# Évaluation de l'intelligibilité de la parole dans les établissements recevant du public sonorisés.
## Les méthodes normalisées sont-elles adaptées ?

*Dans le cadre de la nouvelle norme européenne concernant les systèmes électroacoustiques de secours, cet article présente les différentes méthodes pour caractériser l'intelligibilité dans les lieux réverbérants et bruyants, tels que les établissements recevant du public (ERP).*

**Christophe Lambourg,**
*Signal Développement,*
*Bâtiment @2,*
*avenue du Futuroscope,*
*Téléport 1,*
*86360 Chasseneuil du Poitou,*
*tél. : 05 49 49 64 53,*
*fax : 05 49 49 64 20,*
*e-mail : info@signal-dev.fr*

La grande nouveauté de la norme européenne EN 60849 de 1998 [1], qui traite des systèmes électroacoustiques de secours, tient dans le fait qu'elle spécifie non seulement des obligations de moyens, mais également de résultats quantitatifs : elle fixe en effet une valeur minimum d'intelligibilité définie sur une échelle commune (CIS en anglais pour Common Intelligibility Scale) regroupant des indicateurs de natures diverses. Le choix du critère d'évaluation n'étant pas imposé, il est essentiel de bien connaître les mécanismes mis en œuvre dans le processus de transmission et de compréhension de la parole, les facteurs susceptibles de les affecter, et la manière dont ils sont évalués par les différents critères proposés.

La perception de la parole est un processus complexe qu'il est difficile de quantifier de façon globale. La correspondance entre les critères proposés par la norme est présentée sous la forme d'une superposition de courbes de régression issues d'analyse statistique. Il existe une dispersion importante entre eux, qui peut s'expliquer par la diversité des informations qui sont exploitées pour la compréhension d'un message parlé. En particulier, notre cerveau possède la remarquable faculté de reconstruire des données manquantes - phonèmes, mots ou portions de phrases - à partir du contexte dans lequel elles s'insèrent. En plus de l'information physique qui peut être dégradée par le canal de transmission, il faut donc également considérer les informations grammaticales, syntaxiques et sémantiques du signal de parole. C'est la raison pour laquelle il n'y a pas de méthode de référence pour évaluer l'intelligibilité, à laquelle on pourrait comparer les différents critères existants : par exemple, le résultat d'un test de reconnaissance de mots courts isolés ne permet pas de prédire quelle sera exactement l'intelligibilité dans des conditions d'écoute d'un message signifiant (conditions réalistes), les informations disponibles pour la compréhension étant différentes dans les deux cas.

L'exploitation des données sémantiques contenues dans un message perçu par un auditeur dépend fortement de sa culture et de ses préoccupations du moment, ce qui rend illusoire la prédiction du degré de compréhension du même message par un auditeur choisi au hasard. C'est pourquoi une approche statistique de l'intelligibilité s'avère indispensable. Pour réduire la dispersion inhérente aux facteurs humains, la solution généralement adoptée consiste à considérer uniquement les causes physiques de détérioration de l'intelligibilité, soit en se limitant à une mesure des « performances » du canal de transmission, soit en effectuant des tests de reconnaissance de mots dans lesquels l'influence du contexte sémantique et syntaxique est limitée au maximum.

Les premières études menées sur l'intelligibilité ont été motivées par le développement des systèmes de téléphonie dans les années 30. Les travaux initiés par Fletcher & Harvey, puis repris par French & Steinberg ont conduit au premier indicateur entièrement calculé : l'Indice d'Articulation [2], qui fait l'objet d'une norme ANSI datant de 1969 (S 3.5). Ce critère apparaît dans la norme EN 60849.

Ces études ont en particulier permis d'établir les poids des contributions de chaque bande de fréquences à l'intelligibilité. L'AI est avant tout destiné à chiffrer l'effet du masquage par des bruits parasites, et se trouve mal adapté pour évaluer l'intelligibilité de messages diffusés dans des espaces réverbérants.

Dans les années 70, les travaux menés par Peutz ont montré l'influence primordiale de la compréhension des consonnes sur l'intelligibilité globale. Il a développé un critère basé sur un test de reconnaissance de mots simples : le pourcentage de pertes d'articulation des consonnes (% d'Alcons, [3]) qui est également cité dans la norme EN 60849.

Le concept de fonction de transfert de modulation et l'Indice de Transmission de la Parole (STI), défini par Steeneken & Houtgast [4], permet en théorie de prendre en compte la plupart des formes d'altération physique du signal de parole. Popularisé depuis 1985 par le système de mesure du RASTI [5] commercialisé par Bruel & Kjaer, le développement récent des ordinateurs portables a depuis permis l'apparition de nombreux systèmes réalisant la mesure complète du STI. Depuis 1998, le STI et ses dérivés font l'objet d'une norme européenne [6].

Bien qu'ils ne soient malheureusement pas cités dans la norme EN 60849, il existe une autre classe d'indicateurs développés spécifiquement pour les espaces réverbérants, et qui apparaîtront probablement dans les futures normes concernant l'intelligibilité. Dans les années 60, Lochner & Burger ont mis en évidence l'influence sur l'intelligibilité des temps d'arrivée et de l'énergie relative des différents échos par rapport au premier front d'onde [7]. Ces travaux ont conduit à l'élaboration d'un critère basé sur le concept du rapport Énergie utile/Énergie nuisible, qui a été depuis repris et modifié par de nombreux auteurs (par exemple, [8-11]).

Par la suite, on s'intéressera tout particulièrement aux ERP, pour lesquels une réglementation appelant la norme EN 60849 sera mise en place prochainement. Après une présentation des principaux facteurs potentiels de détérioration de l'intelligibilité dans ce type d'espace, les deux classes de méthodes d'évaluation seront détaillées :

**- Les méthodes directes,** basées sur la réalisation de tests d'écoute de mots ou de phrases. Ces méthodes, généralement lourdes à mettre en œuvre, permettent de mesurer de façon précise l'influence de tous les facteurs physiques de détérioration du signal de parole.

**- Les méthodes indirectes,** basées sur la mesure de critères dits «objectifs». Elles permettent de s'affranchir de la réalisation de tests d'écoute, mais présentent en contre partie des limitations inhérentes à la diversité des facteurs qui doivent être pris en compte. Dans la norme EN 60849, le STI est le seul indicateur de ce type adapté aux grands ERP. Cependant, nous parlerons également des rapports Énergie Utile/Énergie Nuisible.

Pour finir, des exemples de corrélations entre ces critères ainsi que l'échelle commune d'intelligibilité proposée dans la norme seront présentés.

## Les facteurs de détérioration de l'intelligibilité dans les établissements recevant du public

La compréhension des messages parlés diffusés dans les ERP peut être altérée par plusieurs facteurs. Voici la liste des principaux, classés par ordre décroissant d'importance, en se limitant uniquement aux facteurs physiques :

**- Distorsion temporelle :** Les ERP sont souvent de grands espaces dans lesquels la réverbération est responsable d'une dégradation de l'intelligibilité, que l'on peut limiter à l'aide de traitements de correction acoustique ou en utilisant des sources électroacoustiques directives (enceintes colonnes). D'autre part, pour assurer un niveau sonore homogène, il est généralement nécessaire de mettre en œuvre un nombre important de sources électroacoustiques (système multi-diffusion), dont les contributions qui parviennent décalées dans le temps à l'auditeur (diaphonie) peuvent également entraîner une distorsion temporelle du signal de parole. Il faut donc trouver un compromis entre l'homogénéité de la couverture et la diaphonie, qu'il est possible de réduire en utilisant des lignes à retards et en optimisant le nombre et la directivité des sources électroacoustiques.

Dans les ERP tels que les gares, on observe souvent des variations d'intelligibilité avec le timbre de la voix du locuteur, qui peuvent s'expliquer par une dépendance en fréquence du temps de réverbération : dans ce type d'espace, une analyse en fréquence de la réverbération est donc nécessaire. D'autre part, l'importance des volumes ainsi que le caractère non-diffusant des matériaux typiquement utilisés (parois lisses en béton ou en verre) entraînent une durée élevée de la partie précoce de la réverbération, pendant laquelle la densité d'échos reste faible. Une approche purement statistique basée sur des hypothèses de champs diffus s'avère donc généralement insuffisante pour caractériser les phénomènes de distorsion temporelle qui nuisent à l'intelligibilité dans ce type d'espace.

**- Masquage par des bruits parasites :** la présence de bruits masquant, en particulier dans les bandes de fréquences essentielles pour la compréhension (500 Hz et 2 kHz) peut également entraîner une diminution importante de l'intelligibilité des annonces dans les espaces publics. La répartition inhomogène des sources de bruit et les fluctuations parfois élevées des niveaux sonores au cours de la journée rendent alors inenvisageable l'application d'un niveau de diffusion permettant de couvrir tous les bruits, qui serait extrêmement inconfortable en période calme. Lorsque la réduction de bruit à la source ou sur les voies de transfert n'est pas possible, une solution peut consister à utiliser des systèmes d'asservissement du niveau de diffusion des annonces aux variations de niveau de bruit ambiant.

**- Distorsion fréquentielle :** Dans le cas des systèmes de diffusion privilégiant à outrance le rendement acoustique sur la fidélité, les distorsions fréquentielles peuvent contribuer à la dégradation de l'intelligibilité. Cependant, on peut considérer que ce facteur est de moindre importance que les deux premiers, les systèmes

électroacoustiques modernes ayant généralement une bande passante suffisante pour couvrir la gamme de fréquences importante pour l'intelligibilité. Des distorsions fréquentielles peuvent cependant également provenir de la réponse acoustique de la salle.

**- Distorsions non-linéaires :** Ce facteur de dégradation intervient principalement pour les systèmes électroacoustiques à haut rendement, ou sous-dimensionnés par rapport au niveau de diffusion imposé. Il peut être également lié à une compression initiale du signal de parole diffusé ou à un effet Larsen.

On notera que ces facteurs sont souvent interdépendants.

Ainsi, le niveau de bruit est lié à la réverbération. D'autre part, la liste proposée ci-dessus n'est pas exhaustive. En particulier, il ne faut pas oublier l'influence du locuteur, dont la prosodie, le timbre de voix, la vitesse de locution ou la position par rapport au microphone peuvent avoir des effets sur l'intelligibilité. C'est la raison pour laquelle la SNCF généralise par exemple les systèmes automatiques de diffusion d'annonces enregistrées. Un autre facteur potentiellement influent sur l'intelligibilité est lié à l'écoute spatialisée et à la position des sources dans l'espace : notre système auditif possède en effet la capacité de réhausser le niveau d'un signal acoustique noyé dans le bruit, lorsqu'il provient d'une source dont la position est identifiée [12]. Ce processus pourrait donc introduire des différences de performances entre plusieurs systèmes, non mesurables à l'aide des indicateurs objectifs existants basés sur des mesures par microphone omnidirectif.

Les principales difficultés techniques rencontrées dans la sonorisation des ERP sont donc, d'une part, de réduire la distorsion temporelle du signal de parole introduite par la réverbération et la diaphonie, et d'autre part d'assurer un rapport Signal/Bruit suffisant en présence de bruits masquant. Pour évaluer correctement l'intelligibilité, on doit donc être capable de chiffrer l'influence de ces deux facteurs.

## Les tests d'intelligibilité

Depuis Fletcher, de nombreuses méthodes d'évaluation de l'intelligibilité basées sur le décompte de phonèmes, mots ou phrases correctement reconnues par un jury ont été proposées. Ces tests sont construits pour répondre à des exigences parfois contradictoires :

**- Validité :** le résultat du test doit permettre d'estimer l'intelligibilité dans des conditions réelles d'écoute de messages (quantité non-mesurable, voir l'introduction). Pour s'en rapprocher, une solution consiste à employer des listes de stimuli statistiquement représentatives du langage considéré.

**- Reproductibilité/Précision :** pour une configuration donnée, les résultats obtenus doivent varier faiblement d'un auditeur à un autre. Une bonne précision n'est possible qu'à la condition de réduire l'influence des informations syntaxiques et sémantiques contenues dans le signal de parole test, et en utilisant des listes d'égale difficulté.

**- Sensibilité :** il s'agit de l'étendue des niveaux d'intelligibilité qu'un test est capable de mesurer (on peut parler de dynamique du test).

**- Facilité de mise en œuvre :** elle dépend du nombre de personnes qui doivent être interrogées pour obtenir des résultats statistiquement fiables, de la période d'apprentissage nécessaire, de la durée de validité des listes de stimuli, des conditions de passation du test (in situ ou en laboratoire), et des méthodes d'analyse (expertise phonétique nécessaire ou non).

La norme ISO TR 4870 [13] décrit les procédures de calibrage de tests d'intelligibilité qui peuvent être appliqués dans le cadre de la norme EN NF 60849. Elle souligne en particulier l'influence du type de matériel phonétique employé et du nombre total d'éléments que le jury s'attend à entendre sur les résultats des tests. D'autre part, la norme insiste sur la nécessité de se limiter à l'évaluation des détériorations du canal de transmission en réduisant au maximum l'influence des processus cognitifs, sources de dispersion importante. Pour cette raison, seuls les tests sur des mots isolés ou des logatomes (mots sans signification) sont considérés. Il faut cependant noter qu'il existe des tests sur des phrases sans signification, qui permettent également en théorie de s'affranchir de ce biais (voir par exemple [14]).

La norme ISO TR 4870 présente deux classes de tests qui se distinguent par les tâches que le jury doit réaliser et le type de matériel phonétique utilisé :

**- Test sur grand corpus avec choix libre**
Les éléments d'une liste de mots ou de logatomes sont soumis un par un à un jury, qui note ce qu'il a compris. Le résultat final est exprimé par le pourcentage de mots ou de phonèmes correctement identifiés. La norme précise que chaque liste doit contenir un minimum de 50 mots ou logatomes choisis dans un grand corpus initial (entre 500 et 1 000 éléments). Pour ce type de test, les auditeurs ne doivent pas avoir de connaissance a priori des choix possibles (corpus initial). Pour réduire la dispersion des résultats, les listes doivent être d'égale difficulté et «phonétiquement équilibrées», c'est-à-dire contenir les mêmes nombres de types de phonèmes, et si possible dans des proportions représentatives de la parole courante considérée (pour le français, on pourra se référer à [15] ou [16], études citées dans [14]). Dix exemples de listes phonétiquement équilibrées pour le français sont données dans [17].

**- Test sur des petits corpus fermés avec choix forcé**
Les éléments du test sont classés dans des sous-ensembles de 2 à 10 éléments (mots ou logatomes) qui se distinguent uniquement par une consonne, toujours à la même place. On parle souvent de « tests de rime ». Le nombre de sous-ensembles doit être suffisant pour pouvoir tester de manière systématique toutes les consonnes, associés avec au moins plusieurs voyelles différentes. Dans ce type de test, le jury doit identifier parmi la liste des éléments de chaque sous-ensemble, le mot tiré au hasard qu'il vient d'écouter. Les réponses sont notées sur un questionnaire à choix multiples, ce qui permet d'automatiser facilement l'étape de dépouillement des résultats.

Le nombre de choix possibles (nombre d'éléments dans chaque sous-ensemble) conditionne le pourcentage de réponses correctes. Sur 2 alternatives possibles, un auditeur a par exemple 50 % de chances de trouver la bonne réponse par hasard même si l'intelligibilité est nulle. Pour pouvoir comparer les résultats des tests par choix forcé avec ceux obtenus par test sur grand corpus, la correction suivante doit donc être apportée :

$$P = \frac{100}{\text{nbre de sous-ensembles}} \left[ \text{nbre de bonnes réponses} - \frac{\text{nbre de mauvaises réponses}}{\text{nbre d'alternatives} - 1} \right] \quad (1)$$

Quel que soit le type de test, la norme conseille de présenter au moins 3 listes au jury pour chaque difficulté. Elle insiste sur l'importance de faire précéder chaque élément par une phrase porteuse, destinée en particulier à focaliser l'attention de l'auditeur et à exciter la réverbération le cas échéant.

Pour un matériel phonétique (liste), un jury et une difficulté donnés, l'évolution statistique des résultats au cours de tests successifs peut être décomposée en trois parties : la première correspond à la durée d'apprentissage de l'auditeur, pendant laquelle les résultats obtenus s'améliorent de test en test.

Suit ensuite la période utile du test, au cours de laquelle les résultats sont statistiquement stationnaires. Enfin, pour les listes « pseudo-ouvertes », les auditeurs ont tendance à connaître les éléments après un certain nombre de présentations, ce qui entraîne une nouvelle phase d'amélioration des scores.

Le tableau 1 permet de comparer les durées d'apprentissage et de validité typiques pour les deux types de tests. Selon ces critères, on note un net intérêt pour les tests sur petits corpus avec choix forcé.

| | **Essais sur des grands corpus (basés sur environ 20 listes de 50 éléments de test, le contenu de chaque liste étant redistribué)** | | **Essais sur des petits corpus fermés** | |
|---|---|---|---|---|
| | Logatomes | Mots | Logatomes | Mots |
| **Durée typique d'apprentissage de l'auditeur** | 24 heures | 12 heures | 2 heures | 5 minutes |
| **Durée typique de test autorisée après apprentissage de l'auditeur** | 200 heures | 120 heures | Pas de limite | Pas de limite |

**Tabl. 1 : Comparaison des deux types de test (extraits de la norme ISO TR 4870)**

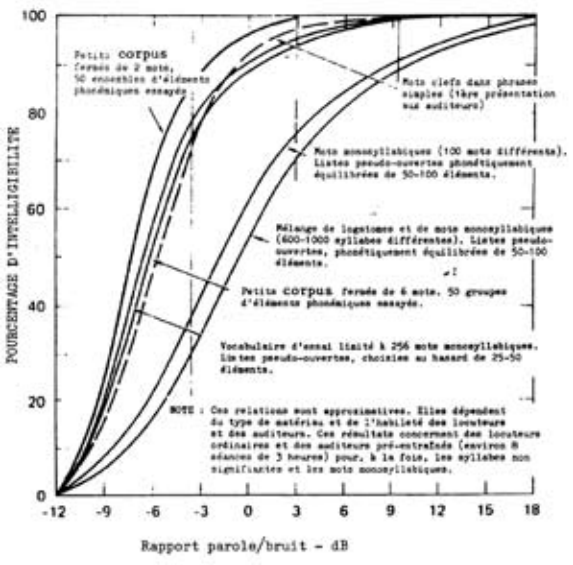


**Fig. 1 : Exemples pour certains tests en langue anglaise, de la relation générale entre le pourcentage d'intelligibilité et le rapport Parole/Bruit en bande large (extrait de la norme ISO TR 4870)**

La période d'apprentissage est très courte, en raison de la simplicité de la tâche à réaliser (remplir un questionnaire à choix multiples). Il n'y a d'autre part aucune limitation sur la durée de validité du matériel phonétique utilisé.
Ce n'est pas le cas en ce qui concerne les essais sur listes « pseudo-ouvertes » qui doivent être renouvelées régulièrement pour un jury donné, et pour lesquels une période d'apprentissage de plusieurs heures est nécessaire. De plus, la meilleure précision des tests par choix forcé permet de réduire considérablement le nombre de personnes à interroger par rapport aux tests sur listes « pseudo-ouvertes » : pour l'étalonnage, la norme préconise un effectif minimum de 5 auditeurs dans le premier cas, contre 10 auditeurs dans le second cas.

La figure 1 représente l'allure des courbes d'étalonnage, à savoir les relations entre le pourcentage d'intelligibilité et le rapport niveau moyen de parole/niveau de bruit obtenues pour différents types de tests. On constate que plus le nombre de choix laissés au jury (taille du corpus apparent) est faible, plus la sensibilité est limitée. Pour le test sur petit corpus fermé de 2 mots par exemple, le pourcentage d'intelligibilité devient maximum pour un rapport Signal/Bruit positif, alors qu'il continue à augmenter pour des tests sur listes pseudo-ouvertes.
Il faut donc retenir que les tests sur petits corpus avec choix forcé sont plus simples à mettre en œuvre et plus précis que les tests sur grand corpus avec choix libre. En revanche, ces derniers ont généralement une meilleure sensibilité.

## Application aux ERP

Dans le cas de l'évaluation de l'intelligibilité dans les ERP, la réalisation de tests d'écoute in situ se heurte au problème de l'évaluation de l'effet de masquage par des bruits parasites réalistes. En effet, il est souvent inenvisageable de réaliser les tests en période d'affluence, pendant laquelle les niveaux de bruits sont représentatifs des conditions d'écoute des usagers. Les tests doivent donc être effectués en période calme, bien souvent durant la nuit, en diffusant un bruit stationnaire pour se rapprocher des conditions réalistes d'écoute.

Une alternative consiste à reproduire en laboratoire des conditions d'écoute in situ. Pour cela, il est nécessaire d'avoir, soit enregistrer des listes sur place en période calme, soit mesurer la réponse impulsionnelle (RI) qui caractérise les transformations linéaires subies par le signal de parole à travers la chaîne électroacoustique et la salle, jusqu'au point d'écoute. Dans ce dernier cas, la diffusion de n'importe quel message parlé peut ensuite être simulée par convolution avec la RI. Il est possible d'évaluer les détériorations de l'intelligibilité entraînées par la présence de bruit parasite en mixant au signal de parole un bruit stationnaire synthétique ou enregistré. La réalisation de tests d'écoute en laboratoire permet d'alléger considérablement la procédure de test. A partir d'une seule mesure de RI et de rapport niveau du signal de parole/niveau de bruit, il est possible de réaliser n'importe quel type de test d'écoute, et de calculer également la plupart des indicateurs « objectifs ». De plus, plusieurs logiciels de prédiction permettent à l'heure actuelle de calculer ces données à partir de modèles géométriques du bâti et des caractéristiques des sources sonores qui s'y trouvent (voir par exemple [18]), rendant ainsi possible la mise en œuvre d'une unique procédure de test, applicable aussi bien dans le cadre d'un diagnostic de l'existant, que pour des études prévisionnelles.

Il faut noter cependant que la RI caractérise uniquement les transformations linéaires. En cas de distorsions importantes de la chaîne électroacoustique, il est nécessaire de réaliser un enregistrement in situ des mots ou des phrases utilisées pour le test. D'autre part, les conditions d'écoute in situ ne peuvent pas être reproduites exactement, même si la mise en œuvre de techniques de prise et restitution du son spatialisé permettent de s'en rapprocher (techniques binaurales [12], ou Ambisonics [19] par exemple).

## Les indicateurs «objectifs»

L'utilisation d'indicateurs physiques «objectifs» permet de s'affranchir des difficultés de mise en œuvre de tests d'écoute. Cependant, il n'existe pas encore d'indicateur universel qui prendrait en compte l'influence de tous les facteurs de détérioration de l'intelligibilité.
Certains critères permettent de séparer les influences des caractéristiques des sources électroacoustiques (directivité, position relative au récepteur…) et de la salle (réverbération) sur l'intelligibilité (par exemple [9], [11]). Cependant, ces procédures reposent généralement sur des hypothèses restrictives concernant l'orientation des enceintes de sonorisation par rapport à l'auditeur, leur nombre, leur directivité, et la nature du champ réverbéré (supposé complètement diffus dans [9] par exemple). Bien que ces approches soient intéressantes pour identifier les facteurs de détérioration de l'intelligibilité, on s'intéressera ici aux critères plus généraux, déterminés à partir du rapport niveau de parole/niveau de bruit moyen et de la réponse impulsionnelle caractéristique des transformations subies par le signal de parole, qui peut être déterminée par la mesure, ou calculée à l'aide de logiciels de prédiction.
Deux familles de critères objectifs adaptés aux espaces réverbérants et bruyants sont considérées par la suite : les Indices de Transmission de la Parole (STI) et les rapports Énergie utile/Énergie nuisible.

### Indices de Transmission de la Parole (STI, STITEL, RASTI)

L'idée originale de Steeneken & Houtgast repose sur les trois hypothèses suivantes :

- La parole peut être représentée par un signal stationnaire, modulé en amplitude par des variations de section du conduit vocal pour former les phonèmes.
- Les détériorations de l'intelligibilité s'expliquent par des pertes de profondeur de ces modulations, qui peuvent être liées aux effets combinés d'une distorsion temporelle, non-linéaire, ou d'un bruit masquant.
- L'intelligibilité peut s'écrire comme la somme pondérée des contributions de chaque bande de fréquence.

La méthode initiale pour déterminer le STI est basée sur la génération de signaux tests, constitués de bruits stationnaires filtrés par bandes d'octaves (fréquences centrales $f_c$ comprises entre 125 Hz et 8 kHz) et modulés en amplitudes à des fréquences $f_m$ comprises entre 0,63 Hz et 12,5 Hz. Les valeurs de la fonction de transfert de modulation (MTF pour Modulation Transfer Function) sont déterminées pour différentes combinaisons ($f_c$, $f_m$). Il s'agit d'une grandeur adimensionnée, comprise entre 0 et 1, qui caractérise le rapport entre les amplitudes de modulation des signaux captés et émis. Cette méthode est employée par exemple par le système de mesure du RASTI de Bruel & Kjaer, paramètre calculé à partir de 9 valeurs de MTF. Il est également possible de déterminer les MTF à partir de mesures de réponses impulsionnelles [20] et de rapport niveau de parole/niveau de bruit mesuré dans le local [21], ou directement par séquences pseudo-aléatoires du type MLS [22]. Dans ce cas, les distorsions non-linéaires ne sont pas prises en compte.

Les étapes de calcul du STI à partir des MTF sont détaillées dans la norme EN NF 60268-16 [6]. La procédure consiste à déterminer à partir de chaque MTF, un rapport signal de parole/bruit apparent (RSB), qui prend en compte l'influence des distorsions du signal, du bruit parasite, des effets de masquage auditif des fréquences basses sur les fréquences plus élevées, et du seuil absolu d'audition. A partir de chaque RSB apparent, est calculé ensuite un indice de transmission, de valeur comprise entre 0, correspondant à un RSB < -15 dB pour lequel le signal de parole utile est supposé être complètement noyé dans le bruit, et un RSB > 15 dB, correspondant à une contribution maximum à l'intelligibilité. Dans chaque bande

d'octave, les indices de transmission pour les différentes fréquences de modulation sont moyennés, pour en déduire des valeurs d'Indices de Transfert de Modulation (MTI). Le STI est ensuite déterminé en effectuant une somme pondérée des MTI, où les poids attribués aux contributions de chaque bande d'octave dépendent du type de locuteur (homme ou femme). La valeur finale s'exprime sous la forme d'une grandeur adimensionnée comprise entre 0 (intelligibilité nulle) et 1 (intelligibilité parfaite).

Avant la parution de la norme EN NF 60268-16, il existait de nombreuses versions du STI qui se différenciaient essentiellement par les poids attribués aux contributions de chaque bande de fréquences, par les hypothèses sur le spectre de voix du locuteur, et par la façon de prendre en compte le bruit parasite et le masquage fréquentiel. En plus du STI, la norme présente également les versions simplifiées (STITEL et RASTI), en précisant les domaines d'applications de chacun de ces trois indicateurs. Ils ne doivent pas être utilisés lorsqu'un traitement non-linéaire du signal de parole est réalisé avant diffusion (glissement de fréquences des systèmes anti-Larsen, compression-expansion de la dynamique, utilisation de vocodeurs de phase…). Le STITEL peut être appliqué lorsque les effets de distorsion temporelle sont faibles. Il correspond à une simplification du STI dans laquelle une seule valeur de MTF est déterminée pour chacune des 7 bandes d'octave 125 Hz-8 kHz (contre 14 dans les 7 bandes d'octaves pour le STI, soit 98 valeurs de MTF en tout). Le RASTI est destiné à caractériser l'intelligibilité de la parole naturelle, ou diffusée par des systèmes électroacoustiques de réponse en fréquence uniforme, et en présence de bruit de fond de spectre régulier. Il repose sur une réduction du nombre de bandes d'octave prises en compte (9 valeurs de MTF réparties sur les deux bandes centrées sur 500 Hz et 2 kHz). La mise en œuvre du RASTI et du STITEL, initialement motivée par la complexité de calcul du STI, ne se justifie plus à l'heure actuelle étant donnée l'augmentation de la puissance informatique des systèmes de mesure portables.
Dans le cadre d'un diagnostic, l'observation de l'évolution des valeurs de MTF avec les fréquences de modulation permet de déterminer quel facteur contribue le plus à la détérioration de l'intelligibilité (réverbération ou bruit de masquage).

### Rapport Énergie utile/Énergie nuisible

Ces critères reposent sur la séparation en deux parties de l'énergie acoustique totale captée au point d'écoute :

**- L'énergie utile à l'intelligibilité,** est associée au signal de parole porté par le front d'onde direct et les premières répliques retardées (contributions des échos précoces et des sources électroacoustiques situées à proximité de l'auditeur).
**- L'énergie nuisible à l'intelligibilité,** regroupe l'énergie portée par les répliques tardives du signal de parole (champ réverbéré) et par le bruit parasite le cas échéant.

Sous ces hypothèses, le rapport Énergie utile/Énergie nuisible, exprimé en dB, permet de caractériser l'intelligibilité. Dans la procédure initiale décrite par Lochner & Burger, la contribution de chaque écho à l'énergie utile ou nuisible dépend non-seulement de l'instant d'arrivée, mais également de son amplitude relative par rapport au premier front d'onde.
Étant donnée la complexité du calcul, des critères simplifiés ont été proposés, éliminant la dépendance en amplitude (clarté, définition). Le seul paramètre est alors la limite temporelle qui permet de distinguer énergies utiles et tardives, qui varie de 30 à 100 msec suivant les auteurs (voir par exemple [9]). Récemment, des critères plus complexes reprenant le critère initial de Lochner & Burger ont été développés [10,11]. Ils permettent de prendre en compte l'effet cumulé des distorsions temporelles et du bruit masquant sur l'intelligibilité.
Contrairement à l'AI et au STI, les rapports énergie utile/énergie nuisible, notés Ux par Bradley [10] (où x désigne la limite temporelle en msec permettant de distinguer les énergies utiles et nuisibles), reposent donc essentiellement sur une description temporelle de la répartition de l'énergie. Ce type de critère est particulièrement intéressant pour la recherche de solutions d'amélioration de l'intelligibilité dans les grands espaces réverbérants. En effet, en fonction des instants d'arrivée et de leur amplitude, on peut déterminer la contribution à l'intelligibilité de chaque écho de la partie précoce de la réverbération, qui peut ensuite être optimisée en modifiant l'emplacement des sources électroacoustiques ou en leur appliquant des lignes à retard par exemple. Comme dans le cas du STI, l'influence des bruits parasites est prise en compte sous la forme d'un rapport niveau moyen du signal de parole/niveau moyen de bruit dans le local. Dans les travaux de recherche menés sur ce type de critères, les dépendances avec la fréquence de la réverbération, de la réponse du système électroacoustique ou du bruit masquant sont généralement négligées. Cependant, il est tout à fait envisageable d'appliquer une procédure similaire à celle du STI, qui consiste à déterminer des indices équivalents aux MTI à partir des valeurs de U dans chaque bande d'octave, puis d'effectuer une somme pondérée des contributions.

## Corrélation entre les critères – Présentation de l'échelle commune d'intelligibilité de la norme EN 60849

De nombreuses études ont porté sur la comparaison des critères objectifs et des résultats de tests d'intelligibilité. Les figures 2 et 3 représentent par exemple les corrélations observées pour le STI et le rapport énergie utile/énergie nuisible U80 par Bradley [10], et par Steeneken & Houtgast [4] pour le STI uniquement. Sur ces figures, on note des divergences assez importantes entre les résultats, notamment concernant les écarts-types par rapport aux courbes de régression obtenues pour le STI.
Elles peuvent s'expliquer par des différences dans l'ensemble des configurations testées (réverbération et bruit masquant), pour lesquelles les critères sont plus ou moins bien adaptés. Ainsi par exemple, la précision du STI est sans doute meilleure que celle des critères énergétiques lorsqu'il y a de fortes variations de réverbération avec la fréquence. En revanche, pour des bruits masquants de niveau élevé ou en présence de fortes réflexions isolées, les critères énergétiques semblent mieux adaptés (lire par exemple [23]). Les mauvais résultats obtenus à l'aide du STI pour les forts bruits masquants proviendraient

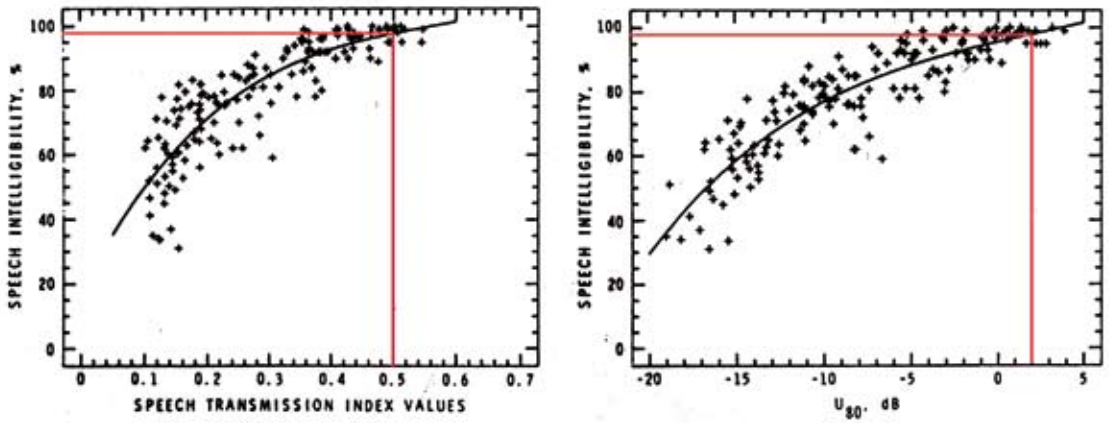


**Fig. 2 : Corrélations entre STI (gauche), et U80 (droite) avec des scores d'intelligibilité (tests de rimes). Résultats obtenus par Bradley [10] pour 160 configurations différentes de réverbération (TR compris entre 0.6 et 3 s) et de bruit (RSB entre – 10 et + 15 dB (A)). L'écart type par rapport à la courbe de régression (polynôme d'ordre 3) est indiqué pour chaque critère.**

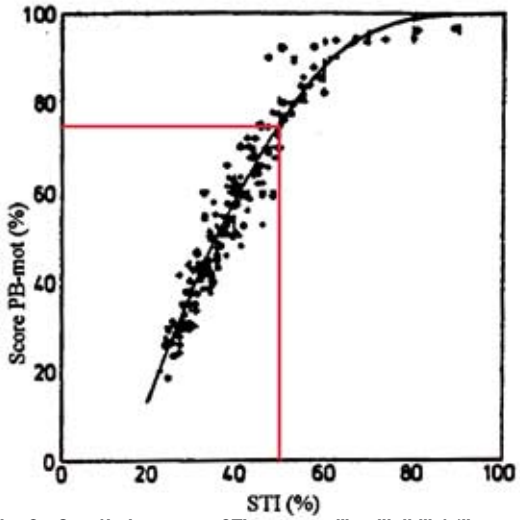


**Fig. 3 : Corrélations entre STI et score d'intelligibilité (liste équilibrées phonétiquement). Résultats obtenus par Steeneken & Houtgast [4] pour 167 configurations de spectres de bruit, de RSB, de réverbération, de compression.**

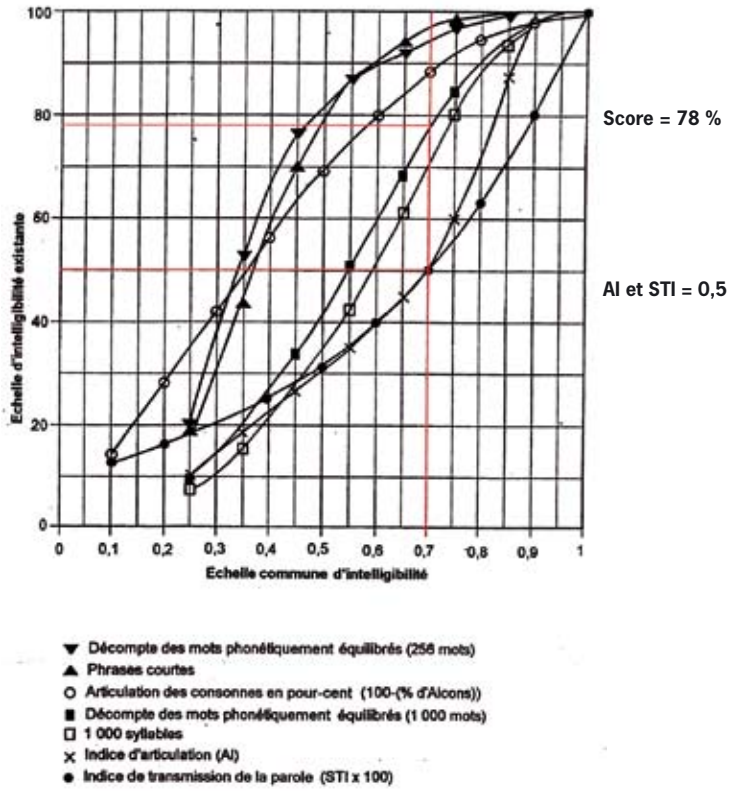


**Fig. 4 : Échelle commune d'intelligibilité (CIS, extrait de la norme EN NF 60849). Les lignes en rouge correspondent à la limite inférieure d'intelligibilité spécifiée dans la norme.**

principalement de la limite basse (indice de transmission nul pour des RSB < -15 dB), qui aurait tendance à sous-estimer l'intelligibilité. Une étude menée au LAUTM par Legros, Randrianarison & Gamba, citée dans [23], a en effet montré que l'intelligibilité était encore non-nulle pour des RSB = -15 dB (A).

Il faut noter d'autre part que les tests de référence utilisés dans les études [4] et [10] ne sont pas les mêmes (test sur listes de mots phonétiquement équilibrés dans [4], et test de rime dans [10]). Ceci peut expliquer les différences d'allure des courbes de corrélation obtenues, les tests n'ayant pas la même dynamique. Ainsi par exemple, un STI égal à 0,5 correspond à des scores d'intelligibilité de 75 % d'après [4], et 95 % d'après [10]. Ces exemples illustrent bien les difficultés qui peuvent être rencontrées dans la recherche d'un critère universel.

La figure 4 représente l'échelle commune d'intelligibilité de la norme EN NF 60849. La limite inférieure, représentée en rouge (0,7 sur l'échelle commune), correspond à un score de 80 % environ pour un essai sur listes équilibrées phonétiquement tirées d'un grand corpus pseudo-ouvert (carrés noirs), et à des valeurs de STI et de AI égales à 0,5.

Ce résultat est cohérent avec ceux présentés par Steeneken & Houtgast dans [4] pour le STI (figure 3). En revanche, il concorde moins bien avec les résultats présentés par Bradley dans [10], pour qui un STI et un AI de 0,5 sont associés à des scores d'intelligibilité différents (respectivement 95 % et 80 %). D'après Bradley, cette valeur de STI correspond à U80 = 2 dB.

Il faut retenir que les correspondances entre les critères peuvent s'écarter de manière significative des courbes données dans la norme. Sur les figures 2 et 3, on peut noter par exemple, des écarts allant jusqu'à 20 % entre certains points et les courbes de régression. L'utilisation d'un seul critère objectif, tel que le STI, peut entraîner une sous-évaluation ou sur-évaluation importante des résultats qui seraient obtenus à l'aide d'un test d'intelligibilité. La principale critique que l'on peut faire à l'échelle commune d'intelligibilité de la norme EN 60849 concerne donc l'absence d'indications sur la dispersion des correspondances entre critères.

## Conclusion

Les enjeux de la nouvelle norme EN NF 60849 sont importants. Lorsqu'elle sera imposée par la réglementation, une valeur minimum de 0,7 sur l'échelle commune d'intelligibilité devra être assurée dans les ERP équipés d'une système de sonorisation de sécurité (correspondant à un «bon niveau d'intelligibilité» d'après [9]). A l'heure actuelle, il est techniquement possible d'atteindre ce résultat dans la plupart des espaces difficiles, tels que les grands ERP réverbérants et bruyants. Cependant, la complexité des solutions à mettre en œuvre et le nombre de facteurs à prendre en compte nécessite l'intervention de spécialistes pour y parvenir : on risque donc d'assister à une révolution dans le monde des installateurs de systèmes de sonorisation dans les ERP, qui font bien souvent à l'heure actuelle une impasse complète sur l'étude acoustique.

Comme on l'a vu, plusieurs critères peuvent être utilisés pour le diagnostic d'intelligibilité. Cependant, la lourdeur de mise en œuvre des tests basés sur le décompte de phrases, de mots ou de phonèmes correctement reconnus par un jury, qui représentent la majorité des méthodes proposées par la norme, les rendent quasiment inapplicables dans la plupart des cas d'étude, à moins d'effectuer les tests en laboratoire à partir de signaux enregistrés ou simulés par convolution. Ce point est d'autant plus critique, que l'application stricte de la norme exige une étude statistique de l'intelligibilité sur toute la zone de couverture ([1] annexe B-3). Pour les études à budget limité, la seule alternative possible consiste donc à mesurer les valeurs de STI, l'autre indicateur «objectif» proposé par la norme (AI) étant mal adapté aux espaces réverbérants. Cependant, comme l'ont montré plusieurs auteurs (notamment [10] et [24]), les corrélations entre le STI et les scores de mots ou de phrases reconnues sont faibles dans certains cas, en particulier en présence de bruit important. En l'absence d'indications dans la norme sur la dispersion entre les critères, la marge d'erreur est délicate à évaluer, et risque d'entraîner des conflits d'experts. La rédaction de nouvelles normes sur l'intelligibilité venant compléter les normes EN NF 60849 et 60268-16 semble donc indispensable. En particulier, la normalisation d'un critère basé sur le rapport Énergie utile/Énergie nuisible, bien adapté aux espaces très réverbérants et bruyants, permettrait de compléter le STI.

## Références bibliographiques

[1] Norme NF EN 60849, « Systèmes électroacoustiques pour services de secours », 1998.

[2] N.R. French & J.-C. Steinberg « Factors governing the intelligibility of speech sound », J. Acoust. Soc. Am. 19 (4), 1947.

[3] V.M.A. Peutz « Articulation Loss of consonants as a criterion for speech transmission in a room », J. Audio Eng. Soc. 19 (11), 1971.

[4] H.J.M. Steeneken & T. Houtgast, « A physical method for measuring speech-transmission quality », J. Acoust. Soc. Am. 67 (1), 1980.

[5] T. Houtgast & H.J.M. Steeneken, « A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria », Acustica 54 (4), 1984.

[6] Norme NF EN 60268-16 « Équipements pour systèmes électroacoustiques. Partie 16 : évaluation objective de l'intelligibilité de la parole au moyen de l'indice de transmission de la parole », 1998.

[7] J.P.A. Lochner & J.-F. Burger, « The influence of reflections on auditorium acoustics », J. Sound Vib. 1, 1964.

[8] H. Kuttruff, Room acoustics, Applied Science Publishers Ltd., 1973.

[9] L.G. Marshall, « An analysis procedure for room acoustics and sound amplification systems based on the early-to-late sound energie ratio », J. Audio Eng. Soc.44 (5), 1996.

[10] J.-S. Bradley, « Predictors of speech intelligibility in rooms », J. Acoust. Soc. Am. 80 (3), 1986.

[11] L. Faiget & R. Ruiz, « Speech intelligibility model including room and loudspeaker influences », J. Acoust. Soc. Am. 105 (6), 1999.

[12] J. Blauert, Spatial Hearing, the psychoacoustics of human sound localization, rev. ed., MIT Press, 1999

[13] Norme ISO TR 4870, « Acoustique – Élaboration et étalonnage des tests d'intelligibilité de parole », 1991.

[14] P. Combescure, « 20 listes de phrases phonétiquement équilibrées », Revue d'Acoustique 56, 1981

[15] J.-P. Haton & M. Lamotte, « Étude statistique des phonèmes et diphonèmes dans le français parlé », Revue d'Acoustique 16, 1971.

[16] L. Sklarczyk, Essai sur la structure phonologique du français, Thèse de Doctorat de l'Université de Pennsylvania, USA, 1961.

[17] P. Zuliani, Intelligibilité de la parole dans des conditions d'écoute difficiles, Thèse de Doctorat de l'Université de Toulouse III, 1988.

[18] M. Kleiner, B.I. Dalenbäck & P. Svensson, « Auralization – An overview », J. Audio Eng. Soc., 41 (11), 1993

[19] M.A. Gerzon, « Ambisonics in multichannel broadcasting and video », J. Audio Eng. Soc., 33 (11), 1985.

[20] M.R. Schroëder, « Modulation transfer functions : definition and measurement », Acustica 49, 1981

[21] T. Houtgast & H.J.M. Steeneken, « A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria », J. Acoust. Soc. Am. 77 (3), 1985.

[22] D.D. Rife, « Modulation transfer function measurements with Maximum-Length Sequences », J. Audio Eng. Soc. 40 (10), 1992.

[23] L. Faiget, Séparation de l'influence du local et de l'enceinte pour la prévision de l'intelligibilité dans des conditions d'écoute difficiles, Thèse de Doctorat de l'Université Paul Sabatier de Toulouse, 1997.