Mesure d'intelligibilité de segments de parole à l'envers en français

Fanny Meunier, Tristan Cenier, Melissa Barkat, et Ivan Magrin-Chagnolleau

Laboratoire Dynamique du Langage
Institut des Sciences de l'Homme
14, avenue Berthelot – 69363 Lyon Cedex 07, France
Tél.: +33 (0)4 72 72 64 12 - Fax: +33 (0)4 72 72 65 90
Mél: fanny.meunier@ish-lyon.cnrs.fr; melissa.barkat@univ-lyon2.fr; ivan.magrin-chagnolleau@univ-lyon2.fr

ABSTRACT

We ran an experiment focusing on cognitive implication of reversed speech segments. Nine durations of reversed segments plus a non-distorted control condition have been considered (varying between 20 ms and 180 ms) in order to test the pattern of intelligibility degradation in French. We observed an overall strong negative correlation between the degree of intelligibility and the size of reversed-speech windows. These results appear to be very comparable to those obtained in English by Greenberg & Arai [Gre01], at least on the slope of intelligibility performance decrease. However, intelligibility loss in French is delayed by twenty milliseconds. Apart from confirming the cognitive ability to restore reversed speech up to a certain point, our study revealed differences that could be interpreted as *'language specific'*.

1. INTRODUCTION

La perception du langage parlé est une tâche menée quotidiennement et représente un haut degré d'implication des fonctions cognitives. Son intelligibilité dépend directement de la clarté du signal émis et de la capacité cérébrale à traiter ce signal. Ces traitements comprennent entre autres : 1) la discrimination langue/non-langue permettant de s'affranchir des phénomènes parasites (réverbération, bruit de fond, discours simultanés, etc.) pendant une conversation dans un environnement sonore chargé ; et 2) la capacité à corriger les erreurs dues à la distorsion du signal.

Différentes études ont montré que le langage parlé reste intelligible malgré certaines détériorations acoustiques : il existe une capacité cognitive à reconstruire un signal de parole ayant subi une détérioration due aux conditions d'émission/réception. Par exemple le phénomène de restauration de phonème, où lorsqu'on remplace un phonème dans un mot par un bruit, les sujets continuent à percevoir le mot dans son intégrité [War70]. Cette capacité de restauration dépend du type de distorsion et de son importance.

Il existe au moins deux limites concernant la capacité du système de restauration du signal de parole, que l'on qualifiera respectivement de physiologique et de cognitive. La limite physiologique est la quantité maximum d'information pouvant être traitée par unité de temps. La limite cognitive se situe à un niveau de traitement supérieur et à pour but d'empêcher une surexploitation du système qui conduirait à des constructions purement fictives nuisant à la compréhension du message reçu. Ces considérations étant établies, il apparaît alors crucial de déterminer les limites de la capacité de restauration.

L'inversion temporelle de la parole a été qualifiée comme *" la forme la plus drastique de détérioration de la parole "* [Sab99]. Il est classiquement admis que de la parole jouée à l'envers soit utilisée comme condition de contrôle dans des expériences visant à montrer l'existence d'une capacité humaine à traiter le langage, par exemple [Mel88]. Cependant, des expériences réalisées en anglais ont montré que le système perceptuel humain était capable dans une certaine mesure de traiter l'inversion de la parole [Sab99]; [Gre01]. D'après ces résultats, l'intelligibilité du signal dépend de la longueur et de la fréquence des fenêtres d'inversion.

Le but de l'expérience présentée est de quantifier les capacités du système cognitif à récupérer l'information lexicales contenues dans des phrases ayant subit des inversions temporelles.

2. EXPÉRIENCE

Le principe de l'expérimentation est de soumettre des sujets volontaires à l'écoute d'une série de phrases plus ou moins modifiées et de leur demander de retranscrire ce qu'ils ont entendu sur un micro-ordinateur.

2.1 Méthode

Matériel

Nous avons utilisé un corpus de cinquante phrases extraites de la base de données acoustiques BD-Sons. Elles sont prononcées par dix locuteurs différents (5 hommes et 5 femmes), sont équilibrées phonétiquement et ont une durée moyenne de trois secondes.

La préparation des stimuli a été effectuée avec le logiciel MATLAB et consiste en un retournement du signal sonore sur son axe temporel sur une durée de x millisecondes toutes les x millisecondes, les valeurs choisies pour le paramètre x étant 20 ; 40 ; 50 ; 60 ; 70 ; 80 ; 100 ; 140 et 180 ms. Le paramètre 0 (phrase intacte) est utilisé comme condition contrôle.

Par ailleurs, nous avons fait un pré-test visant à déterminer la prédictibilité des phrases présentées. On peut effectivement penser que ce paramètre peut avoir une influence sur la performance des sujets : il se peut qu'un sujet ne comprenant pas un mot en devine cependant la

nature par le contexte de la phrase. Pour réaliser ce pré-test, nous avons présenté à 12 sujets – autres que ceux devant passer l'expérience d'intelligibilité – une liste des cinquante phrases amputées de leur dernier mot avec comme consigne de les compléter. Nous avons ensuite utilisé cette mesure afin de construire nos listes (les phrases dans chaque condition étaient équivalente sur ce point).

Comme nous avions 10 conditions expérimentales nous avons réparties nos 50 phrases en 10 sous-groupes de 5 phrases. Ces sous-groupes étaient appariés en fonction : 1) du nombre de syllabes qui les constituent ; 2) de leur degré de prédictibilité ; 3) du logarithme des fréquences des mots ; 4) du nombre de mots dans la phrase ; et enfin 5) du sexe du locuteur qui la prononce.

Nous avons ensuite constitué 5 listes expérimentales comprenant chacune les 50 phrases. Ce qui variait entre les différentes listes était la condition d'inversion appliquée à chaque sous-groupe de 5 phrases. Par exemple, la phrase *'Il se garantira du froid avec un bon capuchon'* était présentée dans la condition d'inversion 40 ms dans la liste 1, 140 ms dans la liste 2, et 0 ms dans la liste 3. Une autre phrase (d'un autre sous-groupe), par exemple *'Le train entre déjà en gare'*, était présentée dans la condition d'inversion 140 ms dans la liste 1, 40 ms dans la liste 2, et 80 ms dans la liste 3, et ainsi de suite pour toutes les phrases et toutes les listes.. Cette manipulation a permis que chaque sujet qui n'écoutait qu'une liste entendait toutes les conditions d'inversion (5 phrases par condition) et chacune des phrases une seule fois. Pour chaque liste, les items étaient répartis selon un ordre pseudo-aléatoire. Cette distribution pseudo-aléatoire assure une bonne répartition des différentes conditions expérimentales sur l'ensemble de la liste ainsi que la non répétition d'une même condition d'inversion sur deux items consécutifs.

Déroulement de l'expérience

Les passations de l'expérience ont eu lieu à l'Institut des Sciences de l'Homme à Lyon. Chaque sujet, équipé d'un casque stéréo, a été placé face à un ordinateur. Les phrases étaient présentées l'une après l'autre dans une des conditions expérimentales (fenêtres d'inversion de 0 (contrôle), 20, 40, 50, 60, 70, 80, 100, 140 ou 180 ms). Le sujet devait taper sur un clavier, immédiatement après écoute, la phrase perçue. Il avait la possibilité d'écouter jusqu'à quatre fois chaque phrase.

Procédure

L'expérience se déroule en trois phases :

1- Présentation à l'écran de la consigne : *Vous allez écouter plusieurs phrases. Pour chaque phrase, vous devrez taper les mots que vous avez reconnus. Ne tapez rien si vous n'avez reconnu aucun mot. Puis appuyez sur la touche " Entrée " pour activer le bouton " Écouter la phrase suivante ". Vous pourrez écouter chaque phrases jusqu'à quatre fois*.

2- Une phase d'apprentissage au cours de laquelle le sujet écoute et retranscrit successivement cinq phrases, avec possibilité d'écouter chacune des phrases jusqu'à quatre fois.

3- Une fois que l'expérimentateur s'est assuré que le sujet avait compris le principe de l'expérience la liste expérimentale proprement dite était diffusée. La tâche du sujet était la même que dans la phase d'apprentissage.

La durée totale de l'expérience était d'environ trente minutes.

Sujets

Nous avons fait passer l'expérience à 28 sujets naïfs par rapport au but de l'étude, étudiants de première année à l'Ecole des Psychologues Praticiens de Lyon. Ils étaient tous de langue maternelle française, âgés de 19 à 30 ans, et ne présentant pas de handicap auditif connus. Nous les avons indemnisés pour leur participation.

2.2 Résultats

Les données sont obtenues en reprenant chaque phrase retranscrite par chaque sujet et en calculant le pourcentage de mots exacts par rapport au nombre total de mots dans la phrase. Nous avons effectué ce codage manuellement acceptant ainsi les fautes d'orthographes.

La courbe d'intelligibilité est obtenue en calculant pour chaque condition d'inversion la moyenne des pourcentages de mots correctement reconnus.

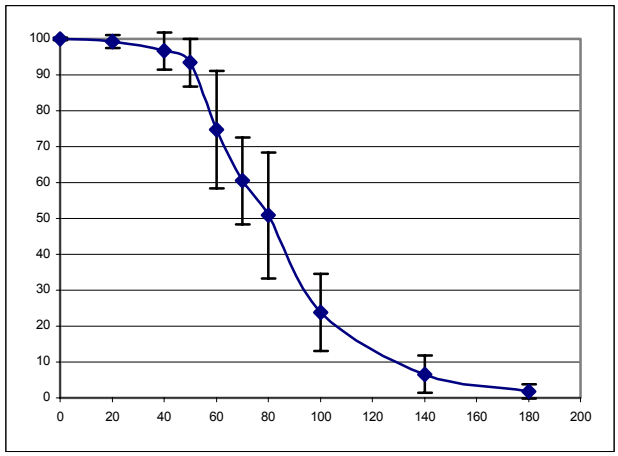

Figure 1 : Intelligibilité en fonction de la taille de la fenêtre d'inversion en ms (0, 20, 40, 50,60, 70, 80, 100, 140, 180). L'intelligibilité est conservée à 90 % entre 0 ms et 50 ms puis on observe une chute rapide des performances des sujets jusqu'à 100 ms, après quoi l'intelligibilité est inférieure à 10 %.

Conformément à ce à quoi nous nous attendions, les phrases utilisées comme contrôle (0 ms d'inversion) présentent un taux d'intelligibilité de 100 %. Le reste de la courbe peut-être analysé selon trois sections : 1) une phase d'intelligibilité conservée]100% ;90%] pour les traitements]0ms ;50ms] ; 2) une phase dite d'ambiguïté où l'on constate une dégradation de l'intelligibilité

[75% ;25%] pour les traitements [60ms ;100ms] ; et 3) une phase d'incompréhension, où l'intelligibilité est fortement diminuée]10% ;2%] pour les derniers traitements [140ms ;180ms].

Nous observons une forte corrélation négative entre le degré d'intelligibilité et la longueur de la fenêtre d'inversion (r=-0.95).

3. DISCUSSION

Nos résultats indiquent une dégradation de l'intelligibilité en fonction de la taille des fenêtres d'inversion du signal.

Une expérience comparable a été réalisée en anglais par Greenberg et Arai [Gre01]. Ces auteurs observent (d'après une estimation basée sur la courbe publiée dans leur article) : 80% d'intelligibilité pour la condition 40 ms ; 60% pour 50 ms ; 50% pour 60 ms ; 30% pour 70 ms ; 25% pour 80 ms. La rupture presque totale de l'intelligibilité était observée à partir de 100 ms d'inversion de signal (6% pour 100 ms, 4% pour 140 ms et 3,5% pour 180%). Greenberg et Arai [Gre01] relatent une corrélation négative entre la dégradation de l'intelligibilité et la taille des fenêtres d'inversion du signal.

Si l'on compare les deux droites de régression pour l'anglais et le français obtenues à partir des moyennes par condition (Figure 2), on observe que les pentes des deux droites sont très similaires (a= -0.67 pour le français, et a= -0.62 pour l'anglais). En revanche on observe une différence concernant l'ordonnée à l'origine (b= 110.58 pour le français et b= 90.74 pour l'anglais). Cette différence reflète un délai temporel d'environ 20 ms dans la dégradation de l'intelligibilité observée en français.

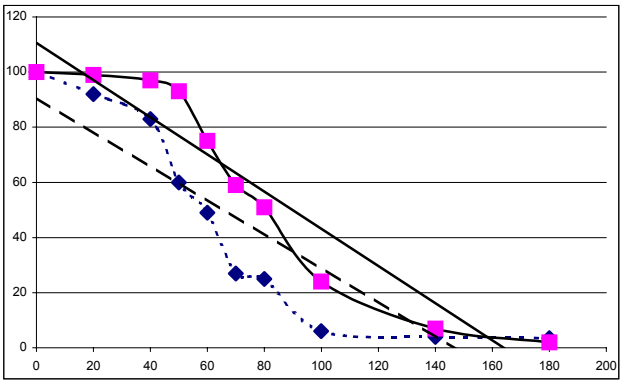

Figure 2 : Courbes d'intelligibilité et droites de régression pour l'anglais (en pointillés) -d'après [GRE01] - et pour le français (traits pleins) en fonction de la taille de la fenêtre d'inversion en ms (0, 20, 40, 50,60, 70, 80, 100, 140, 180).

Par exemple (voir Tableau 1), en anglais avec des fenêtres d'inversion de 60 ms, le taux d'intelligibilité est d'environ 50% alors qu'en français cette valeur est atteinte 20 ms plus tard, c'est-à-dire. avec des fenêtres d'inversion de 80 ms.

Tableau 1 : Taux d'intelligibilité en anglais et en français en fonction de la taille de la fenêtre d'inversion en ms (0, 20, 40, 50,60, 70, 80, 100, 140, 180

Langues / Tailles fenêtres d'inversion	**Anglais**	**Français**
0 ms	100 %	100 %
20 ms	92 %	99 %
40 ms	83 %	97 %
50 ms	60 %	93 %
60 ms	49 %	75 %
70 ms	27 %	59 %
80 ms	25 %	51 %
100 ms	6 %	24 %
140 ms	4 %	7 %
180 ms	3,5 %	2 %

Nos données confirment donc la capacité du système cognitif à restaurer, dans une certaine mesure, des segments du signal de la parole présentés à l'envers, mais révèlent des différences entre notre expérience et celle de Greenberg et Arai [Gre01]. Comment expliquer cette différence de délai temporel ? Nous pouvons envisager deux possibilités : soit une différence méthodologique, soit une différence due à la spécificité des langues utilisées (le français vs. l'anglais).

La différence méthodologique entre les deux expériences peut se situer au niveau du critère d'acceptation d'un mot comme exact : Greenberg rapporte '*the number of correct words per sentence was scored using an algorithm that automatically compensated for minor errors in spelling*' ; pour notre étude nous avons effectué cette cotation manuellement acceptant ainsi même les fautes d'orthographe affectant plusieurs graphèmes, par exemple : '*accour'* au lieu de '*accourent'*. Cet exemple montre un bon décodage phonologique même si la transcription orthographique est erronée. Il n'est pas clair quel était précisément le critère utilisé par Greenberg, mais si son algorithme acceptait uniquement 1 à 2 caractères divergents de la bonne orthographe, il est possible qu'au moins une partie de la différence observée entre les deux expériences soit explicable par cette différence de critère.

L'autre possibilité est que la différence observée soit due à une différence dans le signal de parole des deux langues, comme des différences phonologiques (par exemple dans les structures syllabiques, l'accentuation, les caractéristiques rythmiques,…). Des expériences complémentaires seront nécessaires afin de clarifier les processus sous-jacents aux performances d'intelligibilité.

4. CONCLUSION

Dans cet article, nous avons présenté des expériences perceptuelles sur la capacité cognitive à restaurer de la

parole inversée en français. Des phrases, appartenant à la base de données BD-SONS, ont été inversées temporellement par segments. La taille des segments variant de 20 ms à 180 ms (plus une condition de contrôle où les phrases n'étaient pas modifiées). Nous avons observé une forte corrélation négative entre le degré d'intelligibilité et la taille des segments inversés. Pour des tailles d'inversion allant de 0 à 50 ms, l'intelligibilité reste supérieure à 90%. Pour des tailles d'inversion allant de 60 à 100 ms, l'intelligibilité décroît de 75 à 25%. Et enfin pour des tailles d'inversion supérieures à 140 ms, l'intelligibilité devient inférieure à 10%. Cette expérience est comparable à une expérience menée en anglais par Greenberg et Arai [Gre01]. La pente de la corrélation négative est similaire. Cependant, l'intelligibilité se perd en moyenne 20 ms plus tôt sur les expériences menées en anglais. Cette différence peut s'expliquer par une différence méthodologique, notamment dans la façon de compter le nombre de mots correctement reconnus. Cette différence peut aussi s'expliquer par des différences phonologiques entre le français et l'anglais (par exemple la structure syllabique, l'accentuation, les caractéristiques rythmiques). Des expériences complémentaires seront menées afin de clarifier les processus cognitifs sous-jacents à cette tâche de restauration de parole inversée.

5. NOTES DES AUTEURS ET REMERCIEMENTS

L'ordre des auteurs est arbitraire. Nous remercions François Pellegrino et Leonid Synyukov pour leur aide et contribution. La recherche décrite dans cet article a été réalisée grâce à des financements du Centre National de la Recherche Scientifique – APN 2001 n°26635 attribuée à Melissa Barkat et grâce à une ACI – MSH 2001 attribuée à Fanny Meunier.

BIBLIOGRAPHIE

[Gre01] Greenberg, S. & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum, Proceedings of the 7th Eurospeech Conference on Speech Communication and Technology (Eurospeech-2001), pp. 473-476.

[Mel88] Melher, Jusczyk, Lambertz, Halsted, Bertoncini, Amiel-tison, "A precursor of language acquisition in young infants", Cognition, 29, 143-178.(1988)

[Sab99] Saberi, K. & Perrott, D.R. (1999). Cognitive restoration of reverse speech, Nature, 398: 760.

[War70] Warren, R. M.(1970). Perceptual restoration of missing speech sounds, Science, 167, 392-393.