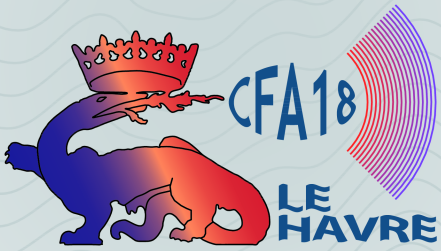
CFA '18 LE HAVRE ▪ 23-27 avril 2018
14ème Congrès Français d'Acoustique



Optimisation d'un critère d'Intelligibilité de la Parole dans un Contexte Bruité Automobile

E. Gentet[a,b], B. David[a], S. Denjean[b], G. Richard[a] et V. Roussarie[b]
[a]LTCI, Télécom ParisTech, Université Paris-Saclay, 46 rue Barrault, 75013 Paris, France
[b]Groupe PSA, Route de Gisy, 78140 Vélizy-Villacoublay, France
enguerrand.gentet@telecom-paristech.fr

Ce travail s'inscrit dans le cadre de l'amélioration de l'intelligibilité des signaux de parole en contexte bruité automobile. Contrairement aux approches de rehaussement il s'agit ici de réallouer l'énergie de la parole pour améliorer sa compréhension, mais sans modifier le Rapport Signal sur Bruit (RSB) global. L'approche étudiée consiste à effectuer cette réallocation en maximisant une fonction objectif basée sur le *Speech Intelligibility Index* (SII). Ce critère est calculé à partir d'une somme pondérée de RSB sur différents canaux fréquentiels et il est alors possible d'appliquer un égaliseur dynamique aux signaux de parole sous une contrainte énergétique donnée. À la différence d'autres méthodes, nous proposons une résolution exacte du problème d'optimisation du SII avec adaptation dynamique aux signaux (parole et bruit). Elle prend en compte la sensibilité acoustique de l'utilisateur et utilise une échelle de puissance adaptée au contexte de manière à ne pas augmenter l'intensité perçue. Les résultats montrent une nette amélioration de l'intelligibilité et des tests d'écoute subjectifs viennent valider la méthode.

1 Introduction

La diffusion de signaux de parole dans les habitacles automobiles est sans cesse grandissante à travers notamment les télécommunications, les émissions de radio, les podcasts, les informations transmises par l'ordinateur de bord ou le système de navigation,... Cependant, malgré les efforts et les avancées mécaniques, beaucoup de bruit persiste au sein de l'habitacle. Son origine peut être de diverses natures (roulement, moteur, frottements aérodynamiques, ...) et sa présence parasite très souvent la bonne écoute des signaux audio. L'objectif de ce travail est donc de traiter les signaux de parole avant leur diffusion dans l'habitacle afin d'améliorer leur intelligibilité en fonction du bruit ambiant.

En débruitage, et plus particulièrement dans le domaine de la réduction de bruit, la parole et le bruit sont mélangés dans le signal à traiter. Au contraire, lorsque le signal de parole et le bruit sont séparés, ce sont des méthodes de contrôle de bruit qui sont majoritairement employées. Notre étude se place à la croisée de ces deux domaines en tentant de contrecarrer l'effet néfaste du bruit en agissant, en amont, sur le signal avant qu'il ne s'intègre au bruit. Contrairement aux approches de rehaussement, il s'agit donc de modifier les signaux de parole afin d'améliorer leur compréhension sans modifier le Rapport Signal sur Bruit (RSB) global. Cette situation suscite de plus en plus d'intérêt [1, 2], à tel point qu'un challenge *Hurricane* [3] a été mis en place sur cette problématique lors de l'édition 2013 de la conférence internationale *Interspeech*.

Nous avons choisi le *Speech Intelligibility Index* (SII) [4] comme outil d'évaluation de l'intelligibilité des signaux de parole. Ce critère étant calculé à partir d'une somme pondérée de RSB sur différents canaux fréquentiels, notre approche consiste à appliquer un égaliseur dynamique aux signaux de parole afin d'optimiser le SII et, à fortiori, d'améliorer l'intelligibilité des signaux.

Cette stratégie a déjà été suivie dans la littérature, que ce soit par une approximation linéaire du SII [5], ou une approximation non linéaire [6]. Dans tous les cas, la contrainte énergétique n'est pas pondérée et, au vu des spectres traités obtenus, il est probable que le niveau perçu ait été augmenté. Cette augmentation du niveau perçu pourrait expliquer l'augmentation significative de l'intelligibilité lors de tests subjectifs. Dans notre contexte automobile, si l'utilisateur souhaite une intensité perçue supérieure, il peut l'augmenter manuellement. Au contraire, si nous l'augmentons, l'utilisateur la ramènera à un niveau confortable. Ainsi, en plus de proposer une méthode de résolution exacte du problème de maximisation du SII, nous proposons une contrainte énergétique supplémentaire qui consiste à maintenir le niveau perçu constant. Les résultats montrent une nette amélioration du SII dans un contexte automobile à grande vitesse et des tests d'écoute subjectifs viennent valider la méthode.

L'article est organisé de la façon suivante. Premièrement la justification du choix du SII comme mesure d'intelligibilité et sa méthode de calcul sont détaillés. Ensuite, la procédure d'optimisation et son exploitation pour traiter les signaux de parole sont présentés. Enfin, nous décrivons la validation de l'approche par tests subjectifs et analysons les résultats obtenus. Une conclusion sur les performances et les limites de notre algorithme est finalement proposée avec des propositions d'améliorations et d'extensions pour le futur.

2 Mesure de l'intelligibilité de la parole

Plusieurs mesures objectives d'intelligibilité de la parole existent et chacune est adaptée à des cadres bien précis. Dans cette partie, nous justifions, dans notre contexte, l'utilisation d'un critère en particulier et nous détaillons sa méthode de calcul.

2.1 Choix de la mesure

Les mesures d'intelligibilité découlent souvent de deux mesures principales qui sont le *Speech Transmission Index* (STI) [7] et le *Speech Intelligibility Index* (SII) [4]. D'un côté, le STI et ses extensions se basent sur la Fonction de Transfert de Modulation (FTM) du canal de transmission et sont donc adaptés à l'étude de l'influence des caractéristiques du canal sur l'intelligibilité de la parole. D'un autre côté, le SII et ses extensions se basent sur le Rapport Signal sur Bruit (RSB) et sont donc plus adaptés a l'étude de l'intelligibilité de la parole dans un environnement bruité. Compte tenu du fait que l'environnement bruyant est la principale cause de la baisse d'intelligibilité des signaux transmis dans un habitacle automobile, le choix du SII comme mesure de référence est totalement justifié. Concernant les diverses extensions proposées par la communauté nous pouvons citer une liste non exhaustive avec l'AI-ST [8] et des mesures objectives modifiées [9], adaptées aux bruits fluctuants, ou encore le CSII [10], adapté aux aides auditives car basé sur le taux de distorsion harmonique. Certaines de ces extensions pourraient s'avérer utiles dans la suite de nos travaux, notamment le CSII car les traitements que nous portons aux signaux peuvent s'apparenter à l'utilisation d'aides auditives introduisant des distorsions. Cependant, le SII classique reste la mesure la plus utilisée dans la communauté scientifique et cet article s'appuiera sur sa définition originale introduite dans la norme ANSI/ASA S3.5 [4] en 1997.

2.2 Méthode de calcul du SII

Les mesures objectives sont basées sur de solides connaissances empiriques et de nombreuses hypothèses. L'hypothèse principale du SII est que la parole est composée de canaux fréquentiels qui sont porteurs d'informations indépendantes. La norme ANSI/ASA S3.5 [4] prévoit le calcul du critère pour différentes décompositions en i^{max} canaux e.g. bandes d'octaves ($i^{max} = 6$), bandes de tiers d'octaves ($i^{max} = 18$) et bandes critiques ($i^{max} = 21$). Ainsi, le SII est calculé à partir des niveaux du spectre équivalent de parole E_i, et des niveaux du spectre équivalent de bruit N_i, dans chaque bande et en décibels. Ces niveaux s'obtiennent en intégrant le périodogramme de chaque signal sur leurs canaux respectifs et en normalisant par la largeur de bande associée $\delta freq_i$. A partir de ces spectres équivalents, deux coefficients par bande sont calculés : les coefficients d'audibilité et de distorsion.

2.2.1 Coefficients d'audibilité

Les coefficients d'audibilité représentent la proportion du spectre audible au-dessus des diverses perturbations qui impactent l'intelligibilité. Ils nécessitent un calcul préalable des niveaux D_i du spectre équivalent de perturbation donné par l'équation suivante :

$$D_i = max(T_i, Z_i)\,, \tag{1}$$

avec T_i les niveaux du seuil d'audition de l'auditeur et Z_i les niveaux du spectre équivalent de masquage. Les Z_i prennent en compte l'étalement du masquage et, en fixant l'hypothèse 1, les Z_i se calculent uniquement à partir de l'ensemble $\{N_j\}_{j\leq i}$ en appliquant une formule fournie par la norme ANSI/ASA S3.5 [4] et qui dépend du type de bandes choisies. Le spectre équivalent d'un bruit stationnaire de type automobile à grande vitesse et le spectre équivalent de masquage associé sont visibles figure 2a.

Hypothèse 1 *Les niveaux N_i du spectre équivalent de bruit sont supérieurs aux niveaux du spectre équivalent d'auto-masquage :*

$$N_i \geq E_i - 24\,\text{dB}\,. \tag{2}$$

Les coefficients d'audibilité A_i sont alors calculés comme indiqué par l'équation suivante :

$$A_i(E_i, D_i) = \begin{cases} 0 & \text{si} \quad E_i \leq E_i^{act}(D_i)\,, \\ \frac{E_i - E_i^{act}(D_i)}{E_i^{lim}(D_i) - E_i^{act}(D_i)} & \text{si} \quad E_i^{act}(D_i) \leq E_i \leq E_i^{lim}(D_i)\,, \\ 1 & \text{si} \quad E_i \geq E_i^{lim}(D_i)\,, \end{cases} \tag{3}$$

$$\text{avec} \;\; E_i^{act}(D_i) = D_i - 15\,\text{dB} \;\; \text{et} \;\; E_i^{lim}(D_i) = D_i + 15\,\text{dB}\,. \tag{4}$$

2.2.2 Coefficients de distorsion

Les coefficients de distorsion L_i prennent en compte la distorsion introduite lorsque les niveaux par bande s'éloignent trop des niveaux d'un spectre équivalent de parole de référence U_i fournit dans la norme ANSI/ASA S3.5 [4]. Ils sont calculés comme indiqué par l'équation suivante :

$$L_i(E_i) = \begin{cases} 1 & \text{si} \quad E_i \leq E_i^{rupt}\,, \\ 1 - \frac{E_i - E_i^{rupt}}{E_i^{fin} - E_i^{rupt}} & \text{si} \quad E_i^{rupt} \leq E_i \leq E_i^{fin}\,, \\ 0 & \text{si} \quad E_i \geq E_i^{fin}\,, \end{cases} \tag{5}$$

$$\text{avec} \quad E_i^{rupt} = U_i + 10\,\text{dB} \quad \text{et} \quad E_i^{fin} = U_i + 170\,\text{dB}\,. \tag{6}$$

2.2.3 Fonction d'importance de bande

Toutes les bandes ne contiennent pas la même quantité d'information relative à la parole, elles n'ont donc pas la même importance vis à vis de l'intelligibilité. Ainsi, une Fonction d'Importance de Bande (FIB), dont les coefficients sont notés I_i est appliquée pour pondérer chaque bande. Plusieurs FIB sont mises à disposition dans la norme ANSI/ASA S3.5 [4] en fonction du matériel vocal utilisé e.g. des syllabes sans sens particulier, des mots monosyllabiques ou des courts passages de discours.

2.2.4 Formule du SII

La formule du SII correspond donc à une somme pondérée de ces différents facteurs dans chaque bande, comme indiqué par l'équation suivante :

$$SII(\{E_i\}, \{D_i\}) = \sum_{i=1}^{i^{max}} I_i \cdot A_i(E_i, D_i) \cdot L_i(D_i)\,. \tag{7}$$

3 Optimisation et traitement

3.1 Définition du problème

Afin de prendre en compte la sensibilité acoustique de l'utilisateur, nous utilisons une échelle de puissance adaptée, le dB(A), de manière à ne pas augmenter l'intensité perçue. L'objectif de l'optimisation est donc de trouver les niveaux optimaux E_i^{opt} qui maximisent le SII pour un bruit donné, sans augmenter le niveau perçu en dB(A) S^{dBA}. On peut donc formuler le problème de la façon suivante :

$$\{E_i^{opt}\} = \arg\max_{\{E_i\}} \sum_{i=1}^{i^{max}} f_i(E_i, D_i)\,, \tag{8}$$

soumis à

$$\sum_i g_i(E_i) \leq 10^{S^{dBA}/10}\,, \tag{9}$$

avec

$$f_i(E_i, D_i) = I_i \cdot A_i(E_i, D_i) \cdot L_i(E_i)\,, \tag{10}$$

$$g_i(E_i) = h_i \cdot \delta freq_i \cdot 10^{E_i/10} = 10^{(E_i + H_i \cdot \Delta freq_i)/10}\,. \tag{11}$$

Les H_i (dB) correspondent aux coefficients de pondération de passage en dB(A). On obtient alors un problème d'allocation de ressource classique, aussi appelé problème du sac à dos.

3.2 Procédure d'optimisation

Les fonctions g_i sont continues et convexes sur tout $\mathbb{R}$. La figure 1 montre l'allure des fonctions f_i dans trois situations, notons que ces fonctions sont :

- constantes et minimales sur $]-\infty, E_i^{act}]$ donc soit $E_i^{opt} > E_i^{act}$, soit la bande i est désactivée,
- décroissantes sur $[E_i^{lim}, +\infty[$ donc $E_i^{opt} \leq E_i^{lim}$,
- continues et concaves sur $[E_i^{act}, E_i^{lim}]$,
- non différentiables en E_i^{rupt} si $E_i^{act} < E_i^{rupt} < E_i^{lim}$.

On obtient donc au maximum 3 intervalles de recherche possibles sur chaque bande : $[E_i^{act}, E_i^{rupt}]$, $[E_i^{rupt}, E_i^{lim}]$ ou alors la bande est désactivée (on fixera $E_i = E_i^{min} = -60\,\text{dB}$). Le nombre maximal de sous problèmes est donc de $3^{i^{max}}$ et il est possible d'utiliser un algorithme de séparation et

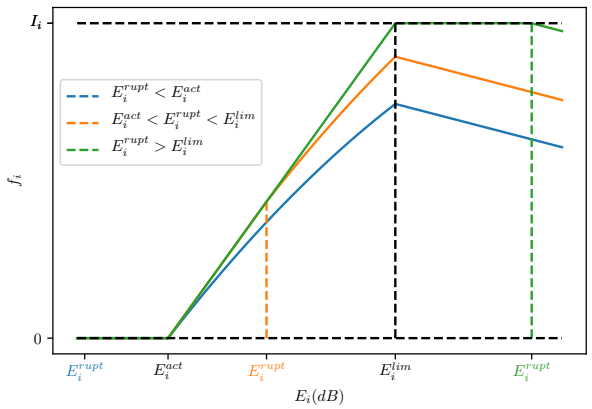


FIGURE 1 – Évolution des fonctions f_i en fonction des niveaux E_i pour trois situations distinctes.

évaluation [11] pour choisir méthodiquement les branches à résoudre et converger relativement rapidement vers la solution globale du problème. Chaque sous problème répond aux critères de résolution du problème continu convexe d'allocation non-linéaire sous contraintes et est donc directement solvable en utilisant la méthode du multiplicateur de Lagrange [12] (Section 3.1).

Le nombre d'itérations nécessaire pour la résolution du problème est trop important pour être effectué en temps réel. Par contre, l'optimisation ne dépendant que des niveaux du spectre équivalent de masquage Z_i, du seuil d'audition de l'auditeur T_i et du niveau de référence de la parole en dB(A), il est donc possible de calculer les spectres équivalents optimaux en avance si les profils sont connus et c'est ce que nous faisons dans la suite.

3.3 Interprétation des résultats

Avant d'étudier les résultats, il est important de vérifier l'hypothèse 1. Considérons par l'absurde que l'hypothèse n'est pas vérifiée et donc que $E_i > N_i + 24\,\text{dB}$. Nous avons déjà vu que l'optimisation du SII imposait $E_i \leq E_i^{lim} = Z_i + 15\,\text{dB}$. En combinant ces deux inégalités, on obtient $Z_i > N_i + 9\,\text{dB}$ et on peut voir sur la figure 2a que ce n'est pas le cas pour notre bruit automobile considéré. Ainsi l'hypothèse 1 est vérifiée et l'optimisation est exacte.

Afin d'obtenir une visualisation complète du processus d'optimisation pour un bruit et un auditeur donnés, il est possible de calculer les spectres équivalents optimaux pour plusieurs niveaux du signal de référence S^{dBA}. Prenons un intervalle allant du niveau minimal pour obtenir un SII non nul S_{min}^{dBA}, au niveau nécessaire pour obtenir un SII maximum S_{max}^{dBA}. Les formules de ces niveaux sont respectivement exprimées par les équations suivantes :

$$S_{min}^{dBA} = \min_i(E_i^{act} + H_i + \Delta freq_i)\,, \tag{12}$$

$$S_{max}^{dBA} = 10 \cdot \log(\sum_{i=1}^{i^{max}} 10^{(E_i^{lim}+H_i+\Delta freq_i)/10})\,. \tag{13}$$

Les résultats de l'optimisation obtenus pour un normo-entendant dans notre bruit automobile grande vitesse sont visibles figure 2b. Une décomposition fréquentielle en bandes de tiers d'octaves a été choisie arbitrairement.

Le processus d'optimisation semble cohérent : plus le niveau du signal S^{dBA} est grand, plus la répartition énergétique active des bandes. Les premières bandes a recevoir de l'énergie sont celles dont l'énergie d'activation E_i^{act} est faible i.e. où le bruit est peu présent. En effet, comme on peut le voir sur la figure 2, la répartition énergétique suit de près l'inverse du profil du spectre équivalent de masquage. On alloue alors de l'énergie aux bandes en fonction de leur rendement qui évolue à cause de la nature non-linéaire de leurs fonctions f_i et de la nature exponentielle de leurs fonctions g_i. Les différents facteurs de pondération introduits par la FIB, les h_i et les $\delta freq_i$, vont aussi conditionner l'allocation énergétique en donnant plus ou moins de rentabilité aux bandes.

La figure 3 représente l'évolution du SII pour un signal de parole moyen à différent niveau comparée à celle pour des signaux optimaux de même niveau en dB(A), mais aussi en dB(Z), i.e. sans pondération ($H_i = 0\text{dB}$). On reconnait les allures de sigmoïdes récurrentes aux courbes d'intelligibilité. Bien que moins importante qu'à dB(Z) constant, l'amélioration du SII à dB(A) constant est toujours notable, surtout dans la zone [30 dB, 70 dB]. On obtient une augmentation maximum de 16% d'intelligibilité théorique pour un niveau de signal à 58 dB.

3.4 Traitement de la parole

L'obtention des spectres équivalents optimaux est une étape importante dans notre approche et nous détaillons ci-dessous comment ils sont exploités pour traiter efficacement les signaux de parole afin d'améliorer leur intelligibilité. Nous avons exploré deux stratégies : une approche locale qui consiste à appliquer un égaliseur dynamique afin d'optimiser le SII sur de courtes fenêtres du signal, et une approche globale qui consiste à appliqué un égaliseur fixe calculé à partir des spectres équivalents long terme.

L'optimisation locale contraint, à chaque instant, le spectre équivalent court terme du signal à correspondre au spectre équivalent optimal. Cependant, si le signal n'a pas d'énergie, i.e. d'information, dans les "bandes optimales", ce sera le bruit qui sera amplifié pour atteindre les niveaux d'énergie optimaux. Même en utilisant des fenêtres de plus en plus larges pour lisser le traitement, la qualité du signal est largement impactée et l'intelligibilité n'en est que réduite.

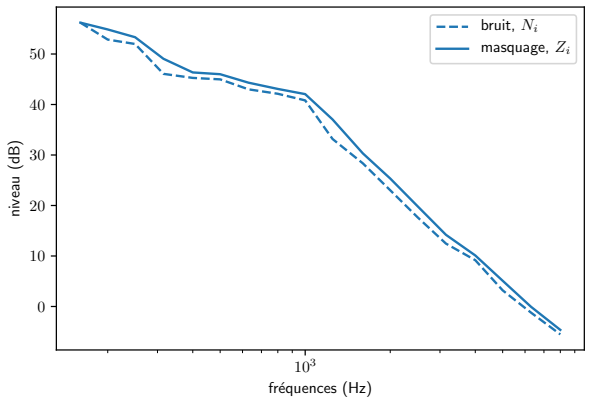
L'optimisation globale consiste à calculer un filtre qui permet de passer du spectre équivalent long terme du signal vers le spectre équivalent optimal. Pour généraliser le plus possible on considère que nous n'avons pas, ou difficilement, accès au spectre équivalent long terme de parole du locuteur, ce qui est le cas lors d'une discussion par exemple. Le filtre est donc calculé à partir d'un spectre équivalent de parole de référence de niveaux U_i' normalisé au bon niveau perçu S^{dBA} comme indiqué par l'équation suivante :

$$U_i' = U_i - 10 \cdot \log(\sum_{i=1}^{i^{max}} h_i \cdot \delta freq_i \cdot 10^{U_i/10}) + S^{dBA}\,. \tag{14}$$

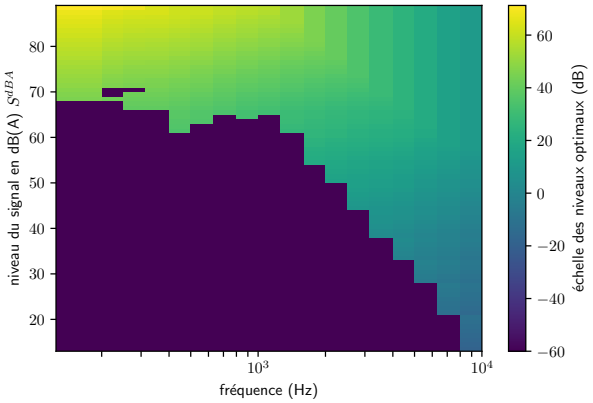
Le calcul des niveaux du filtre équivalent F_i en dB est une simple soustraction comme indiqué par l'équation suivante :

$$F_i = E_i^{opt} - U_i'\,. \tag{15}$$

Ce filtre constant est alors appliqué sur l'intégralité du signal et c'est cette approche, plus douce, qui a été retenue pour traiter nos signaux.



(a) Spectre équivalent de bruit et spectre de masquage correspondant.



(b) Spectres équivalents de parole optimaux à différents niveaux.

FIGURE 2 – Résultats de l'optimisation du SII pour un bruit stationnaire de type automobile à grande vitesse. On considère un auditeur normo-entendant, ou du moins avec des seuils d'audibilité inférieurs aux niveaux de masquage i.e. $T_i \leq Z_i$, ce qui donne $D_i = Z_i$ d'après l'équation 1. La figure (a) présente les niveaux du spectre équivalent de bruit N_i et ceux du spectre équivalent de masquage correspondant Z_i. La figure (b) présente les niveaux des spectres équivalents optimaux pour plusieurs niveaux du signal S^{dBA}.

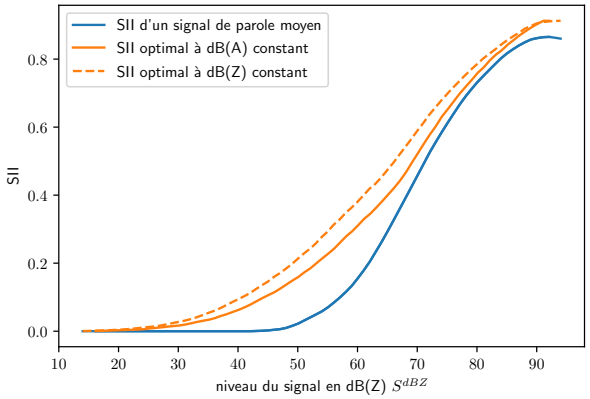


FIGURE 3 – Comparaison du SII calculé à partir d'un spectre de parole moyen, d'un spectre de parole optimal de même niveau en dB(A) et d'un spectre de parole optimal de même niveau en dB(Z). Le spectre de parole moyen est issu de la norme ANSI/ASA S3.5 normalisé à différents niveaux S^{dBZ}.

4 Tests subjectifs

Dans cette partie, nous détaillons les tests subjectifs mis en place pour valider notre approche et nous discutons des résultats obtenus.

4.1 Protocole

4.1.1 Stimuli et présentation

Le matériel vocal choisi suit les recommandations du *Hearing In the Noise Test* (HINT) [13] et est construit avec des phrases au vocabulaire standard représentatives d'un discours classique. Ce sont des phrases tirées du HINT franco-canadien [14] et des phrases de Fournier enregistrées par le Collège National d'Audioprothèse [15]. Le corpus est prononcé par un unique locuteur (homme) de sorte à pouvoir comparer les résultats.

Le type de bruit utilisé classiquement dans les tests d'écoute est synthétisé à partir d'un spectre de parole moyen appelé *Long Terme Average Speech Spectrum* (LTASS) généré à partir de l'ensemble des signaux enregistrés par le locuteur. Cela a l'intérêt de proposer un RSB à peu près constant dans toutes les bandes de fréquence. Dans notre cas, le bruit d'étude est imposé par notre application automobile. Le bruit automobile grande vitesse utilisé est le même que précédemment, il a été enregistré en conditions réelles avec une tête acoustique (HMS IV, HEAD acoustics *GmbH*) et est retransmis en haute fidélité à l'auditeur au moyen de l'interface dédiée et d'un casque audio (Sennheiser^{TM} HD 650) calibré.

4.1.2 Équilibrage des phrases

Une étape cruciale dans la mise en place d'un test d'écoute est l'équilibrage du matériel vocal. Cela consiste à s'assurer que tous les éléments vocaux ont la même difficulté à être répétés dans le bruit considéré. C'est une étape qui demande généralement plusieurs itérations de corrections sur des groupes de personnes différentes [13, 14]. Effectuer cette étape sur l'ensemble des bruits automobiles sur lesquelles nous travaillerons durant notre étude n'est pas concevable, c'est pourquoi nous avons décidé d'effectuer un équilibrage moyen sur un bruit synthétique dont le spectre correspond à la moyenne, en décibels, de bruits automobiles qui décrivent des situations à différentes vitesses, avec ou sans pluie. La méthode d'équilibrage employée est très proche de celle proposée par Nielsen et al. [16] pour la mise en place du HINT danois [17]. Elle a été réalisée en deux itérations avec deux groupes de six sujets chacun. A l'inverse des méthodes classiques, cette méthode fait intervenir le jugement subjectif des sujets dans l'équilibrage ce qui permet, théoriquement, d'obtenir de meilleurs résultats avec moins de sujets.

A l'issue de cette phase d'équilibrage nous avons créé dix listes de vingt phrases. Deux listes sont composées uniquement de phrases nécessitant un équilibrage trop important (correction supérieure à 3 dB), elles serviront de listes d'entrainement pour éviter que l'effet d'apprentissage n'impacte nos résultats. Les huit autres listes sont composées de manière à minimiser la variance des corrections, tout en essayant de conserver les phrases des listes originales ensemble car ces dernières sont équilibrées phonétiquement.

4.1.3 Procédure adaptative d'estimation du seuil d'intelligibilité vocale

Le seuil d'intelligibilité vocale, ou *Speech Reception Threshold* (SRT), correspond au niveau de présentation le plus bas pour que 50% du matériel vocal soit répété. La procédure adaptative d'estimation du SRT qui a été choisie est celle détaillée dans les travaux de Brand et al. [18] qui est une généralisation de la méthode adaptative de Hagerman et Kinnefors [19]. Elle consiste à présenter une première phrase à un niveau très bas relativement au SRT supposé et d'augmenter progressivement le niveau jusqu'à ce que le sujet soit capable de répéter au moins un mot. Les 19 phrases suivantes ne sont présentées qu'une seule fois et leur niveau dépend de la réponse donnée par le sujet pour la phrase précédente. Le pas en dB qui conditionne le changement de présentation d'une phrase sur l'autre est donné par l'équation suivante :

$$pas = \frac{10}{1.41^r} \cdot (0.5 - prec)\,. \tag{16}$$

L'indice r correspond au nombre d'inversions du niveau de présentation ayant eu lieu durant la procédure i.e. le nombre de changements de signe du *pas*. La variable *prec* correspond au score de la phrase précédente i.e. le pourcentage de mots correctement répétés. La répétition d'un mot dont la phrase est présentée à un niveau S^{dB} est considérée comme une épreuve de Bernoulli indépendante de probabilité *proba* décrite par l'équation suivante :

$$proba(S^{dB}, SRT, pente) = \frac{1}{1 + \exp(4 \cdot pente \cdot (SRT - S^{dB}))}\,. \tag{17}$$

Cette probabilité correspond à une sigmoïde centrée en SRT et dont la dérivée au point d'inflexion est notée *pente*. Le niveau de présentation ainsi que la réussite pour la répétition de chaque mot sont sauvegardés durant la présentation de la liste. A la fin de celle ci, les paramètres SRT et *pente* sont alors estimés en utilisant un estimateur du maximum de vraisemblance sur le processus de Bernoulli résultant. Cette procédure permet d'obtenir une estimation du SRT dont l'erreur est inférieure à 1 dB [18].

4.2 Résultats

Les tests ont été réalisés sur une population de 13 normo-entendants dont l'acuité auditive a été vérifiée par un examen d'audiométrie tonale [20]. Les tests se déroulent alors de la façon suivante. La procédure d'estimation du SRT est appliquée sur les deux listes d'entrainement afin d'habituer le sujet à la tâche. On applique ensuite la procédure à deux listes équilibrées en conservant la voix originale pour l'une et en la traitant pour l'autre. Il est important que les listes, les phrases et le traitement soient présentés de façon pseudo-aléatoire afin d'éviter l'effet d'ordre.

Les résultats des tests d'intelligibilité ont été synthétisés dans le tableau 1 et par des diagrammes en boîte de Tukey figure 4. On remarque une nette amélioration du SRT avec une diminution moyenne d'environ 6,9 dB. En notant μ la moyenne des différences entre le SRT estimé sur la voix traitée et celui estimé sur la voix originale, on pose l'hypothèse nulle : $\mu \geq 0$. Un test de Student vient rejeter cette hypothèse de façon très significative avec une valeur-p très petite devant les niveaux de signification classique ($\alpha = 0,05$ ou même $\alpha = 0,01$). Ainsi les SRT estimés

Tableau 1 – Tableau des résultats des tests.

	voix originale	voix traitée
SRT moyen (dB)	57,88	51,01
SRT écart type (dB)	1,07	1,84
valeur-p	1,42E-08	
intervalle de confiance à 95%	]-∞, -5,90[	

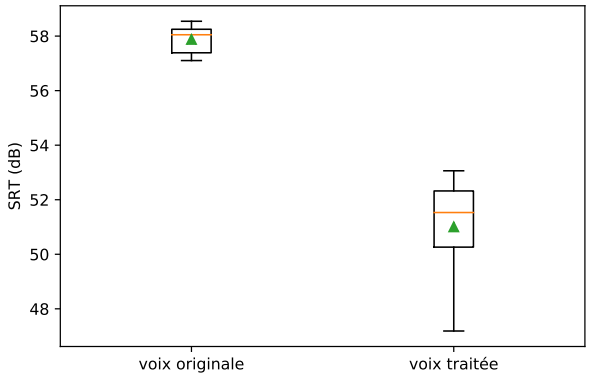


Figure 4 – Profil statistiques des SRT estimés pour les signaux non traités et traités par notre algorithme.

à partir du matériel vocal traité sont significativement et strictement inférieurs à ceux estimés à partir du matériel original. La borne supérieure de l'intervalle de confiance confirme nos observations car au moins 95% des différences entre SRT devraient être d'une amplitude d'au moins 5,9 dB.

On remarque aussi une augmentation relative de 72% de l'écart type des SRT pour les voix traitées par rapport à celui pour les voix originales. Cela indique que des paramètres liés au matériel vocal ou aux sujets semblent influer sur les performances du traitement.

Une question ouverte à chaque sujet concernant leur ressenti sur le naturel et la qualité de la voix n'a pas mis en évidence de différences majeures entre celle traitée et non traitée. Il semblerait que leur inaptitude à répéter certaines parties du matériel était seulement liée au RSB trop faible et non pas à cause d'une quelconque distorsion de la voix.

4.3 Interprétations

En ré-allouant l'énergie vers les hautes fréquences où le bruit est beaucoup moins présent, nous favorisons l'émergence du signal au détriment des composantes basses fréquences qui, de toute façon, seraient masquées par le bruit trop important dans cette zone. Cette procédure permet d'améliorer grandement l'intelligibilité tout en conservant un niveau perçu constant.

En écoutant la voix traitée sans bruit, on est vite dérangé par l'absence de basses fréquences. Par contre, une fois la voix noyée dans le bruit, le retour des sujets semblerait indiquer que le traitement impacte peu le naturel et la qualité de celle-ci. Des tests de qualité restent tout de même nécessaires afin de quantifier et valider ces impressions.

5 Conclusion

En appliquant un égaliseur de fréquences à des signaux de parole de façon à maximiser le SII à intensité perçue constante, nous avons mis en évidence une approche permettant d'améliorer l'intelligibilité théorique de ces signaux diffusés dans un environnement automobile bruité spécifique. Le SII étant un critère mathématique objectif, des tests subjectifs rigoureux ont été mis en place afin de valider l'approche auprès d'une population de 13 sujets normo-entendants. Les résultats obtenus montre une amélioration très significative de l'intelligibilité avec une diminution moyenne du SRT d'environ 6,9 dB pour le matériel traité par notre algorithme.

Par contre, la généralisation de ces résultats sur d'autres profils de bruit nécessite des tests supplémentaires. Il serait notamment intéressant de travailler sur des spectres plus large bande et non stationnaires, comme en présence de pluie, ou en phase d'accélération. L'utilisation d'outils déjà existants pour compléter cette approche peut aussi être intéressante avec, par exemple, l'utilisation de la compression dynamique qui est une méthode très utilisée dans les aides auditives.

Références

[1] K. Nathwani, G. Richard, B. David, P. Prablanc, and V. Roussarie, "Speech intelligibility improvement in car noise environment by voice transformation," *Speech Communication*, vol. 91, pp. 17–27, 2017.

[2] G. Kim and P. C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 8, pp. 2080–2090, 2010.

[3] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility-enhancing speech modifications : the hurricane challenge." in *Interspeech*, 2013, pp. 3552–3556.

[4] A. ANSI, "S3. 5-1997, methods for the calculation of the speech intelligibility index," *New York : American National Standards Institute*, vol. 19, pp. 90–119, 1997.

[5] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations," in *Signal Processing Conference, 2010 18th European*. IEEE, 2010, pp. 1919–1923.

[6] C. H. Taal, J. Jensen, and A. Leijon, "On optimal linear filtering of speech for near-end listening enhancement," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 225–228, 2013.

[7] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *The Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 318–326, 1980.

[8] K. S. Rhebergen, N. J. Versfeld, and W. A. Dreschler, "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," *The Journal of the Acoustical Society of America*, vol. 120, no. 6, pp. 3988–3997, 2006.

[9] J. Ma, Y. Hu, and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387–3405, 2009.

[10] J. M. Kates and K. H. Arehart, "Coherence and the speech intelligibility index," *The journal of the acoustical society of America*, vol. 117, no. 4, pp. 2224–2237, 2005.

[11] S. Ağralı and J. Geunes, "Solving knapsack problems with s-curve return functions," *European Journal of Operational Research*, vol. 193, no. 2, pp. 605–615, 2009.

[12] K. M. Bretthauer and B. Shetty, "The nonlinear knapsack problem–algorithms and applications," *European Journal of Operational Research*, vol. 138, no. 3, pp. 459–472, 2002.

[13] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, 1994.

[14] V. Vaillancourt, C. Laroche, C. Mayer, C. Basque, M. Nali, A. Eriks-Brophy, S. D. Soli, and C. Giguère, "Adaptation of the hint (hearing in noise test) for adult canadian francophone populations : Adaptación del hint (prueba de audición en ruido) para poblaciones de adultos canadienses francófonos," *International Journal of Audiology*, vol. 44, no. 6, pp. 358–361, 2005.

[15] C. N. d'Audioprothèse. (2006) Cd d'audiométrie vocale. [Online]. Available : http://www.college-nat-audio.fr/listes-cd-audiometrie-vocale.html

[16] J. B. Nielsen and T. Dau, "Development of a danish speech intelligibility test," *International journal of audiology*, vol. 48, no. 10, pp. 729–741, 2009.

[17] ——, "The danish hearing in noise test," *International journal of audiology*, vol. 50, no. 3, pp. 202–208, 2011.

[18] T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2801–2810, 2002.

[19] B. Hagerman and C. Kinnefors, "Efficient adaptive methods for measuring speech reception threshold in quiet and in noise," *Scandinavian audiology*, vol. 24, no. 1, pp. 71–77, 1995.

[20] B. S. of Audiology. (2012) Recommended procedure for pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. [Online]. Available : http://www.thebsa.org.uk