

## L'intelligibilité de la parole en présence d'un écho et de bruit

F. SANTON, A. MARCHIONI et P. SUSINI

*Laboratoire de Mécanique et d'Acoustique, CNRS, 31 Chemin J. Aiguier, 13402 Marseille cedex 20, France*

**Abstract:** The intelligibility of speech is a good example of the detection of a real signal in noise. The presence of an echo of the original signal or direct sound (by analogy with the room acoustics) has an obvious interest, because an early echo increases the intelligibility, while a delayed echo deteriorates it.

Recent works have shown that the auditory integration of an echo with a short delay is not complete in the presence of noise. These measurements were performed using intelligibility scores.

We have measured the Speech Reception Threshold (SRP) corresponding to a given intelligibility score for various phonetic transitions and various delays. The results show irregular variations of the SRP according to the delay and phonetic transitions.

### 1. INTRODUCTION

Le problème de l'intelligibilité de la parole en présence d'un écho a été envisagé depuis de longues années. Après les célèbres travaux de HAAS (1950), LOCHNER et BURGER (1961) ont proposé des courbes d'intégration auditive, qui ont été appliquées en particulier au cas de l'acoustique des salles.

Plus récemment, certains auteurs ont envisagé l'intelligibilité en présence d'un écho mais également en présence de bruit (SOULODRE et al, 1989) puis PARIZET et POLACK (1992). Cette disposition présente l'avantage d'une conception plus réaliste de l'environnement sonore. Les auteurs susnommés ont pu mettre en évidence que la présence de bruit entraîne une intégration seulement partielle de l'écho.

En fait, l'origine de notre travail est liée à la mesure de l'amélioration de la détection d'un signal supposé utile dans le bruit, étude effectuée dans le cadre d'un contrat de notre équipe avec la DCN (1989-92). Nous avons décidé d'utiliser non pas des mesures d'intelligibilité, mais des mesures du Seuil de Réception de la Parole (SRP), que nous avons utilisé dans une étude antérieure (SANTON, 1986).

Le matériau phonétique retenu était les paires minimales de PECKELS et ROSSI (1973). Nous avons l'intention de mesurer le SRP sur un ensemble de paires minimales présentées les unes à la suite des autres. En fait, l'absence de variation du SRP en fonction du retard, dans la plage considérée, c'est à dire au-dessous de 60 ms, nous a amenés à considérer différentes paires individuellement, correspondant à certains traits phonétiques.

### 2. DESCRIPTION DE L'EXPERIENCE

#### 2.1. Locuteurs, auditeurs et matériau phonétique

Nous avons utilisé trois locuteurs, deux locuteurs masculins et une voix de synthèse. Les paires minimales étaient enregistrées en chambre anéchoïque sur un magnétophone numérique. Les locuteurs s'efforçaient de prononcer les deux mots d'une même paire sur le même ton. Les mots ainsi enregistrés étaient transférés dans la mémoire de l'ordinateur PDP 11.73. La fréquence d'échantillonnage était fixée à 12,5 kHz, pour des raisons de limitation du système informatique. Ils étaient diffusés sous écouteurs en écoute diotique. Le bruit de masque avait un niveau fixe (en général 50 dB SPL).

Comme matériau phonétique, nous avons utilisé les paires minimales. Dans ces mots, c'est la consonne de tête qui varie. Nous avons également utilisé des mots ou logatomes formés à partir de ces paires prises à l'envers. Il n'était pas dans nos intentions de faire des tests exhaustifs sur toutes les configurations phonétiques possibles, mais d'envisager seulement des exemples particuliers qui nous ont permis de dégager des profils caractéristiques de l'évolution du SRP en fonction du retard.

## 2.2. Présentation de la chaîne d'expérimentation

Le retard de l'écho, de 0 à 60 ms est donné par une ligne à retard programmable. L'écho et le son direct ont le même niveau. La source de bruit aléatoire est fournie par l'analyseur HP 3561 A, puis filtrée à 8 kHz par un filtre KEMO. La source de parole est fournie par l'ordinateur: après conversion numérique / analogique, le spectre de chaque mot est lissé puis coupé à sa fréquence maximale par un filtre de synthèse.

## 2.3. Méthode utilisée

Nous avons choisi une méthode adaptative du type "3 down - 1 up", c'est à dire que le niveau du son direct, ainsi que celui de l'écho, est diminué de 2 dB après trois réponses correctes et augmenté de 2 dB après une réponse fautive. Le seuil correspond à 79 % de réponses correctes. Ce seuil correspond à une intelligibilité de 58 % pour une liste de mots "ouverte". Chaque mesure est en fait la moyenne de trois seuils partiels obtenus à partir de six inversions: on ne garde d'ailleurs que les quatre dernières pour le calcul d'un seuil partiel. L'écart-type relatif au trois seuils partiels permet d'estimer la cohérence des résultats. Nous avons pris l'habitude de rejeter les résultats correspondant à un écart-type supérieur à 3 dB.

## 3. ROLE D'UN ECHO EN ECOUTE DICHOTIQUE

### 3.1. Remarques préliminaires

Comme pour un retard nul, son direct et écho sont en phase, ils s'additionnent de 6 dB. Par contre, pour les autres retards ils sont incohérents et s'additionnent de 3 dB. On peut donc relever de 3 dB le seuil  $N_s$  obtenu à 0 ms pour le comparer aux autres valeurs.

### 3.2. SRP pour les paires minimales, en présence d'un bruit de 50 dB

Les six paires minimales que nous avons envisagées sont évoquées par PARIZET et POLACK, et caractéristiques de différents traits phonétiques.

#### a - Résultats avec le locuteur 1

Les courbes relatives à deux paires de mots prononcés par le premier locuteur et aux trois auditeurs (dont le locuteur lui-même) sont données par les figures 1 et 2.

On voit que les diverses paires minimales ne donnent pas un SRP ayant la même évolution en fonction du retard. Certaines oppositions présentent une progression régulière, plus ou moins rapide du SRP.

Par exemple, pour la paire six / fils, l'augmentation assez rapide du SRP montre une intégration seulement partielle de l'écho. On retrouve bien les résultats de SOULODRE et al ou de PARIZET et POLACK.

Pour d'autres paires comme limbe / nimbe, on constate un effet très particulier: il s'agit du maximum local obtenu pour le seuil  $N_s$  quant le retard est de 10 ms. On a alors affaire à une intégration partielle de l'écho, suivie d'ailleurs par un minimum qui correspond à une intégration particulièrement bonne et amènerait une meilleure intelligibilité que le retard nul.

D'autre part, on ne constate pas de remontée du seuil  $N_s$  au delà d'un retard de 35 ou 40 ms. Cet effet est sans doute lié à la situation en tête du mot de la consonne variable.

Les variations du seuil moyen calculé à partir des six paires pour les trois auditeurs considérés montrent une allure proche de celle envisagée a priori. On peut noter l'effet d'intégration partielle de l'écho de façon très nette.

#### b - Résultats pour le locuteur 2

Les figures 3 et 4 donnent deux exemples des courbes obtenues avec le deuxième locuteur. En ce qui concerne les deux paires envisagées (limbe / nimbe et six / fils), on trouve des résultats comparables à ceux obtenus pour l'auditeur 1, en particulier un pic accentué pour deux auditeurs lors d'un retard de 10 ms dans le cas de la première paire, et au contraire un seuil presque constant pour la seconde.

### *c - Résultats pour le locuteur 3*

Nous avons utilisé, pour la troisième série de tests, un synthétiseur vocal du commerce. Nous espérons nous libérer complètement des indices parasites (intonation, longueur variable des mots) rencontrés chez un locuteur humain. En fait, si effectivement nous n'avons plus rencontré ce type de problème, nous avons eu affaire à des mots peu naturels et dont l'intelligibilité intrinsèque laissait quelque peu à désirer, étant donné que la bande passante était limitée à 4 kHz. En particulier, certaines consonnes (les sifflantes en particulier) ont souffert de ce filtrage drastique. Les mesures ont demandé un effort d'attention très soutenu aux auditeurs...

Les résultats sont donnés par les figures 5 et 6. On remarque une assez bonne ressemblance avec les résultats obtenus avec des locuteurs humains: cependant, certaines paires passent dans le groupe des "paires à pic", en particulier la paire lard / dard qui montre un maximum du SRP pour 5 ou 10 ms suivant les auditeurs.

On remarque d'autre part un niveau pour le SRP sensiblement plus élevé que dans le cas de locuteurs humains, effet lié sans doute au filtrage évoqué plus haut.

### *d - Remarque sur le pourcentage d'erreurs lié à chaque mot*

Dans certaines oppositions, les intelligibilités respectives des deux mots mesurées sur l'ensemble du test peuvent être comparables, pour d'autres l'intelligibilité d'un mot est faible par rapport à l'autre, c'est-à-dire que l'un des mots est plus souvent confondu avec son opposé que l'inverse.

Par exemple, pour la paire bile / pile, il apparaît une répartition à peu près égale des erreurs, sauf pour un retard de 5 ms: dans ce cas, le pourcentage d'erreurs lié à bile (47%) est nettement plus élevé que celui lié à pile (5%). Or ce retard correspond à un maximum de la courbe du SRP.

Par contre, pour la paire limbe / nimbe, il y a beaucoup plus d'erreurs pour le premier mot que pour le second: or pour le retard correspondant au maximum de Ns, les deux pourcentages sont comparables.

## **4. ESSAI D'INTERPRETATION**

### **4.1. Usage des spectrogrammes**

Nous avons réalisé à l'aide d'un microordinateur quelques figures de type spectrographique qui illustrent assez bien l'évolution de la structure constituée par le son direct et l'écho lorsque le retard varie. Il apparaît de façon nette que si le mot "nimbe" est étiré pour le retard de 60 ms, sa structure reste peu altérée. Par contre, c'est pour un retard de 10 ms, qui correspond à un maximum du SRP et à une augmentation des erreurs que la structure montre une détérioration nette, allant à une quasi-disparition de la ligne mélodique ainsi qu'à une apparition hachée des différents harmoniques. Par contre la structure du mot "limbe" reste résistante, pour les retards considérés.

### **4.2. Effet de coloration**

Il apparaît que pour certaines oppositions phonétiques, un maximum du SRP se produit pour une réflexion précoce de l'ordre de 5 ms. C'est justement dans cette région que l'effet de coloration est le plus sensible (ATAL et al, 1962, SANTON, 1991). Il y a par conséquent tout lieu de penser que l'importance de la coloration pour un tel retard justifie la baisse de l'intelligibilité.

## **5. CONCLUSION**

Nous avons montré que, contrairement aux prévisions habituelles, l'intégration auditive pour certains phonèmes passe par un maximum local lorsque le retard de l'écho varie. Cet effet est particulièrement sensible dans la zone des retards courts, de 5 à 10 ms où l'on sait que la coloration est intense, et pour des mots dont la consonne initiale est variable.

Dans le cas où la consonne se trouve à la fin du mot, cet effet est beaucoup moins sensible.

**Remerciements:** Ce travail a eu pour origine un contrat passé entre le L.M.A et le D.C.N que nous tenons à remercier. Nous remercions également le Laboratoire de Phonétique d'Aix-en-Provence qui a bien voulu nous communiquer une bande portant une série de paires minimales.

### **Références:**

ATAL B.S, SCHOEDER M.R et KUTTRUFF H: Perception of coloration in filtered gaussian noise. 4° CIA, Copenhague, 1962

HAAS H: Uber den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache. *Acustica* 1 (1951) 49-58

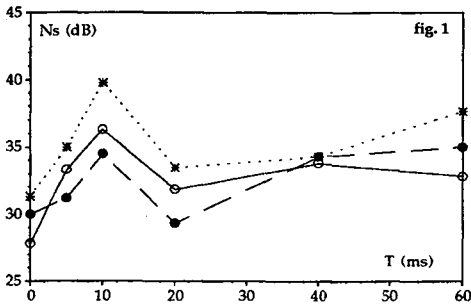
LOCHNER J.P.A et BURGER J.F: The subjective masking of short delayed echoes. *Acustica* 8 (1958) 1-10  
 PARIZET E et POLACK J.D: The influence of an early reflection upon the speech intelligibility in the presence of a background noise. *Acustica* 77 (1992) 21-30

PECKELS J.P et ROSSI M: Le test de diagnostic par paires minimales. *Revue d'acoustique* 27 (1973) 245-262

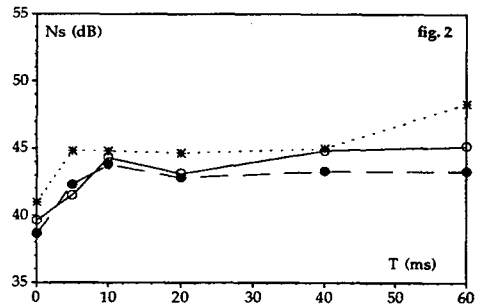
SANTON F: Intelligibilité de la parole masquée et rôle de la direction de la source masquante. *Acustica* 61 (1986) 67-74

SOULODRE G.A, POPPLEWELL N et BRADLEY J.S: Combined effects of early reflections and background noise on speech intelligibility. *Journ. of Sound and Vibr.* 135 (1989) 123-133

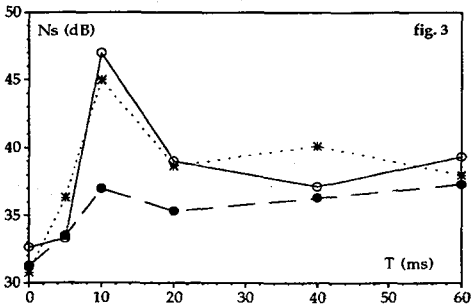
limbe/nimbe, Nb = 50 dB SPL



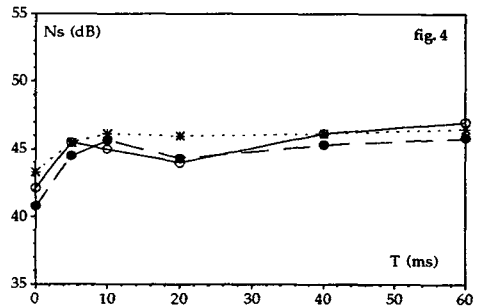
six/fils, Nb = 50 dB SPL



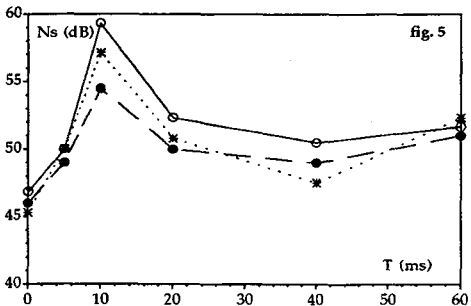
limbe/nimbe, Nb = 50 dB SPL



six/fils, Nb = 50 dB SPL



limbe/nimbe, Nb = 50 dB SPL



six/fils, Nb = 50 dB SPL

