

■ Intelligibilité et Appréciation des Synthèses Vocales en Français

■ Intelligibility and Subjective Ratings of French Voice Synthesizers

Natacha Trudeau, Ève Chaput, Ann Sutton, Elsa Chan et Resi Contardo

Abrégé

Les synthèses vocales en français se multiplient et sont plus fréquemment recommandées aux personnes ne pouvant communiquer par la parole naturelle. Les intervenants disposent de très peu de données objectives pour les guider au moment de sélectionner une synthèse vocale pour un client francophone. Cette étude a pour but de mesurer l'intelligibilité de sept voix (six synthèses vocales françaises et une voix naturelle numérisée), et d'évaluer l'appréciation subjective que des locuteurs du français en ont. Vingt-cinq participants ont été répartis en deux groupes sur la base de leur expérience antérieure avec des appareils de communication avec sortie vocale (15 novices et 10 initiés). Leur tâche consistait à écrire des mots produits par les différentes voix dans deux conditions (Mots isolés et Mots produits à la fin d'un énoncé) et de donner une appréciation subjective de la voix entendue. Au niveau de l'intelligibilité, le degré d'expérience avec les synthèses vocales n'a pas affecté la performance des participants. De façon générale, l'intelligibilité varie d'une voix à l'autre et la présence d'un contexte linguistique facilite significativement la reconnaissance des mots. De plus, une interaction entre le contexte linguistique et la voix permet de constater que la présence d'un énoncé porteur facilite la reconnaissance de mots de façon différente selon la voix, ramenant l'intelligibilité de trois synthèses vocales au même niveau que celle de la voix naturelle numérisée. Au niveau de l'appréciation, les participants ont indiqué des préférences, peu importe leur degré de familiarité avec les synthèses vocales. De plus, bien qu'une relation s'observe entre l'intelligibilité et l'appréciation, les résultats montrent qu'à intelligibilité égale, l'appréciation subjective varie, confirmant l'importance de prendre en considération les deux types de variables lors d'une recommandation en clinique.

Mots clés: suppléance à la communication orale, synthèse vocale, intelligibilité, appréciation

Abstract

French voice synthesizers are now more common and more frequently recommended for individuals who are unable to communicate through natural speech, but clinicians lack objective data to guide them in the choice of a voice synthesizer for a French-speaking client. The purpose of this study was to measure the intelligibility of seven voices (six French synthesized voices and a digitized natural voice), and to assess the subjective ratings that speakers of French gave to these voices. Twenty-five participants were divided in two groups on the basis of their past experience with voice output communication aids (15 novice and 10 familiar). The participants were asked to write words produced by the various voices under two conditions (words alone and words produced at the end of a sentence). In addition, when a word was produced at the end of a sentence, the participants were asked to give a subjective rating of the voice. Results showed that the degree of previous exposure to voice output communication aids did not affect the performance of the participants. General trends revealed that some voices were more intelligible than others, and that the presence of a linguistic context significantly facilitated the recognition of the words. Moreover, an interaction between the linguistic context and voice revealed that the impact of the linguistic context is different between the voices, increasing the intelligibility of some synthesized voices to the same level as that of the digitized voice. Participants also demonstrated preferences for certain voices. Moreover, although a link between intelligibility and subjective ratings was observed, results showed that among voices with similar intelligibility levels, subjective preferences were still present, confirming the importance of taking into account the two types of variables when making a recommendation.

Key Words: Augmentative and alternative communication, voice synthesizer, intelligibility, appreciation

**Natacha Trudeau, Ph.D.,
Ève Chaput, B.Sc., et Ann
Sutton, Ph.D.**

*Centre de Recherche de
Hôpital Sainte-Justine
École d'orthophonie et
d'audiologie,
Université de Montréal
Montréal, Québec Canada*

**Elsa Chan, M.Sc. (A.) et
Resi Contardo, M.A.**

*Centre de réadaptation
Marie-Enfant de Hôpital
Sainte-Justine
Montréal, Québec Canada*

Introduction

Les personnes souffrant de troubles sévères de la communication sont parfois incapables d'utiliser la parole naturelle pour s'exprimer. Dans ces cas, il n'est pas rare qu'un système de suppléance à la communication orale (SCO) soit mis en place pour compenser ou remplacer la parole (Beukelman et Mirenda, 1998; von Tetzchner et Martinsen, 2000). La technologie ayant évolué de façon marquée au cours des dernières années, les possibilités en SCO sont de plus en plus nombreuses, que ce soit en termes d'accès (accès direct, accès par interrupteur), de stockage (capacité de mémoire, encodage des contenus), ou de sortie (visuelle ou auditive). L'une des caractéristiques importantes d'un système de SCO concerne la possibilité de produire le message oralement (sortie vocale), permettant au partenaire de communication d'entendre le message émis par l'utilisateur de l'appareil. Le choix de la voix devrait faire l'objet d'une démarche rigoureuse afin d'assurer une efficacité optimale de la communication. Cette démarche est d'autant plus importante qu'il est, de nos jours, pratique courante de recommander des appareils de SCO offrant une sortie vocale (Beukelman et Mirenda, 1998).

Afin de comparer les différents produits sur le marché, plusieurs chercheurs se sont intéressés à la performance des différentes voix disponibles en anglais. Les études menées sur les synthèses vocales produisant l'anglais se sont principalement penchées sur deux aspects pour comparer les différentes voix, soit l'intelligibilité et l'appréciation subjective des participants. En outre, plusieurs travaux ont exploré les différentes variables extrinsèques à la voix comme telle (variables liées aux participants ou aux stimuli et à leur mode de présentation) pouvant influencer l'évaluation des voix (intelligibilité et appréciation).

Intelligibilité

Pour étudier l'intelligibilité des voix, la plupart des chercheurs demandent aux participants de répéter ou d'écrire ce qu'ils ont entendu (mots ou phrases), ou de choisir, parmi un nombre plus ou moins restreint de choix, celui qui correspond à ce que l'appareil a produit. D'autres, moins nombreux, demandent aux participants de juger si l'énoncé produit est vrai ou faux, ou encore de produire un résumé d'un paragraphe entendu, pour ainsi juger la compréhensibilité des messages transmis par la voix. Dans tous ces types de conditions, la majorité sinon la totalité des études portant sur les synthèses anglaises indiquent que la voix naturelle est plus intelligible que toutes les synthèses vocales (Delogu et coll., 1998; Drager et Reichle, 2001a; Kangas et Allen, 1990; Koul et Allen, 1993; Massey, 1988; Mitchell et Atkins, 1988; Paris et coll., 2000; Reynolds et Fucci, 1998; Reynolds et Givens, 2001; Reynolds et coll., 2000). Par contre, parmi les différentes synthèses vocales anglaises, l'intelligibilité varie grandement d'une synthèse à l'autre (Higgingbotham

et coll., 1994; Hoover et coll., 1987; Hustad et coll., 1998; Koul et Hanners, 1997; McNaughton et coll., 1994; Mirenda et Beukelman, 1987; Mirenda et Beukelman, 1990; Mitchell et Atkins, 1988; Paris et coll., 2000; Raghavendra et Allen, 1993; Rupperecht et coll., 1995; Scherz et Beer, 1995).

Préférence subjective des participants

Certains chercheurs se sont plutôt intéressés à l'appréciation subjective que les participants avaient des différents types de voix. Dans ces études, on demandait aux participants de choisir, entre plusieurs voix, celle qu'ils préféreraient dans différentes mises en situation (Crabtree et coll., 1990; Paris et coll., 2000; Ratcliff et coll., 2002; Sutton et coll., 1995). Le résultat principal est convergent avec les données sur l'intelligibilité et concerne la nette préférence que les participants manifestent à l'égard de la voix naturelle (Crabtree et coll., 1990; Paris et coll., 2000). Par ailleurs, selon la situation présentée, les participants ne choisissaient pas toujours, parmi les différentes synthèses vocales, la même voix comme étant leur préférée.

Variables influençant les performances

La constatation que les préférences d'un même individu peuvent varier en fonction de la situation soulève d'ailleurs la question de facteurs extrinsèques à la synthèse vocale proprement dite et qui pourraient en affecter l'intelligibilité ou l'appréciation : les caractéristiques de la personne qui participe à l'étude, les caractéristiques intrinsèques au stimulus et les caractéristiques du contexte de présentation du stimulus. Parmi les caractéristiques personnelles de participants ayant été étudiées, un impact de l'âge sur l'intelligibilité et l'appréciation des voix synthétiques a été observé. Ainsi, les voix synthétiques sont moins intelligibles pour les enfants que les adultes (McNaughton et coll., 1994; Mirenda et Beukelman, 1987; Mirenda et Beukelman, 1990). De plus, lorsque des enfants doivent choisir une voix pour une situation donnée, ils ne choisissent pas toujours la même voix que des adultes, bien que dans l'ensemble leurs choix soient comparables (Crabtree et coll., 1990). Les aînés réussissent aussi bien que les jeunes adultes à des tâches d'intelligibilité des voix synthétiques (Drager et Reichle., 2001a; Drager et Reichle., 2001b; Humes et coll., 1991; Humes et coll., 1993). De plus, dans une tâche d'évaluation subjective, les adultes plus âgés montrent un niveau de confort différent de celui des jeunes adultes en fonction du débit de la voix synthétique, ayant un seuil de tolérance plus bas à l'égard des débits dits rapides. Ainsi, l'âge semble avoir un impact sur des mesures objectives (intelligibilité) et subjectives (préférence et confort).

Un autre facteur personnel ayant fait l'objet de quelques travaux est le degré d'exposition ou de pratique avec les synthèses vocales. De façon générale, il semble qu'une exposition répétée à des stimuli générés par une synthèse vocale résulte en une amélioration des performances dans des tâches d'intelligibilité, bien que celle-ci ne s'observe pas avec toutes les synthèses vocales

(Delogu et coll., 1998; Hoover et coll., 1987; Koul et Hanners, 1997; McNaughton et coll., 1994; Reynolds et coll., 2000; Venkatagiri, 1994). Au-delà d'un effet de pratique immédiat, il se pourrait aussi que les gens exposés à une synthèse vocale de façon courante (ceux qu'on pourrait qualifier d'experts) obtiennent de meilleures performances dans des tâches d'intelligibilité (Hustad et coll., 1998).

Contrairement à l'âge et au degré d'exposition, le genre des participants a été relativement peu étudié comme facteur pouvant influencer l'intelligibilité des voix synthétiques et aucun effet significatif n'a été observé (Scherz et Beer, 1995). Par contre, dans une tâche de jugement subjectif (attribution d'une voix selon une description du locuteur utilisant la synthèse), les hommes et les femmes, bien que largement semblables dans leurs jugements, différaient sur quelques points (Crabtree et coll., 1990).

Outre les facteurs personnels, certains facteurs liés aux stimuli utilisés peuvent aussi influencer l'intelligibilité des synthèses vocales. D'abord, la plupart des auteurs s'entendent pour dire que la présence d'un contexte linguistique facilite la reconnaissance du stimulus par les participants. Les mots isolés sont plus difficiles à reconnaître que les mots présentés dans une phrase porteuse (Mirenda et Beukelman, 1987), surtout si cette phrase offre une forte possibilité de prévoir le mot final (Hoover et coll., 1987). De façon similaire, des phrases précédées d'un paragraphe offrant un contexte d'interprétation étaient plus faciles à répéter que des phrases présentées isolément (Drager et Reichle, 2001b). Finalement, lorsque les stimuli sont des paragraphes, leur longueur et leur complexité influencent leur intelligibilité (Higgingbotham et coll., 1994).

Buts de l'étude

La recherche sur les synthèses vocales anglaises révèle donc qu'à ce jour, la voix humaine enregistrée demeure la plus intelligible, la plus efficace pour transmettre un message, et celle que les interlocuteurs préfèrent, la jugeant plus naturelle. Par ailleurs, si une synthèse vocale doit être recommandée, il paraît évident que tous les produits sur le marché ne présentent pas les mêmes niveaux de performance, et que cette performance peut être affectée par une quantité de variables. Il serait tout à fait légitime de s'attendre à observer les mêmes phénomènes dans l'évaluation des synthèses vocales françaises. Cependant, aucune étude sur le sujet n'est disponible actuellement. Bien que les études menées dans des milieux anglophones nous donnent des indices à propos des facteurs importants faisant varier l'intelligibilité des synthèses vocales, leurs résultats ne sont pas directement applicables au français parce que les deux langues n'ont pas les mêmes structures phonologiques, syllabiques et prosodiques. De plus, les produits disponibles en français étant différents, les données sur la supériorité de certaines synthèses en anglais n'informent pas les cliniciens souhaitant formuler

une recommandation à cet égard à leurs clients francophones.

L'objectif de la présente étude est donc d'évaluer l'intelligibilité et l'appréciation de synthèses vocales françaises par des participants sans troubles de la communication. Les facteurs du contexte linguistique (mots isolés vs. phrases) et de l'expérience des participants (novices vs. initiés) ont été manipulés et les autres facteurs susceptibles d'influencer les résultats ont été maintenus constants. Les résultats fourniront de l'information objective (par des mesures de reconnaissance des mots cibles) et subjective (par une échelle de l'appréciation de la qualité des synthèses vocales) aux utilisateurs des aides techniques à la communication, aux intervenants et aux instances administratives quant au choix d'appareils incorporant une synthèse vocale.

Hypothèses

Sur la base des études réalisées en anglais, certaines hypothèses ont pu être formulées. Tout d'abord, les voix devraient différer quant à leur intelligibilité et leur appréciation. Plus spécifiquement, il est prévu que la voix naturelle sera plus intelligible et plus appréciée par les participants. De même, parmi les voix synthétiques, il est probable que certaines seront mieux comprises et plus appréciées par les participants. Par contre, puisque aucune étude de ce genre n'a été réalisée en français jusqu'à présent, il est impossible de poser des hypothèses spécifiques quant à la synthèse vocale la plus intelligible ou la plus appréciée. Ensuite, compte tenu de l'effet facilitant du contexte linguistique dans des tâches d'intelligibilité en anglais, les mots en contexte (présentés à la fin d'une phrase) devraient être mieux identifiés que les mots isolés. Finalement, les participants ayant eu une exposition préalable aux synthèses vocales devraient obtenir une performance supérieure que les novices à la tâche d'intelligibilité.

Une association positive entre les deux variables à l'étude, l'intelligibilité et le jugement des participants, est prévue (i.e. les gens devraient préférer des voix qu'ils arrivent à mieux comprendre). Par contre, à intelligibilité égale, les participants pourraient néanmoins manifester une préférence pour certaines voix, se basant sur des caractéristiques indépendantes de l'intelligibilité proprement dite.

Méthodologie

Participants

Vingt-cinq adultes francophones ont participé à l'étude, soit 15 *novices* (12 femmes et 3 hommes), rapportant ne pas avoir eu d'expérience avec des personnes qui communiquent à l'aide de synthèses vocales et 10 *initiés* (9 femmes et 1 homme) travaillant avec des synthèses vocales depuis plus d'un an. Les *novices* étaient âgés de 19 à 58 ans (moyenne de 32 ans), et les *initiés* de 24 à 48 ans (moyenne de 40 ans) Tous les participants étaient de langue maternelle française et

ont rapporté ne pas avoir de problème de la parole, du langage, de l'audition, de troubles psychiatriques ou neurologiques et avoir au moins complété l'école secondaire régulière (secondaire V). Tous les participants ont été soumis à un dépistage audiolinguistique sous écouteurs à 500Hz, 1kHz, 2kHz et 4kHz. Tous, sauf un, ont obtenu des seuils de moins de 25dB à au moins 3 des quatre fréquences testées. Trois participants dans chacun des groupes ont obtenu un seuil entre 30 et 50dB pour une seule fréquence. Étant donné la répartition de ces participants entre les deux groupes, ceux-ci ont tout de même participé aux tâches expérimentales¹.

Synthèses vocales sélectionnées

Trois synthèses vocales ont été sélectionnées, après consultation auprès de professionnels oeuvrant dans une équipe spécialisée en SCO en milieu pédiatrique. Pour chacune des synthèses, une voix masculine et une voix féminine ont été retenues : 1) L & H : Pierre et Véronique (LHM et LHF), 2) Digalo : Robert et Cathy (DM et DF) et 3) Via Voice IBM : Jacques et Jacqueline (VVM et VVF). L'inclusion de voix masculine et féminine d'une même synthèse avait pour but de comparer l'intelligibilité et l'appréciation au sein d'un même produit selon les caractéristiques intrinsèques de la synthèse. Cette comparaison est cliniquement pertinente puisque chaque patient désirera une voix de son sexe. De plus, une voix naturelle (homme adulte de 25 ans ayant le français québécois comme langue maternelle) a été incluse. Ainsi, sept voix différentes ont été présentées aux participants.

Tâche d'intelligibilité

Dans la tâche d'intelligibilité, les participants devaient identifier des mots produits par des synthèses vocales.

Stimuli. Puisque aucune liste de stimuli n'avait été développée pour ce genre d'étude en français, une liste de 112 mots a été créée (voir annexe 1 pour la liste de mots). Tous les mots étaient des mots de la langue française, monosyllabiques de structure consonne-voyelle-consonne (CVC), soit des noms communs, des adjectifs ou des verbes à l'infinitif ou conjugués. Afin d'équilibrer phonétiquement cette liste, toutes les consonnes du français ont été utilisées le même nombre de fois (7) en début et en fin de mot (sauf le /l/, présent 6 fois et le /s/, 8 fois). Chaque mot de la liste a également été placé à la fin d'une courte phrase (5 à 7 mots). Le mot cible à reconnaître était donc présenté dans deux conditions, soit isolément (Mots isolés) et à la fin d'une phrase (Mots en contexte). Les listes de mots et de phrases développées pour cette tâche ont été divisées en sept blocs qui ont par la suite été assignés à chacune des voix pour présentation.

Trois versions du test ont été construites en faisant varier l'attribution des voix aux différents blocs de mots ou de phrases. De cette façon, à chaque version, chaque voix prononçait des stimuli différents. Ainsi, chaque

participant, quelle que soit la version qui lui a été attribuée, a entendu les mêmes sept blocs de mots et de phrases. Ces blocs ont cependant été produits par des synthèses différentes selon la version du test que le participant recevait. De plus, les stimuli composant les blocs étaient présentés dans un ordre aléatoire, ce qui avait pour but de contrôler les effets possibles d'ordre et de fatigue. La répartition phonétique entre les blocs était telle que chaque participant a entendu chaque voix produire chaque consonne au moins une fois au début et une fois à la fin d'un mot cible.

Matériel. Le montage des stimuli a été effectué à l'aide du logiciel « Écrire en symboles™ » (Mayer-Johnson Company Inc., 2002). Les synthèses vocales ont pu produire directement les stimuli via un ordinateur de table PC dans lequel ces dernières sont intégrées, au moyen d'une carte de son Crystal Sound Fusion. De cette façon, il n'y avait pas d'intermédiaire (pré-enregistrement) entre les synthèses vocales et les auditeurs. Pour les stimuli produits par la voix humaine, un enregistrement audio numérique a été réalisé via un microphone, directement sur l'ordinateur puis présenté sur le même ordinateur que les productions synthétiques. Chaque stimulus a été présenté à une intensité fixe sur des haut-parleurs Labtec en utilisant les paramètres octroyés par défaut par le logiciel pour chaque synthèse vocale. Cette méthode se rapproche donc d'une situation réelle de communication où la voix numérisée a été pré-enregistrée et la voix synthétique est générée directement par l'appareil de communication.

Procédure. Le participant a été informé que tous les mots cibles étaient des mots réels de la langue française. Deux cent vingt-quatre stimuli lui ont été présentés, soit 32 par synthèse vocale, dans un ordre aléatoire. À chaque essai, le participant devait écouter un stimulus à la fois en champ libre via les haut-parleurs de l'ordinateur et écrire le mot cible qu'il croyait avoir entendu, sur une feuille-réponse. Tous les participants étaient situés à la même distance des haut-parleurs, le volume étant également toujours le même.

Tâche d'appréciation

Cette tâche avait lieu de façon simultanée avec la tâche d'intelligibilité. Le participant devait donner une cote d'appréciation de la voix pour les mots en contexte (c'est-à-dire pour les courtes phrases). En effet, après une seule écoute de l'énoncé, le sujet écrivait le mot cible tout en donnant la cote d'appréciation.

Stimuli et matériel. Les stimuli étaient les mêmes phrases que celles utilisées pour la reconnaissance des mots en contexte. Le participant a donc eu à juger de la qualité des synthèses vocales pour 112 phrases (16 par voix).

Procédure. Une mise en situation était d'abord présentée : « Imaginez-vous que la voix entendue est la voix que votre frère ou votre sœur va devoir utiliser pour communiquer à la suite d'un accident ou d'une chirurgie. Aimeriez-vous qu'il ou elle ait cette voix? » Le jugement de

la qualité des synthèses vocales se faisait selon une échelle d'appréciation de 1 à 5, où 1 signifiait « pas du tout », 3 « peut-être » et 5 « beaucoup ». À chaque écoute de phrase, le participant donnait une cote sur une feuille réponse.

Déroulement général

Chaque participant a été accueilli dans un local tranquille mais où des sources naturelles de bruits ambiants étaient présentes (i.e. système de ventilation, ordinateur). Chacun y a complété un questionnaire de renseignements généraux (ex., langue maternelle, âge, éducation) avant de passer aux épreuves expérimentales. Le rythme de présentation des stimuli était contrôlé par le participant qui, après avoir écrit sa réponse, indiquait au chercheur qu'il était prêt à passer à l'essai suivant. Les deux tâches ont été réalisées par le participant en une seule rencontre qui durait de 90 à 120 minutes. Une ou deux pauses ont été prises durant le test, selon les besoins de chaque participant.

Traitement et analyses de données

Tâche d'intelligibilité. Les réponses de chaque participant ont été codées correctes ou erronées. La réponse a été considérée correcte si elle correspondait exactement à la forme phonétique de la cible. La fidélité inter-juge et intra-juge a été vérifiée sur 10% des données brutes. Le taux d'accord inter-juge et intra-juge était de 100%, aucune erreur de correction n'ayant été décelée.

Le total des réponses correctes pour chaque participant (dans chacune des conditions et pour chacune des voix) a été calculé. Une analyse de variance à mesures répétées avec un facteur inter-sujet (groupe) et deux facteurs intra-sujets (condition et voix) a été utilisée pour évaluer l'effet des variables groupe (*Novices* vs. *Initiés*), condition (mots isolés vs. mots en contexte) et voix (7 niveaux), et leur interaction sur le nombre de mots rapportés correctement.

Tâche d'appréciation. La cote moyenne accordée par chaque participant à chacune des voix a été calculée. Une analyse de variance à mesures répétées a été utilisée avec un facteur inter-sujet (groupe) à deux niveaux (*Novices* et *Initiés*), et un facteur intra-sujet (voix) à sept niveaux.

Résultats

Les résultats à la tâche d'intelligibilité révèlent l'absence de l'effet de groupe ($F(1,24)=2,85, p=0,10$), de même que de toute interaction de ce facteur avec les autres facteurs à l'étude (condition-groupe $F(1,24)=0,001, p=0,98$, voix-groupe $F(6,24)=0,95, p=0,49$ et condition-voix-groupe $F(6,24)=0,25, p=0,95$). Les *Novices* réussissent aussi bien que les *Initiés* à reconnaître les mots produits, peu importe la condition ou la voix. Par contre, il existe des effets significatifs de la condition ($F(1,24)=226,88, p<0,001$) et de la voix ($F(6,24)=35,33, p<0,001$). La performance est plus élevée au niveau des mots en contexte, avec un taux moyen de

bonnes réponses de 88% comparativement à 64% pour les mots isolés (voir tableau 1). De plus, l'effet de la voix s'observe pour les stimuli sans contexte (mots isolés), ($F(6,24)=35,91, p<0,001$), et les stimuli avec contexte (mots en fin de phrase), ($F(6,24)=17,55, p<0,001$). Une interaction voix-condition est également observée ($F(6,24)=15,21, p<0,001$). Une analyse post hoc (comparaisons par paire, $p<0,05$) révèle qu'au niveau des mots isolés, la voix la plus intelligible est la voix naturelle (85% de bonnes réponses). Les voix DM (74%), DF (69%) et LHM (76%) ont toutes un niveau d'intelligibilité statistiquement comparable et inférieur à celui de la voix naturelle. Les voix LHF (51%), VVM (49%) et VVF (44%), forment à leur tour un groupe ayant un niveau d'intelligibilité comparable, inférieur à celui des autres voix. Par contre, au niveau des mots à la fin d'une phrase, la voix naturelle (96%) et les voix DM (95%), DF (92%) et LHM (91%) sont statistiquement aussi intelligibles les unes que les autres. Les voix LHF (80%), VVM (81%) et VVF (78%) demeurent encore les moins intelligibles (voir figure 1).

Pour la tâche d'appréciation (voir tableau 2), l'effet de groupe n'est pas significatif ($F(1,24)=0,001, p=0,972$), ni l'interaction du facteur groupe avec le facteur voix ($F(6,24)=1,68, p=0,182$). Cependant, un effet de voix est observable ($F(6,24)=47,2, p<0,001$). Une analyse post hoc (comparaisons par paire, $p<0,05$) révèle que les résultats regroupent les voix en trois niveaux d'appréciation. La voix la plus appréciée est la voix naturelle avec une moyenne de 4,3 sur une échelle de 5. Les voix DM (2,9) et DF (3,1) viennent ensuite et les voix LHM (2,2), LHF (2,1), VVM (2,1) et VVF (2,2) obtiennent les moins bonnes évaluations de la part des participants (voir figure 2).

Discussion

Les résultats confirment nos hypothèses concernant les différences attendues en fonction du type de voix. Comme dans bon nombre d'études antérieures, un ordre hiérarchique entre les synthèses vocales et la voix naturelle a pu être établi (Sherz et Beer, 1995; Miranda et Beukelman, 1987, 1990; Fucci et coll. 1995; Hoover et coll., 1987, Higginbotham et coll. 1994). En effet, les participants de l'étude actuelle ont eu plus de succès dans la reconnaissance de stimuli produits par la voix naturelle. Par contre, il est intéressant de constater que même cette voix n'a pas entraîné une reconnaissance parfaite des mots (85% pour les mots isolés et 96% pour les mots en contexte). Il est possible qu'un léger bruit de fond attribuable au système de ventilation ait occasionné certaines confusions phonémiques, surtout lorsque les mots étaient présentés isolément. Par ailleurs, la présence de ce bruit ne remet pas en cause les résultats, puisqu'il était subtil, constant (i.e. le même pour tous les participants et pour tous les stimuli) et pourrait être considéré comme typique du bruit ambiant dans des contextes de conversation dits 'calmes'.

Tableau 1

Pourcentage moyen (%), et écart type des mots reconnus par les *Novices* ($n=15$) et les *Initiés* ($n=10$) et Rang pour chaque voix dans les conditions Mots isolés et Mots en contexte.

Voix		Condition							
		Mots isolés				Mots en contexte			
		Novices	Initiés	Total	Rang	Novices	Initiés	Total	Rang
Naturelle	%	85	85	85	1	95	96	96	1
	ET	10,8	7,9	9,6		6,4	6,6	6,4	
DM	%	75	74	74	2	95	95	95	1
	ET	14,3	13,8	13,8		4,0	3,9	3,9	
DF	%	70	68	69	2	90	96	93	1
	ET	19,5	18,8	18,8		8,5	4,3	7,4	
LHM	%	74	78	76	2	91	91	91	1
	ET	15,7	18,3	16,5		8,5	5,3	7,3	
LHF	%	47	57	51	3	77	85	80	2
	ET	14,1	11,6	13,9		9,6	8,4	9,9	
VVM	%	46	54	49	3	78	85	81	2
	ET	21,4	15,1	19,3		11,6	9,4	11,1	
VVF	%	41	48	44	3	75	82	78	2
	ET	16,3	11,5	14,7		10,4	9,5	10,4	

Note: ET : Écart-type ; Rang : groupements statistiques établis par une comparaison multiple, sur la base de différences significatives entre les différents scores obtenus ($p<0,05$); DM : Voix Robert de Digalo ; DF : Voix Cathy de Digalo ; LHM : Voix Pierre de L&H ; LHF : Voix Véronique de L&H ; VVM : Voix Jacques de Via Voice (IBM) ; VVF : Voix Jacqueline de Via Voice (IBM).

Les différentes synthèses vocales ont donné lieu à des performances différentes à la tâche de reconnaissance des mots. De façon générale, les voix produites par Digalo (DM et DF) étaient les mieux comprises, alors que celles produites par Via Voice (VVM et VVF) l'étaient moins. La différence d'intelligibilité entre la voix masculine et féminine proposées par L&H n'était pas attendue, mais montre l'importance de bien évaluer chaque voix de façon individuelle plutôt que d'inférer de la qualité d'un produit uniquement sur la base de l'évaluation d'une seule des voix offertes.

Les taux de réussite étaient relativement bas pour les synthèses vocales lorsque des mots isolés sont présentés (44% à 75%) par rapport au taux d'intelligibilité de la voix naturelle (85%). À la lumière de ces résultats, force est de constater que les synthèses vocales offrent des possibilités limitées dans des conditions où le contexte communicatif est artificiel et dénué de tout indice (linguistique, social, etc.).

Les différences entre les voix pour la tâche d'appréciation confirment aussi les hypothèses. Non seulement la voix naturelle était-elle la préférée, avec une cote d'appréciation moyenne de 4,3 sur 5, mais elle est en fait la seule voix à obtenir une cote nettement positive (i.e. supérieure à 4). Toutes les synthèses vocales ont obtenu des cotes beaucoup plus faibles. En effet, les voix de la synthèse Digalo obtiennent des cotes plutôt neutres (tout près de 3), alors que les voix de L&H et Via Voice obtiennent des cotes plutôt négatives (plus près de 2).

La seconde hypothèse principale concernait l'effet attendu du contexte linguistique sur la performance des participants. Un effet significatif du contexte a pu être observé au niveau de l'intelligibilité comme dans les études ayant comparé des mots isolés et des phrases en anglais (Miranda et Beukelman, 1987, 1990; Drager et Reichle, 2001b). Ainsi, peu importe la voix, les mots sont mieux compris lorsqu'ils sont précédés par une phrase.

Tableau 2

Cote d'appréciation moyenne (sur 5), Écart type (ET) et rang accordés à chaque voix par les Novices (n=15) et les Initiés (n=10).

Voix	Groupe				Total		Rang
	Novices		Initiés		M	ET	
	M	ET	M	ET			
Naturelle	4,2	0,6	4,3	0,9	4,3	0,7	1
DM	2,8	0,8	3	0,7	2,9	0,7	2
DF	3,1	0,5	3,1	0,6	3,1	0,6	2
LHM	2,3	0,7	2,1	0,4	2,2	0,6	3
LHF	2,1	0,4	2,2	0,4	2,1	0,4	3
VVM	2,2	0,6	2,0	0,4	2,1	0,5	3
VVF	2,3	0,5	2,1	0,4	2,2	0,4	3

Note. M : Moyenne ; ET : Écart-type ; Rang : groupements statistiques établis par une comparaison multiple, sur la base de différences significatives entre les différentes cotes accordées ($p < 0,05$); DM : Voix Robert de Digalo ; DF : Voix Cathy de Digalo ; LHM : Voix Pierre de L&H ; LHF : Voix Véronique de L&H ; VVM : Voix Jacques de Via Voice (IBM) ; VVF : Voix Jacqueline de Via Voice (IBM).

De plus, l'effet du contexte résulte en l'effacement de la différence d'intelligibilité entre la voix naturelle et trois des synthèses (DM, DF et LHM).

La présence d'un contexte semble donc être un facteur non négligeable dans l'intelligibilité des synthèses vocales. En fait, compte tenu du fait que la majorité des échanges se font en contexte (social, linguistique, etc.) il est remarquable que, même lorsque le contexte de présentation est très limité (i.e. phrase porteuse), trois synthèses sont alors aussi intelligibles que la voix naturelle (environ 95% de réussite). Ces résultats permettent un peu plus d'optimisme quant à l'utilisation de voix synthétisées dans des situations réelles.

La troisième hypothèse principale concernait la plus grande facilité avec laquelle des participants initiés aux synthèses vocales pourraient reconnaître les mots produits par celles-ci. L'effet de groupe escompté ne s'est pas manifesté, en contraste avec des études où la performance des sujets s'était vue améliorée après des séances d'exposition aux synthèses vocales (Delogu et coll., 1998; Reynolds et coll., 2000; McNaughton et coll., 1994; Venkatagiri, 1994). Il importe de souligner que l'effet d'exposition aux synthèses et la façon dont celle-ci a été contrôlée et évaluée dans cette étude ne sont pas tout à fait comparables avec les études antérieures. Tout d'abord, lorsque des différences ont été relevées entre des groupes de participants ayant une expérience différente avec les synthèses, les participants experts avaient une expérience spécifique avec les synthèses à l'étude, ce qui n'était pas le cas ici. Ceci soulève la question de la généralisation des habiletés acquises d'une synthèse à l'autre. De plus, les sujets

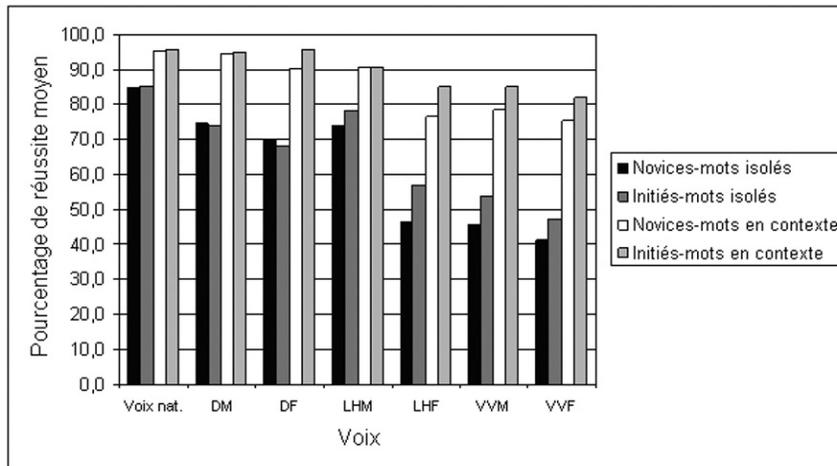


Figure 1. Pourcentages de bonnes réponses pour chaque voix selon le groupe et le contexte

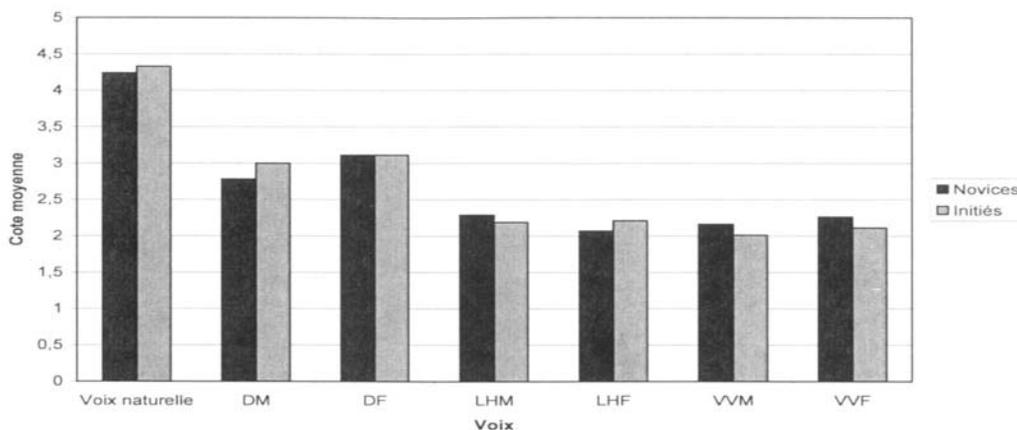


Figure 2. Évaluation subjective de chacune des voix selon le groupe

initiiés aux synthèses vocales n'avaient pas tous la même expérience avec les synthèses vocales et il est fort probable que la régularité de l'exposition est un facteur déterminant dans l'atteinte d'un statut d'expert. En outre, des travaux antérieurs (ex. : Venkatagiri, 1994) ont mis en évidence le fait que l'effet de pratique peut s'observer de façon très rapide en contexte d'exposition répétée aux voix synthétiques (i.e. durant les 20 ou même les 5 premiers essais). Les tâches utilisées reposant sur un nombre important de stimuli (224), il est possible que les novices aient rapidement rattrapé les initiés, d'autant plus que ceux-ci n'avaient pas tous une expérience aussi poussée. Le type d'analyse effectuée ne permet pas d'investiguer la présence d'une éventuelle différence de performance entre les groupes en fonction du nombre d'essais écoulés.

La dernière question à l'étude concernait le lien possible entre l'intelligibilité et l'appréciation subjective des différentes voix. Les résultats montrent que de façon générale, il existe une relation entre ces deux variables, bien qu'elle ne soit pas absolue. En effet, à intelligibilité égale, les participants ont tout de même préféré la voix naturelle à celles de DM, DF et LHM pour les mots présentés à la fin d'une phrase. Cet effet peut être attribué aux caractéristiques uniques de la voix humaine qui lui confèrent un aspect qu'aucune synthèse vocale n'arrive à reproduire à ce jour. Cependant, même au sein du groupe des synthèses vocales les plus intelligibles (DM, DF et LHM), les jugements subjectifs ne sont pas équivalents. La voix LHM, bien qu'aussi intelligible que les deux autres, est jugée moins acceptable qu'elles. La différence d'intelligibilité entre la voix masculine et féminine de L&H ne se traduit donc pas par un écart dans l'appréciation subjective que les participants en font. Bien que la voix humaine soit en quelque sorte inimitable, quelques synthèses vocales semblent toutefois offrir un remplacement plus acceptable. Cette absence de corrélation parfaite entre l'intelligibilité et l'appréciation montre bien qu'il faut considérer ces deux types de variables lors d'une recommandation clinique à l'égard d'un système de SCO.

Retombées cliniques

Parmi les voix évaluées dans cette étude, la voix naturelle numérisée est sans contredit la plus intelligible au niveau des mots isolés ainsi que la plus appréciée sur le plan de la qualité. Il apparaît donc primordial que chaque appareil d'aide à la communication puisse permettre l'enregistrement de la voix naturelle, ne serait-ce que pour les formules de salutation et de politesse et les phrases les plus souvent utilisées dans le quotidien.

La présence d'un contexte linguistique semble quasi indispensable pour assurer une communication efficace. Le taux d'intelligibilité moyen passe globalement pour l'ensemble des voix de 64 % à 88 % en présence d'un contexte linguistique : la synthèse Digalo et la voix de masculine de la synthèse L&H sont statistiquement aussi intelligibles que la voix naturelle numérisée. Ces résultats

confirment l'intérêt d'ajouter un contexte linguistique au message transmis par la synthèse vocale. Cette constatation devrait pousser les cliniciens à encourager l'utilisation d'énoncés plutôt que de mots isolés par les utilisateurs de synthèses vocales. Par contre, comme la production d'énoncés peut être laborieuse pour des personnes ayant, par exemple, des difficultés motrices importantes, il serait essentiel d'exploiter d'autres caractéristiques des appareils de communication pour minimiser les demandes physiques tout en permettant la production d'énoncés. La programmation d'énoncés courants pouvant être produits sans être construits chaque fois par le locuteur serait une façon de gérer ce problème.

Études futures. L'étude actuelle se voulait une première exploration de l'intelligibilité des synthèses vocales francophones. À ce titre, seules quelques voix parmi celles disponibles ont été évaluées. Devant l'éventail grandissant de produits disponibles sur le marché francophone, il serait important, dans une étude future, d'inclure d'autres synthèses vocales pour parvenir à déterminer laquelle ou lesquelles des synthèses vocales doivent être recommandées. Ceci pourrait être fait en comparant seulement les synthèses vocales retenues par cette étude comme étant les plus compréhensibles (Robert, Cathy de Digalo et Pierre de L&H) avec de nouvelles synthèses, permettant ainsi de conserver un matériel de taille raisonnable.

Il serait aussi pertinent d'étudier la capacité à généraliser les acquis faits avec la pratique pour une synthèse à la compréhension d'autres synthèses. Ce type d'étude pourrait aider à mieux comprendre l'effet de pratique immédiat et l'effet de l'exposition régulière et continue aux synthèses vocales et leur interaction.

Conclusion

Cette étude a mis en évidence des profils de performance différents pour différentes synthèses vocales en français. Les progrès récents et continus dans le domaine de la voix artificielle rendent désormais nécessaires des études systématiques permettant aux utilisateurs de prendre des décisions éclairées lors du choix d'une synthèse vocale. En effet, les résultats de cette étude confirment que le choix d'une synthèse vocale est une démarche complexe, pour laquelle des facteurs liés à la qualité de la synthèse et aux préférences personnelles de l'utilisateur doivent être considérés. De plus, étant donné les caractéristiques phonétiques uniques de chaque langue et les variations régionales existant pour une langue donnée, il est essentiel de poursuivre le développement de produits qui répondront aux préférences des utilisateurs de façon accrue et de poursuivre les études sur la question. À ces caractéristiques régionales s'ajouteront également les caractéristiques personnelles des utilisateurs. En effet, la population des personnes utilisant des systèmes de SCO n'étant pas homogène, il est important de bien comprendre les variables qui influencent l'efficacité des synthèses vocales afin de

pouvoir faire des recommandations appropriées à la situation spécifique de chaque personne. Ainsi, bien que des études de groupe soient nécessaires à la démarche d'évaluation des produits disponibles, il n'en demeure pas moins que la démarche clinique devra toujours tenir compte d'une multitude de facteurs complexes afin d'appliquer judicieusement les résultats d'études comme celle-ci à la réalité individuelle des utilisateurs de SCO.

Références bibliographiques

- Beukelman, D. R. & Mirenda, P. (1998). *Augmentative and alternative communication: Management of severe communication disorders in children and adults* (2nd Ed.). Baltimore: Paul H. Brookes.
- Crabtree, M., Mirenda, P., & Beukelman, D. R. (1990). Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication, 6*, 256-261.
- Delogu, C., Conte, S., & Sementina, C. (1998). Cognitive factors in the evaluation of synthetic speech. *Speech Communication, 24*, 153-168.
- Drager, K. D. R., & Reichle, J. E. (2001a). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication, 17*, 109-119.
- Drager, K. D. R., & Reichle, J. E. (2001b). Effects of discourse context on the intelligibility of synthesized speech for young adults and older adult listeners: Applications for AAC. *Journal of Speech, Language, and Hearing Research, 44*, 1052-1057.
- Fucci, D., Reynolds, M. E., Bettagere, R., & Gonzales, M. D. (1995). Synthetic speech intelligibility under several experimental conditions. *Augmentative and Alternative Communication, 11*, 113-117.
- Higgingbotham, D. J., Drazek, A. L., Kowarsky, K., Scally, C. A., & Segal, E. (1994). Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication, 10*, 191-202.
- Hoover, J., Reichle, J. E., Van Tasell, D., & Cole, D. (1987). The intelligibility of synthesized speech: Echo II versus Votrax. *Journal of Speech and Hearing Research, 30*, 425-431.
- Humes, L. E., Nelson, K. J., & Pisoni, D. B. (1991). Recognition of synthetic speech by hearing-impaired elderly listeners. *Journal of Speech and Hearing Research, 34*, 1180-1184.
- Humes, L. E., Nelson, K. J., Pisoni, D. B., & Lively, S. E. (1993). Effects of age on serial recall of natural and synthetic speech. *Journal of Speech and Hearing Research, 36*, 634-639.
- Hustad, K. C., Kent, R. D., & Beukelman, D. R. (1998). DECTalk and MacinTalk speech synthesizers: Intelligibility differences for three listener groups. *Journal of Speech, Language, and Hearing Research, 41*, 744-752.
- Kangas, K., & Allen, G. D. (1990). Intelligibility of synthetic speech for normal-hearing and hearing impaired listeners. *Journal of Speech and Hearing Disorders, 55*, 751-755.
- Koul, R., & Allen, G. D. (1993). Segmental intelligibility and speech interference thresholds of high-quality synthetic speech in presence of noise. *Journal of Speech and Hearing Research, 36*, 790-798.
- Koul, R., & Hanners, J. (1997). Word identification and sentence verification of two synthetic speech systems by individuals with intellectual disabilities. *Augmentative and Alternative Communication, 13*, 99-107.
- Massey, H. J. (1988). Language-impaired children's comprehension of synthesized speech. *Language, Speech, and Hearing Services in Schools, 19*, 401-409.
- Mayer-Johnson Company Inc. (2002). *Écrire en symboles* (version française de Writing with Symbols), Auteur.
- McNaughton, D., Fallon, J., Tod, J., Weiner, F., & Neisworth, J. (1994). Effect of repeated listening experiences on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10*, 161-168.
- Mirenda, P., & Beukelman, D. R. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. *Augmentative and Alternative Communication, 3*, 120-128.
- Mirenda, P., & Beukelman, D. R. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. *Augmentative and Alternative Communication, 6*, 61-68.
- Mitchell, P. R., & Atkins, C. P. (1988). A comparison of the single word intelligibility of two voice output communication aids. *Augmentative and Alternative Communication, 4*, 84-88.
- Paris, C. R., Thomas, M. H., Gilson, R. D., & Kincaid, J. P. (2000). Linguistic cues and memory for synthetic and natural speech. *Human Factors, 42*, 421-431.
- Raghavendra, P., & Allen, G. D. (1993). Comprehension of synthetic speech with three text-to-speech systems using a sentence verification paradigm. *Augmentative and Alternative Communication, 9*, 126-133.
- Ratcliff, A., Coughlin, S., & Lehman, M. (2002). Factors influencing ratings of speech naturalness in augmentative and alternative communication. *Augmentative and Alternative Communication, 18*, 11-19.

Reynolds, M. E., & Fucci, D. (1998). Synthetic speech comprehension: A comparison of children with normal and impaired language skills. *Journal of Speech, Language, and Hearing Research, 41*, 458-466.

Reynolds, M. E., & Givens, J. (2001). Presentation rate in comprehension of natural and synthesized speech. *Perceptual and Motor Skills, 92*, 958-968.

Reynolds, M. E., Isaacs-Duvall, C., Sheward, B., & Rotter, M. (2000). Examination of the effects of listening practice on synthesized speech comprehension. *Augmentative and Alternative Communication, 16*, 250-259.

Rupprecht, S. L., Beukelman, D. R., & Vrtiska, H. (1995). Comparative intelligibility of five synthesized voices. *Augmentative and Alternative Communication, 11*, 244-247.

Scherz, J. W., & Beer, M. B. (1995). Factors affecting intelligibility of synthesized speech. *Augmentative and Alternative Communication, 11*, 74-78.

Sutton, B., King, J., Hux, K., & Beukelman, D. R. (1995). Younger and older adults' rate performance when listening to synthetic speech. *Augmentative and Alternative Communication, 11*, 147-153.

Venkatagiri, H. S. (1994). Effect of sentence length and exposure on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10*, 96-104.

von Tetzchner, S., & Martinsen, H. (2000). *Introduction to augmentative and alternative communication* (2nd Ed.). Philadelphia: Taylor and Francis.

Notes en bas de page

¹ Note: Des analyses statistiques (tests t) ont montré que les individus ayant obtenu des seuils au-dessus de 25 dB ne différaient des autres participants ($p \leq 0,05$) que pour l'une des 14 conditions à l'étude; ils ont reconnu moins de mots isolés produits par la voix naturelle.

Note de l'auteur

La correspondance concernant cet article devrait être adressée à Natacha Trudeau, Centre de recherche de l'Hôpital Sainte-Justine, Centre de réadaptation Marie-Enfant, 5200 Rue Bélanger Est, Montréal, Québec, Canada H1T 1C9. Courrier électronique: natacha.trudeau@umontreal.ca.

Date soumis: le 1 novembre 2004

Date accepté: le 5 octobre 2005



Annexe

Liste des mots et courtes phrases, par phonème initial

Phonème /p/

pape	À Rome j'ai vu le pape.
pub	Je travaille pour une compagnie de pub.
pif	Ça prend un super pif.
passé	Je te prête ma passe.
pause	Tu devrais prendre une pause.
page	Lisez les dix premières pages.
pomme	Ma grand-mère cuisine une tarte aux pommes.

Phonème /b/

bombe	Le village est détruit par une bombe.
baguette	Il lui a offert une baguette.
bœuf	J'ai commandé un hamburger au bœuf.
bave	Sa gueule est pleine de bave.
bosse	Il s'est fait une bosse.
basse	La table est très basse.

Phonème /t/

taupe	Il est aussi aveugle qu'une taupe.
tape	Il a reçu une tape.
tombe	Ne pleurez pas sur ma tombe.
tasse	J'en prendrais bien une tasse.
thèse	Je termine d'écrire ma thèse.
touche	J'ai appuyé sur cette touche.
toge	Le juge a revêtit sa toge.

Phonème /d/

dette	Tu ne pourras pas rembourser tes dettes.
date	Je ne connais pas la date.
dinde	On se réunit pour manger la dinde.
douce	Sa peau est si douce.
douche	Je vais prendre une douche.
dame	Cette toilette est réservée aux dames.
dard	Elle m'a piqué avec son dard.

Phonème /k/

cube	Tu dois prendre tous les cubes.
coude	Je me suis éraflé le coude.
cave	Il y en a dans la cave.
cause	Je ne connais pas la cause.
coche	Tu choisis et tu coches.
quiche	Aujourd'hui c'est de la quiche.
canne	Mon grand-père a besoin d'une canne.

Phonème /g/

guêpe	Elle m'a piqué la guêpe.
goutte	Je n'en veux pas une goutte.
guide	Nous devons suivre le guide.
gaffe	J'ai fait une grosse gaffe.
gauche	Tu dois tourner à gauche.
guerre	Ce fut une longue guerre.
gansé	La ceinture rentre pas dans la gansé.

Phonème /f/

fugue	Beaucoup d'enfants font des fugues.
figue	Il aime grignoter des figues.
femme	Il vit avec une seule femme.
faune	Notre devoir est de protéger la faune.
fer	Elle a une carence de fer.
foule	Je manque d'air dans une foule.
fil	Pour coudre il faut du fil.

Phonème /v/

vide	Il est tombé dans le vide.
vague	Il y a trop de vagues.
vif	Elle a l'esprit vif.
veuf	Hier il est devenu veuf.
vis	Ça tient pas il manque une vis.
vase	Mets les fleurs dans un vase.
vache	Le vétérinaire soigne aussi les vaches.

Annexe (suite)

Liste des mots et courtes phrases, par phonème initial

Phonème /s/		Phonème /m/	
sept	Je vais en prendre sept.	moppe	Le concierge passe la moppe.
sud	J'aime voyager dans le sud.	mise	J'ai perdu ma mise.
sec	L'air est beaucoup trop sec.	mouche	Il y a trop de mouches.
sève	Au printemps nous récoltons la sève.	mage	Ils étaient trois les rois mage.
sage	Il est bien trop sage.	mer	Regarde le bateau sur la mer.
somme	Il me doit une belle somme.	moule	Les gourmands ont mangé toutes les moules.
cenne	Je n'ai plus une cenne.	mule	Il transporte son butin sur la mule.
Phonème /z/		Phonème /n/	
zut	Je l'ai manqué zut.	note	Il a fait une fausse note.
zinc	Il fabrique des pièces coulées en zinc.	nymphé	Il a rêvé à une nymphe.
zouk	Veux-tu danser le zouk?	neuf	Je porte un habit neuf.
zoum	Avec ta caméra fait le zoom.	noce	J'étais présent à vos noces.
zone	Le taxi est pas dans sa zone.	neige	Regarde par la fenêtre il neige.
zen	Mon ami japonais est très zen.	nonne	Elle veut se faire nonne.
zèle	Cet employé fait preuve de zèle.	nerf	Elle me tombe sur les nerfs.
Phonème /-/		Phonème //	
chute	Il a fait une mauvaise chute.	lobe	Je me suis fait percer le lobe.
choc	C'est tout un choc.	lac	Veux-tu te baigner dans le lac?
chèque	Je peux payer par chèque?	ligue	Je fais partie de la ligue.
chaude	La soupe est trop chaude.	langue	Tu t'es fait percé la langue?
chaise	Assis-toi sur ta chaise.	loue	Dans cette forêt il y a une loue.
schème	Il a déduit un schème.	lave	Le village est enseveli sous la lave.
cher	Les enfants coûtent beaucoup trop cher.	loge	J'ai des billets dans une loge.
Phonème /zʃ/		Phonème /r/	
jupe	Elle est trop courte ta jupe.	rampe	Tiens-toi bien à la rampe.
jambe	Je rêve d'avoir des longues jambes.	robe	As-tu vu ma nouvelle robe?
gigue	Sais-tu danser la gigue?	ride	Ma mère n'a pas de rides.
juge	L'avocat parle au juge.	rêve	La nuit est-ce que tu rêves?
jaune	Tu prends le crayon jaune.	rive	J'habite sur l'autre rive.
jour	Dépêche-toi il fait déjà jour.	roc	C'est dur comme du roc.
gèle	À zéro degré Celcius l'eau gèle.	rime	Est-ce que ton poème rime?