pour l'édition originale) et à la théorie de Shannon et Weaver (1949), qui peuvent apporter un éclairage intéressant car ils sont tous les deux contemporains de l'élaboration des mesures subjectives de l'intelligibilité. La théorie mathématique de la communication (Shannon et Weaver, 1949) est d'autant plus intéressante pour notre propos qu'elle a été développée au sein des laboratoires *Bell*, en réponse à la même problématique que celle guidant les recherches sur l'intelligibilité de la parole. Enfin, nous verrons que le modèle de Jakobson (1963) peut se révéler pertinent pour décrire les aspects fonctionnels de la communication qui sont privilégiés dans les tests et mesures d'intelligibilité de la parole utilisés en acoustique et dans le domaine des télécommunications.

FIGURE 1.3 – Le circuit de la parole, d'après F. de Saussure (1995, pp.27-28, 1916 pour l'édition originale). En haut, A et B figurent deux personnes qui s'entretiennent; un concept donné active dans le cerveau de A une image acoustique, dont l'impulsion est ensuite transmise aux organes de la phonation. L'onde acoustique est transmise jusqu'aux organes récepteurs de B. L'image acoustique sera associée au concept correspondant dans le cerveau de B. Si B parle à son tour, le processus se déroulera exactement de la même manière. Ce circuit est représenté de façon plus schématisée en bas de la figure.



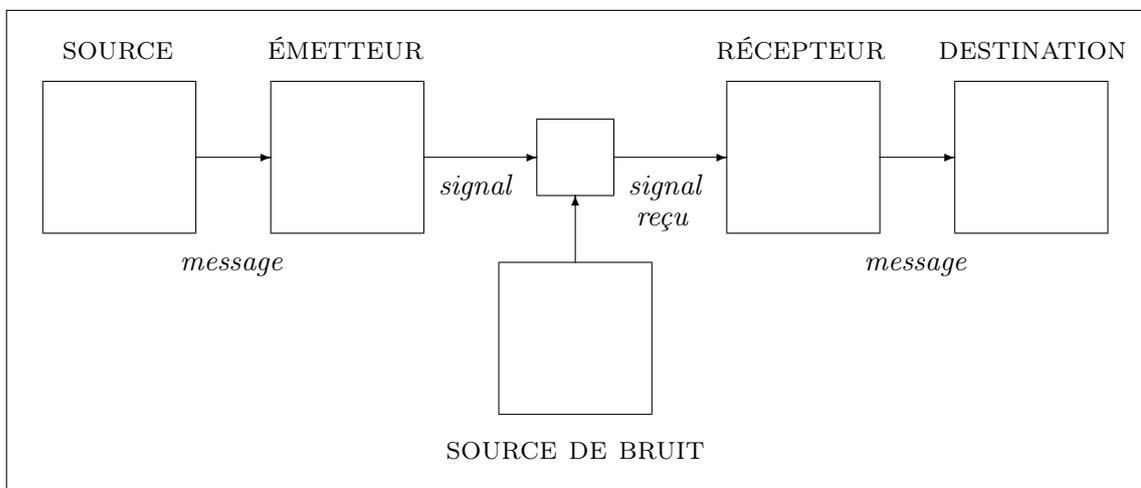
Au vu des méthodes que nous avons présentées dans ce premier chapitre il nous paraît important de souligner un aspect fondamental des tests d'intelligibilité, celui du *paradigme de lecture et de retranscription de stimuli*. De ce point de vue, nous

ne nous intéressons pas aux méthodes de mesure objectives et prédictives de l'intelligibilité. Comme nous l'avons vu, ces dernières sont des outils puissants permettant de retrouver par le calcul ou par des approximations *les scores que l'on obtient par les méthodes subjectives de mesure de l'intelligibilité*. L'essence de la mesure reste donc la même dans les deux cas : pour les tests prédictifs comme pour les tests subjectifs l'intelligibilité de la parole renvoie au taux d'unités linguistiques (syllabes et mots principalement) dont tous les phonèmes sont correctement identifiés par des auditeurs.

Ainsi dans le domaine de l'acoustique et des télécommunications nous pouvons définir les mesures d'intelligibilité de la parole comme des mesures de *transfert du code linguistique*. Le schéma de la communication quelque peu mécaniciste proposé par Ferdinand de Saussure (cf. figure 1.3) donne un éclairage intéressant à notre propos. On peut y voir les deux principaux éléments de toute situation de communication verbale – le *locuteur* et l'*auditeur* – engagés dans une conversation. Le troisième élément permettant la communication est représenté dans la partie inférieure du schéma : il s'agit du *canal de communication* qui permet le transfert des « images acoustiques », c'est-à-dire du versant formel du code linguistique (ce que Saussure appellera le « signifiant » par opposition au « signifié »), entre les organes émetteurs et récepteurs des deux individus.

Du point de vue de ces trois principaux éléments constitutifs d'une situation de communication, les tests d'intelligibilité se focalisent donc sur l'évaluation des *performances du canal de transmission*. Ceci n'est pas étonnant lorsque l'on considère les problématiques auxquelles étaient confrontés les ingénieurs en télécommunication à cette époque, et qui concernaient l'évaluation des performances d'appareils d'émission et de réception ainsi que celles de canaux de communication comme les lignes téléphoniques. Ces problématiques sont particulièrement bien illustrées par le schéma de la communication proposé par Shannon et Weaver (1949, cf. figure 1.4).

FIGURE 1.4 – Schéma de la communication de Shannon et Weaver (1949)



Bougnoux (1993, p. 418) décrit le schéma de Shannon et Weaver de la manière suivante :

La source d'information choisit le message désiré parmi une série de messages possibles [...]. Le message choisi peut consister en mots écrits ou parlés, ou en images, musique, etc. L'*émetteur* transforme ce *message* en *signal* qui est alors envoyé par le canal de communication de l'émetteur au *récepteur*. Dans le cas du téléphone, le canal est un fil métallique, le signal un courant électrique variable parcourant ce fil ; l'émetteur est un ensemble d'éléments (émetteur téléphonique, etc.) qui transforme la pression du son vocal en un courant électrique variable. Dans le cas de la télégraphie, l'émetteur code des mots écrits en séquences de courant interrompues de longueurs variables (points, traits, blancs). Pour le langage parlé, la source d'information est le cerveau, l'émetteur est l'organe vocal qui produit la pression variable sonore (le signal) transmise à travers l'air (le canal). [...] Le récepteur est en quelque sorte un émetteur inversé, changeant le signal reçu en message et amenant ce message à destination. Quand je vous parle, mon cerveau est la source d'information, et le vôtre la destination ; mon système vocal est l'émetteur, et votre oreille avec la huitième paire de nerfs crâniens est le récepteur.

La source des problèmes rencontrés par les ingénieurs en télécommunication (la « source de bruit ») se situait donc au niveau du canal de communication. Dans la perspective de Shannon et Weaver ces perturbations agissent comme un filtre réduisant la quantité d'information reçue en comparaison de celle qui a été émise. Pour évaluer un système de transmission ou un canal de communication il est donc nécessaire de mesurer les différences entre le signal émis et le signal reçu.

Le parallèle avec les tests d'intelligibilité apparaît comme évident. Dans le cas des tests subjectifs les signaux émis correspondent aux stimuli (logatomes, mots, phrases) prononcés par les locuteurs et les signaux reçus aux transcriptions effectuées par les auditeurs. Comme dans le schéma de la communication de Shannon et Weaver, dans les tests d'intelligibilité les performances des émetteurs et des récepteurs (c.-à-d. les organes de production et de réception de la parole) sont considérées comme parfaites, toute différence entre les signaux émis et reçus pouvant ainsi être imputée au canal de communication ¹¹.

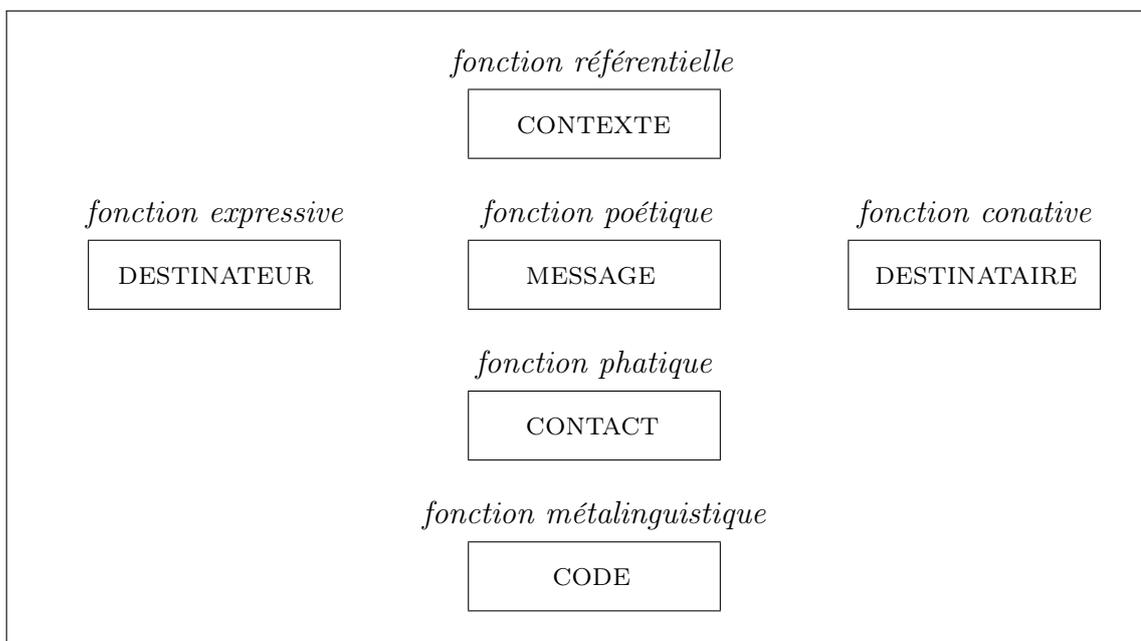
Ce qu'il apparaît également à la lumière du schéma de Shannon et Weaver (1949), c'est que la parole était considérée par les ingénieurs en télécommunication comme un *code*, c'est-à-dire comme un système composé de symboles finis qui peuvent être combinés afin de formuler des messages. Dans cette perspective la parole est en

11. le parallèle que nous faisons entre la théorie mathématique de la communication et les tests d'intelligibilité de la parole trouve également un élément de justification dans l'écrit de Allen (2004). L'auteur y effectue une démonstration mathématique prouvant que le calcul de l'Indice d'Articulation (*Articulation Index* – *AI*, cf. section 1.4.1 p. 23) est quasiment similaire au calcul proposé par Shannon et Weaver (1949) pour mesurer la capacité d'un canal de communication.

tous points comparable au code Morse¹², comme cela transparait des propos de Bougnoux (1993) que nous avons cités plus haut. Dans le cas du code Morse les symboles sont les points, les traits et les blancs, dans le cas de la parole ce sont les phonèmes. Dans le cas du code morse les éléments permettant de distinguer les symboles sont a) l'absence ou la présence de l'onde (\pm *silence*) et b) la durée de l'onde (\pm *court*), et dans le cas de la parole ce sont les traits distinctifs (\pm *voisé*, \pm *nasal*, \pm *antérieur*, etc.).

Ce point de vue sur la communication verbale est donc centré sur le canal de communication, dont on évalue les performances en observant sa capacité à transmettre le code linguistique de l'émetteur vers le récepteur. Plus tard, Jakobson (1963) a proposé un modèle de la communication différent qui permet de mieux cerner les limites de cette perspective (figure 1.5).

FIGURE 1.5 – Les fonctions du langage selon Jakobson (1963)



Jakobson (1963) a en effet adopté un point de vue *fonctionnel* sur la communication verbale, qui nous semble tout à fait pertinent pour notre propos puisque nous sommes particulièrement intéressé par la question de l'*évaluation* – que celle-ci concerne les performances d'un système de transmission ou bien les performances de locuteurs pathologiques, comme nous le verrons dans le prochain chapitre. Aux 6 éléments qu'il juge comme intrinsèques à toute situation de communication (les *destinateur* et *destinataire*, le *contact* établi entre eux, le *code* qu'ils partagent, le

12. à propos de cette comparaison entre le langage oral et la communication télégraphique, l'article « Description of language design » de Joos (1950) est très éclairant sur le point de vue que pouvaient avoir les ingénieurs de cette époque.

contexte de la communication et le *message* concerné), Jakobson fait correspondre 6 fonctions :

- au *destinateur* correspond la fonction *expressive*, celle qui permet au locuteur d'exprimer ses émotions et sentiments ;
- au *destinataire* correspond la fonction *conative*. C'est cette fonction qui permet au destinataire d'*agir* sur le destinataire, par exemple en lui donnant un ordre verbal ;
- au *contact* correspond la fonction *phatique*. Par cette fonction le destinataire établit, maintient ou rompt le contact qui le lie au destinataire. Par exemple lorsque le locuteur ponctue ses propos de « tu vois », « hein », et autres « n'est-ce pas » pour s'assurer que le contact avec le destinataire est toujours établi, c'est la fonction phatique qui est en jeu ;
- au *code* correspond la fonction *métalinguistique*. C'est la fonction métalinguistique qui permet au destinataire de faire référence au code linguistique, c'est-à-dire d'utiliser le langage pour parler du langage – d'où le préfixe *méta*. Ainsi, lorsqu'une personne énonce « J'ai pas dit *tartelette*, j'ai dit *tartiflette* », elle utilise la fonction métalinguistique du langage ;
- au *contexte* correspond la fonction *référentielle*. Par *contexte* il est entendu tous les éléments extra linguistiques qui font partie de la situation de communication et qui sont, au moins en partie, partagés par le destinataire et le destinataire. Lorsque le destinataire, par ses paroles, fait référence à un élément extra linguistique (par exemple lorsqu'il dit « la *tartiflette* est servie. » pour se référer à un plat particulier composé de pommes de terres et de reblochon existant dans le monde réel) c'est la fonction référentielle qui est activée ;
- enfin, au *message* correspond la fonction *poétique*. C'est cette fonction qui permet au destinataire de jouer sur la *forme* du message. Il ne faut pas faire un amalgame entre cette fonction et les canons de la poésie : dès que le locuteur effectue un choix quant à la forme que prend son message, alors la fonction poétique est activée – et ce sans qu'aucun critère (d'esthétisme ou autre) ne vienne décider du caractère véritablement poétique de son message. Lorsqu'il choisit d'énoncer une histoire avec un rythme particulier, lorsqu'il effectue un jeu de mots, lorsqu'il choisit un mot plutôt qu'un autre car il lui préfère sa forme, le locuteur est dans la dimension poétique du langage.

Il convient toutefois de rappeler que, comme l'a dit Jakobson, dans la plupart des énoncés que nous produisons plusieurs fonctions coexistent. Par exemple un énoncé tel que « Humm, j'aime les œufs de Meaux ! » implique d'une part l'expression d'une sensation d'appétit (« Humm »), la référence à un élément du contexte (des œufs), l'activation de la fonction poétique (à travers le jeu de mots induit par la proximité phonétique entre [zødəmo] et [ʒødəmo] (*jeux de mots*) et la mise en jeu de la dimension conative avec l'intention du locuteur de faire rire son interlocuteur.

Si l'on se base sur les différentes fonctions définies par Jakobson, nous pouvons nous apercevoir que les tests d'intelligibilité reposent principalement sur la fonction *métalinguistique* du langage. Ceci est particulièrement clair si l'on observe les phrases porteuses utilisées dans ces tests. Dans la feuille de passation donnée en exemple

dans la figure 1.2 (p. 17), nous trouvons des phrases comme « I am about to say [syllabe cible] » (« Je suis sur le point de dire [syllabe cible] »), « I continue with [syllabe cible] » (« Je continue avec [syllabe cible] »), « These sounds are [syllabe cible] » (« Ces sons sont [syllabe cible] »), etc. Toutes ces phrases utilisent la fonction récursive du langage pour faire référence aux éléments de code représentés par les syllabes cible.

En conclusion, nous avons pu voir dans ce chapitre que les premiers tests d'intelligibilité ont été développés pour répondre à des objectifs particuliers liés à l'évaluation de canaux de communication et de systèmes de transmission. L'éclairage apporté par la théorie mathématique de la communication (Shannon et Weaver, 1949) et par les fonctions du langage définies par Jakobson (1963) nous a permis de mieux cerner le point de vue sur la communication qui avait été adopté lors de la création de ces premiers tests. Ils mettent en effet l'accent sur le code linguistique, dont ils évaluent le transfert en comparant les signaux émis par le locuteur et les signaux retranscrits par les auditeurs – l'objectif étant de quantifier les pertes provoquées par le « bruit » intervenant dans le canal de communication.

Il était nécessaire de bien souligner le contexte dans lequel ont été élaborés les premiers tests d'intelligibilité dans la mesure où ce que nous avons identifié comme le « noyau », l'aspect fondamental commun à tous les tests d'intelligibilité développés dans ce cadre – à savoir le *paradigme de lecture et de retranscription*, a ensuite été porté au domaine clinique pour l'évaluation des performances de patients souffrant de troubles pathologiques de production de la parole (Tikofsky et Tikofsky, 1964).

Chapitre 2

Méthodes d'évaluation et de mesure de l'intelligibilité pour les troubles pathologiques de production de la parole (TPPP)

Sommaire

2.1	Introduction	36
2.2	Tâches d'identification de stimuli	38
2.3	Tâches d'évaluation sur des échelles subjectives	40
2.3.1	Échelles à intervalles réguliers (EIR)	40
2.3.2	Estimations directes de l'intelligibilité (EDI)	42
2.4	Facteurs influençant les scores d'intelligibilité	44
2.4.1	Caractéristiques propres au locuteur	44
2.4.2	Caractéristiques propres à l'auditeur	45
2.4.3	Caractéristiques propres à la situation de communication	48
2.4.4	Caractéristiques propres au matériel linguistique et à la méthode de mesure	49
	Propriétés des stimuli	49
	Procédure de test et calcul des scores	50
2.5	Synthèse et discussion	51

2.1 Introduction

Dans ce chapitre nous traitons des tests d'intelligibilité de la parole utilisés pour quantifier les performances de locuteurs souffrant de troubles pathologiques de production de la parole (TPPP), c'est-à-dire de personnes qui souffrent de troubles pathologiques engendrant des difficultés d'élocution. Ces troubles peuvent être d'origine :

1. *anatomique*. C'est par exemple le cas de personnes présentant des malformations de l'appareil bucco-phonatoire, telles que des *fentes labio-palatines*. La *surdité profonde* s'accompagne également de troubles de production de la parole. Enfin, certaines opérations chirurgicales comme l'ablation du larynx (*laryngectomie*) ou de structures de la cavité buccale – notamment la langue (*glossectomie*) provoquent une perte d'intelligibilité de la parole ;
2. *neurologique*. C'est par exemple le cas des patients atteints de *dysarthrie*, un trouble de la production renvoyant à une faiblesse ou à un manque de coordination dans les mouvements articulatoires (Hosom *et al.*, 2003), et engendré par des atteintes du système nerveux comme la maladie de Parkinson ou la sclérose en plaques.

Contrairement au point de vue des ingénieurs en télécommunications, qui considèrent habituellement qu'un signal de parole tout à fait intelligible est plus ou moins dégradé lors de sa transmission, dans le cas des TPPP le signal est déjà dégradé à la source. La perspective est donc ici très différente de celle des ingénieurs, dans la mesure où le transfert du signal est considéré comme « normal », de même que les capacités auditives et cognitives du récepteur (Kent *et al.*, 1989; Özsancak, 2001).

Aussi dans le cas des TPPP le premier objectif des tests d'intelligibilité est la quantification des *performances des locuteurs*. En effet, les tests d'intelligibilité ont d'abord été conçus dans l'optique de fournir aux cliniciens un moyen de mesurer la *sévérité des cas de TPPP* (Beverly *et al.*, 2010; Kent *et al.*, 1989; Metz *et al.*, 1980; Schiavetti, 1992; Tikofsky et Tikofsky, 1964). Les tests d'intelligibilité continuent à être largement utilisés à cette fin (Beverly *et al.*, 2010; Keintz *et al.*, 2007; Klopfenstein, 2009; Hustad, 2008; Özsancak, 2001) ; l'intelligibilité de la parole est en effet un facteur déterminant pour identifier les TPPP et décider de la prise en charge thérapeutique d'une personne (Ertmer, 2010; Skahan *et al.*, 2007). Plus que cela, l'intelligibilité de la parole est considérée comme un indice de la *performance communicative* des sujets (Beukelman et Yorkston, 1979; DePaul et Kent, 2000), sur lequel peut reposer l'arrêt d'un programme thérapeutique – le sujet étant considéré comme suffisamment performant pour quitter le milieu clinique ou arrêter sa thérapie (DePaul et Kent, 2000; Ertmer, 2010; Skahan *et al.*, 2007).

Enfin, les tests d'intelligibilité de la parole permettent de suivre les progrès des patients dans le temps, ce qui fait que les scores d'intelligibilité sont une variable très souvent mise à contribution en recherche clinique. En effet l'amélioration de l'intelligibilité est le but principal de la plupart des interventions thérapeutiques (Hustad, 2006; Yorkston *et al.*, 1990), et par conséquent l'évolution des scores d'intelligibilité permet d'évaluer le succès :

- d'interventions chirurgicales (*cf.* par ex. Leary *et al.*, 2006; Lohmander *et al.*, 2006) ;
- de traitements médicamenteux (voir par ex. Pinto *et al.*, 2010) ;
- de thérapies orthophoniques (par ex. Neel, 2009; Puyuelo et Rondal, 2005) ;
- de la mise en place de stratégies palliatives, comme par exemple les stratégies de supplémentation de la parole par des gestes (*cf.* Garcia *et al.*, 2000; Garcia et Dagenais, 1998) et divers signaux visuels (*cf.* Hustad, 2007a; Hustad *et al.*, 2003), ou encore la mise en place de systèmes de traitement des signaux de parole produits par les patients (par ex. Hosom *et al.*, 2003; Kain *et al.*, 2007; Neel, 2009).

Dans le domaine clinique, deux grands paradigmes de test sont utilisés pour l'évaluation de l'intelligibilité de la parole (Ertmer, 2010). Ils se distinguent essentiellement par la nature de la tâche assignée aux juges lors des tests. Dans un cas les juges transcrivent les paroles entendues, et dans l'autre ils évaluent directement l'intelligibilité en reportant une note sur une échelle. Samar et Metz (1988, p. 307) se réfèrent à ces deux paradigmes sous les noms de « write-down paradigm » et de « rating-scale paradigm ». Si les procédures de retranscription sont généralement considérées comme fiables et valides, le recours à des échelles subjectives est davantage sujet à controverses (Beltyukova *et al.*, 2008; Ellis et Beltyukova, 2008).

Dans la suite de ce chapitre nous commençons par présenter les tâches de retranscription orthographique, qui s'inscrivent dans la continuité directe des mesures subjectives de l'intelligibilité dont nous avons parlé dans le chapitre précédent (*cf.* section 1.2, p. 14 et section 1.3, p. 18). Ensuite nous abordons les procédures d'évaluation subjective de l'intelligibilité (section 2.3 p. 40), avec en premier lieu les échelles à intervalles réguliers (section 2.3.1, p. 40) puis les procédures d'estimation directe de l'intelligibilité (section 2.3.2, p. 42). Nous présentons enfin les principaux facteurs influençant les scores d'intelligibilité de la parole, et qui ne sont pas liés aux performances des locuteurs en termes de production de la parole (section 2.4, p. 44).

2.2 Tâches d'identification de stimuli

Il existe deux grands types de tests d'identification de stimuli par des auditeurs (Belyukova *et al.*, 2008; Ertmer, 2010) : a) les tâches de retranscription orthographique et b) les tâches de sélection d'items parmi des ensembles fermés de stimuli (écrits ou picturaux).

a) **Les tâches de retranscription** sont essentiellement de la même nature que celles utilisées en acoustique et dans les télécommunications pour effectuer des mesures subjectives de l'intelligibilité (*cf.* chapitre 1). Tikofsky et Tikofsky (1964) sont souvent cités comme les premiers auteurs à avoir porté ce paradigme de test dans le domaine de l'évaluation de la parole pathologique, en l'occurrence pour l'évaluation de la parole dysarthrique.

Hustad (2008, p. 563) donne une définition très claire des tâches de retranscription, et qui correspond bien aux procédures que l'on peut rencontrer dans la littérature :

Il y a différentes façons de mesurer l'intelligibilité. Une des méthodes consiste à demander à des auditeurs naïfs d'effectuer la transcription orthographique d'échantillons de paroles standards (*cf.* Garcia et Cannito, 1996; Giolas et Epstein, 1963; Tikofsky et Tikofsky, 1964; Yorkston et Beukelman, 1981). Dans ce paradigme, les auditeurs entendent un échantillon de parole (généralement de l'ordre de la phrase) et écrivent ce qu'ils pensent que le locuteur a dit. Les mots écrits par les auditeurs sont comptés comme corrects ou incorrects selon qu'ils correspondent aux mots que le locuteur a voulu prononcer. Le pourcentage d'intelligibilité est ensuite calculé en divisant le nombre de mots correctement identifiés par le nombre total de mots, multiplié par 100.¹

Il est possible de distinguer entre différentes tâches de retranscription en fonction :

- des types de stimuli lus par les locuteurs : listes de mots (par ex. dans le protocole d'Évaluation Clinique de la Dysarthrie – *ECD*, Auzou *et al.*, 1998), listes de phrases (par ex. dans l'*ECD*, Auzou *et al.*, ou dans le Frenchay Dysarthria Assesment – *FDA*, Enderby, 1983), paragraphes (par ex. dans l'étude de Hustad, 2008) ;
- des éléments qui sont comptabilisés dans le calcul des scores d'intelligibilité : en général le pourcentage de mots correctement reconnus (par ex. Hustad,

1. « There are several ways that intelligibility can be measured. One method is orthographic transcription of standard speech samples by naive listeners (see Garcia & Cannito, 1996; Giolas & Epstein, 1963; Tikofsky & Tikofsky, 1964; Yorkston & Beukelman, 1981). In this paradigm, listeners hear a speech sample (usually sentence length) and then write down what they thought the speaker said. Constituent transcribed words are scored as either correct or incorrect on the basis of whether they match the intended words of the speaker. Percentage of intelligibility scores are calculated by dividing the number of words identified correctly by the number of words possible, multiplied by 100. » (Hustad, 2008, p. 563).

2008), mais parfois le pourcentage de *phrases* correctement reconnues (par ex. dans le Dysarthia Profile, Robertson, 1982). Dans le cas de la retranscription de phrases ou paragraphes, il est également possible de trouver des procédures qui consistent à n'observer que la retranscription des mots apportant une information sur le contenu sémantique des énoncés (Hustad, 2006) ;

- la façon dont on juge qu'il y a correspondance entre les items prononcés par le locuteur et les items retranscrits par l'auditeur. Si le critère retenu est toujours la correspondance phonémique, il existe une procédure consistant à observer si la correspondance est parfaite (*total phonetic match*) et une autre – beaucoup plus rare – ne tient pas compte de « petits » changements tels que des variations au niveau morphologique affectant les informations de genre et de nombre² (*cf.* Liss *et al.*, 2002).

En dépit des différences que nous venons d'évoquer, et d'après ce que nous avons pu constater dans la littérature, dans la grande majorité des cas les scores de retranscription orthographique se présentent sous la forme de *pourcentage de mots dont tous les phonèmes correspondent avec les sons censés avoir été prononcés par le locuteur*.

Tout comme dans le domaine des télécommunications, les tests de retranscription sont souvent décrits comme étant très lourds à mettre en place. La seconde critique importante qui a été faite à ces tests est qu'ils ne permettent pas de déterminer quels sont les déficits phonétiques sous-jacents expliquant les scores d'intelligibilité (Kent *et al.*, 1989).

b) Les tâches de sélection d'items parmi des ensembles fermés de stimuli permettent de réduire le temps nécessaire aux passations (les auditeurs n'ont plus besoin d'écrire), et parfois d'identifier les phonèmes ou les contrastes phonétiques (les traits distinctifs) qui posent problème aux auditeurs dans les stimuli.

Dans ces tests les auditeurs disposent d'un éventail d'items (soit des images, soit des mots imprimés), parmi lequel ils doivent identifier le stimulus qu'ils entendent (généralement un mot). Le score d'intelligibilité correspond alors au pourcentage de stimuli choisis par l'auditeur et correspondant à ceux prononcés par le locuteur (Beltyukova *et al.*, 2008; Ertmer, 2010; Hosom *et al.*, 2003; Hustad, 2006; Yorkston et Beukelman, 1978).

Pour ce qui est du caractère explicatif de ces tests, l'adaptation du *Word Intelligibility Test* de Kent *et al.* (1989) pour le français par Gentil (1992) fournit un exemple concret. Ce test se présente sous la forme d'un questionnaire à choix multiples. À chaque item un mot est entendu, et quatre choix sont possibles. Les éléments distracteurs ont été sélectionnés avec l'objectif de représenter 17 contrastes phonétiques. À travers les erreurs commises par les auditeurs il est donc possible d'identifier les contrastes phonétiques les plus problématiques. Le test sera plus tard repris par Auzou *et al.* (1998) et par Crochemore et Vannier (2001).

2. ainsi si le locuteur prononce « petit » ([pəti]) et que l'auditeur écrit « petite », il y a correspondance.

2.3 Tâches d'évaluation sur des échelles subjectives

La première raison expliquant l'utilisation d'évaluations sur des échelles subjectives est historique : de nombreux aspects liés à la perception de la voix ou de la parole pathologique ont été et continuent à être évalués sur des échelles subjectives. C'est le cas par exemple pour des dimensions telles que la « nasalité » (*nasality*), la « rugosité » (*roughness*), ou encore la « bizarrerie » (*bizarreness*) de la voix (Schiavetti, 1992; Weismer et Laures, 2002). C'est donc tout naturellement que l'intelligibilité de la parole a trouvé sa place dans ce faisceau de dimensions évaluées de façon directe par les cliniciens.

D'autre part les protocoles d'évaluation sur des échelles subjectives sont souvent considérés comme simples et rapides à mettre en œuvre, deux avantages déterminants en pratique clinique. Ils se présentent donc comme une alternative de choix face à des méthodes demandant davantage de temps³ ou de moyens (Ertmer, 2010; Metz *et al.*, 1980; Samar et Metz, 1988; Schiavetti, 1992).

2.3.1 Échelles à intervalles réguliers (EIR)

Ces procédures consistent à juger des échantillons de parole sur la base d'échelles prédéfinies. Ces échelles représentent le continuum entre une intelligibilité nulle et parfaite, et sont divisées en intervalles réguliers ; le nombre de valeurs possibles est habituellement impair (5, 7 ou 9), ce qui donne la possibilité aux évaluateurs de se référer à une valeur centrale correspondant à une intelligibilité « moyenne » (Schiavetti, 1992). Selon les échelles employées, les valeurs peuvent être accompagnées de descriptions qualitatives (par ex. 1–*Complètement inintelligible* et 5–*Complètement intelligible*) ou non (Schiavetti, 1992; Ertmer, 2010). Le tableau 2.1 présente les descripteurs associés à l'échelle du *NTID Read Intelligibility Test* développée par le *National Technical Institute for the Deaf* (NTID) aux U.S.A.

La validité des échelles à intervalles réguliers pour l'estimation de l'intelligibilité de la parole a souvent été mise en cause dans la littérature, et ce pour deux raisons principales (Belyukova *et al.*, 2008; Ellis et Belyukova, 2008; Özsancak, 2001; Samar et Metz, 1988; Schiavetti, 1992; Whitehill, 2000). En premier lieu, Schiavetti *et al.* (1981) ont présenté des données psychométriques qui tendent à invalider le recours à des échelles à intervalles réguliers pour mesurer le continuum de l'intelligibilité. Leur étude s'appuie sur les travaux de S. S. Stevens, qui distingue entre deux types de continua perceptifs : les continua *prothétiques* et les continua *métathétiques*. Les continua prothétiques impliquent un mécanisme de type *additif* au niveau physiologique – l'excitation sensorielle est « ajoutée » à l'excitation déjà présente, alors que les continua métathétiques reposent sur un traitement physiologique de type *substitutif* – l'excitation sensorielle remplace l'excitation déjà présente (Stevens et Galanter, 1957, p. 377). Les continua prothétiques, tels que la *sonie*⁴, sont consi-

3. Samar et Metz (1988, p. 315) rapportent cependant un temps d'administration *per capita* équivalent pour une tâche de retranscription de mots et une évaluation subjective.

4. intensité perçue.

TABLE 2.1 – Échelle du *NTID Read Intelligibility Test* (adapté de Samar et Metz, 1988)

Score	Descripteur
1	Speech is completely unintelligible ^a
2	Speech is very difficult to understand, only isolated words or phrases are intelligible ^b
3	With difficulty the listener can understand about half the content of the message (Intelligibility may improve over a listening period) ^c
4	Speech is intelligible with the exception of a few words or phrases ^d
5	Speech is completely intelligible ^e

a. la parole est totalement inintelligible.

b. la parole est très difficile à comprendre, seuls des mots isolés ou des phrases sont intelligibles.

c. avec difficulté, l'auditeur peut comprendre environ la moitié du contenu du message (l'intelligibilité peut s'améliorer après une période d'écoute).

d. la parole est intelligible, à l'exception de quelques mots ou phrases.

e. la parole est totalement intelligible.

dérés comme des dimensions *quantitatives* dans lesquelles les valeurs s'actualisent sous la forme de *degrés de magnitude* ; à l'inverse, les continua métathétiques comme la *hauteur*⁵ sont considérés comme des dimensions *qualitatives* dans lesquelles les valeurs s'actualisent sous la forme de *positions* (Whitehill, 2000).

De nombreuses expériences ont démontré qu'il était impossible à des observateurs de partitionner un continuum prothétique en intervalles réguliers : invariablement les sujets tendent à diviser la partie inférieure du continuum en intervalles plus fins que pour la partie supérieure. En d'autres termes, les intervalles le long d'un continuum prothétique ne sont pas équivalents, ce qui rend l'utilisation d'une échelle à intervalles réguliers inadaptée (Schiavetti, 1992, p. 22). Or les données collectées par Schiavetti *et al.* (1981) paraissent démontrer de manière assez claire que l'intelligibilité est un continuum perceptif prothétique plutôt que métathétique, ce qui les amène à conclure que la méthode des échelles à intervalles réguliers est inappropriée dans le cas de l'intelligibilité de la parole et à suggérer plutôt l'utilisation de la méthode d'estimation directe (*Direct Magnitude Estimation* – DME).

Le second argument avancé en défaveur de l'utilisation d'échelles à intervalles réguliers est le manque de sensibilité de cette méthode, en particulier lorsqu'il s'agit de

5. fréquence perçue.

juger des échantillons de parole dont le degré d'intelligibilité est modéré (Schiavetti, 1992). Ce constat résulte de plusieurs études mettant en rapport les scores d'intelligibilité obtenus par des tâches de retranscription avec les scores moyens obtenus par des évaluations sur des échelles à intervalles réguliers.

Ainsi Yorkston et Beukelman (1978) notent que la méthode des intervalles réguliers favorise les extrêmes de l'échelle : les locuteurs tendent à se regrouper en deux catégories, les « plus intelligibles » et les « moins intelligibles » (Yorkston et Beukelman, 1978, p. 504). De même, Samar et Metz (1988) relèvent une corrélation très faible entre les scores d'intelligibilité obtenus par une échelle et ceux obtenus par la retranscription orthographique – et ce particulièrement dans le cas d'une intelligibilité moyenne. Enfin, Beukelman et Yorkston (1980) notent une tendance chez les juges à surestimer l'intelligibilité lorsqu'ils évaluent des locuteurs d'intelligibilité moyenne par la méthode des intervalles réguliers. Tous ces différents points sont d'autant plus importants que les atteintes moyennes sont les plus fréquentes (Samar et Metz, 1988).

En conséquence, un grand nombre de chercheurs préfère recourir aux procédures d'estimation directe de l'intelligibilité de la parole plutôt qu'aux échelles à intervalles réguliers (Beltyukova *et al.*, 2008).

2.3.2 Estimations directes de l'intelligibilité (EDI)

Tout comme pour les EIR, les estimations directes de l'intelligibilité (désormais EDI) consistent à noter des échantillons de parole en leur attribuant une valeur correspondant au taux d'intelligibilité perçu. Cependant les EDI n'imposent pas de valeurs minimum et maximum à chaque extrémité du continuum de l'intelligibilité, et les juges ne sont pas contraints de situer leurs réponses dans une partition linéaire de ce continuum : chaque échantillon est évalué en fonction des taux d'intelligibilité attribués aux autres échantillons (Schiavetti, 1992).

Il existe différentes variantes des EDI. Tout d'abord, ces procédures peuvent ou non comporter un échantillon de parole standard, un « étalon » avec lequel les juges comparent les autres stimuli (Beltyukova *et al.*, 2008; Schiavetti, 1992; Weismer et Laures, 2002). Ce stimulus de référence est habituellement choisi parmi les stimuli d'intelligibilité moyenne (Schiavetti, 1992), mais l'expérimentateur peut aussi bien décider de sélectionner un stimulus d'intelligibilité basse ou haute (Weismer et Laures, 2002). Ces échantillons de référence sont généralement accompagnés d'un *modulus*, c'est-à-dire d'une valeur numérique qui leur est associée ; typiquement, lorsque le standard est choisi parmi les stimuli d'intelligibilité moyenne, le modulus est de 10 ou 100 (Schiavetti, 1992). Parmi les procédures d'EDI comportant un stimulus standard, une autre distinction peut être faite selon que le standard est choisi par l'expérimentateur, ou bien sélectionné par le juge parmi l'ensemble des stimuli (Beltyukova *et al.*, 2008). Enfin, à la fin des procédures d'EDI l'expérimentateur peut éventuellement demander aux juges d'explicitier la manière dont ils ont noté les différents stimuli (Beltyukova *et al.*, 2008).

Selon Özsancak (2001), une des procédures d'EDI les plus utilisées consiste à noter les échantillons de parole en plaçant une croix sur une ligne représentant le

continuum entre une intelligibilité nulle (0%) et une intelligibilité normale (100%). L'emplacement de la croix sur la ligne traduit le degré d'intelligibilité du locuteur.

2.4 Facteurs influençant les scores d'intelligibilité

Dans cette section nous nous intéressons aux facteurs de variation des scores d'intelligibilité qui ne sont pas le fait des performances des locuteurs en terme de production de la parole. Ces facteurs sont extrêmement nombreux (Hustad, 2006, 2008; Keintz *et al.*, 2007). Aussi nous ne prétendons pas à l'exhaustivité : notre objectif est simplement de donner au lecteur une vue d'ensemble sur la diversité des facteurs de variation de l'intelligibilité, que ceux-ci soient liés au locuteur, à l'auditeur, à la situation de communication ou à la procédure de test employée. Nous ne traitons pas des facteurs liés aux caractéristiques du canal de communication (par exemple le type de microphone employé, la réverbération de la salle, etc.) ou aux performances de réception de la parole des auditeurs. En effet, puisque notre sujet de recherche concerne les performances du locuteur uniquement, nous supposons *a priori* que ces derniers facteurs sont contrôlés.

2.4.1 Caractéristiques propres au locuteur

Le dialecte qu'emploie le locuteur peut avoir une incidence sur son intelligibilité (Soli et Wong, 2008). Il est d'ailleurs intéressant de noter que le terme d'*intelligibilité* a aussi une acception en dialectologie – deux parlars étant considérés comme deux variantes d'une seule et même langue si leur *intelligibilité mutuelle* est suffisamment élevée (*cf.* Beijering *et al.*, 2008; Cheng, 1993; Heeringa, 2004; Tang et van Heuven, 2009).

On considère également que l'intelligibilité des locuteurs est fortement liée à la sévérité du trouble dont ils souffrent. Cette question n'est pas évidente : Woisard-Bassols (2011) s'interroge sur la notion de sévérité et note que l'idée de sévérité d'un trouble peut être envisagée sous l'angle des déficiences physiques, comme sous celui des incapacités provoquées par ces déficiences, et même sous celui du handicap social qui en découle. Woisard-Bassols (2011) note aussi que certaines études ont montré une absence de corrélation entre des évaluations subjectives de la sévérité et des mesures subjectives d'intelligibilité de la parole (scores de retranscription orthographique). Nous rajouterons que le fait que la sévérité des troubles soit évaluée par des tests d'intelligibilité peut induire des problèmes méthodologiques, comme nous le verrons lorsque nous aborderons notre problématique liée à la relation entre l'intelligibilité et la compréhension de la parole. Par exemple, Hustad (2007b) reporte un effet hautement significatif de la sévérité sur des scores d'intelligibilité, ce qui n'est pas très surprenant lorsque l'on considère que la sévérité des troubles dont souffrent les locuteurs de cette étude a été évaluée par un autre test d'intelligibilité.

D'autre part, outre l'effort vocal produit par le locuteur et sa langue maternelle – deux facteurs de variation que nous avons déjà introduits dans le chapitre précédent (NF EN ISO 9921, 2004, *cf.* section 1.4.3, p. 24 et section 1.4.4, p. 25) – de nombreuses études ont montré que le débit du locuteur pouvait être un élément influençant les scores d'intelligibilité (pour une revue voir Özsancak, 2001).

Enfin, nous pouvons imaginer que des facteurs physiologiques et psychologiques

comme la fatigue, le stress, l'anxiété, ou diverses émotions peuvent avoir une incidence sur l'intelligibilité du locuteur.

2.4.2 Caractéristiques propres à l'auditeur

Outre les capacités auditives de l'auditeur, Soli et Wong (2008) listent les facteurs suivants :

- comme pour le locuteur, le *dialecte* parlé par l'auditeur ;
- les *compétences linguistiques et cognitives* de l'auditeur (*mémoire* notamment), qui peuvent jouer un rôle plus ou moins important selon la longueur et de la complexité des stimuli utilisés ;
- l'*âge auquel l'auditeur a acquis la langue concernée* ;
- le fait que la langue du test soit la *langue maternelle* de l'auditeur, ou bien seulement une *langue seconde*.

Cependant, les questions qui intéressent Soli et Wong (2008) concernent l'évaluation des performances de l'auditeur, ce qui est différent du cas qui nous occupe. En effet dans le cas de l'évaluation de locuteurs souffrant de TPPP, il arrive fréquemment que l'auditeur et l'évaluateur soient *une seule et même personne*. Ce qui ajoute une autre variable d'importance : l'*expertise de l'auditeur*.

De nombreuses études ont ainsi montré que l'expertise pouvait avoir une incidence sur les scores d'intelligibilité – que ceux-ci soient calculés sur la base de tâches de retranscription ou d'évaluations subjectives (Ertmer, 2010). Si l'on définit le degré d'expertise de l'auditeur comme la somme de connaissances et de compétences que ce dernier peut mettre à profit lors de l'écoute d'un patient, on peut recenser au moins trois grands types d'expertise :

- la familiarité avec les troubles d'élocution dont souffre le locuteur (c.-à-d. avec le type de parole produite) ;
- la familiarité avec le locuteur lui-même ;
- la connaissance – *a priori* – des stimuli.

Plusieurs études ont montré que les scores d'intelligibilité tendent à être plus élevés lorsque les auditeurs sont familiarisés avec les locuteurs, ou bien exposés au type de parole produite. Ces études ont été réalisées auprès de différentes populations de locuteurs : malentendants (ex. Ellis et Belyukova, 2008; McGarr, 1983; Monsen, 1983; Thomas, 1961), dysarthriques (ex. Beukelman et Yorkston, 1980; DePaul et Kent, 2000; Hustad *et al.*, 2003; Tjaden et Liss, 1995), personnes ayant subi une laryngectomie (ex. Williams et Watson, 1985) ou encore les enfants atteints de troubles du développement de la parole (Flipsen, 1995). Le tableau 2.2 reprend sous forme de synthèse les résultats de quelques unes de ces études, qui ont en commun de trouver une incidence positive de l'expertise des auditeurs sur les scores d'intelligibilité.

Cependant, de larges différences apparaissent dans l'ordre de grandeur des effets mis au jour : selon les études l'expertise des auditeurs peut provoquer une augmentation de 3% à 31% des scores obtenus par les patients. Ceci peut s'expliquer par les différences entre les études, qui varient selon plusieurs critères, dont entre autres :

TABLE 2.2 – Exemples d'études montrant un effet de l'expertise des auditeurs sur les scores d'intelligibilité

Étude	Type d'expertise	Incidence sur les scores d'intelligibilité ^a
DePaul et Kent (2000)	Familiarité avec le locuteur (épouse)	de +5% à +30% env. ^b
Ellis et Beltyukova (2008)	Familiarisation avec les stimuli (pré-test)	+3% env. en moyenne
Flipsen (1995)	Familiarité avec le locuteur (parents)	+5% env. en moyenne
Hustad <i>et al.</i> (2003)	Familiarisation avec les stimuli (4 pré-tests)	+11% en moyenne
McGarr (1983)	Entre 1 et 25 ans d'exposition à la parole de malentendants (moy. 6.8)	+25% env. en moyenne
Monsen (1983)	Professeurs pour malentendants ^c	+20% env. en moyenne
Thomas (1961)	Env. 1 an d'exposition à la parole de malentendants	+24% env. en moyenne
	Plus de 1 an d'exposition à la parole de malentendants	+31% env. en moyenne
Tjaden et Liss (1995)	Familiarisation avec les stimuli (pré-test)	+8,9% en moyenne

^a. Par rapport aux scores obtenus auprès d'auditeurs naïfs, c'est-à-dire n'ayant jamais été exposés à la parole pathologique

^b. DePaul et Kent (2000) ont effectué une étude longitudinale : 7 tests ont été échelonnés sur 39 mois

^c. Monsen (1983) étudie les effets de l'expertise selon divers degrés, nous ne rapportons ici les résultats obtenus pour les sujets auditeurs les plus experts

- le profil pathologique des locuteurs : dysarthriques (DePaul et Kent, 2000; Hustad *et al.*, 2003; McGarr, 1983; Monsen, 1983; Thomas, 1961; Tjaden et Liss, 1995), malentendants (Ellis et Beltyukova, 2008), enfants souffrant de troubles du développement de la parole (Flipsen, 1995) ;
- la sévérité des TPPP. Par exemple dans leur étude longitudinale, DePaul et Kent (2000) observent seulement 5% de différence entre les scores obtenus par un auditeur expert (l'épouse du locuteur) et un groupe d'auditeurs naïfs au départ de l'étude. Peu à peu, avec la dégradation de la parole chez le locuteur, la différence s'accroît jusqu'à atteindre environ 30% ;
- le type d'expertise dont bénéficie l'auditeur : familiarité avec le TPPP concerné (McGarr, 1983; Monsen, 1983; Thomas, 1961), avec le locuteur (DePaul et Kent, 2000; Flipsen, 1995), ou encore avec les stimuli (Ellis et Beltyukova, 2008; Hustad *et al.*, 2003; Tjaden et Liss, 1995) ;

- l'expertise peut être acquise naturellement (par ex. par la pratique professionnelle, par le milieu familial, *cf.* DePaul et Kent, 2000; Monsen, 1983; Thomas, 1961), ou bien induite par les expérimentateurs (par ex. par un pré-test de familiarisation avec les stimuli, *cf.* Ellis et Beltyukova, 2008; Hustad *et al.*, 2003; Tjaden et Liss, 1995);
- la procédure utilisée : évaluation sur une échelle subjective (Ellis et Beltyukova, 2008), tâches de retranscription de phrases (DePaul et Kent, 2000; Ellis et Beltyukova, 2008; Flipsen, 1995; Hustad *et al.*, 2003; McGarr, 1983; Monsen, 1983; Tjaden et Liss, 1995) ou de pseudo-mots (Thomas, 1961). Ainsi Ellis et Beltyukova (2008) trouvent un effet significatif de l'expertise des auditeurs lorsque l'intelligibilité est évaluée par une tâche de retranscription, effet non significatif quand l'intelligibilité est évaluée au moyen d'une procédure d'estimation directe.

Malgré ces résultats l'influence de l'expertise des auditeurs sur les scores d'intelligibilité est encore peu connue, et continue d'être étudiée aujourd'hui (par ex. Ellis et Beltyukova, 2008). En effet toutes les études sur la familiarité ou la familiarisation des auditeurs n'ont pas démontré d'effet significatif de l'expertise sur les scores d'intelligibilité (par ex. Cullinan *et al.*, 1986; Doyle *et al.*, 1989; Garcia et Cannito, 1996; Keintz *et al.*, 2007; Yorkston et Beukelman, 1983).

Néanmoins il convient d'être prudent quant à l'interprétation des résultats de certaines de ces études. Comme l'a remarqué Hustad *et al.* (2003), l'étude de Yorkston et Beukelman (1983) par exemple n'impliquait que 9 auditeurs, répartis dans trois groupes selon leur degré de familiarisation avec les stimuli ; peut-être que cette population n'était pas suffisamment importante pour obtenir un effet significatif du point de vue statistique. De même, nous pouvons émettre des doutes quant au niveau d'expertise réel des auditeurs « naïfs » convoqués dans les études de Cullinan *et al.* (1986) et de Doyle *et al.* (1989). Les deux études concernent l'intelligibilité de la parole alaryngée ; le groupe d'auditeurs non experts constitué par Cullinan *et al.* (1986) rassemblait des personnels hospitaliers ayant déjà eu des conversations avec des patients laryngectomisés. Pour ce qui est de l'étude de Doyle *et al.* (1989), le groupe d'auditeurs « naïfs » était constitué par des étudiants en première année universitaire de Sciences de la Parole, du Langage et de l'Audition. Si ces derniers ont déclaré ne jamais avoir été exposés à la parole alaryngée, on peut tout de même imaginer que leur orientation et leur parcours universitaire particulier ait pu jouer un rôle dans leurs performances aux tâches de retranscription – rôle ne concernant pas l'expertise des sujets mais leur motivation, et par là l'attention qu'ils ont porté aux stimuli.

2.4.3 Caractéristiques propres à la situation de communication

À notre connaissance peu d'études ont été consacrées à l'influence de paramètres propres à la situation de communication (par ex. la présence *vs* l'absence d'un auditeur, les conditions matérielles, l'environnement physique) sur l'intelligibilité de locuteurs pathologiques. Pourtant il a été montré que les locuteurs s'adaptent à la situation de communication et que de cette adaptation peut résulter une augmentation de l'intelligibilité. C'est notamment le cas pour l'*effet Lombard* (Lombard, 1911) : plus le bruit ambiant est fort, plus le locuteur va augmenter l'intensité de sa voix pour en élever l'audibilité. Il a été montré par Lane et Tranel (1971) que cet effet s'inscrivait pleinement dans une dimension *communicative*, les auteurs concluant que « le locuteur ne modifie pas l'intensité de sa voix pour mieux communiquer avec lui-même, mais plutôt avec les autres »⁶.

Un autre facteur lié à l'environnement de la communication est celui des informations visuelles accessibles à l'auditeur. Plusieurs études ont montré que les scores d'intelligibilité pouvaient changer de manière significative selon que l'auditeur voit ou ne voit pas le visage du locuteur lorsqu'il parle. Hustad et Cahill (2003) ainsi que Keintz *et al.* (2007) signalent par exemple cet effet pour des locuteurs dysarthriques sévères. Les résultats de ces études concordent avec le fait que de nombreux patients font état d'un handicap particulièrement important lors de conversations téléphoniques (Hunter *et al.*, 1991). Le gain d'intelligibilité apporté par les informations visuelles a d'ailleurs été mis à profit par l'implémentation de stratégies de supplémentation et de complétion de la parole par des indices visuels (notamment des gestes ou des vignettes représentant des lettres de l'alphabet, *cf.* Garcia et Cannito, 1996; Garcia et Dagenais, 1998; Hustad et Beukelman, 2001; Hustad *et al.*, 2003; Hustad, 2007a).

Enfin, un troisième effet important a été mis au jour dans le domaine des TPPP : l'*effet de performance* (*performance effect*). Il renvoie au fait que la performance des locuteurs pathologiques diffère significativement lorsque le locuteur communique dans une situation naturelle (quotidienne) ou bien, au contraire, lors d'une évaluation clinique; cet effet, conduisant soit à une augmentation soit à une baisse des scores d'intelligibilité, a notamment été observé pour des locuteurs dysarthriques (voir Keintz *et al.*, 2007). Récemment Goberman *et al.* (2010) ont essayé de déterminer si cet effet était davantage dû à la situation elle-même ou à la présence d'un microphone. Les auteurs ne trouvent pas d'effet statistiquement significatif relatif à la présence du microphone⁷, ce qui suggère que l'effet de performance serait davantage induit par la situation de test clinique en général.

6. « the speaker does not change his voice level to communicate better with himself, but rather with others » Lane et Tranel (1971, p. 692).

7. cet effet est cependant significatif pour le groupe contrôle, composé de locuteurs non pathologiques.

2.4.4 Caractéristiques propres au matériel linguistique et à la méthode de mesure

Propriétés des stimuli

Diverses études ont mis en évidence que certaines caractéristiques des stimuli, externes au signal de parole lui-même, pouvaient avoir une influence significative sur les scores d'intelligibilité (Garcia et Cannito, 1996; Hustad, 2007a,b). Nous pouvons classer ces caractéristiques en trois catégories :

1. les caractéristiques de *forme*, avec en particulier :
 - (a) le type des unités linguistiques produites par le locuteur (mots, phrases, discours). Il a été montré que, lorsque le signal de parole est dégradé, les scores d'intelligibilité sont généralement plus élevés lorsque le locuteur prononce des phrases plutôt que des mots (voir Hustad, 2007a). Hustad (2007b) a montré que cet effet intervenait aussi à un niveau discursif : l'intelligibilité est significativement supérieure lorsqu'un patient prononce un discours que lorsqu'il prononce des listes de phrases, et *a fortiori* des listes de mots ;
 - (b) la longueur des phrases produites par le locuteur. Cette caractéristique se situe à un niveau plus fin que la précédente : plusieurs études ont montré que, lorsqu'un locuteur prononce des listes de phrases, les scores d'intelligibilité peuvent être influencés par le nombre de mots qu'elles contiennent. Ainsi Yorkston et Beukelman (1981) et Beverly *et al.* (2010) ont montré que plus la longueur des phrases était importante, plus les scores d'intelligibilité augmentaient. Beverly *et al.* (2010) rapportent ainsi une différence significative de 5% entre les scores d'intelligibilité correspondant à des phrases courtes (5-9 mots) *vs* des phrases plus longues (10-15 mots).
2. les caractéristiques de *fond*, notamment :
 - (a) la prédictibilité des mots. Plusieurs études ont testé l'influence de la prédictibilité des mots sur leur intelligibilité (par ex. Beverly *et al.*, 2010; Garcia et Cannito, 1996). La prédictibilité est généralement mesurée par des pré-tests de complétion : les sujets lisent des phrases dans lesquelles certains mots sont remplacés par des blancs. Les sujets doivent compléter chaque phrase avec le ou les mots qui leur semblent le(s) plus probable(s) ; la prédictibilité d'un mot dans un contexte phrastique particulier est ensuite calculée comme son pourcentage d'occurrences dans les réponses écrites des sujets. Beverly *et al.* (2010) et Garcia et Cannito (1996) ont ainsi trouvé une augmentation significative de l'intelligibilité lorsque celle-ci porte sur des phrases dont les éléments sont hautement prédictibles ;
 - (b) l'appartenance des items de test à une thématique commune. Garcia et Cannito (1996) ont ainsi montré que les scores d'intelligibilité étaient significativement plus élevés (5% en moyenne) lorsque les items de test

(en l'occurrence des phrases) étaient liées par une thématique commune (ex. les phrases décrivent toutes des actions effectuées à la maison, ou dans le jardin, etc.) que lorsque les phrases étaient indépendantes de ce point de vue.

3. les caractéristiques d'*usage*, notamment :

- (a) la fréquence des mots. Dans leurs articles fondateurs, Rosenzweig et Postman (1957, 1958) ont montré que la fréquence des mots (pour l'anglais et pour le français) avait un effet statistiquement significatif sur l'intelligibilité de la parole – les mots les plus fréquents ayant tendance à être plus facilement reconnus par les sujets. À noter qu'aujourd'hui de plus en plus d'études montrent que l'effet de fréquence pourrait aussi être présent pour des éléments plus larges que les mots, comme les structures lexico-syntaxiques (voir Arnon et Snider, 2010) ;
- (b) la familiarité des sujets avec les mots. Les scores de familiarité sont obtenus par des évaluations subjectives, généralement sur des échelles à 7 points (de 1 – *pas du tout familier* à 7 – *très familier*, cf. Amano *et al.*, 2009). La familiarité des sujets avec les mots semble par ailleurs être un meilleur index de la difficulté posée par les stimuli que la fréquence des mots (Amano *et al.*, 2009; Sakamoto *et al.*, 2004).

Procédure de test et calcul des scores

Yorkston et Beukelman (1978) ont montré que différents scores pouvaient être obtenus selon la procédure de test employée par les auditeurs. Les auteurs ont montré que les scores obtenus pour un même jeu de stimuli augmentaient de manière significative selon le type de test (classement croissant) :

1. transcription orthographique complète des phrases entendues.
2. complétion de phrases (texte à trous).
3. test à choix multiples.

Yorkston et Beukelman (1980) ont également montré que, dans le cas d'un test à choix multiples, les scores variaient en fonction du nombre d'items possibles – les scores d'intelligibilité diminuant de manière significative en fonction de l'augmentation du nombre d'items.

Enfin, Hustad (2006) a obtenu des résultats indiquant que les scores d'intelligibilité peuvent varier selon la façon dont est comptée la correspondance entre les mots prononcés par le locuteur et les mots écrits par l'auditeur. Ainsi, les scores d'intelligibilité obtenus a) en comptant tous les mots dont tous les phonèmes correspondent et b) en comptant seulement les mots apportant une information sémantique et dont tous les phonèmes correspondent et c) seulement les mots apportant une information sémantique et dont la signification correspond à celle du mot qui a été prononcé, donnent des résultats significativement différents du point de vue statistique.

2.5 Synthèse et discussion

Dans ce chapitre nous avons vu que deux grands types de tests d'intelligibilité de la parole sont utilisés dans le cadre des TPPP : ceux portant sur la reconnaissance d'items (généralement des tests de retranscription orthographique, mais parfois aussi des tests à choix multiples), et ceux consistant en des évaluations subjectives de l'intelligibilité – que ce soit sur des échelles à intervalles réguliers ou par une estimation directe de la magnitude. Ce chapitre aura également été l'occasion de présenter les principaux facteurs qui influencent les scores d'intelligibilité et qui ne sont pas le fruit des performances du locuteur en termes de production de la parole. Nous avons regroupé ces facteurs en quatre catégories distinctes selon qu'ils sont plutôt liés au locuteur (dialecte parlé, sévérité du TPPP, ...), à l'auditeur (dialecte parlé, compétences linguistiques et cognitives, expertise dans les TPPP, ...), à la situation de communication (bruit ambiant, multimodalité de la communication, effet de performance) et enfin au matériel (forme, contenu, fréquence et familiarité des stimuli) et à la procédure de test utilisée (type de retranscription effectuée par l'auditeur, critères de jugement des réponses).

Malgré le nombre et la diversité des procédures de test que nous avons présentées dans ce chapitre, il est important de souligner le fait que, comme le dit Woisard-Bassols (2011) les tests de retranscription orthographique demeurent le « gold standard » des tests d'intelligibilité qui sont utilisés dans le cadre des TPPP. Ils sont en effet reconnus comme les plus fiables et les plus valides, les procédures d'évaluation subjective de l'intelligibilité sur des échelles ou par des estimations directes de magnitude étant bien davantage critiquées dans la littérature. La validité des autres tests d'intelligibilité de la parole est à ce titre jugée en comparant les scores obtenus avec ceux des tâches de retranscription orthographique (Beltyukova *et al.*, 2008; Ellis et Beltyukova, 2008).

Nous avons vu dans le chapitre précédent que le « noyau », le trait commun à l'ensemble des tests subjectifs d'intelligibilité de la parole dans le domaine de l'acoustique et des télécommunications était le *paradigme de lecture et de retranscription*. Il ressort de notre état de l'art que cela est aussi le cas dans le domaine des TPPP. Tout comme les tests objectifs et prédictifs permettent de retrouver les scores d'intelligibilité en se passant de jurys d'auditeurs, les évaluations subjectives employées dans le cadre des TPPP sont souvent présentées comme des procédures permettant d'économiser du temps et des moyens lors de l'évaluation de patients souffrant de TPPP.

Dans la mesure où la validité des différents tests d'intelligibilité est jugée en comparaison du « gold standard » des tâches de retranscription, nous en déduisons également qu'il n'est pas offert ici de perspective sur la communication autre que celle qui est représentée par ce type de tâche. Tout comme pour les tests subjectifs d'intelligibilité utilisés en acoustique et dans les télécommunications, dans le domaine des TPPP l'accent est surtout mis sur le *transfert de code linguistique*⁸.

8. cela se voit également dans les phrases porteuses utilisées dans les tests aux mots, qui mettent

Nous avons été étonné de trouver dans le domaine des TPPP des procédures de test d'intelligibilité qui soient aussi proches de celles utilisées par les ingénieurs en télécommunication, et ce malgré les différences concernant les objectifs d'évaluation. Weismer et Martin (1992) soulignent ces différences en disant :

Le concept d'intelligibilité de la parole, dont les origines remontent aux années 1920, provient des recherches en téléphonie. Ces recherches concernaient la qualité des systèmes de transmission de la voix, et les moyens par lesquels elle pouvait être évaluée. Si l'on se place dans le cadre d'un simple modèle de communication ternaire, composé d'un locuteur, d'un *medium* et d'un récepteur, ces recherches se concentraient clairement sur les effets du *medium* sur l'intelligibilité de la parole.⁹

Or dans le cas des TPPP l'« objet » de l'évaluation est le premier élément de ce schéma ternaire de la communication : le *locuteur*. Est-ce que le succès de la communication peut-être évalué de la même manière selon que l'on s'intéresse aux performances d'une personne ou d'un canal de communication ? Le rôle d'un canal de communication est de transmettre de l'information¹⁰. Or dans le cas de la communication humaine, nous avons vu à travers le modèle de Jakobson (*cf.* figure 1.5, p. 31) que la parole pouvait remplir différentes fonctions, comme l'expression de sensations et de sentiments, la référence¹¹ à des objets, à des idées ou encore l'action sur l'interlocuteur.

Dans le cas des tests de retranscription orthographique, la mesure porte sur le transfert des phonèmes, soit sur le transfert de code linguistique – et non sur la compréhension, par l'auditeur, des références faites par le locuteur, ou encore des intentions et des sentiments de celui-ci. Nous verrons dans le chapitre suivant qu'il

en jeu la fonction métalinguistique du langage, au même titre que ce que nous avons observé dans le chapitre précédent. Ainsi Yorkston *et al.* (1990) utilisent des phrases comme « J'ai pensé que c'était "neat" ou "bathe" ; tu as pensé que c'était "rain". » (« I thought it was "neat" or "bathe"; you thought "rain". ») ou « Épelle "pan" et dis "fuel" et "should". » (« Spell "pan", then say "fuel" and "should". »). Ces phrases utilisent clairement la fonction récursive (métalinguistique) du langage pour faire référence aux signifiants présentés entre guillemets.

9. « The concept of speech intelligibility, which had its origins in the 1920s, comes from the telephony literature. This literature was concerned with the quality of voice transmission systems, and the means by which that quality could be assessed. Within the framework of a simple, three-part model of communication, consisting of a speaker, medium and receiver, the emphasis in this early literature was clearly on the effects of medium characteristics of speech intelligibility. » (Weismer et Martin, 1992).

10. nous utilisons le terme *information* avec le sens que lui donnent Shannon et Weaver (1949). Dans leur conception l'information est une mesure tenant compte à la fois de la quantité de symboles transmis et à la fois du degré de liberté de choix lors de l'émission de chaque symbole. L'*information* est donc uniquement relative au *code*, et non à la signification des messages.

11. à noter que les procédures consistant à ne comptabiliser que les mots apportant une information sémantique qui ont été correctement retranscrits par l'auditeur, ou à juger que la correspondance entre les mots prononcés par le locuteur et les mots écrits par l'auditeur est effective si les deux signifiés sont équivalents (*cf.* section 2.4.4, p. 50) démontrent des tentatives de se rapprocher d'une mesure du transfert de signifiés plutôt que du transfert des signifiants.

est très souvent supposé que les scores de retranscription orthographique permettent de prédire les performances du locuteur en situation de communication.

Pourtant, du point de vue expérimental, la nature de la relation existant entre les scores de retranscription orthographique et l'aptitude du locuteur à se faire comprendre en situation de communication sont encore peu connus (Hustad, 2008). Dans le chapitre suivant nous abordons cette question, autour de laquelle s'articule notre problématique.

Chapitre 3

Problématique : intelligibilité et compréhension de la parole en situation de communication

Sommaire

3.1	Introduction	56
3.2	Perspectives psycholinguistiques sur la perception et la compréhension de la parole	59
3.2.1	Les questions de la perception et de la compréhension de la parole dans la conception du langage <i>en tant que produit</i>	60
	De la nature des unités perceptives de base	61
	De la nature ascendante ou descendante des processus de perception et de compréhension de la parole . . .	65
3.2.2	La question de la compréhension de la parole dans la conception du langage <i>en tant qu'action</i>	71
	Principaux concepts pragmatiques	71
	Outils d'analyse de la pragmatique et compréhension de la parole	75
3.3	Quelle perspective sur la compréhension de la parole dans notre travail?	78
3.4	Relation entre les mesures subjectives d'intelligibilité et la compréhension de la parole : état de la question et hypothèses	80
3.4.1	État de la question dans la littérature	80
3.4.2	Hypothèses	83

3.1 Introduction

Dans les deux chapitres précédents, nous avons présenté les différentes méthodes d'évaluation et de mesure de l'intelligibilité, telles qu'elles ont été développées au départ dans le domaine des télécommunications, puis telles qu'elles sont appliquées dans le domaine clinique dans le cadre de la prise en charge de personnes souffrant de troubles pathologiques de production de la parole. Nous avons vu que dans les deux domaines l'évaluation de l'intelligibilité de la parole se base sur le paradigme central constitué par la lecture d'unités linguistiques pour le locuteur, et par la retranscription de ces unités par l'auditeur.

Pourtant, comme nous l'avons souligné le recours à des tests d'intelligibilité dans ces deux domaines répond à des problématiques bien différentes, centrées sur l'évaluation du *canal de communication* dans le premier cas, et sur l'évaluation des performances du *locuteur* dans le second. Ainsi dans le cas des TPPP l'objectif des tests d'intelligibilité est bien souvent de fournir un indice de la *performance communicative* des patients (Ertmer, 2010; Skahan *et al.*, 2007). En effet le but de la plupart des interventions thérapeutiques est de permettre au patient de mieux communiquer par la parole, autrement dit de mieux *se faire comprendre par des tiers* (Keintz *et al.*, 2007; Neel, 2009; Watson et Schlauch, 2009).

Une relation étroite entre les scores obtenus par les tests d'intelligibilité et la compréhension du locuteur par l'auditeur est donc supposée. Dans la littérature, cela transparait de certaines définitions de l'intelligibilité de la parole qui s'appuient sur l'idée de *compréhension de l'auditeur*. Parmi les articles récents, nous pouvons citer Ertmer (2010, p. 1077), qui reprend la définition de l'intelligibilité donnée par Nicolosi *et al.* (1996) : « cet aspect du signal de parole qui permet aux auditeurs de *comprendre* ce qu'un locuteur est en train de dire »¹. Ou encore Keintz *et al.* (2007, p. 223) qui énoncent que « l'intelligibilité de la parole est définie comme la proportion de parole *comprise* »². Ce type de définition s'applique aussi bien à des tâches d'intelligibilité reposant sur la retranscription orthographique qu'à des évaluations sur des échelles subjectives : par exemple Beverly *et al.* (2010, p. 9) indiquent que dans les deux cas « les résultats sont en général des scores représentant la quantité d'information qu'un auditeur a entendu et *compris* »³.

La relation postulée entre les scores d'intelligibilité et la compréhension de l'auditeur apparaît également lorsque les auteurs donnent une interprétation aux scores d'intelligibilité. McColl (2006) présente les résultats de tâches de retranscription en indiquant que les scores représentent « l'acuité avec laquelle les sujets ont *compris*

1. « that aspect of speech-language output that allows a listener to understand what a speaker is saying » (Nicolosi *et al.*, 1996).

2. « speech intelligibility is defined as the amount of speech understood » (Keintz *et al.*, 2007, p. 223). Bien qu'il n'en soit pas fait mention dans l'article de Keintz *et al.* (2007), cette définition correspond à celle donnée dans la norme NF EN ISO 9921 (2004) – Ergonomie – Évaluation de la communication parlée.

3. « The result tends to be a score based on the amount of information that a listener heard and understood correctly. » (Beverly *et al.*, 2010, p. 9).

les stimuli de parole trachéo-oesophagienne »⁴. De même, Peng *et al.* (2004, p. 1233) interprètent les scores obtenus par un groupe de locuteurs malentendants à des tests d'intelligibilité (retranscription et échelle subjective) en déclarant qu'ils étaient capables *d'être compris par des auditeurs naïfs*⁵ (c.-à-d. non familiers avec la parole de sourds ou de malentendants). Mais le fait que les auteurs utilisent les tests d'intelligibilité pour évaluer la compréhension des auditeurs n'est que rarement explicitée de manière aussi claire que, par exemple, dans l'étude de Watson et Schlauch (2009, p. 163), où l'on peut lire que « pour examiner la *compréhension de la parole*, [Watson et Schlauch ont] choisi d'étudier l'intelligibilité de la parole »⁶.

Une autre manière de mettre au jour la relation postulée entre intelligibilité et compréhension de la parole est d'observer les descripteurs accompagnant les échelles subjectives d'intelligibilité et les protocoles d'évaluation directe. Dans des échelles standardisées comme la *NTID read intelligibility scale* (cf. table 2.1, p. 41), les termes d'intelligibilité et de compréhension semblent tout à fait interchangeables : certains scores sont couplés avec un descripteur impliquant la notion d'intelligibilité de la parole, d'autres avec un descripteur traitant de la compréhension, quand intelligibilité et compréhension ne sont pas directement incluses dans un même descripteur (par ex. « Avec difficulté, l'auditeur peut *comprendre* environ la moitié du contenu du message (l'*intelligibilité* peut s'améliorer après une période d'écoute) »⁷. Ce rapprochement des termes d'intelligibilité et de compréhension se retrouve dans les descripteurs d'échelle utilisés par Neel (2009), par exemple pour le score le plus bas « 1 – *Intelligibilité* très basse (la parole est très difficile à *comprendre*) »⁸. De même, les descripteurs utilisés dans les protocoles d'estimation directe de l'intelligibilité peuvent faire appel à la notion de compréhension. C'est par exemple le cas dans l'étude de McHenry (2011) qui donne aux auditeurs pour consigne d'« indiquer dans quelle proportion ils ont *compris* ce que la personne a dit (de '*rien compris*' [...] à '*tout compris*' [...]) »⁹. Les consignes de Ellis et Belyukova (2008) et de Belyukova *et al.* (2008), visant à donner lieu à une estimation de l'intelligibilité, vont dans le même sens :

- « Je vais vous demander de me donner un chiffre qui corresponde à combien la parole était *compréhensible* pour vous »¹⁰ ;

4. « The mean scores and percent correct represent the accuracy with which the subjects understood the TE speech stimuli » (McColl, 2006, p. 609).

5. « In summary, the results suggest that approximately 70% of a particular set of utterances produced by pediatric CI recipients, with 7 years of device experience, could be understood by unfamiliar listeners » (Peng *et al.*, 2004, p. 1233).

6. « To examine speech understanding, we chose to study speech intelligibility » (Watson et Schlauch, 2009, p. 163).

7. « With difficulty the listener can understand about half the content of the message (Intelligibility may improve over a listening period) ».

8. « 1 - Very poor intelligibility (speech is very difficult to understand) » (Neel, 2009, p. 1023).

9. « indicate how much they understood of what the person said (from 'understood none' at 0 to 'understood all' at 229) » (McHenry, 2011, p. 120).

10. « I will ask you to give me a number that goes with how understandable the speech was for you. » (Ellis et Belyukova, 2008, p. 1118).

- « Donnez des chiffres plus élevés pour la parole qui est moins *compréhensible* ; donnez des chiffres plus bas pour la parole qui est plus *compréhensible* »¹¹.

Nous le voyons, dans la littérature les notions d'intelligibilité et de compréhension semblent se confondre. Cependant notre propos est ici bien moins d'ordre terminologique que méthodologique. Et ce, même si nous préférierions que l'intelligibilité soit toujours définie par la méthode employée pour la mesurer ou l'évaluer, par exemple « mesure du pourcentage de mots correctement retranscrits par un auditeur » pour les tests d'intelligibilité portant sur la retranscription de mots. Nous pensons en effet que, dès que l'on définit l'intelligibilité en termes de compréhension de l'auditeur nous sommes déjà dans l'*interprétation de la mesure*. Comme le dit Hustad (2008), il est possible qu'une tâche orientée vers le *code* comme la retranscription orthographique ne donne pas les mêmes résultats qu'une tâche qui implique un niveau de conceptualisation plus élevé chez l'auditeur. Afin de préciser cette problématique de la relation entre l'intelligibilité et la compréhension de la parole, dans la suite de ce chapitre nous nous posons la question de *ce qu'est la compréhension de la parole*.

11. « Give higher numbers for speech that is less understandable; give lower numbers for speech that is more understandable. » (Belyukova *et al.*, 2008, p. 1129).

3.2 Perspectives psycholinguistiques sur la perception et la compréhension de la parole

La question des représentations et des mécanismes impliqués dans la compréhension de la parole a été traitée de manière différente au sein de deux grandes écoles théoriques (Tanenhaus et Brown-Schmidt, 2008; Treiman *et al.*, 2003). La première école, que Tanenhaus et Brown-Schmidt (2008) appellent la tradition du *langage en tant que produit* (*language-as-product*), s'est focalisée sur les processus cognitifs individuels par lesquels les auditeurs activent les représentations linguistiques – les « produits » de la compréhension du langage. La seconde école, que Tanenhaus et Brown-Schmidt (2008) désignent sous le nom de tradition du *langage en tant qu'action* (*language-as-action*) conçoit le langage comme un moyen d'agir dans une situation de communication, et a insisté sur les rôles que jouent le contexte de communication (par ex. les intentions des participants, le contexte spatio-temporel) et les connaissances de l'auditeur (par ex. de son interlocuteur ou du monde en général) dans la compréhension de la parole (Treiman *et al.*, 2003).

Ces deux paradigmes sont ancrés dans deux écoles théoriques distinctes, et ont utilisé des méthodes d'investigation de natures différentes. Les théoriciens de la tradition du langage en tant que produit s'inscrivent dans la continuité du paradigme du traitement de l'information représenté entre autres par les travaux de Miller (1962), et de la théorie chomskyenne de la grammaire transformationnelle (Miller et Chomsky, 1963). Ces travaux s'appuient sur le constat que l'être humain est capable de traiter une infinité de nouveaux messages linguistiques, et que les processus de reconnaissance et d'interprétation linguistique s'effectuent avec une vitesse remarquable (Treiman *et al.*, 2003). La parole a un caractère linéaire – elle se « déroule » dans le temps, et les auditeurs font des choix provisoires à divers niveaux de représentation (par ex. phonétique, lexical, syntaxique) au fur et à mesure qu'ils reçoivent l'input acoustique (Marslen-Wilson, 1973, 1975). Les travaux des théoriciens du langage comme produit se sont concentrés sur cette dimension temporelle du traitement du code linguistique, en utilisant généralement des mesures de temps de réaction dans des tâches (habituellement d'ordre métalinguistique) de traitement d'unités linguistiques (ex. phonèmes, mots, phrases) en temps réel¹² (Tanenhaus et Brown-Schmidt, 2008).

Les travaux s'inscrivant dans la tradition du langage en tant qu'action ont pour origine les théories pragmatiques développées à l'école d'Oxford (Austin, 1962; Grice, 1957; Searle, 1969), ainsi que les théories et méthodes développées au sein du courant de l'analyse conversationnelle (Schegloff et Sacks, 1973). Leurs travaux se focalisent sur la façon dont les êtres humains utilisent le langage pour exécuter des actions (des *actes de langage*) dans des situations de communication réelles. Ainsi dans ce courant théorique les protocoles se présentent généralement sous la forme d'interactions verbales entre plusieurs personnes qui sont amenées à collaborer ensemble pour

12. ces méthodes incluent en particulier les tâches de *gating*, de décision lexicale ou encore d'amorçage transmodal – voir Nguyen (2001) pour une description éclairante.

une tâche bien définie, et dans un environnement physique contrôlé (ex. présence d'objets réels dans un environnement en partie partagé par les interlocuteurs). Ces objectifs et méthodes répondent au constat selon lequel de nombreux aspects des énoncés verbaux ne peuvent être compris que si l'on tient compte du contexte dans lequel ils sont produits. Dans ce cadre la prise en compte des intentions communicatives des interlocuteurs – notamment de par la nature de la tâche collaborative dans laquelle ils sont impliqués – ainsi que du cadre spatio-temporel dans lequel prend place l'interaction revêt un intérêt particulier (Tanenhaus et Brown-Schmidt, 2008).

Ces deux perspectives sur la compréhension de la parole ne s'opposent pas, toutes deux ont pour objet des mécanismes à des niveaux différents et mettent en jeu des types de protocoles dont chacun présente un intérêt particulier pour notre sujet d'étude. Nous avons soulevé la question de la validité des tests d'intelligibilité pour révéler la dimension particulière de la compréhension de la parole. Les travaux menés dans le paradigme du traitement de l'information peuvent nous éclairer sur une partie des mécanismes physiologiques et cognitifs primaires qui sont en jeu dans le traitement du signal acoustico-phonétique, et nous permettre ainsi de nous interroger sur la pertinence des stimuli et des procédures d'évaluation utilisés. Ensuite, puisque l'objectif des procédures thérapeutiques est de rendre l'individu « fonctionnel », c'est-à-dire performant dans les situations de communication verbale, le cadre théorique de la pragmatique offre une perspective dépassant le niveau du code linguistique, et permettant de mettre au jour toutes les dimensions inhérentes à la plupart des situations de communication.

3.2.1 Les questions de la perception et de la compréhension de la parole dans la conception du langage *en tant que produit*

La tradition du *langage en tant que produit* s'est attachée à définir les représentations et les processus cognitifs permettant à l'être humain de résoudre la quantité d'ambiguïtés qui se présente à lui lorsqu'il écoute un stimulus de parole. Par exemple, pour ne citer que le niveau lexical, lorsqu'un auditeur perçoit l'incipit /taʁ/ d'un mot comme *tartiflette*, beaucoup d'autres candidats lexicaux sont possibles (ex. *Tarzan*, *tariquet*) ; comment cette ambiguïté temporaire est-elle traitée ? Existe-t-il des mécanismes cognitifs permettant d'activer provisoirement des candidats lexicaux puis, au fur et à mesure de l'arrivée de l'input acoustique, d'inhiber les candidats non conformes aux données d'entrée¹³ ? C'est un des problèmes typiques posés au sein de ce courant de recherche, dont les préoccupations sont résolument tournées vers le traitement de l'information.

De nombreuses études ont porté sur les processus de *reconnaissance des mots parlés*. Cet intérêt peut s'expliquer par le fait que la question renvoie à une notion centrale de la linguistique traditionnelle, celle du *signe* linguistique, composé

13. C'est ce qui est postulé dans des modèles tels que celui de la COHORTE (Marslen-Wilson et Welsh, 1978).

d'un *signifiant* (la forme acoustique du mot) et d'un *signifié* (le sens du mot). La façon dont un auditeur parvient à apparier ces deux facettes du signe linguistique a, depuis les débuts de la psycholinguistique, trouvé une place de choix dans les études – jusqu'à occuper une place privilégiée dans le champ de la compréhension de la parole. Ces idées se trouvent particulièrement bien explicitées par Norris *et al.* (2006) lorsqu'ils énoncent : « Comprendre la parole nécessite la conversion du signal sonore en signification(s). Les mots sont les unités de sens exprimées dans le signal sonore. En conséquence, le processus de reconnaissance des mots est le coeur de la compréhension de la parole. »¹⁴.

Les études sur la reconnaissance des mots parlés peuvent apporter un éclairage intéressant pour notre sujet d'étude. En effet nous avons vu dans le chapitre précédent que les mesures subjectives d'intelligibilité accordaient une importance particulière à l'unité qu'est le *phonème*. C'est en effet la correspondance phonémique qui est le critère de jugement dans les tâches de retranscription, et les tests dits *explicatifs* de l'intelligibilité se basent sur l'aptitude de l'auditeur à recouvrir le caractère distinctif des phonèmes prononcés par le locuteur. De plus, la tâche même de retranscription orthographique implique certainement pour l'auditeur de se concentrer sur les aspects segmentaux de la parole. Il est donc important de voir dans quelle mesure cette place qui est accordée au phonème dans les tests d'intelligibilité de la parole se justifie dans la littérature ; nous verrons ainsi dans la section *De la nature des unités perceptives de base* qu'il n'existe pas, aujourd'hui, de consensus autour des unités de base servant à la reconnaissance des mots parlés.

Ensuite, dans notre partie intitulée *De la nature ascendante ou descendante des processus de perception et de compréhension de la parole* (p. 65) nous aborderons la question des attentes perceptives et cognitives de l'auditeur lorsqu'il écoute des énoncés verbaux. Cette question est également d'un intérêt particulier pour le sujet qui nous concerne. Nous avons vu que dans les mesures d'intelligibilité pratiquées pour les TPPP très peu d'attentes sont impliquées chez l'auditeur – l'objectif étant plutôt de *contrôler* les effets produits sur l'intelligibilité par des facteurs externes à la performance du locuteur en termes de production de la parole. Or nous verrons dans cette partie qu'aujourd'hui de plus en plus d'études semblent montrer que la prise en compte d'éléments externes au signal de parole, telles que le *contexte*, est *inhérent* au processus de compréhension de la parole. Contrôler les effets produits par la prise en compte d'un contexte lors du traitement de la parole pourrait donc être vu comme un élément compromettant la validité externe de la mesure.

De la nature des unités perceptives de base

Nous avons vu qu'au sein du paradigme du traitement de l'information, la question de la perception et de la compréhension de la parole a principalement été traitée en considérant les mécanismes permettant d'apparier un signal acoustique avec une

14. « Understanding spoken language requires the conversion of sound to meaning. Words are the units of meaning expressed in sound. The process of word recognition is thus the heart of language understanding. » (Norris *et al.*, 2006, p. 147).

unité signifiante – et plus spécifiquement le mot. De nombreux modèles ont été élaborés pour tenter d'expliquer ces mécanismes. Malgré leurs divergences, la plupart de ces modèles envisagent l'accès au sens comme une succession d'étapes faisant intervenir un ou plusieurs types de représentations infra-lexicales, c'est-à-dire des représentations intermédiaires entre le signal acoustique et la forme que prend le lexique dans la cognition. Ainsi le signal acoustique perçu par l'auditeur est tout d'abord « transformé » sous cette forme intermédiaire (généralement par des opérations de segmentation et de catégorisation) avant d'être apparié à l'unité lexicale dans la mémoire à long terme de l'auditeur, activant ainsi la signification de l'unité linguistique.

La majorité des modèles de traitement de la parole envisagent les représentations infra-lexicales comme composées d'unités plus élémentaires, correspondant généralement à des unités définies par la phonologie structurale comme le phonème, la syllabe ou le trait (Frauenfelder et Nguyen, 2000). Ainsi le phonème a longtemps été considéré comme l'unité de base des représentations infra-lexicales (Nguyen, 2005) ; les raisons principales motivant ce choix sont les suivantes (Frauenfelder et Nguyen, 2000) :

- il permet de concevoir un lexique mental occupant un espace réduit dans la mémoire à long terme, puisque seuls les phonèmes correspondant à chaque unité sont stockés. Cependant on sait aujourd'hui que la mémoire à long terme a une capacité beaucoup plus importante que ce que l'on supposait alors ;
- il permet d'élaborer des modèles en accord avec les théories phonologiques, qui ont eu une grande influence sur les travaux menés en psycholinguistique.

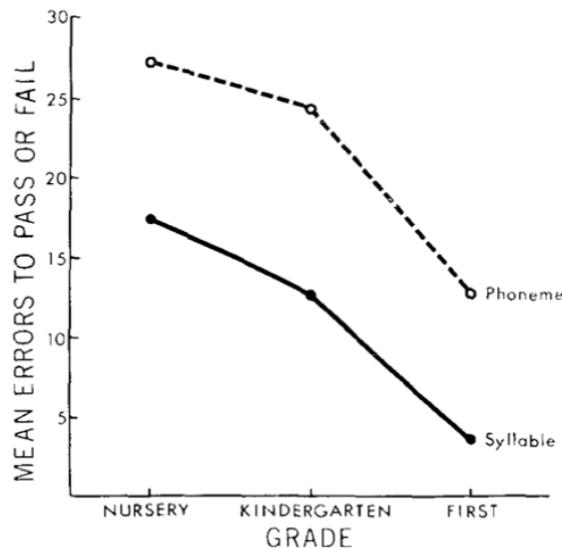
Mais il existe plusieurs arguments en défaveur de l'unité phonémique comme unité de base pour la segmentation de la parole. Tout d'abord, sur le plan acoustique les phénomènes de coarticulation engendrent une confusion des limites des phonèmes qui rend la segmentation difficilement concevable (Frauenfelder et Nguyen, 2000). Ensuite, un argument souvent avancé est que, sur le plan historique, les systèmes de retranscription phonémiques (systèmes alphabétiques) ne sont apparus que très tardivement, semblant démontrer que le phonème n'est peut-être pas l'unité de segmentation la plus « naturelle » qui soit.

Ces deux points trouvent une réponse dans l'hypothèse de l'unité syllabique. En effet pour ce qui est des phénomènes de coarticulation, comme le soulignent Frauenfelder et Nguyen (2000), l'influence inter-phonémique est beaucoup plus importante à l'intérieur de la syllabe qu'elle ne l'est à la frontière entre deux syllabes. Autrement dit, la production d'un phonème est davantage sujette à variation du fait des phonèmes environnants que ne l'est celle de la syllabe face aux syllabes adjacentes. Ensuite, pour revenir à l'argument historique, les systèmes syllabiques sont apparus bien plus tôt que les systèmes alphabétiques ; ce sont en effet les premiers systèmes d'écriture consistant à retranscrire les sons de parole, et non plus les signifiés comme pour les systèmes pictographiques et logographiques.

L'hypothèse syllabique trouve également un argument important sur le plan développemental. Liberman *et al.* (1974) ont montré que des enfants éprouvaient

moins de difficulté à segmenter des unités linguistiques en syllabes qu'à parvenir à un degré d'abstraction phonémique (cf. figure 3.1).

FIGURE 3.1 – Erreurs commises lors de tâches de segmentation en phonèmes et syllabes par des enfants (Liberman *et al.*, 1974)



Cette primauté de la syllabe sur le phonème dans la perception de la parole a également été montrée dans des tâches en temps réel. À travers plusieurs expériences de détection de fragments de parole, Mehler (1981) a montré qu'une séquence cible de phonèmes était détectée plus rapidement lorsqu'elle correspondait à la première syllabe d'un mot que lorsqu'elle n'y correspondait pas. Ainsi, la cible visuelle « bal » sera détectée plus rapidement dans le stimulus [bal/kõ] (*balcon*) que dans [ba/lã] (*balance*). Les temps de réponse ne dépendraient donc pas de la longueur de la séquence (nombre de phonèmes), mais plutôt de la relation syllabique entre la cible et le mot stimulus.

Cependant, Cutler *et al.* (1986) n'ont pas pu obtenir des résultats similaires avec des sujets anglais ; les temps de réponse obtenus pour les cibles « pa » et « pal » dans les stimuli [p^hælis] (*pa/lace*) et [p^hælpitêit] (*pal/pitate*) étaient équivalents. La principale hypothèse avancée pour expliquer ces résultats se trouve dans les différences rythmiques des deux langues : le français serait rythmé par la syllabe, et l'anglais par l'accent.

Norris et Cutler (1988) ont aussi nuancé les éléments en faveur de la syllabe. Selon eux, les conditions d'expérimentation utilisées dans la plupart des travaux en temps réel favorisaient une détection plus rapide des syllabes. En effet, selon ces auteurs, la majorité des listes utilisées permettent aux sujets de détecter des syllabes cibles dès le premier phonème perçu : les syllabes cibles ne partagent aucun phonème avec les autres stimuli dans les listes. Ainsi les sujets ne doivent faire qu'une analyse partielle de la syllabe pour pouvoir l'identifier. Pour expliquer que la détection des

syllabes soit encore plus rapide que celle d'un seul phonème, les auteurs soulignent ensuite que la plupart des tests mettent en jeu des syllabes isolées, alors que les phonèmes sont présentés en contexte; les différences dans les temps de réponse seraient ainsi dues au fait que, pour les syllabes, les sujets n'ont pas à segmenter le signal acoustique. En mettant en place un protocole destiné à éviter ces biais, Norris et Cutler ont observé des temps de réponse plus courts pour les phonèmes que pour les syllabes.

Christophe *et al.* (1991) décrivent une autre expérience intéressante, à savoir la détection de phonème avec biais attentionnel. L'étude consistait à demander à des sujets de détecter des phonèmes cible, alors que les listes de stimuli étaient biaisées au niveau syllabique : sans que les sujets n'en soient conscients, les phonèmes cible apparaissaient plus fréquemment en position d'attaque¹⁵ syllabique, ou bien en position de coda¹⁶ selon les groupes. Par ailleurs, la position des phonèmes cible (rang phonémique dans la séquence) ne variait pas dans les listes. Christophe *et al.* (1991) ont observé que les sujets dont les listes favorisaient la position d'attaque détectaient plus rapidement les phonèmes cible apparaissant dans cette position, et ont trouvé de semblables résultats pour les listes favorisant la position de coda. Selon Christophe *et al.* (1991), « on serait bien en peine d'expliquer ces résultats en refusant l'existence d'une représentation syllabique qui influence le lieu de l'attention des sujets » (Christophe *et al.*, 1991, p. 73).

En sus des hypothèses phonémique et syllabique, certains chercheurs ont postulé que la reconnaissance des mots parlés n'impliquerait pas de représentation infra-lexicale de type segmental. En effet une conception segmentale des représentations infra-lexicales pose les problèmes de l'absence de frontières nettes délimitant chaque segment dans le signal de parole (et donc de la difficulté de segmentation de la chaîne parlée) ainsi que de la variabilité acoustique des segments de parole (et donc de la façon dont sont catégorisés les sons de parole).

Pour tenter de dépasser ces difficultés, Stevens (1986) ainsi que Marslen-Wilson et Warren (1994), entre autres, ont émis l'hypothèse selon laquelle l'identification des mots parlés se ferait à partir de traits asynchrones. C'est-à-dire que les traits acoustiques perçus par l'auditeur seraient directement associés au lexique mental; dans ce cas les traits ne dépendent pas d'une représentation intermédiaire segmentale phonémique avec laquelle ils sont coordonnés dans le temps. Les traits « évoluent » plus librement dans le temps. Ainsi un trait correspondant à un phonème qui ne serait pas encore réalisé selon les modèles phonémiques (par exemple le second phonème d'un mot) peut être perçu par l'auditeur de manière précoce (par exemple lors de la réalisation du premier phonème de ce même mot). Cette conception, qui répond en partie aux problèmes posés par la variabilité liée aux phénomènes de coarticulation, et à la question de la frontière entre les segments, trouve une justification dans les résultats de diverses expériences. Warren et Marslen-Wilson (1987) ont par exemple montré que le public anglophone pouvait détecter le trait /+ nasal/ d'une consonne

15. l'attaque correspond à la première consonne de la syllabe.

16. le coda correspond à la dernière consonne de la syllabe.

dès la fin de la voyelle précédant cette consonne.

Nous n'irons pas plus loin dans la description des données existant en faveur des différents types de représentations infra-lexicales. Notre objectif était simplement de présenter les principales théories et de souligner le fait qu'il n'existe pas, à l'heure actuelle, de consensus autour des unités phonétiques servant de base à la perception de la parole. Certains modèles vont même jusqu'à postuler que dans le lexique mental de l'auditeur les mots ne seraient pas stockés sous la forme de représentations phonologiques abstraites, mais seraient plutôt représentés par des listes d'exemplaires contenant des informations acoustiques très fines (Nguyen *et al.*, 2009). En d'autres termes, l'auditeur disposerait de listes d'exemplaires acoustiques correspondant à chaque mot, listes auxquelles il confronterait l'input acoustique reçu pour reconnaître un mot.

En outre, le sujet des unités perceptives de base est loin d'être le seul point de désaccord théorique. Si la plupart des modèles postulent l'existence d'unités infra-lexicales et de plusieurs niveaux de traitement (Nguyen, 2005), l'interactivité entre les divers niveaux et la nature ascendante (*bottom-up*, de la forme vers le sens) et/ou descendante (*top-down*, du sens vers la forme) de la perception et de la compréhension de la parole sont encore sujettes à de nombreuses études et discussions.

De la nature ascendante ou descendante des processus de perception et de compréhension de la parole

Les modèles psycholinguistiques de traitement de la parole postulent généralement l'existence de traitements de « bas niveau » orientés vers la perception – c'est-à-dire vers le décodage des informations acoustico-phonétiques, et des traitements de « plus haut niveau » impliquant à la fois les informations décodées et les connaissances de l'auditeur, et conduisant à l'interprétation du message. Par exemple Marslen-Wilson (1989) décrit le processus de compréhension de la parole comme étant composé de deux mécanismes :

- le premier s'appuie sur des fonctions de *forme*, et correspond au décodage acoustico-phonétique du signal de parole, par lequel les auditeurs font correspondre l'input au lexique mental ;
- le second implique des fonctions de *contexte* et permet de lier les informations reçues par les fonctions de forme à des représentations de plus haut niveau, pour interpréter le sens du message.

Cette représentation des processus cognitifs engagés dans le traitement de la parole pose, entre autres, la question de l'interactivité entre les différents niveaux. Cette question a donné lieu à de nombreux débats et fait encore l'objet d'investigations scientifiques (Grataloup *et al.*, 2009; Nguyen, 2005). En effet, traditionnellement la relation envisagée entre les niveaux de forme et de sens était de nature séquentielle et ascendante (*bottom-up*, de la forme vers le sens). Cependant au fil des ans de nombreuses études se sont accumulées qui tendent à montrer l'existence de processus « descendants » (*top-down*, du sens vers la forme) influençant les traitements dits « de plus bas niveau », et traduisant des attentes chez l'auditeur.

Dans la perception de la parole

Un effet maintenant bien connu est celui de la *restauration phonémique*. Warren (1970) a conduit une série d'expériences montrant que, lorsque des auditeurs entendent des phrases dans lesquelles certains phonèmes sont remplacés par du bruit, les auditeurs ne sont pas conscients de la transformation qui a été effectuée. En d'autres termes les auditeurs continuent d'« entendre » des phonèmes pourtant absents de la chaîne parlée. Ces résultats, qui seront confirmés par d'autres études comme celles de Samuel (1997), ont été interprétés comme étant la preuve de l'existence d'effets descendants (*top-down*) du lexique vers la perception du signal de parole. Même si le lieu et la nature de ces effets fait encore débat (Nguyen, 2005), la somme des données empiriques en faveur de l'influence d'attentes lexicales sur la perception de la parole s'accroît. Récemment Sivonen *et al.* (2006) ont étudié le phénomène de la restauration phonémique par la méthode des potentiels évoqués¹⁷. Leurs résultats montrent que les mécanismes de bas niveau, automatisés, détectent correctement la présence de bruits en lieu et place de phonèmes, mais que cela n'empêche pas le traitement sémantique du mot de se dérouler normalement – avec cependant un petit délai temporel comparé à un mot ayant conservé l'intégralité de ses phonèmes. Ces données permettent aux auteurs de conclure que la « perception » d'un phonème absent n'est pas un phénomène intervenant au début de la chaîne de traitement partant de la forme vers le sens (*bottom-up*), mais plutôt le reflet des attentes de l'auditeur (effet *top-down*).

De même, les contraintes phonotactiques pesant sur la composition phonémique des unités lexicales pourraient aider à la perception de la parole (Magnen, 2009; Treiman *et al.*, 2003). Hallé *et al.* (1998) ont montré que les auditeurs anglophones utilisent le fait qu'en anglais la suite /tɪ/ n'est pas permise en position initiale pour mieux identifier les phonèmes et les frontières de mots. De même, Massaro et Cohen (1983) ont montré dans une tâche de catégorisation de phonèmes que les auditeurs peuvent avoir tendance à privilégier les catégories de phonèmes qui donnent lieu à une séquence phonétique acceptable dans leur langue. Ainsi lors de la présentation de syllabes de type Consonne-Consonne-Voyelle dont la consonne centrale se situe sur le continuum [ɾ-l], si la consonne initiale est un /s/, les sujets anglophones ont tendance à privilégier la catégorie /l/ (ex. /sla/) qui forme une séquence phonétique acceptable, au contraire de la séquence */sɪa/ qui n'est pas permise en anglais.

Enfin, la connaissance des patterns rythmiques de la langue peut éventuellement servir de support à la perception de la parole. Norris *et al.* (1995) ont par exemple montré que les auditeurs anglophones utilisent le fait que les mots anglais sont généralement accentués sur leur première syllabe pour segmenter les énoncés en mots.

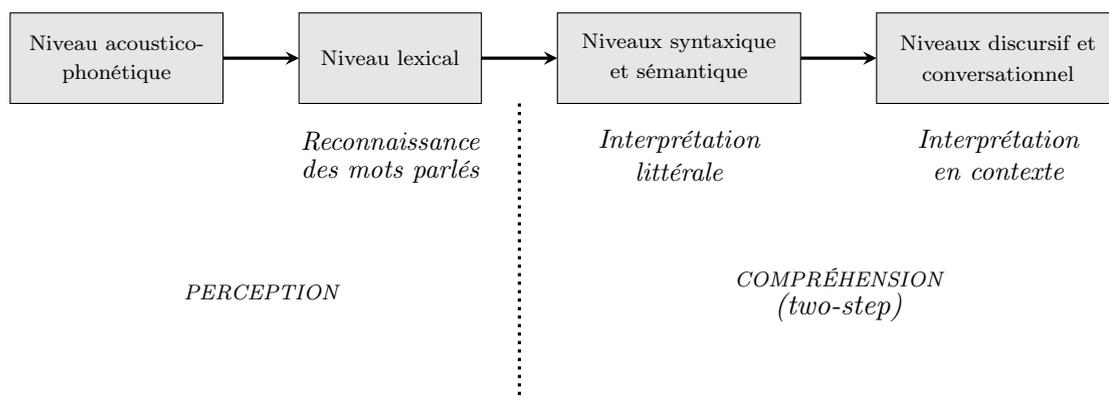
17. les potentiels évoqués (*PE*, ou *ERP* pour l'anglais *Event-Related Potentials*) renvoient à des modifications du potentiel électrique du système nerveux au niveau de l'encéphale, en réaction à des événements externes (stimuli sensoriels) ou internes (activités cognitives). Les potentiels sont généralement enregistrés à l'aide d'électroencéphalogrammes.

Dans la compréhension de la parole

Tout comme pour les premiers modèles de perception de la parole, la question de l'interprétation des messages parlés a d'abord été envisagée à travers le prisme des règles postulées par la linguistique structurale. Les représentations et les mécanismes impliqués s'inscrivent généralement dans plusieurs modules de traitement distincts, correspondant aux différents niveaux de l'analyse linguistique (par ex. les niveaux lexical, morpho-syntaxique, sémantique et discursif). Ces modules se succèdent alors de manière sérielle dans la chaîne de traitements cognitifs, en partant du niveau le plus fin (niveau lexical) vers le niveau le plus global (niveau discursif).

Les niveaux de la syntaxe et de la sémantique ont occupé un rôle particulièrement important. La plupart des modèles obéissent en effet à ce que Culicover et Jackendoff (2006) ont appelé les principes de *compositionnalité frégienne*, c'est-à-dire qu'ils postulent que le sens global d'une phrase résulte d'une part de la signification de ses parties (c.-à-d. des unités lexicales) et d'autre part des agencements syntaxiques qui régissent leur combinaison. Cette conception a donné lieu à des modèles d'interprétation du langage qui impliquent deux grandes étapes (modèles *two-step*, cf. Hagoort et van Berkum, 2007). Durant la première étape le sens littéral (hors-contexte) des phrases est « calculé » en combinant la signification des mots (sémantique lexicale) en accord avec leur agencement syntaxique. Ce n'est qu'ensuite que cette signification littérale est rapprochée des autres sources d'information telles que les connaissances du monde, la situation de communication, ou encore les énoncés précédents – c'est à dire du contexte dans son sens le plus large. Nous illustrons cette perspective sur les mécanismes de compréhension de la parole dans la figure 3.2.

FIGURE 3.2 – Mécanismes et niveaux impliqués dans la perception et la compréhension de la parole – Perspective ascendante (*bottom-up*) et fragmentée (*two-step*)



Peu à peu de nombreuses données sont venues remettre en cause cette modélisation des mécanismes de compréhension de la parole. Tout d'abord, l'existence de modules dédiés au traitement des différents niveaux linguistiques (ex. lexical, syntaxique et sémantique) semble de moins en moins envisageable. Par exemple, il

a été longtemps supposé que les traitements syntaxiques et sémantiques reposaient sur des mécanismes bien distincts, et les questions de la localisation et du fonctionnement de ces deux modules ont donné lieu à de nombreux travaux en neurosciences cognitives (Bornkessel-Schlesewsky et Schlesewsky, 2008).

Or malgré les études en neuroanatomie fonctionnelle, la délimitation des modules syntaxique et sémantique dans le cerveau humain demeure encore vague aujourd'hui (pour une revue de littérature sur cette question voir Bornkessel-Schlesewsky et Friederici, 2007). Les arguments les plus forts en faveur de l'existence de modules distincts ont été apportés par la technique de l'électroencéphalographie (EEG). En 1980 Kutas et Hillyard ont mis au jour la composante N400, une onde négative centro-pariétale apparaissant généralement entre 300 ms et 500 ms après le début d'un mot sémantiquement plein, lorsque ce dernier présente une incongruité sémantique avec l'énoncé dans lequel il apparaît (par ex. « socks » dans « He spread the warm bread with socks »¹⁸ ; Kutas et Hillyard, 1980). Cette découverte a été suivie de travaux qui ont contribué à associer la composante N400 avec la charge cognitive liée aux traitements sémantiques du langage (Bornkessel-Schlesewsky et Schlesewsky, 2008). L'équivalent de la composante N400 pour ce qui est des traitements syntaxiques a été mise au jour par Osterhout et Holcomb (1992) : il s'agit de la composante P600, une onde positive tardive apparaissant approximativement 600 ms après le début du mot critique. Elle est considérée comme le reflet des efforts cognitifs liés aux traitements syntaxiques car son intensité peut être accrue par des incohérences de type syntaxique (pour une brève revue de ces effets voir Bornkessel-Schlesewsky et Schlesewsky, 2008).

La découverte des composantes N400 et P600, ainsi que les travaux confirmant l'influence des incongruités sémantiques et syntaxiques sur leurs intensités respectives, ont d'abord été interprétés comme les signes d'une distinction bien claire entre les traitements sémantiques et syntaxiques dans le cerveau humain. Cependant, un nombre de plus en plus important de travaux font aujourd'hui douter de la réalité d'une compartimentation aussi nette entre traitements syntaxiques et sémantiques (Bornkessel-Schlesewsky et Schlesewsky, 2008; Kuperberg, 2007). Ainsi des études ont révélé l'existence d'effets N400 non attribuables à des violations d'ordre sémantique, de même que des effets P600 apparaissant sans qu'une contrainte de niveau syntaxique ne soit violée (voir Bornkessel-Schlesewsky et Schlesewsky, 2008; Kutas *et al.*, 2006). Les recherches empiriques semblent au contraire montrer que les informations syntaxiques et sémantiques sont entremêlées de façon très étroite dans le cerveau (Bornkessel-Schlesewsky et Schlesewsky, 2009a,b).

Par ailleurs, la perspective purement ascendante (*bottom-up*) de la compréhension de la parole présente dans les modèles *two-step* devient de plus en plus difficile à soutenir en regard des données empiriques qui se sont accumulées ces dernières années (Hagoort et van Berkum, 2007). De nombreux phénomènes mis au jour sont en effet très difficiles à expliquer sans postuler l'existence de flux descendants dans le traitement de la parole. Ces phénomènes impliquent pour l'auditeur des attentes

18. ≈ « Sur le pain, il a étalé de la chaussette ».

(notamment lexicales, mais pas seulement) générées par ses connaissances antérieures (liées à sa connaissance du monde en général, au discours, ou à la situation de communication).

Tout d'abord, en analogie avec l'effet de « restauration phonémique » existant pour la perception de la parole, il existe dans la compréhension un phénomène que l'on a désigné sous le terme d'*illusion sémantique*. Plusieurs études ont rapporté que dans certains cas les auditeurs procèdent à une analyse sémantique incomplète des énoncés entendus. Par exemple, la plupart des sujets ne relèvent pas d'anomalie lorsqu'ils doivent répondre à la question « Combien d'animaux de chaque espèce Moïse a-t-il embarqué sur l'arche? »¹⁹ alors qu'ils savent très bien que c'est Noë, et non Moïse, qui est censé avoir construit l'arche. De même, les sujets ne relèvent aucune incohérence particulière lorsque, au milieu d'une histoire d'accident d'avion, ils traitent l'énoncé « Les autorités ont dû décider où enterrer les survivants »²⁰. Les études ayant traité de ces phénomènes d'illusion sémantique suggèrent que plus les mots « intrus » (ex. *Moïse, survivants*) partagent de traits sémantiques avec le mot correct (*Noë, victimes*) plus l'incohérence sera difficile à relever pour les sujets (Nieuwland et Van Berkum, 2005). Certains auteurs ont ainsi suggéré l'existence d'une « heuristique de plausibilité », qui biaise notre compréhension en favorisant les interprétations sémantiques qui sont les plus consistantes avec les connaissances que l'on a du monde en général (Ferreira *et al.*, 2002; Sanford et Sturt, 2002).

Le phénomène d'illusion sémantique n'est pas le seul à appuyer l'existence de mécanismes top-down dans la compréhension de la parole. De récentes études impliquant l'utilisation de techniques d'électro-encéphalographie (EEG) ont montré que les connaissances encyclopédiques, les contextes discursifs et conversationnels peuvent être pris en compte de manière extrêmement précoce lors de l'interprétation de messages oraux.

Hagoort *et al.* (2004) ont testé les effets provoqués par une incongruité liée aux connaissances encyclopédiques des auditeurs. À l'écoute de la phrase « The Dutch trains are yellow/white/sour and very crowded »²¹, le même accroissement de l'onde N400 a été observé pour une incongruité située uniquement au niveau phrastique (« The Dutch trains are sour and very crowded »²²) et pour une incohérence avec les connaissances des auditeurs (« The Dutch trains are white and very crowded »²³ – Les auditeurs, tous Néerlandais, savaient que les trains néerlandais sont jaunes, et non blancs).

Dans l'étude de Van Berkum *et al.* (2003), les auditeurs entendaient des histoires dans lesquelles la dernière phrase pouvait contenir un mot présentant une incongruité. Ces incongruités n'étaient pas liées au contexte phrastique, mais au contexte discursif, c'est-à-dire aux énoncés précédant la phrase. Une telle phrase pouvait être

19. « How many animals of each kind did Moses take on the Ark? » (Erickson et Mattson, 1981).

20. « The authorities had to decide where to bury the survivors. » (Barton et Sanford, 1993).

21. « Les trains néerlandais sont jaunes/blancs/acides et pleins de monde ».

22. « Les trains néerlandais sont acides et pleins de monde ».

23. « Les trains néerlandais sont blancs et pleins de monde ».

« Jane told the brother that he was exceptionally slow »²⁴ dans un contexte discursif où le frère en question était censé être rapide. Ce type d'incongruité, en comparaison avec un mot en accord avec le contexte (ex. *quick* – c.-à-d. *rapide*), provoquait un effet N400 tout à fait comparable avec le type d'onde suscitée par une anomalie sémantique classique, notamment en ce qui concerne la dimension temporelle.

Nieuwland et Van Berkum (2006) sont allés plus loin et ont montré qu'un mot ne présentant pas d'incongruité au niveau phrastique pouvait déclencher un effet N400 lorsque celui-ci ne s'insérait pas correctement dans le contexte discursif. Ainsi le participe passé *salted* dans la phrase « The peanut was salted »²⁵ ne présente pas d'incongruité sémantique et ne devrait donc pas déclencher d'effet particulier. Les auteurs ont utilisé cette phrase dans le contexte discursif suivant :

A woman saw a dancing peanut who had a big smile on his face. The peanut was singing about a girl he had just met. And judging from the song, the peanut was totally crazy about her. The woman thought it was really cute to see the peanut singing and dancing like that. The peanut was salted, and by the sound of it, this was definitively mutual. He was seeing a little almond.²⁶

Les auteurs ont observé un effet N400 beaucoup plus prégnant lorsque le prédicat inanimé canonique *salted* (*salée*) était utilisé que lorsque le prédicat animé *in love* (*amoureuse*), incongru au niveau phrastique mais approprié au contexte, était utilisé.

Outre ces effets liés aux connaissances des auditeurs et au contexte discursif, Van Berkum *et al.* (2008) ont également montré l'existence d'attentes générées par la prise en compte de la situation de communication, notamment en fonction des caractéristiques sociologiques de l'interlocuteur. Les auteurs ont trouvé des effets N400 tout à fait similaires à ceux rencontrés pour des incongruités sémantiques « classiques », lorsque des mots présentaient une incohérence avec des informations portées par la voix du locuteur, comme :

- le sexe du locuteur ; par exemple lorsque la phrase « if only I looked like Britney Spears in her latest video »²⁷ était prononcée par un homme ;
- l'âge du locuteur ; par exemple lorsque la phrase « Every evening I drink some wine before I go to sleep »²⁸ était prononcée par un enfant ;
- le niveau social du locuteur ; par exemple lorsque la phrase « I have a large tattoo on my back »²⁹ était prononcée par une personne à l'accent typique de

24. « Jane a dit au frère qu'il était exceptionnellement lent. ».

25. « La cacahuète était salée. ».

26. « Une femme vit une cacahuète qui dansait avec un grand sourire. La cacahuète chantait une chanson à propos d'une fille qu'elle venait de rencontrer. À en juger par la chanson, la cacahuète était folle d'elle. La femme pensa que c'était vraiment mignon de voir une cacahuète chanter et danser comme ça. La cacahuète était salée, et d'après ce que l'on pouvait entendre, c'était clairement réciproque. Elle fréquentait une jeune amande. ».

27. « Si seulement je pouvais ressembler à Britney Spears dans son dernier clip ».

28. « Tous les soirs je bois du vin avant d'aller me coucher ».

29. « J'ai un gros tatouage dans le dos ».

la classe haute néerlandaise.

Ces données ont contribué à conforter un point de vue radicalement opposé à la conception *two-step* de la compréhension de la parole : la conception de l'interprétation *en bloc* (*one-step model of language interpretation*, cf. Hagoort et van Berkum, 2007; Kuperberg, 2007; Lotze *et al.*, 2011). Ce modèle postule que toutes les sources d'informations, à quelque niveau que ce soit (par ex. syntaxique, sémantique, prosodique, discursif, encyclopédique) sont prises en compte immédiatement et simultanément lors du « calcul » du sens par le cerveau humain.

Cette perspective, présentée dans sa conception peut-être la plus radicale par Van Berkum *et al.* (2008) lorsqu'ils affirment que « pour le cerveau de l'utilisateur, il n'y a pas de sens en dehors du contexte »³⁰, rejoint la position dominante au sein de la tradition du *langage en tant qu'action*. Les résultats obtenus par les techniques de neuro-imagerie et d'électroencéphalographie convergent en effet avec les observations effectuées dans les études linguistiques, et selon lesquelles le sens s'inscrit toujours en lien avec la situation de communication et avec les connaissances du monde en général (Hagoort et van Berkum, 2007). Or, comme nous allons le voir, au sein la tradition du *langage en tant qu'action* le contexte – qu'il soit lié au discours ou à la situation de communication, est vu comme *inhérent* à la notion même de compréhension.

3.2.2 La question de la compréhension de la parole dans la conception du langage *en tant qu'action*

La tradition du *langage en tant qu'action* puise ses racines dans le courant théorique de la pragmatique. Elle s'appuie également sur les méthodologies qui ont découlé de ce courant dont, en particulier, l'analyse conversationnelle (Tanenhaus et Brown-Schmidt, 2008). Après avoir exposé brièvement les principaux concepts apportés par la pragmatique, nous traitons des outils d'analyse et de leur application à l'étude des mécanismes de compréhension de la parole.

Principaux concepts pragmatiques

La plupart des concepts développés au sein du courant pragmatique sont directement issus de la philosophie du langage (Burton-Roberts, 2007; Levinson, 1983). En effet, la pragmatique est d'abord apparue comme une alternative à la conception chomskyenne d'un langage comme un outil « abstrait », comme une aptitude mentale *dissociable des utilisateurs, des usages et des fonctions du langage* (Cummings, 2005; Levinson, 1983). En opposition avec la tradition chomskyenne, la pragmatique s'est intéressée aux faits de langue qui ne peuvent être décrits et expliqués sans tenir compte du contexte d'usage et des intentions des usagers. Ainsi sont apparus les théories et concepts centraux de la pragmatique tels que la *deixis*, les *actes de parole* ou, plus récemment, la *théorie de la pertinence* (cf. Wilson et Sperber, 2005).

30. « To the brain of the language user, there is no context-free meaning » (Van Berkum *et al.*, 2008, p. 589).

La deixis

La *deixis* (terme emprunté au grec signifiant *montrer* ou *pointer du doigt*) concerne la façon dont les éléments du contexte discursif, spatio-temporel ou conversationnel se réalisent dans les énoncés (Cummings, 2005; Levinson, 1983). Les pronoms personnels de 1^{ère} et 2^{ème} personne ou encore les pronoms démonstratifs sont des exemples prototypiques de ces réalisations. Ainsi, si l'on considère la variété des usages, le pronom « ceci » ne réfère à aucune entité particulière ; il se comporte plutôt comme une « variable » qui est instanciée par le contexte de l'énoncé (par ex. par un geste de pointage co-verbal). On distingue généralement la deixis interpersonnelle et sociale, la deixis temporelle, la deixis spatiale et enfin la deixis discursive (Cummings, 2005).

L'implication conversationnelle

L'implication conversationnelle (*conversational implicature*), parfois appelée plus simplement *implication*, a d'abord été élaborée et exposée par Grice, en particulier lors de ses enseignements à l'Université de Harvard (cf. Grice, 1975, 1978). L'idée générale développée ici est qu'il est possible de « dire plus qu'il n'en est dit ». Par exemple, si l'on considère le dialogue suivant³¹ :

A : Tu peux me dire l'heure ?
B : Le facteur vient de passer.

Il est clair que l'énoncé de B porte davantage de signification que le seul sens que l'on pourrait dégager si l'on ne tenait pas compte du contexte précédent sa réalisation. On pourrait gloser les éléments implicites de l'énoncé de B de la manière suivante (en italique ci-dessous) :

A : Tu peux me dire l'heure ?
B : *Non je ne sais pas l'heure exacte du moment présent, mais je peux fournir une information à partir de laquelle tu peux déduire l'heure approximative, à savoir : Le facteur vient de passer.*

Ce phénomène a une implication directe dans la conception de la relation qu'entretiennent les interlocuteurs A et B. En effet, pour que A comprenne la signification de l'énoncé de B, il est nécessaire que A pense que B *coopère* activement et sincèrement à la conversation, autrement dit qu'il ne dise pas « n'importe quoi », par exemple quelque chose qui n'ait pas de rapport avec le contexte précédent, ou bien quelque chose de faux. Cette notion de *coopération* est centrale dans la théorie de l'implication conversationnelle. Selon Levinson (1983) elle suppose entre autres que les interlocuteurs accordent la réalisation de leurs énoncés avec certains principes : les maximes de *qualité* (sincérité et vérité), de quantité (ni trop ni trop peu d'information), de *pertinence* et de *manière* (éviter l'obscurité et l'ambiguïté, privilégier la brièveté et l'ordonnancement). Plus tard, ces quatre maximes seront remplacées par le seul principe de pertinence (cf. *infra* la *théorie de la pertinence*).

31. adapté de Levinson (1983).

La présupposition

Les présuppositions renvoient aux conditions préalables qui sont impliquées par certains énoncés, souvent en fonction du contexte. Ainsi dans l'exemple suivant³² :

(i) Le docteur a réussi à sauver le bébé.

il est présupposé que le docteur a *essayé* de sauver le bébé. Pourtant, cette implication n'est pas du domaine de la sémantique lexicale : le sens de *essayer* n'est pas inclus dans le verbe *réussir*, comme le montre l'énoncé (ii).

(ii) Les étudiants ont encore réussi à se planter.

Les actes de parole

De la plupart des questions traitées en pragmatique, la théorie des actes de parole est certainement celle qui a suscité le plus d'intérêt (Levinson, 1983). Elle a trouvé des applications théoriques et pratiques dans des domaines aussi variés que la psychologie, la littérature, l'anthropologie, ou encore dans d'autres champs des sciences du langage comme la syntaxe, la sémantique et l'enseignement/apprentissage des langues étrangères et secondes.

Dans la lignée de philosophes comme Wittgenstein, Austin (1962, publication posthume) s'est attaché à démonter les idées du positivisme logique ambiant selon lequel un énoncé doit obligatoirement être vérifiable (c.-à-d. avoir une valeur vraie ou fausse), faute de quoi il n'a pas de sens. Dans ce contexte il lance la théorie des actes de paroles, en soulignant principalement le fait qu'il existe des phrases qui n'ont pas pour objet de *dire*, mais plutôt de *faire*, comme dans les exemples suivants³³ :

(i) Je te parie deux euros qu'il va pleuvoir demain.

(ii) Je m'excuse.

(iii) Je te promets que je rentrerai tôt.

Dans de telles phrases, il n'est pas possible de juger de la véracité de ce qui est rapporté. En effet leur objet n'est pas tourné vers la déclaration mais vers l'action (faire un pari, s'excuser ou promettre), et l'énonciation même de ces phrases crée et constitue l'action visée.

Selon Austin (1962) une personne quand elle produit un énoncé effectue à la fois un acte *locutoire*, un acte *illocutoire* et un acte *perlocutoire*.

L'acte *locutoire* consiste globalement à produire une phrase avec une certaine signification et certaines références au monde, comme cela est entendu en sémantique traditionnelle. Ainsi lorsqu'une personne énonce :

(iv) Le molosse est dans le jardin.

32. les exemples sont adaptés de Cummings (2005)

33. adaptés de Austin (1962).

il produit une phrase dont la signification implique la référence à un chien et un jardin particuliers du monde réel.

Selon Austin (1962), tout en produisant des actes locutoires les locuteurs produisent également des actes *illocutoires* tournés vers leurs interlocuteurs comme *informer*, *ordonner*, *avertir*, etc. Par exemple pour l'énoncé (iv), le locuteur peut effectuer un acte illocutoire consistant à *avertir* son interlocuteur pour qu'il n'entre pas dans le jardin.

Les actes illocutoires produisent des effets dans le monde réel. C'est ce changement substantiel de l'ordre du monde réel qu'Austin (1962) nomme l'acte *perlocutoire*. Ainsi, si lorsqu'il énonce la phrase (iv) le locuteur parvient à empêcher son interlocuteur d'entrer dans le jardin, alors il a conduit un acte perlocutoire.

La structure conversationnelle

En pragmatique, la conversation constitue l'usage linguistique central, ou basique ; elle est l'unité d'étude privilégiée, explorée par la méthode de l'analyse conversationnelle (désormais *AC*). L'*AC* consiste à observer et rechercher les récurrences des événements apparaissant dans des conversations naturelles. L'*AC* est donc de nature essentiellement *inductive*, elle n'implique pas la construction de modèles ou de théories en pré-requis (Levinson, 1983).

Grâce à l'*AC*, la pragmatique s'est attachée à mettre au jour et à expliquer divers mécanismes, dont entre autres :

- les *tours de parole*. Il est très rare que deux personnes parlent en même temps (de l'ordre de 5% du temps de parole, cf. Levinson, 1983), alors que les « blancs » entre les prises de parole ne durent que quelques dixièmes de seconde. Et ces performances impressionnantes ne sont pas altérées par le nombre de participants ou bien par la modalité de la conversation (au téléphone par ex.). Expliquer la manière dont sont gérés les tours de parole requiert de nombreuses observations et analyses, et occupe un champ de recherche entier ;
- les *paires adjacentes* (*adjacency pairs*). Ce sont un autre type d'organisation locale de la conversation : il s'agit des énoncés qui fonctionnent par paires, comme les couples *question/réponse*, *salutation/salutation*, *offre/acceptation*, etc. Ce type d'organisation locale a fait l'objet de divers travaux, notamment de typologie ;
- l'*organisation globale*. Il existe des phénomènes conversationnels qui apparaissent uniquement lorsque l'on observe des groupes supérieurs à trois ou quatre tours de parole. C'est le cas par exemple des rituels d'ouverture ou de fermeture de conversation.

La théorie de la pertinence

Selon Wilson et Sperber (1991) toutes les maximes d'implication conversationnelle définies par Grice (cf. *supra*) peuvent être remplacées par un seul grand principe : celui de la *pertinence*. Selon ce dernier le locuteur essaye d'être aussi pertinent que possible en fonction des circonstances.

Le principe de pertinence concerne aussi bien le locuteur que son interlocuteur. Ceci transparait clairement lorsque Wilson et Sperber (1991) postulent :

Nous supposons également qu'un locuteur qui juge utile de parler va essayer de produire un énoncé aussi pertinent que possible. En conséquence, pour traiter chaque énoncé, un auditeur devrait supposer que le locuteur a essayé d'être aussi pertinent que possible compte-tenu des circonstances. C'est cette supposition que nous appelons le principe de pertinence.³⁴

La théorie de la pertinence est certainement la plus influente dans le domaine de la pragmatique aujourd'hui (Cummings, 2005).

Outils d'analyse de la pragmatique et compréhension de la parole

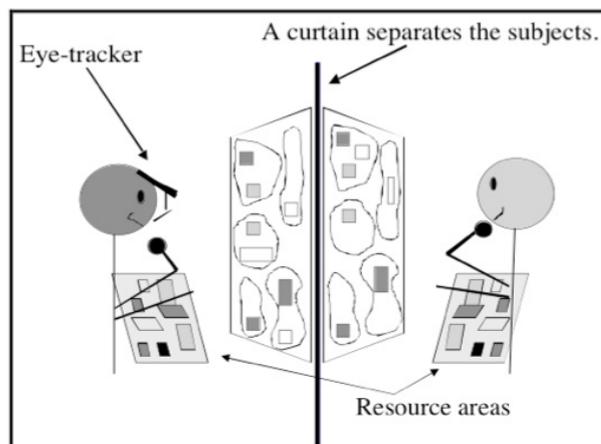
Dans la tradition du langage en tant qu'action, les études sur la compréhension de la parole impliquent en général des protocoles mettant en jeu des tâches « naturelles », c'est-à-dire sans scripts écrits à l'avance, et dans lesquelles les interlocuteurs sont amenés à collaborer pour atteindre un objectif particulier. Ce sont donc les mêmes types de tâches et d'analyses inductives qui sont utilisées pour décrire la structure conversationnelle (*cf.* section précédente) et pour mettre au jour les processus de compréhension de la parole. La collaboration entre les interlocuteurs suppose que ceux-ci disposent d'un « terrain commun » (*common ground*), c'est-à-dire d'un environnement, d'activités, de perceptions, d'émotions, de projets et d'intérêts en partie partagés (Clark, 1996).

Par exemple, pour étudier les stratégies que l'être humain emploie pour résoudre des références langagières ambiguës, Brown-Schmidt *et al.* (2005) ont utilisé un environnement dans lequel les interlocuteurs étaient séparés par un rideau (voir figure 3.3). Ils avaient chacun devant eux un panneau sur lequel des cartes de différentes tailles, formes et couleurs étaient disposées ; les dispositions des cartes sur les deux panneaux étaient en partie similaires. À l'aide d'un ensemble de cartes à leur disposition, les interlocuteurs étaient amenés à communiquer pour atteindre un objectif particulier, comme la reproduction d'un panneau. La compréhension des interlocuteurs était donc jugée à travers les effets que la communication produisait sur le contexte (c.-à-d. sur la reproduction du panneau).

Ce protocole expérimental rappelle ceux utilisés en Analyse Conversationnelle, comme par exemple la tâche de reproduction cartographique utilisée pour la constitution du corpus de parole *HCRC Map Task* (Anderson *et al.*, 1991). Pour constituer le corpus HCRC, Anderson *et al.* (1991) ont imaginé une tâche impliquant 2 interlocuteurs. L'un d'eux tient le rôle d'instructeur, et l'autre doit suivre les instructions.

34. « We also assume that a speaker who thinks it worth speaking at all will try to make his utterance as relevant as possible. A hearer should therefore bring to the processing of every utterance the standing assumption that the speaker has tried to be as relevant as possible in the circumstances. It is this assumption that we call the principle of relevance. » (Wilson et Sperber, 1991).

FIGURE 3.3 – Exemple de situation de test pragmatique pour l'étude de la compréhension de la parole (Brown-Schmidt *et al.*, 2005)

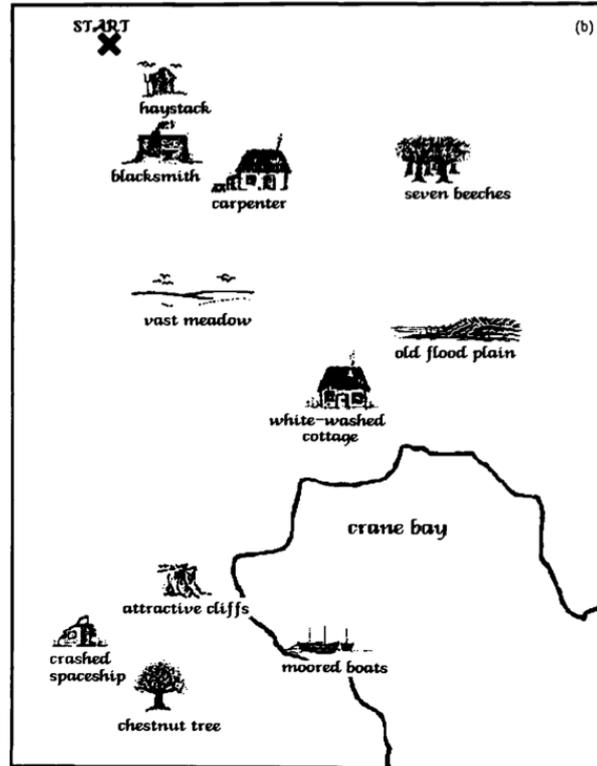


Tous les deux possèdent une carte comme celle présentée en figure 3.4). Les cartes que chacun des interlocuteurs possède contiennent des éléments communs et des éléments différents (par ex. certains noms de lieux sont différents, certains éléments sont présents dans la carte de l'instructeur seulement, etc.). Le but donné aux interlocuteurs est d'amener le « suiveur » à reproduire la carte de l'instructeur.

Selon Tanenhaus et Brown-Schmidt (2008), ce genre de protocole expérimental est utilisé pour deux raisons principales. Tout d'abord, le langage utilisé dans des conversations interactives diffère considérablement du langage très contrôlé qui est utilisé dans le paradigme du traitement de l'information. Les énoncés produits dans des conversations naturelles peuvent donc constituer des données intéressantes à confronter aux modèles développés au sein du courant du traitement de l'information. En outre de nombreux aspects des énoncés ne peuvent être compris sans tenir compte du contexte (temps, lieu, situation, objectifs des interlocuteurs) dans lequel ils ont été produits.

La seconde raison est d'ordre méthodologique. De plus en plus d'études montrent l'influence que peut avoir le type de tâche sur les processus cognitifs ou perceptifs, et ce même en ce qui concerne les traitements dits « de bas niveau » (*cf.* Tanenhaus et Brown-Schmidt, 2008, pour des références dans le domaine de la perception visuelle par exemple). Étudier certains processus de manière isolée, sans tenir compte du contexte comportemental dans lequel les êtres humains sont engagés, pourrait ainsi constituer un biais. De plus, les domaines de la pragmatique et du traitement de l'information peuvent parfois s'avérer complémentaires, en offrant des explications différentes pour un même phénomène.

FIGURE 3.4 – Exemple de matériel utilisé pour la constitution d'un corpus de conversation (Anderson *et al.*, 1991)



3.3 Quelle perspective sur la compréhension de la parole dans notre travail ?

Il est important que nous nous positionnons quant à la perspective que nous avons sur la compréhension de la parole. De ce positionnement dépendent entre autres nos hypothèses (*cf.* la fin de la section 3.4) et les protocoles que nous envisageons d'utiliser pour évaluer la compréhension de la parole par des auditeurs (*cf.* chapitre 4, p. 85).

Le premier élément qui définit notre conception de la compréhension de la parole est l'importance que nous accordons au contexte de communication. Nous adhérons à la vision de Van Berkum *et al.* (2008) lorsqu'ils énoncent que « dans le cerveau de l'utilisateur, il n'existe pas de sens en dehors du contexte »³⁵. Nous pensons en effet que les processus de compréhension de la parole impliquent des attentes chez l'auditeur, et que ces attentes dépendent en particulier du contexte de la communication : par exemple de l'identité des interlocuteurs, du contexte spatio-temporel, du contexte discursif et conversationnel, des objectifs des communicants ou des connaissances qu'ils partagent.

Selon nous ces processus dits *descendants* sont donc intrinsèques aux processus de compréhension de la parole. De plus, nous n'envisageons pas que ces attentes soient de nature secondaire par rapport aux processus ascendants – ce que semblent confirmer les études rapportées par Van Berkum *et al.* (2008). D'un point de vue théorique, nous penchons donc plutôt pour un modèle d'interprétation *en bloc*, dans lequel les éléments relatifs à l'input de parole et les connaissances de l'auditeur sont traitées de manière conjointe (*one-step model of language interpretation*, *cf.* Hagoort et van Berkum, 2007; Kuperberg, 2007; Lotze *et al.*, 2011).

Nous rejoignons également l'idée selon laquelle les attentes de l'auditeur influencent même les processus les plus « primitifs » dans la perception de la parole (Van Berkum *et al.*, 2008). Nous pensons notamment – et c'est une hypothèse – que la question des unités propres à la perception dépend en grande partie des attentes de l'auditeur. On peut en effet imaginer que l'auditeur ne « retire » pas le même matériel phonético-phonologique du signal de parole selon les attentes qu'il a vis-à-vis des propos de son interlocuteur³⁶ et selon ses objectifs³⁷, entre autres.

35. « To the brain of the language user, there is no context-free meaning » (Van Berkum *et al.*, 2008).

36. par exemple, si une personne cherche seulement à s'assurer que son interlocuteur est d'accord avec ses propos, peut-être qu'elle n'écouterait que les informations fournies par les aspects prosodiques des réponses de son interlocuteur (intonation notamment). À l'opposé, lors d'une communication entre un contrôleur aérien et un pilote impliquant l'utilisation d'un vocabulaire codifié, peut-être que les interlocuteurs dirigent davantage leur attention vers les aspects segmentaux de la parole.

37. par exemple si l'auditeur écoute afin de *retranscrire* les paroles entendues, son attention sera particulièrement dirigée vers les phonèmes (dans le cas d'une retranscription phonétique ou alphabétique bien entendu), alors que s'il écoute afin d'*imiter* une personne, alors il pourra intégrer d'autres informations fournies par le signal acoustique – comme un rythme particulier, une tendance à produire des intonations singulières.

Nous pensons donc que la *tâche* dans laquelle est engagé l'auditeur influence fortement les processus de perception et de compréhension de la parole. Si l'on désire obtenir une évaluation *valide* (c.-à-d. *représentative*) de la performance de communication orale de patients auprès d'auditeurs, il est donc nécessaire, selon nous, de placer les auditeurs dans des situations qui s'approchent le plus possible de situations de communication réelles.

C'est sur ce dernier point que les notions et méthodes apportées par le courant du *langage en tant qu'action* peuvent nous être utiles. Nous pensons en effet que, pour une question de validité, il est nécessaire de développer des tests « pragmatiques » pour évaluer la compréhension de la parole. Par *pragmatique* nous signifions que ces tests doivent tendre vers l'implication de tous les aspects non linguistiques inhérents à la plupart des situations de communication, et que nous avons soulignés dans la section précédente : notamment la présence d'un contexte et d'objectifs de communication (en partie) partagés par les interlocuteurs.

Ce point de vue pragmatique sur la compréhension de la parole guidera nos hypothèses et l'élaboration de nos protocoles d'évaluation de la compréhension chez des auditeurs. Avant d'énoncer ces dernières, nous terminons cette partie en présentant un état de l'art des principales études qui ont porté sur la relation existant entre les performances d'auditeurs dans des tâches de retranscription orthographique et leurs performances dans des tâches impliquant un niveau de conceptualisation plus élevé, c'est-à-dire dans des tests visant à évaluer la *compréhension de la parole*.

3.4 Relation entre les mesures subjectives d'intelligibilité et la compréhension de la parole : état de la question et hypothèses

3.4.1 État de la question dans la littérature

À notre connaissance, la question de la validité externe des tâches de retranscription de mots vis-à-vis de la compréhension de la parole n'a que peu été traitée dans la littérature. Beukelman et Yorkston (1979) ont été les premiers à se poser cette question, dans l'article « The relationship between information transfer and speech intelligibility of dysarthric speakers ».

Par « transfert d'information » (*information transfer*), il faut entendre ici *succès de la communication*. Beukelman et Yorkston se situent en effet dans une perspective fonctionnelle, et citent Sarno (1968) lorsqu'elle énonce le besoin de disposer de méthodes d'évaluation qui permettent de « refléter le succès avec lequel le locuteur peut se faire comprendre, malgré le trouble de la parole dont il souffre »³⁸. C'est donc bien de la question de la validité des tests d'intelligibilité qu'il s'agit ici. Pour Beukelman et Yorkston, cette validité est fortement compromise si les scores d'intelligibilité ne correspondent pas fidèlement à une mesure du succès de la communication – évalué par l'aptitude de l'auditeur à interpréter les phrases entendues.

Beukelman et Yorkston ont enregistré 9 locuteurs dysarthriques lisant un texte narratif composé de trois paragraphes, et trois listes de 50 mots – incluant les mots les plus « informatifs » contenus dans le texte narratif. Pour tester l'intelligibilité, deux tâches classiques ont été proposées aux auditeurs :

- la retranscription des paragraphes, phrase par phrase. Le score d'intelligibilité correspondait alors au taux de mots correctement retranscrits (sans tenir compte de graphies erronées) ;
- la retranscription des mots contenus dans les 3 listes. Ici également, le score d'intelligibilité correspondait au taux de mots correctement retranscrits.

Pour tester le succès de la communication, Beukelman et Yorkston ont évalué la compréhension des auditeurs en posant, pour chacun des trois paragraphes, 10 questions relatives au contenu textuel entendu³⁹. Cette procédure correspond au type d'évaluation de la compréhension orale que l'on peut trouver dans des classes de langue étrangère.

Beukelman et Yorkston ont trouvé des corrélations quasi parfaites entre les mesures du transfert d'information et les scores de retranscription de mots (90% pour les listes de mots, 95% pour la retranscription des paragraphes). Ceci permet aux auteurs de conclure que les résultats de tâches de retranscription de mots peuvent être utilisés comme un indice de la *performance communicative des locuteurs*.

38. « [...] reflect how well the speaker can make himself understood in spite of whatever speech impairment he may exhibit » (Sarno, 1968).

39. malheureusement le jeu de questions utilisé par les auteurs n'est pas disponible.

Cependant 13 ans plus tard Weismer et Martin (1992) ont relevé un problème majeur dans l'étude de Beukelman et Yorkston, qui compromet fortement l'interprétation des résultats. Weismer et Martin ont souligné le fait que Beukelman et Yorkston n'ont pas tenu compte dans leurs calculs statistiques de la sévérité des troubles dont souffraient les locuteurs. Ceci peut en effet produire un biais statistique. On peut supposer que les scores d'intelligibilité, tout comme les scores de compréhension, dépendent du degré de l'atteinte dont souffraient les patients. La véritable relation qu'entretiennent les scores d'intelligibilité et de compréhension peut donc être masquée par la corrélation de ces deux variables avec la sévérité des TPPP dont souffraient les patients.

On peut donc supposer que la relation entre l'intelligibilité et la compréhension de la parole serait d'un ordre plus faible si l'on contrôlait la variable « sévérité ». C'est l'hypothèse que Hustad (2008) s'est proposé de tester, en employant globalement le même protocole expérimental que pour l'étude de Beukelman et Yorkston⁴⁰.

L'étude de Hustad (2008), relatée dans l'article « The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers with Dysarthria », impliquait la participation de 12 patients dysarthriques, dont 3 présentaient des atteintes légères (*mild*), 3 des atteintes modérées (*moderate*), 3 des atteintes sévères (*severe*), et 3 des atteintes profondes (*profound*). Les patients ont été enregistrés lors de la lecture de trois passages narratifs. Les auditeurs, au nombre de 144, ont été soumis à deux tâches :

- la retranscription orthographique des trois passages, phrase par phrase. L'intelligibilité correspondait alors au taux de mots correctement retranscrits (sans tenir compte des erreurs d'orthographe, ou des homonymes) ;
- la réponse à 10 questions de compréhension portant sur chaque passage. Pour chaque groupe de 10 questions, 5 étaient de nature plutôt factuelle (ex. « Quand est-ce que cette histoire se passe ? »), 5 étaient au contraire de nature essentiellement inférentielle (ex. « Quel est le sujet de l'histoire ? »). La correction des réponses aux questions de compréhension étaient ensuite évaluée sur des échelles à trois points (0, 1, 2).

Afin d'étudier la corrélation existant entre les scores d'intelligibilité et de compréhension, Hustad a calculé une corrélation partielle, tenant compte de l'influence du degré de sévérité sur les deux variables d'intérêt. Cette procédure statistique permet de retirer de l'indice de corrélation entre les scores d'intelligibilité et de compréhension la force d'association que les deux variables entretiennent avec le degré de sévérité. Hustad trouve ainsi une corrélation résiduelle très basse entre les deux variables : sur l'ensemble des locuteurs la corrélation obtenue entre les scores d'intelligibilité et de compréhension n'est que de 5,6%, et elle n'est pas statistiquement

40. à noter que, puisque l'hypothèse de Hustad ne concernait que la procédure de calcul statistique à employer (c.-à-d. une corrélation partielle plutôt qu'une corrélation bivariée), et que Beukelman et Yorkston avaient classé les locuteurs en groupes de sévérité, Hustad aurait pu se contenter de reprendre les données de Beukelman et Yorkston et d'appliquer un calcul de corrélation partielle pour vérifier son hypothèse.

significative ($P = 0,501$).

Les résultats trouvés par Hustad se situent donc à l'opposé de ceux mis au jour par Beukelman, ce qui amène l'auteure à énoncer : « Le manque de correspondance entre l'intelligibilité et la compréhension de l'auditeur a d'importantes implications cliniques. Une de ces implications est liée à la performance communicative, un concept qui pourrait grandement différer de l'intelligibilité »⁴¹.

Cependant, nous nous demandons si les résultats de Hustad ne sont pas discutables. Il paraît en effet assez contre-intuitif que les scores d'intelligibilité et les scores de compréhension ne soient pas du tout liés, puisque l'on peut supposer que tous les deux reposent, au moins en partie, sur ce que l'auditeur parvient à percevoir de l'input acoustique. Même si, comme le souligne Hustad, les deux tâches ne font sans doute pas appel aux mêmes mécanismes cognitifs chez les auditeurs (l'une concentre l'attention des auditeurs sur la forme des messages, la seconde sur le sens), et même si le *co(n)texte*⁴² et les connaissances de l'auditeur jouent certainement un rôle plus important dans la tâche de compréhension que dans celle d'intelligibilité, les deux mesures devraient être, au moins en partie, liées.

Cette réflexion nous a amené à observer plus en détails la méthodologie employée pour cette étude. Les trois variables principales utilisées par Hustad sont les *scores d'intelligibilité*, les *scores de compréhension*, et la *sévérité de la dysarthrie* dont souffraient les locuteurs. Les scores d'intelligibilité correspondent au pourcentage de mots correctement retranscrits par les auditeurs après écoute de chaque phrase. Les scores de compréhension correspondent au nombre de points (0, 1, 2) obtenus par les auditeurs à chaque question de compréhension (10 pour chacun des 3 paragraphes) – donnant ainsi un score maximal de 60. Reste la variable « sévérité de la dysarthrie ». Hustad (2008, p. 564) dit que les 12 locuteurs dysarthriques ont été répartis en 4 groupes de sévérité uniquement sur la base des scores qu'ils ont obtenus au SIT Yorkston et Beukelman (1996). Or le SIT, le *Sentence Intelligibility Test*, est une tâche d'intelligibilité tout à fait comparable avec celle employée par Hustad pour attribuer les scores d'intelligibilité dans son étude : il s'agit pour les auditeurs de retranscrire des phrases, et le score correspond au nombre de mots correctement retranscrits. On peut donc supposer une corrélation assez forte entre la mesure de la sévérité et les scores d'intelligibilité dans cette étude. Or comme nous l'avons dit Hustad a calculé une corrélation partielle entre les scores d'intelligibilité et de compréhension, calcul tenant compte de la force d'association existant entre chacune de ces deux variables et la sévérité de la dysarthrie. Étant donné la relation très forte qu'il doit exister entre la variable sévérité et la variable intelligibilité dans cette étude, il n'y a peut-être rien d'étonnant à ce que la corrélation résiduelle obtenue pour rendre compte de la relation entre intelligibilité et compréhension soit quasi nulle. Nous pensons donc qu'il y a là un biais statistique, biais opposé à celui de

41. « There are important clinical implications for the mismatch between intelligibility and listener comprehension. One implication relates to communicative functioning, a construct that in fact may be quite different from intelligibility. » (Hustad, 2008, p. 572).

42. par *co(n)texte* nous entendons à la fois le *cotexte* – c'est-à-dire le contexte linguistique – et le *contexte*, qui renvoie à la situation de communication.

l'étude antérieure de Beukelman et Yorkston (1979).

3.4.2 Hypothèses

Nos hypothèses principales concernant la nature et la force de la relation existant entre l'intelligibilité et la compréhension de la parole sont les suivantes : premièrement, nous supposons qu'un test portant sur les mécanismes relevant de la compréhension de la parole, telle que nous l'avons définie en section 3.3 (p. 78), donnera lieu à des performances globalement plus élevées qu'un test de retranscription de mots. En effet, comme nous l'avons souligné dans la section 3.2.1 (voir p. 65), il existe des effets descendants (*top-down*) dans le traitement de la parole. Or ces effets sont certainement moindres dans les tests de retranscription classiques où les effets du contexte extralinguistique sont soigneusement éliminés. Les sources d'information accessibles à l'auditeur étant plus nombreuses dans le cadre de la compréhension de la parole en contexte, les scores seront plus élevés, comme cela a été suggéré par les résultats obtenus dans les études de Beukelman et Yorkston (1979) et deHustad (2008)..

Cette hypothèse est d'une grande importance en regard de l'utilisation des tests pour le suivi clinique de patients. En effet, comme cela a été souligné par le courant du *langage en tant qu'action* (voir section 3.2.2, p. 71), et également suggéré par des études plutôt situées dans le courant dit du traitement de l'information (*cf.* par ex. Hagoort et van Berkum, 2007; Kuperberg, 2007; Lotze *et al.*, 2011, *cf.* section 3.2.1 p. 60) l'intégration des informations contextuelles semble être inhérent au processus de compréhension de la parole. Or, dans la mesure où un objectif majeur de l'évaluation du patient est justement de rendre compte de son aptitude à se faire comprendre verbalement dans des situations de communication, il serait donc nécessaire de disposer aussi de tests mettant en jeu, chez l'auditeur, ces mécanismes descendants. Cette position diffère donc de celle consistant, au contraire, à éviter toute source d'information autre que le signal de parole pour l'auditeur (comme par exemple dans les tests aux logatomes ou les tâches de retranscription de listes de mots) en le mettant dans une situation de laboratoire.

Ensuite, nous supposons que la différence quantitative entre les scores de retranscription et les scores de compréhension sera d'autant plus grande que la dégradation du signal de parole sera importante. L'hypothèse étant que, dans des conditions d'écoute difficiles l'auditeur s'appuiera davantage sur les informations fournies par le contexte extralinguistique. En d'autres termes, selon notre hypothèse la différence entre les scores de compréhension et d'intelligibilité ne sera pas constante, mais variera en fonction de la qualité du signal de parole.

Enfin, en ce qui concerne la force d'association entre les scores d'intelligibilité et de compréhension de la parole, nous supposons, comme nous l'avons souligné dans la section précédente, que des scores obtenus par des tâches de retranscription et des scores obtenus par une tâche de compréhension seront corrélés de manière statistiquement significative. En effet, même si les deux tâches impliquent différents processus cognitifs chez l'auditeur, elles reposent au moins en partie sur l'intégrité relative des signaux de parole reçus par les auditeurs. Nous supposons donc que, même

en tenant compte de l'influence des conditions de dégradation du signal (par ex. de la sévérité des troubles dont souffrent les patients), nous trouverons une corrélation statistiquement significative entre les scores d'intelligibilité et de compréhension de la parole. Cependant, nous supposons également que la force de cette relation (le coefficient de corrélation) ne sera pas aussi forte que cela a pu être suggéré par l'étude de Beukelman et Yorkston (1979). En effet selon nous la nature de la tâche à laquelle est soumise l'auditeur joue un rôle primordial. Or dans une tâche de retranscription l'attention de l'auditeur est plutôt tournée vers la forme du message (à travers la perception des phonèmes, notamment), alors que dans une tâche de compréhension son attention est davantage tournée vers le sens des messages entendus.

Vérifier ces hypothèses implique bien évidemment que nous élaborions un protocole expérimental permettant de faire effectuer à des sujets auditeurs deux types de tâches : l'une consistant en une tâche de retranscription orthographique, telles qu'elles sont habituellement conduites pour évaluer l'intelligibilité de la parole, et l'autre permettant cette fois d'appréhender la dimension de la compréhension de la parole. Puisque les tests d'intelligibilité portant sur la retranscription orthographique utilisent une procédure connue et reproductible, notre réflexion a principalement porté sur le développement d'un test de compréhension de la parole.

Chapitre 4

Choix préliminaires et développement du test EloKanz

Sommaire

4.1	Choix d'une procédure pour l'évaluation de la compréhension de la parole	86
4.1.1	Contraintes	86
4.1.2	Panorama des tests de compréhension de la parole	87
	Évaluations subjectives de la compréhensibilité	87
	Évaluations s'appuyant sur la mesure de temps de réaction	88
	Questionnaires de contenu textuel	89
	Tâches de sélection d'images complexes	90
	Protocoles impliquant l'observation du comportement des auditeurs	91
	Conclusion	93
4.2	Création du logiciel EloKanz	96
4.2.1	Procédure générale	96
4.2.2	Fonctionnement du logiciel EloKanz	98
	Fonctionnement de l'interface pour le locuteur	98
	Fonctionnement de l'interface pour l'auditeur	103

4.1 Choix d'une procédure pour l'évaluation de la compréhension de la parole

4.1.1 Contraintes

Deux aspects principaux ont guidé notre réflexion quant au choix d'un test pour évaluer la compréhension de la parole :

1. le respect d'une conception pragmatique de la compréhension de la parole, telle que nous l'avons présentée dans le chapitre précédent ;
2. le respect des contraintes imposées par la vérification de nos hypothèses de recherche.

Concernant le premier point, nous avons défini la compréhension de la parole comme un processus reposant non seulement sur le décodage de la parole, mais aussi sur la prise en compte d'autres informations liées au contexte dans lequel l'interaction prend place (notamment les objets environnant les interlocuteurs) et pouvant faire l'objet d'attentes de la part des interlocuteurs. Nous souhaitons par conséquent utiliser un test plaçant les interlocuteurs dans un environnement physique qu'ils partagent (en partie ou totalement), et à propos duquel ils peuvent communiquer. De même, nous avons vu à travers les études issues de la pragmatique que toute situation de communication impliquait des *intentions* de communication chez les interlocuteurs. Notre second objectif est donc d'utiliser un test qui implique chez les interlocuteurs d'avoir de véritables intentions de communication – que celles-ci soient naturelles ou induites par une consigne, différant par là des tâches de retranscription orthographique où le locuteur n'a finalement comme objectif communicatif que de « faire écrire » à son interlocuteur des items de langue.

Concernant maintenant le second point, à savoir le respect des contraintes imposées par notre problématique de recherche, nos exigences concernent principalement la forme des scores. En effet, nous souhaitons mettre en rapport des scores de retranscription orthographique avec des scores de compréhension de la parole, comme cela a été fait dans les études de Beukelman et Yorkston (1979) et de Hustad (2008). Nous avons donc besoin que la forme des résultats soit similaire dans les deux tests. Autrement dit, il faudrait, dans l'idéal, que les scores de compréhension se présentent également sous la forme d'un pourcentage d'items « corrects » (c.-à-d. ici *compris*)¹.

La table 4.1 récapitule les différentes contraintes que nous avons définies pour le choix d'un test de compréhension de la parole.

1. nous n'impliquons pas ici que les dimensions de l'intelligibilité et de la compréhension sont directement comparables. Cette contrainte concernant la forme des scores est plutôt là pour pouvoir répliquer le type d'études qui a été mené par le passé (Beukelman et Yorkston, 1979; Hustad, 2008, en particulier). La question de savoir si les scores sont réellement comparables – notamment de par les différences qui existent entre les tâches d'intelligibilité et de compréhension de la parole – sera traitée dans la discussion faisant suite à notre expérience principale (chapitre 6).

TABLE 4.1 – Contraintes définies pour notre test de compréhension de la parole

Contraintes liées à la conception de la compréhension de la parole adoptée dans notre travail		Contrainte liée à notre problématique de recherche
Présence d'un contexte physique (en partie) partagé par les interlocuteurs	Interlocuteurs doués ou dotés de véritables intentions de communication	Scores pouvant être facilement comparés avec des scores de retranscription

4.1.2 Panorama des tests de compréhension de la parole

Dans cette section nous faisons un tour d'horizon des principaux tests de compréhension de la parole qui sont utilisés pour répondre à des problématiques de recherche en psychologie, en psycholinguistique, ou bien dans le cadre d'applications cliniques. Ce panorama n'a pas pour prétention d'être exhaustif : l'idée est de comprendre quels sont les grands types de tests de compréhension de la parole et en quoi certains répondent bien ou moins bien aux critères que nous avons fixés en fonction de notre cadre théorique et de nos objectifs de recherche.

Évaluations subjectives de la compréhensibilité

L'évaluation subjective de la compréhensibilité² est une procédure particulièrement répandue dans les études portant sur les accents étrangers (Wilson et Spaulding, 2010). Il s'agit pour les auditeurs d'évaluer *leur propre difficulté* à comprendre les paroles entendues (Burda *et al.*, 2003; Burda et Hageman, 2005; Munro et Dering, 1995). Le rôle du locuteur consiste à lire à voix haute des stimuli linguistiques (en général des mots et/ou des phrases), et la consigne de l'auditeur est d'évaluer sa propre aptitude à comprendre les paroles entendues, sur une échelle à intervalles réguliers. Les échelles sont parfois munies de descripteurs accompagnant les scores : par exemple, Burda et Hageman (2005) demandent aux auditeurs de noter la compréhensibilité de locuteurs sur une échelle de 1 à 7, la note de 1 correspondant à un locuteur « extrêmement facile à comprendre »³ et la note de 7 à un locuteur « extrêmement difficile à comprendre »⁴.

Cette procédure d'évaluation de la compréhension de la parole ne répond pas à

2. si nous utilisons ici le terme « compréhensibilité », c'est parce que ce type de test est employé pour évaluer les performances des *locuteurs*, contrairement aux autres types de test de compréhension de la parole que nous détaillons dans cette partie. En effet, ici la tâche de l'auditeur consiste à évaluer sa propre compréhension face aux paroles prononcées par un locuteur ; et le score résultant de cette évaluation est considéré comme un indice de la « compréhensibilité » du locuteur.

3. « A rating of 1 indicated the speaker was "extremely easy to understand" » (Burda et Hageman, 2005).

4. « A rating of 7 indicated the speaker was "extremely difficult to understand" » (Burda et Hageman, 2005).

nos attentes. Tout d'abord, la tâche qui est proposée au locuteur concerne uniquement la réalisation des formes linguistiques, et non l'utilisation de ces formes pour répondre à des objectifs et à un contexte particulier de communication. En effet les interlocuteurs ne sont pas dotés d'objectifs particuliers, leurs rôles consistant uniquement à lire à haute voix pour l'un, et à évaluer les productions pour l'autre.

Ensuite, la forme des résultats obtenus par ce type de test n'est pas optimale pour permettre une mise en rapport des scores avec ceux obtenus par une tâche de retranscription. En effet, contrairement aux pourcentages résultant d'une tâche de retranscription, les évaluations subjectives ne fournissent pas de données situées sur une échelle proportionnelle, mais des données de type ordinal⁵.

Enfin, nous pensons que cette tâche présente l'inconvénient de faire tenir un double rôle à l'auditeur, celui d'*évaluateur* et de *participant à la situation de communication*. Or dans notre conception pragmatique de la compréhension de la parole, l'auditeur ne peut évaluer sa propre compréhension des énoncés que *s'il connaît les intentions communicatives du locuteur*, autrement dit s'il est conscient de tout ce que le locuteur a voulu dire ou exprimer par la parole. En tant qu'*évaluateur*, l'auditeur doit donc connaître le contenu des propos exprimés par le locuteur – ce qui biaise sa compréhension. D'un autre côté, si l'auditeur ne connaît pas à l'avance le contenu véhiculé par les paroles du locuteur, alors il est placé dans de véritables conditions de communication – mais il ne peut pas juger de sa compréhension des paroles prononcées. Par conséquent, l'objectivité des indicateurs obtenus par ce type de méthode peut-être discutée⁶.

Évaluations s'appuyant sur la mesure de temps de réaction

Dans le domaine de la synthèse de la parole, de nombreux travaux ont porté sur la compréhension des signaux produits par des auditeurs, à travers des mesures de temps de réponse.

Le *Sentence-By-Sentence Listening Task (SBSLT)* élaboré par Ralston *et al.* (*cf.* notamment Ralston *et al.*, 1991) consiste à faire écouter des phrases à un auditeur qui a pour consigne d'effectuer une tâche d'ordre métalinguistique, comme :

- détecter la présence d'un mot cible dans la phrase ;
- indiquer si la phrase est vraie ou fausse.

5. en statistique, une échelle qualitative ordinale autorise de « hiérarchiser » des individus au sein d'une population (Ex. « très compréhensible » > « compréhensible »), mais certaines opérations mathématiques n'ont pas de sens dans une telle échelle de valeurs (ex. l'addition, la multiplication), limitant ainsi les possibilités de comparaisons. Ces opérations sont par contre possibles pour des données situées sur une échelle proportionnelle.

6. à notre avis ce problème se retrouve aussi dans les évaluations subjectives de l'intelligibilité. Dans la plupart des cas soit le patient prononce un discours libre (*cf.* par ex. les tests proposés par Auzou *et al.*, 1998; Beltyukova *et al.*, 2008; McHenry, 2011; Rousseaux *et al.*, 2001), soit il lit un texte à haute voix (*cf.* dans le Test Lillois de Communication ou TLD, Rousseaux *et al.*, 2001). Dans le cas du discours libre, l'auditeur ne peut connaître *a priori* les propos du locuteur, et il n'est donc pas apte à juger sa propre compréhension. Au contraire, dans le cas du texte lu, l'auditeur connaît *a priori* le contenu textuel – et sa compréhension peut être biaisée.

Dans les deux cas l'auditeur appuie sur un bouton pour donner sa réponse. Pour évaluer la compréhension de l'auditeur, outre la correction de ces réponses, l'idée est de considérer les temps de réaction de l'auditeur, en supposant qu'ils sont un indicateur de la difficulté de l'auditeur à comprendre la phrase (Carlson *et al.*, 1992; Hustad et Beukelman, 2002; Ralston *et al.*, 1991).

Dans ce type de protocole également, nos critères pour le choix d'un test de compréhension de la parole ne sont pas respectés. Les interlocuteurs ne partagent pas un contexte physique à propos duquel ils peuvent communiquer. Ils n'ont pas, non plus, de réelles intentions communicatives puisque le rôle du locuteur (humain ou système) est de produire des phrases qui ont été définies au préalable par l'expérimentateur, et le rôle de l'auditeur est de répondre à des questions d'ordre métalinguistique.

Questionnaires de contenu textuel

Un autre type de test consiste à questionner les auditeurs à propos du contenu textuel auxquels ils ont été exposés (Hustad et Beukelman, 2002; Hustad, 2008; Wilson et Spaulding, 2010). Autrement dit, après que les auditeurs ont écouté un stimulus (quelques phrases généralement), ils ont pour consigne de répondre à des questions d'ordre sémantique visant à vérifier leur compréhension du discours prononcé. C'est une pratique en cela tout à fait similaire aux tests habituellement utilisés en didactique des langues étrangères pour évaluer la compréhension orale des apprenants. Les questions peuvent être de nature factuelle, c'est-à-dire qu'elles portent sur des éléments explicites du discours, et/ou de nature davantage inférentielle⁷. De même, les réponses aux questions peuvent reposer sur des choix multiples, ou bien impliquer une production écrite – les auditeurs ayant par exemple pour consigne d'écrire un résumé du discours entendu (Higginbotham *et al.*, 1994) ou bien d'imaginer une suite à l'histoire (Hustad, 2008). Dans ce dernier cas la correction des productions écrites est généralement évaluée par un jury (Higginbotham *et al.*, 1994; Hustad, 2008).

En regard des hypothèses que nous souhaitons tester, la forme que prennent les résultats dans ce type de test est intéressante : ils peuvent se présenter sous la forme de pourcentages de réponses correctes, forme similaire aux scores de retranscription orthographique et pouvant par conséquent faciliter un processus de comparaison.

7. C'est-à-dire des questions nécessitant pour l'auditeur de mener une réflexion (par des procédés d'induction, de déduction ou d'abduction) s'appuyant sur les éléments qu'il a retenus du discours entendu. En d'autres termes, une question inférentielle ne porte pas directement sur des éléments présents dans le discours entendu (ex. le nom d'une personne, une date, etc.), mais implique une réflexion à un niveau supérieur. Par exemple, si le discours implique un personnage principal s'appelant « David », une question inférentielle pourrait être de demander si le personnage principal est un garçon. Cette question nécessite pour l'auditeur de faire une inférence suivant un procédé de déduction :

- le personnage principal s'appelle « David » (prémisse 1) ;
- en général les personnes s'appelant « David » sont des garçons (prémisse 2) ;
- Donc, le personnage principal est un garçon (conclusion).

Malgré cela, les questionnaires de contenu textuel ne répondent pas à notre conception pragmatique de la compréhension de la parole, dans la mesure où les interlocuteurs ne sont pas véritablement engagés dans une situation de communication impliquant une interaction entre eux (Hustad, 2008).

Tâches de sélection d'images complexes

Certains tests de compréhension de la parole ont été élaborés pour évaluer la capacité de l'auditeur à traiter des énoncés de plus ou moins grande complexité. C'est le cas de batteries de tests pour les personnes aphasiques, dont font partie le *Protocole Montréal-Toulouse d'examen linguistique du sujet aphasique (MT-86, Nespoulous et al., 1992)* et le *Bilingual Aphasia Test (BAT, Paradis, 1989)*. Il faut bien comprendre que, dans ce cas particulier, les déficiences dans la communication se situent principalement du côté de l'auditeur. Ces batteries de tests ont donc pour objectif d'identifier et de caractériser les difficultés *de l'auditeur*, avec une visée diagnostique.

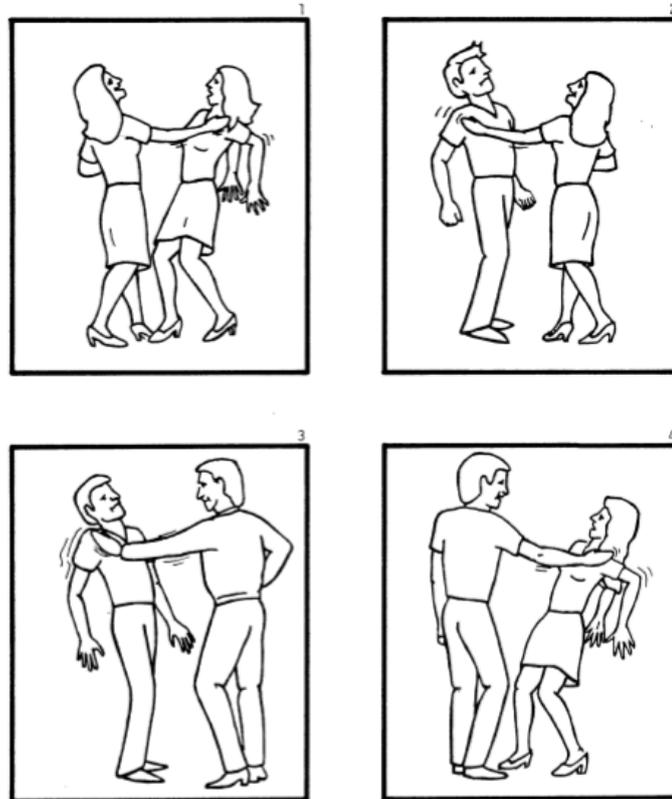
Ces tests visent en particulier la capacité de l'auditeur à interpréter des énoncés dont la complexité syntaxique va croissant, en fonction par exemple de l'utilisation de constructions passives, de propositions enchâssées, etc. Ces batteries proposent généralement, entre autres⁸, une tâche de sélection d'images. Cette tâche implique pour l'auditeur de sélectionner, parmi des ensembles d'images, celles qui correspondent aux énoncés qu'il entend. Dans un des tests du *Bilingual Aphasia Test* (Paradis, 1989) par exemple, l'auditeur doit sélectionner parmi les 4 images présentées en figure 4.1, celles qui correspondent aux énoncés suivants :

- (i) La fille pousse le garçon.
- (ii) Le garçon est poussé par la fille.
- (iii) C'est la fille que pousse le garçon.

Si l'on revient aux critères que nous avons défini pour le choix d'un test de compréhension de la parole (*cf.* table 4.1, p. 87), il nous semble qu'ici l'échange entre le locuteur et l'auditeur s'inscrit davantage dans une véritable situation de communication. En effet, les interlocuteurs partagent un même contexte représenté par un ensemble de référents possibles (le jeu d'images) auquel le locuteur fait référence, référence que doit résoudre l'auditeur par une tâche de sélection. S'agit-il pour autant d'une véritable interaction ? Dans une situation de communication réelle un énoncé déclaratif tel que « La fille pousse le garçon » n'a généralement pas pour but (ni pour résultat) de déclencher une quelconque réaction chez un interlocuteur. Les objectifs des participants sont ici plutôt liés à la passation d'une évaluation ; s'ils étaient véritablement engagés dans une situation de communication et que le locuteur désirait faire sélectionner une image particulière à son interlocuteur, il produirait un énoncé de type injonctif tel que, par exemple « prenez l'image sur laquelle une fille pousse

8. ces batteries de tests incluent généralement d'autres tâches de compréhension orale, comme des tests impliquant l'observation des réactions des auditeurs à des énoncés de type impératif – *cf.* section suivante.

FIGURE 4.1 – Exemples de stimuli utilisés dans le *Bilingual Aphasia Test* (Paradis, 1989)



un garçon ». Sur ce critère particulier, les tests de sélection d'images, tels que pratiqués dans les batteries de tests pour les sujets aphasiques, ne répondent pas à nos attentes.

Protocoles impliquant l'observation du comportement des auditeurs

Les batteries de tests cliniques pour l'évaluation de l'aphasie proposent également différents types de tâches qui sont davantage centrées sur l'interaction verbale entre patients et évaluateurs. C'est notamment le cas du *MT-86* (Nespoulous *et al.*, 1992), du *BAT* (Paradis, 1989) ou encore du *Token Test* (De Renzi et Vignolo, 1962), qui proposent tous des tests de compréhension verbale consistant à observer les réactions comportementales des auditeurs à des énoncés entendus.

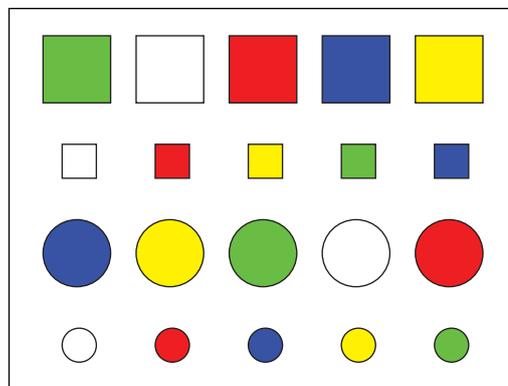
Dans le *BAT* (Paradis, 1989) par exemple le patient est invité à répondre à des ordres verbaux plus ou moins complexes tels que :

- (i) Fermez les yeux.
- (ii) Mettez la fourchette dans le verre.
- (iii) Voici trois crayons. Jetez le jaune par terre, donnez-moi le bleu et prenez le rouge.

Dans le *Token Test* (De Renzi et Vignolo, 1962) les auditeurs sont face à un panneau sur lequel sont représentées des figures géométriques de différentes formes, tailles et couleurs (figure 4.2). Le patient a alors pour consigne de répondre à des ordres verbaux dont la complexité (au niveau syntaxique) est variable, comme dans les énoncés suivants⁹ :

- (iv) Touchez le cercle rouge.
- (v) Touchez le cercle rouge et le carré jaune.
- (vi) Quand je touche le cercle vert, touchez le carré blanc.

FIGURE 4.2 – Exemples de stimuli utilisés pour un test de type *Token Test*



Ce genre de tâche ne se situe pas dans une dimension métalinguistique et ne nécessite pas, pour l'auditeur, de disposer de compétences en matière de compréhension ou de production écrite. C'est sans doute pour cette raison, entre autres, que ces tâches sont aussi utilisées en psychologie du développement et en psychologie animale pour évaluer la compréhension orale d'enfants en bas âge et d'animaux.

Herman *et al.* (1984), par exemple, ont étudié les compétences de compréhension d'un langage artificiel (dont les symboles sont des signaux acoustiques) chez le grand dauphin (*tursiops truncatus*). Pour ce faire, divers objets (balle, filet, etc.) étaient placés dans un bassin, lui même composé de certains éléments identifiables comme des fenêtres, une porte et un panneau. Des instructions plus ou moins complexes étaient données à deux dauphins *via* un émetteur acoustique sous-marin. Ces instructions étaient de complexité syntaxique variable, comme l'illustrent les exemples ci-dessous :

- (vii) Objet direct + action : FENÊTRE-TOUCHER AVEC LA QUEUE (Va toucher l'une des fenêtres avec la queue)

9. Pour des énoncés de ce type, seuls les cercles et carrés de grande taille sont présentés au patient.

- (viii) Complément modifieur + objet + action : GAUCHE-BALLE-BOUCHE (Va vers la balle située à ta gauche et touche-la avec la bouche)
- (ix) Complément modifieur + objet direct + action + modifieur + objet indirect : CERCEAU-BAS-AMENER-SURFACE-TUYAU (Prends le cerceau qui est situé au fond du bassin et amène-le jusqu'au tuyau situé à la surface)

Les phrases étaient considérées comme comprises par les dauphins si ces derniers effectuaient les actions correspondantes. Les scores de compréhension étaient exprimés sous la forme de pourcentages d'actions correctes (ex. de 50% à 57,1% pour les phrases les plus complexes – telles que l'exemple donné en ix).

Savage-Rumbaugh *et al.* (1993) ont utilisé le même type de test pour comparer la compétence de compréhension de la parole chez Alia, une enfant de deux ans, et Kanzi, un bonobo (*pan paniscus*). Entourés de divers objets (directement visibles ou non, voir figure 4.3), Kanzi et Alia avaient pour consigne de réagir à des énoncés verbaux. Ces énoncés pouvaient correspondre :

- soit à des demandes verbales de différents types (phrases impératives, demandes sous forme d'interrogations, de structures syntaxiques plus ou moins complexes, etc.), comme par exemple « amène la pomme de terre dans la chambre », « mange les raisins », etc. ;
- soit à des phrases déclaratives telles que « La surprise se cache dans le lave-vaisselle ». Dans ce cas pour les sujets la réaction correcte consistait précisément à ne pas réagir à l'énoncé.

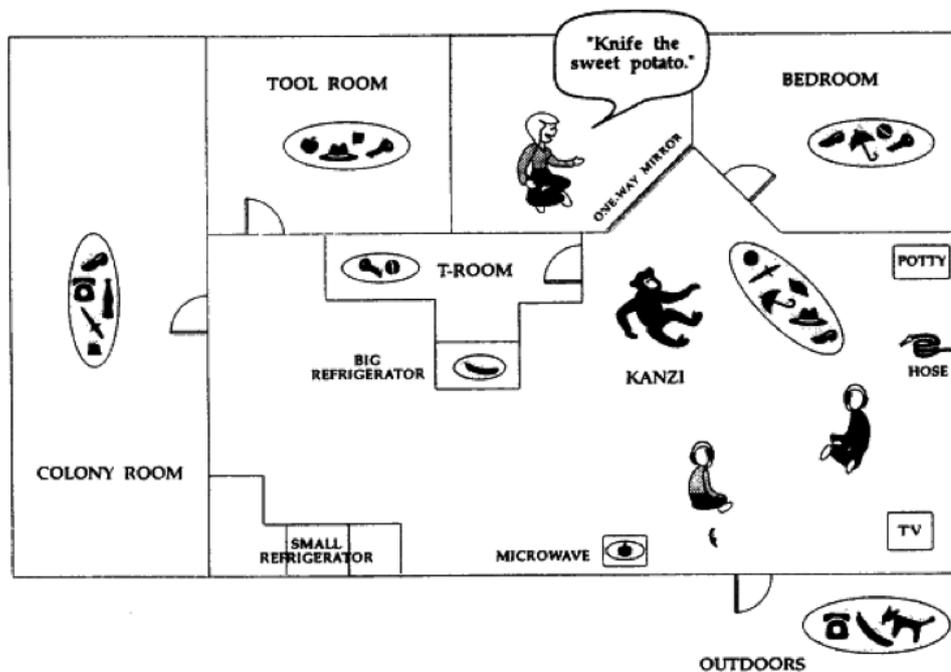
Les réactions des deux sujets étaient considérées soit comme correctes (le sujet effectue l'action demandée), soit comme partiellement correctes (le sujet effectue une partie de l'action commandée), soit incorrectes.

Ce type de protocole répond particulièrement bien à nos attentes. D'une part, les interlocuteurs partagent un même univers physique où sont rassemblés divers objets pouvant jouer le rôle de référents du discours. D'autre part, le locuteur et l'auditeur sont dans une véritable situation d'interaction, les réactions comportementales (les actions) de l'auditeur aux énoncés entendus étant considérés comme l'aboutissement de sa compréhension. Enfin, les scores sont représentés par des pourcentages d'actions correctes, forme tout à fait similaire avec celle sous laquelle se présentent les scores de retranscription orthographique.

Conclusion

Comme nous l'avons souligné, ce panorama des différents types de tests de compréhension de la parole n'avait pas pour prétention d'être exhaustif. D'autres types de tests existent, citons par exemple les méthodes portant sur l'observation des mouvements des yeux de l'auditeur (*eye-tracking*) ou sur l'activité électrique de l'encéphale (*EEG*). Cependant ces dernières méthodes ne permettent pas de juger si une personne comprend des énoncés dans leur ensemble mais plutôt si elle parvient à intégrer des éléments à un niveau plus fin (ex. résolution d'une référence langagière, traitement sémantique d'un mot particulier, etc.).

FIGURE 4.3 – Évaluation de la compréhension de la parole par le bonobo Kanzi (Savage-Rumbaugh *et al.*, 1993)



Parmi les différents tests que nous avons passés en revue, les procédures impliquant l'observation des réactions comportementales des auditeurs a particulièrement retenu notre attention. En effet, il satisfait tout d'abord tous les critères que nous avons définis (*cf.* table 4.2), en lien avec notre conception de la compréhension de la parole et avec nos perspectives de recherche.

Par ailleurs, nous avons beaucoup insisté sur le fait que les tests d'intelligibilité font, le plus souvent, intervenir la fonction métalinguistique du langage, au détriment des autres fonctions. Un test tel que celui utilisé par Savage-Rumbaugh *et al.* (1993) permet de dépasser ce cadre. Savage-Rumbaugh *et al.* testent d'ailleurs non seulement la fonction référentielle du langage (à travers la résolution de références à des objets ou des actions) mais aussi la fonction conative du langage, celle qui permet au locuteur de choisir de simplement affirmer quelque chose (*faire savoir*) ou de donner un ordre à son interlocuteur (*faire faire*).

En revanche, il nous est apparu qu'un tel test nécessitait certains aménagements dans notre cas, en fonction à la fois des contraintes pratiques liées à l'utilisation du test en laboratoire ou en milieu clinique, et aussi en fonction de notre problématique de recherche. En conséquence nous avons décidé d'adapter ce test, en passant notamment par une phase d'informatisation. Le développement du test sous la forme d'un logiciel permet en effet de faciliter la mise en œuvre de conditions expérimentales (tirages aléatoires, équilibrage des conditions, etc.) mais aussi de réduire les contraintes d'espace et de matériel nécessaires à la passation du test (les « vrais »

TABLE 4.2 – Récapitulatif des différentes méthodes d'évaluation de la compréhension orale, en fonction de nos critères de sélection

	Contraintes liées à la conception de la compréhension de la parole adoptée dans notre travail		Contrainte liée à notre problématique de recherche
	Présence d'un contexte physique (en partie) partagé par les interlocuteurs	Interlocuteurs doués ou dotés de véritables intentions de communication	Scores pouvant être facilement comparés avec des scores de retranscription
Évaluations subjectives	Non	Non	Non (données d'échelle)
Mesures de temps de réaction	Non	Non	Non (temps de réaction)
Questionnaires de contenu textuel	Non	Non	Oui (% de réponses correctes)
Tâches de sélection d'images	Oui	Non	Oui (% de réponses correctes)
Observation du comportement	Oui	Oui	Oui (% d'actions correctes)

objets sont remplacés par des images représentées sur un écran). Nous consacrons la section suivante à la description du développement de ce test, que nous avons baptisé « EloKanz » en hommage à Kanzi, le bonobo ayant participé à l'étude de Savage-Rumbaugh *et al.* (1993).

4.2 Création du logiciel EloKanz

Pour développer le test EloKanz, nous avons défini une procédure générale fondée sur l'observation des réactions comportementales des auditeurs à des énoncés entendus, dans la continuité des tests proposés par Savage-Rumbaugh *et al.* (1993). Cette procédure fait intervenir deux interfaces logicielles, l'une pour le locuteur (*cf.* p. 98) et l'autre pour l'auditeur (*cf.* p. 103). La confrontation des données relatives aux consignes données par le locuteur avec celles correspondant aux actions effectuées par l'auditeur permet au programme EloKanz de vérifier la compréhension de l'auditeur.

4.2.1 Procédure générale

Si l'on résume les principaux aspects des tests de compréhension de la parole fondés sur l'observation des réactions comportementales des auditeurs, nous pouvons établir la liste de contraintes suivantes pour le développement de notre test :

- (i) les sujets (locuteur et auditeur) doivent pouvoir communiquer par la parole, et uniquement par la parole. Par exemple, pour supprimer les informations que peuvent apporter les gestes et mimiques du locuteur, Savage-Rumbaugh *et al.* (1993) demandent au locuteur et à l'auditeur de se tenir dans des pièces séparées, ou bien demandent simplement au locuteur de porter un masque et d'effectuer le moins de mouvements possibles lors de l'énonciation des consignes orales ;
- (ii) les sujets doivent pouvoir communiquer à propos d'objets présents dans un environnement qu'ils partagent ;
- (iii) l'auditeur doit pouvoir répondre aux commandes par la manipulation des objets ;

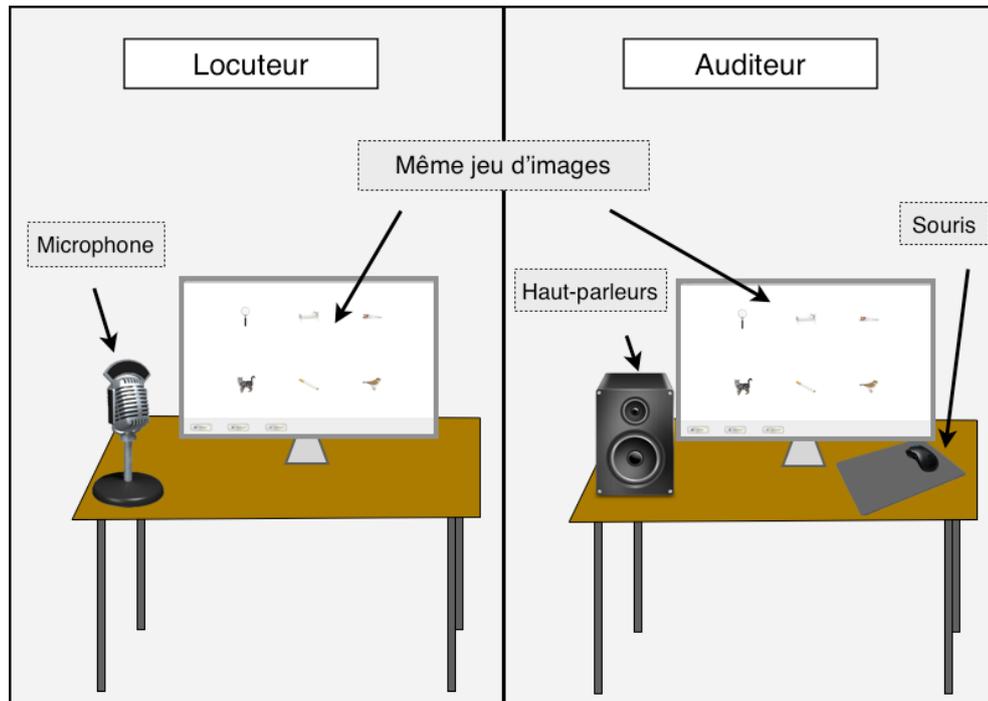
D'autres contraintes, liées à la fois à notre problématique et à nos hypothèses, viennent compléter cette liste :

- (iv) dans notre cas, la personne en position d'« instructeur » (celle qui donne des ordres verbaux à l'auditeur) n'est pas l'évaluateur mais bien *le sujet qui fait l'objet de l'évaluation*, c'est-à-dire le patient. Par conséquent il faut donner au patient des consignes relatives aux actions qu'il doit faire exécuter à l'auditeur ;
- (v) dans le cas de réactions incorrectes ou partiellement incorrectes de la part de l'auditeur, il doit être possible de déterminer quels sont les éléments qui n'ont pas été compris, afin d'être éventuellement en mesure de donner des scores partiels, comme dans la procédure de Savage-Rumbaugh *et al.* (*cf.* 1993) ;
- (vi) les scores doivent pouvoir se présenter sous la forme de pourcentages d'actions correctes, pour faciliter la comparaison avec les performances des auditeurs dans des tâches de retranscription.

Compte-tenu de ces différentes contraintes nous avons défini la procédure générale suivante. Le locuteur et l'auditeur sont tous les deux face à un écran d'ordinateur, sur lequel ils peuvent voir les mêmes objets. Les passations du locuteur et

de l'auditeur se déroulent en différé : ils ne communiquent pas directement, comme nous le verrons par la suite. Le locuteur dispose d'un microphone¹⁰, et l'auditeur de haut-parleurs et d'une souris (cf. figure 4.4).

FIGURE 4.4 – Matériel prévu pour le locuteur et l'auditeur dans le test EloKanz



De façon générale, nous avons prévu qu'une passation de test se déroule de la manière suivante :

- a) du côté « locuteur » : pour chaque item de test, le locuteur reçoit une consigne lui indiquant une action à faire exécuter à son interlocuteur. Ces consignes concernent toutes des *déplacements d'images*, et se présentent sous la forme d'animations : par exemple, le locuteur voit d'abord à l'écran une image de chat se déplacer depuis sa position d'origine vers un emplacement situé sous une image de voiture. Dans ce cas le locuteur a pour consigne de donner une commande orale (par exemple « Mettez le chat sous la voiture ») qui vise à *faire exécuter l'action correspondante à son interlocuteur*. Cette consigne verbale est enregistrée *via* le microphone. Le test se poursuit ainsi, jusqu'à l'épuisement des animations diffusées au locuteur ;
- b) du côté « auditeur » : chaque stimulus correspond à une phrase enregistrée par le locuteur. La phrase est diffusée *via* des haut-parleurs ou bien *via* des écouteurs. L'auditeur a pour consigne d'exécuter l'ordre verbal qui lui a été donné, en utilisant la souris et en agissant sur les images affichées à l'écran. Le test se poursuit

10. le microphone représenté dans la figure 4.4 est juste une illustration : nous n'avons pas utilisé de microphone de ce type pour nos expériences mais un micro-casque unidirectionnel.

ainsi, consigne orale après consigne orale. Le logiciel évalue alors les actions effectuées par l'auditeur (sélection et déplacements d'objets) en les comparant aux animations qui ont été jouées au locuteur lorsqu'il a enregistré ses phrases (nous explicitons cette dernière étape plus loin, à la fin de la section *Fonctionnement de l'interface pour l'auditeur*).

Comme nous pouvons le remarquer, nous avons fait le choix de nous concentrer dans un premier temps sur la vérification de la compréhension d'énoncés verbaux de type impératif. Nous avons fait ce choix pour limiter à la fois le nombre de variables à observer pour vérifier nos hypothèses, et également pour limiter notre travail de développement informatique du test EloKanz.

Pour ce développement nous avons choisi d'utiliser le langage de programmation *Lingo* et l'environnement *Adobe Director*[®] (version 11.5). *Lingo* est un langage événementiel, qui permet de travailler avec des animations et des actions portant sur des objets (appelés *images-objets*). De plus, *Lingo* permet grâce à l'utilisation de plug-ins (notamment *Buddy API*[®] de *Magic Modules*, et *Audio XTRA*[®] de *Tableiro*) d'enregistrer et de récupérer des fichiers de résultats sur le disque dur, ainsi que d'enregistrer des fichiers sonores.

Nous avons été amené à faire de nombreux aménagements et perfectionnements concernant des points précis des interfaces *locuteur* et *auditeur*. Dans la suite de cette section nous présentons donc plus en détail ces deux interfaces, ainsi que la façon dont nous avons informatisé la mise en place des passations et le recueil des données (consignes et résultats). Dans cette partie nous ne traitons pas de la question de la sélection des images, à laquelle nous avons consacré plusieurs études que nous relatons dans le chapitre 5.

4.2.2 Fonctionnement du logiciel EloKanz

Fonctionnement de l'interface pour le locuteur

Préparation de la passation

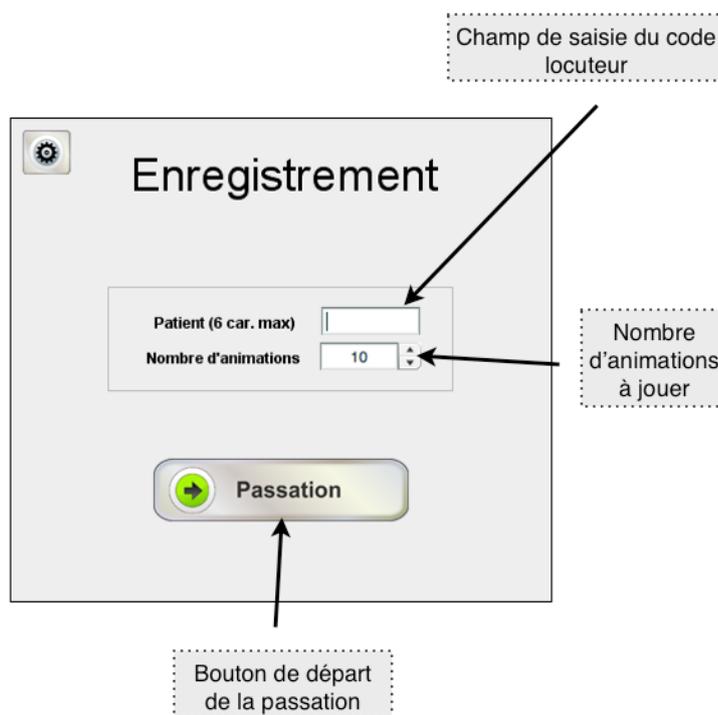
Le côté « locuteur » du logiciel EloKanz doit permettre de visionner des animations tenant le rôle de consignes, et d'enregistrer les commandes verbales que le locuteur prononce à leur suite. Pour créer chaque animation qui sera jouée au locuteur, le programme choisit deux images au hasard (l'une source, l'autre cible), et détermine (de manière aléatoire) si l'image source ira se déplacer à droite, à gauche, au-dessus ou en dessous de l'image cible.

La liste des animations qui seront jouées au locuteur est créée dès que l'expérimentateur a saisi le nom (ou le code) du locuteur et le nombre d'animations à jouer (c.-à-d. le nombre de phrases à enregistrer), et a cliqué sur le bouton « Passation » (cf. figure 4.5).

Session d'enregistrement

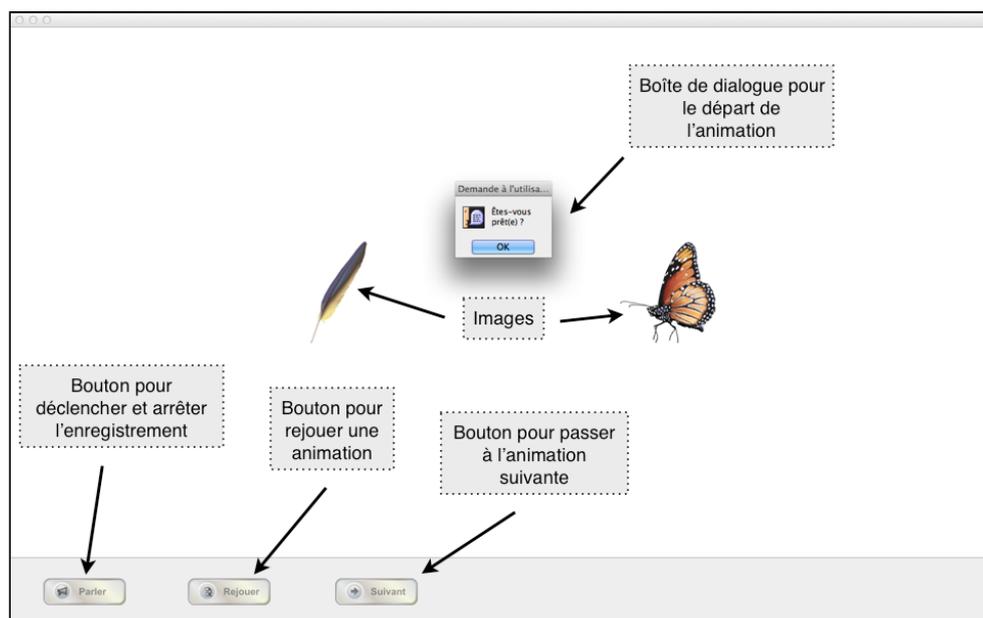
De manière générale, nous avons prévu qu'une session d'enregistrement se déroule de la manière suivante. Tout d'abord, toute passation commence par la lecture

FIGURE 4.5 – Fenêtre de saisie du code locuteur et du nombre d’animations à jouer



automatique de la première animation : le locuteur signale qu’il est prêt en cliquant sur le bouton [OK] dans la boîte de dialogue au centre de l’écran (figure 4.6).

FIGURE 4.6 – Interface prévue pour le locuteur



L'animation se joue ; par exemple le papillon vient se placer au-dessus de la plume¹¹. S'il le désire, le locuteur peut visionner à nouveau cette animation en cliquant sur le bouton [REJOUER]. Lorsqu'il est prêt, le locuteur peut enregistrer sa commande verbale : pour cela il clique alors sur le bouton [PARLER]. Un voyant animé (roue qui tourne) sur le bouton et une étiquette « Stop » signalent alors au locuteur que l'enregistrement a débuté. Le locuteur donne sa consigne verbale, par exemple « Mettez le papillon au-dessus de la plume » *via* un microphone, puis clique sur le même bouton pour terminer l'enregistrement. S'il n'est pas satisfait de son enregistrement, le locuteur a la possibilité de réenregistrer sa consigne autant de fois qu'il le souhaite en cliquant à nouveau sur le bouton [PARLER] : les anciens enregistrements sont automatiquement remplacés par les nouveaux. Enfin, lorsqu'il est satisfait, le locuteur peut cliquer sur le bouton [SUIVANT] pour passer à l'animation suivante. Le locuteur a également la possibilité de passer à l'animation suivante sans avoir enregistré aucune phrase (par exemple s'il n'a pas reconnu un objet). Dans ce cas le nombre d'animations qui lui ont été jouées n'est pas incrémenté ; autrement dit, il devra enregistrer une phrase supplémentaire.

Au début du développement du logiciel, le déplacement des images se faisait avec une vitesse constante depuis la position source vers la position cible. Cependant, ce mode de déplacement nous a paru quelque peu artificiel :

- d'une part, le déplacement semblait uniforme, sans phase d'initiation ou de clôture. Ainsi, dans les jeux vidéos les déplacements d'objets sont habituellement conduits avec un ralentissement en début ou en fin de mouvement (*ease-in* et *ease-out*, Penner, 2002) ;
- d'autre part, ces animations servent à aider le locuteur à imaginer une action qui serait effectuée par son interlocuteur. Or un déplacement simple de l'objet ne paraît pas représenter une action initiée par un acteur humain, donnant par là un caractère assez artificiel aux animations.

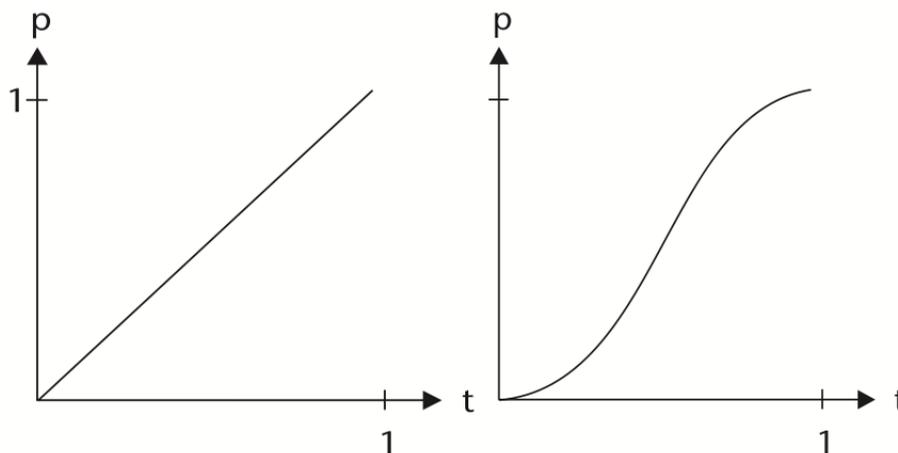
Pour pallier ces deux problèmes, nous avons décidé :

- (i) de modifier l'algorithme de déplacement des images pour lui donner un caractère plus naturel ;
 - (ii) d'ajouter des éléments graphiques permettant d'imaginer que les déplacements sont initiés par un acteur humain.
- (i) Concernant le déplacement des images, Penner (2002) présente plusieurs algorithmes permettant de commencer et de terminer les déplacements en douceur (voir figure 4.7).

Nous avons choisi d'utiliser une fonction sinusoïdale de déplacement (*sinusoïdal easing*), car elle nous semblait donner un rendu très naturel. L'algorithme de déplacement impliquant cette fonction repose sur le calcul de la position de l'objet déplacé

11. plus tard, nous améliorerons ce type d'animation : ce sera une main qui viendra saisir un objet et le déplacer (voir *infra*).

FIGURE 4.7 – Fonctions de déplacement linéaire (à gauche) et de déplacement avec entrée et sortie en douceur (*ease-in* et *ease-out*, à droite) – Penner (2002, pages 201 et 209). Dans chacun de ces graphiques l'axe t représente le temps, et l'axe p représente la position de l'image déplacée.



tout au long de son déplacement. Ce calcul de la position de l'objet est traduit par la fonction mathématique ci-dessous (Penner, 2002, p. 215)¹² :

$$f(t, b, c, d) = \frac{c}{2}(1 - \cos(\pi \frac{t}{d})) + b$$

où :

- d représente la durée prévue pour le déplacement ;
- t représente le temps déjà écoulé depuis le départ du déplacement ;
- b représente la position initiale de l'objet ;
- c représente la distance qu'il reste à parcourir à l'objet pour atteindre la position finale.

Le rendu donné par cet algorithme de déplacement semble bien plus naturel qu'un mouvement linéaire tel que celui que nous avons implémenté au départ. À titre d'exemple, nous avons fourni en annexe DVD 3a_1 des exemples vidéo d'animations¹³.

(ii) Pour que les déplacements d'images ressemblent plus à des actions effectuées par *une personne* sur des objets, nous avons ajouté deux éléments graphiques intervenant dans les animations. Le premier élément est une main qui, pour chaque animation, vient saisir l'image (figure 4.8, images 1 à 4), la déplace vers sa position cible, la relâche (figure 4.9 images 1 à 3) et revient à sa position initiale.

12. dans l'ouvrage de Penner cette fonction est présentée avec le lexique et la syntaxe propres au langage *ActionScript*. Nous la présentons ici sous sa forme mathématique. Dans le programme EloKanz, cette fonction est implémentée en langage *Lingo*.

13. le lecteur peut également essayer la première version du programme EloKanz, donnée en annexe DVD 1a. Cette version du programme fonctionne sous le système d'exploitation Mac OS.

FIGURE 4.8 – Images utilisées pour l’animation de la main, lorsque celle-ci saisit un objet

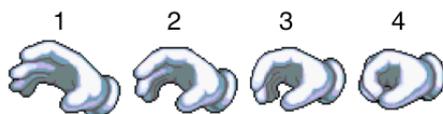


FIGURE 4.9 – Images utilisées pour l’animation de la main, lorsque celle-ci relâche un objet



Le second élément ajouté est un cadre venant entourer les images lorsque celles-ci sont déplacées (figure 4.10).

FIGURE 4.10 – Main déplaçant une image encadrée. Ici la main déplace le serpent au-dessus du seau.



L’effet rendu par l’ajout de ces éléments graphiques est satisfaisant, dans la mesure où l’on a davantage l’impression de regarder une action effectuée par quelqu’un que de simplement assister à un déplacement généré par le programme informatique.

Enregistrement audio des consignes orales données par le locuteur

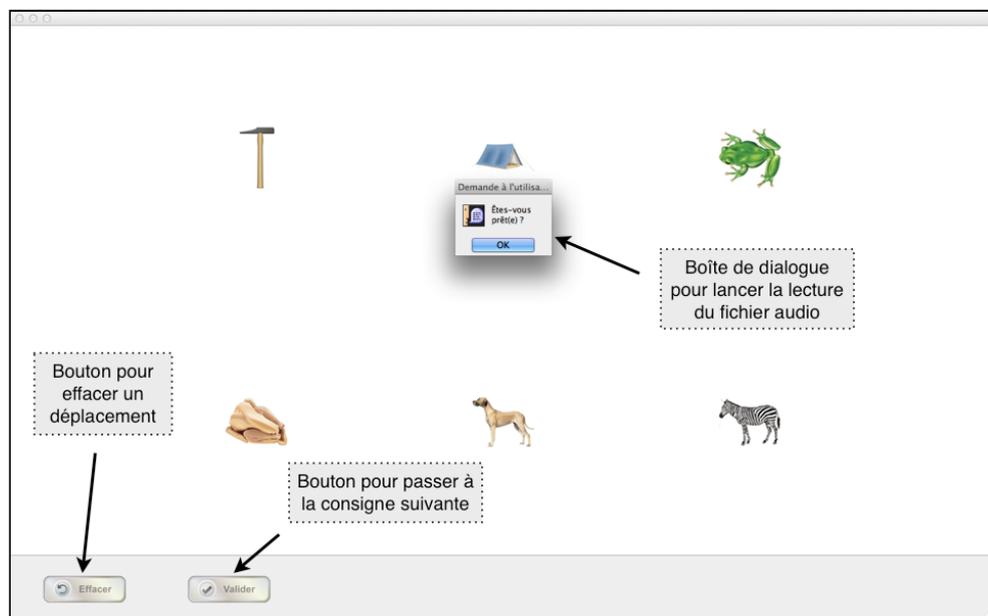
Dans le programme EloKanz, l’enregistrement des données sonores se fait par des fonctions propres au plug-in *Audio XTRA*[®], développé par la société *Tabuleiro*.

Lorsque le locuteur clique sur le bouton [PARLER], un fichier audio de type AIF (*AIFF, Audio Interchange File Format*) est créé. Les données sonores y sont enregistrées avec une fréquence d'échantillonnage de 44100 Hz, et avec une quantification des niveaux sonores de 16 bits, jusqu'à ce que le locuteur clique à nouveau sur le bouton pour terminer l'enregistrement.

Fonctionnement de l'interface pour l'auditeur

L'interface pour l'auditeur est construite selon le même modèle que l'interface du locuteur : des images sont présentées au centre de l'écran et des boutons de contrôle apparaissent en bas de l'interface (figure 4.11).

FIGURE 4.11 – Interface prévue pour l'auditeur



Préparation de la passation

À partir des consignes orales enregistrées par le locuteur, EloKanz crée la liste des stimuli qui seront joués à l'auditeur. Pour cela il associe à chaque consigne orale les images concernées, ainsi que des distracteurs – c'est-à-dire d'autres images qui ne sont pas concernées par les consignes orales données par le locuteur¹⁴. Les images cibles et les distracteurs sont ensuite positionnés de manière aléatoire¹⁵ sur l'écran.

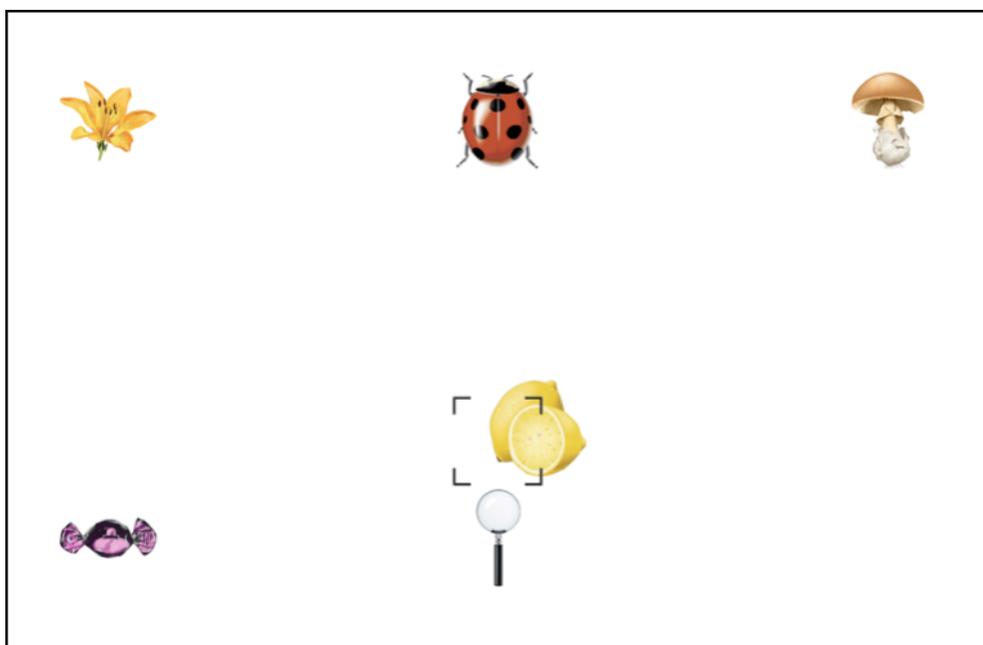
14. ces distracteurs sont sélectionnés au hasard.

15. l'algorithme de positionnement des images évite néanmoins les positions qui correspondent déjà au déplacement donné dans chaque consigne : par exemple si la consigne est de placer le citron à gauche de la loupe, alors le programme fait en sorte de ne pas positionner le citron à gauche de la loupe.

Déroulement d'une session d'écoute

Du côté *auditeur*, la passation se déroule de la manière suivante. Tout d'abord, toute session d'écoute débute par la lecture automatique de la première consigne orale : l'auditeur signale qu'il est prêt en cliquant sur le bouton [OK] dans la boîte de dialogue au centre de l'écran (figure 4.11). Le fichier audio correspondant à la première consigne orale est joué (par ex. « mettez le zèbre sous le marteau »). L'auditeur ne peut pas entendre la consigne à nouveau. Il a le choix entre exécuter un déplacement d'image, ou bien passer directement à la consigne suivante en cliquant sur le bouton [VALIDER] (figure 4.11). Pour effectuer un déplacement l'auditeur place le curseur de la souris au-dessus d'une image. Le curseur prend alors la forme d'une main. L'auditeur peut ensuite déplacer l'image sur un emplacement (gauche, droite, haut et bas) situé à côté de n'importe quelle autre image. Lorsque l'auditeur déplace une image à proximité d'une autre image, des cadres noirs apparaissent pour indiquer les emplacements cible (figure 4.12). L'auditeur peut relâcher l'image au-dessus d'un emplacement cible, auquel cas l'image vient se positionner au centre de cet emplacement. Si l'auditeur n'est pas satisfait de son déplacement, il peut recommencer en cliquant sur le bouton [EFFACER] (figure 4.11). Lorsqu'il le souhaite, l'auditeur peut passer à la consigne suivante en cliquant sur le bouton [VALIDER] (figure 4.11).

FIGURE 4.12 – Exemple de cadre signalant à l'auditeur un emplacement cible. Ici l'image déplacée est le citron, et l'emplacement cible est situé au-dessus de la loupe.



Enregistrement des résultats

EloKanz permet tout d'abord de récupérer l'historique des actions effectuées par l'auditeur, que ces dernières concernent :

- la sélection d’une image (lorsque l’auditeur clique sur une image mais ne la déplace pas) ;
- le déplacement d’une image ;
- l’effacement d’un déplacement ;
- la validation d’un déplacement.

La figure 4.13 présente un exemple de fichier d’historique¹⁶. Les premières lignes indiquent que l’historique correspond aux actions effectuées par l’auditeur « AE09 », qui a passé le test EloKanz le 03/05/2012 à 16h02. Dans chacun des trois paragraphes suivants :

- la première ligne donne le nom du fichier audio qui a été joué à l’auditeur, et les actions qu’il est censé faire pour répondre à la commande orale ;
- la seconde ligne contient les informations relatives à chaque action effectuée par l’auditeur :
 - dans la colonne TIME est écrite l’heure à laquelle chaque action a été effectuée ;
 - dans la colonne ACTION est écrite la nature de chaque action effectuée : « Selection_Pour_Rien_X » (simple sélection de l’image X, sans qu’un déplacement ne soit effectué), « Positionnement_X_Y_Z » (déplacement de l’image X à la position Y par rapport à l’image Z), ou « Validation » (l’auditeur clique sur le bouton « Valider » pour passer à la consigne orale suivante) ;
 - enfin, dans la colonne TIME_MS_DEPUIS_AUDIO est écrite la durée écoulée, en millisecondes, depuis la fin de la diffusion de la consigne orale.

D’autre part, EloKanz enregistre le résultat des actions effectuées par l’auditeur, compte-tenu des consignes orales qui lui ont été jouées. Ces données sont enregistrées dans un fichier similaire à celui présenté dans les figures 4.14 et 4.15. Chaque colonne du fichier contient les informations suivantes :

- codes du locuteur et de l’auditeur ;
- dates et heures des passations du locuteur et de l’auditeur ;
- adresse du fichier son qui a été joué ;
- objet source, position et objet cible correspondant à la consigne donnée par le locuteur (colonnes OBJ1, POS et OBJ2 dans la figure 4.15) ;
- objet source, position et objet cible correspondant au déplacement effectué par l’auditeur (colonnes OBJ1 DONE, POS DONE et OBJ2 DONE dans la figure 4.15) ;
- correspondance (1) ou non-correspondance (0) entre les objets sources, les objets cibles et les positions indiquées par le locuteur, et celles résultant du

16. cet exemple est construit à partir de données réelles, tirées de notre étude de faisabilité (cf. chapitre 7). Par ailleurs, un exemple complet de fichier d’historique est donné en annexe DVD 3a_2.

FIGURE 4.13 – Historique des actions effectuées par l’auditeur

AUDITEUR : #AE09#		
DATE : #3/05/12#		
HEURE : #16:02#		
PN14_2012_02_1_11h42_17.AIF (Mettre Boxeur bas Entonnoir)		
TIME	ACTION	TIME_MS_DEPUIS_AUDIO
16:10	Selection_Pour_Rien_Boxeur	135899
16:10	Selection_Pour_Rien_Boxeur	138756
16:10	Positionnement_Boxeur_bas_Entonnoir	145472
16:10	VALIDATION	151526
PA16_2011_11_22_14h16_18.AIF (Mettre Entonnoir bas Cuillere)		
TIME	ACTION	TIME_MS_DEPUIS_AUDIO
16:11	Positionnement_Entonnoir_bas_Cuillere	7590
16:11	VALIDATION	9652
PN12_2011_11_30_14h35_42.AIF (Mettre Tournevis gauche Boxeur)		
TIME	ACTION	TIME_MS_DEPUIS_AUDIO
16:11	Positionnement_Tournevis_gauche_Boxeur	10578
16:11	VALIDATION	34355
...

déplacement effectué par l’auditeur (colonnes OBJ1 OK, POS OK, OBJ2 OK et RESULTAT dans la figure 4.15) ;

- délai (en millisecondes) entre la fin de la lecture de la consigne, et la première action effectuée par l’auditeur (colonne DELAI FIRST ACTION dans la figure 4.15) ;
- délai (en millisecondes) entre la fin de la lecture de la consigne, et la validation par l’auditeur (colonne DELAI VALIDER dans la figure 4.15) ;
- nombre total d’actions (sélection, déplacement ou effacement) effectuées par l’auditeur (colonne NB ACTIONS dans la figure 4.15) ;
- nombre d’images affichées à l’écran (colonne NB OBJETS dans la figure 4.15) ;
- contexte (distracteurs affichés à l’écran).

FIGURE 4.14 – Exemple de fichier résultat (première partie)

CODE_AUDITEUR	SESSION_AUDITEUR_DAY	SESSION_AUDITEUR_TIME	CODE_LOCUTEUR	SESSION_LOCUTEUR_DAY	SESSION_LOCUTEUR_TIME	FICHIER_SON
AE05	2012_04_30	16h42	PA16	2011_11_22	14h16	PA16_2011_11_22_14h16_12.AIF
AE05	2012_04_30	16h42	PN14	2012_02_1	11h42	PN14_2012_02_1_11h42_7.AIF
AE05	2012_04_30	16h42	PN14	2012_02_1	11h42	PN14_2012_02_1_11h42_17.AIF
AE05	2012_04_30	16h42	PN12	2011_11_30	14h35	PN12_2011_11_30_14h35_36.AIF
AE05	2012_04_30	16h42	PN12	2011_11_30	14h35	PN12_2011_11_30_14h35_42.AIF
AE05	2012_04_30	16h42	PA16	2011_11_22	14h16	PA16_2011_11_22_14h16_19.AIF
AE05	2012_04_30	16h42	PN14	2012_02_1	11h42	PN14_2012_02_1_11h42_11.AIF
AE05	2012_04_30	16h42	PA16	2011_11_22	14h16	PA16_2011_11_22_14h16_23.AIF
AE05	2012_04_30	16h42	PN12	2011_11_30	14h35	PN12_2011_11_30_14h35_44.AIF
AE05	2012_04_30	16h42	PA16	2011_11_22	14h16	PA16_2011_11_22_14h16_18.AIF

En résumé, pour évaluer la compréhension de la parole dans notre travail, nous avons choisi d’utiliser un protocole fondé sur l’observation des actions de l’auditeur en réponse à des énoncés de type impératif. Ce protocole, inspiré de travaux en aphasiologie et en psychologie animale (Savage-Rumbaugh *et al.*, 1993, notamment),

FIGURE 4.15 – Exemple de fichier résultat (seconde partie)

OBJ1	POS	OBJ2	OBJ1 DONE	POS DONE	OBJ2 DONE	OBJ1 OK	POS OK	OBJ2 OK	RESULTAT	DELAI FIRST ACTION	DELAI VALIDER	NB OBJETS	NB ACTIONS	CONTEXTE
Lit	droite	Scie	Lit	droite	Scie	1	1	1	1	7192	94095	6	1	#Cigarette#Tournevis#Bouteille#Tondeuse#
Canne	haut	Brosse	Canne	bas	Brosse	1	0	1	0	7793	12469	6	1	#Lunettes#Telephone#Poire#Cheval#
Boxeur	bas	Entonnoir	Boxeur	bas	Entonnoir	1	1	1	1	7966	14968	6	2	#Fusée#Tracteur#Bonnet#Fleur#
Avocat	gauche	Avion	Avocat	gauche	Avion	1	1	1	1	7120	23804	6	2	#Zebre#Poisson#Champignon#Moulin#
Tournevis	gauche	Boxeur	Tournevis	gauche	Boxeur	1	1	1	1	6846	11872	6	1	#Plume#Cigare#Telephone#Poire#
Marteau	bas	Pipe	Marteau	bas	Pipe	1	1	1	1	5671	10412	6	1	#Poussin#Rateau#Accordeon#Poisson#
Boutons	haut	Boxeur	Boutons	haut	Boxeur	1	1	1	1	7672	10642	6	1	#Skieur#Elephant#Fleur#Fraises#
Fleur	bas	Tente	Fleur	bas	Tente	1	1	1	1	3778	6682	6	1	#Autruche#Poussin#Pipe#Cuillere#
Oignon	gauche	Fourchette	Oignon	droite	Brouette	1	0	0	0	6958	15256	6	2	#Croissant#Quille#Tracteur#Pipe#
Entonnoir	bas	Cuillere	Entonnoir	bas	Cuillere	1	1	1	1	4772	6842	6	1	#Coccinelle#Cerises#Bouteille#Gomme#

répond aux attentes que nous avons définies en fonction de notre conception de la compréhension de la parole (*cf.* section 3.3, p. 78) et de nos hypothèses (*cf.* section 3.4.2, p. 83). Nous avons choisi d’implémenter notre protocole dans un logiciel, que nous avons baptisé « EloKanz » en hommage à Kanzi, le bonobo étudié par Savage-Rumbaugh *et al.* (1993). Ce logiciel implique pour les locuteurs d’assister à des déplacements d’images générés par le logiciel, et d’enregistrer des commandes verbales correspondant à ces déplacements. Le test propose ensuite à l’auditeur de répondre aux commandes en déplaçant les images à l’aide d’une souris. Enfin, pour évaluer la compréhension de l’auditeur, le logiciel observe s’il y a correspondance entre les déplacements d’images que le locuteur a dû verbaliser sous la forme de commandes, et les actions effectuées par l’auditeur en retour.

Deuxième partie

Expériences

Chapitre 5

Études préliminaires

Sommaire

5.1	Sélection du jeu d'images et études de l'accord sur le nom	112
5.1.1	Choix de la batterie d'images	112
	Nos besoins	112
	Les batteries d'images normées existantes	113
	Le jeu d'images du dictionnaire <i>Le Visuel 4</i> ®	118
	Conclusion	120
5.1.2	Test de dénomination écrite des images	121
	Population	121
	Matériel	121
	Procédure	121
	Résultats et discussion	122
5.1.3	Étude de l'accord sur le nom lors de l'utilisation d'EloKanz	129
	Population	129
	Images	129
	Procédure	130
	Résultats et discussion	130
5.2	Test de l'interface locuteur d'EloKanz	132
	Participants	132
	Matériel	132
	Procédure	133
	Résultats	133
5.3	Conclusion et discussion	139

Il est important de souligner que le logiciel EloKanz implique de la part du locuteur la production de *paroles spontanées* (non lues). Dans ce test le locuteur ne dispose pas d'une liste de mots ou de phrases à lire, mais doit produire librement des énoncés pour faire exécuter à son interlocuteur des déplacements d'images. Or, puisque EloKanz juge la compréhension des énoncés en comparant les images désignées par les locuteurs et celles déplacées par les auditeurs, il est fondamental d'étudier *l'accord qui peut exister entre les sujets à propos des noms correspondant aux différentes images*. En effet, si les locuteurs et les auditeurs ne sont pas d'accord sur les concepts représentés par certaines images, ou bien s'ils utilisent différents noms pour désigner les mêmes images, nous ne pouvons pas affirmer avec certitude que les échecs de communication sont uniquement le fait des performances phonatoires des locuteurs. Il est donc nécessaire d'étudier en amont *l'accord sur le nom* pour chaque image utilisée dans EloKanz. Nous consacrons la première partie de ce chapitre à cette question de la sélection des images. Nous expliquons tout d'abord comment nous avons choisi un jeu d'images existant (section 5.1.1), puis nous détaillons deux études que nous avons faites autour de l'accord sur le nom pour ces images (sections 5.1.2 et 5.1.3, p. 121 et 129).

Ensuite, EloKanz implique pour les locuteurs de suivre des consignes assez complexes. D'une part, il doivent être capables d'appréhender des animations correspondant à des déplacements d'images, de les garder à l'esprit puis de formuler un ordre verbal approprié. D'autre part, EloKanz se présente sous la forme d'un logiciel informatique dont la manipulation peut être problématique et dont le fonctionnement peut être gêné par des comportements que nous n'avions pas prévu lors de la programmation initiale (c.-à-d. des comportements engendrant des « bugs » informatiques). Pour toutes ces raisons nous avons décidé de profiter de notre seconde expérience d'accord sur le nom pour étudier dans quelle mesure l'ergonomie du test (consignes et interface logicielle) peut poser des difficultés aux locuteurs. La seconde partie de ce chapitre (section 5.2, p. 132) est consacrée à cette question.

5.1 Sélection du jeu d'images et études de l'accord sur le nom

5.1.1 Choix de la batterie d'images

Nos besoins

Plusieurs critères vont contraindre le choix des images, de par a) la procédure et b) les objectifs d'évaluation :

- a) durant l'utilisation du test EloKanz le locuteur et l'auditeur doivent remplir des tâches qui leur demandent de *reconnaître les concepts* représentés par les images à l'écran. Or une reconnaissance correcte des concepts suppose :

1. que les images soient d'une taille et d'une résolution¹ suffisantes pour permettre une bonne appréhension visuelle ;
 2. que les concepts représentés par les images soient suffisamment familiers aux interlocuteurs pour pouvoir être reconnus ;
 3. que les images représentent fidèlement les concepts.
- b) EloKanz a pour objectif final d'évaluer les performances du locuteur en termes de production de la parole. Ceci implique de supprimer les obstacles à la compréhension verbale qui ne relèvent pas de la performance phonatoire du locuteur. Un tel obstacle pourrait être un désaccord entre le locuteur et l'auditeur vis-à-vis des noms et des concepts associés aux images. En effet si le locuteur voit un « homard » là où l'auditeur voit un « scorpion », ou encore si le locuteur parle d'un « isard² » pour désigner l'animal que l'auditeur nomme habituellement « chamois³ », l'éventuel échec de la communication ne peut être attribué à la seule qualité de la parole produite par le locuteur. En conséquence, nous devons disposer d'un jeu d'images pour lequel l'*accord sur le nom* est maximal entre les interlocuteurs.

Nous avons recherché des batteries d'images qui puissent répondre à ces différents critères parmi les jeux d'images normées utilisés en psycholinguistique (*cf.* section *Les batteries d'images normées existantes*). Ce matériel n'a pas répondu à toutes nos attentes, notamment en ce qui concerne la résolution des images et l'accord sur le nom. En conséquence nous avons fait le choix d'utiliser des images provenant d'une nouvelle source, un dictionnaire visuel produit au Canada (*cf.* section 5.1.1, p. 118). Les images contenues dans ce dictionnaire n'étant pas normées, nous avons conduit plusieurs études d'accord sur le nom (sections 5.1.2 et 5.1.3, p. 121 et 129).

Les batteries d'images normées existantes

Les images de Snodgrass et Vanderwart (1980)

Parmi les batteries d'images utilisées en psycholinguistique, le travail pionnier de Snodgrass et Vanderwart (1980) a permis de recueillir des données normatives pour 260 images⁴ correspondant à des noms anglais concrets (*ex.* *table, chair, banana*). Ces images se présentent en noir et blanc et leur résolution est assez faible, ce qui donne aux traits et courbes des images un aspect plutôt irrégulier (*cf.* exemples en figure 5.1).

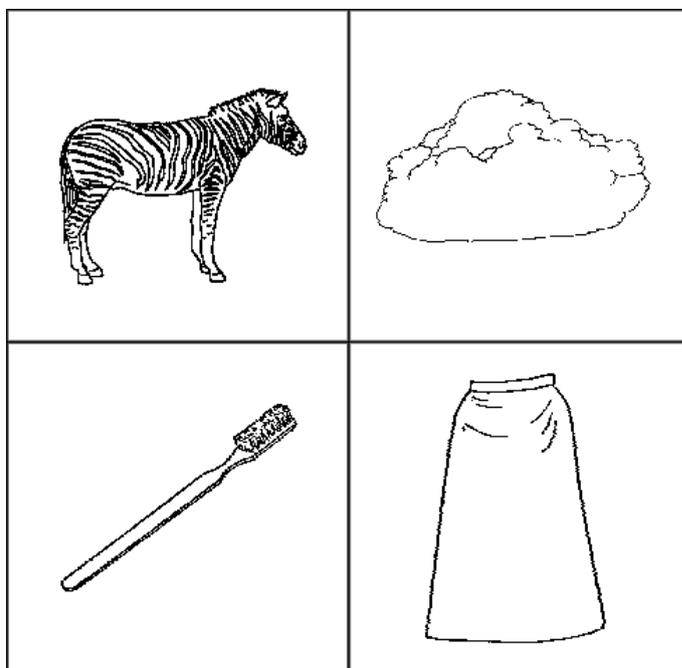
1. la résolution est une mesure, habituellement exprimée en nombre de pixels par pouce (PPP), qui traduit la finesse des détails affichés.

2. nom habituellement donné au chamois dans les Pyrénées.

3. nom habituellement donné à l'isard dans les Alpes.

4. le jeu d'images de Snodgrass et Vanderwart est fourni en annexe DVD 4a. Trois images correspondant à des concepts propres à la culture américaine (batte de baseball, ballon de football américain et casque de football américain) sont absentes de l'annexe.

FIGURE 5.1 – Exemples d’images normées par Snodgrass et Vanderwart (1980). De haut en bas et de gauche à droite : un zèbre (*zebra*), un nuage (*cloud*), une brosse à dents (*toothbrush*), une jupe (*skirt*).



Snodgrass et Vanderwart ont recueilli pour ces images les données d’*accord sur le nom*, d’*accord sur l’image*⁵, de *familiarité*⁶ et de *complexité*⁷. Ces données ont été recueillies auprès d’une population de 219 étudiants anglophones. Pour calculer l’accord sur le nom, les auteurs ont demandé à 42 sujets de nommer par écrit chaque image, et ce de la façon la plus spontanée, précise et concise possible.

Les images d’Alario et Ferrand (1999)

Le travail de Snodgrass et Vanderwart (1980) a ensuite été adapté au français par Alario et Ferrand (1999). Les auteurs ont repris les images de la batterie normée par Snodgrass et Vanderwart, en enlevant du jeu original les images qui représentaient des concepts propres à la culture américaine (ex. batte de baseball, ballon de football américain et casque de football américain). Alario et Ferrand ont ajouté à ce jeu 147 nouvelles images en noir et blanc recueillies par Cychowicz *et al.* (1997) auprès de différentes sources⁸. Pour étudier l’accord sur le nom pour l’ensemble de ces 400

5. l’accord sur l’image était noté par les sujets sur une échelle de 1 à 5. Pour cela ils avaient pour consigne de se représenter mentalement les concepts avant de voir l’image qui les représentaient. Ils jugeaient ensuite la conformité des images avec l’image mentale qu’ils s’en étaient formée.

6. la familiarité avec les concepts représentés était évaluée par les sujets sur une échelle de 1 à 5, selon l’expérience qu’ils en avaient.

7. la complexité était notée par les sujets sur une échelle de 1 à 5. La complexité d’une image était définie comme la proportion de détails, ou d’entrecroisements de lignes qu’elle contenait.

8. ces nouvelles images sont disponibles en annexe DVD 4b. Merci à Ludovic Ferrand de nous

images, Alario et Ferrand ont suivi la même procédure que celle décrite dans l'étude de Snodgrass et Vanderwart (1980). Cette procédure a été effectuée par 28 étudiants en psychologie.

Les images de Rossion et Pourtois (2004)

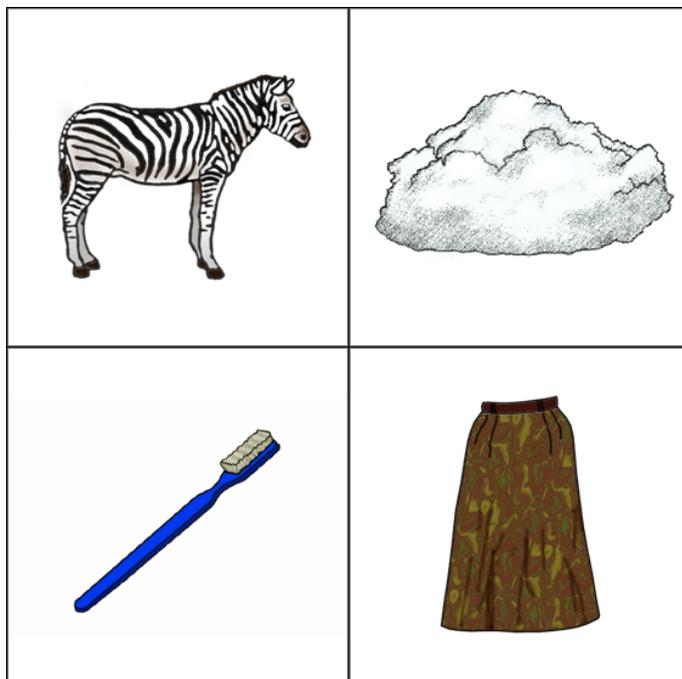
La seconde adaptation du travail de Snodgrass et Vanderwart (1980) pour le français a été conduite par Rossion et Pourtois (2004). Les auteurs ont colorisé l'ensemble des 260 images originales⁹, comme l'illustrent les exemples donnés en figure 5.2. Les nouvelles images ont par ailleurs une résolution supérieure à celle des images de Snodgrass et Vanderwart auxquelles nous avons eu accès, ce qui donne aux images un aspect plus lisse.

Pour étudier l'accord sur le nom des images, Rossion et Pourtois n'ont pas utilisé la même procédure que dans les études de Snodgrass et Vanderwart (1980) et d'Alario et Ferrand (1999). Rossion et Pourtois faisaient en effet l'hypothèse que les images en couleur seraient plus rapidement reconnues que les images en noir et blanc, hypothèse qu'ils ont vérifié en observant les temps de réaction des sujets lors de la dénomination des images. Pour ce faire les auteurs ont demandé à 60 étudiants en psychologie de nommer le plus rapidement possible, à l'oral, chaque image. Les sujets avaient également pour consigne de nommer les objets de la façon la plus brève et la moins ambiguë possible, et ce en utilisant un seul nom.

les avoir transmises.

9. l'ensemble de ces 260 images est disponible en annexe DVD 4c. Merci à Bruno Rossion de nous les avoir transmises.

FIGURE 5.2 – Exemples d’images normées par Rossion et Pourtois (2004). De haut en bas et de gauche à droite : un zèbre, un nuage, une brosse à dents, une jupe.



Les images de Bonin et al. (2003)

Le travail de Bonin *et al.* (2003) a consisté à recueillir des données normatives pour 299 nouvelles images. L’objectif des auteurs était d’augmenter le nombre de jeux d’images normées disponibles pour les études psycholinguistiques. Les images de Bonin *et al.*¹⁰ proviennent de différentes sources. De ce fait, on peut remarquer une certaine hétérogénéité quant à la qualité des illustrations : certaines images ont un aspect net et lisse, alors que d’autres présentent des traits plutôt irréguliers (*cf.* par exemple la différence entre l’*ange* et l’*infirmière* dans la figure 5.3).

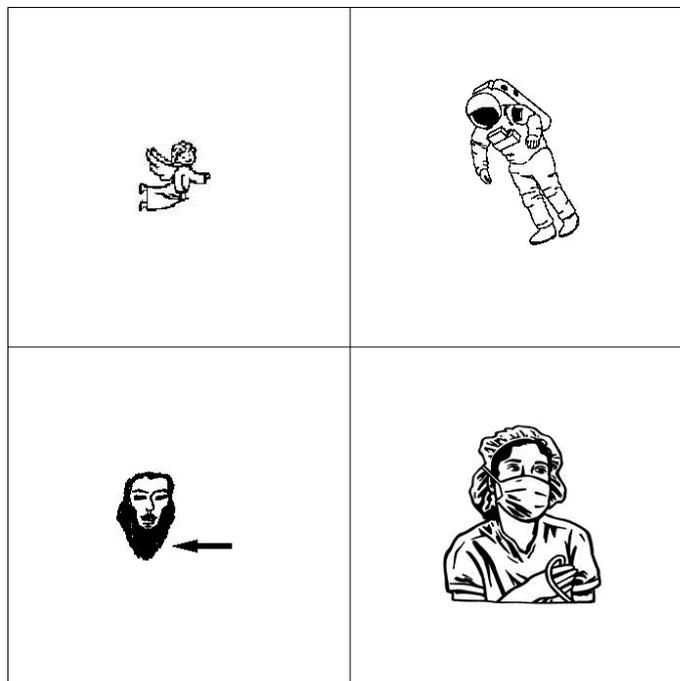
Pour collecter les données d’accord sur le nom, Bonin *et al.* ont fait appel à 30 étudiants en psychologie. La tâche de dénomination suivait la même procédure que celle utilisée par Alario et Ferrand (1999) : pour chaque image les étudiants avaient pour consigne d’écrire le premier nom (d’un ou plusieurs mots) qui leur venait à l’esprit.

Conclusion

En conclusion, parmi tous ces jeux d’images normées nous n’avons pas trouvé de matériel qui corresponde totalement à nos besoins. Tout d’abord, au début de l’élaboration de nos expériences nous pensions diffuser l’interface d’EloKanz sur un

10. ces images sont disponibles en annexe DVD 4d.

FIGURE 5.3 – Exemples d'images normées par Bonin *et al.* (2003). De haut en bas et de gauche à droite : un ange, un astronaute, une barbe, une infirmière.



écran de grande taille, et éventuellement tactile. Or dans la plupart des jeux d'illustrations cités la qualité des images ne nous a pas semblé suffisante pour permettre une diffusion sur un support de taille importante. Agrandir les images aurait provoqué l'apparition de phénomènes de *pixellisation* (crénelage) donnant des contours irréguliers aux images. Nous pensons en particulier aux images d'Alario et Ferrand (1999) et de Bonin *et al.* (2003).

En outre, la plupart des jeux d'images que nous avons cités sont en noir et blanc, alors qu'il semble que la couleur garantisse un meilleur accord sur le nom (Rossion et Pourtois, 2004). Un de nos objectifs étant de disposer d'un jeu d'images pour lequel l'accord sur le nom est maximal, nous avons eu tendance à rejeter les jeux d'images non colorisées (c.-à-d. les images de Alario et Ferrand, 1999; Bonin *et al.*, 2003; Snodgrass et Vanderwart, 1980).

Parmi l'ensemble des jeux d'images que nous avons décrits dans cette partie, celui de Rossion et Pourtois (2004) représentait donc, d'après nous, la solution la mieux adaptée à nos contraintes. Les seuls éléments qui nous ont retenu dans ce choix étaient :

- a) la façon dont Rossion et Pourtois (2004) ont collecté les données d'accord sur le nom. Les auteurs ont demandé aux sujets de dénommer les images le plus rapidement possible, et ce à l'aide d'un seul nom. En contraste, l'utilisation d'EloKanz implique que le locuteur *ait le temps de réfléchir à la formulation de son message linguistique* (cf. la description de l'interface locuteur, section 4.2.2 p. 98). Nous

avons pensé que cette différence concernant les conditions de la production linguistique pourrait influencer sur les dénominations attribuées aux images – et que, par conséquent, nous ne pourrions pas nous fier aux normes d'accord sur le nom collectées par Rossion et Pourtois (2004) pour prévoir l'accord sur le nom que nous obtiendrons lors de l'utilisation d'un test comme EloKanz ;

- b) la qualité des images, qui nous semblait toujours insuffisante pour une diffusion sur un support de grande taille. Nous avons pensé que nous pourrions peut-être trouver des images d'une résolution supérieure en dehors des batteries publiées dans la littérature psycholinguistique.

En conséquence, nous avons fait le choix de sélectionner nous-même d'autres images, pour lesquelles nous collecterions nos propres données d'accord sur le nom. Après nous être intéressé à diverses sources d'images (par ex. *L'imagier du Père Castor*^{© 11} et des batteries de photos ou d'images utilisées en orthophonie), notre choix s'est porté sur les illustrations contenues dans CD-ROM *Le Visuel 4*^{© 12}, un dictionnaire illustré produit par la société *QA International*.

Le jeu d'images du dictionnaire *Le Visuel 4*[©]

Le dictionnaire *Le Visuel 4*[©] contient 6000 illustrations en couleur, organisées par thèmes (*alimentation et cuisine, règne animal, bricolage et jardinage*, etc.). Les illustrations proposées nous ont paru particulièrement réalistes ; elles sont de plus de bonne résolution et forment un ensemble assez homogène sur le plan esthétique (*cf.* les exemples donnés en figure 5.4).

Les images que nous avons présentées dans la section précédente, c'est-à-dire principalement les images de Snodgrass et Vanderwart (1980) – qui ont été reprises ou complétées par Alario et Ferrand (1999), Bonin *et al.* (2003) et Rossion et Pourtois (2004), représentent des concepts appartenant pour l'essentiel aux catégories suivantes : *animaux à quatre pattes, ustensiles de cuisine, mobilier, corps humain, fruits, armes, outils de charpentier, vêtements, maison, instruments de musique, oiseaux, véhicules, jouets, légumes et insectes*.

Selon Snodgrass et Vanderwart (1980) ce sont dans ces différentes catégories que l'on trouve le plus d'exemplaires susceptibles d'être représentés de manière non ambiguë par une image (Snodgrass et Vanderwart appellent ce critère « *imageability* »). Pour cette raison, nous avons été particulièrement intéressé par les illustrations du dictionnaire visuel qui s'inscrivent dans des thématiques similaires, comme¹³ :

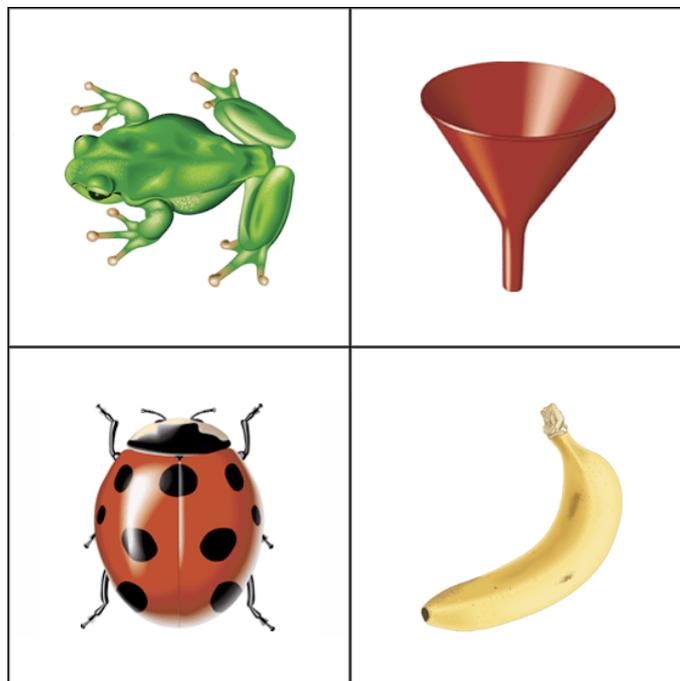
- le règne végétal (plantes, fleurs...);
- le règne animal ;

11. publié par Flammarion.

12. <http://www.ledictionnairevisuel.com> – nous remercions la société *Q.A. International* de nous avoir permis l'utilisation des illustrations contenues dans ce dictionnaire dans le cadre de nos recherches.

13. toutes les thématiques citées correspondent à différentes catégories du CD-ROM *Le Visuel 4*[©].

FIGURE 5.4 – Exemples d'images du Dictionnaire Le Visuel 4[©]. De haut en bas et de gauche à droite : une grenouille, un entonnoir, une coccinelle, une banane.



- l'alimentation et la cuisine (fruits, légumes, ustensiles...);
- la maison (meubles, éléments de la maison...);
- les vêtements;
- la parure et les objets personnels (bijoux, bagages...);
- le bricolage et le jardinage.

Parmi ces différentes catégories, nous avons sélectionné 212 images qui nous ont semblé *a priori* pouvoir faire l'objet d'un consensus de dénomination linguistique. Pour ce faire, nous avons utilisé les mêmes critères de sélection que Snodgrass et Vanderwart (1980), et qui visent à vérifier :

- que le concept visé est facile à illustrer de manière non ambiguë, ce qui exclut notamment les noms de relation (*père, sœur, ...*) ou les noms d'objets indénombrables (*eau, argent, ...*);
- que le dessin représente un exemplaire suffisamment prototypique du concept visé (l'*autruche* serait par exemple un exemplaire peu prototypique du concept d'*oiseau*);
- que le dessin contient suffisamment de détails (traits et couleurs), ou est de taille suffisante pour permettre une bonne reconnaissance du concept représenté;
- que le concept visé soit familier au plus grand nombre de personnes possible (nous pouvons supposer par exemple que peu de gens connaissent la *mu-*

saraigine pygmée de Thompson – microsorex thompsoni, en comparaison de l'animal *chat*).

Conclusion

En résumé, nous n'avons pas trouvé de batterie d'images normées qui permette de satisfaire pleinement nos critères, que ceux-ci concernent la qualité des images (résolution) et/ou les normes recueillies. Ceci nous a conduit à choisir une autre source d'illustrations, le dictionnaire *Le Visuel 4*©. Parmi les illustrations proposées par cet ouvrage, nous avons effectué une première sélection d'images (212 au total) en fonction du caractère *a priori* non ambigu des images, et de la familiarité des concepts représentés.

Dans un second temps, nous avons mené plusieurs études afin de vérifier le caractère non ambigu des images, et pour recueillir les normes d'accord sur le nom – toujours avec l'objectif de recueillir *in fine* un ensemble d'illustrations pour lequel l'accord sur le nom serait maximal. Nous relatons ces différentes études dans la section suivante.

5.1.2 Test de dénomination écrite des images

Comme nous l'avons souligné dans la section précédente, l'utilisation de l'interface *locuteur* d'EloKanz n'impose pas de contraintes de temps aux sujets. En conséquence, pour étudier l'accord sur le nom pour notre jeu d'images nous avons décidé d'employer une procédure qui laisse aux sujets le temps de dénommer les images. Une procédure de dénomination écrite, du type de celles utilisées par Alario et Ferrand (1999) et Bonin *et al.* (2003), nous est ainsi apparue dans un premier temps comme une solution adaptée à nos objectifs.

Population

Notre population était constituée de 26 sujets : 10 hommes et 16 femmes âgés de 18 à 83 ans, tous de langue et de culture françaises. Les sujets exerçaient différentes professions dans des domaines aussi divers que l'aéronautique, l'agriculture, le bâtiment, la robinetterie industrielle, la recherche spatiale, le journalisme, les mathématiques, la médecine, la psychologie, la philosophie ou encore la police judiciaire. Toutes les informations sur les sujets ayant participé à l'étude sont fournies dans le tableau donné en annexe A.1 (p. 277).

Matériel

Nous avons utilisé les 212 illustrations en couleur que nous avons sélectionnées dans le CD-ROM du dictionnaire *Le Visuel 4*[©] (cf. section 5.1.1, p. 118). Ces illustrations représentent des objets manufacturés, des bâtiments, des animaux et des végétaux. Les images présentaient une taille (en hauteur ou en largeur selon la forme de l'objet) de 203 pixels. L'ensemble de ces 212 images est fourni en annexe DVD 4e; le nom de chaque fichier correspond au nom que nous attendions dans cette tâche de dénomination.

Procédure

Pour diffuser ces images, nous avons créé un site Internet dynamique utilisant les langages PHP, HTML et JavaScript, et reposant sur une base de données MySQL. La base de données permettait de recueillir à la fois les informations relatives aux sujets (âge, profession...) ainsi que les noms saisis par les sujets pour désigner les images. Nous avons inclus une version statique¹⁴ (HTML et JavaScript uniquement) de notre site Internet en annexe DVD 5a. Des copies d'écran correspondant à l'interface du site Internet sont également disponibles en annexe A.2 (p. 279).

Les images étaient diffusées dans un ordre aléatoire différent pour chacun des 26 sujets. Pour chaque image, les sujets étaient invités à écrire dans un champ de texte le nom qui, selon eux, correspondait le mieux au concept représenté. Aucune contrainte de temps ne leur était imposée. Par ailleurs, les sujets étaient avertis

14. la version dynamique de notre site web doit être visionnée à partir d'un serveur informatique, ce qui explique notre choix d'inclure une simple version HTML dans l'annexe DVD.

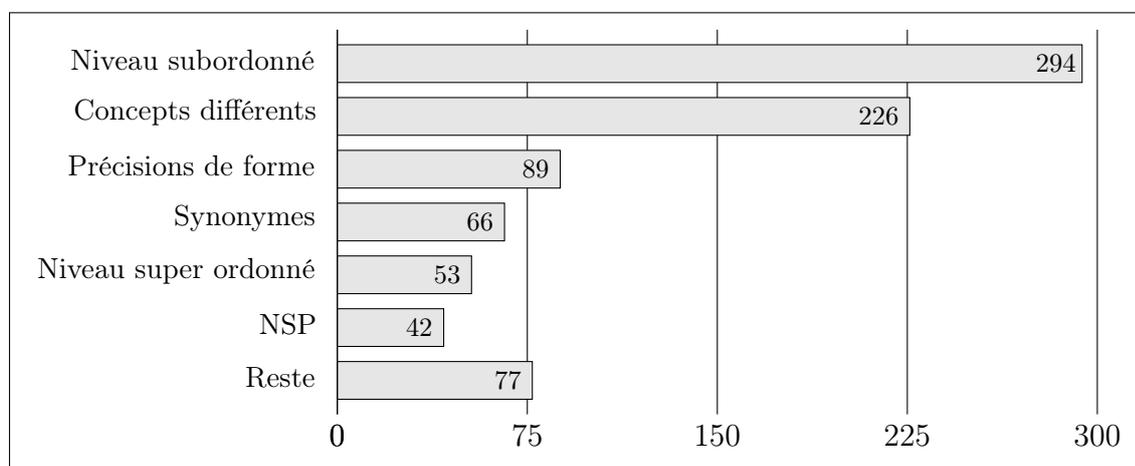
qu'il ne s'agissait pas d'une évaluation de compétences particulières (orthographe notamment), et étaient invités à dénommer les images de façon *simple et directe*. En cas d'impossibilité (objet ou mot non connu, mot « sur le bout de la langue »), les sujets avaient pour consigne de répondre « NSP ». Enfin, nous avons indiqué aux sujets qu'il était possible de nommer les images en utilisant plusieurs mots.

Résultats et discussion

Contrairement à ce que nous avons supposé lors de la sélection des images, nous avons trouvé un accord sur le nom maximal (100%) pour un jeu d'images plutôt réduit : 39 images, soit environ 18% des illustrations seulement. Ainsi, sur 5512 dénominations 873 ne correspondaient pas aux noms attendus¹⁵ (soit environ 16%).

En annexe A.3 (p. 281) un tableau donne, pour chaque image, le nom le plus fréquemment attribué ainsi que le pourcentage d'accord sur le nom. Le graphique en figure 5.5 présente les différents types de désaccords sur le nom rencontrés dans notre étude. Plusieurs de ces catégories font référence aux travaux de Rosch *et al.* (1976) qui ont introduit l'idée selon laquelle les êtres humains catégorisent les objets à différents niveaux : *super ordonné* (ex. /animal/), *de base* (ex. /chien/) et *subordonné* (ex. /dogue allemand/).

FIGURE 5.5 – Types de désaccord sur le nom des images



Les différents types de dénominations alternatives sont expliqués dans la liste ci-dessous :

15. ou bien ne correspondaient pas au nom le plus fréquemment attribué pour l'image. Dans le cas de l'image *abeille.gif* par exemple, le nom le plus fréquemment attribué était « guêpe » – ce qui résulte par ailleurs d'une confusion entre les deux insectes de notre part. Les cas pour lesquels le nom attendu ne correspondait pas au nom le plus fréquent apparaissent en gras dans le tableau donné en annexe A.3 (p. 281).

- Niveau subordonné : le sujet utilise un mot/concept plus spécifique que celui attendu, par exemple *Ariane* pour *fusée*, *ours brun* pour *ours*, *brique de 5* pour *brique*¹⁶ ;
- Concepts différents : le sujet se réfère à un concept autre que celui qui est illustré, par exemple *myrtilles* pour *framboises*, *fer à cheval* pour *aimant*, *cendrier* pour *piscine*¹⁷ ;
- Précisions de forme : le sujet précise non pas le concept, mais la « forme » du référent, en spécifiant par exemple le nombre (*deux huîtres*), la couleur (*chemise bleue*) ou la taille (*petit moulin*) ;
- Synonymes : variantes lexicales, par exemple *piaf* pour *oiseau*, *pointe* pour *clou*, *cagette* pour *cageot*¹⁸ ;
- Niveau super ordonné : le sujet utilise un mot/concept moins spécifique que celui attendu, par exemple *sucrerie* pour *sucette*, *poisson* pour *dauphin*, *ouverture* pour *fenêtre*, *rongeur* pour *souris*¹⁹ ;
- NSP : le sujet est dans l'impossibilité de dénommer l'image (mot/concept inconnu, mot « sur le bout de la langue ») ;
- Reste : cette catégorie regroupe les dénominations alternatives renvoyant à plusieurs types de désaccord²⁰, ainsi que les types de désaccord peu fréquents²¹.

Comme l'illustre la figure 5.5, l'emploi de mots/concepts subordonnés représente le type de dénominations alternatives le plus fréquent dans notre étude. Lorsque l'on observe les mots/concepts subordonnés employés par les sujets (cf. annexe A.7, p. 287), il semble que plusieurs aspects peuvent expliquer ce type de productions. Tout d'abord, il est possible que des sujets aient volontairement employé des termes plus précis pour lever d'éventuelles ambiguïtés. En effet certains pourraient être ambigus en l'absence de contexte. C'est le cas par exemple d'homonymes comme *ampoule* (électrique) et *ampoule* (bulle cutanée), *ancres* (ancres marines) et *encre*²², *fouet* (arme) et *fouet* (pâtissier), *grue* (oiseau) et *grue* (de chantier), ou encore *ballon* (dirigeable) et *ballon* (de jeu). De nombreux mots ont été ainsi agrémentés d'adjectifs

16. la liste des dénominations correspondant à des concepts subordonnés est donnée en annexe A.7 (p. 287).

17. la liste des dénominations désignant des concepts différents de ceux censés être représentés par les images est fournie en annexe A.4 (p. 284).

18. la liste des dénominations entretenant une relation synonymique avec les noms attendus est fournie en annexe A.5 (p. 286).

19. la liste des dénominations correspondant à des concepts super ordonnés est donnée en annexe A.6 (p. 286).

20. cela est par exemple le cas lorsqu'un sujet indique « carton » pour « cible ». Cette dénomination est difficile à classer : tout d'abord, le carton est la matière de la cible. Ensuite, « carton » est aussi un synonyme de « cible » pour les adeptes du tir (d'où l'expression « faire un carton »). Enfin, un « carton » peut également être envisagé comme un concept super ordonné (un hyperonyme) de cible, si l'on considère qu'un carton de cible est un type de carton.

21. par exemple, les dénominations traduisant une relation de partie à tout (méronymie) avec le concept représenté par l'image (« pneu » pour « roue », « abat-jour » pour « lampe ») représente une catégorie de dénominations peu fréquentes.

22. parmi les trois sujets ayant précisé « marine », deux ont orthographié le mot *ancres* avec un « e » (« encre marine ») – la graphie du mot ne leur suffit donc pas à lever l'ambiguïté.

ou de syntagmes prépositionnels²³ permettant de lever une éventuelle ambiguïté.

Ensuite, il nous semble que l'expertise des sujets – au sens large – a pu jouer un rôle dans les dénominations attribuées aux images, et ce même si nous avons explicitement demandé aux sujets de dénommer les images de manière simple et directe. Un premier exemple est celui de l'expertise développée par les sujets dans le cadre de leur profession. Certaines images étaient en relation avec les domaines professionnels des sujets, et il semble que cela ait pu amener les sujets à dénommer parfois les concepts de manière plus précise. Par exemple :

- le sujet 1, qui travaillait alors dans le domaine de l'aéronautique, a utilisé le terme *A380* pour désigner l'image « Avion » ;
- le sujet 3, qui travaillait pour le Centre National d'Études Spatiales (CNES), a utilisé le terme *Ariane* pour désigner l'image « fusée » ;
- le sujet 13, apprenti maçon, a utilisé les termes *brique de 5* et *brouette 120 litres* pour désigner les images « brique » et « brouette » ;
- le sujet 16, entrepreneur en bâtiment, a utilisé les termes *ampoule à vis* et *scie égoïne* pour désigner les images « ampoule » et « scie » ;
- le sujet 18, employée de maison, a utilisé les termes de *chinois* et *crêpière* pour désigner les images « passoire » et « poêle » ;
- le sujet 25, technicien en robinetterie industrielle, a utilisé le terme *palette européenne*²⁴ pour désigner l'image « palette » ;
- le sujet 26, lieutenant de police, a utilisé le terme *berline*²⁵ pour désigner l'image « voiture ».

Un phénomène similaire pourrait concerner l'expertise que les sujets développent dans le cadre de leurs loisirs. Ainsi :

- le sujet 4, amateur de chapeaux, a utilisé le terme *borsalino* pour désigner l'image « chapeau » ;
- le sujet 7, amateur de bicross, a utilisé le terme *BMX* pour désigner l'image « vélo » ;
- le sujet 17, amateur de violon, a utilisé le terme *violon alto* pour désigner l'image « violon » ;
- le sujet 25, passionné de motos, a utilisé le terme *BMW* pour désigner l'image « moto ».

Enfin, l'environnement géographique dans lequel évoluent habituellement les sujets pourrait aussi influencer le niveau de précision utilisé pour dénommer les images. Ainsi le sujet 5, habitant à la frontière espagnole parle de « melon français » pour

23. par ex. brosse à ongles, cartouches de fusil, casque de moto, feuille d'arbre, moulin à vent, tente de camping...

24. le sujet nous a indiqué qu'il recevait des palettes tous les jours dans le cadre de son travail. Selon ses explications, les palettes européennes n'ont pas les mêmes dimensions que les palettes provenant d'autres parties du monde.

25. dans la police il est en effet courant de classer les véhicules en fonction de leur type de carrosserie (*berline, break, coupé...*).

désigner l'image que les autres sujets appellent simplement « melon ». De même le sujet 15, ayant effectué de nombreux séjours en Thaïlande, appelle un « piment thaï » ce que les autres sujets dénomment de manière plus générale « piment ».

Les autres types de désaccord sur le nom sont peut-être moins complexes à analyser. Dans le cas des erreurs de types « Concepts différents » et « Concepts super ordonnés », il semble que les sujets n'ont pas reconnu correctement le concept représenté par l'image, et qu'ils ont :

- soit fait référence à un autre concept (ex. *baleine* pour *dauphin*);
- soit choisi de se placer à un niveau de description moins élevé (ex. *insecte* pour *mouche*, *ustensile* pour *louche*).

Enfin la catégorie « Synonymes » recense simplement les variantes lexicales (ex. *pointe* pour *clou*, *TV* pour *télé*) que les sujets ont utilisées.

En résumé, sur le plan quantitatif nos résultats ne sont pas satisfaisants puisque seules 39 images obtiennent un accord sur le nom de 100%. Si nous devions utiliser ces images dans notre test de compréhension de la parole, le champ des référents possibles serait donc très limité. Cependant, en analysant les sources de désaccord sur le nom nous nous sommes aperçu que la majorité des dénominations « alternatives » montraient que les sujets avaient tendance à se référer à des concepts plus précis que ceux que nous visions (catégories « Niveaux subordonnés » et « Précisions de forme »). En analysant les productions des sujets, nous avons émis l'hypothèse selon laquelle ces précisions étaient apportées principalement du fait :

- de l'absence de contexte, qui contraint les sujets à préciser les concepts pour éviter les ambiguïtés;
- de l'expertise des sujets (liée à leur profession, à leurs loisirs, etc.), qui les amènent à utiliser parfois des concepts plus précis que ceux attendus.

Or, nous nous sommes demandé s'il n'était pas possible de contrôler ces effets. Par exemple, dans EloKanz nous avons prévu que les sujets (locuteurs et auditeurs) partagent un même contexte, représenté par un même jeu d'images (cf. section 4.2.1, p. 96). Nous pourrions donc contrôler le contexte pour qu'aucune ambiguïté (par ex. *feuille d'arbre* versus *feuille de papier*) ne soit présente.

Notre seconde hypothèse, celle de l'expertise, s'appuie sur l'idée selon laquelle « les experts utilisent comme catégories de base celles qui correspondent au niveau subordonné pour les novices²⁶ ». Cependant, nous pensons que ce phénomène pourrait dépendre de la situation de communication : on peut notamment se demander ce qu'il advient lorsque des sujets « experts » se trouvent en face d'une personne qui ne partage pas leur expertise. Ainsi dans une étude sur la communication référentielle, Fussell et Krauss (1992) ont observé que les locuteurs prennent en compte le degré d'expertise des personnes auxquelles ils s'adressent, et que cela se traduit

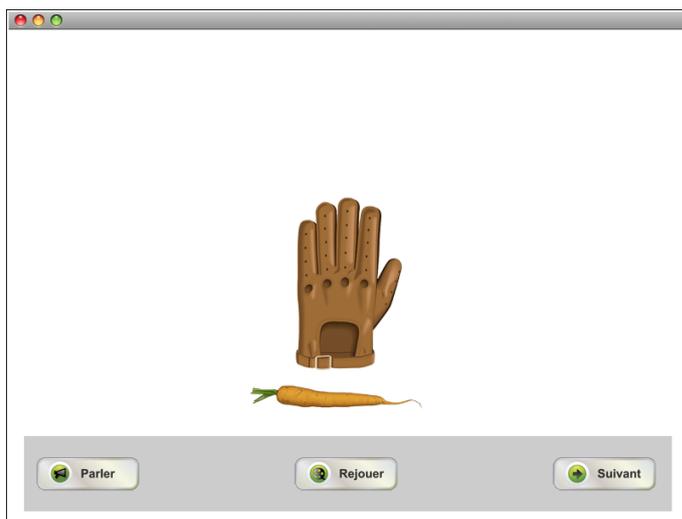
26. « experts use as their basic-level categories what are the subordinate levels for novices » (Rota et Zellner, 2007, p. 179).

dans la relative spécificité du lexique qu'ils emploient. Or, dans notre test de compréhension, nous prévoyons de faire en sorte que les interlocuteurs *ne se voient pas* (cf. section 4.2.1, p. 96). Le locuteur ne connaît donc pas l'identité de la personne à qui il s'adresse, et ne peut donc pas faire d'hypothèse à propos des connaissances et du code linguistique qu'il partage avec ce dernier. Il serait légitime de faire l'hypothèse que, pour optimiser les chances d'être compris par leur interlocuteur, les locuteurs démontrent une tendance à délaissier des mots/concepts trop spécifiques pour se placer au niveau le plus basique.

Pour obtenir des indications quant à la véracité de nos deux hypothèses, nous avons testé à nouveau une petite partie de notre population (5 sujets). Cette nouvelle expérience a eu lieu entre 3 et 4 semaines après le test de dénomination écrite.

Sous l'environnement logiciel *Adobe Flash CS4*® nous avons programmé l'interface présentée en figure 5.6.

FIGURE 5.6 – Interface utilisée pour la pré-étude d'accord sur le nom en situation de communication



Durant le test, les sujets visionnaient des animations faisant intervenir deux images dont l'une, fixe, était centrée sur l'écran. Une image mobile partait depuis la droite de l'écran (hors-champ) puis venait se placer indifféremment à droite, à gauche, au-dessus ou en dessous de la seconde image. L'ensemble de ces animations est disponible en annexe DVD 5b.

Les images de ces animations correspondent à celles pour lesquelles chacun des 5 sujets avait attribué un mot/concept subordonné dans le test de dénomination écrite, auxquelles nous avons ajouté des distracteurs²⁷.

27. les distracteurs n'ont pas fait l'objet d'une sélection particulière.

Les sujets devaient suivre des consignes très proches de celles utilisées pour le test EloKanz²⁸. Pour chaque animation, les sujets devaient faire exécuter à un interlocuteur distant (et imaginaire, *cf. infra*) les actions visionnées à l'écran. Pour ce faire, ils étaient invités à presser un bouton, puis à donner une consigne orale à leur interlocuteur par le biais du microphone de l'ordinateur (ex. « Mettez la carotte sous le gant »). Aucune contrainte de temps n'était imposée : les sujets donnaient leurs consignes quand ils se sentaient prêts, et pouvaient visionner les animations autant de fois qu'ils le souhaitaient. Pour habituer les sujets à cette procédure, chacun a participé à une petite phase d'entraînement portant sur 3 animations. L'ensemble de la consigne que nous avons utilisée est disponible en annexe B.1 (p. 291).

Par ailleurs, nous souhaitons préciser qu'il n'y avait en réalité *personne* qui réalisait les actions que les locuteurs commandaient. Pour nous, il ne s'agissait ici que de placer les sujets dans une situation similaire au test EloKanz, et de récupérer le lexique qu'ils employaient pour dénommer les images. Pour que les sujets ne puissent faire aucune hypothèse concernant les connaissances des auditeurs, il ne leur était donné aucune indication quant aux personnes auxquelles ils s'adressaient.

Nos résultats montrent que, sur 62 images dénommées par des noms/concepts subordonnés dans le test écrit, 39 sont dénommées par des noms/concepts de base dans ce nouveau test, soit 63%. L'ensemble des dénominations attribuées dans les deux tests est fourni en annexe B.3 (p. 293). La tendance des sujets à utiliser des mots/concepts de base plutôt que de niveau subordonné dans cette expérience est donc relativement marquée (cela concerne environ 2 concepts sur 3). À titre d'illustration, quelques exemples sont présentés dans la table 5.1.

TABLE 5.1 – Noms attribués aux images lors de la tâche de dénomination écrite et lors de la tâche de communication (exemples)

Tâche de dénomination écrite	Tâche de communication
Vespa	Scooter
Rainette	Grenouille
Truite	Poisson
Vipère	Serpent
Moineau	Oiseau
Batavia	Salade
Guitare sèche	Guitare
Borsalino	Chapeau
Radis rond	Radis

Malgré la tendance que nous avons soulignée, nos données ne nous permettent pas de confirmer nos deux hypothèses, à savoir :

28. au moment du test nous n'avions pas terminé de développer le logiciel EloKanz. C'est pour cela que nous avons utilisé cette autre interface.

1. que, si les locuteurs ne disposent pas d'informations sur leur interlocuteur, alors ils ont tendance à se référer aux images en se plaçant à un « niveau de spécification de base » ;
2. que les locuteurs ont tendance à ne pas avoir recours à des concepts subordonnés si le contexte qu'ils partagent avec leur interlocuteur ne le justifie pas.

En effet, même si la tendance que nous avons observée est assez forte, plusieurs raisons nous empêchent d'effectuer de telles conclusions. Tout d'abord, notre population est d'un cardinal trop faible pour nous permettre de calculer la significativité des différences observées. Ensuite, le fait que les locuteurs ne savent pas à qui ils s'adressent, et la présence d'un contexte ne sont pas les deux seules différences entre les deux tâches. Par exemple, dans le second test de dénomination les noms sont produits dans des énoncés complets. Or, certaines des dénominations renvoyant à des concepts subordonnés et obtenues dans notre premier test de dénomination étaient de nature syntaxique plus complexe qu'un simple nom (généralement un nom suivi d'un adjectif ou d'un syntagme prépositionnel – ex. *poêle à crêpes*). Peut-être que le fait que ces dénominations doivent s'insérer dans des structures plus larges a incité les sujets à recourir à de simples noms pour éviter la formulation d'énoncés trop complexes (comme *Mettez la tondeuse à gazon au-dessus du cheval de course*, par exemple). Ensuite, le premier test de dénomination était de nature écrite, alors que le second se déroulait à l'oral. Or l'écrit est une modalité qui a une connotation bien plus normative, et donc davantage liée à la notion d'*erreur* que ne l'est la modalité orale. En assimilant la précision des réponses à leur correction, il est possible que les sujets aient eu davantage tendance à vouloir produire des réponses précises dans le test écrit. Enfin, même si nos deux hypothèses étaient vérifiées les procédures de test que nous avons utilisées ne nous permettraient pas de distinguer entre les effets provoqués par le fait que les locuteurs ne connaissaient pas l'identité de leur interlocuteur et ceux provoqués par la présence d'un contexte.

Néanmoins, les résultats que nous avons obtenus dans cette expérience nous ont conforté dans l'idée que les données d'accord sur le nom dépendent grandement de la tâche de dénomination utilisée. C'est pourquoi nous avons décidé de compléter nos données en demandant à 12 nouveaux locuteurs de dénommer les images *directement pendant l'utilisation de l'interface d'EloKanz*. Nous décrivons ce test dans la section suivante.

5.1.3 Étude de l'accord sur le nom lors de l'utilisation d'EloKanz

Nos premiers tests de dénomination nous ont montré que l'accord sur le nom pouvait dépendre pour une grande partie de la tâche de dénomination utilisée. C'est pourquoi, pour compléter nos données d'accord sur le nom pour les images, nous avons demandé à de nouveaux sujets de dénommer nos images directement lors de l'utilisation de l'interface *locuteur* d'EloKanz.

Population

12 sujets, âgés de 19 à 27 ans, ont participé à ce test. Nos critères d'inclusion pour le test étaient que les sujets soient de langue maternelle française, et ne souffrent pas de problèmes d'élocution particuliers. Nous avons également demandé aux participants si ils avaient vécu dans un milieu plutôt rural ou urbain, et quels étaient leurs domaines d'expertise – que ces domaines soient liés à leurs études, leur profession et/ou leur hobbies. Ces questions avaient pour but d'expliquer éventuellement certaines difficultés par rapport aux images utilisées, et/ou d'expliquer des dénominations particulières (*cf.* section précédente). Toutes les informations relatives aux participants sont données dans le tableau en annexe C.1 (295).

Images

Parmi l'ensemble des 212 images utilisées dans notre premier test de dénomination (*cf.* section précédente), nous avons sélectionné les images pour lesquelles seuls des désaccords de type « concepts subordonnés » ou « précisions de forme » intervenaient. En effet, nous n'avons pas jugé utile d'inclure les images ayant donné lieu à d'autres types de désaccords sur le nom – par exemple, si le concept représenté par une image n'a pas été reconnu lors de la première expérience, il n'y a pas de raison que cela ne soit plus le cas lors de l'utilisation d'EloKanz. Ainsi nous n'avons pas conservé les images pour lesquelles, lors de la tâche de dénomination écrite :

- avaient été attribués des concepts différents de ceux illustrés par les images (*cf.* annexe A.4, p. 284) ;
- avaient été employés des synonymes (autres manières de dénommer le même concept, *cf.* annexe A.5, p. 286) ;
- les sujets avaient employé des noms/concepts super ordonnés (annexe A.6, p. 286) ;
- certains sujets n'étaient pas parvenus à attribuer un concept (réponse « NSP ») : *rasoir, poubelle, épée, hachoir, palette, toucan, phare, moules, truelle, canon, huîtres, sucette, microphone, fouet, casque.*

Il résulte de cette sélection un jeu de 115 images, que nous avons inclus en annexe DVD 4f.

Procédure

Dans ce test les sujets étaient placés, individuellement, dans une cabine audiométrique de la plateforme PETRA (*cf.* annexe D.3, p. 302²⁹). Ils étaient assis face à un moniteur d'une taille de 20 pouces. Sur cet écran était diffusée l'interface locuteur d'EloKanz (*cf.* section 4.2.2, p. 98). En outre, les sujets étaient équipés d'un micro-casque de type SHURE BETA 53, et disposaient d'une souris pour interagir avec le programme EloKanz.

La consigne donnée aux sujets est retranscrite en annexe B.1 (p. 291). Les sujets participaient d'abord à une phase d'entraînement durant laquelle ils enregistraient une dizaine de phrases. Ensuite, ils participaient à trois sessions consécutives, chacune consistant en l'enregistrement de 20 phrases.

Résultats et discussion

110 images sur 115 ont été dénommées à 100% par le mot attendu, soit 96,7% des images. Seules 5 illustrations ont fait l'objet de dénominations alternatives, pouvant être classées dans les trois types suivants :

- Concepts différents (*moto* pour *scooter*, *parasol* pour *parapluie*) ;
- Synonyme (*culotte* pour *slip*) ;
- Concepts subordonnés (*tondeuse à gazon* pour *tondeuse* et *tasse à café* pour *tasse*).

Plusieurs aspects de nos résultats retiennent notre attention. Tout d'abord, contrairement aux résultats obtenus lors de la tâche de dénomination écrite (*cf.* section 5.1.2, p. 121), notre jeu d'images génère ici un accord sur le nom quasi parfait entre les participants. Nous pensons que cela est dû pour une part au fait que les participants sont placés dans une situation de communication impliquant un contexte fini (le jeu d'images) et qu'ils s'adressent à une personne dont ils ne connaissent pas l'identité. Cependant, comme cela était le cas pour le test précédent nous ne pouvons pas effectuer ce type de conclusion avec certitude. Tout d'abord la nature des tâches est différente : dans la première les sujets dénommaient les images à l'écrit et en dehors de toute situation de communication alors que dans la seconde les sujets dénomment les images à l'oral, et ces noms sont produits à l'intérieur de structures lexico-syntaxiques. De plus, la population convoquée dans les deux cas était très différente. Par exemple, la plupart des sujets convoqués dans la seconde étude étaient étudiants ; or, nous avons vu lors de la première tâche de dénomination que l'expertise professionnelle des sujets semblait être une source de désaccord sur le nom. Les résultats obtenus dans les deux tests ne sont donc pas comparables de ce point de vue.

29. à travers ce test l'idée était aussi de tester le fonctionnement de l'interface *locuteur* d'EloKanz – comme nous le relatons dans la section 5.2 (p. 132), et d'enregistrer des locuteurs pour avoir des stimuli utilisables dans nos expériences ultérieures. C'est pourquoi dans ce test nous avons placé les sujets dans de véritables conditions d'enregistrement.

Sur le plan des besoins liés au fonctionnement du test EloKanz, les résultats obtenus dans cette seconde étude sont très satisfaisants. En effet, nous avons pu obtenir un jeu important d'images (110) dont l'accord sur le nom est de 100% lors de l'utilisation du logiciel EloKanz. Même si le nombre de sujets de cette étude est assez restreint, nous pouvons penser qu'*a priori* ces images ne poseront pas de difficultés majeures à d'autres sujets locuteurs et auditeurs, et qu'ils auront tendance à utiliser les mêmes noms pour les désigner.

5.2 Test de l'interface locuteur d'EloKanz

Comme nous l'avons spécifié, un des objectifs du test de dénomination présenté dans la section précédente était de vérifier que l'ergonomie de l'interface locuteur d'EloKanz ne posait pas de problème particulier aux sujets. En effet, le test EloKanz ayant été implémenté sous la forme d'un logiciel informatique requérant l'utilisation de matériel particulier (écran, microphone), nous voulions nous assurer qu'il n'existait pas de dysfonctionnement (de « bug ») compromettant son utilisation dans le cadre d'une étude de laboratoire. Ensuite, nous voulions nous assurer que les consignes proposées pour l'enregistrement de phrases ne posait pas de problème particulier aux sujets. Notamment, nous nous demandions :

- si l'épreuve ne semblait pas trop répétitive – et si elle était en cela fatigante ;
- si l'épreuve n'engendrait pas de confusion quant aux prépositions spatiales à employer (droite *vs* gauche par exemple) ;
- si la lecture de certaines images posait des problèmes particuliers aux sujets.

Participants

Comme nous l'avons indiqué, cette étude s'est déroulée en même temps que le second test de dénomination des images (*cf.* section précédente). Il a donc porté sur la même population. Nous souhaitons néanmoins préciser que nous avons volontairement inclus des participants ayant des problèmes de vue (7 sujets sur 12) et bénéficiant d'un moyen de correction (lentilles ou lunettes) au cas où la lecture des animations poserait un problème particulier à ce type de population.

Matériel

Comme nous l'avons spécifié dans la section précédente, chaque participant a été invité à participer à ce test dans une cabine audiométrique de la plateforme PE-TRA, dans laquelle ils étaient assis en face d'un moniteur. Sur l'écran d'ordinateur, les sujets pouvaient voir l'interface locuteur d'EloKanz, avec laquelle ils pouvaient interagir *via* une souris. Pour pouvoir enregistrer les phrases, les sujets étaient équipés d'un micro-casque de type SHURE BETA 53. Ce micro était relié à une table de mixage TASCAM DM-3200, elle-même connectée à l'ordinateur sur lequel tournait le programme EloKanz.

D'autre part, pour essayer d'appréhender les difficultés ressenties par les sujets durant l'utilisation du logiciel EloKanz, nous avons créé le questionnaire inclus en annexe C.2.3 (p. 297). Ce questionnaire propose tout d'abord d'appréhender la difficulté et la lourdeur (en terme de fatigue) ressenties par le locuteur à l'aide d'échelles numérotées de 1 à 7. Pour chacune des deux échelles (difficulté et « lourdeur » ressenties), le sujet est ensuite invité à préciser les raisons qui, selon lui, font que l'épreuve est plus ou moins difficile, ou plus ou moins lourde. Ensuite, quelques questions plus ouvertes permettent de questionner le sujet sur les difficultés rencontrées, sur les éléments qui ne lui semble pas ou peu naturels, sur les images visionnées, ou sur toute autre remarque concernant le test EloKanz. En plus de ce questionnaire,

nous avons utilisé comme indicateur les temps de réaction des participants, récupérés par le programme EloKanz. Les temps de réaction correspondent à la durée (en millisecondes) entre la fin de chaque animation et le moment où le participant clique sur le bouton « parler » pour enregistrer une phrase.

Procédure

Une passation de test se déroulait dans l'ordre suivant :

1. Les consignes générales, données en annexe C.2.1 (p. 296), étaient lues au sujet ;
2. Le sujet était équipé du micro-casque ;
3. Les consignes pour la phase d'entraînement (*cf.* annexe C.2.2, p. 296) étaient lues au sujet ;
4. Le sujet participait à une phase d'entraînement consistant à enregistrer 10 phrases avec le logiciel EloKanz. Nous assistions chaque sujet durant cette phase, et répondions à ses éventuelles questions ;
5. Le sujet enregistrait, seul, trois séries de 20 phrases avec le logiciel EloKanz. À la fin de chaque série nous soumettions au sujet le questionnaire (annexe C.2.3, p. 297).

Résultats

Difficultés par rapport à l'utilisation du logiciel EloKanz

Sur l'ensemble des 720 phrases enregistrées par les sujets, nous n'avons relevé aucune confusion concernant les prépositions spatiales (au-dessus *vs* en dessous, droite *vs* gauche) correspondant aux mouvements indiqués par les animations. De même, sur l'ensemble des 720 animations, seules 3 ont été évitées (les locuteurs ont cliqué sur le bouton « suivant » sans avoir enregistré de phrase).

Sur l'échelle subjective de difficulté (échelle de 1 à 7, de 1- *Très facile* à 7- *Très difficile*, la moyenne des notes données par les participants (toutes sessions confondues) correspondait à 1,64. Le détail de ces notes est donné dans la table 5.2.

Globalement, les notes qu'ont donné les sujets sont donc très basses, indiquant par là qu'ils ont plutôt ressenti le test comme étant facile, voire « très facile ». Quelques écarts existent cependant entre les sujets : certains n'ont donné que des notes de 1 après chaque session (sujets 2, 6, 8 et 9), et les autres des notes plus élevées, sans toutefois dépasser la note de 4. Il faut noter que deux des sujets ayant attribué les notes les plus hautes (sujets 3 et 7) étaient des experts en psychologie expérimentale. Or, la discussion que nous avons eue avec eux après le questionnaire nous a suggéré que ces sujets n'avaient pas simplement noté la difficulté qu'ils avaient *eux-même* ressenti pendant l'épreuve, mais qu'ils ont aussi tenu compte de la difficulté du test *dans l'absolu*, c'est-à-dire la difficulté du test par rapport à l'expérience qu'ils ont des tests habituellement conduits dans leur discipline.

Concernant maintenant la « lourdeur » de la tâche – en termes de fatigue ressentie par les sujets, les résultats obtenus sont indiqués dans la table 5.3. Tout comme pour

TABLE 5.2 – Difficulté ressentie par rapport à l’interface *locuteur* d’EloKanz (échelle de 1 – *Très facile* à 5 – *Très difficile*)

Sujet n°	Session 1	Session 2	Session 3
1	2	1	2
2	1	1	1
3	2	3	1
4	1	1	1
5	1	2	2
6	1	1	1
7	3	4	3
8	1	1	1
9	1	1	1
10	2	2	2
11	2	3	2
12	2	2	1
Moy.	1,58	1,83	1,50

l’échelle de difficulté, les scores obtenus sont dans l’ensemble assez bas (moyenne de 1,61 pour toutes sessions confondues). Six sujets (sujets 1, 4, 6, 10, 11 et 12) n’ont donné que des notes de 1 pour l’ensemble des trois sessions d’enregistrement.

Les scores obtenus pour les deux échelles, tout en étant assez bas, démontrent une certaine variabilité. Cependant, cette variabilité ne semble pas pouvoir être expliquée par le fait que la difficulté ou la fatigue ressentie par les sujets irait croissant selon la durée de l’épreuve, comme nous l’aurions supposé au départ. En effet, dans les deux cas les scores sont en moyenne inférieurs pour la troisième session que pour la première.

La variabilité pourrait s’expliquer en partie par la diversité des difficultés ressenties par les sujets. En effet lorsque nous leur avons demandé de préciser les raisons les ayant conduit à attribuer telle ou telle note sur l’échelle de difficulté, les sujets ont évoqué les points suivants :

- il est parfois difficile de dénommer les images (5 sujets) ;
- il faut se concentrer sur la localisation (droite, gauche, haut, bas ; 4 sujets) ;
- la tâche est répétitive (2 sujets) ;
- il faut se concentrer sur la prononciation (2 sujets).

Dans leurs commentaires les sujets ont rapproché l’aspect « répétitif » de la tâche de leurs difficultés à se concentrer sur la localisation. Les difficultés à dénommer les images ont été expliquées par les sujets en termes de besoin de concentration (« Il faut réfléchir », « C’est lourd »), mais aussi par rapport à des problèmes que leur posaient des images particulières (voir commentaires dans la table 5.4).

TABLE 5.3 – Lourdeur (fatigue) ressentie par rapport à l'utilisation de l'interface *locuteur* d'EloKanz (échelle de 1 – *Pas du tout lourd/fatigant* à 5 – *Très lourd/fatigant*)

Sujet n°	Session 1	Session 2	Session 3
1	1	1	1
2	2	2	2
3	2	1	1
4	1	1	1
5	5	4	3
6	1	1	1
7	3	3	3
8	2	2	1
9	1	1	2
10	1	1	1
11	1	1	1
12	1	1	1
Moy.	1,75	1,58	1,50

TABLE 5.4 : Images posant des problèmes aux sujets dans l'interface *locuteur* d'EloKanz

Image	Nb. sujets	Commentaire(s)
Grue	3	« C'est bien une grue ? » (2) ; « Je cherchais le nom » (1)
Avocat	2	« C'est bien un avocat ? » (2)
Brosse	2	« Faut-il dire <i>brossette</i> ? » (1) ; « Est-ce une brosse à cheveux ? » (1)
Poulet	2	« Il est cru, c'est bizarre » (2)
Salade	2	« Devais-je dire <i>batavia</i> ? » (1) ; « J'ai dit <i>laitue</i> » (1)
Allumettes	1	« J'ai hésité entre <i>allumettes</i> et <i>boîte d'allumettes</i> »
Asperges	1	« C'est un terme peu fréquent »
Autruche	1	« C'est bien ça la race ? »
Escargot	1	« J'ai failli dire <i>cagouille</i> mais j'ai pensé que ça ne serait peut-être pas compris »
Fleur	1	« Je ne savais pas de quelle fleur il s'agissait »
Igloo	1	« C'est peu fréquent »
Lunettes	1	« Faut-il dire <i>paire de lunettes</i> ou <i>lunettes</i> ? »
Marteau	1	« Pour moi le marteau prototypique est muni d'un arrache-clou »
Moulin	1	« On se demande ce qu'un moulin fait là »
Ours	1	« J'ai failli dire <i>grizzli</i> »
Parapluie	1	« Je ne savais pas si c'était un parapluie, un parasol ou une ombrelle »

Images posant des problèmes aux sujets (suite)

Pastèque	1	« J'ai mis du temps à trouver le mot »
Poussin	1	« Faut-il préciser qu'il s'agit d'un poussin de canard ? »
Râteau	1	« Il est un peu petit »
Scoter	1	« J'ai dit <i>moto</i> »
Skieur	1	« Ça va pas avec le reste »
Tondeuse	1	« J'ai cherché un moment comment ça s'appelait »

Les sujets porteurs de corrections optiques (lunettes ou lentilles de contact) ne se sont pas plaints d'une quelconque gêne occasionnée par leur problème de vue. Cependant, si l'on compare les scores obtenus par les sujets porteurs de correction (sujets 2, 3, 5, 6, 7, 8, 12) et par les autres sujets, il apparaît une nette différence pour ce qui est de l'évaluation de la lourdeur de la tâche. La table 5.5 illustre ce point.

TABLE 5.5 – Moyennes des scores pour les échelles de difficulté et de lourdeur de la tâche, pour les sujets porteurs et non porteurs de corrections oculaires

	Sujets non porteurs de correction	Sujets porteurs de correction
Échelle de difficulté	1,60	1,67
Échelle de lourdeur	1,07	2,00

Pour avoir une indication de la significativité de cette différence, nous avons effectué un test de Mann-Whithney. Les résultats, présentés dans la table 5.6, indiquent une différence proche du seuil de significativité ($P = 0,056$). Pour vérifier que la différence de lourdeur ressentie par les sujets porteurs de corrections oculaires et non porteurs de telles corrections, il faudrait certainement effectuer un test auprès d'un groupe plus important de sujets. En effet, même pour un test non paramétrique comme celui de Mann-Whithney le nombre de sujets dans chaque groupe nous semble trop faible pour interpréter correctement les résultats obtenus.

De plus, les temps de réaction moyens³⁰ (cf. table 5.7) n'indiquent pas de différence majeure entre les deux groupes de sujets : les deux moyennes se situent aux alentours de 1720 ms (1696 ms pour les sujets sans problème de vue, 1741 ms pour les sujets utilisant un moyen de correction optique). Ces données ne paraissent pas aller dans le sens de l'existence d'une différence majeure concernant les performances d'utilisation du logiciel par les deux groupes de sujets.

30. un temps de réaction correspond à la durée (en ms) entre la fin de lecture d'une animation et le moment où le sujet clique sur le bouton « parler » pour enregistrer une phrase.

TABLE 5.6 – Différences d'évaluation de la lourdeur de la tâche entre les sujets porteurs de corrections oculaires et non porteurs de telles corrections (test de Mann-Whitney)

Rangs				
	Correction oculaire	N	Rang moyen	Somme des rangs
Évaluation de la lourdeur de la tâche sur une échelle de 1 à 7	Non	5	4,30	21,50
	Oui	7	8,07	56,49
	Total	12		

Tests pour échantillons indépendants

Variable	U de Mann-W.	Sig. bilatérale
Évaluation de la lourdeur de la tâche sur une échelle de 1 à 7	6500	0,056

Dysfonctionnements techniques

Durant ce test, nous n'avons relevé que peu de dysfonctionnements techniques. Le logiciel EloKanz a bien fonctionné lors de la mise en place des séries d'animations, de la lecture des animations et de la récupération des données (*cf.* la description du fonctionnement de l'interface en section 4.2.2, p. 98). En revanche, nous avons noté deux problèmes liés à l'enregistrement des phrases par les locuteurs :

1) certains fichiers d'enregistrement étaient incomplets : une partie des données sonores n'était pas enregistrée. Par exemple après l'enregistrement d'une phrase telle que [metelatomatəsulekamjõ] (« Mettez la tomate sous le camion »), le fichier AIF résultant ne contenait plus que [metesulekamjõ] (« Mettez sous le camion »). Les données manquantes n'étaient pas remplacées par du silence : le fichier apparaissait simplement comme « tronqué ». Nous ne sommes pas parvenu à trouver la source de cette erreur, ce qui nous a empêché de pouvoir la corriger³¹. Cependant, puisque les erreurs d'enregistrement étaient assez peu fréquentes (de l'ordre de 1 pour 40 phrases environ), nous avons décidé d'opter pour la solution suivante. Nous avons créé une fonction permettant de vérifier l'intégralité (en terme de durée) des enregistrements.

31. nous avons passé beaucoup de temps à essayer d'identifier la cause de ce dysfonctionnement, en essayant notamment de reproduire les conditions de son apparition. En vain. Nous n'avons trouvé aucune régularité pouvant expliquer l'occurrence de cette erreur. La seule chose dont nous étions sûr est qu'il était causé par le plug-in *Audio XTRA*. Nous avons essayé de trouver d'autres plug-ins permettant l'enregistrement de données sonores, mais tous étaient incompatibles avec notre version du logiciel *Adobe Director*® et/ou avec le système d'exploitation *Mac OS X*. Nous avons également contacté l'équipe technique de la société *Tabuleiro*, qui nous a indiqué que ce problème était probablement dû à l'obsolescence du plug-in *Audio XTRA* en regard des nouveaux environnements *Mac OS X*. Cependant la société *Tabuleiro* n'entend plus effectuer de mise à jour du plug-in *Audio XTRA*.

TABLE 5.7 – Temps de réaction des locuteurs (en ms) pour l'enregistrement d'une phrase

Sujet n°	Session 1	Session 2	Session 3
1	2334	3542	2095
2	2353	1910	1263
3	2675	2095	1681
4	1374	1276	1276
5	2189	1933	2003
6	1204	1375	1603
7	1121	1963	1358
8	1647	1391	1299
9	1493	1668	1358
10	1529	1412	1034
11	1726	1705	1619
12	2288	2069	1600
Moy.	1828	1862	1477

Le code fonctionne de la manière suivante :

- le programme enregistre la durée (en millisecondes) entre le moment où le locuteur a cliqué sur le bouton [PARLER] et celui où il a cliqué sur [STOP] pour terminer l'enregistrement ;
- ensuite, le programme va rechercher le fichier audio AIF résultant de l'enregistrement. La taille du fichier, en octets, est convertie en millisecondes ;
- enfin, les deux durées sont comparées. S'il existe une différence de plus de 0,2 seconde entre les deux durées, alors un message est affiché à l'écran, indiquant l'erreur et demandant au locuteur de recommencer son enregistrement.

2) deuxièmement, les sujets avaient tendance, avec l'habitude, de cliquer trop tôt sur le bouton « Stop » à la fin des enregistrements. Il arrivait ainsi que le dernier mot prononcé soit tronqué. Pour pallier ce problème, nous avons fait en sorte que, lorsque le locuteur clique sur le bouton « Stop », une seconde de délai soit respectée avant de terminer l'enregistrement.

5.3 Conclusion et discussion

En conclusion, nos résultats répondent selon nous de façon satisfaisante aux objectifs que nous nous étions fixés. D'une part, nous avons pu constituer un jeu de 110 illustrations dont on peut supposer qu'elles ne posent pas de problème de reconnaissance aux sujets, et pour lequel l'accord sur le nom est maximal. D'autre part, les consignes et l'ergonomie de l'interface *locuteur* d'EloKanz ne semblent pas causer de difficulté aux sujets, qui ont dans l'ensemble jugé le test comme facile voire très facile. D'après nos résultats le test pourrait se révéler particulièrement fatigant pour des sujets porteurs de corrections oculaires (lentilles, lunettes) – mais la tendance observée n'était pas statistiquement significative et ne se traduisait pas dans les performances des sujets en termes de temps de réaction aux stimuli visuels. Enfin, le logiciel EloKanz a donné toute satisfaction, même si des problèmes d'enregistrement audio sont parfois intervenus. Nous avons cependant été capable de programmer un algorithme pour repérer automatiquement les erreurs d'enregistrement et afficher un message d'erreur à l'écran – ce qui permet aux sujets locuteurs de prendre connaissance du problème et d'enregistrer la phrase à nouveau.

Par ailleurs, nous voudrions souligner le fait que, lors de nos études de l'accord sur le nom des images, nous avons d'abord considéré une tâche de dénomination écrite, puis nous nous sommes orienté vers une tâche de dénomination en situation de communication, pour finir par tester l'accord sur le nom *directement lors de l'utilisation d'EloKanz*. Cette transition traduit selon nous le fait que nous avons pris peu à peu conscience de *l'importance du rôle de la tâche dans les résultats obtenus* : nous avons en effet observé que les données relevées variaient de manière importante selon la tâche de dénomination employée. Or, puisque notre objectif était de vérifier l'accord sur le nom des images *en vue de leur utilisation dans EloKanz*, nous avons jugé que le test le plus approprié consistait à étudier l'accord sur le nom *directement lors de l'utilisation d'EloKanz par les sujets*.

Cette hypothèse forte sur l'influence de la tâche, de même que sur l'influence de certains paramètres propres à la population convoquée offre selon nous des perspectives intéressantes à notre travail. Par exemple, dans le test de dénomination écrite nous avons observé une tendance des sujets à préciser les concepts représentés par les images, notamment lorsque cette distinction était pertinente dans leur domaine professionnel. Nous pensons qu'il serait intéressant d'étudier plus précisément cette question, pour vérifier qu'il s'agit bien là d'une influence de l'expertise des sujets sur le lexique employé. Cette question est d'autant plus importante qu'elle pourrait questionner la validité des normes d'accord sur le nom disponibles aujourd'hui. En effet ces dernières ont en général été collectées auprès de populations uniquement constituées d'étudiants, et pour une grande part en première année de psychologie (par exemple Alario et Ferrand, 1999; Bonin *et al.*, 2003; Dimitropoulo *et al.*, 2009; Nishimoto *et al.*, 2005; Rossion et Pourtois, 2004; Snodgrass et Vanderwart, 1980). Il est donc possible que ces données ne soient pas totalement représentatives de ce que l'on peut observer auprès de populations appartenant à d'autres catégories socio-professionnelles.

De plus, les différences d'accord sur le nom que nous avons observées entre nos

différents tests de dénomination semblent montrer qu'il existe une influence importante de la tâche de dénomination employée. Ceci pose également la question de la représentativité des données d'accord sur le nom collectées dans les différentes études que nous avons présentées en section 5.1.1 (p. 112). Dans ces dernières l'accord sur le nom est relevé dans des tâches qui n'impliquent pas de contexte de communication : les sujets dénomment simplement les images à l'oral ou à l'écrit. Il est possible que les données d'accord sur le nom relevées dans ces différents tests ne soient pas représentatives de l'accord que l'on peut relever entre interlocuteurs dans une situation de communication. Plus précisément, notre étude nous a conduit à émettre deux hypothèses :

1. dans une situation de communication comme celle en jeu dans le test EloKanz (ou comme dans celles qui sont utilisées dans le courant de la pragmatique, *cf.* p. 75), le contexte physique partagé par les interlocuteurs, c'est-à-dire l'ensemble des référents possibles du discours (le *jeu d'images* dans EloKanz), a une influence sur l'accord sur le nom entre les interlocuteurs. Suivant le courant de la pragmatique et de l'analyse conversationnelle, nous pensons que les interlocuteurs s'appuient sur leur « terrain commun » (*common ground* – *cf.* Clark, 1996) pour interagir verbalement. Ainsi, si le contexte qu'ils partagent ne nécessite pas de distinguer entre deux référents partageant des propriétés communes (ex. une *feuille de papier* et une *feuille d'arbre*, un *citron jaune* et un *citron vert*, etc.), nous pensons qu'ils auront tendance à se référer à ces éléments en utilisant des dénominations situées à un niveau de base (c.-à-d. *feuille* et *citron*).
2. les dénominations utilisées par le locuteur prennent en compte les connaissances linguistiques et encyclopédiques de la personne à qui le message est adressé. Ceci a trait, de nouveau, à la notion pragmatique de *terrain commun* : les interlocuteurs ont besoin de construire un espace de connaissances partagées pour communiquer. Et pour créer cet espace les interlocuteurs effectuent, entre autres, des hypothèses sur les connaissances linguistiques et encyclopédiques qu'ils partagent. EloKanz met en jeu une situation de communication très particulière en cela que le locuteur ne sait presque rien³² de la personne à qui les messages sont adressés. Nous pensons que dans une telle situation le locuteur a tendance à utiliser des mots/concepts situés à un niveau *de base* pour maximiser les chances de situer ses propos dans le cercle des connaissances linguistiques et encyclopédiques que son interlocuteur partage avec lui.

Nos tests d'accord sur le nom n'ont pas été conçus pour vérifier ces deux hypothèses. Nous pensons cependant qu'il serait intéressant d'étudier ces questions de plus près. Si ces hypothèses étaient vérifiées, il serait alors possible d'envisager de contrôler le contexte de communication pour que l'accord sur le nom soit maximal entre les interlocuteurs. Cela permettrait de recourir à des tâches impliquant de la parole spontanée dans de nombreux domaines de recherche sur la production et la

32. si nous écrivons *presque* rien, c'est que les locuteurs peuvent quand même, selon nous, faire des hypothèses assez fortes à propos des connaissances des personnes qui répondront à leurs messages, comme par exemple le fait que ces dernières parlent le français.

compréhension de la parole – et nous pensons en particulier à l'évaluation des performances de patients souffrant de troubles pathologiques de production de la parole (TPPP). En effet, selon Ellis et Beltyukova (2008), très peu d'études relatives aux TPPP utilisent de la parole spontanée car il peut être problématique de déterminer les paroles qui ont voulu être prononcées par le locuteur (par exemple dans le cas de signaux de parole très dégradés). Pourtant, il y aurait un grand intérêt à évaluer des paroles émises spontanément, car nous savons que celles-ci présentent des caractéristiques acoustiques différentes de celles des paroles résultant de tâches de lecture à voix haute, que ce soit au niveau segmental ou prosodique (*cf.* entre autres : Duez, 1982, 1987, 1992; Lindblom, 1990a; Vaissière et Michaud, 2005). Un accord sur le nom maximal permettrait d'utiliser plus facilement de la parole spontanée dans le cadre des TPPP, car même si le signal est très dégradé les expérimentateurs ou les évaluateurs seraient à même de faire des hypothèses fortes quant aux noms que le locuteur a voulu prononcer.

Chapitre 6

Intelligibilité et compréhension de la parole dans le bruit

Sommaire

6.1	Objectifs et hypothèses	145
6.2	Matériel	148
6.2.1	Préparation du programme EloKanz	148
6.2.2	Stimuli de parole	149
	Enregistrement des stimuli	149
	Égalisation de l'intensité des stimuli	149
6.2.3	Stimulus de bruit	155
	Création du fichier de bruit	155
	Étalonnage des différents niveaux de bruit	156
6.3	Population et procédure	157
6.3.1	Population	157
6.3.2	Procédure générale	157
6.3.3	Test d'intelligibilité	158
	Consignes	158
	Calcul des scores	158
6.3.4	Test de compréhension en situation de communication	159
	Consignes	159
	Calcul des scores	160
6.4	Résultats	161
6.4.1	Scores d'intelligibilité et de compréhension en fonction du niveau de bruit	161
6.4.2	Relation entre les scores d'intelligibilité et de compréhension	166
6.4.3	Fiabilité des mesures	166
6.5	Discussion	168
6.5.1	De la force d'association entre les scores d'intelligibilité et les scores de compréhension	168
6.5.2	Des différences entre les scores d'intelligibilité et les scores de compréhension	171

CHAPITRE 6. INTELLIGIBILITÉ ET COMPRÉHENSION DE LA PAROLE DANS
LE BRUIT

Différences de moyenne 171
De la variabilité des deux mesures 173
6.6 Conclusion 175

6.1 Objectifs et hypothèses

L'objectif de cette expérience est d'étudier la relation entre les scores obtenus par une tâche de retranscription orthographique et les scores obtenus dans une tâche destinée à évaluer la compréhension des auditeurs en situation de communication. Nous avons vu (Chapitre 3) que les études qui ont porté sur les liens entre l'intelligibilité et la compréhension de la parole pathologique ont été confrontées au problème du contrôle du degré de sévérité des TPPP, une variable qui influence à la fois les scores d'intelligibilité et les scores de compréhension.

Dans l'étude de Beukelman et Yorkston (1979), qui a révélé une association très forte entre les scores d'intelligibilité et de compréhension de la parole ($r = 0,95$), ce contrôle a fait défaut. Pour cette raison Hustad (2008) a reconduit une étude similaire en contrôlant cette fois l'influence du degré de sévérité des TPPP dont souffraient les locuteurs. Seulement dans l'étude de Hustad la sévérité a été mesurée par un pré-test d'intelligibilité, ce qui pose selon nous un problème méthodologique majeur.

Hustad a calculé un coefficient de corrélation partielle ($r_{AB.C}$) : la force d'association (r) entre les scores d'intelligibilité (variable A) et les scores de compréhension (variable B) est calculée comme si le degré de sévérité (variable C) était demeuré constant pour la série d'observations considérées. Or le test d'intelligibilité utilisé par Hustad pour quantifier le degré de sévérité, le *Sentence Intelligibility Test for Macintosh* (*SIT*, Yorkston et Beukelman, 1996), utilise une procédure tout à fait comparable avec le test utilisé par Hustad pour mesurer l'intelligibilité dans son étude. Dans les deux cas les patients prononcent des phrases, et les scores correspondent au nombre de mots correctement retranscrits par les auditeurs. On peut donc supposer qu'il existe une très forte association (r_{AC}) entre les scores d'intelligibilité (variable A) et le degré de sévérité (variable C) dans l'étude de Hustad – puisque tous les deux ont été mesurés par le même type de test. Le corollaire de cette remarque est que, dans un calcul de corrélation partielle mettant en jeu ces deux variables, la force d'association résiduelle existant entre la première et les scores de compréhension risque d'être très faible. Et c'est effectivement le résultat obtenu par Hustad (2008), qui trouve un coefficient de corrélation de 0,056 et non significatif entre les scores d'intelligibilité et de compréhension.

Il semble donc difficile de donner une interprétation valable aux coefficients de corrélation trouvés par Beukelman et Yorkston (1979) et Hustad (2008). En conséquence nous avons décidé de mener une nouvelle étude des liens existant entre les scores d'intelligibilité et les scores obtenus par une méthode destinée à évaluer la compréhension des auditeurs. Pour pallier le problème du manque de critère objectif de mesure du degré de sévérité des TPPP, nous avons décidé d'observer les différences entre l'intelligibilité et la compréhension de la parole *dans le bruit*.

En effet, lorsque l'on compare des scores d'intelligibilité et de compréhension, on compare les résultats obtenus *par des auditeurs* à deux tâches différentes. La question centrale n'a donc finalement que peu à voir avec *la cause* de la dégradation du signal de parole, que celle-ci intervienne au niveau de *l'émetteur* (locuteur souffrant de TPPP de nature structurelle ou neurologique, locuteur étranger, système

de génération de la parole, etc.) ou au niveau du *canal de communication* (bruit, réverbération, etc.). En effet, puisque les deux tâches – tests d’intelligibilité et de compréhension – sont effectuées par les auditeurs, les seules contraintes pesant sur les stimuli sont :

1. que le signal de parole soit dégradé, et que ces dégradations soient d’importance variable¹ ;
2. que le contenu linguistique soit correct sur les plans lexical, morphosyntaxique et phonologique.

Compte-tenu de ces deux contraintes, rien ne nous empêchait d’utiliser des stimuli de parole produits par des locuteurs *ne souffrant pas de TPPP*, dégradés artificiellement et de manière contrôlée par la diffusion simultanée de bruit. Car, si les problèmes intervenant au niveau de l’émetteur ne peuvent être quantifiés de manière objective, il est en revanche possible d’effectuer des mesures objectives du *niveau de bruit* (cf. par exemple la prise en compte du niveau de bruit dans certaines mesures prédictives de l’intelligibilité de la parole, section 1.4 p. 23).

Nous avons donc décidé d’élaborer un protocole impliquant deux tâches – intelligibilité et compréhension en situation de communication (mesurée à l’aide du programme EloKanz) – et reposant sur l’écoute de signaux de parole *non pathologique*, accompagnés de bruit diffusé à différents niveaux sonores. Notre hypothèse est que, une fois contrôlée la variabilité associée aux différents niveaux de bruit, nous trouverons une corrélation significative entre les scores d’intelligibilité et de compréhension. Nous supposons en effet *qu’une relation existe entre les deux mesures*, qu’elles ne sont pas totalement indépendantes comme le suggèrent les résultats de Hustad (2008). Notamment, nous avons vu dans notre état de l’art que la plupart des modèles psycholinguistiques postulent que la compréhension de la parole repose à la fois sur des fonctions de forme et de sens – que ces dernières fassent l’objet de traitements parallèles ou non. Dans cette optique les scores obtenus par des tests de compréhension de la parole ne devraient pas être totalement indépendants de scores obtenus par un test portant sur la forme des messages – comme les tests de retranscription orthographique.

D’un autre côté, nous pensons que des différences importantes distinguent certainement ces deux mesures – ce qui nous laisse penser que la force d’association que nous trouverons entre les deux mesures ne sera pas aussi importante que dans les résultats obtenus par Beukelman et Yorkston (1979). En effet, les deux mesures – intelligibilité et compréhension en situation de communication – ne sont pas obtenues à partir des mêmes tâches. Une des différences majeures entre les deux tâches est la présence d’un contexte de communication dans le cas de la compréhension : dans le logiciel EloKanz les énoncés produits par le locuteur concernent des référents représentés par des images sur un écran. Nous supposons que la présence d’un contexte facilitera la compréhension des paroles, et donnera donc lieu à des performances plus élevées que dans la tâche de retranscription orthographique. Les études antérieures

1. la variabilité est en effet nécessaire pour étudier la corrélation entre les deux mesures.

de Hustad (2008) et de Beukelman et Yorkston (1979) ont d'ailleurs indiqué que les scores obtenus dans des tâches de compréhension étaient dans l'ensemble plus élevés que dans des tâches de retranscription orthographique. Cependant dans le cas de ces deux études la supériorité des scores de compréhension était expliquée par le fait que, pour comprendre les paroles prononcées par les patients les auditeurs prenaient également en compte le contexte discursif ainsi que leurs connaissances générales sur le monde. Dans notre cas, nous pensons que c'est le contexte physique (le jeu d'images représenté à l'écran) qui va aider les auditeurs à saisir les paroles. De plus, nous pensons que cette différence entre les scores d'intelligibilité et les scores de compréhension ne sera pas du même ordre pour toutes les conditions de bruit. En effet, nous pouvons supposer que les auditeurs ne s'appuieront pas autant sur le contexte dans le cas de stimuli assez faciles à saisir (conditions de faible bruit) que dans le cas de stimuli difficiles à saisir (conditions de bruit fort).

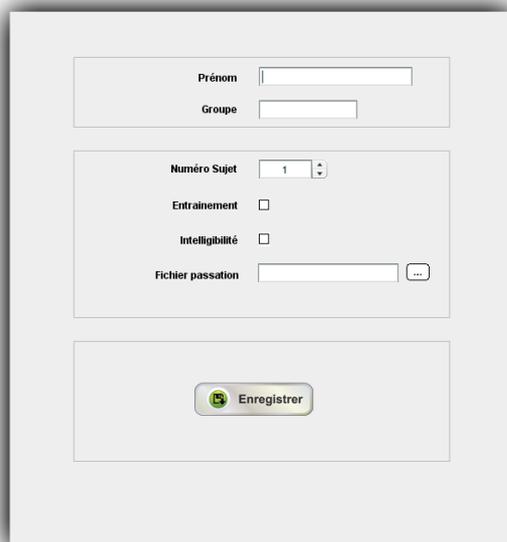
6.2 Matériel

6.2.1 Préparation du programme EloKanz

Pour effectuer cette expérience, nous avons apporté quelques modifications à l'interface *auditeur* du programme EloKanz, pour que le programme diffuse du bruit à différents niveaux avec les différents stimuli de parole. Comme nous le verrons, notre expérience implique de faire écouter à quatre groupes de 10 auditeurs 50 phrases présentées selon 10 conditions de bruit ; nous avons donc modifié le programme EloKanz pour qu'il distribue les 10 conditions de bruit en fonction des phrases et des participants².

De plus, les données correspondant à la passation d'un auditeur (stimuli audio, niveaux de bruit et images) peuvent être reprises depuis l'interface du logiciel (champ « Fichier passation » dans la figure 6.1), ce qui permet de placer différents auditeurs dans les mêmes conditions d'expérimentation. Une option permet également de placer les participants dans les conditions d'entraînement. Ces conditions sont définies dans un fichier contenant les informations sur les niveaux de bruit et les images accompagnant une dizaine de phrases³.

FIGURE 6.1 – Fenêtre de saisie des informations sur l'auditeur



The screenshot shows a software window titled 'auditeur' with a light gray background. It contains several input fields and checkboxes arranged in a structured layout. At the top, there are two text input fields labeled 'Prénom' and 'Groupe'. Below these, there is a dropdown menu for 'Numéro Sujet' with the number '1' selected. Underneath the dropdown are two checkboxes: 'Entraînement' and 'Intelligibilité', both of which are currently unchecked. At the bottom of this section is a text input field for 'Fichier passation' followed by a small button with three dots. At the very bottom of the window is a large, rounded rectangular button labeled 'Enregistrer' with a green play icon on the left.

Enfin, nous avons inclus la possibilité de masquer ou non les images à l'écran, pour placer les sujets dans les conditions du test d'intelligibilité ou de compréhension en contexte.

2. pour ce faire EloKanz crée pour chaque participant un fichier de passation, dont nous donnons un exemple en annexe DVD 3a_3.

3. nous donnons ce fichier en annexe DVD 5c_1.

Le programme complet, tel qu'il a été conçu et utilisé pour cette expérience, est fourni en annexe DVD 1b⁴.

6.2.2 Stimuli de parole

Enregistrement des stimuli

Nous avons enregistré 50 phrases prononcées par une locutrice francophone, âgée de 25 ans, et qui ne présentait pas de trouble d'élocution. Elle a été enregistrée lors de notre seconde étude d'accord sur le nom des images (section 5.1.3, p. 129). La transcription orthographique des 50 phrases est donnée en annexe D.2, et les 50 fichiers audio correspondant sont inclus dans l'annexe DVD 5c_2.

Égalisation de l'intensité des stimuli

Puisque notre objectif était d'évaluer la compréhension et l'intelligibilité des stimuli en faisant varier le niveau de bruit, nous avons tout d'abord procédé à l'égalisation du niveau sonore des 50 phrases audio. Après avoir tenté – sans succès – d'égaliser le niveau des fichiers de manière satisfaisante en fonction de leur valeur efficace (*RMS – Root-mean-square*), nous avons procédé à un pré-test d'égalisation à partir d'évaluations subjectives du niveau sonore. Dans la prochaine sous-partie nous détaillons la procédure d'égalisation en fonction de la valeur efficace des stimuli. Ces détails ne sont exposés que pour illustrer la difficulté qu'il peut y avoir à égaliser des phrases grâce à des mesures objectives pour une problématique comme la nôtre. Nous invitons donc le lecteur, s'il veut directement accéder aux détails de la procédure d'égalisation que nous avons retenue, à lire les informations présentées à partir de la page 153.

Égalisation en fonction de la valeur efficace des stimuli

La valeur efficace des signaux de parole est un critère d'égalisation du niveau sonore habituellement utilisé pour les tests d'intelligibilité de la parole dans le bruit (par exemple Jansen *et al.*, 2010; Nielsen et Dau, 2009; Ozimek *et al.*, 2009; Wagener et Brand, 2005; Wagener *et al.*, 2006). La valeur efficace est définie comme :

$$\sqrt{\frac{1}{(t_2-t_1)} \int_{t_1}^{t_2} [x(t)]^2 dt}$$

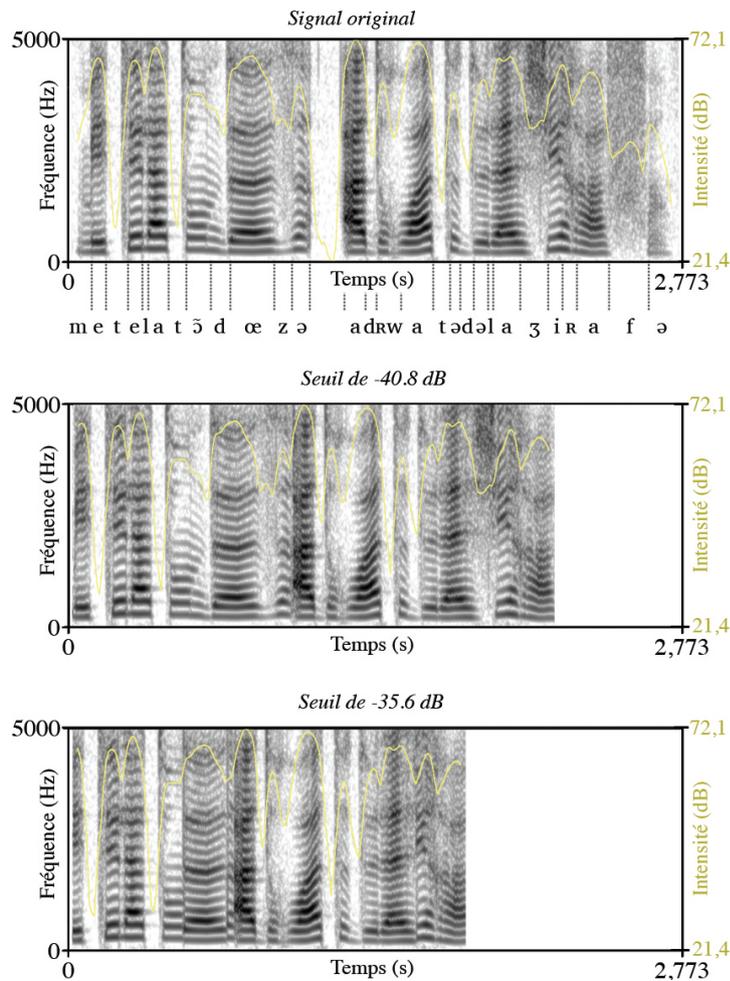
où t_1 et t_2 correspondent aux limites de l'intervalle temporel sur lequel est calculée la valeur efficace, et $x(t)$ à l'amplitude du son. Dans notre cas, l'intervalle (t_1, t_2) renvoyait à la durée totale de chaque stimulus, autrement dit de chaque phrase.

Notre procédure d'égalisation du niveau sonore a consisté à déterminer, parmi les 50 fichiers audio, quel était le stimulus dont la valeur efficace était la plus basse, puis à atténuer le niveau des autres fichiers pour atteindre cette valeur. Cette procédure

4. cette version du programme fonctionne uniquement sous le système d'exploitation Mac OS.

nous a permis d'effectuer uniquement des atténuations de niveau sonore, et d'éviter ainsi d'éventuels problèmes de saturation lors de l'égalisation des fichiers.

FIGURE 6.2 – Spectrogrammes de signaux de parole avant et après atténuation. Le signal original correspond à la phrase « Mettez la tondeuse à droite de la girafe »

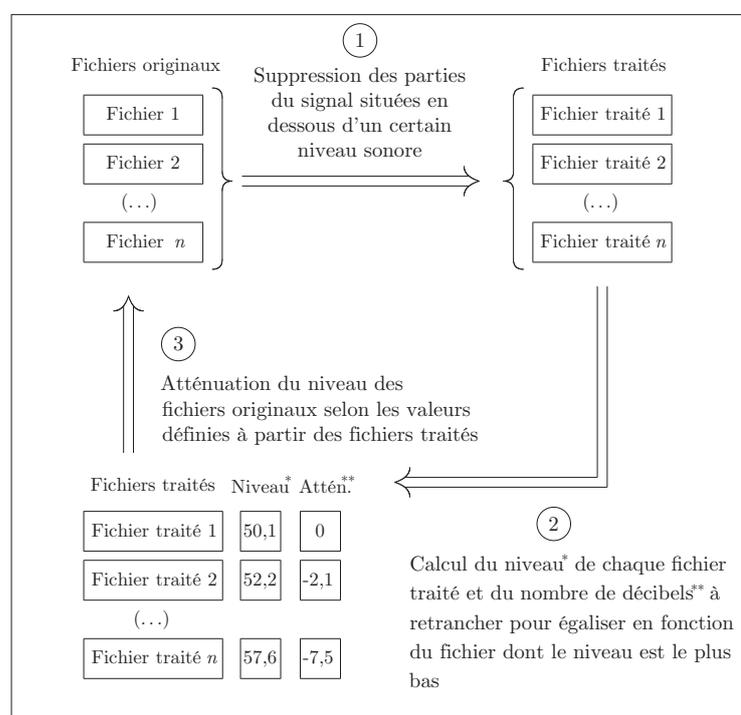


Les écarts de niveau sonore entre les fichiers étaient assez importants, le niveau sonore le plus bas étant de 60,05 dB, et le plus élevé de 67,65 dB. Or, une fois que nous avons abaissé le niveau de tous les fichiers à une valeur de 60,05 dB, les stimuli semblaient très hétérogènes en terme de niveau sonore *perçu* – c'est-à-dire que les fichiers semblaient, à l'écoute, de niveaux très différents. Nous avons fait l'hypothèse que ces problèmes étaient dus à la présence de *pauses*⁵ de différentes durées dans les phrases enregistrées : la valeur efficace étant calculée sur la durée totale des stimuli, les périodes de « silence » influencent les moyennes obtenues pour chaque phrase.

5. la locutrice fait en effet des pauses de différentes durées au niveau des frontières des constituants syntaxiques des énoncés. Cela est souvent le cas entre le complément d'objet direct du verbe *mettre* et le syntagme prépositionnel suivant (ex. [metelatomatə#suləkamjə]).

Pour tenter de pallier ce problème, nous avons créé deux scripts écrits en Shell et en Perl, et utilisant l'utilitaire SoX (*Sound eXchange*⁶). Ces scripts, disponibles en annexe DVD 5c_3, suppriment les parties du signal situées en-dessous d'un certain niveau sonore (*cf.* les exemples donnés en figure 6.2) – ce qui permet d'éliminer les pauses. La procédure consiste ensuite à s'appuyer sur les nouvelles différences de niveau observées entre les fichiers pour procéder à l'égalisation des fichiers originaux. La figure 6.3 illustre cette procédure en détails.

FIGURE 6.3 – Procédure d'égalisation du niveau des stimuli audio en fonction de leur niveau sonore, après suppression des pauses



Nous avons testé plusieurs valeurs de seuil, comprises entre -30 dB et -50 dB. Toutefois, et quel que soit le seuil utilisé, l'égalisation des fichiers résultant de cette procédure ne nous semblait pas satisfaisante : les fichiers semblaient toujours disparates en terme de *niveau sonore perçu*. Pour vérifier que cette impression n'était pas uniquement le fait de notre propre perception, nous avons fait écouter les enregistrements à 6 sujets chercheurs dans le domaine de la parole. Les 6 sujets étaient équipés d'un casque et pouvaient écouter, autant de fois qu'ils le souhaitaient, 3 groupes de 10 phrases :

- 1^{er} groupe : phrases égalisées selon la procédure décrite en figure 6.3, avec un seuil de niveau de -40.8 dB ;

6. <http://sox.sourceforge.net/>

- 2^{ème} groupe : phrases égalisées selon la procédure décrite en figure 6.3, avec un seuil de niveau de -35.6 dB ;
- 3^{ème} groupe : phrases originales.

Les sujets avaient pour consigne de déterminer quel groupe de phrases semblait le plus homogène en terme de niveau sonore perçu.

La moitié des auditeurs a considéré que les fichiers originaux (c'est-à-dire ceux n'ayant subi aucun traitement) formaient un ensemble bien plus homogène que les autres. À partir de ce résultat, nous avons abandonné l'idée de procéder à l'égalisation du niveau sonore de nos stimuli en fonction de mesures objectives.

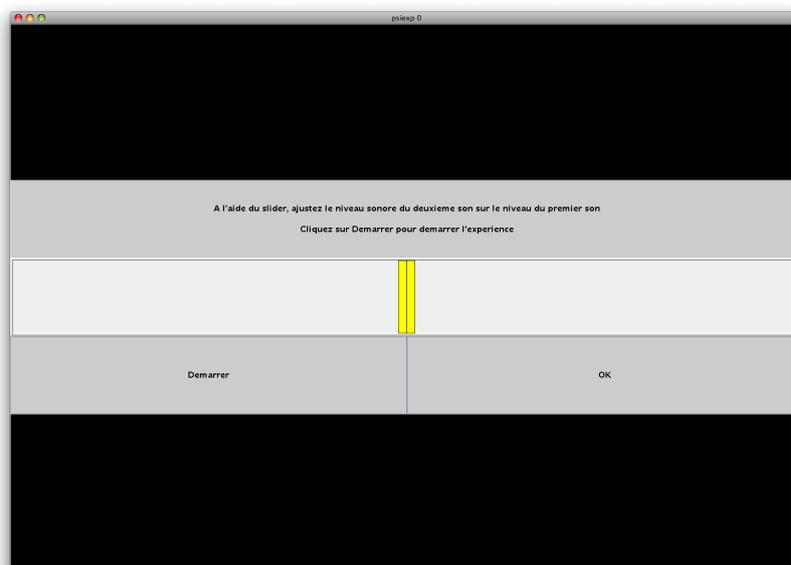
En effet, nous pensons que les mesures objectives du niveau sonore sont trop fortement influencées par le contenu des phrases en terme phonétique. Pour prendre un exemple, les consonnes plosives sourdes ont une énergie très forte en comparaison d'autres consonnes, et provoquent des « pics » de niveau dont la valeur est intégrée à la moyenne calculée sur l'ensemble du signal. Il semble en revanche qu'en perception nous soyons capables de juger directement le niveau « global » de la phrase, sans tenir compte de variations ponctuelles comme celle que nous venons d'évoquer. Pour cette raison, nous avons décidé de ne pas égaliser les fichiers en fonction de leur valeur efficace mesurée, mais de procéder à une égalisation reposant sur des évaluations subjectives du niveau sonore.

Égalisation sur la base d'évaluations subjectives du niveau sonore

Population. 8 sujets francophones, âgés de 24 à 44 ans, ont participé à ce test. Tous étaient chercheurs en sciences du langage, et ont indiqué ne pas présenter de troubles auditifs.

Procédure. Les sujets étaient assis à une distance d'un mètre⁷ de deux haut-parleurs *Tannoy Precision 6D*, dans la cabine audiométrique PETRA-2. Face à eux, un écran d'ordinateur diffusait l'interface présentée en figure 6.4, programmée⁸ à l'aide de *PsiExp* (Smith, 1995). Les sujets entendaient 50 paires de phrases. La première correspondait toujours à celle dont le niveau était le plus bas parmi l'ensemble de nos 50 stimuli, et que nous avons choisi comme fichier étalon pour cette procédure d'égalisation. Pour chaque paire de phrases, les sujets avaient pour consigne d'ajuster, à l'aide d'un curseur (en jaune sur la figure 6.4), le niveau sonore de la deuxième sur le niveau de la première. Les sujets pouvaient écouter chaque paire de phrases autant de fois qu'ils le désiraient.

FIGURE 6.4 – Interface utilisée pour le pré-test d'égalisation du niveau des stimuli sur la base d'évaluations subjectives



Résultats. Le programme a permis de relever toutes les valeurs de gain (positif ou négatif) que les sujets ont appliqué sur chacun des fichiers audio. Ces données sont disponibles en annexe DVD 2b. Comme nous pouvons le voir dans la table 6.1, les

7. dans la procédure de l'expérience visant à comparer l'intelligibilité et la compréhension de la parole dans le bruit (section 6.3), cette distance est équivalente.

8. merci à Julien Tardieu qui a programmé cette interface et qui nous a autorisé à l'utiliser dans le cadre de cette étude.

gains appliqués par les 8 participants sont corrélés de manière hautement significative pour quasiment tous les couples de sujets.

TABLE 6.1 – Corrélations de Pearson entre les gains attribués par chaque couple d’auditeurs lors du pré-test d’égalisation en fonction du niveau sonore perçu

	Sujet 2	Sujet 3	Sujet 4	Sujet 5	Sujet 6	Sujet 7	Sujet 8
Sujet 1	0,521**	0,521**	0,357*	0,517**	0,609**	0,552**	0,669**
Sujet 2		0,322*	0,148	0,534**	0,438**	0,456**	0,544**
Sujet 3			0,417**	0,431**	0,443**	0,643**	0,549**
Sujet 4				0,422**	0,542**	0,470**	0,209
Sujet 5					0,544**	0,267	0,416**
Sujet 6						0,478**	0,652**
Sujet 7							0,651**

** La corrélation est significative au niveau 0,01 (bilatéral).

* La corrélation est significative au niveau 0,05 (bilatéral).

Conclusion.

Le fait que les gains appliqués par les sujets dans l’expérience d’égalisation en fonction du niveau sonore perçu sont corrélés montre qu’il existe une certaine cohésion dans les réponses des sujets. De ce fait, les données recueillies dans ce test apparaissent comme relativement fiables. En conséquence, pour égaliser nos stimuli en terme d’intensité, nous avons appliqué à chaque fichier audio le gain moyen défini par les 8 sujets dans cette expérience. Les fichiers audio égalisés sont inclus dans l’annexe DVD 5c_3.

6.2.3 Stimulus de bruit

Création du fichier de bruit

Le paradigme expérimental consistant à diffuser de la parole dans le bruit a été largement utilisé pour des questions relatives à l'intelligibilité de la parole, notamment pour standardiser le matériel (listes de logatomes, de mots, de phrases) et la procédure (par ex. Amano *et al.*, 2009; Jansen *et al.*, 2010; Nielsen et Dau, 2009; Ozimek *et al.*, 2010, 2009; Wagener et Brand, 2005). Divers types de bruit sont utilisés à cette fin :

- du bruit artificiel et stationnaire, dont le spectre fréquentiel recouvre celui de la parole (par ex. George *et al.*, 2007) ;
- du bruit artificiel modulé en amplitude (par ex. George et Houtgast, 2006) ;
- du brouhaha, constitué par la superposition de diverses productions orales (par ex. Ozimek *et al.*, 2010; Nielsen et Dau, 2009) ;

L'utilisation d'un bruit ressemblant à du brouhaha nous a paru pertinente pour notre étude, car elle permet de rapprocher la situation expérimentale d'une situation naturelle, dans laquelle les auditeurs entendent les paroles prononcées par un locuteur au milieu d'autres conversations (*cf.* l'effet *cocktail-party* mis au jour par Cherry, 1953).

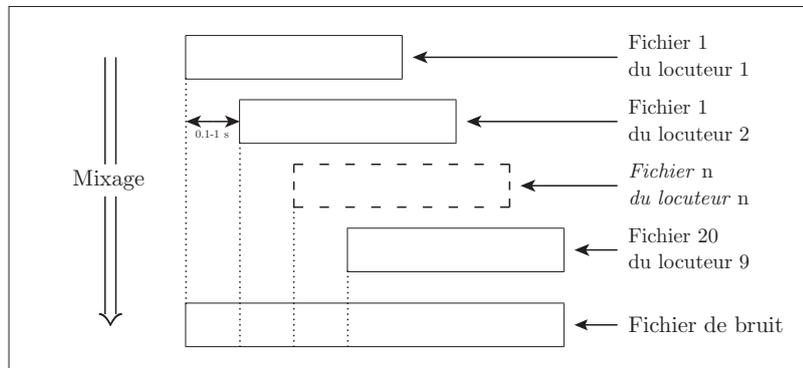
Pour créer un signal audio donnant une impression de brouhaha, nous avons utilisé 180 phrases enregistrées dans le cadre de notre seconde étude d'accord sur le nom (section 5.1.3, p. 129). Ces phrases avaient été prononcées par 9 locutrices différentes (20 phrases par locutrice). Nous avons fusionné toutes les phrases pour obtenir un fichier audio unique. Le mixage a été effectué de manière séquentielle : d'abord le fichier 1 était ajouté au fichier de bruit, puis le fichier 2, etc. jusqu'au fichier 180. Comme toutes les phrases suivaient la même structure lexico-syntaxique (*Mettez SN au-dessus/au-dessous/à droite/à gauche SP*), nous avons introduit un petit décalage temporel, aléatoirement compris entre 0.1 et 1s, avant chaque fichier. Ceci a permis d'éviter que certaines caractéristiques acoustiques des signaux de parole ne se répètent de manière cyclique⁹. De plus, pour ajouter à l'homogénéité du fichier de bruit, nous avons pris soin d'alterner entre les locutrices. Cette procédure est illustrée dans la figure 6.5. Enfin, nous avons supprimé les 5 premières secondes du fichier ainsi créé, afin d'éliminer la partie du fichier dans laquelle le brouhaha commence à se former et dans laquelle il est encore possible de saisir certains mots.

Pour mettre en œuvre cette procédure, nous avons créé un script écrit en langage *Shell* et utilisant l'utilitaire *SoX* (*Sound eXchange*¹⁰). L'algorithme général de fonctionnement du script est donné dans en annexe D.4, et le script complet est inclus en annexe DVD 5c_4.

9. ceci était en effet le cas pour les consonnes constrictives /s/ et /ʃ/, qui se retrouvent souvent au milieu des énoncés, dans les prépositions « en dessous », « au-dessus », « sur », « sous », et « à gauche ».

10. <http://sox.sourceforge.net/>

FIGURE 6.5 – Création du fichier de bruit (brouhaha)



Étalonnage des différents niveaux de bruit

Nous avons choisi de définir 10 niveaux de bruits différents, couvrant le spectre allant d'une intelligibilité quasi nulle à une intelligibilité quasi parfaite. Nous avons donc commencé par déterminer les valeurs minimales et maximales de bruit correspondant à ces deux extrêmes. Nous avons demandé à trois chercheurs de notre laboratoire d'écouter plusieurs phrases diffusées dans le bruit. Les sujets étaient situés à un mètre des haut-parleurs, et avaient pour consigne de répéter les phrases entendues. Pour chaque phrase nous faisons varier le niveau de bruit jusqu'à parvenir à identifier les seuils en deçà et au-delà desquels les phrases n'étaient pas du tout intelligibles ou bien, au contraire, totalement intelligibles. Nous avons ainsi identifié qu'un écart d'environ 13 dB séparait le niveau de bruit pour lequel l'intelligibilité était quasi nulle, du niveau de bruit pour lequel l'intelligibilité était quasi parfaite.

En conséquence, pour définir nos 10 niveaux de bruit nous avons procédé de la manière suivante. Nous avons choisi comme fichier de base le fichier de bruit dont le niveau sonore correspondait au seuil minimal d'intelligibilité. Ensuite, nous avons peu à peu abaissé le niveau du signal de bruit, en suivant des pas de -1,5 dB, pour obtenir 9 autres signaux de bruit. Ainsi, nous avons créé au total 10 fichiers de bruit, dont le niveau variait de 0 dB à -13.5 dB par rapport au fichier original. Les 10 fichiers correspondant sont inclus dans annexe DVD 5c_5.

6.3 Population et procédure

6.3.1 Population

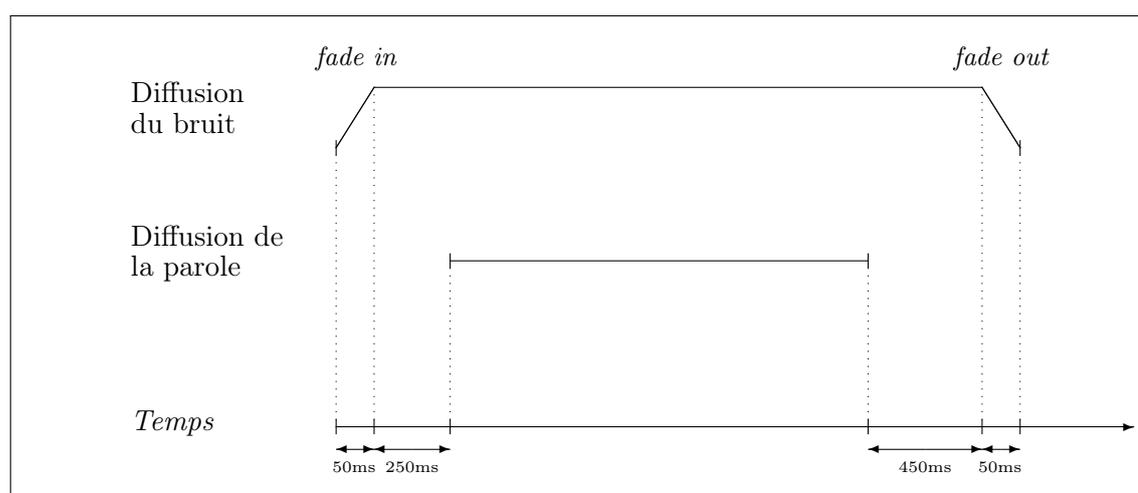
40 sujets adultes francophones natifs ont participé à cette étude. Aucun d'entre eux n'a déclaré souffrir d'un problème d'audition. Les sujets étaient âgés en moyenne de 22 ans (écart-type de 3,85 ans). Des données complémentaires sur les participants (sexe, région d'origine, milieu rural/urbain et profession) sont disponibles en annexe D.1 (p. 299).

6.3.2 Procédure générale

Chaque auditeur a effectué le test individuellement, dans la cabine audiométrique PETRA-2. Ils étaient assis à une distance d'un mètre de deux haut-parleurs *Tannoy Precision 6D*, et en face d'un moniteur couleur d'une taille de 20 pouces.

Les tâches d'intelligibilité et de compréhension étaient administrées successivement pour chaque participant. Pour chacune des tâches, les auditeurs ont entendu 50 phrases diffusées dans le bruit, une à la fois. Contrairement au niveau de présentation des phrases qui était constant, le niveau de bruit variait à différents degrés (niveaux allant de 1 à 10) correspondant aux pas de 1,5 dB prédéfinis. De plus, nous avons fait en sorte que la diffusion du bruit se commence et se termine progressivement (*fade in* et *fade out*), sur des périodes de 50 ms. Cela permet d'éviter qu'un « clic » audible n'apparaisse lors de l'émission du bruit (Amano *et al.*, 2009).

FIGURE 6.6 – Déroulement chronologique de la diffusion de la parole et du bruit pour chaque stimulus



Enfin, les phrases étaient diffusées avec une latence de 300 ms par rapport au bruit, pour capter l'attention des participants. Le déroulement chronologique de la diffusion de bruit et de phrases est illustré dans la figure 6.6.

Nous avons fait l’hypothèse que l’ordre entre les tâches d’intelligibilité et de compréhension affecterait les scores des participants, en particulier pour les scores d’intelligibilité. En effet, comme la tâche de compréhension fournit un contexte (les images), nous avons pensé que cela aiderait les participants pour la tâche d’intelligibilité. Ainsi nous avons fait en sorte que la moitié des participants effectuent la tâche de compréhension en premier, et que l’autre moitié commence par la tâche d’intelligibilité – d’une manière similaire à ce qui a été réalisé dans l’étude de Hustad (2008). Cela a conduit à la constitution de 4 groupes de 10 participants, dont la moitié (groupes 1 et 3) ont commencé par le test de compréhension. D’autre part, comme certaines phrases étaient plus susceptibles d’être affectées par le bruit que d’autres, l’allocation des niveaux de bruit en fonction des phrases a été contrebalancée pour chaque groupe de 10 participants (table 6.2).

TABLE 6.2 – Attribution du niveau de bruit en fonction des phrases, pour un groupe de 10 participants

Phrases	Niveaux de bruit									
	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj. 8	Suj. 9	Suj. 10
1 à 5	1	2	3	4	5	6	7	8	9	10
6 à 10	2	3	4	5	6	7	8	9	10	1
11 à 15	3	4	5	6	7	8	9	10	1	2
16 à 20	4	5	6	7	8	9	10	1	2	3
21 à 25	5	6	7	8	9	10	1	2	3	4
26 à 30	6	7	8	9	10	1	2	3	4	5
31 à 35	7	8	9	10	1	2	3	4	5	6
36 à 40	8	9	10	1	2	3	4	5	6	7
41 à 45	9	10	1	2	3	4	5	6	7	8
46 à 50	10	1	2	3	4	5	6	7	8	9

6.3.3 Test d’intelligibilité

Consignes

Les auditeurs avaient pour consigne d’écrire chaque phrase qu’ils entendaient. Si jamais ils n’entendaient pas un mot, ils étaient encouragés à le deviner. Avant le test les auditeurs étaient familiarisés avec la tâche en écrivant une dizaine de phrases. La consigne complète est donnée en annexe D.5.1 (p. 305).

Calcul des scores

Pour chaque phrase, l’intelligibilité était calculée comme le nombre (0, 1, 2 ou 3) de *mots informationnels* (c.-à-d. les noms correspondant à l’objet 1, à la position, et

à l'objet 2) *correctement retranscrits par l'auditeur*. Ainsi, pour la phrase « Mettez le pantalon à gauche de la fleur » :

- si l'auditeur avait écrit « Mettez le *pantalon* », le score était de 1 ;
- si l'auditeur avait écrit « à *gauche* de la *fleur* », ou « *pantalon... fleur* » le score était de 2 ;
- si l'auditeur avait écrit « le *pantalon* à *gauche* de la *fleur* », le score était de 3.

De plus, comme cela est habituellement le cas dans les tests d'intelligibilité (voir par ex. Hustad, 2008; Keintz *et al.*, 2007; Watson et Schlauch, 2009), nous n'avons pas tenu compte des erreurs orthographiques qui ne portaient pas à conséquence du point de vue de la forme sonore des mots (par ex. *pentalong* pour *pantalon*) : ce type de retranscription était jugé comme correct.

6.3.4 Test de compréhension en situation de communication

Consignes

Comme nous l'avons spécifié, les scores de compréhension de la parole étaient obtenus *via* l'interface du logiciel EloKanz. Pour chaque stimulus, six images étaient affichées à l'écran (*cf.* figure 6.7). Les auditeurs devaient répondre à chaque phrase en agissant comme requis. Par exemple, après avoir entendu la phrase « Mettez le pantalon à gauche de la fleur », les auditeurs étaient censés cliquer sur l'image montrant un pantalon et la déplacer à gauche de l'image représentant une fleur.

FIGURE 6.7 – Interface de l'auditeur



Les images cibles et les distracteurs étaient sélectionnées parmi les 110 images pour lesquelles nous avons trouvé un accord sur le nom à 100%. Comme pour la

tâche d'intelligibilité, les auditeurs étaient familiarisés avec la tâche en écoutant et en réagissant à un jeu de 10 phrases. La consigne complète est donnée en annexe D.5.2.

Calcul des scores

Pour chaque phrase, le score de compréhension (0, 1, 2 ou 3) était calculé automatiquement en vérifiant les actions effectuées par l'auditeur et ensuite en comptant le nombre d'éléments (objet 1, position, objet 2) qui correspondaient à la commande orale. Pour illustrer la procédure, nous donnons quelques exemples d'action effectuées par un auditeur en réponse à la phrase « Mettez le zèbre à gauche de la fleur », et les scores correspondant :

- l'auditeur a cliqué sur l'image représentant un *zèbre* : 1 point ;
- l'auditeur a déplacé une image autre que le zèbre, à *gauche* d'une image autre que la fleur : 1 point (la position est correcte) ;
- l'auditeur a déplacé une image autre que le zèbre, à *gauche* de l'image représentant une *fleur* : 2 points (la position est correcte, le second objet aussi) ;
- l'auditeur a déplacé l'image représentant le *zèbre*, à *gauche* de l'image représentant une *fleur* : 3 points.

Dans la suite du document, nous présentons les scores d'intelligibilité et de compréhension sous la forme de pourcentages de réponses correctes, tel que cela est habituellement effectué dans la littérature.

6.4 Résultats

L'ensemble des données recueillies durant cette étude sont contenues dans différents fichiers donnés en annexe DVD 2c. Elles sont accompagnées d'un fichier expliquant en détail leur contenu (fichier *Lisez_moi.txt*).

6.4.1 Scores d'intelligibilité et de compréhension en fonction du niveau de bruit

Afin de comparer les scores d'intelligibilité et de compréhension, nous avons tout d'abord procédé à un test t de Student pour échantillons appariés. Les résultats sont donnés dans la table 6.3¹¹.

TABLE 6.3 – Différences de moyenne entre les scores d'intelligibilité et de compréhension

Statistiques pour échantillons appariés

Variabes	Moyenne	N	Écart-type	Erreur standard moyenne
Score_Intelligibilite	53,41%	1997	32,77%	0,73%
Score_Comprehension	71,84%	1997	29,22%	0,65%

Corrélations pour échantillons appariés

Variabes	N	Corrélation	Sig.
Score_Intelligibilite & Score_Comprehension	1997	0,61	0,000**

Test pour échantillons appariés

Variables	Différences appariées					t	ddl	Sig.
	Moyenne	Écart-type	Erreur standard moyenne	Intervalle de confiance 95% de la différence				
				inf.	sup.			
Score_Int ^e – Score_Comp ^o	-18,43%	27,53%	0,62%	-19,64%	-17,22%	-29,91%	1996	0,000**

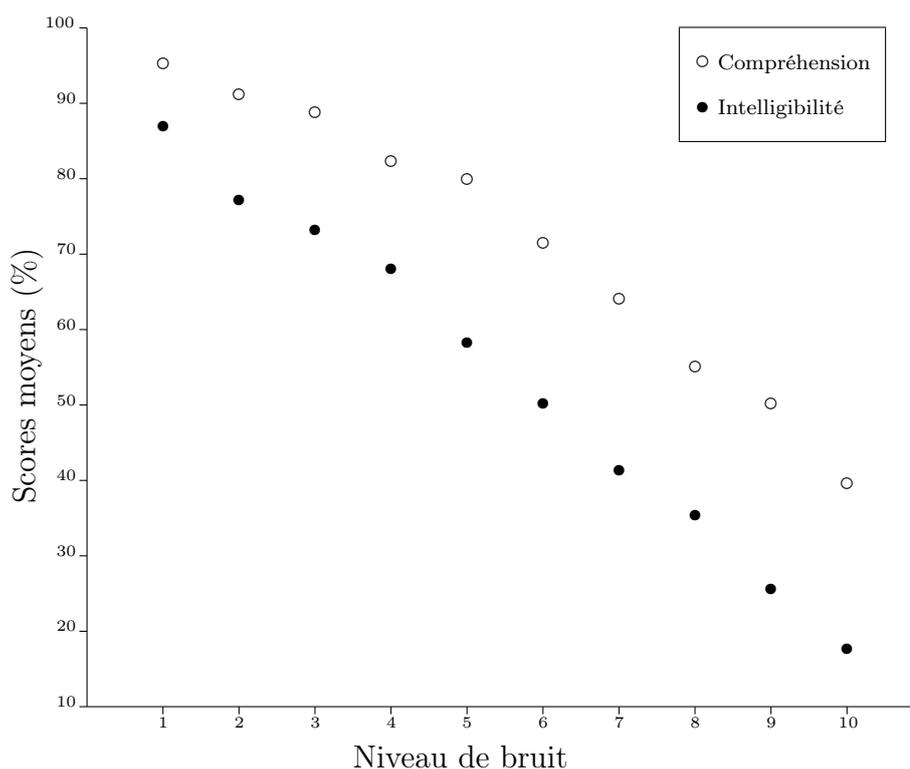
Le calcul des corrélations effectué dans le cadre du test t met au jour une relation hautement significative ($P < 0,001$) et assez forte ($r = 0,61$) entre les scores

11. Dans cette table le nombre N d'observations pour chaque tâche (intelligibilité et compréhension) est de 1997 : un dysfonctionnement du logiciel a causé la perte de 3 observations, dont le total serait autrement de 2000.

d'intelligibilité et de compréhension. Cependant ce calcul ne tient pas compte de l'influence du niveau de bruit, une variable corrélée avec les deux scores. Pour le moment nous nous contentons d'observer les différences entre les scores d'intelligibilité et de compréhension. L'analyse de la corrélation entre les scores est reportée dans la section 6.4.2.

D'autre part les scores de compréhension apparaissent comme significativement différents de ceux d'intelligibilité ($P < 0,001$). Plus précisément, les scores de compréhension sont supérieurs à ceux d'intelligibilité : 71,84% contre 53,41% en moyenne. La figure 6.8 présente l'évolution des scores moyens d'intelligibilité et de compréhension en fonction du niveau de bruit.

FIGURE 6.8 – Évolution des scores d'intelligibilité et de compréhension en fonction du niveau de bruit



L'écart-type est, en moyenne, moins élevé pour les scores de compréhension que pour ceux d'intelligibilité : 29,22% contre 32,77%. Dans la figure 6.9, nous avons représenté les écarts-types pour chaque niveau de bruit ; leur valeur a été divisée par 3, et seule leur partie supérieure est représentée pour éviter les chevauchements entre les écarts-types correspondant aux scores d'intelligibilité et ceux correspondant aux scores de compréhension. Les valeurs brutes sont données dans le tableau 6.4. Nous

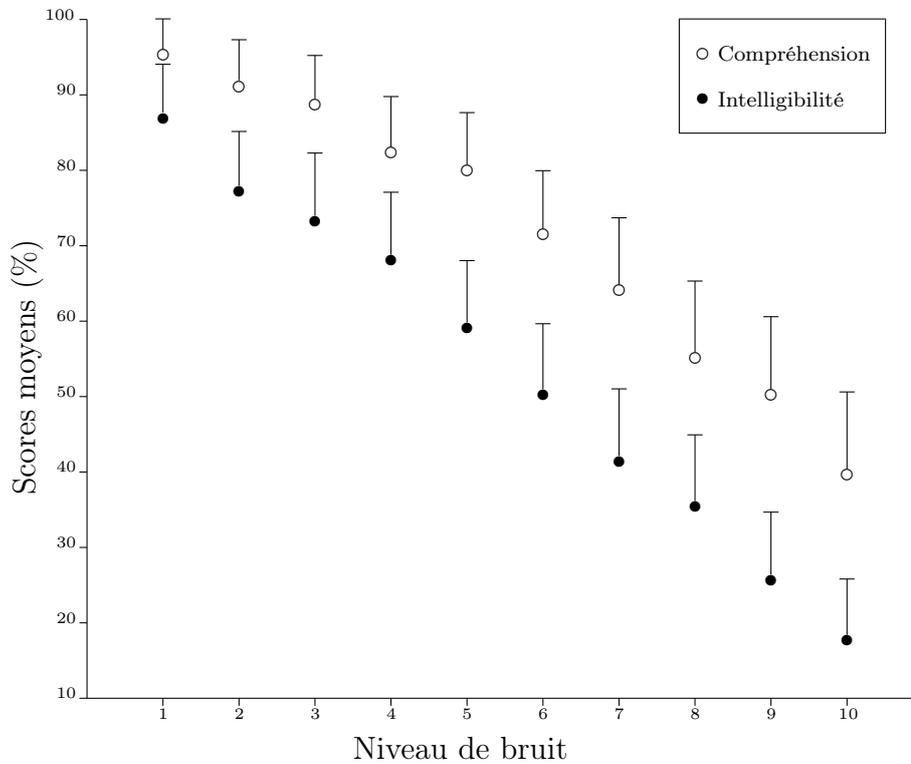
observons dans ce tableau que, contrairement aux valeurs d'écart-type des scores d'intelligibilité qui augmentent puis diminuent en fonction du niveau de bruit, la dispersion des scores de compréhension s'accroît de manière constante en fonction du niveau de bruit ¹².

TABLE 6.4 – Écarts-types des scores (pourcentages) d'intelligibilité et de compréhension en fonction du niveau de bruit

Niveau de bruit	1	2	3	4	5	6	7	8	9	10
Écart-type des scores d'intelligibilité	19.1	21.5	25.0	25.0	24.7	26.1	26.8	26.4	25.0	22.3
Écart-type des scores de compréhension	12.1	16.2	17.1	20.0	20.1	23.0	26.5	28.1	28.8	30.7

12. dans les graphiques et tableaux qui suivent, le niveau de bruit que nous indiquons est une variable ordinaire correspondant au rang du niveau de bruit, de 1 à 10. Si cela suffit à illustrer la variation des scores et des écarts-types d'intelligibilité et de compréhension, il faut toutefois rappeler que c'est en réalité le *rapport signal sur bruit* qui influe sur les scores d'intelligibilité et de compréhension.

FIGURE 6.9 – Évolution des scores d’intelligibilité et de compréhension en fonction du niveau de bruit, avec écart-type



De plus, comme les figures 6.8 et 6.9 le suggèrent, l'écart entre les scores moyens d'intelligibilité et de compréhension semble s'accroître en fonction de l'augmentation du niveau du bruit. Pour illustrer ce phénomène d'une manière plus parlante, nous avons calculé les équations de régression pour chacune des deux mesures.

Nous avons obtenu les deux équations suivantes :

- avec x le niveau de bruit, et y le score moyen d'intelligibilité :

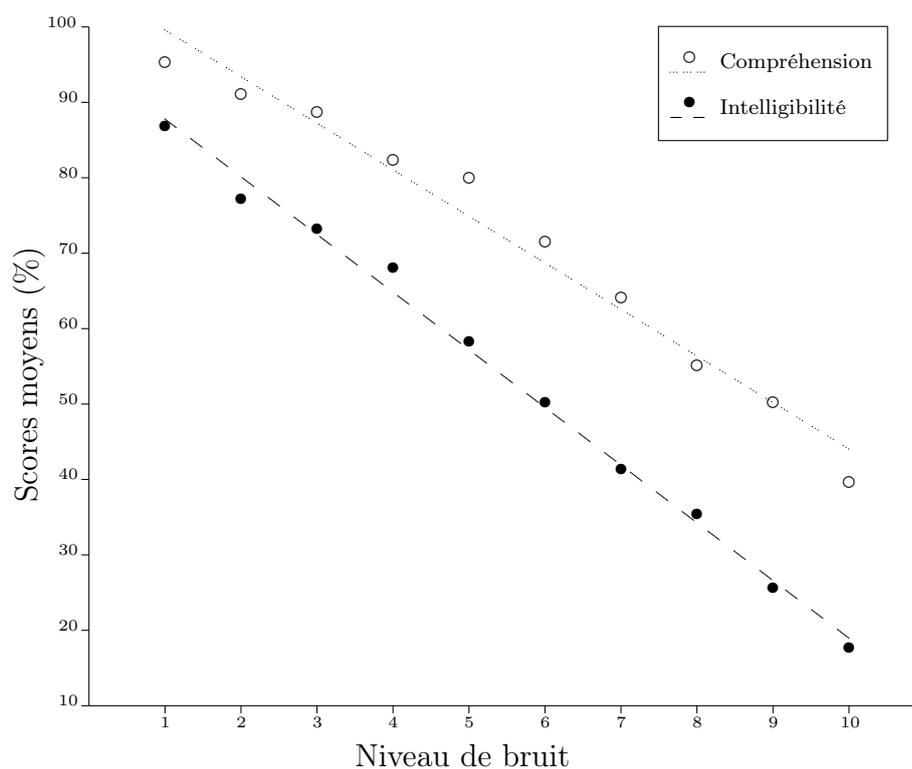
$$y = -0,76x + 9,55$$

- avec x le niveau de bruit, et y le score moyen de compréhension :

$$y = -0,62x + 10,58$$

Nous avons représenté les deux droites correspondant à des deux équations dans la figure 6.10. La figure montre effectivement que, plus l'intégrité du signal de parole est dégradée, plus la différence entre les scores d'intelligibilité et de compréhension s'accroît.

FIGURE 6.10 – Scores d'intelligibilité et de compréhension en fonction du niveau de bruit, avec les droites de régression



6.4.2 Relation entre les scores d'intelligibilité et de compréhension

Pour étudier la force d'association entre les scores d'intelligibilité et de compréhension, nous avons calculé une corrélation partielle tenant compte de la variabilité associée aux différents niveaux de bruit. La corrélation résiduelle est hautement significative ($P < 0,001$), même si sa force demeure relativement modérée, voire faible ($r = 0,346$). Le tableau 6.5 résume les coefficients de corrélation de Pearson trouvés pour les stimuli associés avec chaque niveau de bruit. Toutes ces corrélations sont hautement significatives ($P < 0,001$).

TABLE 6.5 – Coefficients de corrélation entre les scores d'intelligibilité et de compréhension en fonction du niveau de bruit

Niveau de bruit	r
1	0,26
2	0,33
3	0,29
4	0,35
5	0,29
6	0,39
7	0,26
8	0,39
9	0,47
10	0,37

6.4.3 Fiabilité des mesures

L'accord inter auditeurs a été mesuré en calculant la corrélation entre les scores obtenus par les 20 premiers participants (groupe 1) et les 20 derniers (groupe 2). Les participants des deux groupes ont été exposés aux mêmes stimuli, et ont effectué les deux tâches dans le même ordre. Pour mesurer les accords intra auditeurs, 7 participants ont repassé les tests, avec exactement la même procédure. Ce second test a eu lieu entre 9 et 13 semaines après le premier test ($moy. = 11,1$; $écart-type = 1,5$).

Les résultats, présentés dans le tableau 6.6, indiquent que les corrélations sont d'une plus grande force pour l'intelligibilité que pour la compréhension. En outre, pour les deux tâches la variabilité inter auditeurs est supérieure à la variabilité intra auditeurs.

TABLE 6.6 – Accords intra et inter auditeurs pour les scores d'intelligibilité et de compréhension

	<i>Intelligibilité</i>	<i>Compréhension</i>
Corrélations intra auditeurs	0,75	0,61
Corrélations inter auditeurs	0,70	0,55

6.5 Discussion

6.5.1 De la force d'association entre les scores d'intelligibilité et les scores de compréhension

Comme nous l'avons relaté dans notre état de l'art sur la relation entre intelligibilité et compréhension de la parole (*cf.* section 3.4, p. 80), les résultats relevés dans les précédentes études ont montré qu'il existait soit une relation très forte entre les mesures (Beukelman et Yorkston, 1979), soit au contraire une corrélation quasi nulle et non significative entre les deux types de scores (Hustad, 2008). Contrairement à ces résultats, notre étude a mis au jour une relation hautement significative ($P < 0,001$) mais de force relativement faible ($r = 0,346$) entre les scores de retranscription orthographique et les scores de compréhension.

Selon nous, cette différence de résultats pourrait s'expliquer par deux principaux facteurs. Tout d'abord, dans notre étude nous avons contrôlé le taux de dégradation du signal de parole par des mesures objectives du rapport existant entre l'intensité des stimuli et l'intensité du bruit diffusé. Ceci nous a permis d'éviter que la variable « dégradation du signal de parole », qui entretient certainement une relation à la fois avec les scores d'intelligibilité et les scores de compréhension ne vienne parasiter le calcul de corrélation entre les deux mesures. Ce contrôle avait fait défaut dans l'étude pionnière de Beukelman et Yorkston (1979). En conséquence, en ôtant de notre calcul de corrélation l'influence de cette variable parasite, nous nous attendions à trouver une force de corrélation inférieure à celle mise au jour par Beukelman et Yorkston. D'autre part, nous pensons que le contrôle de cette variable dans l'étude de Hustad (2008) avait peut-être, au contraire, *masqué* la relation existant entre les scores d'intelligibilité et de compréhension. En effet dans son étude Hustad a mesuré la dégradation du signal de parole en utilisant un test d'intelligibilité comparable au test utilisé pour mesurer l'intelligibilité. La relation entre la variable « dégradation du signal » et les scores d'intelligibilité était donc peut-être trop forte pour que l'on retrouve une corrélation significative entre les mesures d'intelligibilité et de compréhension. En contrôlant la dégradation du signal *de manière objective*, nous pensions trouver une relation *significative* entre les scores d'intelligibilité et de compréhension, ce que notre étude confirme.

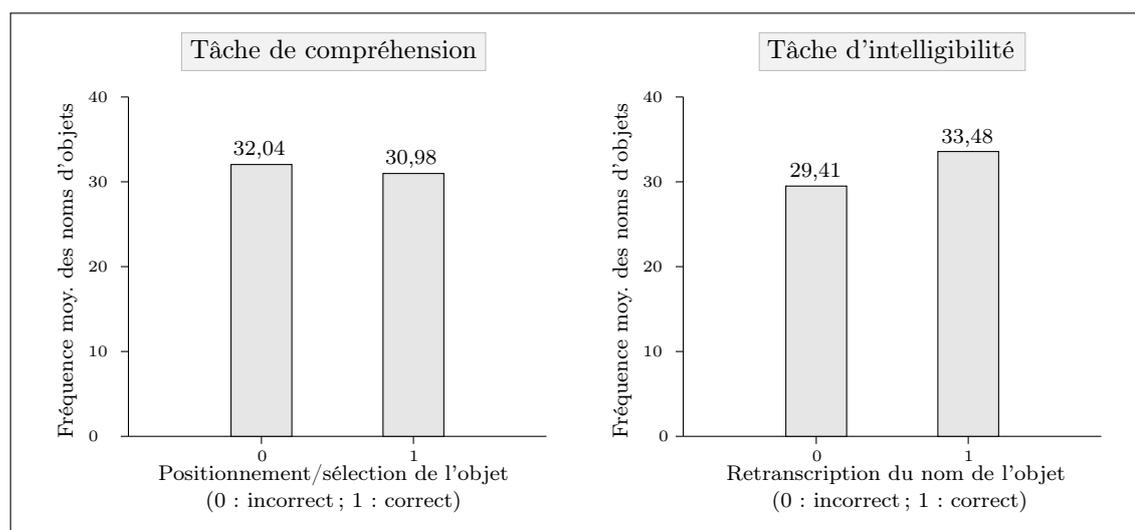
Un second élément pouvant expliquer nos résultats concerne la tâche de compréhension que nous avons employée. Notre test portant sur les actions qu'effectue l'auditeur en réponse aux énoncés, la présence d'un contexte physique (les images) était nécessaire. La présence de ce contexte a certainement contribué au fait que les scores d'intelligibilité et de compréhension n'évoluent pas totalement de la même manière au gré des conditions de notre étude.

Par exemple, Decker *et al.* (1958) avaient montré que dans un test d'intelligibilité pour lequel les auditeurs connaissaient l'ensemble de items possibles (c'est-à-dire dans une tâche de sélection d'items), la fréquence des mots utilisés comme stimuli ne jouait pas de rôle particulier – contrairement aux effets que l'on observe habituellement dans des tâches de retranscription orthographique. Il est possible que cela soit également le cas pour notre étude, c'est-à-dire que la fréquence des mots ait joué un

rôle dans la tâche de retranscription mais pas dans la tâche de compréhension. Ceci pourrait expliquer, en partie, la faiblesse de la corrélation trouvée entre les scores d'intelligibilité et de compréhension.

Pour vérifier cette hypothèse, nous avons relevé la fréquence de chacun des noms utilisés dans les 50 énoncés de cette étude. Pour ce faire, nous avons consulté la base *Lexique 3*, qui fournit les données de fréquence pour 135 000 mots. Nous avons trouvé cette ressource particulièrement intéressante car, pour approximer la fréquence d'occurrence des mots dans les interactions humaines, les auteurs de la base ont fait le choix d'utiliser les fréquences d'occurrence des mots dans des sous-titres de films. Ceci permet de dépasser de nombreux problèmes liés à la constitution de grands corpus oraux (pour plus de détails voir New *et al.*, 2007). La figure 6.11 illustre les fréquences moyennes¹³ des noms lorsque les tâches de retranscription orthographique et d'action sur les objets ont été correctement effectuées par les sujets ou pas.

FIGURE 6.11 – Moyenne de la fréquence des noms (par million d'occurrences), en rapport avec les tâches de compréhension et d'intelligibilité



À la vue de ce graphique, il semble *a priori* que l'effet est plus important dans la tâche d'intelligibilité que pour la tâche de compréhension, les noms pour lesquels la retranscription était correcte ayant une fréquence en moyenne plus élevée que les autres noms. Pour vérifier la significativité de cet effet, nous avons conduit des tests de variance univariée. Pour étudier les effets sur la tâche de retranscription, nous avons utilisé comme variable dépendante le fait que le nom ait été correctement retranscrit ou pas. Pour étudier les effets sur la tâche de compréhension, nous avons

13. en nombre d'occurrences des lemmes par million de mots. Les lemmes correspondent à l'ensemble des formes fléchies. Ainsi la fréquence du lemme « citron » cumule les fréquences des formes au singulier (« citron ») et au pluriel (« citrons »).

utilisé comme variable dépendante le fait que l'objet correspondant au nom ait fait l'objet d'une action correcte (sélection ou déplacement) ou pas. Dans les deux cas, nous avons utilisé comme facteur le niveau de bruit, et comme covariable la fréquence des noms.

Le calcul met au jour une relation statistiquement significative ($P = 0,036$) entre la fréquence des noms et la variable indiquant si la retranscription était correcte ou non. Il convient de dire que ce résultat a été obtenu avec seulement *la moitié* des données de notre test d'intelligibilité : nous n'avons pas inclus les sujets qui avaient effectué la tâche de compréhension *avant* la tâche d'intelligibilité, car ces derniers avaient vu les images avant de participer au test d'intelligibilité. Il est possible qu'avec davantage de données la relation que nous avons mise au jour ici se révèle significative à un plus haut degré.

En revanche, un test de variance univariée ne révèle aucune relation significative ($P = 0,495$) entre la fréquence des noms et le fait que les auditeurs aient correctement réagi aux énoncés ou pas. Il semble donc y avoir une différence au niveau de l'influence de la fréquence des noms dans les deux tâches, ce qui pourrait en partie expliquer pourquoi nous trouvons une relation plutôt faible entre les scores d'intelligibilité et de compréhension dans notre étude¹⁴.

Un second élément lié à la présence du contexte dans la tâche de compréhension pourrait également être déterminant pour cette question. Il s'agit du fait que les contextes imagés accompagnant chaque stimulus n'ont peut-être pas le même « poids » selon les images qui les constituent et selon le stimulus qu'ils accompagnent. Certains contextes pourraient être plus facilitants que d'autres¹⁵, introduisant par là une certaine variabilité dans les scores de compréhension.

Cependant, répondre à cette question est extrêmement complexe. En effet, pour pouvoir attribuer un « poids » à un contexte de communication, il est nécessaire d'identifier précisément les éléments qui sont utiles pour la compréhension de l'auditeur, et de savoir dans quelle mesure chacun contribue à cette même compréhension. Or cette question-là est loin d'être simple, comme nous l'avons vu dans notre état de l'art sur la question de la compréhension de la parole (*cf.* section 3.2, p. 59). Par ailleurs, notre réflexion à ce sujet ne se limite pas au travail présenté dans cette partie. Comme nous le verrons dans le chapitre suivant, cette question a aussi des conséquences sur la fiabilité de la méthode employée pour mesurer la compréhension de la parole. C'est pourquoi nous proposons de traiter cette question dans la partie 7.4 (p. 220) consacrée à cet aspect.

14. il pourrait être également intéressant de vérifier le rôle qu'a pu jouer la fréquence des syllabes (tout au moins celle des premières syllabes des noms cibles).

15. nous pouvons imaginer par exemple que si les distracteurs correspondent à des noms phonétiquement très éloignés de ceux associés aux images cibles, le contexte sera plus facilitant que si les noms des distracteurs et des images cibles ont au contraire une forme très proche.

6.5.2 Des différences entre les scores d'intelligibilité et les scores de compréhension

Différences de moyenne

La première différence qui apparaît lorsque l'on observe les scores de compréhension et d'intelligibilité révélés par notre étude (*cf.* par exemple la figure 6.8), c'est la supériorité des scores de compréhension. Les statistiques décrivant les deux variables sont données dans la tableau 6.7. Dans nos données il existe une différence moyenne de 18,4% entre les scores de compréhension et les scores d'intelligibilité.

TABLE 6.7 – Scores d'intelligibilité et de compréhension : statistiques descriptives

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Moyenne</i>	<i>Écart type</i>
Scores de compréhension	2000	0%	100%	71,83%	29,22%
Scores d'intelligibilité	1997	0%	100%	53,41%	32,77%

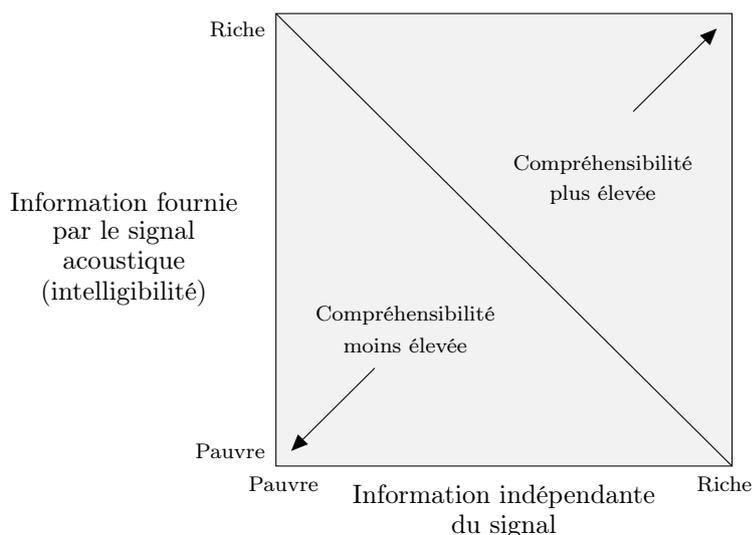
Sur ce critère, nos résultats rejoignent ceux de l'étude de Hustad (2008) sur les différences entre l'intelligibilité et la compréhension de la parole pathologique. Hustad trouve en effet des scores de compréhension significativement supérieurs à ceux d'intelligibilité, et ce pour toutes les conditions de son étude. Hustad interprète cette différence comme le signe que, dans la tâche de compréhension qu'elle a proposé (*cf.* section 3.4.1, p. 80), les auditeurs étaient amenés à puiser dans leurs connaissances sur le monde en général pour donner un sens aux paroles entendues – alors que dans la tâche d'intelligibilité l'attention des auditeurs étaient bien davantage tournée vers la *forme* du message. Si l'on interprète d'une manière similaire nos résultats, dans notre cas ce ne sont pas les connaissances des auditeurs sur le monde qui peuvent avoir bénéficié aux auditeurs dans la tâche de compréhension. En effet, dans notre cas les éléments cruciaux pour la compréhension des phrases (objets et positions) n'étaient ni déductibles du contexte phrastique dans lequel ils apparaissaient, ni déductibles des énoncés précédents. Dans notre étude la source additionnelle d'informations présente dans la tâche de compréhension est *le contexte* constitué par le jeu d'images présenté à l'écran. En effet celui-ci a pour principal effet de *circonscrire* l'ensemble des référents possibles du discours – avec tous les bénéfices que l'on peut supposer pour la perception et la compréhension du contenu linguistique des messages.

Dans tous les cas, il est important de prendre en compte cette différence qui existe entre des tâches tournées vers la perception des unités phoniques et des tâches se rapprochant davantage de situations de communication réelles. En effet dans toute situation de communication l'auditeur peut s'aider de ses connaissances sur le monde en général, mais aussi des connaissances qu'il a de la situation de communication (contexte, intentions des interlocuteurs, connaissances partagées, etc. – *cf.* section 3.2.2, p. 71) pour interpréter le sens d'un message linguistique. C'est l'idée même de

compréhensibilité, telle que l'ont définie Barefoot *et al.* (1993) : « l'acuité avec laquelle un auditeur comprend des énoncés produits par un locuteur *dans un contexte de communication* »¹⁶. Ainsi, comme l'ont souligné Barefoot *et al.* (1993) et Yorkston *et al.* (1996), dans le cas des mesures de la compréhensibilité de la parole le point de vue est *fonctionnel, opérationnel*. S'il existe des différences significatives entre les résultats obtenus par des tests de compréhensibilité et des tâches d'intelligibilité, alors la validité des scores d'intelligibilité en tant qu'indicateurs des performances du locuteur dans un contexte de communication est fortement compromise.

Par exemple, nos résultats ont montré que la différence de moyenne entre les scores d'intelligibilité et de compréhension était d'autant plus élevée que le niveau de bruit était important (*cf.* figure 6.10). Ainsi, pour la condition dans laquelle le niveau de bruit était au plus bas, la différence entre les deux scores n'est que de 8,41%, alors que dans la condition où le bruit était le plus fort cette différence est de 22 %. Selon nous, ce phénomène pourrait être lié à la façon dont les auditeurs ont pris en compte le contexte présent dans notre tâche de compréhension. En effet, nous pouvons émettre l'hypothèse que plus l'intégrité du signal est compromise par la diffusion du bruit, plus les auditeurs vont utiliser les informations fournies par le contexte pour répondre correctement aux stimuli. C'est ce que postule par exemple le *mutuality model* de Lindblom (1990b), selon lequel plus un signal de parole est dégradé, plus le processus de compréhension de la parole devient dépendant des informations présentes *en dehors* du signal de parole (*cf.* figure 6.12).

FIGURE 6.12 – Contribution de l'information acoustique et de l'information indépendante du signal pour la compréhensibilité du locuteur dysarthrique (adapté de Yorkston *et al.*, 1996; Lindblom, 1990b)



16. « the extent to which a listener understands utterances produced by a speaker in a communication context » (Barefoot *et al.*, 1993, p. 31).

Des études ont mis au jour ce phénomène pour certaines caractéristiques sémantiques et/ou structurelles des énoncés (par ex. leur longueur et leur prédictibilité, *cf.* Beverly *et al.*, 2010). Notre étude semble montrer l'existence d'un même phénomène pour ce qui est des informations fournies par le contexte de communication, c'est-à-dire par la présence d'un ensemble de référents du discours (les images) connu des auditeurs.

Cette évolution de la contribution des informations extra acoustiques pour la compréhension de la parole en fonction des conditions de dégradation du signal n'est pas seulement intéressante du point de vue théorique. Les implications cliniques sont également importantes. Comme l'ont souligné Yorkston *et al.* (1996), ce phénomène indique que ce sont les locuteurs pathologiques *les plus sévèrement atteints* qui pourraient bénéficier le plus des apports fournis par les informations contextuelles. C'est ce que montre également l'étude de Hustad (2008), qui révèle une différence plutôt réduite entre les scores de compréhension et d'intelligibilité lorsque les locuteurs souffrent de TPPP légers, que ce n'est le cas pour les cas plus sévères. Une des conséquences cliniques directes pourrait être que les stratégies visant à améliorer la communication en accompagnant la parole pathologique d'indices contextuels (*cf.* par exemple Hustad et Beukelman, 2001) devraient certainement être mises en place *en priorité pour cas de TPPP les plus sévères*.

De la variabilité des deux mesures

Même si l'écart type des scores de compréhension est plus élevé que celui des scores d'intelligibilité (29,2% contre 32,8%), dans notre étude les scores de compréhension semblent sujets à une plus grande variabilité que les scores d'intelligibilité – ce qui rejoint les résultats de Beukelman et Yorkston (1979). En effet les accords inter et intra juges sont supérieurs pour la tâche d'intelligibilité que pour la tâche de compréhension. Que ce soient pour les corrélations intra ou interjuges, les scores sont supérieurs d'environ 15% dans le test d'intelligibilité. De plus, comme cela est souvent le cas, la variabilité inter individuelle est plus forte que la variabilité intra individuelle (env. 5% de différence, pour l'intelligibilité comme pour la compréhension).

Cette différence ne peut être expliquée par la variabilité du contexte présent dans la tâche de compréhension, car nous avons pris soin de garder les mêmes contextes imagiers pour chacun des 4 groupes de participants, et pour les post-tests destinés à étudier les accords intra auditeurs. En considérant que le contexte est demeuré stable, une hypothèse pourrait être que les différents auditeurs (et dans une moindre mesure *chaque auditeur dans des circonstances différentes*) n'ont pas utilisé de la même manière les informations fournies par le contexte. Peut-être que cela est le signe de stratégies différentes (par ex. *étudier d'abord le contexte puis écouter la phrase et agir*, vs. *écouter la phrase puis observer le contexte et agir*), ou bien que l'intérêt que chaque sujet a porté au contexte était variable¹⁷. Par ailleurs nos don-

17. pour pouvoir mieux cerner cet effet, il aurait certainement été pertinent d'effectuer des pré-tests visant à évaluer la capacité de chacun des auditeurs à effectuer des inférences.

nées montrent que la variabilité des scores de compréhension s'accroît en fonction de la dégradation du signal. Ceci pourrait être le signe que plus les conditions d'écoute deviennent difficiles, plus les différences de stratégies d'écoute deviennent marquées.

Il serait intéressant d'étudier cette question des stratégies employées par les auditeurs en utilisant une technique comme l'*eye-tracking*, qui permettrait d'avoir des indications à la fois quantitatives et qualitatives sur la façon dont les auditeurs prennent en compte le contexte dans un tel type de tâche.

6.6 Conclusion

Parmi l'ensemble des résultats que nous avons discutés, il nous paraît particulièrement important de retenir les différences que nous avons trouvées entre les scores obtenus par une tâche de retranscription et ceux obtenus par une tâche de compréhension en contexte. La corrélation entre les deux mesures, bien que hautement significative, s'est révélée être de force relativement faible. Et ce, même si les deux tâches reposaient sur les mêmes stimuli, et que la façon dont nous avons compté les scores était très similaire dans les deux tests, comme nous le rappelons ci-dessous :

- pour le test d'intelligibilité, nous avons compté si les noms correspondant à l'objet à déplacer (1 point), à la position (*dessous, dessus, droite, gauche* – 1 point) et à l'objet cible (1 point) étaient correctement retranscrits par les auditeurs ;
- pour le test de compréhension en contexte, nous avons compté si l'objet à déplacer était sélectionné (1 point) puis déplacé à la position (*dessous, dessus, droite, gauche* – 1 point) par rapport à l'objet cible (1 point).

Malgré cette similitude entre les deux procédures, il est important de réaliser que d'après l'indice de corrélation que nous avons trouvé entre les deux scores ($r = 0,346$) *il ne serait pas possible de prédire de façon satisfaisante les scores de compréhension en contexte à partir des scores de retranscription*. Les implications de cette assertion sont importantes. Comme l'ont dit Beukelman et Yorkston (1979) lorsqu'ils étudiaient pour la première fois la relation entre l'intelligibilité et la compréhension de la parole pathologique, « si les scores d'intelligibilité ne sont pas étroitement liés aux taux de transfert d'information (c.-à-d. à la *compréhension*) ou si la relation entre les deux mesures est variable [...] alors l'utilisation de tests d'intelligibilité pour estimer la fonctionnalité communicative globale est discutable »¹⁸. Si l'on désire évaluer la fonctionnalité de la communication orale chez des patients, il serait donc nécessaire de développer de nouveaux tests orientés vers cette dimension particulière. Un test comme EloKanz pourrait constituer une solution appropriée. Il a été en effet conçu dans l'objectif d'étudier les performances d'un locuteur dans une véritable situation de communication. De plus, durant la saisie de nos données nous nous sommes aperçu que, du côté de l'auditeur, le temps nécessaire pour la passation d'un test de retranscription (passation + vérification et calcul des scores) était plus de 5 fois supérieur au temps nécessaire pour la passation d'un test *via* l'interface *auditeur* d'EloKanz¹⁹.

Cependant, il ne va pas de soi qu'un test tel qu'EloKanz soit praticable en milieu clinique. En effet EloKanz nécessite l'utilisation de matériel particulier (logiciel et périphériques), et implique pour les sujets – locuteurs et auditeurs – d'avoir des comportements différents de ceux engagés par des tests d'intelligibilité usuels. Dans

18. « If intelligibility scores are not closely related to information transfer or if the relationship is variable [...], the use of intelligibility scores as an overall estimate of communicative function is questionable. » (Beukelman et Yorkston, 1979, p. 190).

19. pour 60 phrases, un test de retranscription nécessitait 42 minutes et 48 secondes en moyenne, contre 8 minutes et 48 secondes pour le test *via* l'interface d'EloKanz.

une tâche de retranscription par exemple le rôle du patient consiste simplement à lire un texte à voix haute, alors qu'EloKanz implique l'appréhension d'animations visuelles et la verbalisation de commandes – deux tâches qui ne sont peut-être pas réalisables par tous les sujets. Nous pensons en particulier aux patients souffrant de troubles d'origine neurologique, dont les déficiences articulatoires sont souvent associées à d'autres difficultés. En conséquence, pour étudier plus en avant ces aspects, nous avons décidé de réaliser une étude exploratoire de faisabilité de la méthode EloKanz en milieu clinique.

Chapitre 7

Études de faisabilité clinique et de validation du test EloKanz

Sommaire

7.1	Introduction	179
7.2	Étude de la faisabilité côté <i>locuteur</i> : enregistrement des patients	180
7.2.1	Objectifs et hypothèses	180
7.2.2	Population	181
	Critères d'inclusion	181
	Critères d'exclusion	181
	Conclusion : patients inclus dans l'étude	182
7.2.3	Méthode	183
	Déroulement d'une passation	183
7.2.4	Résultats	185
	Performances générales des patients	185
	Facteurs influençant les performances des patients	189
7.2.5	Discussion	200
	Avant-propos : validité des durées observées en tant qu'indicateurs de difficulté des patients	200
	De la praticabilité du test en milieu clinique	201
7.3	Étude de la faisabilité côté <i>auditeur</i> : écoute des patients	205
7.3.1	Objectifs et hypothèse	205
7.3.2	Population	206
7.3.3	Méthode	207
	Pré-test pour le choix des stimuli	207
	Écoute des 6 locuteurs sélectionnés <i>via</i> EloKanz	212
7.3.4	Résultats	214
	Performances générales des auditeurs	214
	Influence de l'âge sur les performances des auditeurs	215
7.3.5	Discussion	218
	Performances des auditeurs	218

Quelques perspectives d'amélioration du test EloKanz . . .	218
7.4 Éléments de validation du test EloKanz	220
7.4.1 Objectifs et hypothèses	220
7.4.2 Résultats	222
Validité interne	222
Validité externe	226
7.4.3 Discussion	228
Validité interne de la mesure	228
Validité externe de la mesure	229

7.1 Introduction

Nous avons eu l'opportunité de réaliser cette étude exploratoire dans une structure dédiée aux troubles de la parole et de la déglutition, accueillant des patients souffrant de troubles d'origines anatomique comme neurologique. Nous relatons cette étude dans deux parties distinctes :

- dans la section 7.2 (p. 180) nous abordons la partie *locuteurs* de l'étude de faisabilité, dans laquelle nous nous sommes attaché à explorer les difficultés éprouvées par les patients lors de l'utilisation de l'interface *locuteur* d'EloKanz ;
- dans la section 7.3 (p. 205) nous traitons de la partie *auditeurs* de l'étude de faisabilité, dans laquelle nous avons étudié les performances de personnels médicaux lors de l'utilisation de l'interface *auditeur* d'EloKanz.

Enfin, pour conclure cette partie, nous questionnons la validité du test EloKanz, en comparant notamment les résultats qu'il permet de relever avec les scores d'évaluations subjectives de la compréhensibilité de patients réalisées par plusieurs médecins et un orthophoniste habitués à ce genre de procédure (section 7.4, p. 220).

7.2 Étude de la faisabilité côté *locuteur* : enregistrement des patients

7.2.1 Objectifs et hypothèses

Le principal objectif de cette partie de l'étude est d'observer si, de manière générale, des patients souffrant de troubles pathologiques de production de la parole sont capables de passer un test comme EloKanz. Ce test nécessite en effet de répondre à des consignes complexes, passant par l'appréhension visuelle d'animations, leur mémorisation et leur formulation sous la forme de commandes verbales. On peut supposer que pour de telles tâches des sujets souffrant de pathologies parfois très lourdes rencontrent davantage de difficultés que des sujets sains. Nous prévoyons de plus d'effectuer des enregistrements d'une cinquantaine de phrases, ce qui peut être lourd pour des sujets déjà potentiellement affaiblis par une pathologie. Nous pensons que la complexité et la lourdeur de l'épreuve visée feront :

- que tous les patients ne parviendront pas à compléter la session d'enregistrement ;
- que les patients auront, globalement, plus de difficulté à compléter le test que des sujets sains tels que ceux que nous avons enregistrés dans le cadre de notre seconde étude de l'accord sur le nom (*cf.* chapitre 5 page 129, annexe C.1 page 295 pour les informations sur les sujets).

Si, comme nous le pensons, tous les sujets ne parviennent pas à compléter correctement la session d'enregistrement des phrases, il sera alors intéressant d'étudier les profils des patients concernés. Nous pensons par exemple qu'un test comme le *Mini-Mental State* (*MMSE* – Folstein *et al.*, 1975) permettrait peut-être de prédire les performances des patients lors de l'utilisation d'EloKanz. Le *MMSE* contient en effet différents items permettant d'investiguer rapidement, entre autres, les capacités mnésiques et langagières des patients (*cf.* annexe E.6, p. 315). On peut également faire l'hypothèse que *l'âge* des patients joue aussi un rôle dans leur capacité à passer le test EloKanz. Enfin, il existe peut-être des différences significatives entre les performances de patients souffrant de TPPP d'origine *anatomique* et celles de patients souffrant de TPPP d'origine *neurologique* – ces derniers troubles étant souvent associés à d'autres types de difficultés cognitives ou motrices. Connaître précisément les différents profils de patients nous permettrait de définir des critères d'inclusion précis pour l'utilisation d'un test comme EloKanz.

Enfin, nous avons comme objectif de rendre compte des principaux *types d'erreurs* commises par les patients. Nous pensons en effet que différentes sortes d'erreurs vont survenir (par exemple des erreurs liées à *l'utilisation de l'interface du logiciel* et des erreurs liées aux *consignes de formulation des commandes verbales*) et que ces erreurs pourraient concerner différents profils de patients – selon leur âge ou leur pathologie, par exemple. Le relevé de ces différents types d'erreurs nous permettrait de modifier éventuellement l'environnement du test et/ou les consignes pour essayer de les prévenir.

7.2.2 Population

Dans l'optique de réaliser des tests statistiques sur nos données, nous avons choisi d'inclure 30 participants pour cette étude. Ces inclusions concernaient des patients venus dans le service dans le cadre de consultations ou de bilans. Afin de pouvoir vérifier nos hypothèses quand aux difficultés que pourraient rencontrer des patients souffrant de TPPP d'origine neurologique, nous avons fait le choix de constituer deux groupes :

- 15 patients sujets à des atteintes *structurales* ;
- 15 patients présentant des TPPP d'origine *neurologique*.

Au vu de nos objectifs, nous avons défini une liste de critères permettant de décider de l'inclusion, ou au contraire de l'exclusion d'un patient de l'étude.

Critères d'inclusion

Nos critères d'inclusion pour l'étude de faisabilité du test EloKanz étaient les suivants :

- (a) le patient présentait un trouble pathologique de production de la parole (TPPP) ;
- (b) le patient présentait un état général compatible avec les contraintes nécessaires pour passer le test ;
- (c) le patient était âgé de plus de 18 ans ;
- (d) le patient était francophone natif ;
- (e) le patient donnait son consentement éclairé (*cf.* annexe E.5 p. 314) pour participer à l'étude.

Critères d'exclusion

Nos critères d'exclusion pour l'étude de faisabilité du test EloKanz étaient les suivants :

- (a) le patient présentait certains troubles cognitifs associés au TPPP¹, par exemple :
 - (i) le patient produisait des énoncés incorrects du point de vue des niveaux linguistiques autres que phonétique : lexicale (ex. manque du mot), morphosyntaxe (ex. agrammatisme), phonologie (ex. inversion de sons, de syllabes) ;
 - (ii) le patient présentait une déficience portant sur ses capacités mnésiques, ne lui permettant pas de réaliser correctement les tâches demandées ;

1. les critères listés ici se sont révélés trop difficiles à investiguer en amont de l'étude. En conséquence nous les avons vérifiés pendant la phase d'entraînement.

- (iii) le patient présentait un trouble touchant aux capacités cognitives associées à la vision, qui influe sur sa capacité à reconnaître un objet, et à le situer dans l'espace par rapport à un point de référence (droite, gauche, haut, bas);
- (b) le patient parlait une variété de français parlée en dehors de la France métropolitaine (Suisse, Belgique, Québec...);
- (c) le patient avait une capacité visuelle trop limitée pour participer au test.

Conclusion : patients inclus dans l'étude

Nous avons inclus 30 patients (10 femmes et 20 hommes), âgés de 33 à 86 ans (*moyenne* = 63; *écart-type* = 13,22). Les profils pathologiques des participants étaient dans l'ensemble très divers, mêmes si plusieurs patients souffraient de la même pathologie (il y avait par exemple trois cas de maladie de Parkinson et deux cas de laryngectomies totales). Nous donnons l'ensemble des données relatives aux patients en annexe E.1 (p. 307). Le tableau contenu dans cet annexe présente également les scores que chaque patient a obtenu au test du *Mini-Mental State* – *MMSE* (*min.* = 20; *max.* = 30; *moyenne* = 26,9; *écart-type* = 2,55).

Par ailleurs, nous avons également inclus cinq autres patients pour vérifier dans une étude préliminaire que la consigne, la forme de la procédure et les aspects techniques ne posaient pas de difficulté majeure. Aucune difficulté n'ayant été relevée, nous ne traitons pas de ces 5 patients dans la suite de ce chapitre.

7.2.3 Méthode

Déroulement d'une passation

La passation d'un test d'enregistrement suivait 7 étapes consécutives :

1. Recueil du consentement éclairé du patient
2. Passation du test du *Mini-Mental State* (MMSE)
3. Installation du patient
4. Lecture de la consigne au patient
5. Phase d'entraînement
6. Phase d'enregistrement de 50 phrases
7. Questionnaire sur la difficulté d'utilisation du logiciel

Recueil du consentement éclairé du patient

Tout d'abord la fiche d'information sur le protocole de recherche clinique (*cf.* annexe E.4, p. 313) était lue au patient. Le patient pouvait ensuite demander un éclaircissement sur certains points, suite à quoi le formulaire de consentement éclairé (*cf.* annexe E.5, p. 314) était lu au patient. S'il acceptait de participer à l'étude, le patient signait le formulaire.

Passation du test du Mini-Mental State (MMSE)

Le patient passait ensuite le test du *Mini-Mental State*. Comme nous l'avons spécifié, ce test est communément utilisé en milieu clinique pour donner une indication de l'état des fonctions cognitives et mnésiques d'un patient. Le test se présente sous la forme d'un questionnaire contenant différentes parties relatives à l'orientation spatio-temporelle, la mémorisation, les compétences linguistiques et praxiques. Une copie du questionnaire utilisé est disponible en annexe E.6, p. 315.

Installation du patient

Le patient était ensuite invité à s'installer dans une salle dédiée au bilan de voix et de parole (*cf.* photographie en annexe E.7, p. 316). Le patient était assis devant un bureau sur lequel reposait un ordinateur portable de type *MacBook pro 15"*, sur lequel tournait le logiciel EloKanz. Contrairement aux conditions de notre expérience de compréhension de la parole dans le bruit, la taille réduite de la salle et du bureau ne permettaient pas de manipuler une souris d'ordinateur. Les utilisateurs interagissaient donc avec l'ordinateur directement *via* son pavé tactile (c.-à-d. le pavé situé immédiatement sous le clavier de l'ordinateur, voir photo p. 316).

Le patient était ensuite équipé d'un microphone avec serre-tête de marque SHURE, modèle SM10A (microphone électrodynamique unidirectionnel cardioïde). Ce microphone était placé à environ 1 centimètre de la commissure droite des lèvres du patient.

Lecture de la consigne au patient

Une fois le patient installé et équipé, la consigne donnée en annexe E.8 (p. 317) lui était lue. La lecture des consignes impliquait l'utilisation de l'interface *locuteur* d'EloKanz, pour illustrer le propos.

Phase d'entraînement

La phase d'entraînement portait sur 10 items (10 animations). Les heures de départ et de fin de l'entraînement étaient notées. Durant l'entraînement, si le patient ne répondait pas correctement aux consignes, cela lui était explicitement signalé, et des indications lui étaient données en vue de l'aider à corriger ce comportement. Par ailleurs toutes les difficultés rencontrées par le patient étaient notées sur la grille d'observation (*cf.* annexe E.2, p. 309).

La phase d'entraînement était aussi l'occasion de vérifier que les enregistrements de la voix du patient se déroulaient correctement. Entre la phase d'entraînement et la phase d'enregistrement, le bon fonctionnement du dispositif d'enregistrement était vérifié par l'écoute d'une phrase enregistrée par le patient ².

Phase d'enregistrement de 50 phrases

La session d'enregistrement portait sur 50 phrases. Les heures de départ et de fin de la session d'enregistrement étaient notées. Durant la session, les difficultés rencontrées par le patient étaient relevées sur la grille d'observation prévue à cet effet (*cf.* annexe E.2, p. 309). Contrairement à la phase d'entraînement, dans cette phase les comportements non conformes aux consignes n'étaient pas signalés au patient.

Questionnaire sur la difficulté d'utilisation du logiciel

À la fin de l'enregistrement des phrases, le patient devait répondre à la question contenue dans le questionnaire donné en annexe E.3 (p. 312) :

« Est-il facile ou difficile pour vous d'utiliser ce logiciel ? ».

Les patients devaient situer leur réponse entre 1 « Très facile » et 5 « Très difficile ».

Enfin, le patient était invité à exposer les difficultés particulières qu'il aurait pu ressentir, ou bien à formuler d'éventuels commentaires à propos du test.

2. l'écoute se faisait au casque, pour ne pas exposer le patient à sa propre voix et risquer de lui faire perdre confiance en lui.

7.2.4 Résultats

21 patients sur 30 ont réussi à enregistrer l'ensemble des 50 phrases. Les fichiers audio correspondant aux enregistrements, les données relevées sur la grille d'observation ainsi que celles enregistrées par le programme EloKanz sont disponibles en annexe DVD 2d_1. Elles sont accompagnées d'un fichier explicatif.

Par ailleurs, nous tenons à signaler que le dysfonctionnement qui survient parfois durant les enregistrements audio (*cf.* section 5.2, p. 138) a eu, durant ce test, une fréquence bien plus élevée que cela n'a été le cas durant nos autres expériences. En conséquence, nous avons noté dans les fichiers de données la fréquence d'apparition du dysfonctionnement pour chacun des patients. Puisque ce dysfonctionnement nécessite pour le locuteur d'enregistrer à nouveau la phrase, certains des résultats que nous présentons dans cette partie doivent être interprétés avec une grande prudence. On peut en effet imaginer que les durées de passation ou les évaluations de la difficulté d'utilisation du logiciel puissent avoir été influencées par la fréquence d'apparition du dysfonctionnement – fréquence qui variait pour chaque locuteur.

Dans la suite de cette partie, certains des résultats que nous présentons sont relatifs aux performances des patients durant la phase d'entraînement, d'autres à la phase durant laquelle les patients avaient pour consigne d'enregistrer 50 phrases, et enfin d'autres résultats sont propres à la passation du test en entier (entraînement + enregistrement de 50 phrases). Pour clarifier notre propos, nous tenons à indiquer dès à présent au lecteur la terminologie que nous employons pour nous référer à ces trois différents moments sur lesquels a porté notre étude :

- la **phase d'entraînement** désigne la phase durant laquelle les patients ont procédé à l'enregistrement de 10 phrases d'essai ;
- la **phase d'enregistrement** désigne la phase durant laquelle les patients ont procédé à l'enregistrement de 50 phrases ;
- la **passation du test** désigne l'ensemble du test : phase d'entraînement + phase d'enregistrement.

Performances générales des patients

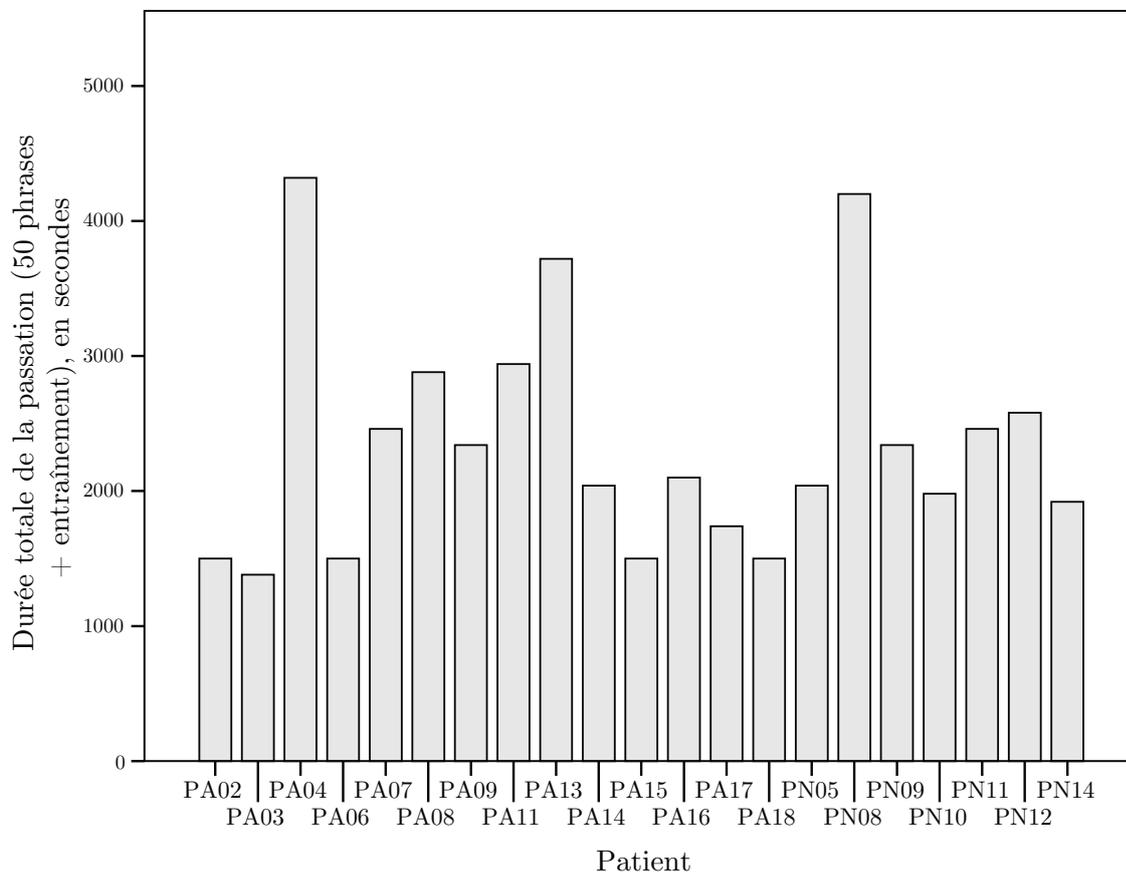
Comme nous l'avons indiqué, sur l'ensemble des 30 patients inclus dans l'étude, 21 (70%) ont réussi à compléter le test en enregistrant 50 phrases. Parmi les 9 patients qui n'ont pas réussi :

- 6 ont été exclus de l'étude après la phase d'entraînement car ils n'arrivaient pas du tout à répondre aux consignes,
- et 3 ont abandonné pendant la phase d'enregistrement.

La figure 7.1 présente les durées totales nécessaires à la passation du test pour les 21 patients qui ont été jusqu'au bout des enregistrements. Pour l'ensemble de ces patients, la durée moyenne nécessaire à la passation est d'environ 40 minutes.

La durée nécessaire à l'enregistrement *d'une phrase* est de 35,2 secondes³.

FIGURE 7.1 – Durée totale de la passation pour chaque patient



D'autre part, les éléments de non respect des consignes que nous avons relevés dans ce test peuvent être regroupés dans les deux grandes catégories que nous indiquons ci-dessous.

a) Non respect des consignes d'utilisation du logiciel :

- le patient ne clique pas sur le bouton [Ok] lorsqu'il est prêt (il clique directement sur le bouton [Parler]) ;
- le patient ne clique pas sur le bouton [Rejouer] même s'il se plaint d'avoir oublié l'animation qui a été jouée ;
- le patient ne clique pas sur le bouton [Suivant] pour passer à l'animation suivante ;
- le patient ne clique pas sur le bouton [Suivant], même s'il se plaint de pas reconnaître une image ;

3. cette valeur correspond à la durée de passation divisée par le nombre de phrases enregistrées. Dans ce calcul sont aussi comptées les données correspondant aux 3 patients qui ont abandonné pendant la phase d'enregistrement.

- le patient ne clique pas sur le bouton [Parler] avant de donner sa commande verbale ;
- le patient n’attend pas de voir que la petite roue sur le bouton [Parler] tourne pour donner sa commande verbale ;
- le patient ne clique pas sur le bouton [Stop] pour terminer l’enregistrement de sa commande verbale ;
- le patient clique sur le bouton [Stop] avant d’avoir terminé l’enregistrement de sa commande verbale.

b) Non respect des consignes de formulation des commandes verbales :

- le patient fait une phrase déclarative (ex. « L’objet X se déplace à gauche de l’objet Y ») au lieu de donner un ordre ;
- le patient fait une phrase inappropriée du point de vue pragmatique ou sémantique (ex. « C’est un bel accordéon, j’aimerais bien en jouer », « Le kangourou joue avec la quille ») ;
- le patient n’utilise pas de verbe de déplacement (le patient utilise par ex. la forme verbale « trouvez » au lieu de « mettez » ou « déplacez ») ;
- le patient utilise d’autres prépositions spatiales que celles attendues (ex. le patient utilise « à côté » au lieu de « à gauche ») ;
- le patient utilise plusieurs prépositions spatiales au lieu d’une seule pour une même commande orale (ex. le patient dit « en haut à gauche » au lieu de « à gauche ») ;
- le patient parle de l’objet immobile (l’objet cible) au lieu de l’objet qui doit être déplacé ;
- la commande verbale donnée par le patient ne correspond pas au déplacement illustré par l’animation qu’il a visionnée ;
- le patient n’utilise pas de verbe dans sa commande orale.

Du point de vue de l’ergonomie matérielle et logicielle, les données recueillies montrent que les patients ont éprouvé des difficultés par rapport aux éléments suivants :

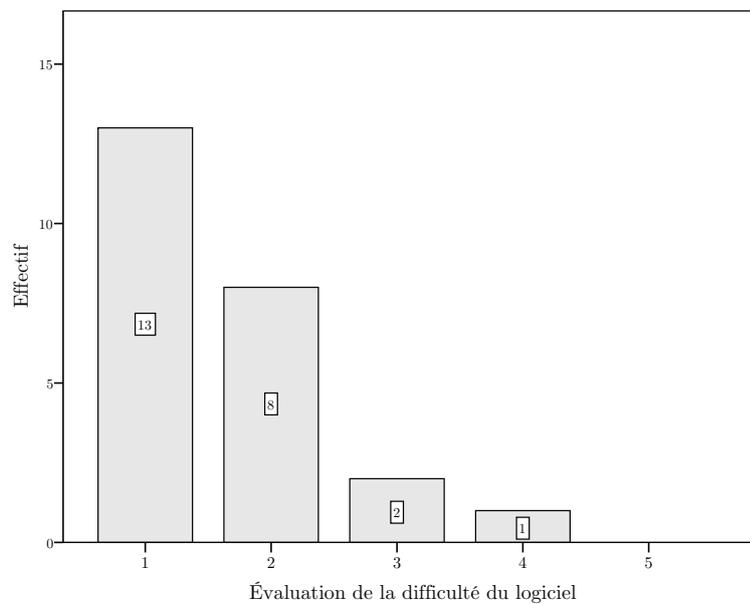
- la manipulation du pavé tactile remplaçant la souris (difficulté à cliquer ou à diriger le curseur, clics involontaires) ;
- la reconnaissance de certaines images ;
- l’utilisation des boutons du logiciel EloKanz (notamment la taille et l’éloignement du bouton [Ok] sur lequel les patients doivent cliquer pour indiquer qu’ils sont prêts – cf. figure 4.6 p. 99) ;
- la taille de certaines images ;
- le positionnement du microphone.

Parmi ces éléments, les problèmes posés par l’utilisation du pavé tactile et par la reconnaissance de certaines images ont concerné la majorité des patients (resp. 87,5% et 79% des sujets pendant la phase d’enregistrement). Ces deux problèmes apparaissent ainsi comme beaucoup plus pregnants par rapport aux difficultés causées

par l'utilisation des boutons du logiciel, la taille des images ou le positionnement du microphone (resp. 16,7%, 12,5% et 8,3% des sujets). Pour plus de détails sur les éléments ergonomiques qui ont gêné les patients, nous renvoyons le lecteur aux questions et commentaires que les patients ont effectué pendant la lecture des consignes, la phase d'entraînement et la phase d'enregistrement ainsi que lors du questionnaire post-test, qui sont disponibles en annexes E.9 (p. 319) et E.10 (p. 321).

Enfin, comme nous l'avons indiqué nous avons évalué la difficulté ressentie par les patients lors de la passation du test en leur demandant de situer cette difficulté sur une échelle de 1 (« Très facile ») à 5 (« Très difficile »). Les scores attribués par les 24 patients ayant continué le test après la phase d'entraînement sont reportés dans la figure 7.2. Comme nous pouvons le voir, la majorité (87,5%) des scores donnés par ces patients se situe entre les valeurs de 1 et 2.

FIGURE 7.2 – Évaluation de la difficulté d'utilisation du logiciel par les patients, sur une échelle de 1 (« très facile ») à 5 (« très difficile »)



Facteurs influençant les performances des patients

Nous avons fait les hypothèses que les facteurs suivants pouvaient influencer les performances des sujets locuteurs lors de l'utilisation d'EloKanz :

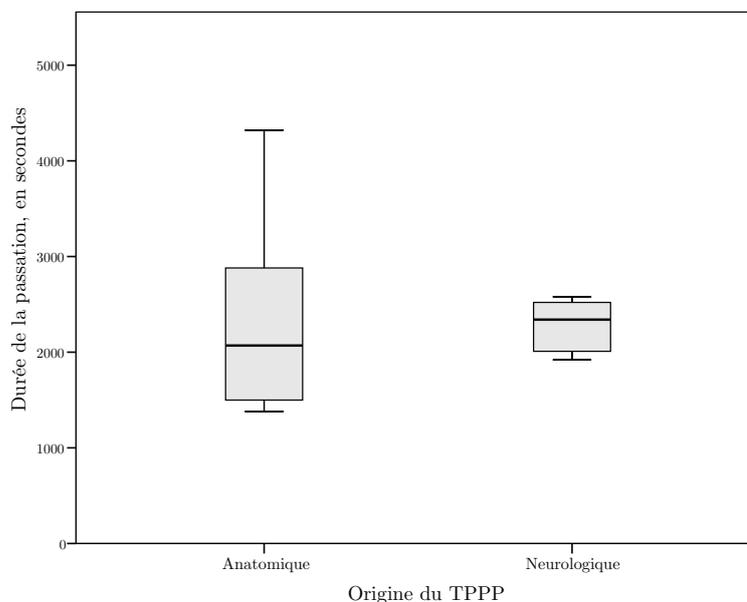
- l'origine du TPPP dont souffre le patient : anatomique ou neurologique ;
- l'état des fonctions cognitives du patient (tel que mesuré par le test du *Mini-Mental State* – *MMS*) ;
- l'âge du patient.

Nous allons donner dans un premier temps les graphiques et résultats d'analyses statistiques correspondant à la vérification de notre première hypothèse, qui demande principalement d'effectuer des comparaisons de moyennes entre les groupes de patients atteints de TPPP d'origine anatomique et neurologique. Dans un second temps nous nous intéressons aux deuxième et troisième hypothèses, dont la vérification demande cette fois de recourir pour une large part à des calculs de corrélations entre les scores des patients au MMSE ou l'âge des patients et différentes variables traduisant leurs performances lors de l'utilisation de l'interface *locuteur* d'EloKanz.

Différences entre les patients atteints de TPPP d'origine anatomique et neurologique

Du point de vue de la durée nécessaire à la totalité de la passation, il n'existe pas *a priori* de différence significative entre les groupes de patients souffrant de TPPP d'origine anatomique et de patients souffrant de TPPP d'origine neurologique, comme le suggère la figure 7.3.

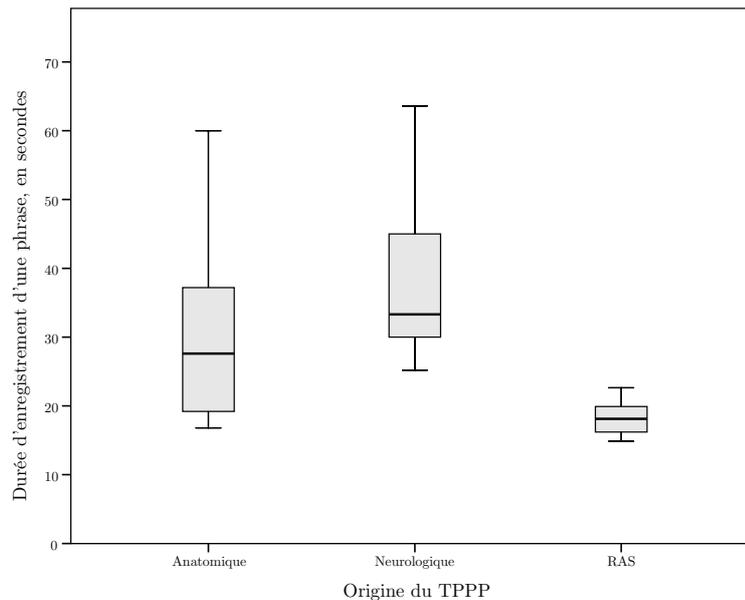
FIGURE 7.3 – Durée de la passation selon l'origine des TPPP



Les moyennes des deux groupes sont en effet très proches, et les valeurs comprises entre le premier et le troisième quartile du groupe « neurologique » sont incluses dans celles du groupe « anatomique ».

Si l'on observe la durée moyenne nécessaire à l'enregistrement d'une phrase (pendant la phase d'enregistrement) pour les deux groupes de patients (*cf.* figure 7.4), il semble exister une légère différence entre les deux populations – les patients souffrant de TPPP nécessitant en moyenne un peu plus de temps que les patients souffrant de TPPP d'origine anatomique pour enregistrer une phrase.

FIGURE 7.4 – Durée nécessaire à l'enregistrement d'une phrase selon l'origine des TPPP, et en comparaison de sujets sains (RAS)



Nous avons rajouté dans le graphique 7.4 les durées moyennes pour l'enregistrement d'une phrase relevées lors de notre seconde expérience d'accord sur le nom réalisée auprès de sujets sains (*cf.* section 5.1.3, p. 129). La différence entre les valeurs de ce groupe et des deux autres groupes est cette fois nettement plus marquée.

Pour calculer la significativité des différences que nous avons évoquées entre a) d'une part, le groupe de patients souffrant de TPPP d'origine anatomique et le groupe de patients souffrant de TPPP d'origine neurologique et b) d'autre part, le groupe de sujets sains convoqués pour notre étude d'accord sur le nom et le groupe de patients convoqué pour la présente étude, nous avons conduit deux tests de Mann-Whitney. En effet, comme nous n'avons que peu de données (24 valeurs pour la phase d'enregistrement), nous préférons utiliser un test non paramétrique. Les résultats sont donnés dans la table 7.1. La différence obtenue pour les groupes de patients souffrant de TPPP d'origine anatomique *vs.* neurologique ne sont donc pas statistiquement significatives, au contraire des différences obtenues en comparant les performances de sujets sains *vs.* les performances de sujets atteints de TPPP.

Intéressons-nous maintenant au respect des consignes du test. Les figures 7.5 et 7.6 indiquent le nombre d'éléments de non respect des consignes (erreurs vis-à-vis des consignes d'utilisation du logiciel *et* de la formulation des commandes verbales), durant les phases d'entraînement et d'enregistrement pour les deux groupes de patients.

TABLE 7.1 – Durées moyennes nécessaires à l’enregistrement d’une phrase, selon l’origine du TPPP et en comparaison de sujets sains

Rangs				
	Groupes de sujets	<i>N</i>	Rang moyen	Somme des rangs
Durée moyenne nécessaire à l’enre- gistrement d’une phrase pendant la phase d’entr ^t	TPPP Anat.	14	10,50	147,00
	TPPP Neuro.	10	15,30	153,00
	Total	24		
	– TPPP	12	8,25	99,00
	+ TPPP	24	23,63	567,00
	Total	36		

Tests pour échantillons indépendants

Variable	Critère de groupe	U de Mann-W.	Sig. bilatérale
Durée moyenne nécessaire à l’enre- gistrement d’une phrase pendant la phase d’entr ^t	Origine du TPPP	42,000	0,101
	– TPPP / + TPPP	21,000	0,000**

FIGURE 7.5 – Nombre d’éléments de non respect des consignes survenus pendant la phase d’entraînement, selon l’origine des TPPP

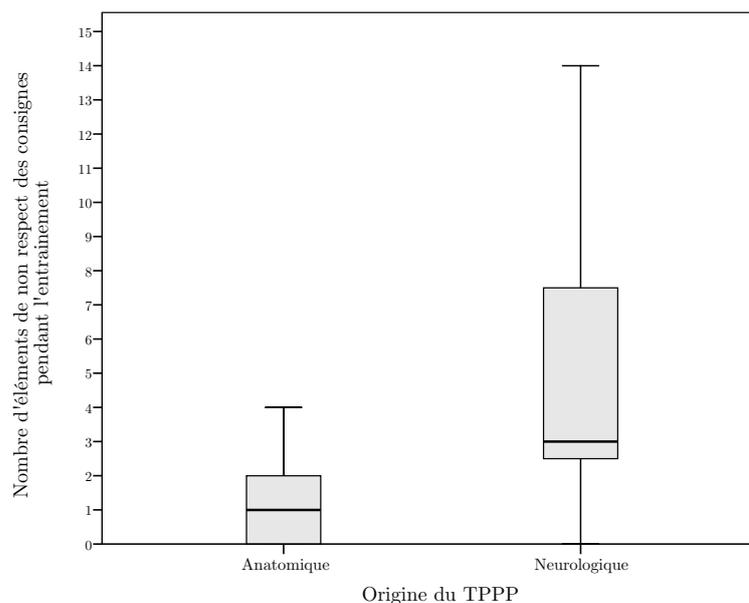
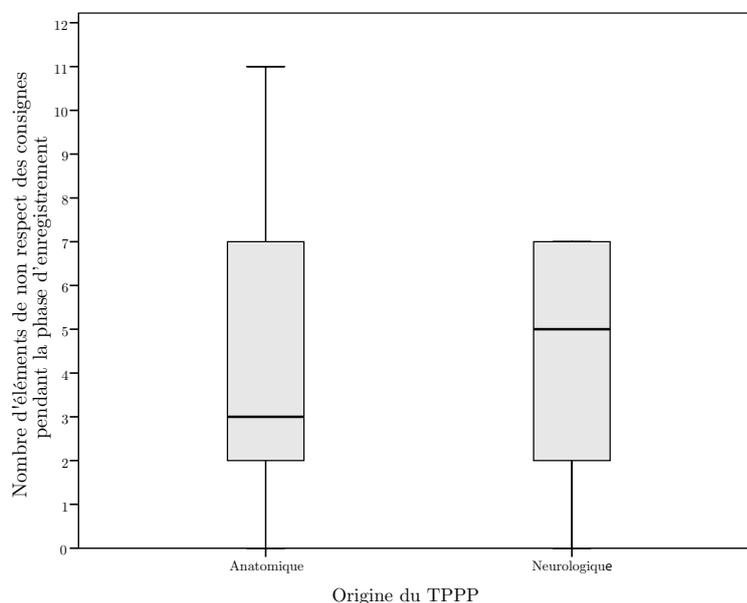


FIGURE 7.6 – Nombre d'éléments de non respect des consignes survenus pendant la phase d'enregistrement, selon l'origine des TPPP



Il semble qu'il existe une différence significative entre les deux groupes de patients pour ce qui concerne la phase d'entraînement uniquement. Nous pouvons vérifier cela en utilisant un test de Mann-Whitney pour comparer les deux échantillons de notre population, pour les deux phases du test – entraînement et enregistrement.

TABLE 7.2 – Nombre d'éléments de non respect des consignes selon l'origine des TPPP

Rangs				
	Origine du TPPP	N	Rang moyen	Somme des rangs
Nombre d'éléments de non respect des consignes pendant l'entraînement	Anatomique	15	10,50	157,50
	Neurologique	15	20,50	307,50
	Total	30		
Nombre d'éléments de non respect des consignes pendant la phase d'enreg. ^t	Anatomique	14	11,75	164,50
	Neurologique	10	13,55	135,50
	Total	24		

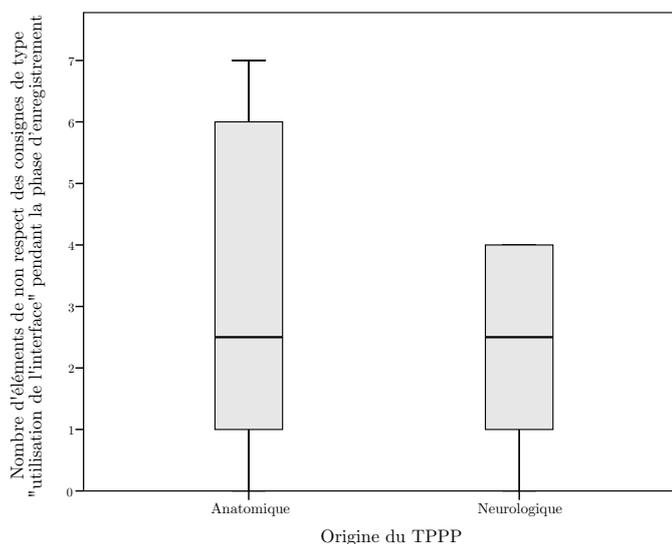
Tests pour échantillons indépendants

Variable	U de Mann-W.	Sig. bilatérale
Nombre d'éléments de non respect des consignes pendant l'entraînement	37,500	0,002**
Nombre d'éléments de non respect des consignes pendant la phase d'enregistrement	59,500	0,535

Les résultats sont indiqués dans la table 7.2. Comme nous le supposions, la différence entre les deux groupes de patients n'est pas significative pour ce qui concerne la phase d'enregistrement. En revanche, la différence entre les deux groupes est hautement significative ($P = 0,002$) pour les données relevées lors de la phase d'entraînement. D'autre part, comme nous l'avons indiqué dans la première partie des résultats, nous avons relevé des erreurs de non respect des consignes de deux différents types :

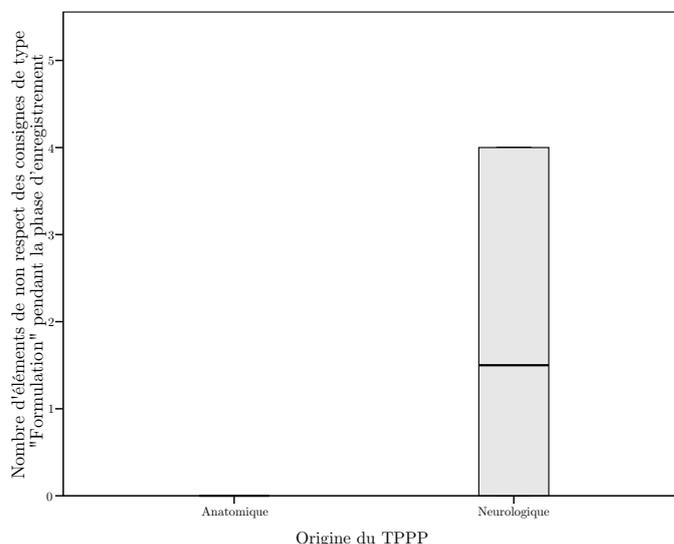
- les erreurs ayant trait à l'utilisation du logiciel EloKanz (ex. *ne pas cliquer sur le bouton adéquat au moment voulu*) ;
- les erreurs se rapportant à la formulation des commandes verbales (ex. *faire une phrase déclarative alors que la consigne est de donner un ordre verbal*).

FIGURE 7.7 – Nombre d'éléments de non respect des consignes de type « utilisation de l'interface » survenus pendant la phase d'enregistrement, selon l'origine des TPPP



Nous avons comparé l'occurrence de ces deux types d'erreurs durant la phase d'enregistrement, par les deux groupes de patients. Les graphiques correspondants sont donnés dans les figures 7.7 (erreurs d'utilisation du logiciel) et 7.8 (erreurs de formulation des commandes verbales). Il semble qu'il n'y ait pas de grande différence entre les deux groupes pour ce qui concerne les erreurs relatives à l'utilisation du logiciel EloKanz. En revanche, la différence très nette pour ce qui est des erreurs de formulation des commandes verbales.

FIGURE 7.8 – Nombre d'éléments de non respect des consignes de type « formulation » survenus pendant la phase d'enregistrement, selon l'origine des TPPP



D'autre part, comme nous l'avons spécifié lors de la présentation de l'expérience visant à tester l'interface *locuteur* d'EloKanz, le programme enregistre l'heure précise à laquelle le locuteur clique sur le bouton « parler » pour donner sa consigne orale. Il nous est ainsi possible de récupérer les temps de réaction des patients, c'est-à-dire les durées situées entre la fin de chaque animation et le moment précis où ils ont commencé à enregistrer leur commande orale. La figure 7.9 présente les temps de réaction moyens pour :

- les patients atteints de TPPP d'origine anatomique ;
- les patients atteints de TPPP d'origine neurologique ;
- les sujets sains convoqués dans notre seconde étude d'accord sur le nom (*cf.* section 5.1.3 p. 129).

Comme cela était le cas pour les durées nécessaires à l'enregistrement d'une phrase, les différences entre les deux groupes de patients sont moins marquées que les différences existant entre les sujets sains de notre étude d'accord sur le nom et la population pathologique convoquée pour cette étude. Cependant, le test de Mann-Whitney confirme que, dans les deux cas les différences sont statistiquement significatives (*cf.* table 7.3).

FIGURE 7.9 – Temps de réaction moyens avant l’enregistrement d’une phrase, selon l’origine des TPPP et en comparaison de sujets sains (« RAS »)

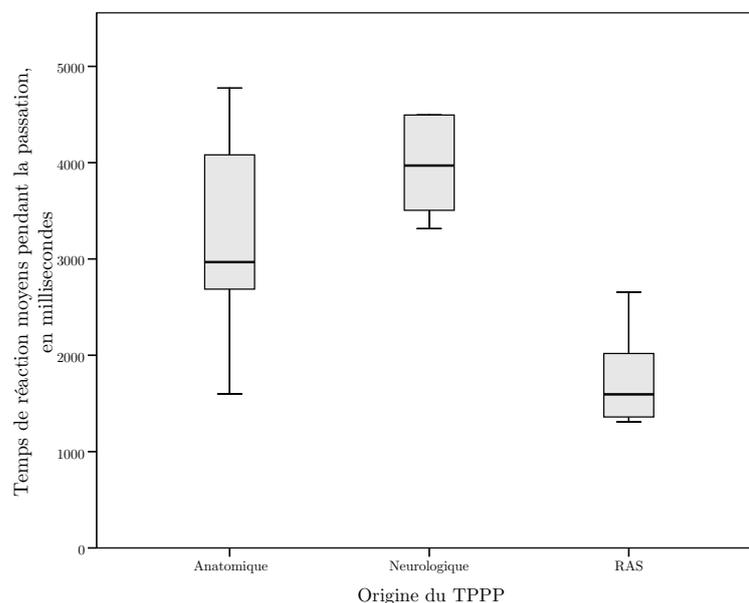


TABLE 7.3 – Temps de réaction moyens lors de l’enregistrement d’une phrase, selon l’origine des TPPP et en comparaison de sujets sains

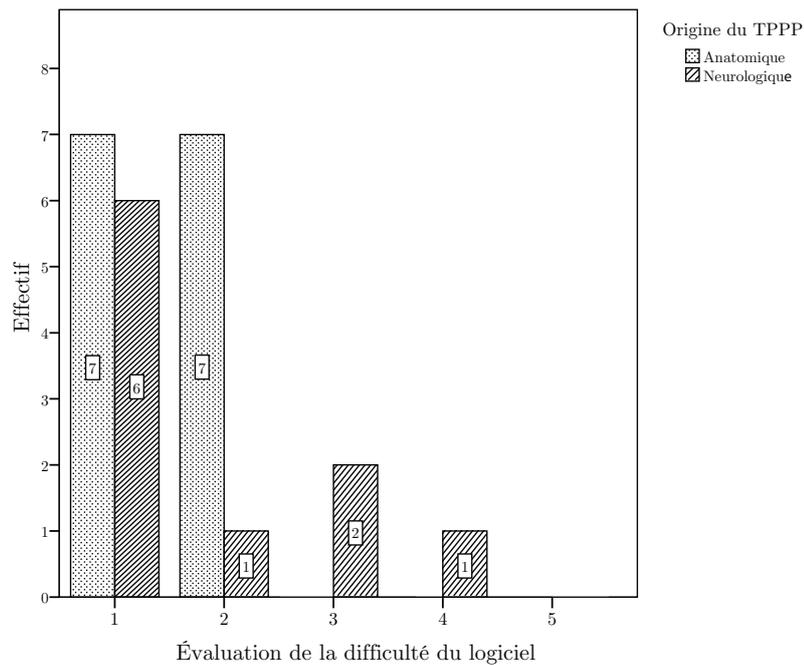
Rangs				
	Groupes de sujets	N	Rang moyen	Somme des rangs
Temps de réaction moyens lors de l’enregistrement d’une phrase pendant la phase d’enr. ^t	TPPP Anat.	14	10,00	140,00
	TPPP Neuro.	10	16	160,00
	Total	24		
	– TPPP	12	7,08	85,00
	+ TPPP	24	24,21	581,00
	Total	36		

Tests pour échantillons indépendants			
Variable	Critère de groupe	U de Mann-W.	Sig. bilatérale
Temps de réaction moyens lors de l’enregistrement d’une phrase pendant la phase d’enr. ^t	Origine du TPPP	35,000	0,040*
	– TPPP / + TPPP	7,000	0,000**

Enfin, il semble exister des différences entre nos deux populations de sujets pour ce qui concerne l’évaluation de la difficulté d’utilisation du logiciel. Comme nous

pouvons le voir sur la figure 7.10, les patients souffrant de TPPP d'origine neurologique sont les seuls à avoir attribué des notes supérieures à 2 sur l'échelle de difficulté que nous avons proposée. En revanche, le test de Mann-Whitney n'indique pas de différence significative entre les deux groupes.

FIGURE 7.10 – Évaluation de la difficulté du logiciel par les patients sur une échelle de 1 (« Très facile ») à 5 (« Très difficile »), selon l'origine des TPPP



Autres facteurs pouvant influencer les performances des sujets

Dans cette partie nous allons exposer les résultats d’analyses statistiques (corrélations principalement) qui visent à vérifier nos hypothèses selon lesquelles l’âge des patients, ainsi que l’état de leurs fonctions cognitives et mnésiques peuvent jouer un rôle dans leurs performances lors de l’utilisation de l’interface *locuteur* d’EloKanz.

Pour cela nous allons reprendre l’ensemble des variables susceptibles de servir d’indicateurs des performances des patients que nous avons traitées dans la partie précédente, à savoir :

- la durée totale nécessaire à la passation du test ;
- la durée moyenne nécessaire à l’enregistrement d’une phrase pendant la phase d’enregistrement ;
- le nombre d’éléments de non respect des consignes du test durant les phases d’entraînement et d’enregistrement ;
- les temps de réaction moyens avant l’enregistrement d’une phrase.

Comme nous l’avons dit dans l’introduction de cette partie, il est fréquemment arrivé durant les enregistrements qu’un dysfonctionnement survienne et contraigne le locuteur à enregistrer sa phrase à nouveau. La fréquence d’occurrence de ce dysfonctionnement n’ayant pas été la même pour tous les patients, cette fréquence a pu influencer sur la durée totale des passations ainsi que sur la durée moyenne nécessaire à l’enregistrement d’une phrase. Pour vérifier nos hypothèses quant à l’influence des scores MMS et de l’âge, nous avons calculé les scores de corrélations partielles tenant compte de l’influence qu’a pu avoir cette troisième variable. Les données sont présentées dans la table 7.4.

TABLE 7.4 – Corrélations partielles tenant compte de la fréquence d’apparition du dysfonctionnement des enregistrements audio

<i>(Variable de contrôle : fréquence des dysfonctionnements audio)</i>		Durée totale nécessaire à la passation	Durée moyenne nécessaire à l’enregistrement d’une phrase
Score MMS	Corrélation	-0,437*	-0,408
	Sig. (bilatérale)	0,042	0,060
	<i>ddl</i>	20	20
Âge	Corrélation	0,586**	0,493*
	Sig. (bilatérale)	0,004	0,020
	<i>ddl</i>	20	20

Nous voyons que les corrélations suivantes sont statistiquement significatives :

- Score obtenu par le patient au MMSE, et durée totale de la passation. Le coefficient est négatif, ce qui indique que plus les scores obtenus par les patients au MMSE sont bas, plus la durée nécessaire à la passation a tendance à augmenter ;

- Âge du patient, et durée totale de la passation. Le coefficient est positif, ce qui indique que plus l'âge du patient est élevé, plus la durée nécessaire à la passation a tendance à augmenter ;
- Âge du patient, et durée moyenne nécessaire à l'enregistrement d'une phrase. Ici aussi le coefficient est positif, ce qui indique que plus l'âge du patient augmente, plus la durée nécessaire à l'enregistrement d'une phrase s'accroît.

Nous avons également effectué des calculs de corrélation pour étudier la relation existant entre l'âge des patients ou leur score au MMSE et le nombre et types d'erreurs (erreurs d'utilisation du logiciel ou erreurs de formulation des commandes verbales) qu'ils ont commises. Les résultats obtenus pour les 29 patients qui ont été jusqu'au bout de la phase d'entraînement, et pour les 21 patients qui ont complété la phase de passation sont indiqués dans la table 7.5.

TABLE 7.5 – Nombre d'éléments de non respect des consignes en fonction de l'âge et des scores des patients au MMSE (corrélations)

		MMS	Âge
Nombre d'erreurs total pendant la phase d'entraînement	Corrélation	-0,531**	0,302
	Sig. (bilatérale)	0,003	0,111
	<i>N</i>	29	29
Nombre d'erreurs total pendant la phase d'enregistrement	Corrélation	-0,021	0,096
	Sig. (bilatérale)	0,927	0,680
	<i>N</i>	21	21
Nombre d'erreurs de type « utilisation » pendant l'enregistrement	Corrélation	-0,416	0,418
	Sig. (bilatérale)	0,061	0,059
	<i>N</i>	21	21
Nombre d'erreurs de type « formulation » pendant l'enregistrement	Corrélation	-0,358	-0,270
	Sig. (bilatérale)	0,112	0,237
	<i>N</i>	21	21

Comme nous pouvons le voir, seule la relation entre les scores obtenus au MMSE et le nombre total d'erreurs pendant la phase d'entraînement apparaît comme statistiquement significative. Cette corrélation est de force moyenne et de polarité négative : moins les scores au MMS sont élevés, plus le nombre d'erreurs pendant la phase d'entraînement a tendance à augmenter.

Dans nos hypothèses nous avons également indiqué que les temps de réaction des patients (durées situées entre la fin d'une animation et l'enregistrement d'une phrase) pourraient être influencés par l'état des fonctions cognitives et mnésiques des patients ou encore par leur âge. Un calcul de corrélation entre les temps de réaction d'une part, et les scores de MMS et les âges des patients d'autre part donne les résultats listés dans la table 7.6.

Toutes les corrélations mises au jour sont statistiquement significatives. Leurs forces sont moyennes. La corrélation entre les scores de MMS et les temps de réaction est de polarité négative : plus les scores au MMS sont bas, et plus les temps

TABLE 7.6 – Temps de réaction moyen lors de l’enregistrement d’une phrase en fonction de l’âge et des scores des patients au MMSE (corrélations)

		Temps de réaction moyen lors de l’enregistrement d’une phrase
Score MMS	Corrélation de Pearson	-0,433*
	Sig. (bilatérale)	0,035
	<i>N</i>	24
Âge	Corrélation de Pearson	0,531**
	Sig. (bilatérale)	0,008
	<i>N</i>	24

de réaction ont tendance à augmenter. Au contraire, dans le cas de la corrélation entre l’âge des patients et les temps de réaction les deux variables ont tendance à augmenter et à diminuer ensemble.

Enfin, nous avons voulu étudier la relation entre les scores attribués par les patients sur l’échelle de difficulté d’utilisation du logiciel, et l’âge et les scores MMS des patients. Puisque les scores donnés par les patients correspondent à une variable de type *ordinal*, nous n’avons pas calculé de corrélation de Pearson mais des corrélations de Kendall (*Tau-B*). Les résultats pour l’ensemble des 24 patients qui ont participé à la phase de passation du test sont donnés dans la table 7.7.

TABLE 7.7 – Évaluations subjectives de la difficulté du logiciel en fonction de l’âge et des scores des patients au MMSE (*Tau-B* de Kendall)

		Évaluation subjective de la difficulté du logiciel (échelle de 1 – <i>Très facile</i> à 5 – <i>Très difficile</i>)
Score MMS	Coefficient de corrélation	-0,233
	Sig. (bilatérale)	0,190
	<i>N</i>	24
Âge	Coefficient de corrélation	0,206
	Sig. (bilatérale)	0,220
	<i>N</i>	24

Comme nous pouvons le constater, aucune des corrélations que nous avons mises au jour n’est statistiquement significative.

7.2.5 Discussion

Dans cette discussion nous commençons par traiter de la validité des variables de durée (temps de réaction et durées d'enregistrement) que nous avons choisies comme indicateurs de la difficulté qu'ont les patients à passer le test EloKanz. Nous discutons ensuite des performances des patients à la lumière de ces variables, mais aussi en observant les erreurs qu'ils ont commises par rapport aux consignes d'utilisation du logiciel et de formulation des commandes verbales. Nous traitons à la fois des facteurs pouvant expliquer ces performances (âges, scores MMS des patients et origines anatomique ou neurologique des TPPP), et à la fois des solutions que nous envisagerions pour restreindre les erreurs et améliorer les performances des patients. Dans un dernier temps nous traitons des principaux problèmes posés par l'ergonomie du test, et de la possibilité d'y pallier par des solutions matérielles et logicielles, ou bien par un remaniement des consignes.

Avant-propos : validité des durées observées en tant qu'indicateurs de difficulté des patients

Lorsque nous avons comparé les temps de réaction et les durées d'enregistrement chez les locuteurs, nous avons fait l'hypothèse implicite que ces variables sont liées aux difficultés que les patients éprouvent lors de la passation du test. Nous pensons qu'il est nécessaire d'obtenir des indications quant à la validité de cette hypothèse. Pour cela nous avons effectué des calculs de corrélation entre ces variables et les évaluations que les patients ont faites de leur propre difficulté. Les résultats sont présentés dans la table 7.8.

TABLE 7.8 – Évaluation de la difficulté du logiciel par les patients en fonction de l'âge et des scores obtenus par les patients aux MMSE (*Tau-B* de Kendall)

		Évaluation subjective de la difficulté du logiciel (échelle de 1 – <i>Très facile</i> à 5 – <i>Très difficile</i>)
Temps de réaction	Coefficient de corrélation	0,375*
	Sig. (bilatérale)	0,024
	<i>N</i>	24
Durée moyenne d'enregistrement d'une phrase	Coefficient de corrélation	0,445**
	Sig. (bilatérale)	0,008
	<i>N</i>	24

Même si elles sont de force faible à modérée, les corrélations apparaissent comme statistiquement significatives, ce qui nous conforte dans le choix que nous avons fait d'observer les temps de réaction et les durées nécessaires à l'enregistrement des phrases pour traduire les difficultés posées par le test aux patients.

De la praticabilité du test en milieu clinique

De manière générale, même si les patients ont jugé l'utilisation d'EloKanz comme étant plutôt facile – voire très facile, le logiciel leur a posé de nombreux problèmes. Cela se traduit en premier lieu par le temps nécessaire à l'enregistrement des phrases, nettement supérieur à celui que nous avons observé lors de notre seconde expérience d'accord sur le nom auprès de sujets sains (*cf.* section 5.1.3, p. 129). Ces résultats doivent cependant être fortement nuancés en prenant en compte le fait que, contrairement aux sujets sains qui ont manipulé une souris, les patients ont manipulé un pavé tactile – et plusieurs d'entre eux se sont plaints de la difficulté de cette manipulation. Il est donc possible que cette différence liée au matériel utilisé explique en partie nos observations. De plus, dans notre expérience d'accord sur le nom les sujets étaient en moyenne bien plus jeunes (*moyenne* = 22,2; *écart-type* = 3,11) que les patients convoqués dans cette étude (*moyenne* = 63; *écart-type* = 13,22), et certainement plus familiers avec l'outil informatique que les patients.

Quoi qu'il en soit, si l'on inclut les phases de lecture des consignes et d'entraînement, le temps nécessaire à l'enregistrement de 50 phrases était en moyenne d'environ 40 minutes, ce qui ne paraît pas si long compte-tenu des difficultés que peuvent poser la nature de la tâche et l'utilisation du matériel informatique et du logiciel. Cela dit de grandes différences apparaissent entre les durées nécessaires à chaque sujet, la passation la plus courte ayant duré 23 minutes seulement, contre 1 heure et 12 minutes pour la plus longue. D'après nos résultats, l'âge des patients, ainsi que leur état de fonctionnement cognitif tel que mesuré par le *Mini-Mental State Examination* semblent déterminants.

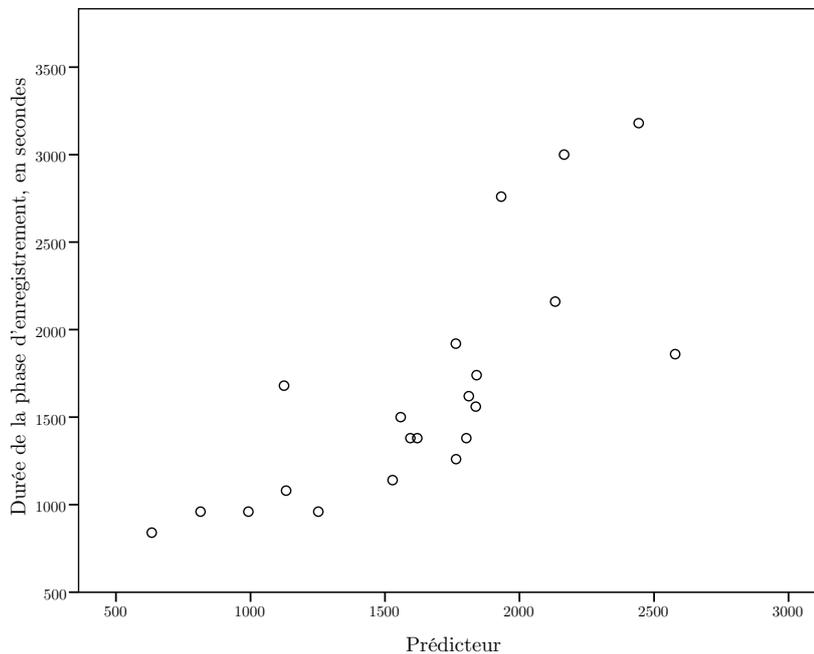
Nous pouvons donc imaginer qu'il soit possible, en amont du test, d'obtenir une approximation du temps nécessaire à l'enregistrement des phrases en tenant compte de ces deux indicateurs. Ceci permettrait éventuellement d'inclure seulement les patients pour lesquels la durée d'enregistrement resterait dans les limites acceptables imposées par l'environnement clinique. Pour vérifier cette hypothèse, nous avons effectué une régression linéaire⁴ avec comme variable dépendante la durée de la phase d'enregistrement, et comme variables explicatives le score au MMSE, l'âge du patient ainsi que la fréquence des problèmes d'enregistrement audio (variable qui a influé sur la durée d'enregistrement). L'équation ainsi obtenue est la suivante (approximation) :

$$\text{Prédicteur de durée} = 25x + 50y - 58z + 1090$$

où x est l'âge du patient, y le nombre de problèmes d'enregistrement survenus et z le score du patient au *Mini-Mental State*. Le prédicteur ainsi obtenu est corrélé de manière forte ($r = 0,76$) et hautement significative ($P < 0,001$) avec la durée nécessaire à l'enregistrement de 50 phrases. Le graphe de dispersion donné en figure 7.11 illustre la covariation des durées prédites avec les véritables durées relevées dans notre étude.

4. la linéarité de la relation est également une hypothèse. Il serait peut-être possible d'obtenir de meilleurs résultats avec d'autres types de fonctions mathématiques.

FIGURE 7.11 – Prédiction de la durée nécessaire à la phase d'enregistrement



Il paraît donc *a priori* possible de prédire avec une assez grande précision les durées nécessaires à l'enregistrement des patients. Cependant, nous pensons que nos données ne sont pas suffisantes pour effectuer ce type de calcul avec beaucoup de précision et de fiabilité. Il faudrait refaire ce type de calcul avec les données relatives à davantage de passations, et si possible dans des conditions où aucune variable comme le *nombre de dysfonctionnements audio* ne vienne parasiter les résultats. En outre, il nous paraîtrait intéressant d'étudier l'influence d'autres variables telles que la familiarité des patients avec l'outil informatique – un sujet qui apparaît dans de nombreuses remarques que les patients ont formulées.

Ensuite, nos résultats montrent que les patients ont commis de nombreuses erreurs par rapport aux consignes d'utilisation du logiciel et de formulation des commandes verbales. Beaucoup de ces erreurs sont réhibitoires pour la praticabilité de notre test puisqu'elles conduisent à des enregistrements qui ne sont absolument pas satisfaisants du point de la forme (ex. signaux de parole tronqués car le patient a commencé à parler avant de cliquer sur le bouton [Parler]) ou du fond (ex. commandes verbales ne correspondant pas aux animations qui ont été jouées aux patients). Il était donc nécessaire pour nous a) d'observer si ces erreurs se retrouvaient chez tous les profils de patients et b) de réfléchir à des moyens de prévenir ces erreurs en modifiant l'ergonomie du logiciel ou les consignes.

- a) Concernant la question des facteurs influençant les erreurs commises par les patients, nos résultats sont particulièrement difficiles à interpréter. D'une part, nous pensons que la façon dont nous avons comptabilisé les erreurs n'est peut-être pas satisfaisante : nous avons attribué le même poids pour des erreurs de types par-

fois très différents⁵, et qui ont une plus ou moins grande prédisposition à être répétées de manière systématique par un même patient⁶. De plus, la distribution des scores au MMSE dans les deux groupes de patients – TPPP neurologiques et TPPP anatomiques – n’est pas égale. Ainsi, lorsque nous trouvons une différence significative entre les deux groupes de patients pour le nombre d’erreurs commises pendant la phase d’entraînement (*cf.* table 7.2, p. 192), il est possible que cela soit dû en partie au fait que le groupe de patients aux TPPP d’origine neurologique compte davantage de sujets avec un MMSE bas que le groupe de patients aux TPPP d’origine anatomique. La relation entre les scores de MMSE et le nombre d’erreurs commises pendant l’entraînement apparaît en effet comme hautement significative (*cf.* table 7.5, p. 198). D’ailleurs, lorsque 4 patients avec un MMSE faible et appartenant au groupe neurologique sont exclus après l’entraînement, la différence entre les deux groupes n’est plus significative (*cf.* les résultats pour la phase d’enregistrement dans la table 7.2, p. 192). Pour la phase d’enregistrement où les distributions des scores de MMSE sont sensiblement équivalentes, nous trouvons néanmoins une différence nette entre les deux groupes pour ce qui est des erreurs de *formulation des commandes verbales* – les patients souffrant de TPPP d’origine neurologique faisant plus d’erreurs que les autres. Il serait intéressant de confirmer cette tendance en testant une population plus importante et dont les distributions des scores au MMSE seraient les plus proches possibles pour chacun des deux groupes concernés. Si la tendance était confirmée, cela pourrait peut-être conduire à l’élaboration de consignes de test sensiblement différentes selon l’origine des pathologies des locuteurs – les consignes adaptées aux troubles neurologiques laissant peut-être moins de liberté d’énonciation aux patients. Enfin, l’âge des patients ne semble pas jouer de rôle déterminant dans la fréquence des erreurs commises.

- b) Les erreurs que nous avons observées concernent à la fois la manipulation du logiciel et à la fois la formulation des commandes verbales. Pour ce qui est de la manipulation du logiciel, les erreurs les plus fréquentes sont aussi les plus gênantes pour la praticabilité de la méthode : les patients oublient de déclencher ou bien d’arrêter l’enregistrement. Nous pourrions essayer de corriger ce problème en insistant peut-être davantage, dans les consignes et lors de l’entraînement, sur cette manipulation qui n’est sans doute pas explicitée à la hauteur de son importance. Nous pourrions également imaginer que le programme détecte automatiquement la présence de parole sans que le bouton [Parler] n’ait été touché, auquel cas il afficherait un message d’alerte à l’écran. Les erreurs de formulation ne nous paraissent pas pouvoir être corrigées sans perdre de la spontanéité dans les comportements verbaux des patients. Par exemple, les erreurs de prépositions spatiales pourraient éventuellement être davantage contrôlées en indiquant par écrit 4 prépositions (dessus, dessous, droite, gauche) directement dans l’interface

5. par exemple « le patient fait une phrase déclarative décrivant le déplacement des objets » est comptée avec le même poids que « le patient oublie de cliquer sur le bouton [parler] ».

6. par exemple il est possible qu’un même patient répète de manière systématique une erreur concernant le verbe utilisé pour la commande verbale.

d'EloKanz.

Enfin, les problèmes d'ergonomie concernaient principalement l'utilisation du pavé tactile de l'ordinateur, la reconnaissance des images et l'utilisation des boutons du logiciel (bouton « Ok » en particulier). La principale difficulté posée par un périphérique de type *souris* vient de la dissociation entre la zone d'action (le périmètre de déplacement de la souris) et la zone de visualisation (l'écran – Lefèvre, 2010). Avec le pavé tactile s'ajoute un autre facteur de difficulté : le manque de pratique des sujets. Il nous semble qu'une solution intéressante pourrait être l'utilisation d'un écran tactile. Cela réglerait la question de la dissociation des zones d'action et de visualisation, et offrirait une solution plus naturelle pour les personnes non familiarisées avec l'outil informatique. Il est d'ailleurs arrivé pendant l'étude que certains patients touchent directement l'écran au lieu de cliquer sur les boutons à l'aide du curseur, ce qui semble montrer que chez ces sujets la manipulation directe des boutons sur l'écran serait plus naturelle. Utiliser un écran tactile éviterait aussi que les patients fassent des allers-retours entre les boutons situés en bas de l'écran et le bouton « Ok » situé au milieu de l'écran – manipulation dont se sont plaints certains patients. Nous pourrions également agrandir la taille des boutons de l'interface d'EloKanz, solution qui a été suggérée par plusieurs patients. Enfin, concernant la reconnaissabilité des images certains traits semblent revenir plus fréquemment dans les images posant des difficultés aux patients. Parmi ces éléments se trouvent en particulier la pâleur (par ex. pour le bonnet *blanc*, les boutons *blancs*, la balance *grise* et la fusée *blanche*) et le fait que les objets représentés sont de forme fine et allongée (ex. *couteau*, *râteau*, *pinceau*, *plume*). Il serait envisageable de coloriser différemment certaines images pour qu'elles se détachent mieux du fond blanc qui est utilisé dans EloKanz. Par contre, il ne nous semble pas possible d'agrandir la taille des images dont la forme est trop allongée. En effet, pour que le logiciel fonctionne les images doivent faire la même taille maximale, en hauteur ou en largeur selon l'objet représenté. Or, même si les patients ont jugé ces images « trop petites », elles font en réalité la même taille que les autres, mais comme nous l'avons dit elles représentent des objets de forme plutôt élancée (*plume*, *passoire*, *couteau*, *quille*...) ce qui fait qu'elles ont une surface bien inférieure que celle des images représentant des objets de forme plutôt carrée ou ronde – comme la tomate par exemple. La seule solution que nous envisageons pour l'instant serait tout simplement de supprimer ces images, et de les remplacer par des images représentant d'autres objets.

7.3 Étude de la faisabilité côté *auditeur* : écoute des patients

7.3.1 Objectifs et hypothèse

Comme nous l'avons indiqué en introduction, le test effectué auprès des auditeurs nous a servi à la fois à étudier *la praticabilité* du test EloKanz, et d'autre part à en étudier *la validité en tant qu'outil d'évaluation des performances de locuteurs pathologiques*. Nous traitons ici seulement de la partie *praticabilité* de l'étude, la seconde partie étant abordée quant à elle dans la section 7.4 (p. 220).

Les objectifs de l'étude de la praticabilité du test EloKanz pour les auditeurs sont :

- a) de vérifier que les auditeurs parviennent à respecter les consignes d'utilisation du logiciel et de réponse aux commandes verbales données par les patients ;
- b) d'étudier la rapidité de mise en œuvre du test, puisque cela est un facteur déterminant pour la pratique clinique (Ertmer, 2010; Metz *et al.*, 1980; Samar et Metz, 1988; Schiavetti, 1992) ;
- c) de quantifier et d'analyser les difficultés que peuvent rencontrer des auditeurs vis-à-vis de l'ergonomie du test EloKanz. EloKanz est en effet un test informatisé qui suppose entre autres la manipulation d'objets *via* un périphérique d'entrée de type souris ou pavé tactile. Or tous les individus ne bénéficient pas de la même expertise face à la manipulation de ces périphériques : nous faisons notamment l'hypothèse que l'âge pourrait être un facteur influant sur les performances des auditeurs durant l'utilisation d'EloKanz.

Dans la suite de cette partie nous commençons par présenter la population que nous avons convoquée pour l'étude (section 7.3.2, p. 206). Dans la section 7.3.3 (p. 207) nous traitons du pré-test qui nous a permis de sélectionner nos stimuli, ainsi que du déroulement des passations. Enfin dans la partie 7.3.4 (p. 214) nous exposons nos résultats avant de les discuter (section 7.3.5, p. 218).

7.3.2 Population

Notre population pour cette étude est constituée de 24 sujets francophones natifs âgés de 19 à 56 ans (*moyenne* = 32,5 ; *écart-type* = 13,4). En fonction d'hypothèses à propos de l'influence de l'expertise des auditeurs sur les scores (*cf.* plus loin section 7.4, p. 220), nous avons composé notre population de la manière suivante :

- 12 sujets ayant été exposés à de la parole pathologique. Ce groupe était composé de sujets âgés de 19 à 56 ans (*moyenne* = 44,7 ; *écart-type* = 11,3). En moyenne, les sujets de ce groupe avaient été exposés pendant 16,4 ans à de la parole pathologique (*min.* = 1 ; *max.* = 35 ; *écart-type* = 11,5) ;
- 12 sujets n'ayant jamais été exposés à de la parole pathologique. Ce groupe était composé de sujets âgés de 19 à 33 ans (*moyenne* = 22,4 ; *écart-type* = 3,9).

Le groupe de sujets exposés aux TPPP était constitué de personnels médicaux, dont les professions entraient dans la liste suivante :

- aide-soignante (4 sujets) ;
- infirmière (2 sujets : 1 étudiante, 1 diplômée d'État) ;
- médecin phoniatre (2 sujets) ;
- diététicienne (1 sujet) ;
- ergothérapeute (1 sujet) ;
- interne en médecine (1 sujet) ;
- orthophoniste (1 sujet).

Le groupe de sujets non exposés aux TPPP était constitué quant à lui uniquement d'étudiantes en première année d'orthophonie. L'ensemble des données relatives aux auditeurs inclus dans l'étude est disponible en annexe E.13, p. 328.

7.3.3 Méthode

Dans cette partie nous abordons deux aspects de la méthodologie utilisée pour cette étude :

- 1) nous relatons tout d'abord le pré-test (*cf. infra*) que nous avons conduit afin de sélectionner nos stimuli, c'est-à-dire le jeu de phrases prononcées par les patients que nous avons ensuite fait écouter aux auditeurs ;
- 2) nous décrivons ensuite (p. 212) la façon dont se sont déroulées les passations de test, c'est-à-dire que nous présentons les consignes données aux auditeurs, puis les conditions matérielles, et enfin les différentes phases des passations du test de faisabilité.

Pré-test pour le choix des stimuli

L'objectif de ce pré-test était de recueillir des données subjectives quant à la compréhensibilité des patients que nous avons enregistré dans l'expérience précédente (section 7.2, p. 180). Nous voulions à terme constituer un ensemble de stimuli dont la compréhensibilité s'étend le plus possible sur le continuum allant d'une compréhensibilité (quasi) nulle à une compréhensibilité (quasi) parfaite. Pour ce faire, nous avons conduit un test d'évaluation subjective de la compréhensibilité, afin d'identifier deux patients dont la compréhensibilité est très basse, deux patients dont la compréhensibilité est moyenne, et deux patients dont la compréhensibilité est très haute.

Population

Trois professionnels évoluant dans le domaine des TPPP ont participé à cette pré-étude. Les informations (sexe, âge, profession) les concernant sont données dans la table 7.9.

TABLE 7.9 – Participants au pré-test d'évaluation subjective de la compréhensibilité des patients

N°	Code	Sexe	Âge	Profession
1	EV1	F	49	Médecin phoniatre
2	EV2	F	49	Orthophoniste
3	EV3	F	39	Orthophoniste

Matériel et procédure

Nous avons choisi comme stimuli 220 phrases prononcées par les 22 patients qui ont totalement complété la phase d'enregistrement (c'est-à-dire ceux qui ont réussi à enregistrer 50 phrases), soit 10 phrases pour chaque patient. Chaque phrase a été sélectionnée selon les critères suivants :

- dans la phrase le patient ne fait pas d'erreur vis-à-vis des noms attribués aux images, ni de la forme verbale et des prépositions spatiales qu'il emploie ;

- les conditions d’enregistrement sont correctes : il n’y a pas, notamment, de bruit de souffle qui ne vienne couvrir la parole ;
- la phrase ne met pas en jeu les mêmes images et la même direction de déplacement qu’une autre phrase sélectionnée.

L’ensemble des fichiers audio correspondant aux 220 phrases sélectionnées est disponible en annexe DVD 5d_1.

Chacun des sujets a participé au test individuellement. Ils étaient assis dans la salle dédiée aux bilans de la voix (*cf.* photographie en annexe E.7, p. 316). En face d’eux était situé un ordinateur portable de type *Mac Book Pro 15”* relié à un boîtier de réponse *Cedrus* de type *RB-730* (*cf.* figure 7.12), et deux enceintes de marque *Yamaha*, modèle *NS-10m studio* (120 Watts). L’ordinateur était connecté aux enceintes *via* une carte son externe *Edirol FA-66* et un amplificateur *NAD Electronics 310*.

FIGURE 7.12 – Boîtier de réponse utilisé pour le pré-test d’évaluation subjective de la compréhension



D’autre part, pour cette expérience nous avons programmé une interface à l’aide du programme *Cedrus SuperLab 4.0*[©]. Cette interface, fournie en annexe DVD 5d_2, permet :

- de présenter les consignes aux sujets (*cf.* figure 7.13) ;
- de diffuser chaque stimulus, accompagné d’un écran rappelant les valeurs possibles de l’échelle et les descripteurs des valeurs extrêmes (*cf.* figure 7.14)
- de récupérer les notes attribuées par les trois évaluateurs pour chaque stimulus.

FIGURE 7.13 – Consignes diffusées pour le pré-test d'évaluation de la compréhension

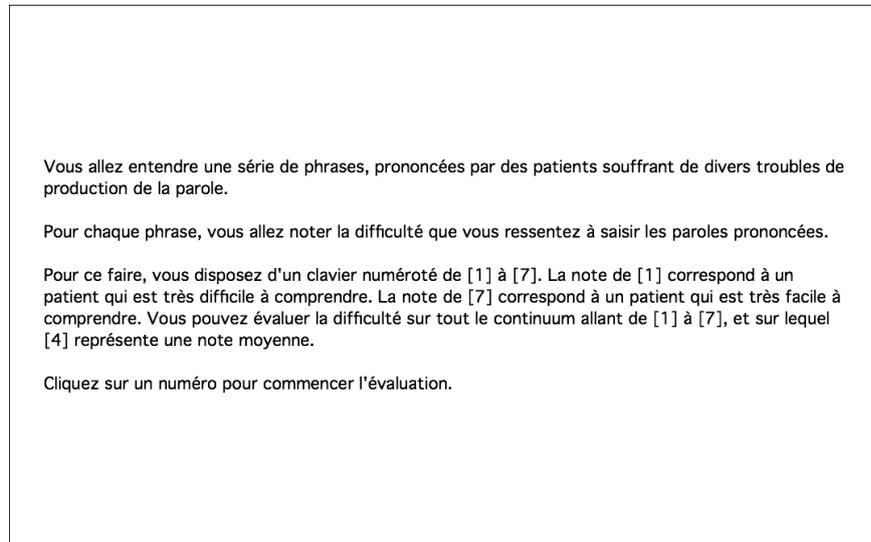
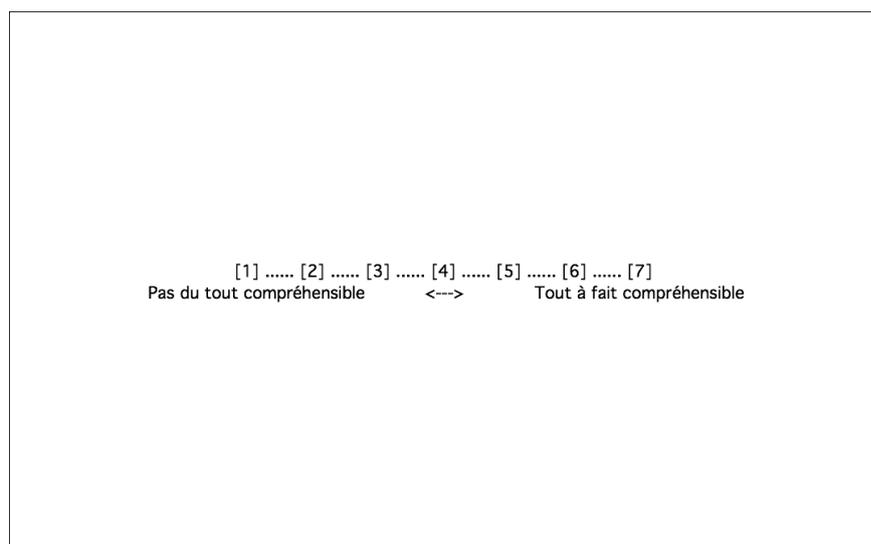


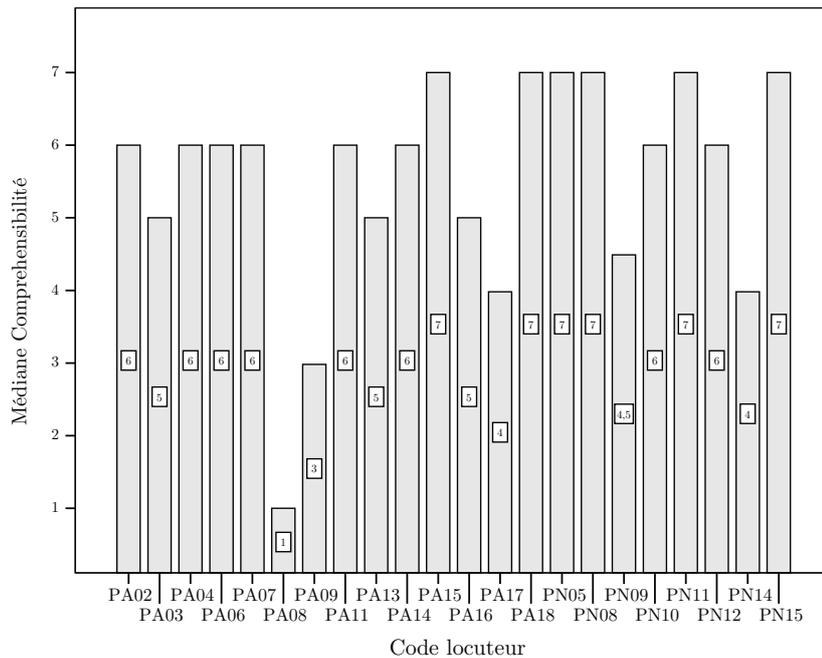
FIGURE 7.14 – Écran diffusé avec chaque phrase dans le pré-test d'évaluation de la compréhension



Résultats

L'ensemble des résultats de ce pré-test est disponible en annexe DVD 2d_2, accompagné d'un fichier explicatif. Dans la figure 7.15 nous avons représenté les médianes des notes obtenues par chacun des 22 patients.

FIGURE 7.15 – Médianes des scores de compréhensibilité obtenus par une évaluation subjective, pour chaque locuteur



Nous avons également testé la fiabilité des scores obtenus, en évaluant l'accord existant entre les notes attribuées par chacun des trois évaluateurs. Parce que les données recueillies dans ce pré-test (les scores de compréhensibilité) sont de type *ordinal*, nous avons effectué un calcul de corrélation de Kendall (*Tau-B*) entre les scores attribués par les 3 évaluateurs. Les résultats sont présentés dans la table 7.10⁷.

Conclusion

En conclusion, les évaluations effectuées semblent plutôt fiables : il existe des corrélations hautement significatives ($P < 0,01$) entre les scores attribués par chaque couple d'évaluateurs, et la force de ces corrélations va d'une valeur moyenne à une valeur élevée.

Si l'on observe les médianes des notes obtenues par chacun des patients (*cf.* figure 7.15), il apparaît que la plupart d'entre eux ont été jugés comme étant bien voire

7. le nombre de valeurs pour les accords entre EV1 et EV2, et entre EV2 et EV3 n'est que de 168 car l'évaluateur EV2 n'a pas terminé sa session de test : elle n'a noté que 168 phrases sur 220 (env. 76,4%). EV2 a cependant évalué chacun des 22 patients.

TABLE 7.10 – Accords inter juges lors de l'évaluation de la compréhensibilité des patients sur une échelle (*Tau-B* de Kendall)

Accord entre EV1 et EV2	Coefficient de corrélation	0,563**
	Sig. (bilatérale)	0,000
	<i>N</i>	168
Accord entre EV2 et EV3	Coefficient de corrélation	0,554**
	Sig. (bilatérale)	0,000
	<i>N</i>	168
Accord entre EV1 et EV3	Coefficient de corrélation	0,727**
	Sig. (bilatérale)	0,000
	<i>N</i>	220

très bien compréhensibles. Seulement deux patients ont obtenu des notes dont la médiane se situe en dessous de 4 : PA08 et PA09. Ceci a limité notre choix quant à la sélection des stimuli pour l'étude de faisabilité. Nous souhaitons en effet sélectionner autant de patients souffrant de TPPP d'origine anatomique que de patients souffrant de TPPP d'origine neurologique, pour chaque groupe de compréhensibilité (faible, moyenne, forte). Cela ne sera pas possible pour le groupe de patients dont la compréhensibilité a été évaluée comme faible.

En conséquence, nous avons sélectionné pour la suite de notre étude les locuteurs suivants :

- pour le groupe de locuteurs évalués comme *difficilement compréhensibles* : PA08 (*médiane* = 1) et PA09 (*médiane* = 3) ;
- pour le groupe de locuteurs évalués comme *moyennement compréhensibles* : PA17 (*médiane* = 4) et PN09 (*médiane* = 4,5) ;
- pour le groupe de locuteurs évalués comme *facilement compréhensibles* : PA06 (*médiane* = 6) et PN10 (*médiane* = 6).

Pour la suite de notre étude de faisabilité, nous avons utilisé les 60 phrases prononcées par ces 6 locuteurs (10 par locuteur) et que nous avons utilisées dans ce pré-test. Dans la suite de cette section, nous détaillons la procédure suivie par chaque auditeur pour procéder à l'écoute de ces 60 phrases.

Écoute des 6 locuteurs sélectionnés *via* EloKanz

Comme nous l'avons indiqué, le pré-test que nous avons présenté n'avait pour objet que de sélectionner 6 locuteurs jugés comme difficilement compréhensibles, moyennement compréhensibles et facilement compréhensibles. L'idée étant de faire ensuite écouter à notre population d'auditeurs des phrases prononcées par ces locuteurs dans l'interface d'EloKanz. Ces sessions de test suivaient 5 étapes consécutives :

1. installation de l'auditeur
2. lecture de la consigne à l'auditeur
3. entraînement de l'auditeur (écoute de 10 phrases)
4. session d'écoute de 60 phrases
5. questionnaire sur la difficulté d'utilisation du logiciel

Installation de l'auditeur

L'auditeur était invité à s'installer dans la salle dédiée au bilan de voix et de parole. Il était assis devant le bureau sur lequel reposait un ordinateur portable *MacBook pro 15"* équipé du logiciel EloKanz. Le son était diffusé *via* les deux haut-parleurs *Yamaha NS-10m Studio 120 watts*.

Nous avons fait en sorte que le niveau sonore soit égal pour chacun des 24 participants. Ce niveau avait été calibré au préalable à l'aide d'un sonomètre : nous avons effectué les réglages nécessaires pour que le niveau de sortie à 1 mètre de distance des haut-parleurs soit le même niveau que celui mesuré à 1 mètre de la bouche d'une personne parlant dans la salle d'enregistrement.

Lecture de la consigne à l'auditeur

La consigne donnée en annexe E.12 (p. 326) était ensuite lue à l'auditeur.

Entraînement de l'auditeur (écoute de 10 phrases)

La phase d'entraînement portait sur 10 phrases prononcées par trois patients différents : PA16, PN12 et PN14. Les fichiers audio correspondant à ces phrases sont donnés en annexe DVD 5d_3. Les phrases étaient diffusées de telle manière que les auditeurs ne pouvaient pas entendre 3 phrases prononcées par le même patient à la suite, comme cela est aussi le cas pour la phase d'écoute de 60 phrases.

Les heures de départ et de fin de l'entraînement étaient relevées. Si l'auditeur rencontrait des difficultés pendant l'entraînement, cela était noté sur la grille d'observation (annexe E.11). Si l'auditeur ne répondait pas correctement aux consignes, cela lui était explicitement signalé, et des indications lui étaient données pour l'aider à corriger ce comportement.

Session d'écoute de 60 phrases

L'auditeur avait pour consigne d'écouter et de réagir aux 60 phrases *via* l'interface d'EloKanz.

Les heures de départ et de fin de la session d'écoute étaient relevées. Durant la session, si l'auditeur rencontrait des difficultés cela était noté sur la grille d'observation. Contrairement à la phase d'entraînement, durant la passation les comportements non conformes aux consignes n'étaient pas signalés à l'auditeur.

Questionnaire sur la difficulté d'utilisation du logiciel

À la fin de la passation, l'auditeur devait répondre à la question contenue dans le questionnaire donné en annexe E.3 (p. 312) :

« Est-il facile ou difficile pour vous d'utiliser ce logiciel ? »

Les auditeurs devaient situer leur réponse à cette question sur une échelle de 1 (« Très facile ») à 5 (« Très difficile »). Enfin, l'auditeur pouvait signaler s'il avait rencontré des difficultés particulières, ou formuler d'éventuels commentaires.

7.3.4 Résultats

L'ensemble des résultats de cette expérience est disponible en annexe DVD 2d_3. D'autre part, certains des résultats que nous exposons dans cette partie ne sont relatifs qu'à un moment particulier de la passation du test. C'est pourquoi nous tenons dans un premier temps à présenter au lecteur la terminologie que nous employons pour nous référer à ces différents moments :

- **la phase d'entraînement** renvoie à la phase durant laquelle les auditeurs ont écouté et réagi aux 10 phrases d'essai prononcées par les 3 patients PA16, PN12 et PN14 ;
- **la phase d'écoute** renvoie à la phase durant laquelle les auditeurs ont écouté 60 phrases prononcées par les 6 patients PA06, PA08, PA09, PA17, PN09 et PN10 ;
- **la passation du test** désigne l'ensemble du test : phase d'entraînement + phase d'écoute.

Nous désirons également signaler qu'un problème technique a contraint deux auditeurs à arrêter les sessions d'écoute avant d'être arrivés à leur terme. Il s'agit des auditeurs AE11 et AE13. Les données correspondant à ces deux auditeurs ne seront donc pas intégrées à certains des calculs statistiques présentés dans cette partie. Le même problème est survenu pendant la phase d'entraînement de l'auditeur AE09 – et les données relatives à l'entraînement de cet auditeur ne seront pas intégrées à nos calculs.

Performances générales des auditeurs

Respect des consignes

Globalement, les auditeurs ont bien répondu aux consignes. Ainsi, durant la phase d'entraînement aucune erreur n'a été relevée. En revanche, durant la phase d'écoute, 6 auditeurs ont éprouvé des difficultés par rapport aux consignes :

- 4 auditeurs ne se souvenaient plus du comportement à adopter pour le cas où ils ont *partiellement* compris la phrase prononcée par le patient (ex. si ils n'ont compris que le premier nom, ils doivent quand même effectuer un déplacement impliquant l'utilisation de cette information) ;
- 2 auditeurs ne se souvenaient plus de la consigne en ce qui concerne l'effacement d'un déplacement (l'auditeur a la possibilité d'effacer puis de recommencer son déplacement) ;
- enfin, un auditeur a demandé s'il pouvait écouter une phrase à nouveau, ce qui s'oppose à ce qui est expliqué dans la consigne (l'auditeur ne peut écouter chaque phrase qu'une seule fois).

Durée de passation

En moyenne, la durée de la passation (entraînement + écoute) se situe en dessous de 20 minutes. La phase d'entraînement a duré, en moyenne, 2 minutes et 50 secondes (*min.* = 2 mn; *max.* = 5 mn), et la phase d'écoute des 60 phrases a duré, en moyenne, 12 minutes et 49 secondes (*min.* = 10 mn; *max.* = 19 mn).

Ergonomie

La majorité des auditeurs (21 auditeurs, soit 87% de notre population) a rencontré des difficultés avec l'ergonomie du test. Les différents problèmes relevés sont les suivants :

- l'auditeur doit refaire son déplacement car il n'a pas relâché l'objet suffisamment près de l'emplacement cible (*cf.* figure 4.12 p. 104) ;
- l'auditeur a des difficultés à reconnaître une image particulière.
- l'auditeur doit essayer de cliquer plusieurs fois sur le pavé tactile avant d'y parvenir ;
- l'auditeur a des difficultés à viser le bouton sur lequel il doit cliquer ;
- l'auditeur a des difficultés à utiliser le pavé tactile pour déplacer un objet.

Parmi ces difficultés, les premières sont celles qui ont concerné le plus de sujets : par exemple 16 sujets (66,7% de notre population) ont eu des problèmes avec les emplacements cible, 9 sujets (37,5%) avec certaines images, et 8 sujets (33,3%) avec la réalisation de clics sur le pavé tactile.

Pour plus de détails sur les éléments de l'ergonomie et des consignes qui ont gêné les auditeurs, nous renvoyons le lecteur aux questions et commentaires que les auditeurs ont effectué pendant la lecture des consignes, la phase d'entraînement et la phase d'écoute ainsi que lors du questionnaire post-test, qui sont disponibles en annexes E.14 (p. 329) et E.15 (p. 330).

Évaluation de la difficulté d'utilisation du logiciel

La majorité des auditeurs a attribué une note de 1 ou de 2 au logiciel sur l'échelle de difficulté que nous avons proposée, allant de 1 – *Très facile* à 5 – *Très difficile*. Ainsi, les résultats illustrés dans la figure 7.16 sont similaires à ceux de l'évaluation de l'interface *locuteur* faite par les patients (*cf.* figure 7.2 p. 188).

Influence de l'âge sur les performances des auditeurs

Nous avons émis l'hypothèse selon laquelle l'âge pourrait être un facteur déterminant dans les performances des auditeurs vis-à-vis de l'utilisation d'EloKanz. Pour vérifier cette hypothèse, nous avons dans un premier temps effectué des calculs de corrélation entre l'âge des auditeurs et le nombre de difficultés qu'ils ont éprouvées par rapport au suivi des consignes ou par rapport à l'ergonomie du test. Les résultats sont présentés dans la table 7.11.

Seule la relation entre l'âge des auditeurs et le nombre d'erreurs commises par rapport aux consignes durant la phase d'écoute apparaît comme statistiquement significative. Elle est de force moyenne et de polarité positive, ce qui signifie que plus l'âge des auditeurs augmente, plus le nombre d'erreurs a tendance à augmenter.

Ensuite, nous avons calculé les corrélations entre l'âge des auditeurs et les durées des phases d'entraînement et d'écoute. Les résultats sont donnés dans la table 7.12. Seule la relation entre l'âge des auditeurs et la durée de la phase d'entraînement apparaît comme statistiquement significative. Elle est de force modérée et de polarité

FIGURE 7.16 – Évaluation de la difficulté d'utilisation du logiciel par les auditeurs, sur une échelle de 1 (« très facile ») à 5 (« très difficile »)

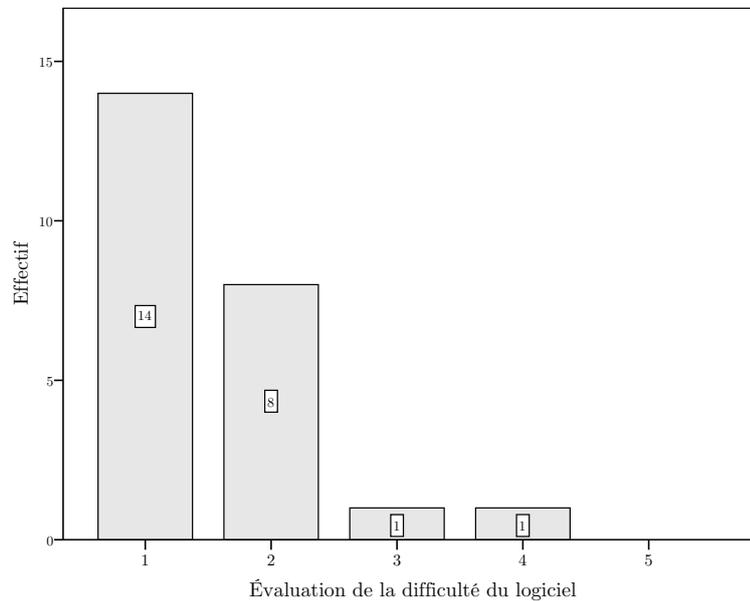


TABLE 7.11 – Nombre de problèmes liés aux consignes et à l'ergonomie en fonction de l'âge des auditeurs (corrélations)

		Nombre d'erreurs commises par rapport aux consignes	Nombre de problèmes éprouvés par rapport à l'ergonomie
Âge	Corrélation de Pearson	0,507*	0,054
	Sig. (bilatérale)	0,016	0,813
	<i>N</i>	22	22

positive : plus l'âge de l'auditeur augmente, plus la durée de la phase d'entraînement s'accroît. La corrélation entre l'âge des auditeurs et la durée de la session d'écoute est cependant assez proche du seuil de significativité ($P = 0,059$), et d'une force comparable à la corrélation trouvée pour la phase d'entraînement.

TABLE 7.12 – Durées des phases d'entraînement et d'écoute en fonction de l'âge des auditeurs (corrélations)

		Durée de la phase d'entraînement	Durée de la phase d'écoute
Âge	Corrélation de Pearson	0,423*	0,409
	Sig. (bilatérale)	0,044	0,059
	<i>N</i>	23	22

Enfin, nous avons calculé la corrélation entre l'âge des auditeurs et l'évaluation

TABLE 7.13 – Évaluation de la difficulté du logiciel par les auditeurs en fonction de leur âge (*Tau-B* de Kendall)

		Évaluation subjective de la difficulté du logiciel (échelle de 1 – <i>Très facile</i> à 5 – <i>Très difficile</i>)
Âge	Coefficient de corrélation	0,427*
	Sig. (bilatérale)	0,013
	<i>N</i>	24

de la difficulté du logiciel sur l'échelle que nous leur avons proposée. Les résultats apparaissent dans la table 7.13.

La relation apparaît comme statistiquement significative, et de force modérée. La polarité est positive, ce qui signifie que plus l'âge des auditeurs augmente, plus les notes attribuées sur l'échelle de difficulté tendent à augmenter aussi.

7.3.5 Discussion

Performances des auditeurs

Dans l'ensemble, les auditeurs n'ont commis que très peu d'erreurs par rapport aux consignes (les 3/4 d'entre eux n'ont fait *aucune* erreur). De plus, la durée de passation semble relativement courte : moins de 20 minutes en moyenne pour la phase d'entraînement et l'écoute de 60 phrases – sachant en outre que les scores de compréhensibilité des patients sont ensuite calculés quasi instantanément par le logiciel. Enfin, la grande majorité des auditeurs a jugé le test comme étant facile, voire très facile⁸. En revanche, l'ergonomie du test constitue un problème beaucoup plus important, qui a affecté la majorité des sujets testés.

L'âge des auditeurs semble avoir une incidence sur le nombre d'erreurs commises par rapport aux consignes, ainsi que sur la durée nécessaire à la phase d'entraînement – et dans une moindre mesure à celle nécessaire à la phase d'enregistrement. Cependant, nous pensons que la familiarité des auditeurs avec l'outil informatique a également pu jouer un rôle vis-à-vis des durées de passation. Certains auditeurs ont d'ailleurs évoqué leur manque d'expertise dans ce domaine (« Ce n'est pas de ma génération. », « Je n'ai pas d'ordinateur chez moi. » – *cf.* remarques en annexe E.15, p. 330). Pour mieux cerner le rôle qu'a pu jouer l'âge dans les performances des auditeurs, il aurait certainement fallu prendre en compte cette variable, par exemple en demandant aux auditeurs d'auto-évaluer leur familiarité avec l'outil informatique sur une échelle de Likert.

L'âge ne semble pas, par contre, être une variable associée au nombre de problèmes d'ergonomie rencontrés par les auditeurs. Ces problèmes, causés pour une large part par le pavé tactile, les images et les boutons du logiciel, ont affecté la majorité des auditeurs – sans distinction particulière liée à l'âge.

Quelques perspectives d'amélioration du test EloKanz

Concernant le respect des consignes, il est apparu que le problème principal était lié aux actions que doivent effectuer les auditeurs lorsqu'ils n'ont que partiellement compris une phrase. Il est vrai que cette partie des consignes est complexe et mériterait certainement une attention renforcée. Nous pensons notamment à l'utilisation d'une mise en situation des auditeurs pendant la phase d'entraînement. Le deuxième point qu'il nous paraît important de noter est celui de la latéralité. Trois auditeurs ont déclaré avoir parfois des difficultés à reconnaître leur droite de leur gauche. Les résultats de l'étude montrent cependant que ce problème est marginal – avec seulement 4 cas possibles de confusions droite/gauche, sur 484 observations⁹. Néanmoins, il nous paraît important de pallier cette difficulté car cela ajouterait au

8. seul un auditeur a attribué un score de 4 sur l'échelle de difficulté. Cependant, cet auditeur exerce la profession d'ergothérapeute. Il est possible qu'il ait, par conséquent, un regard différent sur la facilité d'utilisation du logiciel.

9. de plus, il n'est pas possible d'être certain qu'il s'agit de confusions dans la mesure où les actions s'inscrivaient en réponse à des signaux de parole dégradés.

confort des personnes qui ne se sentent pas à l'aise avec la question de la latéralité, et aussi parce que certains personnels médicaux sont habitués à inverser les notions de droite et de gauche en se mettant à la place de la personne qu'ils ont en face d'eux¹⁰. Ainsi nous pensons qu'il serait intéressant de rajouter des étiquettes « droite » et « gauche » directement dans l'interface *auditeur* d'EloKanz.

Concernant l'ergonomie, tout comme pour la partie *locuteurs* de l'étude il paraît évident que le pavé tactile doit être remplacé par un périphérique plus facile à manier. Nous avons pensé à cette question dès le début de l'étude, mais les conditions matérielles (taille de la salle et du bureau – cf. photographie en annexe E.7 p. 316) ne permettaient pas d'installer un tapis de souris autre part que sur les jambes de l'utilisateur – ce qui aurait également pu causer d'autres soucis. Ensuite, beaucoup d'utilisateurs ont été amenés à recommencer leurs déplacements d'objets car ils relâchaient les objets trop tôt, c'est-à-dire trop loin des emplacements cible (cf. copie d'écran en figure 4.12 p. 104). Ce problème pourrait être dû à une mauvaise manipulation du pavé tactile. Nous pensons d'un autre côté que ce problème pourrait être corrigé dans le logiciel, en élargissant la zone de sensibilité des emplacements cible. D'autre part, plusieurs auditeurs ont évoqué la possibilité d'utiliser le clavier pour valider leurs déplacements, et débiter l'écoute des enregistrements audio – ce qui éviterait des allers-retours entre les boutons du logiciel. Nous pensons que cette solution est très intéressante et ferait certainement gagner du temps. Enfin, tout comme les patients, les auditeurs ont été gênés par des images dont la couleur se rapprochait trop du blanc utilisé pour le fond de l'écran, et/ou qui étaient trop fines (objets élancés comme le râteau, la cuillère, etc.). À un même problème, même solution : nous pensons qu'il serait intéressant d'utiliser des couleurs différentes pour les images trop pâles, et de supprimer les images qui représentent des objets dont la forme est trop élancée.

10. cf. par exemple le commentaire de l'auditeur AE13 « Je fais par rapport à ma droite à moi ? ».

7.4 Éléments de validation du test EloKanz

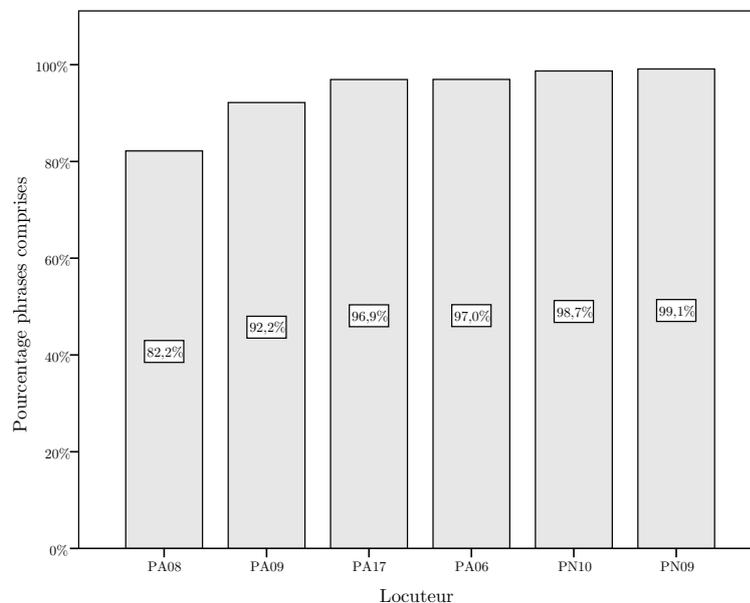
7.4.1 Objectifs et hypothèses

Dans cette partie nous nous intéressons à la validité du test EloKanz, que celle-ci concerne les aspects internes ou externes des mesures obtenues :

- concernant la *validité interne*, nous nous poserons les questions suivantes : est-ce que les mesures sont reproductibles ? Autrement dit, existe-t-il un bon accord entre les scores obtenus par différents auditeurs pour un même ensemble de stimuli ? Questionner la validité interne de la méthode sera donc l'occasion d'étudier les facteurs susceptibles d'ajouter de la variabilité aux scores obtenus.
- la validité *externe*, quant à elle pose la question suivante : est-ce que les scores obtenus sont bien représentatifs de la dimension qu'ils sont censés mesurer ?

Avant d'exposer nos hypothèses concernant les validités interne et externe du test, nous souhaitons traiter des indicateurs que nous avons retenus pour la compréhension de la parole chez les auditeurs. Dans la figure 7.17 sont présentés les pourcentages de phrases comprises pour chacun des 6 locuteurs de notre étude, c'est-à-dire les pourcentages de phrases pour lesquelles les auditeurs ont correctement effectué les actions commandées par les patients. Comme nous pouvons le constater, il existe un important effet de plafond : la grande majorité des phrases ont été comprises par les auditeurs.

FIGURE 7.17 – Pourcentage de phrases comprises par les auditeurs, pour chaque locuteur



Cet effet de plafond nous empêche d'effectuer des calculs statistiques impliquant l'analyse de la variation des scores obtenus, puisque ces derniers varient très peu. Nous ne sommes donc pas en mesure de pouvoir étudier la validité des scores de

compréhension, telle qu'évaluée par l'observation actions effectuées par les auditeurs en réponse aux énoncés.

En conséquence, nous avons choisi de nous intéresser plutôt à la validité des *temps de réaction* des auditeurs, pris comme des indices des difficultés qu'ont eu ces derniers à comprendre les énoncés – une mesure souvent employée en tant que telle dans la littérature (Carlson *et al.*, 1992; Hustad et Beukelman, 2002; Ralston *et al.*, 1991, *cf.* la section *Évaluations s'appuyant sur la mesure de temps de réaction*, p. 88).

Pour ce qui est de la *validité interne* de la mesure (p. 222), nous observerons donc dans un premier temps l'accord qui existe entre les temps de réaction des auditeurs. Ensuite, nous traiterons de l'influence que peut avoir l'exposition des auditeurs aux TPPP. À notre connaissance, l'effet de l'exposition des auditeurs aux TPPP n'a été montré que pour des scores obtenus à des tests d'intelligibilité (*cf.* table 2.2, p. 46), et non pour des données de temps de réaction à des énoncés lors de tâches de compréhension. Nous faisons l'hypothèse que l'exposition des auditeurs aux TPPP améliore la compréhension des auditeurs, ce qui va réduire leurs temps de réaction de manière significative. Enfin, nous étudierons l'influence de facteurs propres aux stimuli utilisés. Nous avons évoqué dans la discussion relative à notre expérience de compréhension de la parole dans le bruit (section 6.5.1, p. 168) le fait que les contextes imagés accompagnant chaque stimulus n'avaient peut-être pas le même « poids » selon les images qui le constituaient et selon le stimulus qu'ils accompagnaient. Nous faisons notamment les hypothèses suivantes :

- la proximité phonologique entre les mots cibles et les mots correspondant aux images dans le contexte est un facteur de difficulté pour la compréhension (par exemple la compréhension du mot « radis » sera davantage gênée par le distracteur « râteau » que par le distracteur « citron »). En conséquence plus cette proximité sera importante plus les temps de réaction vont avoir tendance à augmenter ;
- le contexte est d'autant plus facilitant que les noms des images qui le constituent n'appellent pas le même déterminant que les images cible. Par exemple si le premier nom cible est masculin (« Mettez *le* râteau... ») et que tous les autres noms du jeu d'image sont féminins, alors l'auditeur peut déduire plus facilement quel est le premier nom qui a été donné puisque le déterminant masculin lui donne un indice supplémentaire. Nous faisons donc l'hypothèse que moins le nombre d'images appelant un même déterminant que les images cibles sera important, plus la compréhension des auditeurs sera facilitée et plus les temps de réaction auront tendance à baisser ;
- enfin, nous faisons l'hypothèse que la fréquence des noms n'aura pas d'influence significative sur le temps de réaction, puisque c'est ce que nous avons observé lors de notre expérience de compréhension de la parole dans le bruit (*cf.* section 6.5.1, p.168).

Concernant cette fois la *validité externe* de la mesure (p. 226), nous mettrons en rapport les temps de réaction avec les scores d'évaluation de la compréhensibilité des patients effectuée par les trois experts lors du test présenté dans la section *Pré-test*

pour le choix des stimuli (p. 207).

7.4.2 Résultats

Validité interne

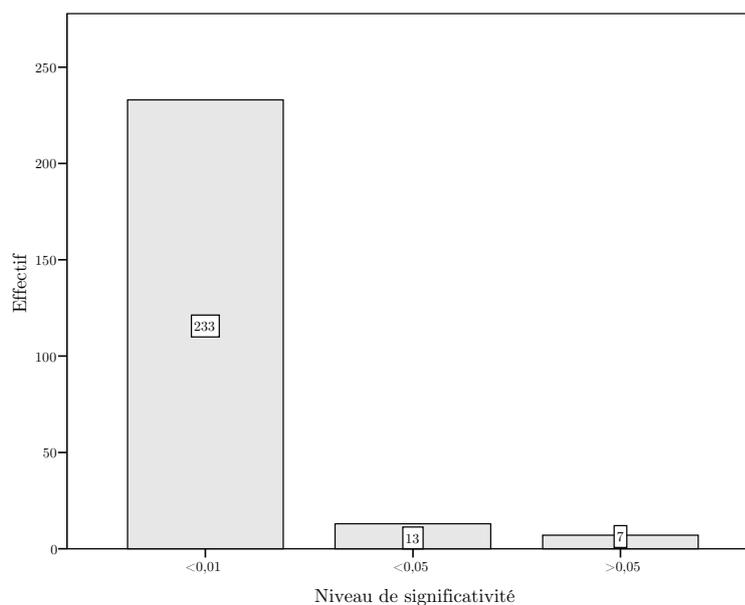
Accords inter auditeurs

L'accord entre les auditeurs a été calculé pour les temps de réaction uniquement. Comme nous l'avons dit la variable booléenne correspondant à la compréhension des phrases (0 = *phrase non comprise* ; 1 = *phrase comprise*) obtient, majoritairement, des valeurs de 1. Par conséquent cette variable ne varie pas assez pour nous permettre de calculer l'accord entre les auditeurs par un test du Kappa par exemple.

Concernant les temps de réaction, nous avons observé 24 auditeurs dans cette étude – ce qui signifie que le calcul des accords entre chaque couple d'auditeurs donne lieu à 276 corrélations ($\frac{24 \times 23}{2}$), et 253 si l'on ôte les données relatives à l'auditeur AE11 pour lequel trop peu de valeurs de temps de réaction ont été relevées. Plutôt que de donner les valeurs de significativité et de force pour chacune de ces corrélations, nous préférons indiquer :

- dans la figure 7.18 le nombre de corrélations hautement significatives ($P < 0,01$), significatives ($P < 0,05$) et non significatives ($P > 0,05$) ;
- dans la figure 7.19 la distribution des coefficients de corrélation de Pearson obtenus.

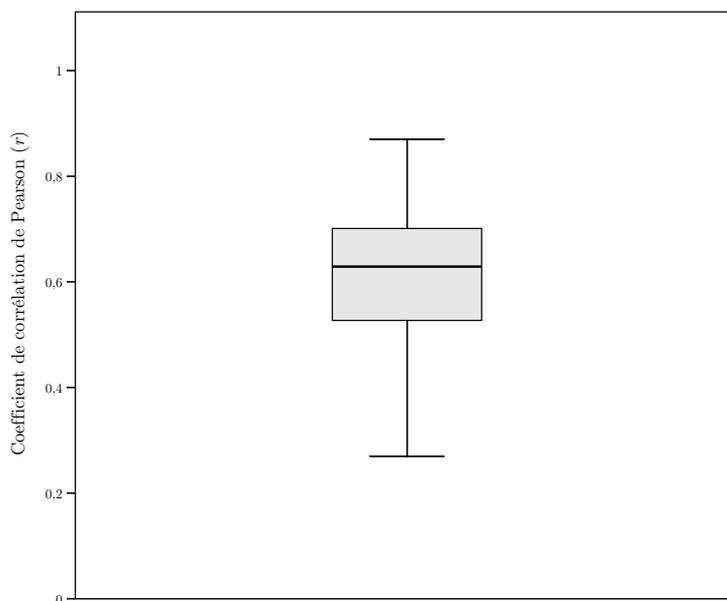
FIGURE 7.18 – Corrélations inter auditeurs, par niveau de significativité



Pour calculer ces corrélations, nous avons tenu compte uniquement des observations pour lesquelles les phrases avaient été comprises par les auditeurs. En effet, si nous avions également pris en compte les observations pour lesquelles les phrases

n'avaient pas été comprises, cela aurait ajouté de la variabilité et donc nui à la fiabilité des accords relevés.

FIGURE 7.19 – Coefficients de corrélation de Pearson obtenus pour les temps de réaction de 253 couples d'auditeurs



Influence de l'exposition des auditeurs aux TPPP

La figure 7.20 illustre les pourcentages de phrases comprises pour les auditeurs qui ont été exposés à de la parole pathologique *vs* les auditeurs qui ont déclaré ne jamais avoir été exposés.

La figure 7.21 montre les temps de réaction moyens des auditeurs appartenant à ces deux groupes.

Dans ces deux figures il n'apparaît pas de différence majeure entre les auditeurs exposés et non exposés aux TPPP. Cependant la question de l'influence de l'exposition est de nature complexe. Nous avons vu dans l'étude de faisabilité pour les auditeurs (section 7.3, p. 205) que l'âge avait une influence *négative* sur les performances des auditeurs pour l'utilisation du logiciel. Or, le nombre d'années d'exposition – dont nous faisons l'hypothèse qu'il a une influence *positive* sur les temps de réaction – augmente avec l'âge. Pour vérifier l'effet produit par l'exposition, il serait donc nécessaire de prendre en compte ces deux variables en même temps. Pour cela nous avons effectué un calcul de régression linéaire, avec :

- comme variable dépendante les temps de réaction des auditeurs ;
- comme étiquettes d'observation les phrases entendues ;
- comme variables indépendantes l'âge des auditeurs et le nombre d'années pendant lesquelles ils ont été exposés aux TPPP.

Les résultats, présentés dans la table 7.14, suggèrent que les deux variables ont un effet statistiquement significatif sur les temps de réaction des auditeurs. L'âge donne

FIGURE 7.20 – Pourcentages de phrases comprises chez les auditeurs exposés aux TPPP ou non exposés aux TPPP

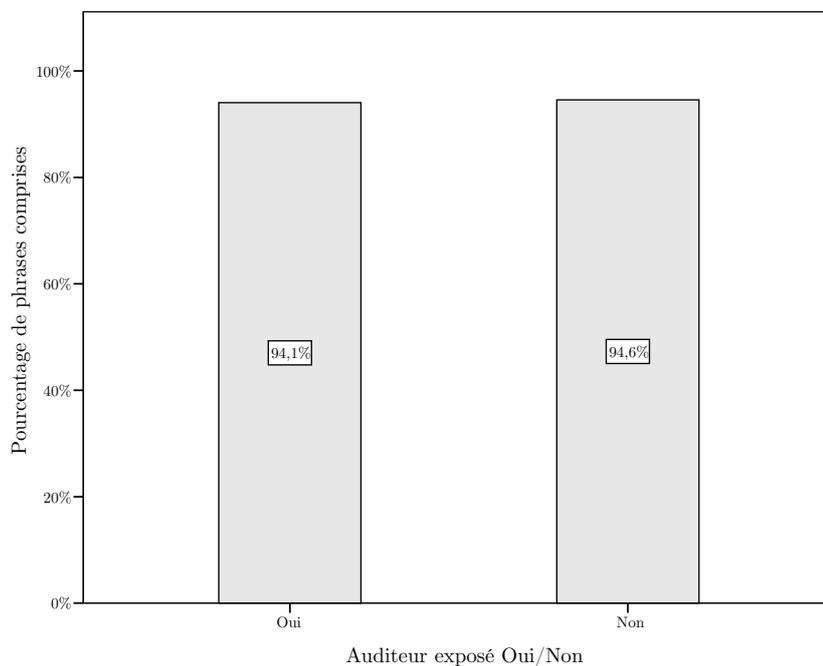
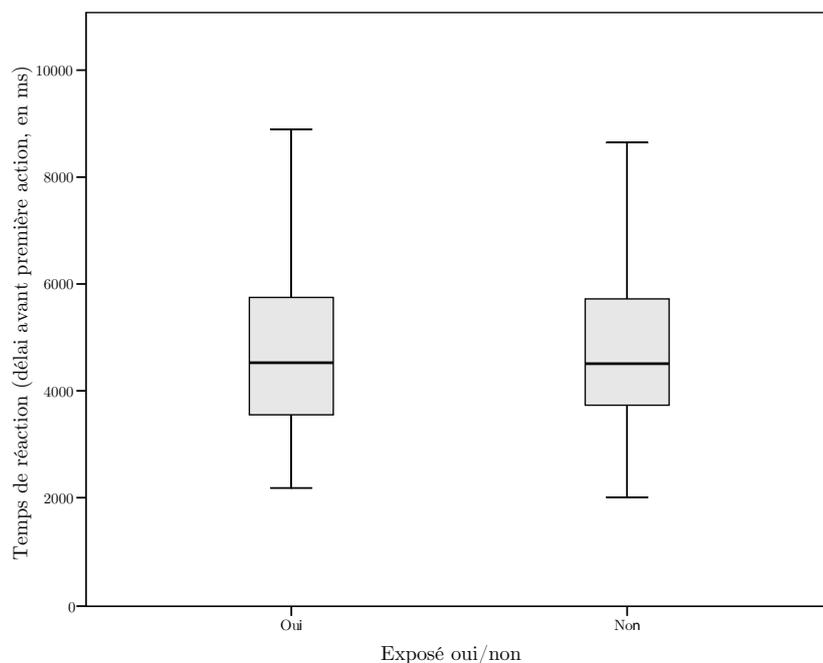


FIGURE 7.21 – Temps de réaction moyens (en ms) avant l'exécution de la première action pour les auditeurs exposés aux TPPP *vs* non exposés



lieu a un coefficient positif, ce qui indique que plus l'âge augmente plus les temps de réaction ont aussi tendance a augmenter. Le coefficient lié au nombre d'années

d'exposition est négatif, ce qui indique que plus les auditeurs ont été exposés aux TPPP, plus les temps de réaction ont tendance à baisser. Par ailleurs, les coefficients standardisés mis au jour ont des valeurs assez proches (valeurs absolues de 0,17 et 0,15), ce qui pourrait éventuellement expliquer que les deux facteurs se sont « annulés » mutuellement et que l'influence de l'exposition des auditeurs n'était pas visible dans le graphique 7.21.

TABLE 7.14 – Variabilité des temps de réaction des auditeurs expliquée par l'âge et les années d'exposition aux TPPP (régression linéaire)

Calcul de régression linéaire

Modèle	Coefficients non standardisés		Coefficients standardisés	t	Sig.
	A	Erreur standard	Bêta		
(Constante)	4346,46	298,40		14,57	0,000
Âge de l'auditeur	29,96	12,03	0,17	2,49	0,013*
Années d'exposition	-30,07	14,12	-0,15	-2,13	0,033*

Influence de facteurs propres aux stimuli

Voici comment nous avons quantifié les variables liées aux stimuli, et dont nous faisons l'hypothèse qu'elles peuvent avoir une influence sur les temps de réaction des auditeurs :

- la *proximité phonologique* entre un nom cible et ses concurrents a été calculée sur la base de l'algorithme de Levenshtein, un algorithme permettant de quantifier la distance entre deux chaînes de symboles, utilisé notamment en dialectologie pour mesurer les distances entre unités linguistiques (*cf.* Beijering *et al.*, 2008; Heeringa, 2004). Le principe de l'algorithme de Levenshtein est d'attribuer un poids aux modifications nécessaires pour passer d'une chaîne A à une chaîne B. Ces modifications peuvent être l'ajout, la suppression ou la substitution d'un symbole. Ainsi pour passer de la chaîne /pəti/ (*petit*) à /pətit/ (*petite*) il faut effectuer *un* ajout de phonème, ce qui donne une distance égale à 1. Si l'on désire maintenant passer de la chaîne /pəti/ (*petit*) à /apeti/ (*appétit*) il faut effectuer *un* ajout de phonème, et *une* substitution de phonème (/ə/ est remplacé par /e/), ce qui donne une distance de 2. L'algorithme de Levenshtein calcule ainsi les différents « chemins » qu'il y a pour passer d'une chaîne à l'autre, et retient le chemin le plus court. Nous avons modifié l'algorithme de Levenshtein car lors de la substitution d'un phonème par un autre, l'algorithme ne permettait pas de prendre en compte le fait que deux phonèmes peuvent être plus ou moins proches selon le nombre de traits distinctifs qu'ils partagent (par exemple le phonème /b/ est plus proche de /p/ que de /f/). Nous avons donc fait en sorte que le poids attribué par l'algorithme pour une substitution de phonème s'appuie sur le nombre de traits distinctifs partagés. Cet algorithme est implémenté dans un programme informatique écrit en Perl, que nous donnons en annexe DVD 5d_5;

- le nombre de déterminants communs aux noms cibles et aux noms des images du contexte a été calculé comme suit. Pour le premier nom cible, nous avons compté le nombre d’éléments qui appellent le même déterminant (c.-à-d. *le*, *l’*, *la* ou *les*) parmi les noms correspondant aux images dans le contexte et le deuxième nom cible. Pour le second nom cible, nous avons compté le nombre d’éléments qui appellent le même déterminant (c.-à-d. *du*, *de la*, *de l’* ou *des*) parmi les noms correspondant aux images dans le contexte ;
- la fréquence des noms cibles correspond à la fréquence de ces noms (des lemmes) dans le corpus de sous-titres de films de la base *Lexique 3*, comme cela était le cas lorsque nous avons étudié la fréquence des noms dans notre expérience de compréhension de la parole dans le bruit (cf. section 6.5.1 p. 168).

Les résultats d’un calcul de régression linéaire prenant en compte ces facteurs, ainsi que les facteurs liés aux auditeurs (âge et nombre d’années d’exposition aux TPPP), sont présentés dans la table 7.15. Ce calcul ne prend en compte que les variables qui se sont révélées statistiquement significatives.

TABLE 7.15 – Variabilité des temps de réaction des auditeurs expliquée par les variables liées aux auditeurs et aux stimuli (régression linéaire)

Calcul de régression linéaire

Modèle	Coefficients non standardisés		Coefficients standardisés	t	Sig.
	A	Erreur standard	Bêta		
(Constante)	4373,27	399,07		10,96	0,000
Âge de l’auditeur	30,60	11,81	0,18	2,49	0,010**
Années d’exposition	-31,39	13,86	-0,16	-2,27	0,024*
Proximité phono. nom 1	295,07	64,44	0,13	4,58	0,000**
Proximité phono. nom 2	-186,87	62,33	-0,08	-3,00	0,003**
Fréquence nom 1	-5,67	1,87	-0,08	-3,04	0,002**

Validité externe

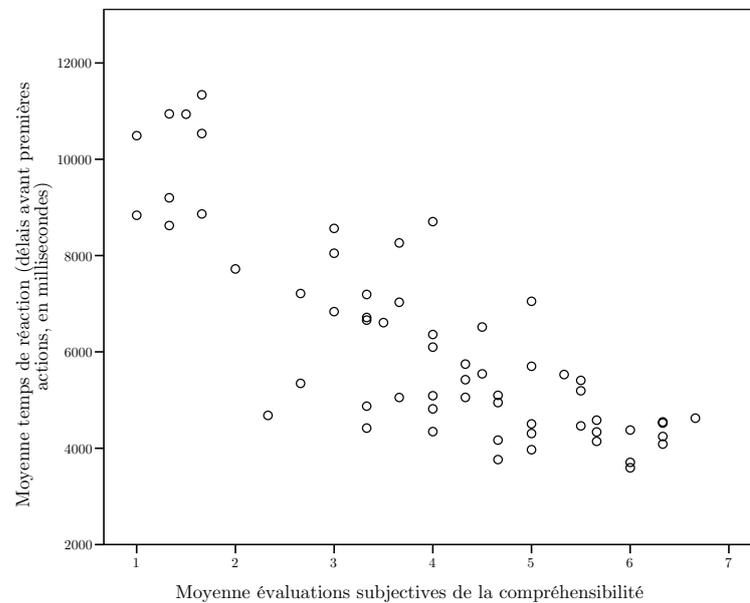
Dans la table 7.16 sont présentés les résultats d’un calcul de corrélation de Pearson entre la moyenne des scores d’évaluation subjective de compréhensibilité obtenus pour chaque phrase, et la moyenne des temps de réaction pour ces phrases dans Elo-Kanz.

TABLE 7.16 – Temps de réaction des auditeurs en fonction des scores d'évaluations de la compréhensibilité sur une échelle (corrélation)

		Évaluation subjective de la compréhensibilité des stimuli (échelle de 1 à 7)
Temps de réaction dans EloKanz (délai avant première action)	Corrélation de Pearson	-0,813**
	Sig. (bilatérale)	0,000
	<i>N</i>	60

La figure 7.22 présente le nuage de points obtenu en mettant en rapport ces deux variables.

FIGURE 7.22 – Temps de réaction moyens des auditeurs en fonction des scores moyens d'évaluation subjective de la compréhensibilité



7.4.3 Discussion

Validité interne de la mesure

Globalement, nous avons trouvé entre les temps de réaction des auditeurs des coefficients de corrélation moyens à forts, et pour la grande majorité hautement significatifs. Ceci démontre une certaine cohésion dans les réponses des participants.

Néanmoins, l'âge et l'exposition des auditeurs aux TPPP pourraient avoir un effet sur leurs performances. C'est ce qu'ont révélé, dans un premier temps, nos calculs de régression linéaire. Il nous est difficile d'étudier plus en avant ce point particulier car l'âge et l'exposition ont un effet inverse sur les performances des sujets. Il faudrait observer des données pour lesquelles seul le nombre d'années d'exposition varie, l'âge demeurant constant.

De plus, notre second calcul de régression linéaire (*cf.* table 7.15 p. 226) révèle que de nombreuses variables liées aux stimuli pourraient avoir une influence sur les temps de réaction obtenus par les auditeurs :

1. *la proximité phonologique entre les noms cible et les distracteurs.* Le calcul de régression linéaire indique que plus le premier nom cible a une forme proche de celles des distracteurs, plus le temps de réaction de l'auditeur a tendance à augmenter – ce qui correspond à l'hypothèse que nous avons formulée. En revanche, plus le second nom cible est similaire aux distracteurs, plus le temps de réaction a tendance à *baisser*. Ne nous attendions pas à ce résultat. Plusieurs possibilités peuvent l'expliquer, la première étant que ce résultat soit dû au hasard – mais la hauteur de la significativité obtenue ($P = 0,003$) ne joue pas en faveur de cette interprétation. Une seconde possibilité pourrait être la mise en place de stratégies particulières par les auditeurs lorsque le second nom est difficile à saisir. Nous pourrions imaginer en effet que les auditeurs, lorsqu'ils réalisent que le deuxième nom est difficile à saisir, cliquent rapidement sur le premier objet pour valider cette partie de la réponse, puis se concentrent sur la compréhension des autres éléments de la phrase¹¹. Ceci expliquerait que les temps de réaction (clics sur le premier objet) aient tendance à diminuer lorsque le second nom est difficile à saisir – notamment du fait de la proximité phonologique entre ce nom et les distracteurs. Cependant, ceci n'est qu'une hypothèse qui demanderait à être vérifiée de manière plus rigoureuse ;
2. *le nombre de déterminants communs aux noms cible et aux distracteurs.* Nos résultats montrent que moins le premier nom a de concurrents pour ce qui est du déterminant parmi les distracteurs, plus les temps de réaction baissent. Cela pourrait être le signe que différentes stratégies sont utilisées par les auditeurs, stratégies dont fait partie le procédé consistant à éliminer les images dont le nom n'est pas compatible avec le déterminant entendu. Nous pourrions également imaginer que l'importance de ce type de stratégie dépende de la difficulté des

11. nous avons observé que des auditeurs adoptaient parfois de tels comportements : ceux-ci sélectionnaient le premier objet, puis le faisaient tourner sur l'écran en ayant visiblement l'air de réfléchir, et effectuaient enfin leur déplacement.

conditions d'écoute – en supposant que plus la parole est difficile à comprendre, plus l'auditeur va avoir tendance à recourir à de telles stratégies¹² ;

3. *la fréquence des noms cible*. Nous ne nous attendions pas à ce résultat, dans la mesure où nous n'avions pas noté d'effet de la fréquence des noms dans notre expérience de compréhension de la parole dans le bruit (*cf.* section 6.5.1, p.168). Nous n'avons pas d'hypothèse forte pour expliquer pourquoi nous n'avons pas relevé d'effet dans la première expérience et pourquoi nous en observons un dans cette seconde.

Validité externe de la mesure

Nous avons relevé une très forte corrélation entre les temps de réaction des auditeurs et les notes de compréhensibilité attribuées par les trois experts dans le pré-test d'évaluation subjective (*cf.* section *Pré-test pour le choix des stimuli*, p. 207). Ceci contribue à envisager les temps de réaction dans le test EloKanz comme une mesure valide de la compréhensibilité de la parole. Toutefois, comme pour les études que nous avons évoquées dans notre état de l'art sur la relation entre l'intelligibilité et la compréhension de la parole (Beukelman et Yorkston, 1979; Hustad, 2008, *cf.* section 3.4.1, p. 80), il est possible que la variable « sévérité du TPPP » masque la vraie nature de la relation entretenue par les deux variables. Pour vérifier que ce n'est pas le cas dans notre étude, nous avons effectué une corrélation partielle entre les temps de réaction et les scores d'évaluation subjective de la compréhensibilité, avec comme variable de contrôle l'identifiant du locuteur. Le résultat montre une corrélation toujours hautement significative ($P < 0,01$) et de force toujours très élevée ($r = -0,805$).

D'autre part, nous pensons qu'il serait intéressant d'étudier les différences de résultats obtenus par la mesure de temps de réaction et des évaluations subjectives. Ainsi la figure 7.23 présente les temps de réaction moyens obtenus pour chaque locuteur, et la figure 7.24 illustre les médianes des scores d'évaluation subjective de la compréhensibilité pour chaque locuteur.

Dans chacune de ces deux figures nous indiquons par des astérisques les différences significatives (*) et hautement significatives (**) entre les locuteurs. Nous voyons que, si les extrêmes restent les mêmes dans les deux figures (PA06 et PN10 d'un côté et PA08 de l'autre), au centre des valeurs les différences observées ne sont pas les mêmes. Par exemple il n'existe pas de différence significative concernant les temps de réaction entre PA09 et PN09, alors que cette différence est hautement significative pour ce qui est de l'évaluation subjective de la compréhensibilité ($P = 0,007$). Nous pensons qu'il serait intéressant d'étudier ce type de différence plus en détail.

12. à ce propos, si nous séparons les données pour chaque locuteur, alors l'influence du nombre de déterminants communs aux noms cible et aux distracteurs ne reste significative que pour les deux locuteurs les plus difficiles à comprendre, PA08 et PA09 – ce qui va dans le sens de notre hypothèse.

FIGURE 7.23 – Différences significatives de temps de réaction chez les auditeurs, selon les locuteurs, pour la condition où l'objet 1 est correctement identifié

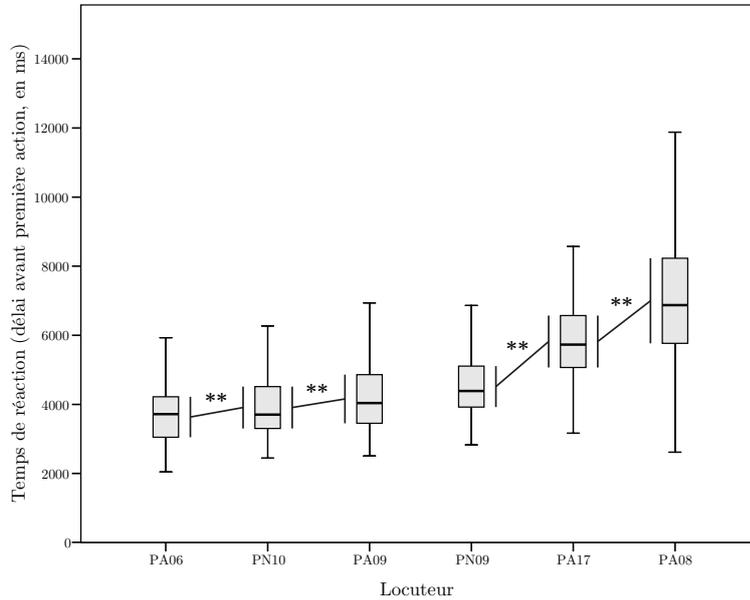
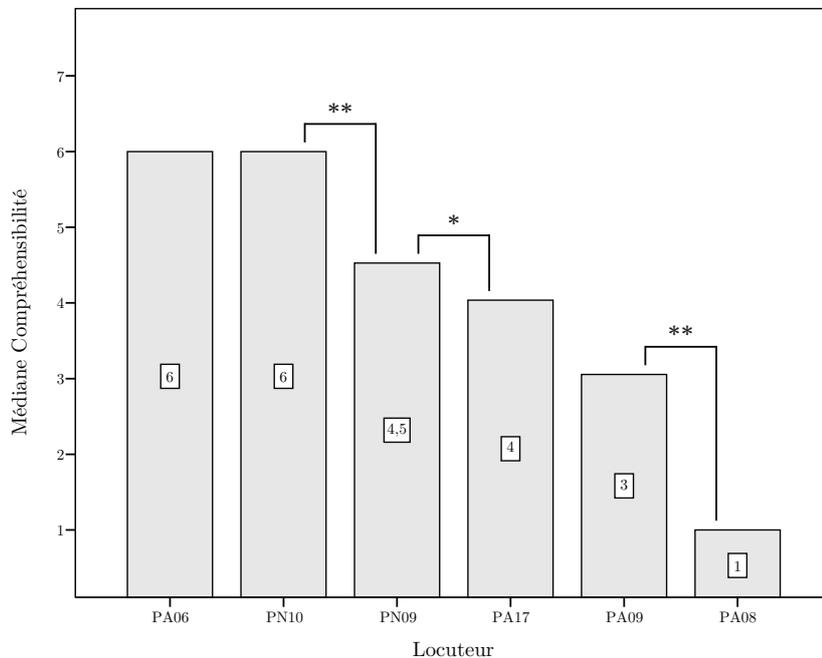


FIGURE 7.24 – Médianes des scores de compréhensibilité pour chaque locuteur, avec les différences significatives



Cependant, dans notre étude nous n'avons convoqué que 3 experts car il s'agissait principalement d'obtenir des indications pour la sélection des stimuli. Pour étudier précisément les différences entre les données obtenues par des évaluations subjectives et par des temps de réaction il faudrait certainement convoquer une population

d'experts plus importante.

Discussion générale : des différences entre l'intelligibilité et la compréhension de la parole en situation de communication

Dans cette discussion nous souhaitons revenir sur ce qui constitue selon nous le point central de ce travail : l'étude de la relation existant entre l'intelligibilité et la compréhension de la parole en situation de communication. Les résultats que nous avons obtenus montrent clairement que les scores de retranscription orthographique ne sont pas de bons prédicteurs des performances des auditeurs dans la tâche de compréhension que nous avons utilisée. En effet, notre étude a montré qu'il n'existe que de *faibles corrélations* entre les scores obtenus par une même population dans les deux types de tâches. Ces résultats, qui sont issus d'un protocole impliquant un contrôle strict des conditions de dégradation du signal de parole, contrastent avec les données qui ont été observées dans les études antérieures (Beukelman et Yorkston, 1979; Hustad, 2008).

Nos données indiquent que des différences existent bel et bien entre les tâches de retranscription orthographique et les tâches visant à évaluer la compréhension de l'auditeur en situation de communication, en dépit du fait que les deux procédures de test portent sur les mêmes stimuli de parole. Et l'importance de ces différences est telle qu'il n'est pas possible d'utiliser les résultats obtenus dans la première pour faire des hypothèses sur les résultats qu'obtiendraient les mêmes auditeurs dans la seconde.

Les implications cliniques de ce résultat sont très importantes, car les scores d'intelligibilité sont souvent utilisés comme des indicateurs de la performance communicative verbale des patients. Or nous observons que les résultats obtenus dépendent pour une grande partie de la tâche impliquée, ce qui signifie en d'autres termes que les scores d'intelligibilité mesurent davantage la capacité du patient à faire transcrire ses paroles qu'à être compris dans une situation de communication. Notre conclusion confirme le constat effectué par le personnel médical et qui est à l'origine de notre travail de thèse : les indicateurs d'intelligibilité habituellement utilisés ne représentent pas fidèlement l'aptitude du patient à communiquer. Nos résultats s'incrinvent donc comme un argument en faveur du développement de tests spécifiquement conçus pour évaluer cette dimension en particulier.

De plus, compte-tenu de l'importance des tâches convoquées dans les tests sur les performances obtenues par les locuteurs, il semble nécessaire de développer des procédures de test qui se rapprochent le plus possible de situations de communication naturelles. Bien entendu, d'un point de vue pratique cela suppose aussi que l'on prenne en compte les contraintes liées au domaine des troubles pathologiques de production de la parole (TPPP) : notamment les difficultés que peut représenter un tel type de test pour les patients, mais aussi les contraintes de temps ou de matériel auxquelles sont soumis les personnels médicaux. L'étude de faisabilité clinique que nous avons conduite a montré toute la difficulté de la question ; mais cette étude a aussi montré qu'il était envisageable de pallier les principales difficultés par un environnement, des consignes et des critères d'inclusion et d'exclusion adaptés.

D'autre part et d'un point de vue théorique, le fait que les scores d'intelligibilité ne soient pas représentatifs des performances du locuteur en situation de communication nous amène à nous interroger sur les aspects *communicatifs* et *cognitifs* qui différencient les deux types de tâches. En effet, on peut s'interroger tout d'abord sur les aspects du « fonctionnement communicatif verbal » que chacune des deux tâches permet d'appréhender chez le locuteur. Et, puisque notre problématique touche aux troubles *pathologiques* de production de la parole, il nous paraît intéressant de rapprocher ces aspects de la notion de *handicap*. C'est pourquoi nous souhaitons dans un premier temps traiter de la relation entre les tâches d'intelligibilité et le *handicap de communication verbale*.

Ensuite, il nous paraît important de souligner le fait que les différences que nous avons observées entre les tâches de retranscription orthographique et de compréhension de la parole en situation de communication concernaient *les performances des auditeurs*. Dans les deux procédures les stimuli de parole utilisés ne variaient pas : seule la tâche proposée à l'auditeur différait. Le fait que nous ayons relevé des différences importantes entre les scores obtenus peut donc indiquer que chacune des deux procédures met en jeu des processus de traitement cognitif différents chez l'auditeur. Nous traitons cet aspect dans la deuxième et dernière partie de cette discussion (p. 239).

Intelligibilité et handicap de communication verbale

Dans notre thèse nous avons insisté sur l'importance qu'il y a de considérer les objectifs pour lesquels les méthodes d'évaluation de l'intelligibilité ou de la compréhension de la parole sont utilisées. Ainsi dans notre premier chapitre nous avons vu que les tests d'intelligibilité ont été, au départ, élaborés en vue d'évaluer les performances de systèmes de transmission de la parole ou de canaux de communication (ex. salles). Dans cette optique particulière la parole est vue comme un *code* parmi d'autres, et l'évaluation des *media* de communication se fait en comparant les symboles émis et les symboles reçus – les phonèmes dans le cas de la parole. La majorité des mesures subjectives de l'intelligibilité conduites pour ces objectifs s'inscrivent ainsi dans un paradigme de lecture d'unités linguistiques (logatomes ou mots) et de retranscription orthographique.

Comme nous l'avons vu à travers le chapitre 2, ce paradigme de lecture et de retranscription a été ensuite porté au domaine clinique pour l'évaluation des troubles pathologiques de production de la parole (TPPP). Dans notre thèse nous sommes interrogé sur les différences entre ces deux domaines et leurs objectifs d'évaluation. En effet dans le cas de l'acoustique et de l'ingénierie des télécommunications le *canal de communication* est au centre de l'évaluation alors que pour le cas des TPPP ce sont les performances du *locuteur* qui sont en question. De plus, l'objectif de l'évaluation des *media* de communication est en général d'en améliorer les performances. Dans le cas des TPPP l'évaluation répond à plusieurs objectifs que nous proposons d'aborder sous l'éclairage du modèle apporté par Wood (1980) pour la classification internationale du handicap (CIH – cf. figure 7.25).

FIGURE 7.25 – Modèle de Wood (1980) pour la classification internationale du handicap (CIH)



Dans le modèle de Wood :

- les *déficiences* renvoient à des pertes, ou des altérations de structures ou de fonctions anatomiques, physiologiques ou psychologiques ;
- les *incapacités* correspondent à la réduction partielle ou totale de la capacité d'accomplir une activité d'une façon ou dans les limites considérées comme normales pour un être humain ;
- le *désavantage social* (ou *handicap*) renvoie au fait que l'individu est limité ou interdit dans l'accomplissement d'un rôle normal (les « rôles » normaux étant fonction de l'âge de l'individu, de son sexe, de son milieu socio-culturel, etc.).

Si l'on met en rapport le modèle de Wood avec le cas d'un patient atteint de TPPP, dû par exemple à l'ablation chirurgicale d'une partie de sa langue :

- la *déficience* est la perte d'une partie de son anatomie, à savoir une partie de la langue ;
- les *incapacités* correspondent à la réduction de la capacité du patient à accomplir des mouvements articulatoires ;
- le *désavantage social* (ou *handicap*) correspond à toutes les limitations ou les interdits que le patient subit dans les rôles sociaux qu'il est amené à tenir, du fait de ses difficultés d'articulation.

Nous avons vu à travers notre état de l'art sur les méthodes d'évaluation utilisées dans le cadre des TPPP (chapitre 2) que les tests d'intelligibilité sont utilisés pour donner des indications quantitatives – et plus rarement qualitatives – à propos de chacun de ces trois niveaux :

- les scores d'intelligibilité sont parfois présentés comme des indices de la sévérité des déficiences dont les patients sont atteints (Keintz *et al.*, 2007; Hustad,

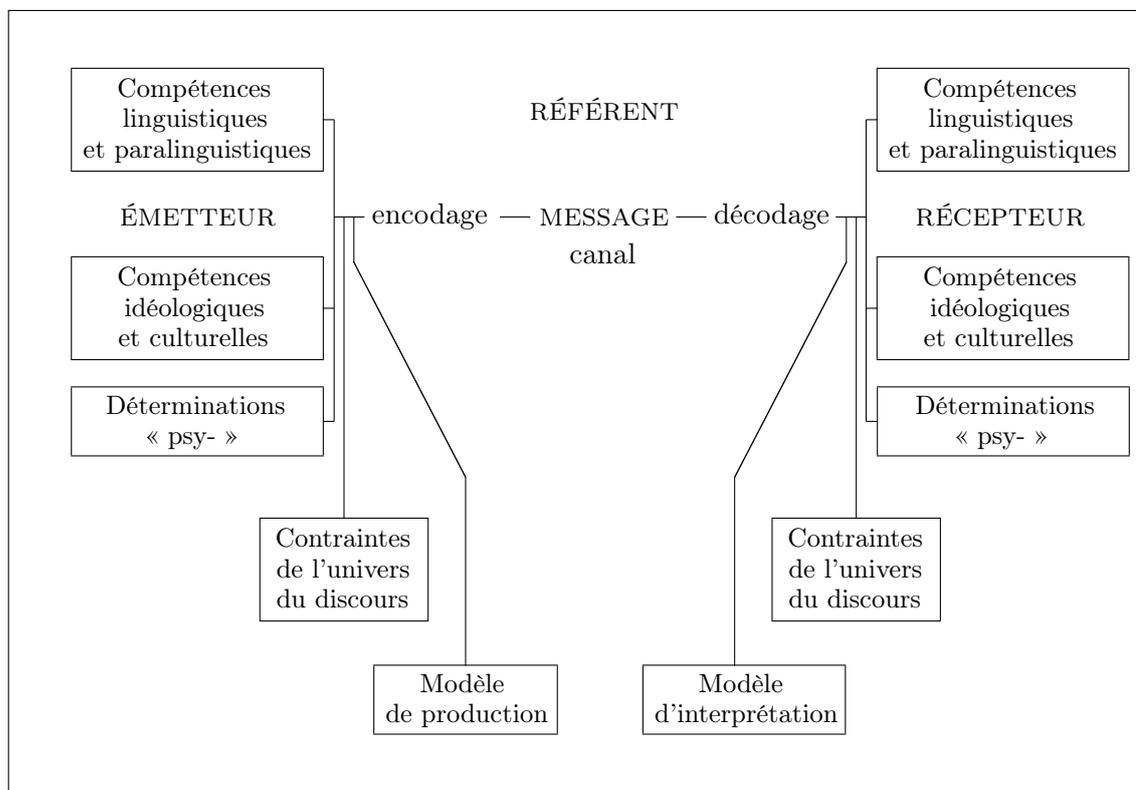
2007b, 2008; Özsancak, 2001). En effet dans beaucoup de cas nous ne disposons pas de critères objectifs pour quantifier ces déficiences, par exemple dans le cas d'atteintes neurologiques ou bien quand nous nous intéressons à des patients souffrant de troubles d'origines différentes. Nous ne sommes pas capable par exemple d'estimer qu'un patient est atteint d'une déficience de x % dans les fonctions neurologiques liées à l'exécution des mouvements phonatoires, et encore moins de rapprocher cette déficience avec celle d'un autre patient ayant perdu y % de la masse de sa langue. Les scores d'intelligibilité sont donc parfois utilisés pour quantifier ces déficiences, ce qui peut causer des biais méthodologiques. Nous avons par exemple relevé dans notre travail deux cas d'études dans lesquelles les scores d'intelligibilités sont utilisés à la fois pour représenter la sévérité des déficiences et l'intelligibilité des patients (Hustad, 2007b, 2008). Ceci pose un problème majeur dans la mesure où la sévérité est envisagée comme une variable *indépendante* de l'intelligibilité des patients ;

- les tests d'intelligibilité sont également utilisés pour quantifier les *incapacités* dont souffrent les patients pour accomplir des mouvements articulatoires, l'évolution de l'intelligibilité au cours du temps permettant de juger de l'efficacité de thérapies orthophoniques ou d'interventions chirurgicales (Hustad, 2006; Leary *et al.*, 2006; Lohmander *et al.*, 2006; Neel, 2009; Puyuelo et Rondal, 2005). Plus rarement, les tests d'intelligibilité permettent aussi d'obtenir des indications *qualitatives* sur ces incapacités – notamment sur l'aptitude du patient à produire les contrastes phonétiques jugés comme nécessaires pour la bonne identification des phonèmes (Kent *et al.*, 1989) ;
- enfin, les tests d'intelligibilité sont utilisés pour rendre compte de la *performance communicative* des patients, c'est-à-dire leur aptitude à se faire comprendre par des tiers dans des situations de communication (Ertmer, 2010; Skahan *et al.*, 2007).

Ce dernier point, la question de l'évaluation de la performance communicative d'une personne, est très complexe. Tout d'abord, la communication humaine est multimodale – aux paroles venant s'ajouter les gestes, mimiques et postures des interlocuteurs. Ce sont entre autres ces aspects que Kerbrat-Orecchioni (1980) désignera sous le terme de « compétences paralinguistiques » aux côtés des autres compétences en jeu chez les interlocuteurs dans une situation de communication (compétences linguistiques, idéologiques et culturelles – *cf.* le schéma de la communication de Kerbrat-Orecchioni en figure 7.26).

De fait, dans notre travail de thèse nous nous sommes limité à la question de la compétence de communication *verbale* en éliminant volontairement les aspects non verbaux de la communication. Notre problématique concerne en effet les troubles de production de la *parole* uniquement, et de plus pour être à même de formuler des hypothèses suffisamment précises quant aux différences existant entre l'intelligibilité et la compréhension de la parole en situation de communication nous nous devons de limiter le nombre de variables en jeu. Nous pensons cependant que les performances du patient en termes de posturo-mimo-gestualité devraient également être prises en compte pour obtenir une image précise et complète de la performance communicative du patient. En ce sens, du point de vue des modalités communicatives, nous nous

FIGURE 7.26 – Schéma de la communication de Kerbrat-Orecchioni (1980, p. 19)



sommes limité à des situations de communication *uniquement verbales*, comme cela peut être le cas lors d'une conversation téléphonique par exemple. Notre travail mériterait cependant d'être développé pour aller vers la prise en compte des éléments non verbaux de la communication.

Ensuite, vouloir quantifier ou qualifier le handicap de communication verbale suppose que l'on ait préalablement défini les *fonctions* que remplit la communication verbale. Aujourd'hui dans le secteur de la santé le handicap tend d'ailleurs à ne plus être défini en termes de « déficiences » ou d'« incapacités » mais de *fonctionnement* (cf. Woisard-Bassols, 2011). Jamet (2003) définit le fonctionnement comme suit :

Le terme *fonctionnement* [...] renvoie à l'ensemble des fonctions de la personne humaine, c'est-à-dire les fonctions organiques, les fonctions liées aux activités de la personne et les fonctions qui permettent sa participation au sein de la société.

Cette perspective *fonctionnelle* sur la communication verbale nous a naturellement amené à considérer les fonctions du langage définies par Jakobson (1963) comme un cadre pertinent pour notre réflexion. Nous avons souligné le fait que les tests d'intelligibilité portant sur la retranscription d'unités linguistiques se concentrent sur la fonction *métalinguistique* du langage – ce qui n'est pas étonnant lorsque l'on considère que ces méthodes ont été développées au départ pour mesurer l'efficacité dans le transfert de *code* par des canaux ou des systèmes de transmission.

On peut cependant imaginer que, dans le cas des TPPP, toutes les autres fonctions définies par Jakobson puisse être touchées au point de présenter un handicap pour les sujets :

- la fonction *référentielle* : le sujet peut avoir des difficultés pour faire comprendre à un tiers les références qu'il fait, par la parole, à des objets, à des personnes, à des événements...
- la fonction *conative* : le sujet peut être gêné dans sa capacité à agir sur son interlocuteur. On peut imaginer par exemple qu'un sujet qui a perdu ses cordes vocales ne puisse plus utiliser le caractère distinctif que peut jouer l'intonation en français pour indiquer à son interlocuteur qu'il lui donne un ordre (« tu viens ! »), qu'il lui pose une question (« tu viens ? ») ou qu'il affirme simplement quelque chose (« tu viens. ») ;
- la fonction *expressive* : nous pensons en particulier au handicap qui peut être causé par les difficultés du sujet à maîtriser les aspects prosodiques de la parole. Le rythme et l'intonation notamment, qui peuvent être largement mis à contribution dans l'expression de sensations et sentiments ;
- la fonction *phatique* : cette fonction s'appuie également pour une large part sur des indices prosodiques : intonations montantes pour s'assurer que la communication avec l'interlocuteur est toujours établie, pauses remplies pour conserver le tour de parole, etc. On peut donc imaginer que cette fonction soit particulièrement touchée chez des personnes ayant perdu des aptitudes de production de la parole au niveau suprasegmental ;
- la fonction *poétique* peut être touchée par des dysfonctionnements au niveau segmental comme suprasegmental : tout ce qui limite le locuteur dans les choix qu'il peut faire quant à la forme de son message (par exemple adopter un rythme saccadé ou un débit lent, jouer sur le voisement d'une consonne, choisir d'utiliser un mot en particulier, etc.) représente un handicap vis-à-vis de la fonction poétique du langage.

À travers le prisme des différentes fonctions du langage nous voyons que la question de l'évaluation du handicap de communication verbale dans le cas des TPPP est extrêmement complexe. Outre le fait que nous produisons la plupart du temps des énoncés mettant en jeu plusieurs des fonctions définies par Jakobson (1963), il se pose la questions du « poids » qu'il faudrait attribuer à chacune de ces fonctions pour rendre compte du handicap de communication verbale. Nous pensons que ce poids ne peut être défini dans l'absolu : tout comme Wood (1980) dit que le handicap ne peut être défini sans tenir compte de variables socio-culturelles propres à l'individu, nous pensons que le handicap de communication verbale ne peut être évalué sans prendre en compte la façon dont *chaque* individu utilise la langue. Ainsi le caractère de la personne peut avoir son importance : une personne ayant des difficultés à exprimer des sentiments par la parole (fonction *expressive* du langage) sera d'autant plus gênée dans son fonctionnement social qu'elle dispose d'un caractère porté vers l'expressivité des émotions et sentiments – s'opposant en cela à une personne de caractère plutôt froid et introverti. De même, la nature des activités que la personne mène à titre privé ou professionnel peut jouer sur l'importance de chacune de ces fonctions pour elle : il suffit de s'imaginer la perte que pourrait représenter

l'incapacité de voiser les consonnes pour un comique très porté sur les jeux de forme – dans la veine de feu Raymond Devos, par exemple – pour comprendre que la fonction *poétique* du langage est cruciale pour certaines personnes. Enfin, la complexité de la question s'accroît à nouveau si l'on considère qu'en plus de l'importance de chacune de ces fonctions pour la personne, il faudrait également prendre en compte leur aptitude à utiliser des stratégies pour pallier leurs difficultés – et également observer dans quelle mesure les sujets acceptent ces stratégies palliatives. Dans le cas de la fonction *conative* par exemple, le locuteur ne doit pas nécessairement utiliser le caractère distinctif de l'intonation en français pour faire comprendre ses intentions à son interlocuteur : il peut ainsi utiliser une stratégie consistant à marquer systématiquement le caractère interrogatif de ses énoncés en utilisant des constructions comme « est-ce que » (« Est-ce que tu viens ? »).

Une des caractéristiques de notre travail a été d'adopter un point de vue *fonctionnel* sur la question du handicap de communication induit par les TPPP, et l'éclairage apporté par les fonctions du langage de Jakobson nous a semblé particulièrement pertinent pour appréhender toute la complexité de cette question. En revanche cette perspective montre également les limites de notre travail de thèse, qui s'est concentré sur la comparaison de tâches portant sur le transfert de code linguistique – c'est à-dire des tâches métalinguistiques – et des tâches de communication référentielle – notre test de compréhension verbale consistant à observer si l'auditeur parvient à résoudre les références que fait le locuteur à propos d'images représentées sur un écran, et d'actions de déplacement. D'un autre côté, le fait que nous nous soyons concentré sur ce type de tâche nous a amené à nous poser la question des *processus cognitifs* impliqués chez l'auditeur dans une tâche plutôt orientée vers le code, et une tâche orientée vers la compréhension des énoncés.

Aspects cognitifs en jeu dans la compréhension de stimuli verbaux en contexte

Le fait que les scores d'intelligibilité et de compréhension présentent de telles différences, alors qu'ils s'agit des mêmes auditeurs qui ont traité les mêmes stimuli verbaux, interroge sur les aspects cognitifs qui sont impliqués chez l'auditeur dans ces deux types de tâches.

Nous avons vu dans notre état de l'art sur les études psycholinguistiques de la perception et de la compréhension de la parole que de nombreux travaux ont montré l'existence de processus descendants dans le traitement des stimuli de parole. Il semble maintenant acquis que les informations fournies par le contexte de communication, notamment, développent des attentes chez l'auditeur.

Dans nos expériences cela transparait en premier lieu du fait que les performances des auditeurs dans la tâche de compréhension en contexte sont largement supérieures à leurs performances dans la tâche de retranscription orthographique – suggérant par là que les auditeurs s'appuient sur le contexte pour traiter les stimuli de parole. De plus, nous avons observé que cette différence entre les scores n'est pas fixe mais

évolue en fonction des conditions de dégradation du signal de parole. La différence entre les scores obtenus dans les deux tâches s'accroît en effet en fonction du niveau de bruit.

Cette dernière observation va dans le sens de la théorie de Lindblom (1990b, cf. p. 172) qui postule que l'auditeur utilise des stratégies d'écoute différentes en fonction des conditions de dégradation du signal de parole. Selon Lindblom (1990b), plus les conditions d'écoute sont difficiles, plus l'auditeur va s'appuyer sur les informations extra linguistiques. D'après ce que nous avons pu observer dans l'expérience de compréhension de la parole dans le bruit, non seulement cette tendance apparaît clairement, mais nous rajouterions également que, plus les conditions de dégradation du signal augmentent, et plus les différences de stratégies *inter auditeurs* apparaissent. Les écarts-types entre les scores obtenus par les différents auditeurs s'accroissent en effet de manière constante en fonction de niveau de bruit, une tendance que l'on ne retrouve pas dans les scores d'intelligibilité.

Nous avons fait également ces deux observations lors de notre étude de faisabilité clinique. Dans ce cas nous ne pouvions contrôler de manière objective la dégradation du signal, mais, si l'on observe la façon dont les informations contextuelles ont été prises en compte par les auditeurs, de nettes différences apparaissent en fonction de la compréhensibilité des locuteurs. Ainsi, lorsque les auditeurs ont écouté les phrases prononcées par les locuteurs les moins compréhensibles, les variables liées à la prise en compte du contexte (par exemple la proximité phonologique entre les mots cibles et les distracteurs) a joué un rôle nettement supérieur que lors de l'écoute des patients aisément compréhensibles.

Nous pensons qu'il serait intéressant de continuer à étudier les différentes stratégies d'écoute des auditeurs, à la fois d'un point de vue théorique, mais surtout d'un point de vue pratique. En effet, si l'on parvient à définir précisément les stratégies d'écoute qu'utilisent les auditeurs pour comprendre, en contexte, des signaux de parole très dégradés, alors on peut imaginer mettre à profit ces stratégies en développant *chez le patient* une aptitude à fournir à l'auditeur les indices les plus pertinents pour faciliter la compréhension de l'auditeur.

Conclusion

Nous avons apporté plusieurs éléments de réponse à la demande exprimée par les orthophonistes et les phoniâtres qui constatent dans leur pratique que les tests destinés à évaluer les performances des locuteurs, les *tests d'intelligibilité*, ne traduisent pas fidèlement la performance de communication des patients souffrant de troubles pathologiques de production de la parole. Cette observation remet en cause la représentativité des tests d'intelligibilité, concentrés sur la perception du code phonémique par les auditeurs, vis-à-vis de leur aptitude à comprendre le patient dans des situations de communication concrètes au quotidien.

La question centrale de notre thèse concernait donc davantage le traitement de la parole pathologique par des auditeurs que la performance de production de parole des patients. Il s'agissait en effet de savoir si les traitements cognitifs engagés par l'auditeur dans les tests d'intelligibilité sont représentatifs de ceux qu'il utilise dans une véritable situation de communication avec le patient.

Au vu de notre état de l'art en psycholinguistique, le lien postulé entre l'intelligibilité du patient et la compréhension de l'auditeur n'est absolument pas évident. Nous avons vu que la question des unités phonologiques servant de point d'ancrage à l'auditeur pour percevoir la parole suscite de nombreux travaux et débats au sein du paradigme du traitement de l'information, certains modèles récents allant jusqu'à suggérer que le lexique ne serait pas constitué de représentations phonologiques abstraites mais d'exemplaires acoustiques. Les tâches de retranscription orthographique, qui privilégient les traitements liés à la perception phonémique chez l'auditeur, constituent de ce point de vue un *choix*, dont nous ne pouvons pas encore évaluer la pertinence.

D'autre part, de nombreuses données psycholinguistiques montrent que lorsque l'auditeur traite des signaux de parole il intègre les informations liées au contexte de communication (identité de l'interlocuteur, objets présents dans l'environnement, objectifs de la communication...) et ce dès les premiers traitements perceptifs. Ces données suggèrent que la prise en compte des informations contextuelles serait *inhérente* au processus de compréhension de la parole, tout comme cela est postulé au sein du courant de la pragmatique. Dans cette perspective, les tests d'intelligibilité, qui reposent sur des tâches de retranscription hors contexte, ne sont pas représentatifs d'une situation de communication canonique – et il est donc possible que les scores d'intelligibilité ne soient pas de bons prédicteurs de la performance communicative des locuteurs.

Notre objectif a donc été dans un premier temps de vérifier cette hypothèse, afin

de confirmer ou d'infirmer l'observation faite par le personnel médical et qui est à l'origine de notre travail. Comme nous l'avons vu, cette question de la relation entre scores d'intelligibilité et compréhension de l'auditeur a été l'objet de quelques études scientifiques seulement. Beukelman et Yorkston (1979) ainsi que Hustad (2008) se sont intéressés à la parole dysarthrique et ont comparé des scores d'intelligibilité et de compréhension orale. Les deux études ont mis au jour des résultats opposés, à savoir une corrélation parfaite *vs.* une absence de corrélation entre les scores. Ceci peut certainement s'expliquer de par les difficultés qu'il y a à contrôler la sévérité des troubles, une variable qui influence à la fois les scores d'intelligibilité et de compréhension et vient donc parasiter le calcul de corrélation.

Pour pouvoir dépasser ce problème nous avons choisi d'utiliser comme stimuli des signaux de parole *non pathologique*, que nous avons dégradés artificiellement et de manière contrôlée en les diffusant avec des signaux de bruit (brouhaha) dont le niveau était variable. Pour pouvoir évaluer la compréhension des auditeurs, nous avons créé le logiciel EloKanz, qui permet d'enregistrer des commandes verbales de déplacement d'images et de vérifier leur compréhensibilité en observant les réactions des auditeurs.

Nos résultats montrent que les scores d'intelligibilité (retranscription orthographique) sont faiblement corrélés aux scores de compréhension. Ceci contraste avec les résultats des études précédentes, puisqu'ils montrent qu'une relation existe bel et bien entre l'intelligibilité et la compréhension de la parole en situation de communication, mais que de grandes différences séparent ces deux types de tâches. Nous avons tenté d'expliquer ces différences de par les traitements cognitifs que les deux tâches impliquent chez les auditeurs. En particulier, nos données paraissent indiquer que la fréquence des noms ne jouerait pas le même rôle lors de la transcription orthographique et lors de la réponse aux commandes verbales ; de même, il semble que la tâche de compréhension implique l'utilisation de stratégies d'autant plus variées entre les auditeurs que les conditions d'écoute se dégradent, alors que la variabilité des scores d'intelligibilité diminue pour ces mêmes conditions.

Ces différences montrent par ailleurs que les scores de retranscription ne constituent pas des indicateurs fiables de la performance des patients dans des situations de communication. Les implications cliniques sont importantes, dans la mesure où seuls des tests d'intelligibilité sont pratiqués aujourd'hui pour évaluer la performance communicative des patients. Il paraît donc nécessaire de développer des tests spécialement destinés à l'évaluation de cette dimension particulière.

Le logiciel EloKanz pourrait représenter une solution intéressante, d'autant que nos expériences ont montré que la passation des auditeurs pouvait être 5 fois plus rapide que pour une tâche de retranscription – et nous savons que la rapidité de mise en œuvre est un facteur déterminant pour la praticabilité clinique. Néanmoins, EloKanz implique pour les sujets de répondre à des consignes complexes, ce qui pourrait compromettre son utilisation, et plus particulièrement par les locuteurs qui souffrent parfois de pathologies lourdes. C'est pourquoi nous avons décidé de mener une étude de faisabilité clinique du logiciel dans une unité hospitalière dédiée aux troubles de la voix et de la déglutition.

Nos résultats ont montré que la majorité des patients étaient capables de compléter l'enregistrement de 60 phrases *via* le logiciel EloKanz, même si d'importantes différences existent entre leurs performances en termes de réponse aux consignes et de durée de passation. Cependant, l'étude des facteurs de variation a montré qu'il était possible de prédire ces performances en prenant en compte de l'âge des patients et leur score au *Mini Mental State Examination* (Folstein *et al.*, 1975). Une étude auprès d'une population plus importante pourrait permettre de définir des critères d'inclusion adaptés à la pratique clinique. De plus, nous avons montré que les temps de réaction des auditeurs relevés par le logiciel EloKanz étaient fortement corrélés à une évaluation subjective de la compréhensibilité des patients par des cliniciens habitués à ce genre de procédure. Ce dernier résultat contribue à considérer la tâche proposée dans EloKanz comme valide. Cependant, pour pouvoir prendre en compte les temps de réaction des auditeurs comme des indicateurs fiables de la compréhensibilité des patients, les effets de différentes variables devraient être contrôlés. Nos données montrent en effet que des variables liées aux auditeurs (âge, nombre d'années d'exposition à la parole pathologique) et aux stimuli (fréquence des noms cibles, proximité phonologique entre noms cibles et distracteurs) ont un effet statistiquement significatif sur les temps de réaction des auditeurs.

Une première perspective « pratique » serait donc de poursuivre ce travail en convoquant un panel d'auditeurs plus large, afin de mieux cerner l'influence des variables que nous avons mises au jour et de pouvoir envisager la standardisation du test EloKanz. Également, ce test est pour l'instant limité du point de vue des aspects linguistiques et communicatifs qu'il met en jeu puisqu'il se concentre sur la fonction *référentielle* du langage, et ne convoque qu'une seule structure syntaxique et prosodique. Même si cette restriction était nécessaire au vu de nos objectifs de recherche, nous pensons qu'un travail de développement est nécessaire pour intégrer d'autres composantes communicatives et linguistiques.

Ensuite, les différences que nous avons observées entre l'intelligibilité et la compréhension de la parole en situation de communication posent des hypothèses fortes quant aux stratégies d'écoute utilisées par les auditeurs, qui semblent dépendre en grande partie des objectifs de la tâche et des conditions de dégradation de la parole – une observation qui rejoint celle effectuée par Liss (2007) pour la parole de locuteurs dysarthriques. Une perspective intéressante des points de vue théorique et pratique serait d'observer de plus près ces stratégies, en particulier lorsque les auditeurs communiquent avec des patients souffrant de TPPP. Ce travail permettrait d'une part de contribuer à adopter une perspective *dynamique* sur les processus de traitement de la parole, puisque dépendant du contexte, des objectifs de communication, ainsi que des conditions de dégradation de la parole. D'autre part, caractériser ces stratégies permettrait enfin de définir, pour le locuteur, les stratégies palliatives qui sont les plus efficaces en fonction de la situation de communication et de la gravité de l'atteinte. Ce travail pourrait donc s'inscrire dans la continuité des travaux de développement de stratégies de complétion de la parole par des gestes ou des indices contextuels (voir par ex. Hustad et Beukelman, 2001, 2002; Hustad *et al.*, 2003; Garcia et Cannito, 1996; Garcia et Dagenais, 1998).

Bibliographie

- ALARIO, F.-X. et FERRAND, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavioural Research Methods, Instruments & Computers*, 31(3):531–552.
- ALLEN, J. B. (2004). The articulation index is a shannon channel capacity. In PRESSNITZER, D., de CHEVEIGNÉ, A., MCADAMS, S. et COLLET, L., éditeurs : *Auditory signal processing: Physiology, psychoacoustics, and models*, pages 314–320. Springer Verlag, New York.
- AMANO, S., SAKAMOTO, S., KONDO, T. et SUZUKI, Y. (2009). Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese. *Speech Communication*, 51(1):76–82.
- ANDERSON, A. H., BADER, M., BARD, E. G., BOYLE, E., DOHERTY, G., GARROD, S., ISARD, S., KOWTKO, J., MCALLISTER, J., MILLER, J., SOTILLO, C., THOMPSON, H. S. et WEINIERT, R. (1991). The HCRC Map Task Corpus. *Language & Speech*, 34(4):351–366.
- ANDERSON, B. W. et KALB, J. T. (1987). English verification of the STI method for estimating speech intelligibility of a communications channel. *Journal of the Acoustical Society of America*, 81(6):1982–1985.
- ANSI S3.5 (1997). *American national standard methods for the calculation of the Speech Intelligibility Index*. American National Standards Institute, New York.
- ARNON, I. et SNIDER, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.
- AUSTIN, J. L. (1962). *How to do things with words*. Harvard University Press, Cambridge.
- AUZOU, P., ÖZSANCAK, C., JAN, M., LÉONARDON, S., MÉNARD, J. F., GAILLARD, M. J., EUSTACHE, F. et HANNEQUIN, D. (1998). Evaluation Clinique de la Dysarthrie : présentation et validation d’une méthode. *Revue Neurologique*, 154(7):523–530.
- BAREFOOT, S. M., BOCHNER, J. H., JOHNSON, B. A. et vom EIGEN, B. A. (1993). Rating deaf speakers’ comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology*, 2(3):31–35.

- BARTON, S. B. et SANFORD, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21(4):477–487.
- BEIJERING, K., GOOSKENS, C. et HEERINGA, W. (2008). Modeling intelligibility and perceived linguistic distances by means of the Levenshtein algorithm. In van KOPPEN, M. et BOTMA, B., éditeurs : *Linguistics in the Netherlands 2008*, pages 13–24, Amsterdam. John Benjamins.
- BELL, A. G. (1876). Researches in telephony. *Proceedings of the American Academy of Arts and Sciences*, 12(1-10):8.
- BELTYUKOVA, S. A., STONE, G. M. et ELLIS, L. W. (2008). Rasch analysis of word identification and magnitude estimation scaling responses in measuring naïve listeners' judgments of speech intelligibility of children with severe-to-profound hearing impairments. *Journal of Speech, Language, and Hearing Research*, 51(5): 1124–1137.
- BENOIT, C., MOHAMADI, T. et KANDEL, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37(5):1195–1203.
- BEUKELMAN, D. R. et YORKSTON, K. M. (1979). The relationship between information transfer and speech intelligibility of dysarthric speakers. *Journal of Communication Disorders*, 12(3):189–196.
- BEUKELMAN, D. R. et YORKSTON, K. M. (1980). Influence of passage familiarity on intelligibility estimates of dysarthric speech. *Journal of Communication Disorders*, 13(1):33–41.
- BEVERLY, D., CANNITO, M. P., CHORNA, L., WOLF, T., SUITER, D. M. et BENE, E. R. (2010). Influence of stimulus sentence characteristics on speech intelligibility scores in hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology*, 18(4):9–13.
- BONIN, P., PEEREMAN, R., MALARDIER, N., MÉOT, A. et CHALARD, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments & Computers*, 35(1):158–167.
- BORNKESSEL-SCHLESEWSKY, I. D. et FRIEDERICI, A. D. (2007). Neuroimaging studies of sentence and discourse comprehension. In GASKELL, M. G., éditeur : *The Oxford Handbook of Psycholinguistics*, pages 407–424, Oxford. Oxford University Press.
- BORNKESSEL-SCHLESEWSKY, I. D. et SCHLESEWSKY, M. (2008). An alternative perspective on 'semantic p600' effects in language comprehension. *Brain Research Reviews*, 59(1):55–73.

- BORNKESSEL-SCHLESEWSKY, I. D. et SCHLESEWSKY, M. (2009a). *Processing syntax and morphology: A neurocognitive perspective*. Oxford University Press, Oxford.
- BORNKESSEL-SCHLESEWSKY, I. D. et SCHLESEWSKY, M. (2009b). The role of prominence information in the real time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, 3(1):19–58.
- BOUGNOUX, D. (1993). *Sciences de l'information et de la communication*. Textes essentiels. Larousse, Paris.
- BROWN-SCHMIDT, S., CAMPANA, E. et TANENHAUS, M. K. (2005). Real-time reference resolution in a referential communication task. In TRUESWELL, J. C. et TANENHAUS, M. K., éditeurs : *Processing world-situated language: Bridging the language-as-action and language-as-product traditions*, Cambridge. MIT Press.
- BURDA, A. N. et HAGEMAN, C. F. (2005). Perception of accented speech by residents in assisted living facilities. *Journal of Medical Speech-Language Pathology*, 13(1): 7–14.
- BURDA, A. N., SCHERTZ, J. A., HAGEMAN, C. F. et EDWARDS, H. T. (2003). Age and understanding speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills*, 97(1):11–20.
- BURTON-ROBERTS, N. (2007). *Pragmatics*. Palgrave Advances in Linguistics. Palgrave Macmillan, New York.
- CALLIOPE (1989). *La parole et son traitement automatique*. Masson, Paris.
- CAMPBELL, G. A. (1910). Telephonic intelligibility. *Philosophical Magazine*, 19(6): 152–159.
- CARLSON, R., GRANSTRÖM, B., NEOVIUS, L. et NORD, L. (1992). The 'listening speed' paradigm for synthesis evaluation. In HUBER, D., éditeur : *Fonetik '92, the Sixth Swedish Phonetics Conference held in Gothenburg*, pages 63–66, Göteborg. Chalmers University of Technology.
- CARTIER, M. et ROSSI, M. (1975). Le test de diagnostic par paires minimales : mise en œuvre et résultats. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 2:1–30.
- CHENG, C.-C. (1993). *Quantifying Dialect Mutual Intelligibility* – Thèse de doctorat. Illinois University.
- CHERRY, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5):975–979.
- CHRISTOPHE, A., PALLIER, C., BERTONCINI, J. et MEHLER, J. (1991). A la recherche d'une unité : segmentation et traitement de la parole. *L'Année Psychologique*, 91(1):59–86.

- CIENKOWSKI, K. M. et SPEAKS, C. (2000). Subjective *vs.* objective intelligibility of sentences in listeners with hearing loss. *Journal of Speech and Hearing Research*, 43(5):1205–1210.
- CLARK, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge.
- COX, R. M. et MCDANIEL, D. M. (1989). Development of the Speech Intelligibility Rating (SIR) test for hearing aid comparisons. *Journal of Speech and Hearing Research*, 32(2):347–352.
- CROCHEMORE, E. et VANNIER, F. (2001). Analyse phonétique de la parole dysarthrique. In AUZOU, P., ÖZSANCAK, C. et BRUN, V., éditeurs : *Les dysarthries*, pages 71–82. Masson, Paris.
- CULICOVER, P. W. et JACKENDOFF, R. (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences*, 10(9):413–418.
- CULLINAN, W. L., BROWN, C. S. et BLALOCK, P. D. (1986). Ratings of intelligibility of esophageal and tracheoesophageal speech. *Journal of Communication Disorders*, 19(3):185–195.
- CUMMINGS, L. (2005). *Pragmatics. A multidisciplinary perspective*. Edinburgh University Press, Édimbourg.
- CUTLER, A., MEHLER, J., NORRIS, D. et SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4):385–400.
- CYCOWICZ, Y. M., FRIEDMAN, D., ROTHSTEIN, M. et SNODGRASS, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65(2):171–237.
- DE BODT, M. S., HERNÁNDEZ-DÍAZ HUICI, M. E. et VAN DE HEYNING, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, 35(3):283–292.
- DE RENZI, E. et VIGNOLO, L. A. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, 85(4):665–678.
- de SAUSSURE, F. (1995). *Cours de linguistique générale*. Payot & Rivages, Paris. 1916 pour l'édition originale.
- DECKER, L., RUBENSTEIN, H. et POLLACK, I. (1958). Word frequency and speech intelligibility for unknown and known message sets. *Journal of the Acoustical Society of America*, 30(7):672–673.
- DEPAUL, R. et KENT, R. D. (2000). A longitudinal case study of ALS: Effects of listener familiarity and proficiency on intelligibility judgments. *American Journal of Speech-Language Pathology*, 9(3):230–240.

- DIMITROPOULO, M., DUÑABEITIA, J. A., BLITSAS, P. et CARREIRAS, M. (2009). A standardized set of 260 pictures for modern Greek: Norms for name agreement, age of acquisition and visual complexity. *Behavior Research Methods*, 41(2):584–589.
- DOYLE, P. C., DANHAUER, J. L. et REED, C. G. (1988). Listeners' perceptions of consonants produced by esophageal and tracheoesophageal talkers. *Journal of Speech and Hearing Disorders*, 53(4):400–407.
- DOYLE, P. C. et LEEPER, H. A. (1997). Dysarthric speech: A comparison of computerized speech recognition and listener intelligibility. *Journal of Rehabilitation Research and Development*, 34(3):309–316.
- DOYLE, P. C., SWIFT, E. R. et HAAF, R. G. (1989). Effects of listener sophistication on judgments of tracheoesophageal talker intelligibility. *Journal of Communication Disorders*, 22(2):105–113.
- DUEZ, D. (1982). Silent and non silent pauses in three speech styles. *Language and Speech*, 25(1):11–28.
- DUEZ, D. (1987). *Contribution à l'étude de la structuration temporelle de la parole en français* – Thèse de Doctorat. Université de Provence.
- DUEZ, D. (1992). Second formant locus-nucleus patterns: An investigation of spontaneous French speech. *Speech Communication*, 11(4-5):417–427.
- EGAN, J. P. (1948). Articulation testing methods. *Laryngoscope*, 58(9):955–981.
- ELLIS, L. W. et BELTYUKOVA, S. A. (2008). Effects of training on naïve listeners' judgments of the speech intelligibility of children with severe-to-profound hearing loss. *Journal of Speech, Language, and Hearing Research*, 51(5):1114–1123.
- ENDERBY, P. M. (1983). *Frenchay Dysarthria Assessment*. College-Hill Press, San Diego.
- ERICKSON, T. et MATTSON, M. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5):540–551.
- ERTMER, D. J. (2010). Relationships between speech intelligibility and word articulation scores in children with hearing loss. *Journal of Speech and Hearing Research*, 53(5):1075–1086.
- FAIGET, L. (1997). *Séparation de l'influence du local et de l'enceinte pour la prévision de l'intelligibilité dans des conditions d'écoute difficiles* – Thèse de doctorat. Université de Toulouse III.
- FAIGET, L. et RUIZ, R. (1999). Speech intelligibility model including room and loudspeaker influences. *Journal of the Acoustical Society of America*, 105(6):3345–3354.

- FAIRBANKS, G. (1958). Test of phonemic differentiation: The Rhyme Test. *The Journal of the Acoustical Society of America*, 30(7):596–600.
- FERREIRA, F., BAILEY, K. et FERRARO, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1):11–15.
- FLETCHER, H. et STEINBERG, J. C. (1929). Articulation testing methods. *Bell System Technical Journal*, 8(4):806–854.
- FLIPSEN, P. (1995). Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders*, 28(1):3–19.
- FOLSTEIN, M. F., FOLSTEIN, S. E. et MCHUGH, P. R. (1975). “Mini-Mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198.
- FRAUENFELDER, U. H. et NGUYEN, N. (2000). La reconnaissance des mots parlés. In RONDAL, J. et SERON, X., éditeurs : *Troubles du langage : bases théoriques, diagnostic et rééducation*, pages 213–240. Mardaga, Bruxelles.
- FRENCH, N. R. et STEINBERG, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America*, 19(1):90–119.
- FUSSELL, S. R. et KRAUSS, R. M. (1992). Coordination of knowledge in communication: Effects of speakers’ assumptions about what others know. *Journal of Personality and Social Psychology*, 62(3):378–391.
- GABOR, D. (1947). New possibilities in speech transmission. *Journal of the Institute of Electrical Engineering*, 94(3):369–390.
- GARCIA, J. M. et CANNITO, M. P. (1996). Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *Journal of Speech and Hearing Research*, 39(4):750–760.
- GARCIA, J. M., CANNITO, M. P. et DAGENAIS, P. A. (2000). Hand gestures: Perspectives and preliminary implications for adults with acquired dysarthria. *American Journal of Speech-Language Pathology*, 9(2):107–115.
- GARCIA, J. M. et DAGENAIS, P. A. (1998). Dysarthric sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language and Hearing Research*, 41(6):1282–1293.
- GENTIL, M. (1992). Phonetic intelligibility testing in dysarthria for the use of French language clinicians. *Clinical Linguistics and Phonetics*, 6(3):179–189.
- GEORGE, E. L. J. et HOUTGAST, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 120(4):2295–2311.

- GEORGE, E. L. J., ZEKVELD, A. A., KRAMER, S. E., GOVERTS, S. T., FESTEN, J. M. et HOUTGAST, T. (2007). Auditory and non-auditory factors affecting speech reception in noise by older listeners. *Journal of the Acoustical Society of America*, 121(4):2362–2375.
- GIOLAS, T. G. et EPSTEIN, A. (1963). Comparative intelligibility of word lists and continuous discourse. *Journal of Speech and Hearing Research*, 6(4):349–358.
- GNANSIA, D., PRESSNITZER, D., PÉAN, V., MEYER, B. et LORENZI, C. (2010). Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. *Hearing Research*, 265(1-2):46–53.
- GOBERMAN, A. M., RECKER, B. et PARVENN, S. (2010). Performance effect: Does the presence of a microphone influence parkinsonian speech? *Journal of Medical Speech-Language Pathology*, 18(4):40–45.
- GRATALOUP, C., HOEN, M., VEUILLET, E., COLLET, L., PELLEGRINO, F. et MEUNIER, F. (2009). Speech restoration: An interactive process. *Journal of Speech, Language, and Hearing Research*, 52(4):827–838.
- GRICE, H. P. (1957). Meaning. *Philosophical Review*, 66(3):377–388.
- GRICE, H. P. (1975). Logic and conversation. In COLE, P. & Morgan, J., éditeur : *Syntax and Semantics*, volume 3, pages 41–58, New York. New York Academic Press.
- GRICE, H. P. (1978). Further notes on logic and conversation. In COLE, P., éditeur : *Syntax and Semantics*, volume 9, pages 113–127, New York. New York Academic Press.
- HABIB, M. G., WALTZMAN, S. B., TAJUDEEN, B. et SVIRSKY, M. A. (2010). Speech production intelligibility of early implanted pediatric cochlear implant users. *International Journal of Pediatric Otorhinolaryngology*, 74(8):855–859.
- HAGOORT, P., HALD, L., BASTIAANSEN, M. et PETERSSON, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669):438–441.
- HAGOORT, P. et van BERKUM, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):801–811.
- HALEY, K. L. et MARTIN, G. (2010). Production variability and single word intelligibility in aphasia and apraxia of speech. *Journal of Communication Disorders*, 44(1):103–115.
- HALLÉ, P., SEGUI, J., FRAUENFELDER, U. et MEUNIER, C. (1998). The processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception & Performance*, 24(2):592–608.

- HAWLEY, M. E. (1977a). Editor's comments on papers 1 through 5. In HAWLEY, M. E., éditeur : *Speech intelligibility and speaker recognition*, Benchmark Papers in Acoustics, pages 12–17. Halsted Press, New York.
- HAWLEY, M. E. (1977b). Editor's comments on papers 12, 13, and 5. In HAWLEY, M. E., éditeur : *Speech intelligibility and speaker recognition*, Benchmark Papers in Acoustics, pages 120–127. Halsted Press, New York.
- HAWLEY, M. E. (1977c). Introduction. In HAWLEY, M. E., éditeur : *Speech intelligibility and speaker recognition*, Benchmark Papers in Acoustics, pages 1–10. Halsted Press, New York.
- HEERINGA, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance* – Thèse de doctorat. University of Groningen.
- HERMAN, L. M., RICHARDS, D. G. et WOLZ, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, 16(2):129–219.
- HIGGINBOTHAM, D. J., DRAZEK, A. L., KOWARSKY, K., SCALLY, C. et SEGAL, E. (1994). Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication*, 10(3):191–202.
- HORNSBY, B. W. (2004). The Speech Intelligibility Index: What is it and what's it good for? *Hearing Journal*, 57(10):10–17.
- HOSOM, J.-P., KAIN, A. B., MISHRA, T., van SANTEN, J. P., FRIED-OKEN, M. et STAEHELY, J. (2003). Intelligibility of modifications to dysarthric speech. In *Proceedings of the 2003 IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, pages 878–881. Hong Kong.
- HOUSE, A. S., WILLIAMS, C. E., HECKER, M. H. L. et KRYTER, K. D. (1965). Articulation testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, 37(1):158–166.
- HOUTGAST, T. et STEENEKEN, H. J. M. (2002). *Past, present and future of the Speech Transmission Index*. TNO Human Factors, Soesterberg.
- HUNTER, L., PRING, T. et MARTIN, S. (1991). The use of strategies to increase speech intelligibility in cerebral palsy: An experimental evaluation. *British Journal of Disorders of Communication*, 26(2):163–174.
- HUSTAD, K. C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech-Language Pathology*, 15(3):268–277.
- HUSTAD, K. C. (2007a). Contribution of two sources of listener knowledge to intelligibility of speakers with cerebral palsy. *Journal of Speech, Language, and Hearing Research*, 50(5):1228–1240.

- HUSTAD, K. C. (2007b). Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatica et Logopaedica*, 59(6):306–317.
- HUSTAD, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 51(3):562–573.
- HUSTAD, K. C. et BEUKELMAN, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech and Hearing Research*, 44(3):497–510.
- HUSTAD, K. C. et BEUKELMAN, D. R. (2002). Listener comprehension of severely dysarthric speech: Effects of linguistic cues and stimulus cohesion. *Journal of Speech and Hearing Research*, 45(3):545–558.
- HUSTAD, K. C. et CAHILL, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 12(2):198–208.
- HUSTAD, K. C., JONES, T. et DAILEY, S. (2003). Implementing speech supplementation strategies: Effects on intelligibility and speech rate of individuals with chronic severe dysarthria. *Journal of Speech and Hearing Research*, 46(2):462–474.
- ISO/TR 4870 (1991). *Acoustique - Élaboration et étalonnage des tests d'intelligibilité de la parole*. Association Française de Normalisation, Saint-Denis.
- JAKOBSON, R. (1963). *Essais de linguistique générale - Tome I : Les fondations du langage*. Les Éditions de Minuit, Paris.
- JAMET, F. (2003). De la Classification internationale du handicap (CIH) à la Classification internationale du fonctionnement de la santé et du handicap (CIF). *La nouvelle revue de l' AIS*, 22:163–171.
- JANSEN, S., LUTS, H., WAGENER, K. C., FRACHET, B. et WOUTERS, J. (2010). The French digit triplet test: A hearing screening tool for speech intelligibility in noise. *International Journal of Audiology*, 49(5):378–387.
- JONGMANS, P., HILGERS, F. J. M., POLS, L. C. W. et van AS-BROOKS, C. J. (2006). The intelligibility of tracheoesophageal speech, with an emphasis on the voiced-voiceless distinction. *Logopedics Phoniatrics Vocology*, 31(4):172–181.
- JOOS, M. (1950). Description of language design. *The Journal of the Acoustical Society of America*, 22(6):701–707.
- KAIN, A. B., HOSOM, J.-P., NIU, X., van SANTEN, J. P., FRIED-OKEN, M. et STAEHEL, J. (2007). Improving the intelligibility of dysarthric speech. *Speech Communication*, 49(9):743–759.

- KEINTZ, C. K., BUNTON, K. et HOIT, J. D. (2007). Influence of visual information on the intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 16(3):222–234.
- KENT, R. D., WEISMER, G., KENT, J. F. et ROSENBECK, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, 54(4):482–499.
- KERBRAT-ORECCHIONI, C. (1980). *L'énonciation. De la subjectivité dans le discours*. Armand Colin, Paris.
- KLOPFENSTEIN, M. (2009). Interaction between prosody and intelligibility. *International Journal of Speech-Language Pathology*, 11(4):326–331.
- KRYTER, K. D. (1962). Methods for the calculation and use of the Articulation Index. *The Journal of the Acoustical Society of America*, 34(11):1689–1697.
- KUPERBERG, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, Special Issue 2007(1146):23–49.
- KUTAS, M. et HILLYARD, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.
- KUTAS, M., VAN PETTEN, C. et KLUENDER, R. (2006). Psycholinguistics electrified II (1994-2005). In TRAXLER, M. et GERNSBACHER, M. A., éditeurs : *Handbook of Psycholinguistics*, Londres. Elsevier.
- LANE, H. L. et TRANEL, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14(4):677–709.
- LEARY, S. M., GILPIN, P., LOCKLEY, L., RODRIGUEZ, L., JARRETT, L. et STEVENSON, V. L. (2006). Intrathecal baclofen therapy improves functional intelligibility of speech in cerebral palsy. *Clinical Rehabilitation*, 20(3):228–231.
- LEFÈVRE, M. (2010). *Création d'un outil informatique de rééducation cognitive pour les patients atteints de la Maladie d'Alzheimer – Mémoire pour l'obtention du certificat de capacité d'orthophonie*. Faculté de médecine de Toulouse-Rangueil, Toulouse.
- LEVINSON, S. C. (1983). *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.
- LIBERMAN, I. Y., SHANKWEILER, D., FISCHER, F. W. et CARTER, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18(2):201–212.
- LINDBLOM, B. (1990a). Explaining phonetic variation: A sketch of the H&H theory. In HARDCASTLE, W. et MARCHAL, A., éditeurs : *Speech production and speech modelling*, pages 403–439, Dordrecht, Boston et Londres. Kluwer Academic Publishers.

- LINDBLOM, B. (1990b). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication*, 6(4):220–230.
- LISS, J. M. (2007). Perception of dysarthric speech. In *Motor speech disorders: essays for Ray Kent*, pages 187–219, San Diego. Plural publishing.
- LISS, J. M., SPITZER, S., CAVINESS, J. N. et ADLER, C. (2002). Effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America*, 112(6):3022–3030.
- LOHMANDER, A., FRIEDE, H., ELANDER, A., PERSSON, C. et LILJA, J. (2006). Speech development in patients with unilateral cleft lip and palate treated with different delays in closure of the hard palate after early velar repair: a longitudinal perspective. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery*, 40(5):267–274.
- LOMBARD, E. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille et du Larynx*, 37(2):101–109.
- LOTZE, N., TUNE, S., SCHLESEWSKY, M. et BORNKESSEL-SCHLESEWSKY, I. (2011). Meaningful physical changes mediate lexical-semantic integration: Top-down and form-based bottom-up information sources interact in the n400. *Neuropsychologia*, 49(13):3573–3582.
- MAGNEN, C. (2009). *Approche dynamique de la perception de la parole : catégorisation de la substance et de la variabilité phonétique en langue maternelle par les francophones et en langue étrangère par les hispanophones* – Thèse de doctorat. Université de Toulouse II.
- MARSLÉN-WILSON, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244(5417):522–523.
- MARSLÉN-WILSON, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.
- MARSLÉN-WILSON, W. D. (1989). Access and integration: Projecting sound onto meaning. In MARSLÉN-WILSON, W. D., éditeur : *Lexical representation and processing*, pages 3–24, Cambridge. MIT Press.
- MARSLÉN-WILSON, W. D. et WARREN, P. (1994). Levels of perceptual representation and process in lexical access – Words, phonemes and features. *Psychological Review*, 101(4):653–675.
- MARSLÉN-WILSON, W. D. et WELSH, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, 10(1): 29–63.
- MASSARO, D. W. et COHEN, M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2(1):15–35.

- MCCOLL, D. A. (2006). Intelligibility of tracheoesophageal speech in noise. *Journal of Voice*, 20(4):605–615.
- MCGARR, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, 26(3):451–458.
- MCHENRY, M. A. (2011). An exploration of listener variability in intelligibility judgments. *American Journal of Speech-Language Pathology*, 20(2):119–123.
- MEHLER, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20(3):298–305.
- METZ, D., SCHIAVETTI, N. et SITLER, R. (1980). Toward an objective description of the dependent and independent variables associated with intelligibility assessments of hearing impaired adults. In SUBTELNY, J., éditeur : *Speech Assessment and Speech Improvement for the Hearing Impaired*, pages 72–81. A.G. Bell Association for the Deaf, Washington.
- MILLER, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, 17(11):748–762.
- MILLER, G. A. et CHOMSKY, N. (1963). Finitary models of language users. In LUCE, R. D., BUSH, R. R. et GALANTER, E., éditeurs : *Handbook of mathematical psychology*. Wiley, New York.
- MIRALLES, J. L. et CERVERA, T. (1995). Voice intelligibility in patients who have undergone laryngectomies. *Journal of Speech and Hearing Research*, 38(3):564–571.
- MONSEN, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders*, 48(3):286–296.
- MUNRO, M. J. et DERWING, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38(3):289–306.
- NEEL, A. T. (2009). Effects of loud and amplified speech on sentence and word intelligibility in Parkinson disease. *Journal of Speech, Language and Hearing Research*, 52(4):1021–1033.
- NESPOULOUS, J.-L., LECOURS, A. R., LAFOND, D., LEMAY, M. A., PUEL, M., JOANETTE, Y., COT, F. et RASCOL, A. (1992). *Protocole Montréal-Toulouse d'examen linguistique de l'aphasie : MT-86 module standard initial, M1b*. Ortho Édition, Isbergues.
- NEW, B., BRYBAERT, M., VERONIS, J. et PALLIER, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4):661–677.
- NF EN ISO 9921 (2004). *Ergonomie - Évaluation de la communication parlée*. Association Française de Normalisation, Saint-Denis.

- NGUYEN, N. (2001). Rôle de la coarticulation dans la reconnaissance des mots. *L'Année Psychologique*, 101(1):125–154.
- NGUYEN, N. (2005). La perception de la parole. *In Phonologie et phonétique : forme et substance*, pages 425–447. Hermès, Paris.
- NGUYEN, N., WAUQUIER, S. et TULLER, B. (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. *In PELLEGRINO, F., MARSICO, E., CHITORAN, I. et COUPÉ, C., éditeurs : Approaches to Phonological Complexity*. Mouton de Gruyter, Berlin.
- NICOLOSI, L., HARRYMAN, E. et KRESHECK, J. (1996). *Terminology of Communication Disorders*. Williams and Wilkins, Baltimore.
- NIELSEN, J. B. et DAU, T. (2009). Development of a Danish speech intelligibility test. *International Journal of Audiology*, 48(10):729–741.
- NIEUWLAND, M. S. et VAN BERKUM, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3):691–701.
- NIEUWLAND, M. S. et VAN BERKUM, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7):1098–1111.
- NISHIMOTO, T., MIYAWAKI, K., UEDA, T. et UNE, Y. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods*, 37(3):398–416.
- NORRIS, D. et CUTLER, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, 43(6):541–550.
- NORRIS, D., CUTLER, A., MCQUEEN, J. M. et BUTTERFIELD, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, 53(2):146–193.
- NORRIS, D., MCQUEEN, J. M. et CUTLER, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(5):1209–1228.
- OSTERHOUT, L. et HOLCOMB, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6):785–806.
- OZIMEK, E., KUTZNER, D., SEK, A. et WICHER, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal of Audiology*, 48(7):443–443.
- OZIMEK, E., WARZYBOK, A. et KUTZNER, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology*, 49(6):444–454.

- ÖZSANCAK, C. (2001). L'intelligibilité. In AUZOU, P., ÖZSANCAK, C. et BRUN, V., éditeurs : *Les dysarthries*, pages 50–60. Masson, Paris.
- PARADIS, M. (1989). *Bilingual Aphasia Test (French version)*. Lawrence Erlbaum, Mahwah.
- PECKELS, J.-P. et ROSSI, M. (1973). Le test de diagnostic par paires minimales, adaptation au français du « Diagnostic Rhyme Test » de Voiers. *Revue d'Acoustique*, 27:245–262.
- PENG, J., BEI, C. et SUN, H. (2011). Relationship between Chinese speech intelligibility and Speech Transmission Index in rooms based on auralization. *Speech Communication*, 53(7):986–990.
- PENG, S.-C., SPENCER, L. J. et TOMBLIN, J. B. (2004). Speech intelligibility of pediatric cochlear implant recipients with 7 years of device experience. *Journal of Speech and Hearing Research*, 47(6):1227–1236.
- PENNER, R. (2002). *Robert Penner's Programming Macromedia Flash MX*. Osborne, Blacklick.
- PEUTZ, V. M. A. (1971). Articulation loss of consonants as a criterion for speech transmission in a room. *Journal of Audio Engineering Society*, 19(11):23–27.
- PINTO, S., GHIO, A., TESTON, A. et VIALLET, F. (2010). La dysarthrie au cours de la Maladie de Parkinson. Histoire naturelle de ses composantes : dysphonie, dysprosodie et dysarthrie. *Revue Neurologique*, 166(10):800–810.
- PUYUELO, M. et RONDAL, J. A. (2005). Speech rehabilitation in 10 Spanish-speaking children with severe cerebral palsy: A 4-year longitudinal study. *Pediatric Rehabilitation*, 8(2):113–116.
- RALSTON, J., PISONI, D., LIVELY, S., GREENE, B. et MULLENIX, J. (1991). Comprehension of synthetic speech produced by rule: Word monitoring and sentence-by-sentence listening times. *Human Factors*, 33(4):471–491.
- ROBERTSON, S. J. (1982). *Dysarthria Profile*. Communication Skill Builders, Tucson.
- ROSCH, E., MERVIS, C. B., GRAY, W. D., JOHNSON, D. M. et BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439.
- ROSENZWEIG, M. R. et POSTMAN, L. (1957). Intelligibility as a function of frequency of usage. *Journal of Experimental Psychology*, 54(6):412–422.
- ROSENZWEIG, M. R. et POSTMAN, L. (1958). Frequency of usage and the perception of words. *Science*, 127(3293):263–266.
- ROSSION, B. et POURTOIS, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33(2):217–236.

- ROTA, L. M. et ZELLNER, D. A. (2007). The categorization effect in hedonic contrast: Experts differ from novices. *Psychonomic Bulletin & Review*, 14(1):179–183.
- ROUSSEAUX, M., DUBUFFET, V., TISNE, S., HENGUELLE, C. et LEFEUVRE, M. (2001). Evaluation de l'intelligibilité : présentation du Test Lillois de Dysarthrie (TLD). In AUZOU, P., ÖZSANCAK, C. et BRUN, V., éditeurs : *Les dysarthries*, pages 61–71. Masson, Paris.
- SADEK-KHALIL, D. (1997). De l'intelligibilité. In *Apport de la linguistique à la pédagogie et apport de la pédagogie à la linguistique*, pages 51–69. Éditions du Papyrus, Montreuil.
- SAKAMOTO, S., IWAOKA, N., SUZUKI, Y., AMANO, S. et KONDO, T. (2004). Complementary relationship between familiarity and SNR in word intelligibility test. *Acoustical Science and Technology*, 25(4):290–292.
- SAMAR, V. J. et METZ, D. E. (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research*, 31(3):307–316.
- SAMUEL, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, 32(2):97–127.
- SANFORD, A. et STURT, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9):382–386.
- SARNO, M. T. (1968). Speech impairment in Parkinson's disease. *Archives of Physical Medicine and Rehabilitation*, 49(5):269–275.
- SAUNDERS, G. H. et CIENKOWSKI, K. M. (2002). A test to measure subjective and objective speech intelligibility. *Journal of the American Academy of Audiology*, 13(1):38–49.
- SAVAGE-RUMBAUGH, E. S., MURPHY, J., SEVCIK, R. A., BRAKKE, K. E., WILLIAMS, S. L. et RUMBAUGH, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58(3/4):1–221.
- SCHEGLOFF, E. A. et SACKS, H. (1973). Opening up closings. *Semiotica*, 8(4):289–327.
- SCHIAVETTI, N. (1992). Scaling procedures for the measurement of speech intelligibility. In KENT, R. D., éditeur : *Intelligibility in speech disorders: Theory, measurement and management*, pages 11–34. John Benjamins, Philadelphie.
- SCHIAVETTI, N., METZ, D. E. et SITLER, R. W. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired. *Journal of Speech and Hearing Research*, 24(3):441–445.

- SEARL, J. P., CARPENTER, M. A. et BANTA, C. L. (2001). Intelligibility of stops and fricatives in tracheoesophageal speech. *Journal of Communication Disorders*, 34(4):305–321.
- SEARLE, J. R. (1969). *Speech acts. An essay in the philosophy of language*. Cambridge University Press, Cambridge.
- SHANNON, C. E. et WEAVER, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
- SIVONEN, P., MAESS, B., LATTNER, S. et FRIEDERICI, A. D. (2006). Phonemic restoration in a sentence context: Evidence from early and late ERP effects. *Brain Research*, 1121(1):177–189.
- SKAHAN, S. M., WATSON, M. et LOF, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology*, 16(3):246–259.
- SMITH, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM music workstation. In *Society for Music Perception and Cognition Conference 95*, pages 83–84, Berkeley. University of California.
- SNODGRASS, J. G. et VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2):174–215.
- SOLI, S. D. et WONG, L. L. (2008). Assessment of speech intelligibility in noise with the Hearing in Noise Test. *International Journal of Audiology*, 47(6):356–361.
- STEENEKEN, H. J. M. (2001). The measurement of speech intelligibility. In *Proceedings of the Institute of Acoustics 2001*, volume 23, Stratford-upon-Avon.
- STEENEKEN, H. J. M. et HOUTGAST, T. (2002). Validation of the revised STI_r method. *Speech Communication*, 38(3-4):413–425.
- STEER, M. D. (1945). Speech intelligibility in naval aviation. *Journal of Speech and Hearing Disorders*, 10(3):215–219.
- STEVENS, K. N. (1986). Models of phonetic recognition II: A feature-based model of speech recognition. In *Proceedings of the Montreal Satellite Symposium on Speech Recognition, XIIth International Congress on Acoustics*, pages 66–67.
- STEVENS, S. S. et GALANTER, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 54(6):377–411.
- TANENHAUS, M. K. et BROWN-SCHMIDT, S. (2008). Language processing in the natural world. In MOORE, B. C. M., TYLER, L. K. et MARSLEN-WILSON, W. D., éditeurs : *The perception of speech: From sound to meaning.*, volume 363 de *Philosophical transactions of the royal society B: Biological sciences*, pages 1105–1122.

- TANG, C. et van HEUVEN, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, 119(5):709–732.
- THOMAS, W. G. (1961). *Factors relating to the intelligibility of the speech of deaf children* – Thèse de Master. Washington University.
- TIKOFSKY, R. S. et TIKOFSKY, R. P. (1964). Intelligibility measures of dysarthric speech. *Journal of Speech and Hearing Research*, 7(4):325–333.
- TJADEN, K. et LISS, J. M. (1995). The influence of familiarity on judgments of treated speech. *American Journal of Speech-Language Pathology*, 4(1):39–48.
- TREIMAN, R., CLIFTON, C. J., MEYER, A. S. et WURM, L. H. (2003). Psycholinguistics: Language comprehension and production. In *Comprehensive Handbook of Psychology, Volume 4: Experimental Psychology*, pages 527–548. John Wiley & Sons, New York.
- VAISSIÈRE, J. et MICHAUD, A. (2005). Prosodic constituents in French: A data-driven approach. In FONAGY, I., KAWAGUCHI, Y. et MORIGUCHI, T., éditeurs : *Prosody and Syntax*, pages 47–64, Amsterdam. John Brujamm.
- VAN BERKUM, J. J., ZWITSERLOOD, P., HAGOORT, P. et BROWN, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17(3):701–718.
- VAN BERKUM, J. J. A., van den BRINK, D., TESINK, C. M. J. Y., KOS, M. et HAGOORT, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4):580–591.
- VOIERS, W. D. (1983). Evaluating processed speech using the Diagnostic Rhyme Test. *Speech Technology*, 1(4):30–39.
- WAGENER, K. C. et BRAND, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, 44(3):144–156.
- WAGENER, K. C., BRAND, T. et KOLLMEIER, B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, 45(1):26–33.
- WARREN, P. et MARSLEN-WILSON, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41(3):262–275.
- WARREN, R. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917):392–393.
- WATSON, P. J. et SCHLAUCH, R. S. (2009). Fundamental frequency variation with an electrolarynx improves speech understanding: A case study. *American Journal of Speech-Language Pathology*, 18(2):162–167.

- WEISMER, G. et LAURES, J. S. (2002). Direct magnitude estimates of speech intelligibility in dysarthria: Effects of a chosen standard. *Journal of Speech, Language and Hearing Research*, 45(3):421–433.
- WEISMER, G. et MARTIN, R. E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In KENT, R. D., éditeur : *Intelligibility in speech disorders: Theory, measurement, and management*, pages 67–118, Philadelphie. John Benjamins.
- WHITEHILL, T. L. (2000). Assessing intelligibility in speakers with cleft palate: A critical review of the literature. *The Cleft Palate-Craniofacial Journal*, 39(1):50–58.
- WILLIAMS, S. et WATSON, J. (1985). Differences in speaking proficiencies in three laryngectomy groups. *Archives of Otolaryngology*, 111(4):216–219.
- WILSON, D. et SPERBER, D. (1991). Inference and implicature. In DAVIS, S., éditeur : *Pragmatics: A Reader*, pages 377–393, New York. Oxford University Press.
- WILSON, D. et SPERBER, D. (2005). *Relevance Theory*. Blackwell, Oxford.
- WILSON, E. O. et SPAULDING, T. J. (2010). Effects of noise and speech intelligibility on listener comprehension and processing time of Korean-accented English. *Journal of Speech, Language and Hearing Research*, 53(6):1543–1554.
- WOISARD-BASSOLS, V. (2011). *Impact de l'intelligibilité dans les troubles de la production de la parole pathologique* – Thèse de doctorat. Université de Provence.
- WOOD, P. H. N. (1980). Comment mesurer les conséquences de la maladie. *Chronique OMS*, 34:400–405.
- YORKSTON, K. et BEUKELMAN, D. R. (1996). *Sentence Intelligibility Test for Macintosh*. Communication Disorders Software, Lincoln.
- YORKSTON, K. M. et BEUKELMAN, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, 11(6):499–512.
- YORKSTON, K. M. et BEUKELMAN, D. R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders*, 13(1):15–31.
- YORKSTON, K. M. et BEUKELMAN, D. R. (1981). *Assessment of Intelligibility of Dysarthric Speech*. Pro-Ed, Austin.
- YORKSTON, K. M. et BEUKELMAN, D. R. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In BERRY, W., éditeur : *Clinical dysarthria*, pages 155–163, San Diego. College Hill.

- YORKSTON, K. M., HAMMEN, V. L., BEUKELMAN, D. R. et TRAYNOR, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders*, 55(3):550–560.
- YORKSTON, K. M., STRAND, E. A. et KENNEDY, M. R. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology*, 5(1):55–66.
- YUNUSOVA, Y., WEISMER, G., KENT, R. D. et RUSCHE, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. *Journal of Speech, Language and Hearing Research*, 48(6):1294–1310.
- ZULIANI, P. (1988). *Intelligibilité de la parole dans des conditions d'écoute difficiles* – Thèse de doctorat. Université Toulouse III.

Index des citations

- ANSI S3.5 (1997), 24, 27
Alario et Ferrand (1999), 114–118, 121, 139
Allen (2004), 30
Amano *et al.* (2009), 50, 155, 157
Anderson *et al.* (1991), 75, 77
Anderson et Kalb (1987), 27
Arnon et Snider (2010), 50
Austin (1962), 59, 73, 74
Auzou *et al.* (1998), 38, 39, 88
Barefoot *et al.* (1993), 172
Barton et Sanford (1993), 69
Beijering *et al.* (2008), 44, 225
Bell (1876), 12
Beltyukova *et al.* (2008), 37–40, 42, 51, 57, 58, 88
Benoit *et al.* (1994), 9
Beukelman et Yorkston (1979), 3, 36, 80, 81, 83, 84, 86, 145–147, 168, 173, 175, 229, 233, 242
Beukelman et Yorkston (1980), 42, 45
Beverly *et al.* (2010), 36, 49, 56, 173
Bonin *et al.* (2003), 116–118, 121, 139
Bornkessel-Schlesewsky et Friederici (2007), 68
Bornkessel-Schlesewsky et Schlesewsky (2008), 68
Bornkessel-Schlesewsky et Schlesewsky (2009a), 68
Bornkessel-Schlesewsky et Schlesewsky (2009b), 68
Bougnoux (1993), 29, 31
Brown-Schmidt *et al.* (2005), 75, 76
Burda *et al.* (2003), 87
Burda et Hageman (2005), 87
Burton-Roberts (2007), 71
CALLIOPE (1989), 16, 21
Campbell (1910), 13–16
Carlson *et al.* (1992), 89, 221
Cartier et Rossi (1975), 22
Cheng (1993), 44
Cherry (1953), 155
Christophe *et al.* (1991), 64
Cienkowski et Speaks (2000), 9
Clark (1996), 75, 140
Cox et McDaniel (1989), 9
Crochemore et Vannier (2001), 39
Culicover et Jackendoff (2006), 67
Cullinan *et al.* (1986), 47
Cummings (2005), 71–73, 75
Cutler *et al.* (1986), 63
Cycowicz *et al.* (1997), 114
DePaul et Kent (2000), 36, 45–47
Decker *et al.* (1958), 168
De Bodt *et al.* (2002), 9
De Renzi et Vignolo (1962), 91, 92
de Saussure (1995), 28
Dimitropoulo *et al.* (2009), 139
Doyle *et al.* (1988), 9
Doyle *et al.* (1989), 47
Doyle et Leeper (1997), 9
Duez (1982), 141
Duez (1987), 141
Duez (1992), 141
Egan (1948), 18, 19
Ellis et Beltyukova (2008), 37, 40, 45–47, 51, 57, 141
Enderby (1983), 38
Erickson et Mattson (1981), 69
Ertmer (2010), 1, 36–40, 45, 56, 205, 236
Faiget et Ruiz (1999), 26
Faiget (1997), 18, 25, 26
Fairbanks (1958), 20, 21
Ferreira *et al.* (2002), 69

- Fletcher et Steinberg (1929), 13, 15–17, 23
Flipsen (1995), 45–47
Folstein *et al.* (1975), 180, 243
Frauenfelder et Nguyen (2000), 62
French et Steinberg (1947), 23
Fussell et Krauss (1992), 125
Gabor (1947), 14
Garcia *et al.* (2000), 37
Garcia et Cannito (1996), 38, 47–49, 243
Garcia et Dagenais (1998), 9, 37, 48, 243
Gentil (1992), 39
George *et al.* (2007), 155
George et Houtgast (2006), 155
Giolas et Epstein (1963), 38
Gnansia *et al.* (2010), 13
Goberman *et al.* (2010), 48
Grataloup *et al.* (2009), 65
Grice (1957), 59
Grice (1975), 72
Grice (1978), 72
Habib *et al.* (2010), 13
Hagoort *et al.* (2004), 69
Hagoort et van Berkum (2007), 67, 68, 71, 78, 83
Haley et Martin (2010), 13
Hallé *et al.* (1998), 66
Hawley (1977a), 9, 14, 15
Hawley (1977b), 23
Hawley (1977c), 2, 12
Heeringa (2004), 44, 225
Herman *et al.* (1984), 92
Higginbotham *et al.* (1994), 89
Hornsby (2004), 23, 24
Hosom *et al.* (2003), 36, 37, 39
House *et al.* (1965), 20, 21
Houtgast et Steeneken (2002), 24
Hunter *et al.* (1991), 48
Hustad *et al.* (2003), 37, 45–48, 243
Hustad et Beukelman (2001), 48, 173, 243
Hustad et Beukelman (2002), 89, 221, 243
Hustad et Cahill (2003), 48
Hustad (2006), 36, 39, 44, 50, 236
Hustad (2007a), 37, 48, 49
Hustad (2007b), 44, 49, 235, 236
Hustad (2008), 1, 3, 36, 38, 44, 53, 58, 81–83, 86, 89, 90, 145–147, 158, 159, 168, 171, 173, 229, 233, 236, 242
ISO/TR 4870 (1991), 19
Jakobson (1963), 13, 28, 31, 33, 237–239
Jamet (2003), 237
Jansen *et al.* (2010), 149, 155
Jongmans *et al.* (2006), 9
Joos (1950), 31
Kain *et al.* (2007), 37
Keintz *et al.* (2007), 36, 44, 47, 48, 56, 159, 235
Kent *et al.* (1989), 2, 10, 14, 36, 39, 236
Kerbrat-Orecchioni (1980), 236, 237
Klopfenstein (2009), 36
Kryter (1962), 24
Kuperberg (2007), 68, 71, 78, 83
Kutas *et al.* (2006), 68
Kutas et Hillyard (1980), 68
Lane et Tranel (1971), 48
Leary *et al.* (2006), 37, 236
Lefèvre (2010), 204
Levinson (1983), 71–74
Lieberman *et al.* (1974), 62, 63
Lindblom (1990a), 141
Lindblom (1990b), 172, 240
Liss *et al.* (2002), 39
Liss (2007), 243
Lohmander *et al.* (2006), 37, 236
Lombard (1911), 48
Lotze *et al.* (2011), 71, 78, 83
Magnen (2009), 66
Marslen-Wilson et Warren (1994), 64
Marslen-Wilson et Welsh (1978), 60
Marslen-Wilson (1973), 59
Marslen-Wilson (1975), 59
Marslen-Wilson (1989), 65
Massaro et Cohen (1983), 66
McColl (2006), 56, 57
McGarr (1983), 45–47

- McHenry (2011), 57, 88
Mehler (1981), 63
Metz *et al.* (1980), 36, 40, 205
Miller et Chomsky (1963), 59
Miller (1962), 59
Miralles et Cervera (1995), 9
Monsen (1983), 45–47
Munro et Derwing (1995), 87
NF EN ISO 9921 (2004), 24, 25, 27, 44, 56
Neel (2009), 37, 56, 57, 236
Nespoulous *et al.* (1992), 90, 91
New *et al.* (2007), 169
Nguyen *et al.* (2009), 65
Nguyen (2001), 59
Nguyen (2005), 62, 65, 66
Nicolosi *et al.* (1996), 56
Nielsen et Dau (2009), 149, 155
Nieuwland et Van Berkum (2005), 69
Nieuwland et Van Berkum (2006), 70
Nishimoto *et al.* (2005), 139
Norris *et al.* (1995), 66
Norris *et al.* (2006), 61
Norris et Cutler (1988), 63
Osterhout et Holcomb (1992), 68
Ozimek *et al.* (2009), 149, 155
Ozimek *et al.* (2010), 155
Özsancak (2001), 9, 10, 36, 40, 42, 44, 236
Paradis (1989), 90, 91
Peckels et Rossi (1973), 21
Peng *et al.* (2004), 57
Peng *et al.* (2011), 13
Penner (2002), 100, 101
Peutz (1971), 25
Pinto *et al.* (2010), 37
Puyuelo et Rondal (2005), 37, 236
Ralston *et al.* (1991), 88, 89, 221
Robertson (1982), 39
Rosch *et al.* (1976), 122
Rosenzweig et Postman (1957), 50
Rosenzweig et Postman (1958), 50
Rossion et Pourtois (2004), 115–118, 139
Rota et Zellner (2007), 125
Rousseaux *et al.* (2001), 88
Sadek-Khalil (1997), 10
Sakamoto *et al.* (2004), 50
Samar et Metz (1988), 37, 40–42, 205
Samuel (1997), 66
Sanford et Sturt (2002), 69
Sarno (1968), 80
Saunders et Cienkowski (2002), 9
Savage-Rumbaugh *et al.* (1993), 3, 93–96, 106, 107
Schegloff et Sacks (1973), 59
Schiavetti *et al.* (1981), 40, 41
Schiavetti (1992), 2, 14, 36, 40–42, 205
Searl *et al.* (2001), 9
Searle (1969), 59
Shannon et Weaver (1949), 13, 28–30, 33, 52
Sivonen *et al.* (2006), 66
Skahan *et al.* (2007), 1, 36, 56, 236
Smith (1995), 153
Snodgrass et Vanderwart (1980), 113–115, 117–119, 139
Soli et Wong (2008), 44, 45
Steeneken et Houtgast (2002), 25
Steeneken (2001), 9
Steer (1945), 2
Stevens et Galanter (1957), 40
Stevens (1986), 64
Tanenhaus et Brown-Schmidt (2008), 59, 60, 71, 76
Tang et van Heuven (2009), 44
Thomas (1961), 45–47
Tikofsky et Tikofsky (1964), 33, 36, 38
Tjaden et Liss (1995), 45–47
Treiman *et al.* (2003), 59, 66
Vaissière et Michaud (2005), 141
Van Berkum *et al.* (2003), 69
Van Berkum *et al.* (2008), 70, 71, 78
Voiers (1983), 21
Wagener *et al.* (2006), 149
Wagener et Brand (2005), 149, 155
Warren et Marslen-Wilson (1987), 64
Warren (1970), 66
Watson et Schlauch (2009), 56, 57, 159
Weismer et Laures (2002), 40, 42
Weismer et Martin (1992), 2, 14, 52, 81

- Whitehill (2000), 40, 41
Williams et Watson (1985), 45
Wilson et Spaulding (2010), 87, 89
Wilson et Sperber (1991), 74, 75
Wilson et Sperber (2005), 71
Woisard-Bassols (2011), 44, 51, 237
Wood (1980), 235, 238
Yorkston *et al.* (1990), 36, 52
Yorkston *et al.* (1996), 172, 173
Yorkston et Beukelman (1978), 39, 42,
50
Yorkston et Beukelman (1980), 50
Yorkston et Beukelman (1981), 38, 49
Yorkston et Beukelman (1983), 47
Yorkston et Beukelman (1996), 82, 145
Yunusova *et al.* (2005), 9
Zuliani (1988), 18, 19

Liste des figures

CHAPITRE 1 – *Développement des tests d’intelligibilité de la parole dans les domaines de l’acoustique et des télécommunications* **11**

1.1	Transmission d’un message de Salem à Boston par A. G. Bell, en mars 1877 (gravure)	12
1.2	Feuille de passation du Test d’Articulation de Fletcher et Steinberg (1929)	17
1.3	Le circuit de la parole, d’après F. de Saussure (1995, 1916 pour l’édition originale)	28
1.4	Schéma de la communication de Shannon et Weaver (1949)	29
1.5	Les fonctions du langage selon Jakobson (1963)	31

CHAPITRE 2 – *Méthodes d’évaluation et de mesure de l’intelligibilité pour les troubles pathologiques de production de la parole (TPPP)* **35**

Pas de figure dans ce chapitre.

CHAPITRE 3 – *Problématique : intelligibilité et compréhension de la parole en situation de communication* **55**

3.1	Erreurs commises lors de tâches de segmentation en phonèmes et syllabes par des enfants (Liberman <i>et al.</i> , 1974)	63
3.2	Mécanismes et niveaux impliqués dans la perception et la compréhension de la parole : perspective ascendante (<i>bottom-up</i>) et fragmentée (<i>two-step</i>)	67
3.3	Exemple de situation de test pragmatique pour l’étude de la compréhension de la parole (Brown-Schmidt <i>et al.</i> , 2005)	76
3.4	Exemple de matériel utilisé pour la constitution d’un corpus de conversation (Anderson <i>et al.</i> , 1991)	77

CHAPITRE 4 – Choix préliminaires et développement du test EloKanz 85

4.1	Exemples de stimuli utilisés dans le <i>Bilingual Aphasia Test</i> (Paradis, 1989)	91
4.2	Exemples de stimuli utilisés pour un test de type <i>Token Test</i>	92
4.3	Évaluation de la compréhension de la parole par le bonobo Kanzi (Savage-Rumbaugh <i>et al.</i> , 1993)	94
4.4	EloKanz – Matériel prévu pour le locuteur et l’auditeur	97
4.5	EloKanz – Fenêtre de saisie du code locuteur et du nombre d’animations à jouer	99
4.6	EloKanz – Interface prévue pour le locuteur	99
4.7	EloKanz – Fonctions de déplacement des images dans les animations (tirées de Penner, 2002)	101
4.8	EloKanz – Animation de la main (1)	102
4.9	EloKanz – Animation de la main (2)	102
4.10	EloKanz – Animation de la main (3)	102
4.11	EloKanz – Interface prévue pour l’auditeur	103
4.12	EloKanz – Exemple de cadre signalant à l’auditeur un emplacement cible	104
4.13	EloKanz – Historique des actions effectuées par l’auditeur	106
4.14	EloKanz – Exemple de fichier résultat (première partie)	106
4.15	EloKanz – Exemple de fichier résultat (seconde partie)	107

CHAPITRE 5 – Études préliminaires 111

5.1	Exemples d’images normées par Snodgrass et Vanderwart (1980)	114
5.2	Exemples d’images normées par Rossion et Pourtois (2004)	116
5.3	Exemples d’images normées par Bonin <i>et al.</i> (2003)	117
5.4	Exemples d’images du Dictionnaire Le Visuel 4 [©]	119
5.5	Types de désaccord sur le nom des images pré-sélectionnées pour le test EloKanz	122
5.6	Interface utilisée pour la pré-étude d’accord sur le nom en situation de communication	126

CHAPITRE 6 – Intelligibilité et compréhension de la parole dans le bruit 143

6.1	Fenêtre de saisie des informations sur l’auditeur	148
6.2	Spectrogrammes de signaux de parole avant et après atténuation	150
6.3	Procédure d’égalisation du niveau des stimuli audio en fonction de leur niveau sonore	151
6.4	Interface utilisée pour le pré-test d’égalisation du niveau des stimuli sur la base d’évaluations subjectives	153

6.5	Création du fichier de bruit (brouhaha)	156
6.6	Déroulement chronologique de la diffusion de la parole et du bruit pour chaque stimulus	157
6.7	Interface de l'auditeur	159
6.8	Évolution des scores d'intelligibilité et de compréhension en fonction du niveau de bruit	162
6.9	Évolution des scores d'intelligibilité et de compréhension en fonction du niveau de bruit, avec écart-type	164
6.10	Scores d'intelligibilité et de compréhension en fonction du niveau de bruit, avec les droites de régression	165
6.11	Moyenne de la fréquence des noms (par million d'occurrences), en rapport avec les tâches de compréhension et d'intelligibilité	169
6.12	Contribution de l'information acoustique et de l'information indépendante du signal pour la compréhensibilité du locuteur dysarthrique (adapté de Yorkston <i>et al.</i> , 1996; Lindblom, 1990b)	172
CHAPITRE 7 – Études de faisabilité clinique et de validation du test EloKanz		177
<hr/>		
7.1	Durée totale de la passation pour chaque patient	186
7.2	Évaluation de la difficulté d'utilisation du logiciel par les patients	188
7.3	Durée de la passation selon l'origine des TPPP	189
7.4	Durée nécessaire à l'enregistrement d'une phrase selon l'origine des TPPP, et en comparaison de sujets sains	190
7.5	Nombre d'éléments de non respect des consignes survenus pendant la phase d'entraînement, selon l'origine des TPPP	191
7.6	Nombre d'éléments de non respect des consignes survenus pendant la phase d'enregistrement, selon l'origine des TPPP	192
7.7	Nombre d'éléments de non respect des consignes de type « utilisation de l'interface » survenus pendant la phase d'enregistrement, selon l'origine des TPPP	193
7.8	Nombre d'éléments de non respect des consignes de type « formulation » survenus pendant la phase d'enregistrement, selon l'origine des TPPP	194
7.9	Temps de réaction moyens avant l'enregistrement d'une phrase, selon l'origine des TPPP et en comparaison de sujets sains	195
7.10	Évaluation de la difficulté du logiciel par les patients, selon l'origine des TPPP	196
7.11	Prédiction de la durée nécessaire à la phase d'enregistrement	202
<i>Étude de la faisabilité du test EloKanz pour les auditeurs</i>		
7.12	Boitier de réponse utilisé pour le pré-test d'évaluation subjective de la compréhensibilité	208
7.13	Consignes diffusées pour le pré-test d'évaluation de la compréhensibilité	209

7.14	Écran diffusé avec chaque phrase dans le pré-test d'évaluation de la compréhension	209
7.15	Médianes des scores de compréhension obtenus par une évaluation subjective, pour chaque locuteur	210
7.16	Évaluation de la difficulté d'utilisation du logiciel par les auditeurs	216

Éléments de validation du test EloKanz

7.17	Pourcentage de phrases comprises par les auditeurs, pour chaque locuteur	220
7.18	Corrélations inter auditeurs, par niveau de significativité	222
7.19	Coefficients de corrélation de Pearson obtenus pour les temps de réaction de 253 couples d'auditeurs	223
7.20	Pourcentages de phrases comprises chez les auditeurs exposés aux TPPP ou non exposés aux TPPP	224
7.21	Temps de réaction moyens (en ms) avant l'exécution de la première action pour les auditeurs exposés aux TPPP <i>vs</i> non exposés	224
7.22	Temps de réaction moyens des auditeurs en fonction des scores moyens d'évaluation subjective de la compréhension	227
7.23	Différences significatives de temps de réaction chez les auditeurs, selon les locuteurs	230
7.24	Médianes des scores de compréhension pour chaque locuteur, avec les différences significatives	230
7.25	Modèle de Wood (1980) pour la classification internationale du handicap (CIH)	235
7.26	Schéma de la communication de Kerbrat-Orecchioni	237

Liste des tableaux

CHAPITRE 1 – <i>Développement des tests d’intelligibilité de la parole dans les domaines de l’acoustique et des télécommunications</i>	11
--	-----------

1.1 Exemples de stimuli du test de Campbell (1910)	15
1.2 Liste de mots de Zuliani (1988)	18
1.3 Durées typiques d’apprentissage pour des tests aux mots <i>vs</i> aux logotomes (ISO/TR 4870, 1991)	19
1.4 Exemples d’items dans les 5 listes du Test de Rime de Fairbanks (1958)	20
1.5 Exemples d’items du test de diagnostic par paires minimales	22
1.6 Relations entre différents indices d’intelligibilité (NF EN ISO 9921, 2004)	27

CHAPITRE 2 – <i>Méthodes d’évaluation et de mesure de l’intelligibilité pour les troubles pathologiques de production de la parole (TPPP)</i>	35
---	-----------

2.1 Échelle du <i>NTID Read Intelligibility Test</i>	41
2.2 Exemples d’études montrant un effet de l’expertise des auditeurs sur les scores d’intelligibilité	46

CHAPITRE 3 – <i>Problématique : intelligibilité et compréhension de la parole en situation de communication</i>	55
---	-----------

Pas de tableau dans ce chapitre.

CHAPITRE 4 – <i>Choix préliminaires et développement du test EloKanz</i>	85
--	-----------

4.1 Contraintes définies pour notre test de compréhension de la parole . .	87
4.2 Récapitulatif des différentes méthodes d’évaluation de la compréhension orale, en fonction de nos critères de sélection	95

 CHAPITRE 5 – *Études préliminaires* 111

5.1	Noms attribués aux images lors de la tâche de dénomination écrite et lors de la tâche de communication (exemples)	127
5.2	Difficulté ressentie par rapport à l’interface <i>locuteur</i> d’EloKanz (échelle de 1 – <i>Très facile</i> à 5 – <i>Très difficile</i>)	134
5.3	Lourdeur (fatigue) ressentie par rapport à l’utilisation de l’interface <i>locuteur</i> d’EloKanz (échelle de 1 – <i>Pas du tout lourd/fatigant</i> à 5 – <i>Très lourd/fatigant</i>)	135
5.4	Images posant des problèmes aux sujets dans l’interface <i>locuteur</i> d’EloKanz	135
5.5	Moyennes des scores pour les échelles de difficulté et de lourdeur de la tâche, pour les sujets porteurs et non porteurs de corrections oculaires	136
5.6	Différences d’évaluation de la lourdeur de la tâche entre les sujets porteurs de corrections oculaires et non porteurs de telles corrections (test de Mann-Whitney)	137
5.7	Temps de réaction des locuteurs (en ms) pour l’enregistrement d’une phrase	138

 CHAPITRE 6 – *Intelligibilité et compréhension de la parole dans le bruit* 143

6.1	Corrélations de Pearson entre les gains attribués par chaque couple d’auditeurs lors du pré-test d’égalisation en fonction du niveau sonore perçu	154
6.2	Attribution du niveau de bruit en fonction des phrases, pour un groupe de 10 participants	158
6.3	Différences de moyenne entre les scores d’intelligibilité et de compréhension	161
6.4	Écarts-types des scores (pourcentages) d’intelligibilité et de compréhension en fonction du niveau de bruit	163
6.5	Coefficients de corrélation entre les scores d’intelligibilité et de compréhension en fonction du niveau de bruit	166
6.6	Accords intra et inter auditeurs pour les scores d’intelligibilité et de compréhension	167
6.7	Scores d’intelligibilité et de compréhension : statistiques descriptives .	171

 CHAPITRE 7 – *Études de faisabilité clinique et de validation du test EloKanz* 177

7.1	Durées moyennes nécessaires à l’enregistrement d’une phrase, selon l’origine du TPPP et en comparaison de sujets sains	191
7.2	Nombre d’éléments de non respect des consignes selon l’origine des TPPP	192

7.3	Temps de réaction moyens lors de l'enregistrement d'une phrase, selon l'origine des TPPP et en comparaison de sujets sains	195
7.4	Corrélations partielles tenant compte de la fréquence d'apparition du dysfonctionnement des enregistrements audio	197
7.5	Nombre d'éléments de non respect des consignes en fonction de l'âge et des scores des patients au MMSE (corrélations)	198
7.6	Temps de réaction moyen lors de l'enregistrement d'une phrase en fonction de l'âge et des scores des patients au MMSE (corrélations) .	199
7.7	Évaluations subjectives de la difficulté du logiciel en fonction de l'âge et des scores des patients au MMSE (corrélations)	199
7.8	Évaluation de la difficulté du logiciel par les patients en fonction de l'âge et des scores obtenus par les patients aux MMSE (corrélations) .	200

Étude de la faisabilité du test EloKanz pour les auditeurs

7.9	Participants au pré-test d'évaluation subjective de la compréhensibilité des patients	207
7.10	Accords inter juges lors de l'évaluation de la compréhensibilité des patients sur une échelle (corrélations)	211
7.11	Nombre de problèmes liés aux consignes et à l'ergonomie en fonction de l'âge des auditeurs (corrélations)	216
7.12	Durées des phases d'entraînement et d'écoute en fonction de l'âge des auditeurs (corrélations)	216
7.13	Évaluation de la difficulté du logiciel par les auditeurs en fonction de leur âge (corrélations)	217

Éléments de validation du test EloKanz

7.14	Variabilité des temps de réaction des auditeurs expliquée par l'âge et les années d'exposition aux TPPP (régression linéaire)	225
7.15	Variabilité des temps de réaction des auditeurs expliquée par les variables liées aux auditeurs et aux stimuli (régression linéaire)	226
7.16	Temps de réaction des auditeurs en fonction des scores d'évaluations de la compréhensibilité sur une échelle (corrélation)	227

Annexes

Annexe A

Première expérience d'accord sur le nom

A.1 Sujets testés

Sujet	Sexe	Âge	Profession	Domaine	Région	Milieu
1	F	28	Doctorante	Sciences du langage / Aéronautique	Midi-Pyrénées	Urbain
2	F	29	Post-doctorante	Psychologie (autisme)	Rhône-Alpes	Urbain
3	F	27	Doctorante	Psychologie (espace)	Nord-Pas-De-Calais	Rural
4	M	78	Retraité	Agriculture	Midi-Pyrénées	Rural
5	F	29	Doctorante	Sciences du langage (morphologie)	Languedoc-Roussillon	Rural
6	F	39	Employée de maison	N/P	Midi-Pyrénées	Rural
7	M	31	Saisonnier	Ski/commerces	Midi-Pyrénées	Rural
8	M	24	Ingénieur	Général	Midi-Pyrénées	Urbain
9	M	61	Retraité Educ. Nat.	Mathématiques	Midi-Pyrénées	Rural
10	F	59	Aide-soignante	Médical	Midi-Pyrénées	Urbain
11	F	48	Secrétaire	Médical	Midi-Pyrénées	Urbain
12	F	28	Enseignante-chercheuse	Sciences du langage	Centre	Rural

Sujets testés (suite)

Sujet	Sexe	Âge	Profession	Domaine	Région	Milieu
13	M	18	Apprenti maçon	Bâtiment	Midi-Pyrénées	Rural
14	F	19	Agent de service hospitalier	Médical	Midi-Pyrénées	Rural
15	F	30	Saisonnnière	Ski/hôtellerie	Midi-Pyrénées	Rural
16	M	50	Entrepreneur	Maçonnerie	Midi-Pyrénées	Rural
17	F	21	Étudiante	Économie	Haute-Normandie	Urbain
18	F	43	Vendeuse	Agriculture	Midi-Pyrénées	Rural
			Employée de maison	N/P		
19	F	62	Retraitée	Journalisme	Midi-Pyrénées	Rural
20	M	22	Étudiant	Économie	Languedoc-Roussillon	Urbain
21	F	31	Télé-enquêtrice	Beaux-arts	Midi-Pyrénées	Urbain
22	F	33	Enseignante	Français Langue Étrangère	Midi-Pyrénées	Urbain
23	F	83	Retraitée	Philosophie, religion	Midi-Pyrénées	Rural
24	M	61	Radiologue	Médical	Midi-Pyrénées	Rural
25	M	32	Technicien	Robinetterie industrielle	Midi-Pyrénées	Rural
26	F	31	Lieutenant	Police judiciaire	Île-de-France	Urbain

A.2 Interface utilisée

A.2.1 Présentation de la consigne

Merci pour votre participation à ce test. Avant de commencer, lisez attentivement les instructions suivantes :

1. Le test commencera par la saisie de vos informations personnelles.
2. Après la saisie de vos informations, vous arriverez sur une page contenant une image et un champ "nom", comme ci-dessous :



Nom :

3. Entrez le nom qui, selon vous, correspond le plus à l'image, puis validez en cliquant sur le bouton. Une autre image sera alors affichée, pour laquelle vous devrez également spécifier un nom, et ainsi de suite jusqu'à la fin du test.

A.2.2 Saisie des informations personnelles

Avant de commencer le test, veuillez s'il vous plaît remplir attentivement les informations ci-dessous:

Votre âge :

Votre profession :

Email ou numéro de téléphone¹:

Votre région²:

Vous vivez en milieu : Urbain Rural Autre

¹Cette information sera effacée dès le traitement des données du test
²Région dans laquelle vous avez vécu le plus longtemps. A défaut, choisissez "Aucune..."

A.2.3 Saisie des dénominations

image 1 sur 213



Nom :

A.3 Noms les plus fréquemment attribués pour chaque image

N°	Nom attendu	Nom le plus fréquent	Accord	N°	Nom attendu	Nom le plus fréquent	Accord
1	autruche	autruche	100	107	dauphin	dauphin	89
2	balance	balance	100	108	feuille	feuille	89
3	banane	banane	100	109	fléchette	fléchette	89
4	bonbon	bonbon	100	110	fromage	fromage	89
5	bonnet	bonnet	100	111	guitare	guitare	89
6	canne	canne	100	112	harpe	harpe	89
7	carotte	carotte	100	113	melon	melon	89
8	cerises	cerises	100	114	moto	moto	89
9	cigare	cigare	100	115	ours	ours	89
10	crabe	crabe	100	116	piment	piment	89
11	fourchette	fourchette	100	117	slip	slip	89
12	fraise	fraise	100	118	vélo	vélo	89
13	gant	gant	100	119	boussole	boussole	89
14	girafe	girafe	100	120	avion	avion	85
15	gomme	gomme	100	121	béquille	béquille	85
16	hache	hache	100	122	boxeur	boxeur	85
17	igloo	igloo	100	123	cartouches	cartouches	85
18	lion	lion	100	124	château	château	85
19	loupe	loupe	100	125	clou	clou	85
20	marteau	marteau	100	126	épée	épée	85
21	oeuf	oeuf	100	127	fauteuil	fauteuil	85
22	oignon	oignon	100	128	fenêtre	fenêtre	85
23	oreille	oreille	100	129	framboises	framboises	85
24	pastèque	pastèque	100	130	haltères	haltères	85
25	peigne	peigne	100	131	jumelles	jumelles	85
26	pelle	pelle	100	132	lustre	lustre	85
27	pinceau	pinceau	100	133	micro	micro	85
28	pipe	pipe	100	134	piscine	piscine	85
29	poireaux	poireaux	100	135	poêle	poêle	85
30	poulet	poulet	100	136	poubelle	poubelle	85
31	râteau	râteau	100	137	quille	quille	85
32	seau	seau	100	138	salade	salade	85
33	skieur	skieur	100	139	scorpion	scorpion	85
34	tomate	tomate	100	140	taupe	taupe	85
35	tortue	tortue	100	141	tente	tente	85
36	vache	vache	100	142	voiture	voiture	85
37	zèbre	zèbre	100	143	aimant	aimant	81

...le tableau continue à la page suivante

ANNEXE A. PREMIÈRE EXPÉRIENCE D'ACCORD SUR LE NOM

N°	Nom attendu	Nom le plus fréquent	Accord	N°	Nom attendu	Nom le plus fréquent	Accord
38	croissant	croissant	100	144	ampoule	ampoule	81
39	poussin	poussin	100	145	asperges	asperges	81
40	arbre	arbre	96	146	brique	brique	81
41	boutons	boutons	96	147	carton	carton	81
42	brosse	brosse	96	148	étoile	étoile	81
43	casquette	casquette	96	149	grenouille	grenouille	81
44	chat	chat	96	150	mouton	mouton	81
45	chenille	chenille	96	151	porte	porte	81
46	cheval	cheval	96	152	rasoir	rasoir	81
47	chien	chien	96	153	tondeuse	tondeuse	81
48	citron	citron	96	154	triangle	triangle	81
49	coq	coq	96	155	tampon	tampon	80
50	couteau	couteau	96	156	bouteille	bouteille	77
51	échelle	échelle	96	157	casque	casque	77
52	gâteau	gâteau	96	158	cassette	cassette	77
53	lit	lit	96	159	castor	castor	77
54	louche	louche	96	160	fouet	fouet	77
55	lunettes	lunettes	96	161	orange	orange	77
56	moules	moules	96	162	piscine	piscine	77
57	phare	phare	96	163	poivron	poivron	77
58	pigeon	pigeon	96	164	scooter	scooter	77
59	plume	plume	96	165	souris	souris	77
60	poire	poire	96	166	trompette	trompette	77
61	radis	radis	96	167	volcan	volcan	77
62	renard	renard	96	168	balai	balai	73
63	requin	requin	96	169	camion	camion	73
64	ressort	ressort	96	170	cuillère	cuillère	73
65	torchon	torchon	96	171	cycliste	cycliste	73
66	trèfle	trèfle	96	172	mouche	mouche	73
67	trombones	trombones	96	173	soleil	soleil	73
68	truille	truille	96	174	canard	canard	69
69	chemise	chemise	96	175	chameau	chameau	69
70	brouette	brouette	92	176	friteuse	friteuse	69
71	canon	canon	92	177	palette	palette	69
72	chapeau	chapeau	92	178	piano	piano	69
73	chaussette	chaussette	92	179	roulette	roulette	69
74	ciseaux	ciseaux	92	180	oiseau	oiseau	68
75	dé	dé	92	181	abeille	guêpe	65
76	drapeau	drapeau	92	182	ancree	ancree	65
77	éponge	éponge	92	183	cagoule	cagoule	65
78	fougère	fougère	92	184	lampe	lampe	65

...le tableau continue à la page suivante

A.3. Noms les plus fréquemment attribués pour chaque image

N°	Nom attendu	Nom le plus fréquent	Accord	N°	Nom attendu	Nom le plus fréquent	Accord
79	fourmi	fourmi	92	185	tasse	tasse	65
80	fusée	fusée	92	186	violon	violon	65
81	glacière	glacière	92	187	brancard	brancard	62
82	gorille	gorille	92	188	carré	carré	62
83	grue	grue	92	189	disque	disque	62
84	huîtres	huîtres	92	190	hachoir	hachoir	62
85	kiwi	kiwi	92	191	moulin	moulin	62
86	maison	maison	92	192	orque	orque	62
87	pain	pain	92	193	poisson	poisson	62
88	palmier	palmier	92	194	saucisse	saucisse	62
89	paon	paon	92	195	cylindre	cylindre	58
90	passoire	passoire	92	196	église	église	58
91	rideaux	rideaux	92	197	fleur	fleur	58
92	robe	robe	92	198	lime	lime	58
93	scie	scie	92	199	plongeur	plongeur	58
94	serpent	serpent	92	200	bottes	bottes	54
95	sucette	sucette	92	201	pneu	roue	52
96	théière	théière	92	202	assiette	assiette	50
97	tigre	tigre	92	203	ballon	ballon	50
98	tracteur	tracteur	92	204	maillet	maillet	50
99	valise	valise	92	205	pendule	horloge	50
100	vis	vis	92	206	coupole	dôme	50
101	lapin	lapin	92	207	batteur	batteur	46
102	ceinture	ceinture	89	208	montagne	montagne	46
103	chaise	chaise	89	209	morse	morse	46
104	chaussure	chaussure	89	210	toucan	toucan	46
105	cible	cible	89	211	cageot	cageot	27
106	cube	cube	89	212	télé	télé	23

A.4 Noms correspondant à d'autres concepts autres que ceux visés par les images

Image	Autres concepts attribués
Abeille	Guêpe, frelon
Aimant	Fer à cheval
Ancre	Grappin, harpon
Assiette	Plat
Balai	Abat-jour, lampe, lustre, ventouse WC, lampe en suspension
Ballon	Dôme
Batteur	Mixeur, fouet électrique, robot mixeur
Boussole	Chronomètre
Brancard	Lit de camp, civière
Brique	Parpaing
Cagoule	Passe-montagne
Carré	Case, tableau
Carton	Boîte
Cartouches	Rouge à lèvres, balles
Cassette	Enregistreur
Castor	Blaireau, ragondin, raton-laveur, loutre
Ceinture	Collier
Chameau	Dromadaire
Château	Forteresse
Chenille	Ver
Coupole	Toit, clocher, chapelle
Cube	Carré
Cylindre	Cône, récipient, abat-jour, cube, tube, verre, cloche
Dauphin	Requin, baleine
Drapeau	Fanion
Église	Cathédrale, abbaye, petit manoir
Framboises	Mûres, myrtilles, groseilles
Friteuse	Cuiseur vapeur, bouilloire, cuiseur à riz, instrument de mesure
Gateau	Bonbonnière
Glacière	Congélateur
Harpe	Cithare, arc
Jumelles	Longue-vue
Kiwi	Fruit de la passion
Lampe	Lampadaire
Lapin	Lièvre
Lustre	Échelle, chandelier
Maillet	Masse, marteau
Maison	Pavillon

...le tableau continue à la page suivante

A.4. Noms correspondant à d'autres concepts autres que ceux visés par les images

Image	Autres concepts attribués
Melon	Courge
Morse	Phoque, éléphant de mer, lion de mer, otarie, ornithorynque
Moto	Mobylette
Mouche	Bourdon, taon
Mouton	Brebis
Orange	Mandarine, clémentine
Orque	Dauphin, baleine, cachalot
Palmier	Cocotier
Pendule	Horloge
Piment	Poivron
Piscine	Cendrier, bouillotte
Piscine (2)	Instrument de pesée
Pneu	Roue
Poivron	Piment
Renard	Loup
Requin	Maquereau
Robe	Blouse
Roulette	Roue, boulier
Saucisse	Boudin, boudin blanc, knacki
Scorpion	Homard, écrevisse, scarabée
Soleil	Lune, ovule, rond avec du poil, cellule, cercle
Taupe	Mulot, fourmilier
Théière	Carafe, cafetière
Tigre	Léopard, panthère
Torchon	Serviette
Trombones	Agraphes
Trompette	Trombone, clarinette, saxophone
Violon	Violoncelle
Vis	Clou

A.5 Noms d'images désignées par des synonymes

Image	Synonymes employés
Béquille	Canne anglaise
Cageot	Cagette, caisse, caissette
Chaussette	Socquette
Chaussure	Soulier
Clou	Pointe
Cycliste	Coureur, vélocycliste
Disque	Vinyle
Haltères	Poids
Lime	Râpe
Micro	Microphone
Oiseau	Piaf
Plongeur	Homme-grenouille, scaphandrier
Rideaux	Tentures
Télé	Télévision, téléviseur, TV
Valise	Bagage
Voiture	Auto

A.6 Images donnant lieu à des noms/concepts super ordonnés

Image	Noms/concepts super ordonnés
Batteur	Robot ménager
Béquille	Canne
Brique	Matériau
Cageot	Boîte, bac en bois
Carton	Boîte
Chaise	Siège
Coupole	Toît
Dauphin	Poisson
Église	Bâtiment
Fauteuil	Siège
Fenêtre	Ouverture
Fléchette	Lancette
Fouet	Batteur, batteur manuel
Fougère	Plante
Fourmi	Insecte
Framboises	Fruits rouges
Glacière	Boîte isotherme

...le tableau continue à la page suivante

Image	Noms/concepts super ordonnés
Gorille	Singe
Hachoir	Couteau, couteau de boucher
Kiwi	Fruit
Louche	Ustensile
Lustre	Luminaire
Melon	Cucurbitacée
Montagne	Relief
Mouche	Insecte
Orque	Mammifère
Pigeon	Oiseau
Ressort	Forme hélicoïdale
Souris	Rongeur
Sucette	Sucrerie
Toucan	Oiseau, oiseau exotique
Valise	Bagage
Voiture	Véhicule

A.7 Images donnant lieu à des noms/concepts subordonnés

Image	Noms/concepts subordonnés
Ampoule	Ampoule à vis, ampoule électrique, ampoule à incandescence
Ancre	Ancre marine
Arbre	Chêne
Asperges	Asperges vertes
Assiette	Assiette creuse, assiette à soupe
Avion	A380, Avion de ligne
Balai	Balai-brosse
Ballon	Ballon de foot, ballon de football
Batteur	Batteur électrique
Bottes	Cuissardes, bottes de pêche, jambières
Bouteille	Bouteille d'eau minérale, Bouteille de Perrier, Perrier
Boxeur	Boxeur de boxe thaïe, boxeur amateur
Brique	Brique de 5
Brosse	Brosse à ongles
Brouette	Brouette 120 litres
Cagoule	Passe-montagne
Camion	Semi, camion-benne
Canard	Colvert, canard colvert
Cartouches	Cartouches de fusil
Casque	Casque de moto, casque intégral

...le tableau continue à la page suivante

Image	Noms/concepts subordonnés
Casquette	Casquette d'hiver
Cassette	Cassette audio
Chapeau	Borsalino
Chaussure	Chaussure en cuir, chaussure de ville
Cheval	Cheval de course
Chien	Dogue allemand
Cible	Cible de tir
Citron	Citron jaune
Dôme	Dôme d'église orthodoxe
Couteau	Couteau de cuisine
Cuillère	Petite cuillère, cuillère à café
Dé	Dé à jouer
Éponge	Spontex, éponge Scotch Brite
Fauteuil	Club, Fauteuil club
Feuille	Feuille d'arbre
Fleur	Lys, iris, hibiscus, orchidée, fleur de lys
Fouet	Fouet pâtissier
Fourmi	Fourmi rouge
Fusée	Ariane, fusée Ariane
Grenouille	Rainette
Grue	Grue de chantier
Guitare	Guitare sèche, guitare acoustique
Haltères	Haltères d'haltérophilie
Lampe	Lampe de chevet
Lapin	Lapin de Garenne
Lime	Lime à bois
Lit	Lit simple
Maison	Pavillon, maison résidentielle
Melon	Melon français
Moto	BMW
Moulin	Moulin à vent
Oiseau	Moineau, fauvette
Ours	Ours brun
Pain	Flûte de pain
Palette	Palette européenne
Passoire	Chinois
Pendule	Francomtoise
Piano	Piano à queue
Piment	Piment rouge, piment thaï
Piscine (2)	Piscine hors-sol
Plongeur	Plongeur sous-marin
Plume	Plume de colvert

...le tableau continue à la page suivante

A.7. Images donnant lieu à des noms/concepts subordonnés

Image	Noms/concepts subordonnés
Roue	Roue de voiture
Poêle	Crêpière, poêle à crêpes
Poire	Poire Williams
Poivron	Poivron rouge
Porte	Porte d'entrée
Quille	Quille de bowling
Radis	Radis rond
Robe	Robe de grand-mère
Roulette	Roulette de jeu, roulette de casino
Salade	Laitue, batavia
Saucisse	Saucisses de Strasbourg, saucisse de Francfort
Scie	Scie égoïne, scie à bois
Scoter	Vespa
Serpent	Vipère, coulœuvre
Slip	Slip kangourou
Tampon	Tampon encreur
Tasse	Tasse à café
Tente	Canadienne, tente de camping, tente canadienne
Tondeuse	Tondeuse à gazon, tondeuse à mains
Tracteur	Tracteur agricole, tracteur Fiat
Vélo	BMX, VTT
Violon	Violon alto
Vis	Vis à bois
Voiture	Berline

Annexe B

Pré-test pour la seconde expérience d'accord sur le nom

B.1 Consigne

On va faire un test qui n'a rien à voir avec celui que vous avez fait la dernière fois. Cette fois, il ne s'agit pas de dire comment vous appelez telle ou telle chose, mais de communiquer avec quelqu'un. C'est-à-dire qu'en ce moment, il y a une autre personne qui, comme vous, est derrière un écran d'ordinateur. Je vous montre ce qu'elle voit.

[L'expérimentateur montre au sujet l'image de l'interface auditeur - cf. annexe B.2]

Vous devrez lui faire exécuter des actions sur les objets qui sont représentés sur l'écran. Vous ferez cela en parlant à voix haute, par le biais du micro de cet ordinateur.

Je vais vous dire comment va se dérouler le test.

D'abord, il y aura une phase d'entraînement, pendant laquelle on va faire quelques essais pour préciser ce que vous devez faire, et aussi pour voir si vous n'avez pas de problème avec l'ergonomie du logiciel.

Ensuite je téléphonerai à la deuxième personne pour lui dire de se tenir prête. Et puis on pourra commencer le test lui-même, qui ne durera pas très longtemps.

Vous avez des questions ?

[La phase d'entraînement commence¹. La première animation est jouée.]

Vous allez voir plusieurs animations comme celle que l'on vient de voir. À chaque fois qu'une animation est jouée, vous avez la possibilité de la visionner à nouveau en cliquant sur le bouton « rejouer ». Essayez.

1. Le fichier contenant les animations d'entraînement est disponible en annexe DVD 5b_6

Votre tâche est de faire exécuter l'action que vous venez de voir à la personne à qui vous parlez. C'est-à-dire que vous lui donnerez un ordre verbal pour qu'il ou elle exécute l'action sur les objets qu'il ou elle voit sur son écran. Par exemple, pour cette première animation, vous lui direz « mettez la carotte sous le gant ».

Quand vous êtes prêt(e) à parler, vous cliquez sur le bouton « parler ». Vous voyez que le bouton change, ça ça veut dire que l'autre personne entend ce que vous dites. Une fois que vous avez donné l'ordre, vous appuyez à nouveau sur le bouton pour qu'il revienne sur son état normal. Quand vous voyez l'animation, vous n'avez pas besoin de vous presser, vous pouvez prendre votre temps puis appuyer sur le bouton pour donner l'ordre. Par contre, une fois l'ordre donné, il faut attendre quelques secondes, le temps que l'autre personne agisse sur les objets qu'il ou elle voit sur son écran.

Le système n'est pas encore interactif, ce qui fait qu'on ne peut pas savoir en temps réel quand l'autre personne est prête à répondre à un nouvel ordre. Pour cette raison, à chaque ordre donné, vous devrez patienter une douzaine de secondes. Je compterai pour vous.

Allez on va commencer l'entraînement.

B.2 Interface auditeur montrée aux sujets



B.3 Résultats

Sujet	Nom attribué dans la première étude d'accord sur le nom	Nom utilisé en situation de communication
3	Fauvette	Petit oiseau marron
3	Club	Fauteuil marron en cuir
3	Ariane	Fusée blanche
3	Assiette creuse	Assiette creuse jaune et verte
4	Borsalino	Chapeau
4	Vespa	Scooter
4	Colvert	Colvert
5	Fauteuil club	Fauteuil club
5	Melon français	Melon
5	Chinois	Passoire
5	Guitare sèche	Guitare
5	Cuissardes	Cuissardes
5	Camion benne	Camion benne
5	Passe-montagne	Passe-montagne
5	Casque de moto	Casque
5	Fourmi rouge	Fourmi
5	Poire Williams	Poire
5	Ampoule à vis	Ampoule
5	Boxeur de boxe thaïe	Boxeur
5	Assiette creuse	Assiette
5	Chaussure en cuir	Chaussure
5	Poivron rouge	Poivron rouge
5	Piano à queue	Piano à queue
5	Télévision à écran plat	Télé
5	Saucisses de Strasbourg	Saucisses
5	Moineau	Oiseau
5	Ours brun	Ours brun
17	Vipère	Serpent
17	Maison résidentielle	Maison
17	Plume de colvert	Plume
17	Dogue allemand	Chien
17	Fusée Ariane	Fusée
17	Ablette	Poisson
17	Violon alto	Violon
17	Cheval de course	Cheval
17	Feuillu	Arbre
17	Guitare acoustique	Guitare
17	Église orthodoxe	Église
17	Lime à bois	Lime

...le tableau continue à la page suivante

ANNEXE B. PRÉ-TEST POUR LA SECONDE EXPÉRIENCE D'ACCORD SUR LE NOM

Sujet	Nom attribué dans la première étude d'accord sur le nom	Nom utilisé en situation de communication
17	Campagnol	Rat
17	Piano à queue	Piano
17	Camion-benne	Camion
17	Saucisses de Strasbourg	Saucisses
17	Chaussure de ville	Chaussure
21	Rainette	Grenouille
21	Batavia	Salade
21	Orchidée	Lys
21	Vespa	Scooter
21	Franc-comtoise	Franc-comtoise
21	Radis rond	Radis
21	Brosse à ongles	Brosse à ongles
21	Vinyle	Vinyle
21	Assiette à soupe	Assiette à soupe
21	Slip kangarou	Slip kangourou
21	Camion-benne	Camion-benne
21	Cuillère à café	Cuillère
21	Truite	Poisson

Annexe C

Seconde expérience d'accord sur le nom, et test de l'ergonomie de l'interface *locuteur* d'EloKanz

C.1 Sujets

N°	Sexe	Âge	Région	Milieu	Profess.	Domaines expertise	Pb. vue
1	F	19	Lang.- Roussill.	Rural	Étudiante L1	Linguistique, psychologie, épicerie	non
2	F	19	Midi- pyrénées	Rural	Étudiante L1, Baby-Sitter	Linguistique, piano	oui (lent. ^a)
3	F	27	Poitou- Charent.	Rural	Ingénieur recherche	Psychologie, cinéma	oui (lun. ^b)
4	F	21	Bretagne	Rural	Étudiante L1, caissière	Anglais, équitation, Japon	non
5	F	20	Midi- Pyrénées	Urbain	Étudiante L1	Linguistique	oui (lun. ^b)
6	F	27	Midi- Pyrénées	Rural	A.T.E.R.	Linguistique, jeux de soc., musique	oui (lent. ^a)
7	F	26	Lang.- Roussill.	Rural	A.T.E.R.	Psychologie	oui (lent. ^a)
8	F	22	Québec	Urbain	Étudiante L1	Linguistique, psychologie, ski, vélo	oui (lent. ^a)
9	F	24	Midi- Pyrénées	Rural	Chargée d'études	Linguistique, musique, tennis,	non

...le tableau continue à la page suivante

Participants au premier test d'ergonomie (suite)

						cinéma	
10	F	19	Midi-Pyrénées	Urbain	Étudiante L1	Linguistique, équitation	non
11	F	23	Midi-Pyrénées	Rural	Étudiante L1, caissière	Linguistique, boulangerie	non
12	F	20	Midi-Pyrénées	Urbain	Étudiante L1, baby-sitter	Linguistique, littérature	oui (lent. ^a)

^a Porte des lentilles de contact

^b Porte des lunettes de vue

C.2 Consignes utilisées

C.2.1 Consigne générale

Avant de commencer, je vais vous expliquer assez brièvement de but de l'expérience. On est en train de développer un logiciel pour évaluer les performances orales de personnes qui ont des troubles d'élocution ; le but du jeu est d'arriver à observer si ces personnes se font plus ou moins bien comprendre par d'autres personnes lorsqu'ils parlent. Le principe de notre logiciel est très simple : celui qui parle doit faire exécuter des actions à celui qui écoute. Ces actions concernent des déplacements d'objets qui sont représentés à l'écran. Pour donner un exemple simple, on peut imaginer qu'il y ait à l'écran une voiture, un chapeau et un chat. Celui qui parle va demander à son interlocuteur de placer la voiture en dessous du chat. Et nous on estime que si l'auditeur effectue l'action correspondante, c'est qu'il a compris la phrase.

La communication entre celui qui parle et celui qui écoute n'est pas directe, c'est-à-dire que vous allez donner des ordres, qu'on va enregistrer (vous allez être équipé(e) d'un micro), et ce n'est que plus tard que l'on fera écouter les enregistrements à d'autres personnes pour voir si elles comprennent les phrases.

Je vais vous dire comment va se dérouler le test. D'abord, il y aura une phase d'entraînement, pendant laquelle on va faire quelques essais pour bien préciser ce que vous devez faire, et aussi pour voir si vous n'avez pas de problème avec l'ergonomie du logiciel. Ensuite, on va faire trois séries d'une vingtaine d'enregistrements.

Avez-vous des questions ?

C.2.2 Consigne donnée pour la phase d'entraînement

Quand vous êtes prêt(e), vous cliquez ici (réponse [OK] donnée à la boîte de dialogue « Êtes-vous prêt(e) ? »). Voilà vous venez de voir une animation ; au cours du test vous en verrez une quinzaine par étape. À chaque fois qu'une animation est jouée, vous avez la possibilité de la visionner à nouveau en cliquant sur le bouton « rejouer », si vous n'avez pas bien vu ou que vous vous souvenez plus de l'animation. Essayez.

Votre tâche est comme je vous l'ai dit de faire exécuter l'action que vous venez de voir à la personne à qui vous parlez. C'est-à-dire que vous lui donnez un ordre verbal pour qu'il ou elle exécute l'action correspondante sur les objets. Par exemple, pour cette première animation, vous pourriez lui dire par exemple (donner un exemple d'énoncé correspondant à l'animation visionnée par le sujet). Les personnes qui écouteront vos ordres verront exactement les mêmes objets que vous voyez, ils auront exactement le même type d'écran. Sauf que eux auront la possibilité de bouger les images pour répondre à vos ordres.

Vous avez compris ?

Quand vous êtes prêt(e) à parler, vous cliquez sur le bouton « parler ». Vous voyez que le bouton change, il y a une petite roue qui tourne, qui veut dire que l'enregistrement a débuté. À chaque fois que vous cliquez sur le bouton parler, attendez bien de voir cette petite roue qui tourne avant de parler. Et une fois que vous avez parlé, vous appuyez à nouveau sur le bouton pour qu'il revienne sur l'état normal.

Quand vous voyez l'animation, vous n'avez pas besoin de vous presser, vous pouvez prendre votre temps puis appuyer sur le bouton pour donner l'ordre. Et une fois que vous avez enregistré, vous cliquez sur « suivant » pour passer à la prochaine animation. Une fois le test fini, vous verrez écrit sur l'écran « Merci pour votre participation », et à ce moment-là vous pouvez m'appeler.

Ça va, on peut commencer l'entraînement ?

C.2.3 Questionnaire administré après chaque session

Si vous deviez noter la difficulté de l'épreuve sur une échelle de 1 à 7 – 1 étant le plus facile, et 7 le plus difficile – quelle note donneriez-vous :

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Pourquoi ?

Si vous deviez noter la « lourdeur » de l'épreuve (si cela vous a paru lourd, fatigant) sur une échelle de 1 à 7 – 1 étant le moins lourd, et 7 le plus lourd – quelle note donneriez-vous :

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Pourquoi ?

Est-ce qu'il y a des aspects de l'épreuve qui vous ont semblé difficiles ?

Est-ce qu'il y a des aspects de l'épreuve qui vous ont semblé bizarres, peu naturels ou bien déconcertants ?

Avez-vous eu des difficultés par rapport à certaines images en particulier ?

Avez-vous d'autres remarques ?

Annexe D

Expérience sur l'intelligibilité et la compréhension de la parole dans le bruit

D.1 Sujets

Sujet	Sexe	Âge	Région	Milieu	Profession
1	F	18	Midi-Pyrénées	Rural	Étudiante (L1 esp. / SDL)
2	F	22	Poitou-Charentes	Rural	Étudiante (M2 SDL/FLES)
3	M	19	Limousin	Rural	Étudiant (L2 SDL)
4	F	20	Aquitaine	Rural	Étudiante (L1 Psycho)
5	F	28	Midi-Pyrénées	Urbain	<i>Sans activité</i>
6	F	20	Midi-Pyrénées	Urbain	Étudiante (L1 Lettres/SDL)
7	F	17	Aquitaine	Urbain	Étudiante (L1 anglais/SDL)
8	F	25	Centre	Rural	Étudiante (Doct. Psycho)
9	F	19	Midi-Pyrénées	Rural	Étudiante (L1 LSF/SDL)
10	F	17	Île-de-France	Rural	Étudiante (L1 Psycho/SDL)
11	F	18	Midi-Pyrénées	Urbain	Étudiante (L1 Socio/SDL)
12	F	19	Midi-Pyrénées	Rural	Étudiante (L3 SDL + anglais)
13	F	18	Midi-Pyrénées	Rural	Étudiante (L1 anglais/SDL)
14	H	23	Midi-Pyrénées	Rural	Étudiant (M2 SDL)
15	F	28	Midi-Pyrénées	Urbain	Étudiante (Doct. SDL)
16	F	21	Midi-Pyrénées	Rural	Étudiante (L2 SDL)
17	F	18	Midi-Pyrénées	Rural	Étudiante (L1 LSF/SDL)
18	F	20	Aquitaine	Urbain	Étudiante (L2 SDL)
19	F	26	Midi-Pyrénées	Rural	Étudiante (Doct. SDL)
20	F	20	P.A.C.A.	Urbain	Étudiante (L1 SDL)
21	F	20	Midi-Pyrénées	Rural	Étudiante (L2 SDL)
22	F	19	Midi-Pyrénées	Rural	Étudiante (L1 SDL)
23	F	21	Midi-Pyrénées	Rural	Étudiante (M1 anglais)

...le tableau continue à la page suivante

ANNEXE D. EXPÉRIENCE SUR L'INTELLIGIBILITÉ ET LA COMPRÉHENSION
DE LA PAROLE DANS LE BRUIT

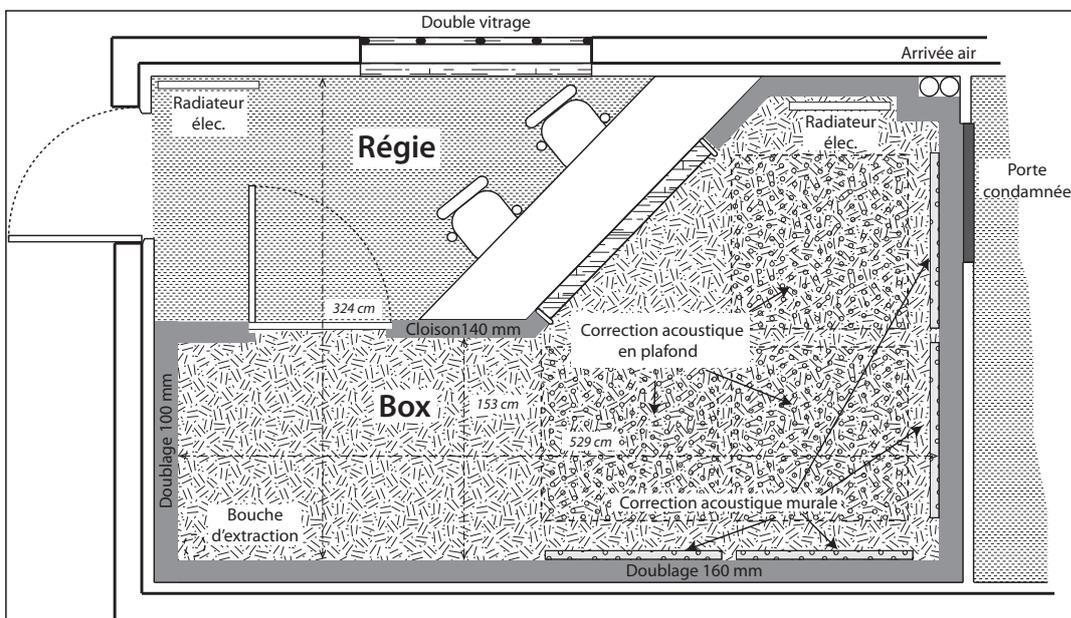
Tableau des sujets (suite)

24	H	24	La Réunion	Rural	Étudiant (M2 SDL)
25	F	22	Centre	Rural	Étudiante (L2 Psycho)
26	F	23	Midi-Pyrénées	Rural	Étudiante (M1 SDL)
27	F	27	Midi-Pyrénées	Urbain	Étudiante (M1 SDL)
28	F	25	Aquitaine	Rural	Étudiante (M1 SDL)
29	F	23	Languedoc-Roussillon	Rural	Étudiante (M1 SDL)
30	F	20	Centre	Urbain	Étudiante (L1 SDL)
31	F	27	Midi-Pyrénées	Rural	Étudiante (Doct. SDL)
32	F	25	Limousin	Urbain	Étudiante (M1 SDL)
33	H	26	Midi-Pyrénées	Rural	Ingénieur informatique
34	F	35	Rhône-Alpes	Urbain	Étudiante (L3 Psycho)
35	F	17	Midi-Pyrénées	Urbain	Étudiante (L1 SDL)
36	F	21	Midi-Pyrénées	Rural	Étudiante (L1 SDL)
37	F	20	Midi-Pyrénées	Rural	Étudiante (L1 SDL)
38	F	20	Midi-Pyrénées	Rural	Étudiante (L1 SDL)
39	F	20	Nord-Pas-de-Calais	Rural	Étudiante (L1 SDL)
40	F	25	Midi-Pyrénées	Rural	Étudiante (Doct. Psycho/SDL)

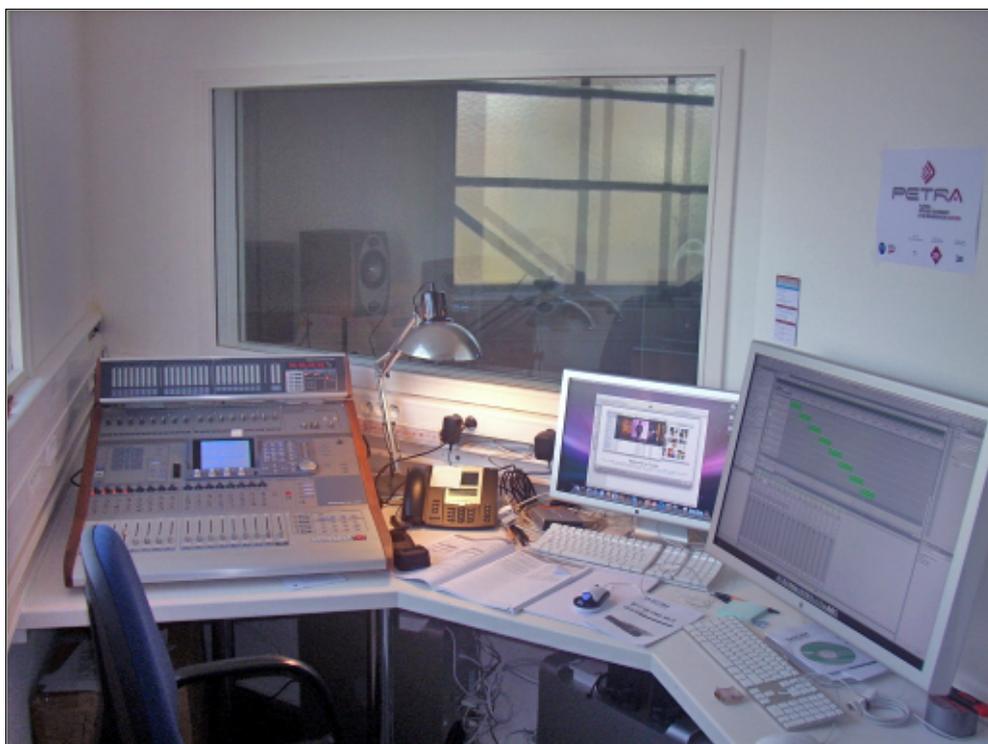
D.2 Phrases utilisées comme stimuli

1. Mettez le lit au dessous de la banane.
2. Mettez la grue au dessus de la pipe.
3. Mettez le dé à droite du gant.
4. Mettez le crabe à droite du serpent.
5. Mettez la canne à droite du kangourou.
6. Mettez l'éléphant à gauche de l'hérisson.
7. Mettez la grenouille au dessous du pantalon.
8. Mettez l'avion au dessous de la fourchette.
9. Mettez la tomate au dessous du camion.
10. Mettez le piano au dessous du marteau.
11. Mettez la cuillère au dessous de la passoire.
12. Mettez le chien au dessus du couteau.
13. Mettez la tondeuse à droite de la girafe.
14. Mettez le bonbon au dessous de la fleur.
15. Mettez la grenouille au dessus du couteau.
16. Mettez la tente au dessous du bonbon.
17. Mettez la bouteille au dessous de l'ananas.
18. Mettez le croissant au dessous du chat.
19. Mettez la canne à droite du boxeur.
20. Mettez le lit à droite de la pelle.
21. Mettez la brouette au dessous de la canne.
22. Mettez le dé au dessus du chien.
23. Mettez le radis au dessus de l'oignon.
24. Mettez le kangourou à gauche du poulet.
25. Mettez le chat à gauche du poisson.
26. Mettez la fourchette à droite du tampon.
27. Mettez le kangourou au dessus de l'ampoule.
28. Mettez la fraise au dessous de l'ours.
29. Mettez l'aspirateur au dessous de la pastèque.
30. Mettez le pantalon à gauche de la fleur.
31. Mettez la brosse au dessus du lion.
32. Mettez le scooter au dessous de l'entonnoir.
33. Mettez la cuillère à gauche de l'éléphant.
34. Mettez le trèfle au dessous du chien.
35. Mettez l'hélicoptère au dessous du pantalon.
36. Mettez la quille au dessus de l'hérisson.
37. Mettez l'oignon à gauche du moulin.
38. Mettez le crabe au dessus des lunettes.
39. Mettez le pantalon à gauche de l'ours.
40. Mettez les boutons à gauche de la boîte d'allumettes.
41. Mettez le marteau au dessus de la fourchette.
42. Mettez la poire au dessus du cheval.
43. Placez le tampon au dessus du kangourou.
44. Placez la fusée au dessous de la fourchette.
45. Mettez le dé au dessus du papillon.
46. Mettez le bonbon au dessous de la poire.
47. Mettez la fourchette au dessous de la tondeuse.
48. Mettez la loupe au dessus du poulet.
49. Mettez le bonbon au dessus de l'autruche.
50. Mettez le parapluie à gauche du camion.

D.3 Cabine Petra-2



Cabine Petra-2 – Plan d'architecte



Cabine Petra-2 – Régie



Cabine Petra-2 – Box

D.4 Algorithme général pour la concaténation des fichiers audio

```
# Programme de concaténation des fichiers audio
# Entrée : URL du dossier contenant les fichiers audio source
# Concatène 180 fichiers audio, avec un décalage temporel variable entre chacun des fichiers.
# 1. On alterne entre les locuteurs (1 à 9)
  tableau_Alterne; # Tableau contenant les URL des fichiers audio
  Pour chaque indice de fichier (1 à 20) faire
    Pour chaque locuteur (0 à 8) faire
      indice_fichier_a_ranger <- (20 * indice_locuteur) + indice_fichier;
      Ranger l'URL du fichier indice_fichier_a_ranger dans le tableau tableau_alterne;
    Fin pour;
  Fin pour;
# 2. On crée un fichier audio temporaire
  Temp_Audio; # Fichier audio temporaire
# 3. On efface le silence en début et en fin des fichiers, et on les concatène
  Pour chaque fichier dont l'URL est contenue dans le tableau tableau_Alterne faire
    Effacer le silence (seuil de -40dB) en début de fichier, avec l'utilitaire SoX;
    Inverser le fichier, avec l'utilitaire SoX;
    Effacer le silence (seuil de -40dB) en début de fichier, avec l'utilitaire SoX;
    Inverser le fichier, avec l'utilitaire SoX;
    RandomGap <- Nombre aléatoire, à une décimale, compris entre 0.1 et 1.0;
    Mixer le fichier avec le fichier Temp_Audio, après un décalage de RandomGap seconde
  (commande SoX);
  Fin pour;
# 4. On coupe le fichier audio temporaire : on enlève les 5 premières secondes, pour ne pas
conserver la montée en amplitude apparaissant au début du signal.
# 5. On transforme le fichier Temp_Audio en fichier mono, et on l'enregistre sous le nom
« babble.wav »
```

D.5 Consignes

D.5.1 Consigne pour le test d'intelligibilité

Nous faisons un test sur la compréhension de la parole en milieu bruyant. De manière globale, le test se passe de la manière suivante. Via ces deux haut-parleurs, vous allez entendre des phrases enregistrées. Ces phrases correspondent toujours à des commandes de déplacement, du type « mettez le chat au-dessus de la voiture », « placez le fromage à droite de la cafetière », « mettez le carton au-dessous du balai », etc. En même temps que ces phrases, vous allez entendre du bruit (un brouhaha) qui va masquer partiellement les phrases. Le but du jeu pour vous c'est d'écrire ce que vous entendez sur cette feuille. Le numéro de chaque phrase est inscrit sur l'écran, et sur votre feuille.

Si vous n'êtes pas sûr(e) d'un mot, essayez quand même d'écrire ce que vous avez entendu, ou de deviner de quel mot il s'agit, si possible.

Pour que vous compreniez mieux, nous allons d'abord faire une phase d'entraînement, sur une dizaine de phrases. Ensuite on commencera le test proprement dit, qui porte sur une cinquantaine de phrases.

Avez-vous des questions ?

D.5.2 Consigne pour le test de compréhension

[L'expérimentateur lance le premier écran, où sont présentés les objets correspondant à la première animation]

Nous faisons un test sur la compréhension de la parole en milieu bruyant. De manière globale, le test se passe de la manière suivante. Sur l'écran que vous avez en face de vous, sont représentés des objets. Via ces deux haut-parleurs, vous allez entendre des phrases enregistrées. Ces phrases consistent en des ordres vous demandant de déplacer tel ou tel objet à tel endroit, par exemple « mettez le chat au-dessus de la voiture », « placez le fromage à droite de la cafetière », « mettez le carton au-dessous du balai », etc. La difficulté réside dans le fait qu'en même temps que les phrases, un brouhaha est diffusé, plus ou moins fort selon les phrases. Pour vous le but du jeu est de répondre aux commandes orales en déplaçant les objets représentés à l'écran.

Si vous n'êtes pas sûr(e) de ce que vous avez entendu, essayez quand même de deviner. Par exemple si vous avez entendu seulement le nom d'un objet, ou d'une position (droite, gauche, dessus, dessous), essayez quand même de faire un mouvement qui implique ces informations :

- *si vous n'avez entendu que le nom du premier objet, par exemple « mettez le chat... », sélectionnez l'image correspondant au chat et déplacez-là n'importe où à côté de n'importe quel autre objet ;*
- *si vous n'avez entendu que la position, par exemple « ...droite... », sélectionnez n'importe quel objet et déplacez-le à droite de n'importe quel autre objet ;*

- *si vous n'avez entendu que le dernier objet, par exemple « ...du chat », sélectionnez n'importe quel objet et déplacez-le n'importe où à côté du chat.*

Par contre si vous n'avez rien compris du tout, cliquez directement sur le bouton Valider.

Pour que vous compreniez mieux, on va d'abord faire une phase d'entraînement, sur une dizaine de phrases. Ensuite on commencera le test proprement dit, qui porte sur une cinquantaine de phrases.

Vous avez des questions ?

Annexe E

Études de faisabilité et de validation du test Elokanz

E.1 Patients inclus dans l'étude

N°	Code	Sexe	Âge	Pathologie	MMS
1	PN01	H	69	Paralysie labio- glosso- vélo- pharyngo- laryngée	28
2	PN02	H	74	Paralysie des dernières paires crâniennes	26
3	PN03	H	75	Maladie de Parkinson	20
4	PN04	H	68	Maladie de Steele Richardson	26
5	PN05	F	58	Maladie de Parkinson	27
6	PN06	H	71	Syndrome pseudo bulbaire	22
7	PN07	F	83	Syndrome cérébelleux	21
8	PN08	H	68	?	29
9	PN09	F	66	Dysphonie spasmodique	30
10	PN10	H	70	Maladie de Parkinson	27
11	PN11	F	39	Sclérose en plaques	27
12	PN12	H	53	Chorée de Huntington	29
13	PN13	F	72	Ataxie cérébelleuse	25
14	PN14	H	63	Tremblement vocal, résection linguale, surdité	29
15	PN15	H	33	Chorée de Huntington	25
16	PA02	H	45	Dysphonie spasmodique	29
17	PA03	F	46	Séquelles thérapeutiques cancer cavité buccale	30
18	PA04	H	79	Séquelles thérapeutiques cancer oropharynx	24
19	PA06	H	52	Séquelles thérapeutiques cancer	29
20	PA07	H	73	Séquelles thérapeutiques	27

Patients inclus dans l'étude (suite)

				cancer larynx (chirurgie partielle)	
21	PA08	H	63	Séquelles thérapeutiques cancer – laryngectomie totale	27
22	PA09	H	66	Séquelles thérapeutiques cancer oro-pharyngé	27
23	PA11	H	76	Séquelles thérapeutiques cancer larynx	25
24	PA12	F	86	Tumeur laryngée bénigne	29
25	PA13	H	73	Séquelles thérapeutiques cancer cavité buccale	28
26	PA14	F	64	Dysphonie	28
27	PA15	H	58	Séquelles thérapeutiques cancer larynx (chirurgie partielle)	28
28	PA16	F	48	Séquelles thérapeutiques cancer – laryngectomie totale	28
29	PA17	H	49	Séquelles thérapeutiques cancer (insuffisance vélaire gauche)	27
30	PA18	F	50	Dysphonie	30

E.2 Grille d'observation pour l'enregistrement des patients

CODE PATIENT :
A. Phase de lecture des consignes
Heure de départ de la lecture des consignes (hh:mm:ss) :
(Questions posées par le patient, remarques) :
Heure de fin de lecture des consignes (hh:mm:ss) :
B. Phase d'entraînement
Heure de départ de l'entraînement (hh:mm:ss) :
B.1 Difficultés observées propres à l'ergonomie
a. Environnement périphérique (micro, souris) :
b. Environnement logiciel (boutons, ergonomie) :
<i>(la grille continue à la page suivante)</i>

Grille d'observation (suite)

c. Écran (taille des images, emplacement des images, etc.) :

B.2 Difficultés observées, propres au patient (motifs d'exclusion pour l'étude)

- Le patient ne parvient pas à identifier les images, à cause :
 - d'un trouble visuel
 - d'un trouble cognitif
 - autre :
- Le patient ne parvient pas à mémoriser les animations
- Le patient ne parvient pas à situer les images les unes par rapport aux autres (droite, gauche, haut, bas)
- Le patient ne formule pas des messages corrects du point de vue linguistique, notamment au(x) niveau(x) :
 - de la syntaxe
 - de la morphologie
 - de la phonologie
 - du lexique

B.3 Autres difficultés observées (questions posées par le patient, remarques)

Heure de fin de l'entraînement (hh:mm:ss) :

(la grille continue à la page suivante)

Grille d'observation (suite)

C. Phase d'enregistrement
Heure de départ de l'enregistrement (hh:mm:ss) :
C.1 Difficultés observées propres à l'environnement
a. Environnement périphérique (micro, souris) :
b. Environnement logiciel (boutons, ergonomie) :
b. Écran (taille des images, emplacement des images, etc.) :
C.1 Autres difficultés observées (remarques)
Heure de fin de l'enregistrement (hh:mm:ss) :

E.3 Questionnaire post-test (patients et auditeurs)

Est-il facile ou difficile pour vous d'utiliser ce logiciel ?

1

Très
facile

2

Assez
facile

3

Ni facile,
ni difficile

4

Assez
difficile

5

Très
difficile

(Difficultés rapportées par le sujet, commentaires)

E.4 Lettre d'informations aux patients

OBJET : Information pour la participation à l'étude de faisabilité du logiciel EloKanz

Service	Investigateur principal
Unité de la voix et de la déglutition Service d'otorhinolaryngologie et de chirurgie cervicofaciale Hôpital Larrey Rangueil C.H.U. Toulouse 31059 Toulouse Cedex 9	Virginie Woisard woisard.v@chu-toulouse.fr 05 87 77 17 18

Madame, Monsieur,

La présente étude a pour objectif de tester le logiciel EloKanz, un outil conçu pour la prise en charge de personnes souffrant de troubles pathologiques de production de la parole. Le logiciel EloKanz permet d'enregistrer des phrases prononcées par des patients souffrant de troubles pathologiques de production de la parole, et ensuite d'évaluer la facilité avec laquelle ces phrases sont comprises par des auditeurs. Cette étude vise plus précisément à vérifier que le logiciel EloKanz est bien adapté à la situation, autrement dit que son ergonomie ne pose pas de difficulté majeure aux patients et au personnel médical.

L'étude dure au maximum une heure, durant laquelle vous serez amené(e) à rester dans la salle dédiée au bilan de la voix et de la parole, dans l'Unité de la Voix et de la Déglutition de l'hôpital Larrey. Vous serez pris(e) en charge par deux étudiantes en orthophonie, qui vous expliqueront le fonctionnement du logiciel avant de vous laisser enregistrer une cinquantaine de phrases. Les étudiantes vous poseront également quelques questions, et évalueront l'aisance avec laquelle vous utilisez le logiciel EloKanz.

Cette étude ne présente aucun risque particulier pour votre santé, elle consistera uniquement en un enregistrement. Elle s'intègre dans le bilan justifié par vos difficultés à parler. Elle ne justifie aucune venue supplémentaire dans le service. Elle est compatible avec votre participation éventuelle à d'autres protocoles de recherche.

Votre participation à l'étude contribue à l'élaboration d'un outil fiable et efficace pour l'évaluation des progrès thérapeutiques des patients souffrant de troubles pathologiques de production de la parole, outil élaboré dans le cadre d'une thèse universitaire en sciences du langage.

Sachez également que **vous avez la possibilité de vous soustraire à l'étude à tout moment, et ce sans en donner la raison.**

À tout moment, vous pouvez vous adresser au responsable de l'étude pour l'obtention d'information complémentaire.

Toutes les données recueillies dans cette étude et permettant de vous identifier resteront confidentielles. Seules les données anonymisées, et les résultats analysés pourront éventuellement faire l'objet de publications scientifiques.

E.5 Formulaire de consentement éclairé

Formulaire de consentement éclairé pour la participation à l'étude de faisabilité du logiciel EloKanz

Service	Investigateur principal
Unité de la voix et de la déglutition Service d'otorhinolaryngologie et de chirurgie cervicofaciale Hôpital Larrey Rangueil C.H.U. Toulouse 31059 Toulouse Cedex 9	Virginie Woisard woisard.v@chu-toulouse.fr 05 87 77 17 18

J'ai bien été informé(e) oralement et par écrit des buts de l'étude portant sur la faisabilité du logiciel EloKanz, de son déroulement, des effets attendus, des avantages et inconvénients possibles ainsi que des risques éventuels.

J'ai lu et compris la lettre d'information pour l'étude susnommée. J'ai reçu des réponses satisfaisantes aux questions concernant ma participation à cette étude. Je peux garder la lettre d'information et je reçois une copie de ma déclaration écrite de consentement.

Je consens à participer,

(Date et signature de la personne donnant son consentement)

(Date et signature de l'investigateur)

E.6 Feuille de passation du *Mini-Mental State* – *MMSE*

Mini-Mental State Examination (MMSE)

Nom et prénom du patient : _____ Date : _____

Je vais vous poser quelques questions pour apprécier comment fonctionne votre mémoire.
Les unes sont très simples, les autres un peu moins. Vous devez répondre du mieux que vous pouvez.

Réponse correcte : Noter 1
Réponse incorrecte : Noter 0

ORIENTATION	
Quelle est la date complète d'aujourd'hui ? <i>Si la réponse est incorrecte ou incomplète, posez les questions restées sans réponse dans l'ordre suivant :</i>	1. En quelle année sommes-nous ? <input type="text"/>
	2. En quelle saison ? <input type="text"/>
	3. En quel mois ? <input type="text"/>
	4. Quel jour du mois ? <input type="text"/>
	5. Quel jour de la semaine ? <input type="text"/>
	Sous-total <input type="text"/> / 5
Je vais vous poser maintenant quelques questions sur l'endroit où nous nous trouvons.	6. Quel est le nom du cabinet (hôpital, centre, etc...) où nous sommes ? <input type="text"/>
	7. Dans quelle ville se trouve-t-il ? <input type="text"/>
	8. Quel est le nom du département dans lequel est située cette ville ? <input type="text"/>
	9. Dans quelle province ou région est situé ce département ? <input type="text"/>
	10. A quel étage sommes-nous ? <input type="text"/>
	Sous-total <input type="text"/> / 5
APPRENTISSAGE	
Je vais vous dire 3 mots; je voudrais que vous me les répétiez et que vous essayiez de les retenir car je vous les redemanderai tout à l'heure :	11. Cigare <input type="text"/>
	12. Fleur <input type="text"/>
	13. Porte <input type="text"/>
	(Faire répéter les 3 mots)
	Sous-total <input type="text"/> / 3
ATTENTION ET CALCUL :	
Voulez-vous compter à partir de 100 en retirant 7 à chaque fois ?	14. 100 – 7 = 93 <input type="text"/>
	15. 93 – 7 = 86 <input type="text"/>
	16. 86 – 7 = 79 <input type="text"/>
	17. 79 – 7 = 72 <input type="text"/>
	18. 72 – 7 = 65 <input type="text"/>
	Sous-total <input type="text"/> / 5
Pour tous les sujets, même ceux qui ont obtenu le maximum de points, demander : Voulez-vous épeler le MONDE à l'envers : EDNOM <i>Noter le nombre de lettres données dans l'ordre correct, mais ce score ne doit pas figurer dans le score total.</i>	19. Cigare <input type="text"/>
	20. Fleur <input type="text"/>
	21. Porte <input type="text"/>
	Sous-total <input type="text"/> / 3
RAPPEL	
Pouvez-vous me dire quels étaient les 3 mots que je vous ai demandé de répéter et de retenir tout à l'heure ?	
LANGAGE	
22. Montrer un crayon : Quel est le nom de cet objet ?* <input type="text"/>	
23. Montrer une montre : Quel est le nom de cet objet ?** <input type="text"/>	
24. Ecoutez bien et répétez après moi*** : « pas de MAIS, de SI, ni de ET » <input type="text"/>	
25. Posez une feuille de papier sur le bureau, la montrer au patient en lui disant : Ecoutez bien et faites ce que je vais vous dire : Prenez cette feuille de papier de la main droite <input type="text"/>	
26. Pliez-la en deux <input type="text"/>	
27. Et jetez-là par terre**** <input type="text"/>	
28. Tendre au patient la feuille de papier ci-jointe sur laquelle est écrit en gros caractères « FERMEZ LES YEUX » et dire au patient : « Faites ce qui est écrit » <input type="text"/>	
29. Tendre au patient une feuille de papier et un stylo en disant : « Voulez-vous m'écrire une phrase, ce que vous voulez, mais une phrase entière»***** <input type="text"/>	
	Sous-total <input type="text"/> / 8
PRAXIES CONSTRUCTIVES	
30. Tendre au patient la feuille de papier ci-jointe et lui demander : « Voulez-vous recopier ce dessin » <input type="text"/>	
	Sous-total <input type="text"/> / 1

SCORE TOTAL (1 À 30)

TOTAL / 30

E.7 Salle d'enregistrement et d'écoute



E.8 Consigne donnée aux patients

Comme nous vous l'avons expliqué rapidement¹, nous sommes en train de tester un logiciel qui sert à évaluer si vous arrivez à vous faire comprendre à l'oral. À l'aide de ce logiciel, nous allons vous enregistrer, puis faire écouter plus tard les enregistrements à d'autres personnes pour voir si elles vous comprennent. L'enregistrement va se dérouler de la manière suivante.

[L'investigateur lance l'écran où sont présentés les objets]

Vous allez voir plusieurs écrans de ce type, où sont représentés deux objets. Pour chaque écran, vous aller indiquer quand vous êtes prêt(e) en cliquant ici.

[L'investigateur choisit la réponse « OK » dans la boîte de dialogue « Êtes-vous prêt(e) ? »]

Vous venez de voir une animation. Si vous n'avez pas bien vu cette animation, ou que vous ne vous en souvenez plus, à tout moment vous avez la possibilité de la visionner à nouveau en cliquant sur le bouton « rejouer ».

[L'investigateur demande au patient de cliquer sur le bouton « rejouer »]

Votre objectif est de faire exécuter l'action que vous venez de voir à la personne à qui vous parlez. C'est-à-dire que vous aller donner un ordre verbal pour qu'il ou elle exécute l'action correspondante sur les objets. Par exemple, pour cette première animation, vous pourriez lui dire par exemple :

[L'investigateur donne un exemple d'énoncé correspondant à l'animation visionnée par le patient : Mettez X sur/sous/à droite/à gauche de Y]

Les animations correspondent toujours à des déplacements d'objets, c'est-à-dire qu'à chaque fois un des deux objets va être déplacé à gauche, à droite, au-dessus ou en dessous du second objet. Plus tard, les personnes qui écouteront vos ordres verront exactement les mêmes objets que vous voyez, ils auront exactement le même type d'écran. Sauf que eux auront la possibilité de bouger les images pour répondre à vos ordres.

Si vous ne reconnaissez pas les objets qui sont représentés à l'écran, vous pouvez directement cliquer sur le bouton « suivant » pour passer à l'animation suivante. Sinon, quand vous êtes prêt(e) à parler, vous cliquez sur le bouton « parler ».

[L'investigateur invite le patient à cliquer sur le bouton « parler »]

1. lors de la lecture de la fiche d'information au patient (annexe E.4)

Vous voyez que le bouton change, il y a une petite roue qui tourne, qui veut dire que l'enregistrement a bien commencé. À chaque fois que vous cliquez sur le bouton parler, attendez bien de voir cette petite roue qui tourne avant de parler. Et une fois que vous avez parlé, vous appuyez à nouveau sur le bouton pour qu'il revienne sur son état normal. Si jamais vous vous êtes trompé en parlant, vous avez la possibilité de recommencer l'enregistrement. Il vous suffit de cliquer à nouveau sur le bouton parler pour recommencer. L'ancien enregistrement sera effacé.

[L'investigateur s'assure que le patient n'a pas de questions]

Nous allons d'abord faire une phase d'entraînement, pendant laquelle on va faire quelques essais pour bien préciser ce que vous devez faire, et aussi pour voir si vous n'avez pas de problème avec le logiciel. Après cet entraînement, on enregistrera une cinquantaine de phrases.

[L'investigateur demande au patient s'il est prêt à commencer l'entraînement]

E.9 Questions et remarques faites par les patients durant la passation

Types de questions et remarques	Questions et remarques faites pendant la phase d'explication des consignes	Questions et remarques faites pendant la phase d'entraînement	Questions et remarques faites pendant la phase d'enregistrement
Images	<ul style="list-style-type: none"> - « C'est bien une brosse ? » (PN02) - « On dirait une espadrille. » (PN02 à propos de la brosse) 	<ul style="list-style-type: none"> - « C'est quoi, c'est un bonnet ? » (PN08) - « On dirait que la tomate est plus grosse que le poisson. » (PN09) - « Est-ce que c'est bien une brosse ? » (PA02) - « C'est un couteau ou un poignard ? » (PA14) 	<ul style="list-style-type: none"> - « C'est une capsule ? » (PN08 à propos de la fusée) - « Ça c'est une gomme ? » (PN09) - « C'est une feuille de quoi ça ? » (PN12) - « Comment ça s'appelle ? » (PN15 à propos de la coccinelle) - « Bon tant pis je vais tricher. » (PN15) - « C'est une poire ou un kiwi ? » (PA11 à propos de l'avocat) - « Ah c'est l'avocat, ça y est » (PA11) - « Qu'es aquò ? C'est quoi ce truc ? » (PA13 à propos de la gomme)
Pavé tactile	<ul style="list-style-type: none"> - « J'aime pas beaucoup ce truc. » (PN14) - « Je me sers d'une souris d'habitude. » (PN14) - « J'ai déjà utilisé un ordinateur, avec une souris. » (PA18) 	<ul style="list-style-type: none"> - « Je me suis habitué à la souris. » (PA09) - « Je préfère la souris classique. » (PN05) 	<ul style="list-style-type: none"> - « Je préfère la souris. » (PN10) - « J'aurais dû amener un marteau ! » (PA08) - « Ça va finir par m'énervé ! » (PA08) - « La prochaine fois, il faut mettre une souris, ça devient pénible ! » (PA09)

Questions et remarques faites par les patients durant la passation (suite)

Boutons		<ul style="list-style-type: none"> - « C'est embêtant de devoir remonter jusqu'au rectangle 'Êtes-vous prêt?' » (PN05) - « Si le 'Ok' était plus gros, ça irait plus vite et ça serait plus facile. » (PA07) 	
Ordinateur	<ul style="list-style-type: none"> - « Je n'ai jamais utilisé d'ordinateur. » (PN06, PN07, PA12) - « Je ne suis pas habitué à ces <i>Mac</i>. » (PA07) - « J'ai déjà utilisé un ordinateur. » (PA15, PA16) 	<ul style="list-style-type: none"> - « Je suis très malhabile en informatique, je ne m'en sers plus depuis que je ne travaille plus. » (PA13) 	

E.10 Commentaires post-test des patients

Types de commentaires	Transcription des commentaires
Commentaires négatifs sur les images	<ul style="list-style-type: none"> - « On ne reconnaît pas tous les objets : avocat, fusée, poire. » (PN08) - « La taille des images est insuffisante. » (PN08) - « Grossir les images aurait pu être bien. » (PN14) - « La gomme n'est pas reconnaissable par rapport à la couleur. » (PA02) - « Le paon est difficile à reconnaître » (PA04) - « Les images sont assez parlantes, sauf une ou deux. » (PA13) - « La pastèque est difficile à reconnaître. » (PA17) - « L'image de la balance est plus difficile à reconnaître. » (PA18)
Commentaires positifs sur les images	<ul style="list-style-type: none"> - « La taille des images convient. » (PA14) - « Les images sont bien claires. » (PA14) - « Les images sont faciles à reconnaître, on les voit bien. » (PA15)
Commentaires négatifs sur l'ordinateur en général	<ul style="list-style-type: none"> - « Ce serait très facile si je savais utiliser un ordinateur. » (PN02) - « Je ne me sers pas de l'ordinateur chez moi. » (PN04) - « Il faut un apprentissage de l'ordinateur. » (PA04)
Commentaires positifs sur le test (aspect ludique)	<ul style="list-style-type: none"> - « C'est ludique » (PN11) - « C'est comme un jeu, ce n'est pas fastidieux. » (PA13) - « C'est ludique donc c'est sympa. » (PA16) - « C'est plus sympa que l'enregistrement de voix. » (PA03)
Commentaires négatifs sur le pavé tactile	<ul style="list-style-type: none"> - « La manipulation tactile est difficile. » (PN05) - « C'est fatigant de refaire le même geste. » (PN05) - « Une souris serait plus pratique et plus rapide. » (PN10) - « La souris aurait été préférable. » (PN14) - « Ce serait plus facile avec une souris car j'ai plus l'habitude et ce serait plus adapté à mes doigts. » (PA08) - « Ce serait plus rapide avec une souris. » (PA09) - « Il faut se familiariser avec les touches et le tactile. » (PA14)
Commentaires positifs sur le pavé tactile	<ul style="list-style-type: none"> - « On s'habitue vite au pavé tactile. » (PN08) - « Une fois que je suis habituée, je n'ai pas de problèmes pour cliquer. » (PN11)
Commentaires négatifs sur la répétitivité	<ul style="list-style-type: none"> - « C'est répétitif, c'est un peu barbant. » (PA02) - « La méthode est répétitive. » (PA18)
Commentaire négatif sur les boutons	<ul style="list-style-type: none"> - « Le bouton 'Ok' est trop petit. » (PA07)

E.11 Grille d'observation pour l'écoute des patients par les auditeurs

CODE AUDITEUR :
A. Phase de lecture des consignes
Heure de départ de la lecture des consignes (hh:mm:ss) :
(Questions posées par l'auditeur, remarques) :
Heure de fin de lecture des consignes (hh:mm:ss) :
B. Phase d'entraînement
Heure de départ de l'entraînement (hh:mm:ss) :
B.1 Difficultés observées par rapport aux images
<input type="checkbox"/> L'auditeur ne parvient pas à identifier les images suivantes à cause de leur taille :
<input type="checkbox"/> L'auditeur ne parvient pas à identifier les images suivantes, pour d'autres raisons (spécifier) :
<i>(la grille continue à la page suivante)</i>

Grille d'observation (suite)

L'auditeur a des difficultés à situer les images les unes par rapport aux autres (droite, gauche, haut, bas) – spécifier la raison :

B.2 Autres difficultés observées (questions posées par l'auditeur, remarques)

Heure de fin de l'entraînement (hh:mm:ss) :

C. Phase d'écoute

Heure de départ de la session d'écoute (hh:mm:ss) :

C.1 Difficultés observées propres à l'environnement

a. Environnement périphérique (haut-parleurs, souris, écran) :

b. Environnement logiciel (boutons, ergonomie) :

(la grille continue à la page suivante)

Grille d'observation (suite)

C.2 Difficultés observées par rapport aux images

- L'auditeur ne parvient pas à identifier les images suivantes à cause de leur taille :

- L'auditeur ne parvient pas à identifier les images suivantes, pour d'autres raisons (spécifier) :

- L'auditeur a des difficultés à situer les images les unes par rapport aux autres (droite, gauche, haut, bas) – spécifier la raison :

C.3 Autres difficultés observées (remarques)

(la grille continue à la page suivante)

E.11. Grille d'observation pour l'écoute des patients par les auditeurs

Grille d'observation (suite)

Heure de fin de la session d'écoute (hh:mm:ss) :
--

E.12 Consigne donnée aux auditeurs

Cette étude a pour objectif de tester le logiciel EloKanz, un outil conçu pour la prise en charge de personnes souffrant de troubles pathologiques de production de la parole (TPPP). Le logiciel EloKanz permet d'enregistrer des phrases prononcées par des patients souffrant de TPPP, et ensuite d'évaluer si ces phrases sont bien comprises par des auditeurs. Cette étude vise plus précisément à vérifier que le logiciel EloKanz est utilisable en milieu clinique, autrement dit que son ergonomie ne pose pas de difficulté majeure aux patients et au personnel médical.

Le principe du logiciel est le suivant. Sur l'écran que vous avez en face de vous, sont représentés des objets.

[L'investigateur lance l'écran où sont présentés les objets]

Via ces deux haut-parleurs (ce casque ???), vous allez entendre des phrases enregistrées par un / des (???) patient(s). Vous n'allez entendre chaque phrase qu'une seule fois. Chacune de ces phrases correspond à un ordre vous demandant de déplacer tel ou tel objet à tel endroit. C'est-à-dire que le patient va vous demander de déplacer l'un des objets à droite, à gauche, au-dessus ou en dessous d'un autre objet. Seules ces quatre positions sont possibles.

[L'investigateur donne un exemple d'énoncé plausible par rapport aux images présentées à l'écran : Mettez X sur/sous/à droite/à gauche de Y]

Pour l'exemple, nous allons écouter ensemble la première phrase. Pour déclencher la diffusion de la phrase, vous devez cliquer sur « OK ».

[L'investigateur choisit la réponse « OK » dans la boîte de dialogue « Êtes-vous prêt(e) ? »]

[L'investigateur et l'auditeur écoutent la phrase]

Si vous n'avez rien compris à ce qui a été dit par le patient, vous allez directement cliquer sur le bouton « valider », sans effectuer aucune action. À l'opposé, si vous avez bien compris la phrase prononcée par le patient, vous devez alors y répondre en effectuant l'action correspondante.

[L'investigateur invite l'auditeur à sélectionner une image, et à la déplacer en haut, en bas, à gauche ou à droite d'un second objet]

Après chaque déplacement, les objets ne sont plus mobiles. Si vous vous êtes trompé(e), si par exemple vous avez fait un faux mouvement ou que vous changez d'avis, il vous faudra alors cliquer sur le bouton « effacer ».

[L'investigateur invite l'auditeur à cliquer sur le bouton « effacer »]

Vous pouvez alors recommencer un déplacement.

Si jamais vous n'avez compris la phrase que partiellement, c'est-à-dire que vous n'avez entendu que l'objet à déplacer, ou la position, ou l'objet cible, alors il faut que vous utilisiez ces informations :

- si vous n'avez entendu que le premier objet, alors vous le sélectionnez et vous le déplacez n'importe où à côté de n'importe quel autre objet ;*
- si vous n'avez entendu que la position (par ex. « ...à droite de... »), alors vous sélectionnez n'importe quel objet et vous le déplacez à droite de n'importe quel autre objet ;*
- si vous n'avez entendu que le second objet, alors vous sélectionnez n'importe quel objet et vous le déplacez n'importe où à côté de cet objet.*

Par contre, comme je vous l'ai dit, si jamais vous n'avez rien compris du tout, alors vous validez directement.

[L'investigateur s'assure que l'auditeur n'a pas de questions]

Nous allons d'abord faire une phase d'entraînement, pendant laquelle on va faire une dizaine d'essais pour bien préciser ce que vous devez faire, et aussi pour voir si vous n'avez pas de problème avec le logiciel. Après cet entraînement, vous écouterez une cinquantaine de phrases.

[L'investigateur demande à l'auditeur s'il est prêt à commencer l'entraînement]

E.13 Sujets auditeurs inclus dans l'étude

N°	Code	Profession	Âge	Années d'exposition aux TPPP
1	AE01	Infirmière	54	32
2	AE02	Diététicienne	31	7
3	AE03	Aide-soignante	50	15
4	AE04	Médecin phoniatre	49	20
5	AE05	Interne en médecine	32	1
6	AE07	Aide-soignante	38	6
7	AE08	Aide-soignante	48	13
8	AE09	Ergothérapeute	51	27
9	AE10	Médecin phoniatre	52	25
10	AE11	Aide-soignante	55	2
11	AE12	Étudiante infirmière	19	1
12	AE13	Orthophoniste	56	35
13	AN01	Étudiante en 1 ^{ère} année d'orthophonie	33	0
14	AN02	Étudiante en 1 ^{ère} année d'orthophonie	26	0
15	AN03	Étudiante en 1 ^{ère} année d'orthophonie	22	0
16	AN04	Étudiante en 1 ^{ère} année d'orthophonie	19	0
17	AN05	Étudiante en 1 ^{ère} année d'orthophonie	23	0
18	AN06	Étudiante en 1 ^{ère} année d'orthophonie	22	0
19	AN07	Étudiante en 1 ^{ère} année d'orthophonie	19	0
20	AN08	Étudiante en 1 ^{ère} année d'orthophonie	19	0
21	AN09	Étudiante en 1 ^{ère} année d'orthophonie	19	0
22	AN10	Étudiante en 1 ^{ère} année d'orthophonie	25	0
23	AN11	Étudiante en 1 ^{ère} année d'orthophonie	21	0
24	AN12	Étudiante en 1 ^{ère} année d'orthophonie	21	0

E.14 Questions et remarques des auditeurs liées aux consignes

Moment de la passation	Transcription des questions et remarques
Pendant la phase d'explication des consignes	<ul style="list-style-type: none"> - « Et après, je ne pourrai plus vous poser de questions ? » (AE07) - « Je peux réécouter ? » (AE11) - « Je fais par rapport à ma droite à moi ? » (AE13) - « À quoi sert EloKanz ? » (AE13) - « On a un temps limite pour chaque phrase ? » (AN01) - « Il faut inverser la droite et la gauche ? » (AN05) - « On a le temps ou il faut aller vite ? » (AN06) - « Il faut bouger plusieurs objets dans la même animation ? » (AN06) - « Je peux pas réentendre la phrase ? » (AN10)
Pendant la phase d'entraînement	<ul style="list-style-type: none"> - « Ça commence sérieusement ou c'est toujours l'entraînement ? » (AE01) - « Je ne sais pas trop me situer dans l'autre position. » (AE03) - « Il y a un temps limite ? » (AN05) - « Je ne pourrais pas mettre l'avion à droite de l'avocat s'il me demande de mettre l'avocat à gauche de l'avion ? » (AN12) - « Ça ne me dit pas si c'est juste ou faux ? » (AN05)
Pendant la phase d'écoute	<ul style="list-style-type: none"> - « J'ai compris que <i>grenouille</i>, je clique sur 'Valider' ? » (AE03) - « J'ai rien compris là, je fais quoi ? » (AE03) - « C'est toujours les mêmes patients ? » (AE05) - « Combien il y a de patients ? » (AE05) - « Quand on n'a rien compris, c'est 'Effacer' ? Euh non, 'Valider' ? » (AE05) - « Je prends n'importe quel objet. » (AE08) - « C'est bien ça que je dois faire si je n'ai compris que la position ? » (AE09) - « On peut réécouter ? » (AE10) - « C'est bien ça que je dois faire si je n'ai compris que la position ? » (AE10) - « C'est bien ça que je dois faire si je n'ai compris qu'une partie de la phrase ? » (AE11) - « Pour refaire le déplacement, je clique sur 'Effacer' c'est bien ça ? » (AN10)
Questionnaire post-test	<ul style="list-style-type: none"> - « Au début, je croyais que <i>sur</i> signifiait <i>par-dessus</i>. » (AN03)

E.15 Questions et remarques des auditeurs à propos de l'ergonomie

Moment de la passation	Transcription des questions et remarques
Pendant la phase d'explication des consignes	<ul style="list-style-type: none"> - « L'ordinateur, ce n'est pas trop mon truc. » (AE03) - « Là ça marche si j'appuie ? » (AE04, en désignant le bouton 'Effacer' inactif) - « On déplace en appuyant où ? » (AE08) - « Ce n'est pas de ma génération. » (AE08) - « Je n'ai pas d'ordinateur chez moi. » (AE11) - « Qu'est-ce que c'est ? » (AN01, à propos de la fourchette) - « Je ne suis pas habituée aux ordinateurs portables. » (AN02)
Pendant la phase d'entraînement	<ul style="list-style-type: none"> - « Je ne suis pas bien latéralisée. » (AE04) - « Il ne faut pas avoir un problème de droite et de gauche. » (AE07) - « C'est pas évident, c'est pas pratique, je vais perdre un temps fou à faire ça moi. » (AE11) - « Il ne faut pas que je confonde ma droite et ma gauche. » (AN10)
Pendant la phase d'écoute	<ul style="list-style-type: none"> - « C'est ça le poulet ? » (AE03) - « Est-ce que ça c'est un poulet ? » (AE05) - « Ça c'est bien un tampon ? » (AE05) - « C'est l'installation qui n'est pas très confortable. » (AE09) - « On est mal placé. » (AE11) - « Il faudrait une souris. » (AE11) - « C'est pas pratique, moi je n'y arrive pas, je fais comme je peux. » (AE11)
Questionnaire post-test	<ul style="list-style-type: none"> - « Certaines images sont trop pâles. » (AE01) - « L'image de la fourchette est trop petite. » (AE02) - « Les images sont colorées donc le logiciel est très bien. » (AE02) - « C'est plus amusant que d'écrire » (AE02) - « Il faudrait que la fourchette soit plus grosse. » (AE03) - « L'image du poulet serait mieux s'il était vivant. » (AE03) - « C'est difficile pour les gens mal latéralisés. » (AE04) - « Il faudrait que les cibles autour des objets soient permanentes. » (AE04) - « Le pavé tactile n'est pas assez efficace : il faut plusieurs mouvements pour y arriver. » (AE05) - « L'image du paon ne convient pas car la queue est trop grande. » (AE07) - « Les images sont bien. » (AE08) - « Il faudrait une vraie souris. » (AE09) - « On est assis trop loin. » (AE09) - « Les objets sont trop pâles. » (AE09) - « C'est un peu répétitif. » (AE09) - « Il faudrait un siège plus approprié. » (AE09) - « La cible est bien pour savoir où placer correctement l'objet. » (AE10) - « L'apprentissage est nécessaire pour la manipulation. » (AE10) - « Avec une souris ce serait plus simple. » (AE11) - « Ma pensée est plus rapide que mon geste à cause du pavé tactile. » (AE11)

Questions et remarques des auditeurs à propos de l'ergonomie (suite)

Questionnaire post-test (suite)	<ul style="list-style-type: none">- « La taille des images est un peu petite. » (AE11)- « La taille des images est bonne. » (AE12)- « Ce serait mieux d'appuyer sur 'Entrée' pour valider. » (AE13)- « Je préférerais utiliser le clavier plutôt que le pavé tactile. » (AN03)- « C'est comme un jeu. » (AN04)- « C'est répétitif. » (AN05)- « La taille des images est bien. » (AN06)- « Ce serait plus rapide et plus pratique d'appuyer sur 'Entrée' pour déclencher la voix. » (AN07)- « Je préférerais une souris. » (AN08)- « Les images sont bien. » (AN10)
---------------------------------	---