MOTS CLÉS
TROUBLE SÉVÈRE DU LANGAGE
SUPPLÉANCE À LA COMMUNICATION ORALE
SYNTHÈSES VOCALES FRANCOPHONES
INTELLIGIBILITÉ
APPRÉCIATION
VOIX HUMAINE
ÂGE

Évaluation de neuf synthèses vocales françaises basée sur l'intelligibilité et l'appréciation

Assessment of nine French synthesized voices based on intelligibility and quality

Patricia Côté-Giroux
Natacha Trudeau
Christine Valiquette
Ann Sutton
Elsa Chan
Catherine Hébert

Abrégé

Introduction : Grâce à l'avancement de la technologie dans le domaine de la communication humaine, plusieurs synthèses de voix françaises ont vu le jour. Celles-ci sont de plus en plus recommandées par les spécialistes pour les personnes atteintes de troubles de la communication. L'objectif de ce projet est d'identifier les synthèses de voix francophones les plus intelligibles selon la condition de production (mots, phrases) et d'évaluer l'appréciation de ces différentes voix. Méthode : Soixante et un participants répartis en trois groupes d'âge (14-20 ans, n = 20; 21-40 ans, n = 20; 41-60 ans, n = 21) ont été recrutés. La tâche consistait 1) à identifier des mots (isolés et contenus dans des énoncés) produits par neuf synthèses vocales et par une voix humaine dans deux conditions (mots isolés et mots en contexte) et 2) à donner leur appréciation globale pour chaque synthèse vocale. Résultats : Les résultats des analyses statistiques démontrent qu'il n'y a pas d'effet de genre ou d'âge sur l'intelligibilité et l'appréciation des synthèses vocales. La performance est plus élevée pour les mots en contexte (90%) comparativement aux mots isolés (71%). De plus, les résultats révèlent que deux synthèses vocales en condition de mots isolés (> 84%) et cinq en condition de mots en contexte (> 92%) sont aussi intelligibles que la voix humaine. Une différence significative a été trouvée entre les niveaux d'appréciation attribués aux synthèses vocales. Il existe également une corrélation positive entre l'intelligibilité des productions de mots et l'appréciation subjective de ces productions. Conclusion : Cette étude met en évidence une hiérarchie de l'intelligibilité et du niveau d'appréciation des différentes synthèses vocales francophones permettant aux professionnels d'obtenir des balises objectives pouvant les guider lors de l'attribution de systèmes et logiciels de communication relatifs à chaque client.

Abstract

Introduction: Technological advancements in human communication have led to the development of several French synthesized voices, which specialists are recommending more and more often to people with communication disorders. This study aimed to determine which French synthesized voice was the most intelligible in various productions (words, sentences), and to assess people's ratings of these voices. Method: We recruited sixty-one participants and split them into three age groups (14-20 years, n = 20; 21-40 years, n = 20; 41-60 years, n = 21). The task consisted of 1) identifying words (in isolation and utterances) produced by nine synthesized voices, as well as words produced by one human voice (in isolation and in context); and 2) giving an overall rating to each synthesized voice. Results: Statistical analysis shows no effect of sex or age on intelligibility or voice rating. The best performance was noted with words in context (90%) as compared to isolated words (71%). In addition, results indicate that two synthesized voices producing words in isolation (> 84%) and five synthesized voices producing words in context (> 92%) were equally as intelligible as the human voice. We noted a significant difference between the rating levels given to the synthesized voices. There was also a positive correlation between the intelligibility of the produced words and the subjective ratings given to these productions. Conclusion: This study outlines a hierarchy in the intelligibility and rating levels of various French synthesized voices, which will give professionals objective benchmarks to guide decision-making when they recommend communication systems and software to their clients.

Patricia Côté-Giroux, M.Sc.[1]
Natacha Trudeau, Ph.D.[1]
Christine Valiquette, M.P.O.[2]
Ann Sutton, Ph.D.[1]
Elsa Chan, M.P.O.[2]
Catherine Hébert[1]

1. Centre de recherche du CHU Sainte-Justine, École d'orthophonie et d'audiologie, Université de Montréal, Montréal, Québec, Canada
2. Centre de réadaptation Marie-Enfant, CHU Sainte-Justine, Montréal, Québec, Canada

INTRODUCTION

Une synthèse vocale (ou voix synthétique) implique un processus informatique de composition sonore permettant la transformation d'un texte en voix artificielle (Dutoit, Couvreur, Malfrère, Pagel, & Ris, 2002). Plusieurs générations de synthèses vocales (TTS; text-to-speech synthesizer) ont vu le jour depuis quelques décennies (Klatt, 1987; Mirenda & Beukelman, 1990; Breen, 1992). La première, appelée « synthèse vocale par formant », fit son entrée dès 1965 et demeura populaire jusqu'au milieu des années 80. En s'appuyant sur des algorithmes, cette technique permet de générer un signal sonore synthétique à l'aide des caractéristiques spectrales d'un signal de parole naturel. La deuxième génération constituée de voix de synthèse semi-synthétiques fut développée afin d'entreposer de façon permanente des bribes de parole naturelle dans une mémoire informatique. Cette méthode, plus précisément appelée « synthèse vocale par diphones, ou modèle à concaténation », consiste à unir des segments élémentaires de parole naturelle afin de former n'importe quel énoncé synthétique voulu (Dutoit et coll., 2002).

On assiste aujourd'hui à l'émergence d'une nouvelle génération de synthèse vocale à diphones élaborée avec une technique de « sélection d'unités de parole dans une grande base de données » (Hunt & Black, 1996; Dutoit, 2002). Afin de représenter le plus fidèlement possible la coarticulation et la prosodie unique à chaque voix, l'échantillonnage d'une même unité phonétique se fait à partir de plusieurs enregistrements contenant cette unité.

L'utilisation des synthèses vocales demeure une application importante dans des appareils de communication pour les personnes ne pouvant pas communiquer par la parole naturelle à cause d'un trouble moteur (ex. la paralysie cérébrale) ou langagier. Il est souvent recommandé que les appareils de suppléance à la communication intègrent une synthèse vocale (voir Beukelman & Mirenda, 2005, pour un survol du domaine). Cette pratique permet à la personne de communiquer par la modalité orale rendant ainsi sa communication plus « naturelle ». Or, l'utilité de l'appareil intégrant une voix dépend en grande partie de la qualité de la synthèse vocale. La communication par le biais de l'appareil est plus efficace si la parole produite par la synthèse vocale est comprise par l'interlocuteur et encore plus si ce dernier trouve la voix agréable à écouter. Les personnes qui souhaitent se procurer une aide technique de suppléance à la communication (SC) incluant une synthèse vocale doivent choisir parmi un large éventail de technologies. L'actualisation des données sur l'intelligibilité et l'appréciation des voix synthétiques disponibles présentement sur le marché est nécessaire afin de guider les intervenants dans la recommandation et le choix d'une synthèse vocale.

Les mots ou énoncés produits en dehors de tout contexte sont intelligibles lorsque l'interlocuteur les identifie correctement. Plus spécifiquement, l'intelligibilité correspond à la façon plus ou moins appropriée (claire et accessible) dont le signal acoustique est transmis (Drager & Reichle, 2001). L'intelligibilité des voix de synthèse s'évalue en demandant aux participants de transcrire leurs réponses dans un formulaire (Pisoni, Nusbaum, & Greene,1985; Manous, Pisoni, Dedina, & Nusbaum, 1986; Crabtree, Mirenda, & Beukelman, 1990; Mirenda & Beukelman, 1990; Hustad, Kent, & Beukelman, 1998; Gong & Lai, 2001; Roring, Hines, & Charness, 2007), de répéter le stimulus entendu (Mirenda & Beukelman, 1990; Von Berg, Panorka, Uken, & Qeadan, 2009) ou de répondre à des questions précises concernant le stimulus présenté (Pisoni et coll., 1985; Drager & Reichle, 2001).

Pour ce qui est d'évaluer l'appréciation des voix de synthèse, on demande aux participants de juger le stimulus entendu selon une échelle d'appréciation (Nass & Lee, 2001; Ratcliff, Coughlin, & Lehman, 2002; Von Berg et coll., 2009). Toutefois, apprécier globalement une voix selon une échelle numérique pourrait ne pas refléter précisément la qualité de celle-ci.

Par ailleurs, le contexte dans lequel sont présentés les stimuli influence l'intelligibilité (Mirenda & Beukelman, 1987, 1990; Winters & Pisoni, 2003, 2004). En effet, l'auditeur compenserait une faible intelligibilité des synthèses vocales en utilisant les informations linguistiques supplémentaires fournies par le contexte, ce qui n'est pas le cas lorsque des mots isolés sont entendus. Plusieurs situations d'écoute ont été utilisées dans les recherches afin de mesurer l'impact de facteurs contextuels sur l'intelligibilité de la voix humaine et des voix de synthèses. Certaines expérimentations ont été effectuées en émettant des stimuli en présence de bruit ambiant (Drager et coll., 2007), en modifiant la longueur et la complexité des énoncés entendus (Higginbotham, Drazek, Kowarsky; Scally, & Segal, 1994; Venkatagiri, 1994), en misant sur la prévisibilité des phrases (revu par Drager & Reichle, 2001) ou encore en contrôlant le débit des stimuli (mots isolés ou phrases) (Higginbotham, 1994). Les résultats de ces études ont montré à quel point le contexte dans lequel les stimuli sont présentés influence les performances des auditeurs.

Des caractéristiques de l'interlocuteur peuvent aussi influencer l'intelligibilité d'une synthèse vocale entendue. L'âge en est une chez des auditeurs adultes. Toutefois, les résultats des études n'arrivent pas aux mêmes conclusions. D'une part, certaines recherches montrent que l'âge n'est pas un facteur déterminant pour une bonne identification

de stimuli produits synthétiquement (Mirenda & Beukelman, 1990; Humes, Nelson, Pisoni, & Lively, 1993) et d'autre part, il y a celles qui concluent que l'âge joue un rôle dans la perception auditive des synthèses vocales (Kangas et Allen, 1990; Roring et coll., 2007). Certains facteurs peuvent expliquer ces divergences. D'abord, la complexité des stimuli peut avoir entraîné ces différences dans le sens où certains mots, méconnus des participants moins âgés, peuvent avoir été jugés inintelligibles. D'autre part, l'écart d'âge au sein et entre chaque groupe d'âge peut influencer les résultats. En effet, les équipes ayant comparés des groupes plus étendus en âge (enfant, adolescents, jeunes adultes et adultes plus âgés) montrent généralement moins d'effet d'âge que ceux ayant comparé seulement des adultes plus ou moins jeunes. Ceci pouvant être expliqué par d'autres facteurs tels l'audition ou le milieu socio-économique qui viennent interagir avec les performances de ces adultes et qui ont moins d'impact chez les groupes de participants plus jeunes. Roring et coll. (2007) ont montré que le contexte a une influence sur les performances des participants en fonction de leur âge. Les jeunes adultes comprenaient mieux les voix de synthèse lors de l'écoute de mots isolés que les personnes âgées. Toutefois, cette différence de performances entre les groupes n'existait plus lorsqu'un contexte était fourni (Roring et coll., 2007). De plus, Kangas et Allen (1990) ainsi que Humes, Nelson et Pisoni (1991) rapportent qu'une perte auditive chez les adultes ou personnes âgées module la perception des synthèses vocales. Ces derniers ont rapporté que l'identification adéquate des stimuli par les personnes âgées est corrélée négativement à la perte auditive. De plus, selon l'équipe de Lai (2000), une meilleure identification des productions des voix de synthèse a été observée lorsqu'elles sont écoutées par des participants ayant atteint un plus haut niveau de scolarité. Toutefois, cet effet du niveau de scolarité pourrait être dû à la méthodologie utilisée. La moitié des participants devaient prendre des notes pendant l'écoute des voix de synthèse. Une prise de note efficace pouvait être influencée par un plus haut niveau de scolarité et pouvait par le fait même contribuer à une meilleure identification des productions des voix. L'expérience antérieure avec une synthèse vocale (degré d'exposition) ainsi que l'effet d'entraînement seraient des facteurs non négligeables puisque tous deux sont corrélés positivement avec une meilleure intelligibilité des productions synthétiques (Schwab, Eileen, Nusbaum, Howard, Pisoni & David, 1985; Koul, 2003; Lai, Wood, & Considine, 2000). Finalement, le genre des participants ne serait pas associé à l'identification correcte des stimuli produits par les synthèses vocales (Gong & Lai 2001; Ellis, Spiegel, & Benjamin, 2002; Roring et coll., 2007).

Certaines études démontrent que les synthèses vocales anglaises sont moins intelligibles que la voix humaine (Mirenda & Beukelman, 1987; Koul & Allen, 1993). Cela peut être dû au fait que les synthèses vocales étudiées étaient d'anciennes générations et, par conséquent, moins intelligibles que celles qui se retrouvent maintenant sur le marché. De plus, comme les indices prosodiques de ces voix de synthèse étaient parfois absents ou peu naturels, il est possible que l'écoute de ces synthèses vocales comparée à celle de la voix humaine exigeait davantage d'attention de la part des participants. La prosodie du discours est un élément important permettant une meilleure identification des productions d'une synthèse vocale (Schroder, 2001). Elle réfère à la modulation de la hauteur (fréquence fondamentale) et de l'intensité de la voix, aux pauses, silences et hésitations de la parole ainsi qu'à la durée syllabique (Bourhis, 2010).

La majorité des études sur l'intelligibilité des synthèses vocales portent sur la langue anglaise. Toutefois, une étude de Trudeau, Chaput, Sutton, Chan et Contardo (2006) a évalué l'intelligibilité des synthèses vocales françaises utilisées à cette époque. Ces auteurs ont demandé aux participants d'écrire le mot-cible après l'avoir écouté dans deux conditions de présentation : sans contexte (mots isolés) et avec contexte (mots en fin de phrase). Le nombre de mots correctement écrits était calculé pour les deux conditions. Ces auteurs ont constaté que le nombre de bonnes réponses était plus élevé pour la voix humaine (85% - mots isolés et 96% - mots en contexte) par rapport aux voix synthétiques Pierre de L&H (76-91%), Robert (74-95%) et Cathy (69-92%) de Digalo (les trois voix les plus intelligibles). Les autres synthèses vocales à l'étude avaient une moyenne de bonnes réponses de moins de 51% pour les mots isolés et de 81% pour les mots en contexte. L'avancée technologique ayant favorisé l'émergence de nouvelles voix de synthèse, les résultats de cette étude sont maintenant jugés désuets. Les nouvelles voix de synthèse disponibles sur le marché doivent être évaluées afin de promouvoir leur utilisation auprès de la clientèle atteinte de déficience motrice ou langagière.

L'identification adéquate des mots prononcés par une synthèse vocale (son intelligibilité) ne reflète en rien l'appréciation de la voix par les interlocuteurs. En effet, une voix de synthèse peut être identifiée correctement sans pour autant être naturelle ou agréable à entendre. L'appréciation est une notion subjective dont la définition varie. Pour certains chercheurs, il s'agit du naturel d'une voix et de l'attraction qu'elle exerce sur l'interlocuteur (Nusbaum, Francis & Henly, 1995; Paris, Thomas, Gilson & Kincaid, 2000. Pour d'autres, elle se définit par la qualité des contours mélodiques du discours produit (Terken, 1993; Winters & Pisoni, 2004). Dans les études portant sur l'appréciation des synthèses vocales, les participants

devaient juger subjectivement chaque voix et ce, dans différentes conditions de présentation des stimuli : voyelles (Nusbaum et coll., 1995), mots isolés (Humes et coll., 1993), mots en contexte (Trudeau, Chaput, Sutton, Chan, & Contardo, 2006) ou en paragraphe (Crabtree et coll., 1990; Nass & Lee, 2001; Ratcliff et coll., 2002; Von Berg et coll., 2009). Les résultats de ces études convergent avec les résultats des études sur l'intelligibilité et montrent que les participants préfèrent la voix humaine aux synthèses vocales. Les travaux réalisés sur l'appréciation des voix artificielles mettent de l'avant l'importance des indices suprasegmentaux dans l'évaluation subjective des voix (Nusbaum et coll., 1995; Paris et coll., 2000).

En ce qui concerne l'appréciation des synthèses vocales françaises dans l'étude de Trudeau et coll. (2006), les participants devaient donner une cote globale représentant leur appréciation de la voix après l'écoute de chaque stimulus. Les voix artificielles se sont révélées peu appréciées des participants : tandis que la voix humaine obtenait une cote d'appréciation moyenne de 4,3 sur 5, les six synthèses vocales obtenaient des cotes beaucoup plus faibles (< 3,1). À intelligibilité semblable, une synthèse vocale peut être préférée à une autre, il est donc important de prendre aussi en considération l'appréciation lors du choix d'un appareil.

LA PRÉSENTE ÉTUDE

Bien que les études menées sur les synthèses vocales anglophones fournissent des indices à propos de leur intelligibilité et de leur appréciation, ces résultats ne sont pas directement applicables aux synthèses vocales francophones. Les deux langues ne partagent pas les mêmes structures phonologiques, syllabiques et prosodiques. De plus, puisque les produits diffèrent d'une langue à l'autre, l'utilité des données sur la qualité des voix de synthèse en anglais est faible pour les cliniciens francophones souhaitant recommander une synthèse vocale à leurs clients. L'étude de Trudeau et coll. (2006) peut servir de modèle. Pour tenir compte de facteurs pouvant influencer l'intelligibilité tels que l'âge et l'acuité auditive, les participants ont été répartis selon trois groupes d'âge (14-19 ans, 20-39 ans, 40-60 ans), équilibrés pour le genre et n'ayant pas de trouble d'audition. Les stimuli ont été présentés dans deux conditions d'écoute (mots isolés et mots en fin de phrase). De plus, nous avons bonifié la tâche d'appréciation de Trudeau et coll. (2006) pour rendre plus naturel le contexte fourni (texte continu versus phrase simple) et avons demandé aux participants d'élaborer leur appréciation en choisissant des qualificatifs pour chaque synthèse vocale, en plus de donner une cote globale.

L'objectif de la présente étude est d'identifier quelles synthèses de voix francophones (françaises et québécoises) féminines ou masculines sont les plus intelligibles et les plus appréciées. Nous croyons que l'intelligibilité varie d'une voix à l'autre, que la présence d'un contexte linguistique facilite l'identification des mots et que les participants apprécieront les voix qui sont plus intelligibles (ils devraient préférer des voix qu'ils arrivent à mieux saisir). Par contre, à intelligibilité égale, l'appréciation pourrait varier d'une voix à l'autre.

MÉTHODES

Participants

Soixante et un participants répartis en trois groupes d'âge (A : 14-19 ans, moyenne=16 ans, n=20; B : 20-39 ans, moyenne=29, n=20; C : 40-60 ans, moyenne=51, n=21), dont 25 hommes et 36 femmes ont été recrutés par des affiches posées dans des endroits publics sur un campus universitaire. Les critères d'inclusion étaient a) avoir comme langue d'usage le français; b) ne pas avoir de trouble de langage ou d'audition et c) ne pas avoir d'expérience avec des appareils de communication incorporant une synthèse vocale. Cinquante huit des soixante-et-un participants avaient comme langue maternelle le français et tous avaient le français comme langue d'usage. Afin d'évaluer le deuxième critère, le participant était soumis à un dépistage auditif sous écouteurs à 500, 1000, 2000 et 3000 Hz. Le critère d'exclusion était un seuil supérieur à 25 dB pour une fréquence aux deux oreilles. Deux participants ont obtenu des seuils entre 30 et 50 dB pour deux et trois fréquences à une oreille. Ils ont tout de même participé à l'expérimentation, puisque celle-ci se déroulait en champ libre.

Tâche d'intelligibilité

Stimuli et conditions. Les stimuli étaient les 112 mots utilisés dans l'étude de Trudeau et coll. (2006). Il s'agissait de noms communs monosyllabiques, comportant les 16 consonnes de la langue française, en position initiale et finale de mot (pour consulter la liste des stimuli, voir Trudeau et coll., 2006). Comme le nombre de synthèses incluses dans l'étude actuelle (10) est supérieure au nombre de synthèses utilisées par Trudeau et coll. (7), 48 mots on dû être répétés afin que chaque synthèse produise 16 mots (ce qui permet la production de chaque consonne en position finale et initiale de mot). Compte tenu du fait qu'il n'y avait pas d'effet de mot dans l'étude antérieure, les mots répétés ont été choisis aléatoirement et les blocs de 16 mots ont été équilibrés phonétiquement. De cette façon, toutes les consonnes du français ont été utilisées le même nombre de fois en début et en fin de mot au sein d'un même bloc. Chacune des 10 voix produisaient 16 mots isolés et 16 mots mis en contexte pour un total de

320 stimuli. La liste complète de stimuli dans les deux conditions était présentée dans un ordre aléatoire afin d'éviter un effet d'entraînement lors des séances. De plus, dix versions équivalentes de la liste ont été conçues et réparties également au sein des 61 participants afin que tous les mots puissent être dits par chaque synthèse vocale.

Matériel et équipement

Neuf voix synthétiques ainsi qu'une voix humaine québécoise ont été utilisées dans cette étude (Tableau 1) : Bruno et Louise de Acapela, Juliette de AT&T Labs Inc., Pierre de L&H, Charlotte et Olivier de Loquendo, Félix, Virginie et Sophie de Nuance RealSpeak. Les compagnies développant ces synthèses vocales sont des chefs de file européens et américains dans le domaine des communications, tant au plan médical qu'au plan informatique. Les voix de synthèse ont été choisies en fonction de leur fréquente utilisation dans le domaine des troubles du langage, de leur coût varié, de leurs différents accents (québécois et français) et de leur genre (voix d'homme ou de femme). Toutes les voix de synthèse évaluées dans cette étude sont des voix élaborées à partir d'une sélection d'unités de parole dans une grande base de données, sauf celle de Pierre. Bien que cette dernière appartienne à une génération plus ancienne de synthèses vocales, elle a été reprise dans cette étude puisqu'elle a été cotée comme étant la plus intelligible et la plus appréciée dans l'étude de Trudeau et coll. (2006). Chaque item (mot isolé ou mot dans une phrase), programmé pour être dit par chaque voix, a été enregistré dans le logiciel Goldwave (version 5.51). Ensuite, les items ont été entrés dans le logiciel SD Pro (version 6.1) ou dans le logiciel The Grid (version 2.4). Les listes de 320 stimuli ont été conçues et lus par le logiciel Windows Media Player. Les paramètres par défaut (volume et débit) de chaque synthèse vocale ont été choisis pour l'enregistrement. Toutes les manipulations reliées au projet ont été effectuées avec un ordinateur portable Toshiba Notebook.

Procédure

Après une familiarisation à la tâche, pour chaque essai, le participant écoutait le stimulus présenté en champ libre à travers les haut-parleurs de l'ordinateur et inscrivait sur la feuille réponse le mot (condition du mot isolé) ou le dernier mot de la phrase (condition du mot en contexte). L'expérimentatrice contrôlait la présentation des stimuli

Tableau 1.

Caractéristiques des synthèses vocales incluses dans l'étude

Voix/nom de la synthèse vocale	Caractéristiques		
	Compagnie	Genre	Dialecte
Humaine	N/A	M	Québécois
Louise	Acapela	F	Québécois
Virginie	Nuance (RealSpeak)	F	Français
Sophie	Nuance (RealSpeak)	F	Français
Bruno	Acapela	M	Français
Olivier	Loquendo	M	Québécois
Juliette	AT&T Labs Inc.	F	Français
Charlotte	Loquendo	F	Québécois
Félix	Nuance (RealSpeak)	M	Québécois
Pierre	L&H	M	Français

et attendait que le participant lui indique qu'il était prêt avant d'envoyer chaque stimulus.

Tâche d'appréciation

Stimuli. Quatre paragraphes inspirés de Chesneau (2007) ont été créés pour la tâche d'appréciation. Ils ont été choisis pour représenter le plus fidèlement possible un contexte naturel d'écoute entre le participant et un utilisateur d'aide technique de suppléance à la communication. Ils contenaient cinq ou six phrases pour un total de 75 à 90 mots et avaient une structure similaire (Annexe A).

Matériel et équipement. Les voix synthétiques et humaine de la tâche d'intelligibilité ont été utilisées pour la tâche d'appréciation. Chaque paragraphe, d'une durée approximative d'une minute, a été produit par chacune des 10 voix. Dix versions différentes du test ont été conçues de manière à varier l'ordre de présentation des voix et les paragraphes attribués à chaque voix. Chaque paragraphe devait être présenté au moins une fois et au plus trois fois dans chaque version du test.

Procédure. Une mise en situation a été présentée au participant : «*Imaginez-vous que la voix entendue est la voix que votre frère ou votre sœur va devoir utiliser à la suite d'un accident ou d'une chirurgie. Sur une échelle de 1 à 7, à quel point aimeriez-vous qu'il ou elle ait cette voix? 1= pas du tout, 4= moyennement et 7= beaucoup.*» Ensuite le participant écoutait le paragraphe et jugeait la voix de deux façons : premièrement en attribuant une cote globale d'appréciation sur 7 et deuxièmement, en qualifiant la voix parmi des adjectifs bipolaires tels que chaleureuse ou froide, dure ou douce, monotone ou expressive et fluide ou saccadée sur une feuille réponse préparée préalablement à cet effet.

Procédure générale

L'approbation du comité d'éthique du CHU Sainte-Justine a été obtenue avant le recrutement des participants. Tous les participants ont été vus individuellement ou en groupe de deux, dans un local isolé et calme, mais où les sources de bruits de fond étaient tout de même présentes (i.e. ventilation, corridor adjacent, ordinateur). La séance a été menée par une expérimentatrice ayant reçu une formation afin d'administrer correctement le protocole. Dès son arrivée, le participant remplissait le formulaire d'information et de consentement ainsi que deux questionnaires d'informations générales. Les participants de moins de 18 ans devaient avoir obtenu le consentement d'un parent pour participer au projet. À la suite du dépistage auditif, la tâche d'intelligibilité a été administrée, puis celle d'appréciation. Cet ordre a été choisi afin d'éviter que l'écoute des voix dans la tâche d'appréciation précède celle de l'intelligibilité, ce qui aurait pu aider à reconnaître plus clairement les voix de synthèse, contribuant ainsi à augmenter le pourcentage de bonnes réponses. Le participant a été muni de feuilles réponses préalablement conçues pour chaque tâche. À la suite de la séance, une compensation de 20$ a été remise au participant afin de couvrir les frais de déplacement.

Analyses statistiques

Intelligibilité. Les réponses correctes de chaque participant ont été codées avec un score de 1 et les réponses erronées avec un score de 0. La réponse a été considérée correcte si elle correspondait parfaitement à la cible phonétique entendue. Les fautes d'orthographe n'étaient pas considérées comme des erreurs. Le total des réponses correctes pour chaque participant (dans chaque condition et pour chacune des dix voix) a été calculé. Une analyse de variance à mesures répétées avec deux facteurs inter-sujet (groupe et genre) et deux facteurs intra-sujet (conditions et voix) a été effectuée afin d'évaluer l'effet de ces variables et leur interaction sur l'intelligibilité des mots et énoncés présentés.

Appréciation. Une analyse de variance à mesures répétées utilisant les mêmes facteurs inter-sujets et un seul facteur intra-sujet (voix) a été effectuée afin de comparer la cote moyenne accordée par les participants à chaque voix. De plus, des tests Khi carré ont permis d'étudier le lien entre le niveau d'appréciation et les quatre caractéristiques pouvant être attribuées aux voix. Ensuite, une analyse de corrélation bivariée a permis d'explorer la relation entre l'intelligibilité et l'appréciation des synthèses vocales.

RÉSULTATS

Les données brutes pour chaque groupe, condition et tâche sont présentées à l'annexe B. Les analyses concernant l'intelligibilité montrent une absence d'effet d'âge et de genre et un effet significatif de voix, de condition ainsi qu'une interaction entre ces deux paramètres (Figure 1 et Tableau 2).

En général, les mots en contexte sont plus intelligibles que les mots isolés (90 versus 71%). Des comparaisons par paires révèlent des regroupements de voix selon leur intelligibilité dans les deux conditions pour les scores moyens (maximum = 16). Sans contexte, les différences permettent de regrouper les voix en cinq sous-groupes, des plus intelligibles au moins intelligibles: la voix humaine (14,31) et Louise (14,25) sont significativement plus intelligibles que toutes les autres voix sauf la voix Virginie (13,43); les voix Virginie, Sophie (12,51), Bruno (12,39) et Olivier (12,15) sont significativement plus intelligibles que les voix Juliette (11,18), Charlotte (10,48), Félix (8,56) et Pierre (3,95); les voix Juliette et Charlotte sont significativement plus intelligibles que les voix Félix et

Tableau 2.

Résultats de l'ANOVA pour les données sur l'intelligibilité des synthèses vocales

Variables	F	ddl	ddl erreur	p
Âge	0,36	2	46	0,69
Genre	0,11	1	46	0,74
Voix	129,03	9	414	< 0,001
Condition	525,7	1	46	< 0,001
Voix X Condition	23,05	9	38	< 0,001
Voix X Âge	0,598	18	78	0,89
Voix X Genre	1,07	9	38	0,406
Condition X Âge	0,283	2	46	0,755
Condition X Genre	0,065	1	46	0,80
Voix X Âge X Genre	0,981	18	78	0,489
Voix X Âge X Condition	0,844	18	78	0,645
Voix X Genre X Condition	0,364	9	38	0,945
Condition X Âge X Genre	0,465	2	46	0,631
Voix X Contexte X Âge X Genre	0,635	18	78	0,861

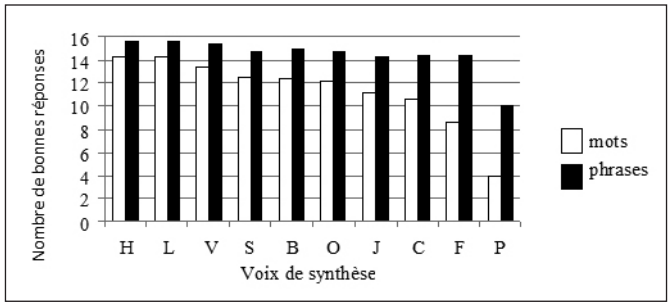

Figure 1.
Intelligibilité. Nombre moyen de bonnes réponses (maximum = 16) pour chaque voix.
Note. H = voix humaine; L = Louise; V = Virginie; S = Sophie; B = Bruno; O = Olivier; J = Juliette; C = Charlotte; F = Félix; P = Pierre

Pierre; et la voix Félix est significativement plus intelligible que la voix Pierre.

Avec contexte, les différences d'intelligibilité permettent de regrouper les voix en trois sous-groupes : la voix humaine (15,52) et les voix Louise (15,57), Virginie (15,31), Sophie (14,74) et Bruno (14,93) sont significativement plus intelligibles que toutes les autres voix sauf la voix Olivier (14,75); et les voix Olivier, Juliette (14,23), Charlotte (14,39) et Félix (14,38) sont significativement plus intelligibles que la voix Pierre (9,98). Dans les deux conditions, il y a une voix dont l'intelligibilité chevauche deux sous-groupes. Dans la condition sans contexte, la voix Virginie n'est pas significativement différente des voix les plus intelligibles (la voix humaine et Louise) ni des voix Sophie, Bruno et Olivier. Dans la condition avec contexte, la voix Olivier n'est pas significativement différente de toutes les autres voix sauf de la voix Pierre. L'apport du contexte est plus prononcé pour les voix les moins intelligibles. Une corrélation négative très forte (r = -0,94, p < 0,001) est observée entre l'augmentation du score moyen entre les deux conditions (i.e., la différence des deux scores) et le score moyen sans contexte.

Tableau 3.

Résultats de l'ANOVA pour les données sur l'appréciation des synthèses vocales

Variables	F	ddl	ddl erreur	p
Âge	1,73	2	54	0,187
Genre	0,36	2	54	0,550
Voix	62,28	9	486	< 0,001
Voix X Âge	0,968	18	76	0,504
Voix X Genre	1,516	9	37	0,179
Voix X Âge X Genre	1,105	18	76	0,421
Âge X Genre	0,595	2	54	0,556

En ce qui concerne l'appréciation, l'analyse montre l'absence d'effet d'âge et de genre ainsi qu'un effet significatif de voix (Figure 2 et Tableau 3).

Des comparaisons par paires démontrent des différences significatives des cotes moyennes (maximum 7) qui permettent de former six sous-groupes selon les résultats relatifs à l'appréciation : la voix humaine (5,74) est significativement plus appréciée que toutes les autres voix sauf la voix Virginie (5,65); la voix Virginie est

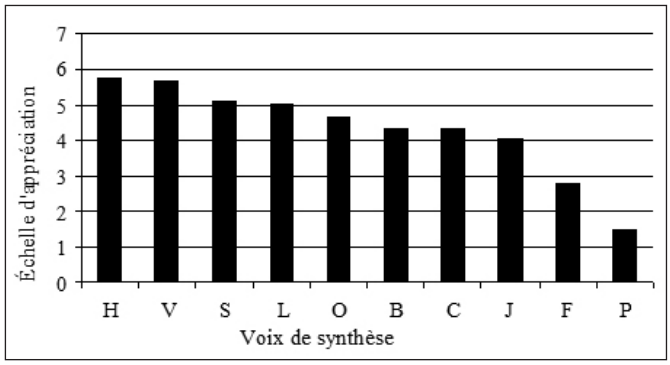

Figure 2.
Appréciation. Cote moyenne pour chaque voix (maximum = 7).
Note. H = voix humaine; V = Virginie; S = Sophie; L = Louise; O = Olivier; B = Bruno; C = Charlotte; J = Juliette; F = Félix; P = Pierre

significativement plus appréciée que toutes les autres voix sauf la voix Sophie (5,10); les voix Sophie et Louise (5,04) sont significativement plus appréciée que les voix Bruno (4,34), Charlotte (4,33), Juliette (4,07), Félix (2,81) et Pierre (1,48) mais pas la voix Olivier (4,64); les voix Olivier, Bruno, Charlotte et Juliette sont significativement préférées aux voix Félix et Pierre; et la voix Félix est significativement préférée à la voix Pierre (1,48). Il y a donc trois voix (Virginie, Sophie, et Olivier) dont l'appréciation chevauche les cotes de plus d'un sous-groupe.

Nous avons ensuite exploré la relation existant entre l'appréciation subjective globale et les caractéristiques des voix. Pour ce faire, les sept échelons d'appréciation ont été regroupés en trois niveaux : appréciation faible (cotes 1, 2), moyenne (cotes 3, 4, 5) et forte (cotes 6, 7). Un test de Khi carré s'est révélé significatif pour chacune des caractéristiques évaluées : chaleureuse ou froide [$\chi^2(2) = 128,47, p < 0,001$]; dure ou douce [$\chi^2(2) = 104,78, p < 0,001$]; monotone ou expressive [$\chi^2(2) = 119,49, p < 0,001$]; fluide ou saccadée [$\chi^2(2) = 108,49, p < 0,001$]. Ainsi, les voix peu appréciées sont jugées froides (88,5%), monotones (87,5%), saccadées (72,1%) et dures (68,3%). Les voix les plus appréciées sont douces (91,3%), fluides (85%), chaleureuses (81,5%) et expressives (79,2%) (Tableau 2).

L'analyse de corrélation bivariée de Pearson montre une corrélation positive modérée significative entre l'intelligibilité d'une voix et son appréciation : plus la voix est intelligible, plus elle est appréciée ($r = 0,429, p < 0,01$)

DISCUSSION

L'intelligibilité varie selon la voix et la condition de présentation des stimuli. Certaines synthèses vocales sont aussi intelligibles que la voix humaine et ce, dans les deux conditions d'écoute (mots présentés isolément et en contexte). En effet, les voix Virginie et Louise sont aussi intelligibles que la voix humaine avec et sans contexte. Toutefois, bien que d'autres synthèses vocales ne soient pas aussi intelligibles que la voix humaine, le contexte augmente grandement leur intelligibilité. Avec contexte, trois voix sont devenues aussi intelligibles que la voix humaine, et le score moyen d'une seule voix (Pierre) était

Tableau 4.

Distribution des jugements (en pourcentage) des caractéristiques selon le niveau d'appréciation attribué à la voix par chaque participant

Caractéristiques	Très appréciée	Appréciée	Peu appréciée
Chaleureuse/	81,5	51,8	11,5
Froide	18,5	48,2	88,5
Dure/	8,7	37,3	68,3
Douce	91,3	62,7	31,7
Monotone/	20,8	53,9	87,5
Expressive	79,2	46,1	12,5
Fluide/	85,0	43,8	27,9
Saccadé	15,0	56,2	71,1

en bas du taux d'intelligibilité de 75 %.

Le résultat montrant que certaines voix, même sans contexte, sont aussi intelligibles que la voix humaine ne concorde pas avec les résultats des études antérieures (Nye & Gaitenby, 1973; Mirenda & Beukelman, 1987, 1990; Koul & Allen, 1993; Trudeau et coll., 2006). Trudeau et coll. (2006) ont observé un effacement de la différence d'intelligibilité de la voix humaine (96%) et la voix Pierre (91%), mais seulement pour la condition avec contexte. De plus, ils ont trouvé un taux d'intelligibilité pour les mots isolés avec la voix Pierre de 76%. Ces pourcentages concernant la voix Pierre sont nettement supérieurs à ceux de notre étude (24%, n = 16, pour les mots isolés et 63%, n = 16, pour les mots en contexte). Cette différence d'intelligibilité pourrait être engendrée par les synthèses vocales avec lesquelles la voix Pierre a été comparée. En effet, toutes les voix de synthèses évaluées dans l'étude de 2006 étaient de l'ancienne génération donc étaient de qualité semblable ou inférieure à la voix Pierre. L'écoute des stimuli produits par la voix Pierre demandait alors moins d'effort attentionnel relativement aux autres synthèses présentées et provoquait par le fait même une augmentation de l'intelligibilité réelle de la voix Pierre. Dans l'étude actuelle, les autres voix de synthèse (de nouvelles générations) étaient plus faciles à comprendre que la synthèse Pierre, ce qui a pu influencer les attentes et le niveau d'attention des participants (entendre des voix très intelligibles), rendant l'écoute de la voix Pierre plus difficile. Une autre explication possible de ce résultat est que les conditions de présentation n'étaient pas les mêmes, soient la qualité des haut-parleurs, le niveau d'intensité et la vitesse de présentation des stimuli. Dans les deux études, les paramètres par défaut des logiciels et des synthèses vocales ont été utilisés. Ainsi, puisque les logiciels n'étaient pas les mêmes, certains paramètres ajustés différemment ont pu influencer la qualité des enregistrements.

De plus, la voix humaine n'a pas entraîné de reconnaissance parfaite des mots et énoncés. Ceci peut être dû aux bruits ambiants tels le système de ventilation ou l'ordinateur. Cependant, le bruit ambiant ne remet pas en cause les résultats puisqu'il était subtil et constant tout au long des séances, reflétant ainsi la conversation dans un environnement calme.

L'effet du contexte sur l'intelligibilité retrouvé dans la présente étude concorde avec des études antérieures (Mirenda & Beukelman, 1987, 1990; Hoover et coll., 1987). Toutes voix confondues, le pourcentage de mots identifiés correctement passe de 71% pour les mots sans contexte à 90% pour les mots avec contexte. À prime abord, on constate que l'augmentation absolue la plus importante survient pour les voix les moins intelligibles (Pierre; 24 à 63% et Félix; 54 à 90%). Toutefois, puisque la condition de présentation des mots isolés engendre un effet plafond pour certaines voix, une augmentation des pourcentages était virtuellement impossible pour celles-ci. Il pourrait donc être plus adéquat d'évaluer le taux d'amélioration engendré par l'utilisation du contexte (i.e. l'augmentation observée en fonction de l'augmentation possible). Par exemple, pour Pierre, avec un taux pour les mots isolés de 24%, une augmentation de 76% était possible. Or, avec un taux en contexte de 63%, on constate donc une augmentation observée de 39%, montrant un effet du contexte qui correspond à 51% de l'effet possible. On se rend alors compte qu'en appliquant cette formule, le pourcentage d'amélioration possible varie entre 51 et 78%. Le moins bon pourcentage étant attribué à la voix Pierre. On stipule donc que les participants ont profité de 51% du contexte pour identifier les mots avec la voix Pierre tandis qu'ils ont profité davantage du contexte avec plusieurs autres voix et ce de façon assez constante.

Toutes voix de synthèse confondues, le taux d'intelligibilité des mots sans contexte s'étend de 24% à 89%. Ceci indique une performance maximale considérablement élevé (voix humaine 89%). L'équipe de Trudeau et coll. (2006) avait montré un plafond du niveau d'intelligibilité plus faible, soit de 75% (voix humaine : 85%). Cette augmentation d'intelligibilité de synthèses vocales françaises depuis 2006 reflète une avancée rapide de la technologie dans le domaine des communications. Il est intéressant de noter que, malgré l'absence de contexte (mots isolés), deux voix de synthèse distinctes apparaissent aussi intelligibles (> 84%) que la voix humaine, et pour cinq autres voix de synthèse, un contexte limité (courts énoncés) augmente leur intelligibilité à un niveau semblable (> 92%) à celui de la voix humaine. Ce résultat conduit à un optimisme certain quant à l'utilisation des synthèses vocales dans diverses situations de la vie courante.

Les résultats confirment que certaines voix sont plus appréciées que d'autres. La voix humaine et la synthèse Virginie sont les préférées (cote moyenne 5,76 et 5,61, respectivement), et les deux voix de synthèse Pierre et Félix ont obtenu des évaluations négatives (inférieures à 3 sur l'échelle de 7). En général, les voix synthétiques évaluées dans cette étude sont plus appréciées que celles de l'étude de Trudeau et coll. (2006), qui avaient trouvé que la voix humaine était la seule à obtenir une cote nettement positive. Les résultats de la présente étude montrent que les voix les plus appréciées (voix humaine, Virginie) sont souvent les plus intelligibles dans les deux contextes tandis que celles les moins appréciées (Félix, Pierre) sont aussi les moins intelligibles. Tel que l'ont montré Sangsue et coll. (1997), l'utilisation des

adjectifs bipolaires hautement contradictoires pour qualifier les voix forçait le participant à plus de précision et ainsi opter un peu plus pour l'un ou pour l'autre des adjectifs. Les résultats révèlent que l'appréciation globale reflète certaines caractéristiques des voix. En effet, plus l'évaluation subjective était positive, plus il était probable que la voix soit jugée chaude, douce, expressive et fluide. Trois participants ont ajouté des commentaires sur leur feuille-réponse à propos des voix Pierre (robotique, peu compréhensible, parle sur le bout de la langue), Juliette (timbre très apprécié) et de la voix humaine (accent québécois prononcé, très naturelle). S'il est difficile de définir «le naturel» d'une voix (Nusbaum et coll., 1995), les adjectifs choisis pour qualifier les différentes synthèses vocales ont permis une juste évaluation de leur qualité dans le contexte où les participants devaient s'imaginer que les voix appartenaient à un proche.

Bien que le but de l'étude ne visait pas à faire une comparaison directe des voix de synthèse québécoises et françaises, les résultats montrent que des quatre voix les plus intelligibles en contexte, deux sont québécoises (voix humaine et Louise) et deux sont françaises (Bruno et Virginie) et que des deux voix les plus appréciées, Virginie est française et la voix humaine est québécoise. En condition de mots isolés, deux des trois voix les plus intelligibles sont québécoises. Les particularités propre à chaque dialecte (ex : affrication, diphtonguaison, régionalismes lexicaux, etc.) ne semblent pas influencer l'intelligibilité ou l'appréciation des synthèses vocales. Ceci peut être attribuable au fait que les participants, de nationalités multiples et résidents à Montréal, sont amenés à côtoyer quotidiennement des personnes parlant d'autres dialectes de la langue française. Bien qu'aucune étude ne l'indique, il serait intéressant d'évaluer l'impact des particularités de chaque dialecte sur l'intelligibilité et l'appréciation de différentes voix de synthèse.

Les résultats de cette étude mettent en évidence une hiérarchie des différentes synthèses vocales francophones tant du point de vue de l'intelligibilité que de l'appréciation. Nos résultats suggèrent une très bonne intelligibilité et appréciation des voix Acapela (Bruno et Louise), Nuance-Realspeak (Virginie), SAPI 4,0 (Sophie) et Loquendo (Olivier). La voix L&H (Pierre) s'est toutefois avérée peu intelligible et faiblement appréciée. Ceci révèle des progrès récents dans le domaine des voix artificielles et montre l'importance de ce type d'études afin de mieux guider les professionnels des troubles du langage dans l'attribution des systèmes de suppléance à la communication. De plus, le coût des voix de synthèse ne semble pas être un facteur déterminant dans le choix des voix à intégrer dans les appareils car toutes les voix coûtent moins de 50$, indépendamment du niveau d'intelligibilité.

D'autres recherches sont nécessaires pour explorer l'intelligibilité et l'appréciation des synthèses vocales françaises dans d'autres contextes (ex. conversation, au téléphone), par d'autres interlocuteurs (ex. utilisateurs de synthèses vocales et leurs proches) et avec d'autres types de stimuli (ex. messages sociaux; informations précises; vocabulaire moins fréquent) afin de mieux préciser les éléments et caractéristiques qui feront en sorte que les voix de synthèse répondent le mieux possible aux besoins des utilisateurs.

RÉFÉRENCES

Bourhis, V. (2010). La prosodie comme indice de contextualisation du discours didactique. *Colloque International « Spécifictés et diversité des interactions didactiques : disciplines, finalités, contextes »*, INRP, CNRS, Lyon

Breen, A. (1992). Speech synthesis models: A review. *Electronics & Communication Engineering Journal*, *4,* 19-31.

Chesneau, S. (2007). Effets du vieillissement et d'une lésion cérébrale gauche sur la compréhension de textes (Thèse de doctorat). Université de Montréal, Montréal.

Crabtree, M., Mirenda P., & Beukelman, D. R. (1990). Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication*, *6,* 256-261.

Drager, K. D., Anderson, J. L, Debarros, J., Hayes, E., Liebman, J., & Panek, E. (2007). Speech synthesis in background noise: effects of message formulation and visual information on the intelligibility of American English DECTalk. Augmentative and Alternative Communication, *23,* 177-86.

Drager, K. D. R., & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners. *Journal of Speech, Language, and Hearing Research*, *44,* 1052-1057.

Drager, K. D. R., & Reichle, J. E. (2001). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication*, *17,* 109-119.

Dutoit, T., Couvreur, L, Malfrère, F., Pagel, V., & Ris, C. (2002). Synthèse vocale et reconnaisance de la parole : Droites gauches et mondes parallèles. *Actes du 6è Congrès Français d'Acoustique*, Lille, France.

Ellis, L. W., Spiegel, B., & Benjamin, B. (2002). Effects of speakers' augmented characteristics and listeners' sex on intelligibility and acceptability of synthesized speech. *Perceptual and Motor Skills*, *94,* 1081-8.

Gong, L., & Lai, J. (2001). Shall we mix synthetic speech and human speech?: Impact on users' performance, perception, and attitude. *Proceedings of the SIGCHI conference on Human factors in computing systems.* Seattle, Washington, United States, ACM.

Higginbotham, D. J., Drazek, A. L. Kowarsky; K., Scally, C., & Segal, E. (1994). Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication*, *10,* 191-202.

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,* 1180-4.

Humes, L. E., Nelson, K. J., Pisoni, D. B., & Lively, S. E. (1993). Effects of age on serial recall of natural and synthetic speech. J*ournal of Speech and Hearing Research*, *36,* 634-9.

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,* 744-52.

Kangas, K. A., & Allen, G. D. (1990). Intelligibility of synthetic speech for normal-hearing and hearing-impaired listeners. J*ournal of Speech and Hearing Disorders*, *55,* 751-755.

Klatt, D. H. (1987). Review of text-to-speech conversion for English. T*he Journal of the Acoustical Society of America*, *82,* 737-93.

Koul, R. K., & 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. (2003). Synthetic speech perception in individuals with and without disabilities. *Augmentative and Alternative Communication*, *19,* 49-58.

Lai, J., Wood, D., & Considine, M. (2000). The effect of task conditions on

the comprehensibility of synthetic speech. *Proceedings of the SIGCHI conference on Human factors in computing systems.* The Hague, The Netherlands, ACM.

Manous, L. M., Pisoni, D. B., Dedina, M. J., & Nusbaum, H. C. (1986). Comprehension of natural and synthetic speech using a sentence verification task. *The Journal of the Acoustical Society of America*, *79*(S1), S25.

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.

Nusbaum, H. C., Francis, A. L., & Henly, A. S. (1995). Measuring the naturalness of synthetic speech. *International Journal of Speech Technology*, *1*, 7-19.

Nye, P. W., & Gaitenby, J. H. (1973). Consonant intelligibility in synthetic speech and in natural speech control. *In Haskins Laboratories Status Report on Speech Research, SR-33,* 77-91. New Haven, CT: Haskins Laboratories.

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.

Pisoni, D. B., Nusbaum, H. C., & Greene, BG. (1985). Perception of synthetic speech generated by rule. *Proceedings of the IEEE*, *73,* 1665-1676.

Pisoni, D., & Hunnicutt, S. (1980). Perceptual evaluation of MITalk: The MIT unrestricted text-to-speech system. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '80.* Cleveland, Ohio, USA.

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.

Roring, R. W., Hines, F. G., & Charness, N. (2007). Age differences in identifying words in synthetic speech. *The Journal of the Human Factors and Ergonomics Society*, *49,* 25-31.

Sangsue, J., Siegwart, H., Cosnier, J., Cornu, J., & Scherer, K. R. (1997). *Développement d'un questionnaire d'évaluation subjective de la qualité de la voix et de la parole* (QEV), FPSE, Université de Genève.

Schroder, M. (2001). Emotional Speech Synthesis: A Review. Proceedings of Eurospeech. ISCA, Germany, 561–564

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1995). Some Effects of Training on the Perception of Synthetic Speech. *The Journal of the Human Factors and Ergonomics Society*, *27,* 395-408.

Terken, J. (1993). Synthesizing natural-sounding intonation for Dutch: Rules and perceptual evaluation. *Computer Speech and Language*, *7,* 27-48.

Trudeau, N., Chaput, E., Sutton, A., Chan, E., & Contardo, R. (2006). Intelligibility and subjective ratings of French voice synthesizers. *Journal of Speech-Language Pathology and Audiology*, *30,* 158-168.

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

Von Berg, S., Panorka, A., Uken, D., & Qeadan, F. (2009). DECtalk and Verivox : Intelligibility, likeability and rate preference differences for four listener groups. *Augmentative and Alternative Communication*, *25,* 7-18.

Winters, S. J., & Pisoni, D. B. (2003-2004). Perception and comprehension of synthetic speech. *Research on spoken language processing.* Progress report no. 26, Indiana University, USA.

REMERCIEMENTS

Cette étude a été financée par le Programme de recherche en réadaptation pédiatrique 2008-2009 du CHU Sainte-Justine, le Comité d'organisation du programme des stages d'été (COPSE) et le Fonds de recherche en santé du Québec (FRSQ). Un merci particulier au technicien en informatique Guillaume Gagnon ainsi qu'à tous les participants de l'étude.

NOTE DES AUTEURS

Prière d'adresser toute correspondance à : Patricia Côté-Giroux : 412-5 Boulevard St-Joseph Est, Montréal, Québec, Canada. H2J 1J5. Courriel : patricia.giroux@umontreal.ca

Date soumis : le 12 février 2010
Date accepté : le 18 août 2010

ANNEXE A

Paragraphe 1

Laura a senti l'avion s'élever rapidement. C'était une magnifique journée ensoleillée, un vent léger finissait de disperser la brume qui couvrait la ville plus tôt en matinée. Le pilote a annoncé que l'avion se dirigeait vers le nord-ouest pour contourner ensuite l'Angleterre en direction du Groenland. De là, il ne restait que quelques heures pour atteindre sa destination : New York. Laura a regardé à travers le hublot. Elle pouvait entendre la pluie s'abattre sur la fenêtre.

Paragraphe 2

Martin a senti avec joie l'auto prendre de la vitesse. Cette nouvelle voiture répondait parfaitement bien aux sollicitations du chauffeur. Il faisait un temps splendide, idéal pour rouler, une petite brise allait finir de dégager le ciel qui s'était couvert dans la journée. Sa voisine, qui connaissait le chemin, a annoncé qu'il faudrait contourner la ville par le nord puis se diriger vers l'est de la province. Alors, il ne resterait que peu de route à faire avant d'arriver sur leur lieu de vacances : Boston.

Paragraphe 3

Patrick a vérifié son billet avant de le remettre au contrôleur du train. Dans quelques heures, il serait chez sa tante, à Halifax. Il a choisi un siège confortable, baigné par les rayons du soleil, et il a attendu que le train quitte la gare. Il a sorti de son sac la lettre que sa cousine Jeanne lui avait envoyée. Son regard s'est attardé sur la photographie jointe à la lettre. Il a contemplé longuement la petite pharmacie familiale qu'il venait d'acheter là-bas. Il avait terriblement hâte d'arriver pour rencontrer ses futurs employés.

Paragraphe 4

Le vélo bleu de Marion était posé contre le mur de brique. La jeune femme avait emprunté la route en direction de Percé à l'aube. Elle ne s'était arrêtée qu'une fois trempée par la fine pluie qui ne cessait de tomber. Elle aurait voulu continuer mais une faim de loup l'en empêchait. Elle est entrée dans un petit café, pour commander une bonne soupe chaude. Tandis qu'un vieil homme préparait son repas, Marion a soudainement vu quelqu'un s'emparer de sa bicyclette.

ANNEXE B

Scores moyens des niveaux d'intelligibilité (maximum = 16), écarts-type (en parenthèses) et pourcentages équivalents pour chaque synthèse vocale dans les deux conditions (sans contexte et avec contexte) pour les trois groupes d'âge

Voix/ Nom de la synthèse vocale		Groupe								
		14-19 ans		20-39 ans		40-60 ans		Total		Total
		Sans	Avec	Sans	Avec	Sans	Avec	Sans	Avec	
Humaine	Score	14,4 (1,3)	15,4 (1,1)	14,5 (1,0)	15,5 (1,2)	14,0 (1,9)	15,7 (0,7)	14,3 (1,5)	15,5 (1,0)	14,9 (1,0)
	Pourcentage	90,0	96,3	90,6	96,9	87,5	98,1	89,4	96,9	93,1
Louise	Score	14,5 (1,6)	15,4 (0,8)	14,0 (1,3)	15,7 (0,7)	14,2 (1,2)	15,7 (0,7)	14,2 (1,3)	15,6 (0,7)	14,9 (0,9)
	Pourcentage	90,6	96,3	87,5	98,1	88,8	98,1	88,8	97,5	93,1
Virginie	Score	12,6 (3,3)	15,1 (1,6)	14,1 (1,5)	15,6 (0,6)	13,5 (1,6)	15,2 (1,3)	13,4 (2,3)	15,3 (1,2)	14,4 (1,4)
	Pourcentage	78,8	94,4	88,1	97,5	84,4	95,0	83,8	95,6	90,0
Sophie	Score	12,7 (1,8)	14,5 (1,5)	12,8 (1,7)	14,9 (1,3)	12,1 (1,8)	14,8 (1,2)	12,5 (1,8)	14,7 (1,3)	13,6 (1,3)
	Pourcentage	79,4	90,6	80,0	93,1	75,6	92,5	78,1	91,9	85,0
Bruno	Score	12,2 (1,3)	14,4 (1,7)	12,7 (1,6)	15,4 (0,8)	12,2 (1,7)	15,0 (1,0)	12,4 (1,6)	14,9 (1,3)	13,7 (1,1)
	Pourcentage	76,3	90,0	79,4	96,3	76,3	93,8	77,5	93,1	85,6
Olivier	Score	12,7 (2,5)	14,6 (1,4)	12,3 (1,7)	15,1 (1,0)	11,5 (1,9)	14,6 (1,2)	12,2 (2,1)	14,8 (1,2)	13,5 (1,4)
	Pourcentage	79,4	91,3	76,9	94,4	71,9	91,3	76,3	92,5	84,4
Juliette	Score	11,0 (2,1)	14,3 (1,4)	11,6 (1,9)	14,4 (1,2)	11,1 (2,0)	14,1 (1,2)	11,2 (2,0)	14,3 (1,2)	12,7 (1,2)
	Pourcentage	68,8	89,4	72,5	90,0	69,4	88,1	70,0	89,4	79,4
Charlotte	Score	10,0 (2,1)	14,0 (1,5)	10,9 (1,8)	14,7 (1,5)	10,6 (1,8)	14,4 (1,4)	10,5 (1,9)	14,4 (1,5)	12,4 (1,4)
	Pourcentage	62,5	87,5	68,1	91,9	66,3	90,0	65,6	90,0	77,5
Félix	Score	8,4 (2,1)	14,4 (1,8)	9,1 (2,8)	14,5 (0,9)	8,2 (1,7)	14,3 (1,2)	8,6 (2,2)	14,4 (1,3)	11,5 (1,4)
	Pourcentage	52,5	90,0	56,9	90,6	51,3	89,4	53,8	90,0	71,9
Pierre	Score	4,0 (1,7)	9,1 (2,8)	4,0 (2,2)	10,9 (2,5)	3,9 (1,5)	10,1 (2,2)	3,9 (1,8)	10,0 (2,6)	7,0 (1,6)
	Pourcentage	25,0	56,9	25,0	68,1	24,4	63,1	24,4	62,5	43,8
Total	Pourcentage							70,8	89,9	80,4