Université Pierre et Marie Curie

Mémoire pour le Master Sciences et Technologie de l'UPMC, Mention SDI, Spécialité MIS, Parcours ATIAM.

Qualité perçue de parole transmise par voie téléphonique large-bande

Coté Nicolas

12 septembre 2005

Laboratoire d'accueil :

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

En collaboration avec l'Institut für Kommunikationsakustik

Maître de stage :
Alexander RAAKE

Effectué du 7 Mars au 9 Septembre 2005

Remerciements

Je tiens tout d'abord à remercier mon maître de stage Alexander RAAKE, pour sa disponibilité malgré ses quelques péripéties. Son sens de l'humour et sa rigueur m'ont permis de travailler dans les meilleures conditions. De plus, c'est grâce à lui si j'ai pu effectuer une collaboration avec l'IKA à l'université de Bochum en Allemagne.

Je tiens à remercier également Sébastian Möller de l'IKA, qui m'a accueilli durant un mois au sein du laboratoire de Bochum et a su jouer le rôle d'un maître de stage tout en me faisant découvrir la région de la Ruhr.

J'aimerais remercier également Brian FG Katz et Christophe d'Alessandro du LIMSI ainsi que Yann Krebber et Jens Blauert de l'IKA, tout comme les autres personnes qui m'ont aidées durant mon stage ; Marcel Wältermann, Rosa Pegam ainsi que les stagiaires du LIMSI.

Enfin un grand merci à toutes les personnes qui m'ont accordées le temps nécessaire aux tests de perception.

Tables des matières

Chapitre 1 Introduction

Ce mémoire étudie la transmission de signaux téléphoniques large-bande. La transmission en large-bande correspond à l'élargissement de la bande passante utilisée pour la transmission du signal de parole. En effet, la bande passante habituellement utilisée en téléphonie va de 300 Hz à 3400 Hz, or la bande élargie permet la transmission de la voix de 50 Hz jusqu'à 7000 Hz. Le but de cette étude est de parvenir à quantifier l'amélioration de la qualité téléphonique par l'utilisation de ce type de signal. Son importance tiens dans le fait qu'il existe pour l'instant peu de protocoles permettant l'étude de signaux large-bande. Cette étude est donc incluse dans un travail de psychoacoustique sur la perception de la parole, mais aussi de traitement du signal sur les modes de transmission de la parole.

L'étude de la qualité de la parole est un domaine important de la psychoacoustique pour ses applications dans la synthèse sonore, la médecine ou comme ici, pour la téléphonie. Les applications en téléphonie ont évolués parallèlement à l'apparition des nouvelles modalités de communication ; l'apparition des téléphones mobiles ou la transmission de la voix par Internet a conduit à l'adaptation (amélioration, création pour certains) de modèles permettant la prédiction de la qualité d'un signal de parole sur une ligne téléphonique. Certains modèles permettent de prendre en compte la perception de l'auditeur, mais l'évaluation de la qualité se fait toujours sur les paramètres physiques d'une connexion. Cependant ces modèles ne prennent pas en compte la qualité induite par la transmission d'une bande passante différente, les apports de la transmission large-bande restant peu étudiés à ce jour [Raake, 2005].

Or, un des sujets de travail de l'union international des télécommunications (ITU) correspond à l'évolution du modèle E [ITU-T Rec. G.107, 2005], modèle de planification de système de téléphonie, pour permettre l'évaluation de signaux téléphonique large-bande.

Le travail de ce mémoire est donc basé sur l'étude de la perception de signaux téléphoniques bande-étroite et large-bande par un auditeur, puis de l'appliquer à l'amélioration du modèle E. Ce travail a donné lieu à une collaboration entre le LIMSI et l'IKA, laboratoire d'acoustique de l'université de Bochum en Allemagne.

Ce travail a nécessité dans un premier temps, une étude des différents facteurs intervenant dans l'évaluation de la qualité de la voix. Une partie importante a été également consacrée à l'étude des modes d'évaluation de la qualité d'un signal de parole. Ceux-ci peuvent être basés sur des paramètres physiques comme le modèle E, ou sur le jugement des auditeurs, comme pour les tests de perception. Un chapitre de l'état de l'art met en exergue cette partie théorique de mon travail qui a précédé la mise en place de tests de perception développés dans le troisième chapitre. Celui-ci présente le protocole utilisé pour la réalisation de trois tests de jugement de la qualité. En effet, les études étant peu nombreuses sur le sujet il a été nécessaire d'expérimenter plusieurs types d'approches du problème. Les chapitres quatre à six montrent chacun les résultats obtenus pour l'un des tests suivant :

- Un test normalisé par l'ITU appelé test MOS [ITU-T Rec. P.800, 1996], a été réalisé à l'IKA, en langue Allemande. Il utilisait des conditions bande-étroite, la bande passante habituelle de la téléphonie, et également en bande élargie mixée avec des conditions bande-étroite.
- Un test utilisant l'échelle de quantification de la qualité du modèle E, a été réalisé dans les deux laboratoires, LIMSI et IKA.
- Un test de comparaison par paires, réalisé en Français au LIMSI.

Enfin le dernier chapitre correspond à une première extension du modèle E à la large-bande, pour permettre une évaluation des transmissions utilisant l'une ou l'autre des deux bandes passantes sur la même échelle de qualité.

Pays :	**Allemagne**			**France**			
Type de test :	Echelle MOS bande-étroite	Echelle MOS bande-étroite large-bande	Echelle R	Echelle R	Echelle R	Comparaison par paires	Comparaison par paires
Nombre de locuteur :	4	4	2	2	2	1	1
Nombre de conditions :	9	18	18	18	18	13	13
Entraînement :	non	non	non	non	**oui**	non	**oui**
Nombre de sujet :	22	22	21	15	6	15	7

Tableau 1.1 : Description des trois tests de perception effectués pour l'étude de la transmission large-bande.

Chapitre 2 Etat de l'art

Cette partie va permettre de décrire toutes les problématiques liées à la quantification de la qualité d'un signal de parole au cours d'une communication téléphonique. Pour cela, tout au long de ce chapitre, il est question d'un message qu'une première personne souhaite transmettre à autrui. Les caractéristiques de ce message seront détaillées au cours du chapitre.

Dans un premier temps, une description de la téléphonie est donnée en début de chapitre. Puis, le message utilisé correspondant à un signal de parole, une partie sera consacrée à la manière dont le signal est créé, puis perçu par un auditeur. Par la suite, dans une seconde partie, la transmission à travers un système de communication est introduite, retraçant les différents procédés qui ont amené la téléphonie large-bande. Enfin pour mener à bien cette étude, il a été nécessaire d'étudier les différents outils mis à notre disposition pour évaluer la qualité d'un système de communication. Une définition de la qualité de la parole est également donnée, ainsi que certaines notions sur les tests subjectifs de perception.

2.1 La téléphonie, pourquoi évoluer vers la « large-bande » ?

Le but de ce mémoire est donc de quantifier l'apport d'une amélioration que connaît la téléphonie, la « large-bande ». Celle-ci permet d'augmenter la bande passante utilisée au cours d'une conversation. En effet la bande passante utilisée habituellement en téléphonie est 300-3400 Hz, elle défini le débit de base d'une ligne téléphonique qui est de 64 kbits/s. Cependant les nouvelles technologies liées aux réseaux permettent une utilisation plus flexible de la transmission de la parole, grâce à un choix assez large de codeur et de bande passante, qui a facilité l'apparition d'une nouvelle bande passante 50-7000 Hz, améliorant la qualité du signal de parole transmis.

La téléphonie peut être décrit comme un système de communication permettant de transmettre de la parole. Elle peut, en premier lieu être vue comme un système créant une liaison entre deux personnes et permettant la transmission d'un message. Son but (comme celui de tout système comportant de la parole) devrais être de maximiser la compréhension de ce message. Le message étant défini comme un signal de parole transmis de la bouche d'un locuteur jusqu'à l'oreille d'un auditeur, le rôle des deux personnes changeant au cours de la conversation. La compréhension ou non de ce message est le résultat de plusieurs étapes dont une d'interprétation du message par l'auditeur. Celle-ci est décrite par une taxinomie décrite dans [Möller, 2000, p. 26-27] et [Jekosch, 2000], et dépend du contexte du locuteur ainsi que des connaissances du sujet.

Ainsi, la compréhension du message délivré par un locuteur dépend de :

- La compréhensibilité du message, liée directement au locuteur ou au système et à sa capacité à donner une information, de transmettre les phonèmes (souvent en fonction de son articulation, tout dépendant du contexte de locution).
- L'intelligibilité, qui correspond à la possibilité d'établir un sens au message transmis avec l'ensemble des phonèmes du message.
- La communicabilité, qui est la compréhension de l'ensemble des messages, dans les deux sens de la liaison.

Il est important de remarquer que, la compréhension du message dépend du locuteur, du contexte de locution, ainsi que des connaissances du sujet sur le message.

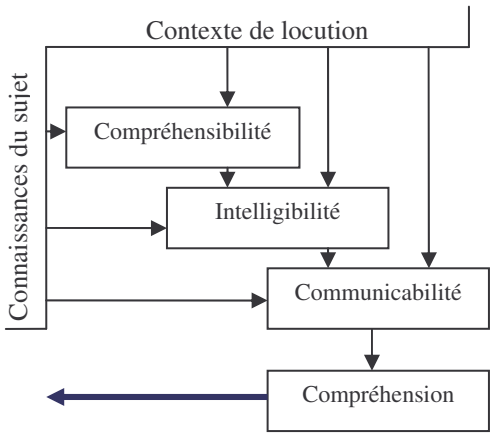

Figure 2.1 : Les facteurs agissant dans la compréhension d'un message [Raake, 2005].

La téléphonie bande-étroite permet la compréhension d'un message mais il est nécessaire de prendre en compte d'autres critères pour définir la qualité globale d'un système de communication. Ainsi, la téléphonie doit par exemple répondre à certaines règles définies par un organisme de standardisation : l'Union International des Télécommunications (UIT). L'ITU définit le terme de qualité de service (QoS) qui correspond au degré de satisfaction d'un utilisateur [ITU-T Rec. E.800, 1994]. Ces règles permettent une réglementation des protocoles utilisés et facilite l'interconnexion des différents réseaux.

Or, dans [Möller, 2000], la qualité liée à la transmission du message est définie comme une partie de la qualité de service. Cette partie, appelée « facteurs de communication de la parole », est composée de trois éléments :

- La qualité de la transmission de la voix (chaque élément ayant un impact sur la parole durant l'écoute).
- L'efficacité de la conversation (les capacités du système lors d'une conversation).
- Les facilités de communication (facteurs liés à l'interlocuteur).

La téléphonie est donc une application qui peut évoluer, en améliorant l'un des deux premiers critères. L'étude de la large-bande va principalement porter sur le premier facteur. Or, pour connaître en quoi l'augmentation de la bande passante permet d'améliorer la qualité globale de la téléphonie, il est nécessaire d'étudier le message vocal transmis. Le prochain paragraphe porte donc sur l'étude de la parole : sa production par un locuteur et sa perception par un auditeur.

2.2 La parole, importance de la bande passante

Le message à transmettre, comme il a été dit précédemment, est un signal de parole. Or, il est nécessaire d'adapter le système de communication au type de message transmis. Ce paragraphe va montrer en quoi il est nécessaire d'augmenter la bande-étroite de la téléphonie

au profit de la bande élargie. De plus, il faut prendre en compte la qualité du système dans son intégralité, donc les « facteurs de communication ». La communication étant basée sur la création du message « la production » et sa réception « la perception ».

2.2.1 La production vocale

Tout d'abord, un son a une forme physique qui se propage dans un milieu par le biais d'ondes. Ces ondes sont liées au canal de transmission (air, câble) mais surtout au producteur de ce son. Elles peuvent alors être quantifiées sur une échelle de fréquences. Le signal de la parole a donc des caractéristiques temporelles mais aussi fréquentielles. Dans le cas de la parole humaine, les fréquences dépendent de la forme et de la position de certains organes du corps humain. La parole peut alors être vue comme un signal source (corde vocales, glotte) qui est filtré par des tuyaux formés par les conduits vocaux (comme le conduit nasal). La manière dont le son d'origine est filtré dépend de la signification que veut lui donner le locuteur. Ainsi on peut voir que la production vocale est composée de sons ayant des composantes fréquentielles très spécifiques, dont on donne souvent comme valeur la fréquence fondamentale (F_0, correspondant au signal porteur), et les premiers formants (F_i, piques dans l'amplitude spectrale dus aux résonances du conduit vocal). Certains phonèmes comme les voyelles se caractérisent très facilement par ses formants. Les consonnes sont produites de manières différentes [speech communications], elles peuvent être sonores (« l », « r »), nasales (« m », »n ») ou fricatives (« h », « f »). Ces dernières, comme le « s » ou le « f », produisent de l'énergie essentiellement dans les hautes fréquences ainsi que dans un formant très bas, autour de 150 Hz. De même la résonance nasale se situe au alentour de 250 Hz.

Les fréquences utilisées par la parole humaine, peuvent donc être comprises entre 110 et 7 kHz (speech communications). La bande-étroite 300-3400 Hz utilisée par la téléphonie permet de faire passer les 3 premiers formants, et ainsi de garantir une intelligibilité de la parole du locuteur, mais ne permet pas de transmettre l'intégralité des fréquences présentes dans un signal de parole. Par exemple, il est très difficile de différencier un « s » d'un « f » prononcé seul, lors d'une conversation téléphonique.

2.2.2 La perception de la parole

Il est nécessaire de rappeler que les fréquences audibles par une oreille humaine sont habituellement situées entre 20 Hz et 20 kHz : ce qui à première vu semble très loin de la bande-étroite de la téléphonie. De plus, suite à de nombreuses utilisations, le cerveau humain, a créé une référence de la qualité « sonore » de la voix humaine, transmise à travers un système de téléphonie [Duncanson, 1969]. L'évaluation de la qualité est alors biaisée par la référence de la téléphonie fixe, fortement liée aux fréquences de la bande-étroite.

Afin de mieux appréhender le choix de la bande élargie pour la téléphonie, le paragraphe suivant rappelle la notion de « bande critique » ;

Les bandes critiques [Zwicker, 1961] correspondent à une répartition dans le spectre des fréquences d'un ensemble de bandes de fréquences. Ces bandes sont des regroupements des excitations sonores ayant des fréquences voisines et perceptivement proches au sein de certaines bandes fréquentielles. Il est possible de passer de l'échelle des fréquences à celle des bandes critiques grâce à la fonction suivante [Zwicker et Fastl, 1999] :

$$z_{(barks)} = 13 \cdot \tan^{-1} \cdot (0{,}76 \cdot f_{(kHz)}) + 3{,}5 \cdot \tan^{-1} (f_{(kHz)} / 7{,}5)^2 \quad (2.1)$$

Cette échelle en bande critique est une échelle perceptive (correspondant à la Tonie), dont l'unité, le « Bark » regroupe un ensemble variable de fréquences, 1 Bark correspondant à une bande passante de 100 Hz à 3500 Hz. Une bande passante peut alors être obtenue en Barks :

$$z_{bp} = z(f_h) - z(f_b) \quad (2.2)$$

Les fréquences audibles vont de 0 (20 Hz) jusqu'à 24 Barks (16 kHz). La téléphonie bande-étroite représente 14 Barks (de 3 à 16 Barks), soit plus de la moitié de l'échelle audible. L'utilisation des bandes critiques permet de connaître la valeur perceptive d'une bande passante. Par exemple, dans [Gleiss, 1970] une bande passante de 180-2800 Hz est considérée de même qualité qu'une seconde bande passante de 280-3550 Hz. Un simple calcul suffit pour voir que la différence en Barks de ces deux bandes passantes est 0, car les deux fréquences de coupures subissent une simple translation d'environ 1 Bark.

Cette échelle permet alors d'analyser le choix de la bande-étroite en téléphonie classique et l'apport de la large-bande. La fréquence de coupure basse, baisse de 300 Hz jusqu'à 50 Hz, soit une augmentation de la bande passante de 2,5 Barks, tandis que la fréquence de coupure haute, permet une augmentation de 4 Barks.

La fréquence centrale, permet également de connaître le poids de fréquences basses et hautes dans les deux bandes passantes.

$$f_c = \sqrt{f_b \cdot f_h} \quad (2.3)$$

Pour la bande-étroite : f_c=1010 Hz, et pour la bande élargie : f_c=590 Hz.

Ces deux valeurs montrent que la bande élargie comporte perceptivement plus de basse fréquence que la bande-étroite. Cette description des bandes passantes par le biais de la fréquence centrale et de la largeur spectrale en Barks est utilisée dans [Raake, 2005], afin de lier la perception fréquentielle à la qualité d'un signal de parole. En effet, le sujet de ce mémoire portant sur la qualité de la voix pour différentes bandes de fréquence, il sera intéressant d'utiliser ces deux principes dans les chapitres suivants.

Par ailleurs, il a été vu précédemment que l'intelligibilité était fortement liée aux premiers formants. Dans [Gleiss, 1970] il est montré que la sensation naturelle de la voix dépend fortement du premier formant et l'intelligibilité du second formant. En effet, les différents phonèmes sont perçus en fonction :

- Du rapport entre les formants.
- De leurs variations dans le temps.

De plus, le premier formant étant approximativement entre 270 et 730 Hz pour les hommes, et entre 310 et 850Hz pour les femmes, le choix de la fréquence de coupure basse à 300 Hz a été choisie judicieusement. La bande-étroite permet un bon compromis entre intelligibilité et qualité du son.

Pour autant, celle-ci ne permet pas de transmettre la fréquence fondamentale F_0, qui dans [Gleiss, 1989] semble être liée à la sensation naturelle de la voix. Celle-ci comprise entre 110 et 200 Hz pour un adulte, et montant jusque 300 Hz pour un enfant, permet de transmettre la prosodie [O'Shaughnessy, 2000, p.101-107], comme les intonations ou les émotions.

Théoriquement, la perception humaine permet de reconstruire cette fréquence fondamentale en son absence, et de percevoir tout de même les différentes intonations exprimées par le locuteur. Malgré cela, dans [Gleiss, 1970] on voit également que la perception de la parole à travers un système de téléphonie dépend énormément de la présence des fréquences basses ; Une petite différence de 45 Hz, de 225 à 180 Hz, sur la fréquence de coupure basse améliore nettement la sensation naturelle de la voix. La perception de la fréquence fondamentale semble donc avoir une importance dans l'évaluation de la qualité de la parole.

Enfin, dans [Moore et Tan, 2003], lors de l'augmentation de la fréquence de coupure basse de 123 à 208 Hz, une dégradation est ressentie sur la perception de la voix humaine, celle-ci semble moins naturelle. Une dégradation est obtenue également lors d'une diminution de la fréquence de coupure de 5500 à 3500 Hz. Mais il montre également qu'il est nécessaire d'améliorer la bande-étroite aux deux extrémités. En effet, il montre que pour une fréquence de coupure basse proche de 300 Hz, le changement de la fréquence de coupure haute (de 7000 à 3500 Hz) n'a que peu d'effets. De même, pour une fréquence de coupure haute à 3500 Hz, un changement de la fréquence basse de 55 à 300 Hz à peu d'effets également. Il est donc impossible de compenser une coupure trop forte, en haut ou en bas spectre, en agrandissant l'autre coté.

La bande-étroite introduit donc une dégradation de la sensation naturelle de la voix par :

- L'atténuation du premier formant.
- L'absence de transmission de F_0.
- L'absence de transmission des hautes fréquences.

En conclusion, la parole humaine produit des fréquences qui en partie ne sont pas comprises dans la bande-étroite, et qui sont nécessaires pour obtenir une voix humaine naturelle. La téléphonie large-bande, qui permet la transmission de la majorité des fréquences produites par la voix, nous espérons que celle-ci permettra de rendre la voix d'un interlocuteur plus naturelle.

2.3 Les évolutions de la téléphonie

Cette troisième partie va montrer les différentes étapes qui ont fait apparaître la large-bande. Car, bien que le premier système large-bande soit apparu en 1988, ce type de transmission est encore aujourd'hui très peu répandu. En effet, il a fallu une suite d'évolution dans les télécommunications pour permettre l'utilisation de CoDec large-bande dans l'industrie.

Historique :

La bande passante traditionnelle utilisée en téléphonie fixe, appelée « bande-étroite », est 300-3400Hz. Elle résulte d'un filtre standardisé par l'ITU [ITU-T, Rec. G.712, 1992]. Or, le réseau national de téléphonie, à l'origine analogique a subi un premier changement en passant à l'ère numérique : Passant du Réseau Téléphonique Commuté (RTC) au Réseau Numérique à Intégration de Service (RNIS). Pour autant ce réseau peut être défini par certaines caractéristiques comme la connexion physique, entre deux interlocuteurs, qui durant toute la durée de la conversation, reste active. Ceci n'est plus vrai avec l'apparition de la « Voix sur IP » ou « Voice over Internet Protocol » (VoIP), qui permet la transmission de

données sous forme de paquets. Le paragraphe suivant introduit le concept de transmission de la voix sur Internet, et les problématiques induites.

2.3.1 La VoIP

Tout d'abord la VoIP est apparue grâce à la démocratisation des connexions Internet Haut débits ou des réseaux Ethernet pour les réseaux locaux.

La VoIP est d'abord un système de mise en paquets des données vocales [Tanenbaum, 2003]. En effet, lorsque la connexion est établie, la parole codée (par un CoDec) est ensuite mise sous forme de paquets de données (de 10 ms à 40 ms) puis envoyée de la source jusqu'à la destination par le biais du réseau Internet (et de son protocole). Des entêtes sont ajoutés aux paquets de donnés vocales (DATA) par les différents protocole employés et augmentent successivement la taille du fichier.

Par exemple :

- Le RTP[1], est un protocole qui permet le transport de données en temps réel en indiquant l'horaire d'envoie des paquets, ce qui permet de calculer les paquets transmis et ceux perdus.
- Le UDP[2], indique le port de la source et de la destination.

La voix sur IP, permet de réutiliser les réseaux existant et ainsi de réduire les coûts liés à la téléphonie. Elle permet également d'introduire de nouvelles modalités de communication, par l'utilisation d'interfaces multimédia, mais introduit des dégradations nouvelles, comme les pertes de paquets ou les bruits introduits par l'utilisation de « codeur - décodeur ». Ceux-ci, les CoDecs, sont détaillés dans le paragraphe suivant.

2.3.2 Les CoDecs

Les CoDecs sont des codeurs de données, qui permettent un transport plus facile sur certains canaux de transmission (comme le hertzien pour les téléphones portables, ou Internet pour la VoIP), en réduisant la taille des données. Un décodeur de l'autre côté de cette chaîne de transmission, permet de retrouver les données envoyées avec parfois certaines dégradations.

L'utilisation de CoDec réduit le débit des transmissions en exploitant les propriétés de la production vocale et de la psychoacoustique. En effet, le débit d'une ligne fixe est de 64 kbits/s. Celle-ci utilise le CoDec G.711 qui correspond à une simple digitalisation du signal analogique en « Pulse Code Modulation » (PCM[3]) [ITU-T Rec. G.711, 1988]. En revanche certains CoDec peuvent atteindre des débits de l'ordre de quelques kbits/s.

Les CoDec peuvent être classifiés en trois groupes :

---

[1] RTP : Real-time Transport Protocol.
[2] UDP : User Datagram Protocol.
[3] PCM : ce codage correspond à une quantification non linéaire de type loi A /μ.

- Les codeurs en forme d'ondes (utilisent l'onde sans compression).
- Les codeurs paramétriques (utilisent un modèle de production vocale).
- Les codeurs hybrides (combinaison des deux).

Il existe différents types de CoDec suivant le canal utilisé :

- RNIS : PCM fonctionnant à un débit fixe (Ex : G.711 à 64 kbits/s).
- DCME[1] : Adaptative Differential Pulse Code Modulation (ADPCM) fonctionnant à des débits variables suivant la disponibilité du réseau (Ex : G.726 à 40, 32, 24 ou 16 kbits/s).
- Réseaux Mobiles : GSM[2] (Ex : GSM 06.60 « Enhanced Full Rate ») [ETSI, Rec. GSM 06.60, 1996], UMTS réseau mobile de troisième génération permettant la transmission de voix et de données.
- VoIP : entre autre Code Excited Linear Prediction (CELP), (Ex: G.729, iLBC).

On remarque que le type de codage effectué sur la voix est souvent adapté au canal de transmission (en fonction du débit disponible). Mais l'utilisation de CoDec introduit une dégradation de la qualité vocale, des distorsions. Il est aujourd'hui possible de quantifier la dégradation introduite par un CoDec sur la qualité du signal de parole grâce notamment à des modèles objectifs de mesure [ITU-T Rec. G.107, 2005]. Le but de cette étude est de quantifier la dégradation introduite par un codeur large-bande, mais également de trouver l'amélioration de qualité par rapport à des codeurs bande-étroite.

2.3.3 La « WideBand »

Le codage large-bande n'est pas un domaine d'étude très récent, le premier CoDec utilisant une bande passante élargie, le G.722 [ITU-T Rec. G.722, 1988], fut développé dans les années 1980 pour être utilisé sur le réseau RNIS. Mais le codage de la parole en large-bande entraîne des techniques différentes de la bande-étroite. En effet, il y a une plus forte dynamique spectrale pour la parole en bande large. De plus la voix est plus inharmonique dans les hautes fréquences, comme pour les fricatives, en raison des caractéristiques morphologiques. Mais, plusieurs études, efforts de développement et standardisation ont permis de créer quelques CoDecs de meilleur qualité et moins coûteux en débit.

Aujourd'hui l'augmentation du débit sur les différents canaux que sont l'UMTS ou la VoIP permet l'utilisation de CoDec de meilleure qualité. Par exemple, le CoDec « Adaptative Multi – Rate WideBand » [AMR-WB, ITU-T Rec. G.722.2, 2002], qui est à bande élargie, peut être utilisé sur tout les réseaux [ETSI, TS 126 173, 2004]. Basé sur un codage du type A-CELP, il utilise le phénomène de masquage fréquentiel et atteint des débits variant de 6,60 à 23,85 kbits/s. Le codage de l'AMR-WB, suivant le débit, correspond aux bandes passantes entre 50Hz et 6600 Hz pour le plus gros débit, et entre 50Hz et 6000Hz pour les débits inférieurs. Cette augmentation de bande passante permet ainsi une augmentation de la sensation naturelle de la voix humaine.

---

[1] DCME : Digital Circuit Multiplication Equipement.
[2] GSM : Global System for Mobil communications.

L'utilisation d'un système utilisant un codage large-bande, permet donc d'augmenter la qualité de la téléphonie. Pour autant, il est nécessaire de quantifier l'apport de ce type de codage. Les prochains paragraphes apportent quelques points de théorie sur la manière de quantifier la qualité d'un signal de parole.

2.4 Les problématiques liées à l'évaluation de la qualité

Pour évaluer la qualité d'un système de télécommunications, il est courant d'utiliser des outils de traitement du signal comme le rapport signal sur bruit. Or, lorsque le message transmis correspond à une voix humaine et que celui-ci est appliqué à la téléphonie, il est nécessaire de prendre en compte les attentes des utilisateurs. Il est donc important de prendre en compte le facteur humain dans ce type d'application, la notion de qualité de la parole étant liée à l'auditeur.

Le prochain paragraphe va donc donner une définition de la qualité. Puis, les paragraphes suivant vont donner les outils d'évaluation de qualité ;

- Les outils objectifs utilisés dans l'industrie.
- Les outils subjectifs, permettant d'étudier la perception humaine et d'améliorer les outils objectifs.

En effet, l'industrie a besoin d'outils permettant de connaître de manière objective la qualité d'un système transmettant de la parole. Or la mise en place de ces outils objectifs nécessite de connaître la perception de l'auditeur, par le biais de test perceptif de jugement de la qualité. Pour une partie des modèles objectifs existant ces connaissances psychoacoustiques sont alors intégrées.

2.4.1 Qualité de la parole

Ce paragraphe étudie la manière dont, le message transmis à l'auditeur, est perçu puis interprété. Le paragraphe précédent, sur la parole, a montré l'importance de la bande passante en tant que caractéristique physique. Ici, ce sont les facteurs psychologiques de la perception qui sont étudiés.

2.4.1.1 Les attentes du locuteur

Il a été vu précédemment que la référence interne de l'auditeur influençait sa perception (Cf. paragraphe 2.2.1). Les conversations téléphoniques peuvent avoir lieux dans des contextes très différents (par exemple, dans une gare ou chez soi) influençant la compréhension du message. De plus, il faut rappeler que le message transmis est un son, donc un phénomène acoustique qui porte des informations délivrées par le locuteur. L'onde perçue par l'auditeur est ensuite analysée par l'auditeur [Jekosch, 2000]. Ici, des informations supplémentaires au signal traité par l'auditeur, autres que la simple compréhension du message, sont introduites (comme les émotions). Le schéma suivant représente l'influence des attentes de l'auditeur dans l'évaluation d'un stimulus.

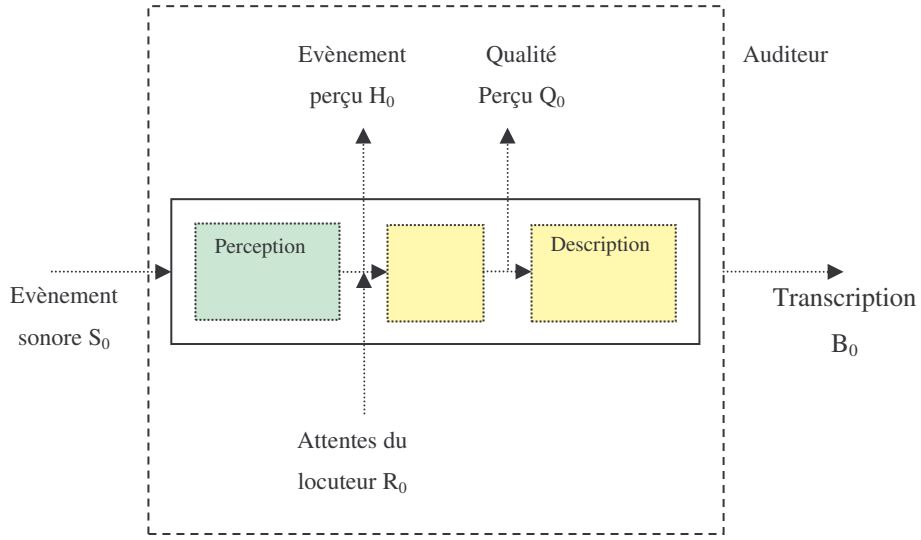

Figure 2.2 : Représentation schématique de la perception d'un son [Raake, 2005], basée sur des travaux de [Blauert, 1997] et [Jekosch, 2000, 2004].

Les cinq évènements sont :

- Le son physique ou évènement sonore (S_0).
- L'évènement perçu au niveau du cerveau (H_0).
- Les attentes de l'auditeur (R_0).
- La qualité perçue (Q_0)
- La transcription par l'auditeur de l'évènement (B_0).

C'est donc B_0, la valeur qui, dans un test de qualité est sensé donner du sens au phénomène physique, la valeur qui doit au mieux correspondre à la qualité perçue par l'auditeur. Cette valeur, dans le cadre de test de perception, est une valeur numérique donnée sur une échelle de perception. Elle est ensuite reliée aux valeurs physiques du signal, quand cela est possible. C'est donc B_0 qui va nous donner une valeur caractéristique de la connexion transportant S_0.

C'est donc l'étape d'interprétation, qui met en relation les connaissances de l'auditeur sur le message (R_0), et H_0 l'évènement perçu, qui donne la « qualité » perçue (Q_0). En fonction de ce principe, une définition de la qualité est donnée dans [Jekosch, 2000] comme étant :

> « *Le résultat de l'évaluation de la nature d'une entité perçue en fonction de sa nature désirée* ».

Dans cette étude portant sur différents systèmes de communication, la nature d'une entité perçue correspond au message acoustique transmis, et la nature désirée correspond à la référence interne de l'auditeur. Un exemple assez fréquent est la première communication téléphonique avec une personne dont on connaît la voix. Il arrive alors fréquemment de ne pas

reconnaître cette personne. En reprenant [Mariani, 2002] on peut dire que la bande passante téléphonique modifie le timbre de la voix et ainsi la rend moins naturelle. Nous pouvons donc définir la qualité de la voix, comme étant lié à son timbre. La qualité du « son vocal » est associé à une « coloration » particulière de la voix. En effet, le timbre est relié à l'identité de la source sonore, donc à l'identification du locuteur. Le passage à la large-bande, en évitant toute coloration de la voix, permet une meilleure identification du timbre du locuteur, plus proche du timbre réel.

Il faut noter également que dans [Raake, 2000, p. 217], l'utilisation de plusieurs terminaux dans l'évaluation de stimuli, montre que les attentes de l'auditeur face au terminal d'écoute jouent un grand rôle dans la valeur globale de la qualité de la parole.

En conclusion, ce paragraphe montre la nécessité de prendre en compte les attentes de l'auditeur face aux sons perçus par un système de communication, et donc aussi, face aux stimuli qui lui sont présentés lors de test de perception.

2.4.2.2 Multi - dimensions

On définit habituellement un son grâce à ses attributs perceptifs ; Sonie, hauteur, timbre, durée et localisation [Letowski, 1989]. Ce son, peut donc être décomposé sur plusieurs dimensions perceptives. De même, on peut aussi décrire la qualité d'un système de communication sur plusieurs dimensions ou attributs, qui correspondent aux caractéristiques perçu du signal. Par exemple, le chapitre précédent a montré que le passage à la bande élargie donnait à la voix une sensation plus naturelle, on voit également dans [Mariani, 2002] que la sensation naturelle est accentuée par l'ajout de bruit ambiant, situé dans les hautes fréquences et ainsi donne une sensation de présence du locuteur lointain. La présence du locuteur lointain et la sensation naturelle de la voix sont deux dimensions différentes.

Lors de la perception de qualité, de manière intuitif (naturelle) et contrôlée (test de qualité), l'espace interne des dimensions est comparée aux valeurs des dimensions perçues pour une connexion.

Le but de cette étude est de quantifier la qualité globale d'un système de communication. Il est donc nécessaire de regrouper toutes les dimensions sur une échelle à une seule dimension. De plus, il faut prendre en compte tous les éléments extérieurs qui peuvent jouer sur la qualité globale du stimulus. Par exemple, le type de présentation des stimuli pendant un test de perception, comme le système d'écoute, peut jouer un rôle sur la perception de la qualité. Enfin, la difficulté est de connaître les variables à prendre en compte lors de la mise en place de test de jugement. Par exemple, dans le cas de cette étude, la durée et la sonie ne sont pas des dimensions prises en compte.

En effet, toutes nos dégradations sont continues dans le temps, et tous les stimuli ont le même volume. Mais, un réseau de télécommunication introduit plusieurs bruits qui peuvent être de natures différentes, et qui varient suivant la bande passante utilisée. Les bruits dégradant le signal peuvent être classés en trois groupes :

- Le bruit de ligne.
- Le bruit corrélé au signal (introduit par la quantification).
- Le bruit de fond (du lieux d'écoute ou de d'envoi).

En conclusion, en prenant en compte les différentes caractéristiques vues précédemment, on peut dire de la qualité :

> « [Qu'] *elle est attribuée à un service par un utilisateur dans une situation spécifique, et reflète les attentes de l'utilisateur, sa motivation et son attitude* » [Jekosch, 2000].

Maintenant qu'une définition de la qualité a été donnée, il est nécessaire de l'appliquer aux outils d'évaluation de système de communications. Ceux-ci sont décrits dans le paragraphe suivant.

2.4.2 L'évaluation de la qualité de la parole de systèmes de télécommunications

Le but des outils d'évaluation de la qualité des systèmes de téléphonie, est de donner une valeur numérique qui permette de quantifier la qualité de ces systèmes et ainsi de les comparer. Cette partie est composée de deux paragraphes, le premier décrivant un outil objectif d'évaluation, et le second correspond à une description des tests de perception qui permettent d'évaluer certaines connexions.

2.4.2.1 Les modèles objectifs d'évaluation de la qualité de la parole

Les recherches en psychoacoustique sur la qualité de la parole ont, entre autre, comme but de créer des outils d'évaluation objectifs. Ceux-ci donnent une valeur de la qualité globale de la parole perçue à travers un système de communications. Ainsi, cette étude a comme objectif de contribuer au développement d'un modèle permettant la prédiction de la qualité perçue à travers un système large-bande, et cela sur la base de paramètres physiques. De tels modèles sont déjà utilisés pour la planification des systèmes téléphoniques à bande-étroite comme par exemple le Modèle E détaillé plus loin [ITU-T Rec.G.107, 2005].

Mais vue que la perception auditive est interne à l'auditeur, la qualité d'un système de téléphonie ne peut pas être mesurée, de manière absolue, par une méthode objective. Les tests de jugement de la qualité par des auditeurs sont les seules évaluations valables et sûres d'un système de parole. Cependant ces tests sont assez difficiles à mettre en place et assez coûteux. Des méthodes instrumentales de jugement de la qualité ont donc été développées, essayant de prévoir la qualité de la parole en se basant sur la mesure de signaux caractéristiques.

Il existe différentes catégories de modèles suivant le domaine d'utilisation. Cependant il est possible de définir ces modèles suivant plusieurs critères [Möller et Raake, 2002]:

- Le but de l'application testé.
- Les dimensions de la qualité à quantifier.
- Les composants considérés dans le réseau.
- Les paramètres d'entrée du modèle.
- La part de psychoacoustique utilisée dans le modèle.

On peut ensuite les classer suivant trois types de modèles [Möller et Raake, 2002]:

- Modèles basés sur des mesures comparatives à base de signaux (Ex : PESQ).
- Modèles de planification de réseaux.
- Modèles de surveillance de réseau.

Un exemple de modèle basé sur des mesures du signal est le modèle « Perceptual Evaluation of Speech Quality » (PESQ) [ITU-T Rec. P.862, 2001]. A la base, ce modèle était destiné à prévoir la qualité d'un seul composant d'un système de télécommunication complet. Mais il peut juger certains autres problèmes, comme certains types de pertes paquets. Ce modèle fonctionne par comparaison entre un signal non processé et un signal qui est passé à travers un système. Il utilise une représentation du signal sous forme psychoacoustique. Certains travaux ont été effectués pour étendre ce modèle à la transmission large-bande [FT R&D, ITU-T, COM- D.046, 2005].

Un modèle de surveillance permet d'évaluer la performance d'un réseau existant, en mesurant « on - line » certains des paramètres utilisés pour planifier un système. Ainsi, il est possible d'évaluer la qualité d'un système en fonctionnement.

2.4.2.2 Le modèle E (exemple de modèle de planification de réseaux)

Le but de ce genre de modèle est donc bien de connaître tous les paramètres liés à une connexion, de la bouche du locuteur jusqu'à l'oreille de l'auditeur. Ainsi le modèle peut prédire la qualité globale perçue par un utilisateur lors d'une communication. Le modèle E est principalement basé sur des résultats de jugements de qualité (tests de perception) [pour un historique de la littérature à ce sujet, voir Möller, 2000].

Ce type de modèle permet d'étudier toutes les dégradations perçues par les utilisateurs, mais n'utilise qu'une seule dimension. La prédiction est obtenue sur une échelle « psychologique » d'évaluation.

Pour planifier un réseau complet il est difficile de connaître la qualité globale avant la mise en place de ce système. Le modèle E permet la « *prédiction de la qualité vocale transmise par une ligne téléphonique* » [Möller et Raake, 2002]. Ce type de modèle estime la qualité globale du système grâce à des mesures instrumentales et une description du système par de nombreux paramètres scalaires. Le modèle E est la compilation de différents modèles de l' « European Telecommunications Standards Institute » [ETSI ETR 250, 1996]. Il est également le modèle recommandé par l'ITU-T pour la conception de réseaux de qualité [ITU-T Rec. G.107, 2005].

Le paramètre le plus important pour cette étude, l' « Impairment Factor » *Ie*, est décrit dans le paragraphe suivant.

Equipement Impairment Factor :

Le facteur de dégradation *Ie* permet de quantifier les dégradations dues à la phase de « codage - décodage » donc dues au CoDec [ITU-T Rec. G.113, Appendice I, 2003 ; ITU-T Rec. P.833, 2001]. Il est défini pour quantifier les dégradations se produisant pendant la transmission. Celles-ci sont typiquement non linéaires et peuvent varier dans le temps. Chaque *Ie* est spécifique aux caractéristiques perceptives d'un CoDec, il est calculé en fonction de la dégradation affecté au signal en comparaison au CoDec de référence : le G.711.

L'hypothèse fondamentale du modèle E postule que toutes les catégories de dégradation sont additives sur une échelle de qualité appropriée (voir plus loin). Selon ce modèle, les différentes dégradations individuelles, lors de l'utilisation de plusieurs CoDec, peuvent être additionnées sur l'échelle R du modèle E de la manière suivante :

$$Ie_{\tan dem} = \sum_{i} Ie_{individual}{}^{i} \tag{2.3}$$

Cela est nécessaire pour un calcul global, puisque dans [FT R&D, ITU-T, COM- D.049, 2005], on voit qu'il devient commun d'utiliser plusieurs types de transmission sur une même liaison et donc plusieurs CoDec différents.

L'échelle R :

La valeur de sortie du modèle E, se situe sur cette échelle perceptive. Elle permet de situer certains CoDec dont l'évaluation est standardisée. Elle permet d'exprimer la qualité grâce à 5 valeurs scalaires :

$$R = Ro - Is - Id - Ie,eff + A \tag{2.4}$$

- *Ro* représente le SNR[1] par rapport au point 0 dBr[2], en prenant en compte le niveau de la voix et les différents bruits présents sur la ligne (une valeur objective de la qualité du signal).
- *Is* est la dégradation pendant la parole (de manière simultanée).
- *Id* correspond aux problèmes de délais ou d'écho.
- *A* est l'« advantage of access » ou avantage d'utilisation. Par exemple le réseau mobile à l'avantage de permettre une mobilité.
- *Ie,eff* est l'« Equipement Impairment factors ». Il représente la dégradation de la qualité de la parole subie durant la transmission à travers un ou plusieurs CoDecs (sans ou avec pertes de paquets ; cf. Appendice I, ITU - T G.113).

L'échelle R est comprise entre 0 et 100, où 0 correspond à la plus mauvaise qualité et 100 la meilleure. La valeur sur l'échelle R de la référence de la téléphonie bande-étroite, le G.711, utilisée sur un canal RNIS propre de bruit vaut :

$$R = 93{,}2$$

Le but de ce mémoire est donc de trouver une méthode capable d'évaluer un système large-bande sans passer par des tests de jugement de qualité mais en utilisant des paramètres physiques. L'objectif est donc de pouvoir utiliser le Modèle E pour la parole large-bande et de trouver une extension de l'échelle R, correspondant à l'amélioration introduite comparé à la bande étroite. Le prochain paragraphe introduit les tests de jugement perceptifs, et décrit le type de test adapté à notre étude.

---

[1] SNR : Signal Noise Ratio, rapport signal sur bruit.
[2] dBr : valeur identique au dB_{SPL}, appliquée à la téléphonie.

2.4.2.3 Téléphonométrie

La psychoacoustique, lorsqu'elle est appliquée aux télécommunications, est basée sur des tests de perception sonore. Ainsi, cette étude, bien qu'ayant pour but le traitement du signal, est basée sur des données subjective (Cf. B_0, vu précédemment). Cette valeur obtenue qui est une valeur numérique, doit être valable (mesure ce qui est variable dans le test) et non biaisée (la même mesure doit être trouvée lorsque le test est répété). Il faut donc prendre en compte toute la chaîne de transmission du stimulus, de la bouche du locuteur jusqu'à la valeur obtenue lors de l'évaluation, et ainsi choisir judicieusement toutes les caractéristiques des tests pour obtenir des valeurs que l'on puisse analyser [Blauert, 1997]. En effet, tous les choix faits à propos d'un test de perception peuvent jouer un rôle sur le choix du sujet.

Les tests de perception, suivant la classification de [Raake, 2005], peuvent être classés en quatre groupes, suivant deux variables :

- Le sujet de l'étude :
    - Analytique (description de la perception en dimensions).
    - Utilitaire (jugement de qualité globale).
- Le type de méthode utilisé :
    - Orienté objet (systèmes de communication).
    - Orienté sujet (rassemblement des informations sur la perception humaine).

Le type de test « utilitaire et orienté objet » a été choisi, car il doit permettre d'obtenir un jugement de la qualité globale sur un système de téléphonie. De plus, les interfaces de jugement utilisaient une échelle monodimensionnelle pour juger directement la qualité de système de communication de la parole.

2.4.2.4 Les tests de perception :

Une définition d'un test de perception est donnée par [Jekosch, 2000] :

« Une méthode courante pour étudier une ou plusieurs dimensions de la qualité d'une parole perçue, distinguables empiriquement, dans le but d'obtenir une valeur quantitative des dimensions étudiées ».

Les tests de perception, dans le cadre d'étude de système de téléphonie sont standardisés par l'ITU : ITU-T Rec. P.800, 1996 et Rec. P.830, 1996.

Le paragraphe précédent a défini le type de test à utiliser dans le cadre de cette étude ; les tests utilitaires. Ceux-ci peuvent être définies et classifiés par trois éléments [Möller, 2000, p. 48-49] :

- L'échelle utilisée : nominale, ordinale à intervalle constante (Ex : MOS), ou du type ratio (à intervalles constantes, avec un point zéro ; Ex : CR-10 [Borg, 1998]).
- La méthode de présentation utilisée :
    - Elle peut être à « estimation absolue », lorsque le stimulus est présenté seul (Ex : ACR).
    - Ou à « comparaison par paires » quand les stimuli sont présentés par paires (Ex : DCR).

- La modalité utilisée pour la présentation des sons : Ecoute seul, conversation, parlé et écoute.

Sur ce dernier point, les tests effectués au cours de l'étude utilisaient la modalité « écoute seul ». Ces tests sont appelés « Listening Only Tests » (LOTs). Ils permettent sur une durée fixe, d'évaluer un nombre plus important de stimuli par rapport aux deux autres modalités. Bien qu'une présentation des stimuli par la modalité « conversation » permette d'être plus proche d'une utilisation réelle, celle-ci implique des difficultés supplémentaires dans la mise en place des tests. De plus, les effets étudiés ici, peuvent être perçus pendant une simple écoute des signaux transmis à travers le système.

Le type de test le plus utilisé dans l'évaluation subjective des systèmes de communication est le test « Absolute Category Rating » (ACR) [ITU-T Rec. P.800, 1996]. Il utilise habituellement une échelle sur 5 points, avec ou sans annotations, appelée échelle MOS, pour « Mean Opinion Score ».

Excellent	Bon	correct	faible	mauvais
5	4	3	2	1

Figure 2.3 : Echelle ACR à 5 points (MOS).

Cependant cette échelle recommandée par l'ITU pose certains problèmes [Möller, 2000, p. 68-72]:

Tout d'abord elle est utilisée comme une échelle à intervalle (où il est possible de faire des calculs statistiques correspondants), mais a été conçue en tant qu'échelle ordinale. De plus, cette échelle peut être utilisée de façons différentes entre les sujets. Cette échelle avec annotations présente également un problème de non linéarité entre les 5 points annotés : les catégories de l'échelle n'étant pas perçues comme également espacées. L'espace d'évaluation varie également suivant les langues. Enfin, elle introduit une saturation au niveau des extrémités de l'échelle en 1 et 5.

Les valeurs MOS peuvent ensuite être converties en valeur sur l'échelle R, du modèle E [ITU-T Rec. G.107, 2005, Annexe B], grâce à la formule suivante :

$$MOS = \begin{cases} 1 & pour : R < 0 \\ 1 + 0{,}035 * R + R(R-60)(100-R) * 7 \cdot 10^{-6} & pour : 0 < R < 100 \\ 4{,}5 & pour : R > 100 \end{cases} \quad (2.5)$$

Valable pour : $6{,}5 < R < 100$.

Or, le but de l'étude est de trouver une amélioration par rapport au système RNIS de référence, il sera nécessaire dans le chapitre 7 de transformer cette formule de conversion.

Ce chapitre d'état de l'art a défini le travail à accomplir pour permettre de quantifier la qualité d'un signal large-bande transmis par voie téléphonique. La définition de la qualité

donnée en 2.4.1.2 et des outils permettant son évaluation subjective, sera utilisée pour la mise en place d'un protocole expérimental rigoureux et fiable. De plus, la première partie de ce paragraphe a expliqué en quoi la large-bande permettait de réellement améliorer la qualité globale de la téléphonie. Les résultats des tests de perception effectués vont pouvoir être interprétés en utilisant ces connaissances, et ainsi donner une valeur aux résultats trouvés.

Chapitre 3 Méthode

Ce troisième chapitre, va permettre de montrer les différents paramètres qui ont joué un rôle dans la mise en place des différents tests de perception. Ces sept tests, ont été effectués pour moitié à l'Institut für Kommunikationsakustik, de l'université de Bochum en Allemagne, et pour l'autre moitié au LIMSI, à l'université d'Orsay. Ce nombre important de test a permis de travailler sur plusieurs problématiques, toujours en rapport avec la transmission large-bande.

3.1 Conditions choisies

Dans la littérature nous pouvons observer comme dans [Krebber, 1995] qu'une amélioration moyenne de la qualité de 1,3 à 1,5 points MOS est obtenue pour un passage de la bande-étroite à la large-bande. Pour effectuer cette comparaison de manière fiable, les conditions représentant les différents systèmes à évaluer, doivent être choisies de manière à étudier les différentes étapes de la compréhension du message. Certaines conditions permettent une meilleure intelligibilité du message, d'autres améliorent la sensation naturelle. En outre, l'utilisation de conditions avec différentes bandes passantes permet de vérifier le modèle de dégradation proposé dans [Raake, 2005]

Plusieurs types de conditions ont donc été choisis :

- Des conditions introduisant des dégradations dues aux codeurs.
- Des conditions sans utilisation de codeurs mais avec des bandes passantes différentes.

18 conditions au total ont été choisies, réparties en 9 conditions concernant la bande-étroite, et également 9 conditions pour la bande élargie.

Ce choix de 18 conditions comprend :

- Des codeurs large-bande, dont un l'AMR-WB (noté ici G.722.2), utilisé à différents débits suivant la recommandation P.833 de l'ITU. En effet, dans [FT R&D, ITU-T, COM D.033, 2005], nous observons que la qualité globale varie avec le débit. La dégradation *Ie* augmentant lorsque le débit diminue.
- Une référence large-bande PCM. Cette condition représente notre référence pour l'extension de l'échelle R du modèle E.
- Des conditions de référence bande-étroite suivant la recommandation P.833 de l'ITU. Ces références utilisent un filtre représentant les caractéristiques électroacoustiques d'un téléphone fixe habituel, appelé filtre « Intermediate Reference System » (IRS_R ou IRS_S pour receive ou send) [ITU-T Rec. P.830 Annexe D, 1996]. Il permet de simuler la réponse fréquentielle des téléphones du commerce.
- Plusieurs bandes passantes entre 600-2000 Hz et 50-7000 Hz.

Numéro de condition	Nom du CoDec	Type de CoDec	Débits (kbits/s)	Bande Passante	Impairment Factor (*IE*)	Fréquence centrale (Hz)	ΔZ (barks)
1	G.711	PCM	64	G.712, IRS	0.0	-	-
2	G.726	ADPCM	32	G.712, IRS	7.0	-	-
3	G.729	CS-ACELP	8	G.712, IRS	11.0	-	-
4	IS-54	VSELP	8	G.712, IRS	20.0	-	-
5	G.726	ADPCM	16	G.712, IRS	50.0	-	-
6	-	-	-	300-3400 Hz	-	1010	13,4
7	-	-	-	600-3400 Hz	-	1428	10,7
8	-	-	-	300-2000 Hz	-	775	10,2
9	-	-	-	600-2000 Hz	-	1095	7,5
10	G.722	ADPCM	64	Plate	?	-	-
11	G.722.2	ACELP	23,05	Plate	?	-	-
12	G.722.2	ACELP	12,65	Plate	?	-	-
13	G.722.2	ACELP	6,6	Plate	?	-	-
14	-	-	-	50-7000 Hz	-	592	20
15	-	-	-	200-7000 Hz	-	1183	18,5
16	-	-	-	600-7000 Hz	-	2049	14,9
17	-	-	-	100-5000 Hz	-	707	17,6
18	-	-	-	50-3400 Hz	-	412	15,8

Tableau 3.1 : Caractéristiques des conditions.

: Conditions de référence.

La condition de référence bande-étroite avec deux fois le CoDec G.726 à 24 kbits/s a été utilisée pour éviter une condition avec un niveau de bruit additif très perceptible. Le G.726 à 16 kbits/s, auquel en théorie est attribuée la même valeur *Ie* = 50 qu'au tandem (*Ie* = 25+25=50), introduit un bruit remarqué. Une telle condition pourrait mener à introduire des dimensions perceptives différentes des autres conditions.

Le chapitre précédent a montré qu'il fallait prendre en compte la totalité de la chaîne de transmission dont le matériel source (Les locuteurs). Les effets de certaines dégradations peuvent bien sur être totalement différents suivant le locuteur. L'augmentation de la fréquence de coupure haute permettant une augmentation de qualité de parole pour un locuteur féminin, et la diminution de la fréquence de coupure basse, une augmentation de qualité de la parole pour un locuteur masculin [Pascal, 1988]. En effet, les caractéristiques de la liaison sont perçues si le signal non processé est approprié pour mettre en valeur ces caractéristiques. locuteurs ont ainsi été choisis : 2 français (un homme et une femme), et 4 allemands (2 hommes et 2 femmes).

Auparavant, ce type de test utilisait des conditions « Modulated Noise Reference Unit » [MNRU, ITU-T Rec. P.810, 1996] comme références, mais elles ne sont plus utilisées

actuellement. Ces conditions introduisent une distorsions corrélé au le signal, perceptivement très différente des dégradations introduites par les codeurs [Hall, J. L., 2001].

3.2 Echelles

3.2.1 ACR type MOS

Tout d'abord, pour obtenir une valeur globale, des échelles monodimensionnelles ont été utilisées. Ainsi, des Tests en utilisant l'échelle MOS avec annotations, développée en 2.4.2.4 (figure 2.3), ont été effectués pour comparer les résultats avec ceux de la littérature. Deux tests utilisant cette échelle ont été utilisé : l'un utilisant exclusivement des conditions bande-étroite, et un second utilisant toutes les conditions, large-bande et bande-étroite. Cette méthode reprend les approches décrites dans [FT R&D, ITU-T, COM- D.046] et [Raake A., ITU-T, COM-D.028]. La littérature montre dans ce cas que les conditions bande-étroite obtiennent des valeurs MOS inférieures à un test utilisant seulement des conditions bande-étroite. La comparaison de ces deux tests a permis de quantifier l'amélioration apportée par l'utilisation de la large-bande. De plus, pour ces deux tests, quatre locuteurs ont été utilisés (2 locuteurs et 2 locutrices). Ces deux tests ont été effectués en Allemagne à l'IKA.

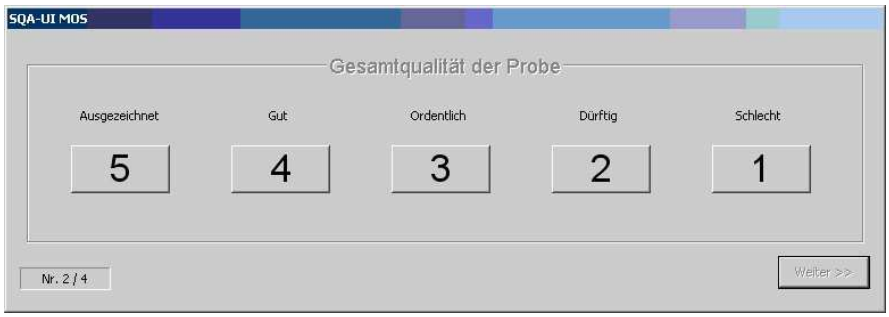

Figure 3.1 : Interface graphique du test utilisant l'échelle MOS.

L'auditeur écoute, une seule fois chaque stimulus, puis estime la qualité du stimulus en appuyant sur l'une des valeurs de l'échelle MOS. Cette valeur est validée lorsque le sujet appuie sur « suivant », ce qui joue automatiquement le prochain stimulus.

3.2.2 Echelle R du modèle E

Cette étude ayant également pour but de déterminer un protocole adapté à l'évaluation de transmission large-bande, un second type de test a été choisi ;

Le choix s'est tourné vers l'échelle R qui pour [Raake A., ITU-T, COM-D.028] semble idéale pour un test large-bande, mais aucun protocole n'a pour le moment été standardisé. L'échelle étant très large (de 0 à 100), pour évaluer les conditions en fonction des références bande-étroite, plusieurs CoDec dont le *Ie* est connu, ont été placés sur cette échelle :

- La référence du modèle E, le G.711 (*Ie*=0, Cf. paragraphe 2.4.2.2). Cette référence est positionnée à 100, maximum de l'échelle R pour la bande-étroite. Ainsi, tout résultat supérieur à cette valeur correspond à une amélioration de la bande-étroite.
- Le CoDec G.729 (*Ie*=10), positionné à 90 sur l'échelle.
- Le CoDec IS-54 (*Ie*=20), positionné à 80.
- Le tandem G.726 (24 kbits/s) * G.726 (24 kbits/s) (*Ie*=50), positionné à 50.

Ce second test, a été effectué dans les deux pays, avec à chaque fois, deux locuteurs différents, un homme et une femme, prononçant la même phrase. Ce choix a été fait pour éviter des différences entre les conditions, liées à un contenu phonétiquement différent entre différentes phrases.

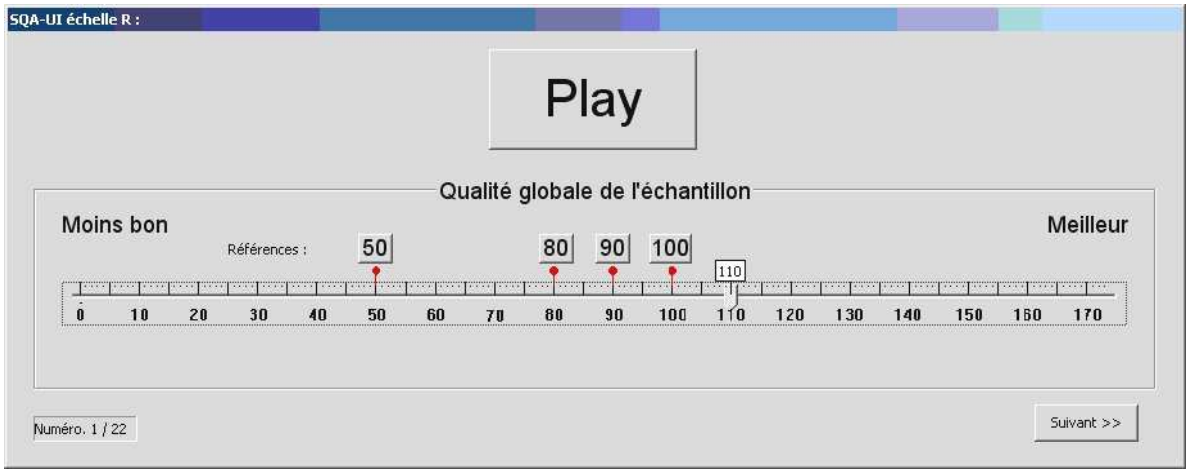

Figure 3.2 : Interface graphique du test utilisant l'échelle R du modèle E.

L'auditeur écoute le son à évaluer grâce au bouton PLAY, et les références en appuyant sur le bouton correspondant. Chaque son peut être écouté plusieurs fois, ce qui permet une meilleure évaluation de chaque stimulus. L'échelle R est agrandie jusque 174 pour éviter tout problème de saturation comme sur l'échelle MOS. Enfin, l'évaluation de la qualité du stimulus est choisie en positionnant le curseur sur la valeur souhaitée entre 1 et 174. Ici, le sujet devait, après validation de l'évaluation (bouton SUIVANT), appuyer sur le bouton PLAY pour écouter le stimulus suivant.

3.2.3 Echelle nominale

Enfin, un dernier test de comparaison entre les conditions, a été réalisé en France. Il évaluait le choix des sujets entre les conditions dégradées large-bande et non dégradées en bande-étroite. Seul 13 conditions ont été utilisées pour ce test afin de réduire le nombre de stimulus (Tableau 3.2).

Figure 3.3 : Interface graphique du test de comparaison par paires.

Le sujet devait écouter les deux sons avec les boutons « Play 1 » et « Play 2 », puis évaluer le stimulus ayant la meilleure qualité en appuyant sur le bouton « 1 » ou « 2 ». Chaque stimulus pouvait être joué plusieurs fois.

N°	Type	Test MOS bande-étroite	Test MOS large-bande	Test échelle R	Test comparaison par paire
1	G.711	X	X	X	X
2	G.726 (32 kbits/s)	X	X	X	X
3	G.729A	X	X	X	X
4	IS-54	X	X	X	X
5	G.726 (24 kbits/s)*2	X	X	X	-
6	300-3400 Hz	X	X	X	-
7	600-3400 Hz	X	X	X	X
8	300-2000 Hz	X	X	X	X
9	600-2000 Hz	X	X	X	-
10	G.722	-	X	X	X
11	AMR-WB (23.05 kbits/s)	-	X	X	X
12	AMR-WB (12.65 kbits/s)	-	X	X	X
13	AMR-WB (6.6 kbits/s)	-	X	X	-
14	50-7000 Hz	-	X	X	X
15	200-7000 Hz	-	X	X	X
16	600-7000 Hz	-	X	X	X
17	100-5000 Hz	-	X	X	-
18	50-3400 Hz	-	X	X	X

Tableau 3.2 : Répartition des conditions dans chaque test de perception.

3.3 Création du corpus de test

Il a fallu d'abord choisir le contenu des phrases utilisées comme stimuli. Le choix de phrases enregistrées auparavant par [Raake, 2005], du type EUROM[1] [Gibbon, 1992], a permis de disposer de matériel sonore phonétiquement riche et correspondant plus ou moins à des phrases typiques pour une conversation téléphonique. Ces phrases, dans cette version raccourcie, ont une durée de 8 à 10 secondes [Möller, 2000]. Elles ont été enregistrées avec un microphone électrostatique AKG C414 mis en directivité omnidirectionnelle et placé à environ 30 cm de la bouche des locuteurs [UTI-T Rec. P.800 Annexe B, 1996].

Les fichiers enregistrés ont ensuite été processés à travers le système temps réel de simulation de communication téléphonique de l'IKA, appelé RASS[2] [Möller, 2000 ; krebber, 2002 ; Raake, 2005]. Ce système basé sur un programme informatique et un ensemble de DSP

[1] EUROM : Base de donnée européenne de parole.
[2] RASS : Remote Access Simulation System (système de simulation d'accès à distance).

permet de faire varier tous les paramètres d'un réseau RTC/RNIS, avec une possibilité de simuler les conditions de transmission par IP.

Les fichiers d'entrée de ce système sont :

- Les fichiers sons, avec une fréquence d'échantillonnage de 32 kHz et quantifié sur 16 bits.
- Un fichier texte comprenant tous les paramètres du système (correspondant aux paramètres du modèle E), chaque ligne correspondant à une condition.

Tous les CoDecs ont été simulés à l'aide des DSP, sauf l'AMR-WB, qui n'est pas implanté dans le simulateur. Pour cette condition, la simulation du CoDec a été effectuée sur un second ordinateur où un codeur et un décodeur ont été compilés [ETSI TS 126 173, 2004]. Le nombre de fichiers à processer était de 144 au total ; 8 locuteurs par 18 conditions (2 locuteurs Allemand furent processés 2 fois par simplification). Ces fichiers audio d'entrée ont été égalisés en volume avant codage pour obtenir le même niveau de sortie. Celui-ci est vérifié afin d'obtenir un niveau ASL[1] de -26 dB_{m0V}. En fait, l'utilisation de valeur ASL [ITU-T Rec. P.56, 1993] permet de donner, perceptivement, une sensation de volume égale. Ce calcul utilise les bandes critiques vues précédemment.

3.4 Protocole expérimental

3.4.1 Entraînement des sujets

Pour réussir à comparer des conditions bande-étroite et large-bande, une des problématiques était l'absence de référence de la téléphonie large-bande pour l'utilisateur. On a donc étudié les différences de jugements de la qualité entre des sujets entraînés, « ayant une expérience », et des sujets sans expérience avec un service téléphonique large-bande. Ainsi, nous avons décidé de faire passer un entraînement à certains sujets. Celui-ci nous a permis de créer une référence plus stable de ce que pouvait être la téléphonie large-bande. Deux types de sujets ont donc été utilisés, différenciés par une écoute de séquences sonores, avant le passage des tests. Celles-ci étaient composées d'une conférence qui s'est déroulée à l'Ircam à l'occasion de la semaine du son en janvier 2005. Ces séquences étaient de cinq fois dix minutes. Chaque sujet ayant écouté une séquence de dix minutes par jour, durant les cinq jours précédant les tests. Le deuxième groupe de sujets n'avait pas à effectuer cette écoute que nous appellerons par la suite « entraînement ». Ces deux groupes de sujets étant différenciés seulement pour les tests effectués en France.

Les sujets étaient au nombre de 27 en France, dont 7 personnes ayant suivi l'entraînement. Aucun entraînement n'a été effectué sur les 23 sujets effectuant les tests en Allemagne. Les sujets Allemand étaient âgés de 20 à 30 ans, avec une moyenne de 24 ans et un écart type de 2,3 ans, les Français étaient âgés de 20 à 53 ans avec une moyenne de 26 ans et un écart type de 7 ans. Pour une majeure partie ils ne présentaient aucun trouble de l'audition. Les personnes ayant des pertes auditives ont étés exclues de l'étude.

[1] ASL : Active Speech Level, c'est un niveau de parole perceptif.

3.4.2 Interface électroacoustique

Dans [Raake, 2005] et [Dimolitsas, 1995] il est montré que l'interface utilisateur a une grande importance dans les résultats des sujets aux stimuli. Il existe, dans l'évaluation globale d'un stimulus, une interdépendance entre l'interface de présentation et la bande passante transmise par le système. Il est donc nécessaire pour contribuer au réalisme de la situation de conversation téléphonique, d'utiliser une interface ayant une ressemblance avec un téléphone ordinaire, ce réalisme étant déjà détérioré par l'utilisation d'une chambre sourde dans un laboratoire comme lieu de test.

Il a été montré qu'un téléphone à main « idéal » pourrait correspondre au Stax Phone (ou Hi-fi téléphone) développé par [Raake, 2000] :

> *« Le Hi-Fi téléphone [_] semble être approprié pour donner l'impression aux sujets d'écouter un interlocuteur pendant une conversation téléphonique ».*

Que celle-ci soit en bande-étroite ou bande élargie. Ce téléphone a été utilisé pour donner à la situation d'écoute un certain réalisme. Typiquement, les téléphones du commerce ne permettent pas l'utilisation de conditions large-bande, en raison des pertes fréquentielles suite au couplage insuffisant, surtout dans les basses fréquences, entre le haut parleur et l'oreille de l'auditeur.

Figure 3.4 : « Hifi téléphone », combinaison d'un téléphone fixe « type 7 » Allemand et d'un écouteur Stax [Raake, 2000].

Ce téléphone, Hifi téléphone, est composé d'un écouteur de casque électrostatique « Stax Lambda Pro », installé sur un téléphone du commerce « type 7 » Allemand. Ce téléphone ne permet pas d'avoir une tonalité avant l'écoute du stimulus ainsi qu'une sonnerie d'appel. Enfin, le niveau d'écoute a été calibré pour obtenir un niveau de 79 dB_{SPL} à l'oreille du sujet.

Chapitre 4 Résultats du test ACR type MOS

Les tests MOS permettent d'évaluer les conditions décrites dans le chapitre précédent et de les confronter aux résultats de la littérature. La première partie de ce chapitre exposera les résultats du test bande-étroite (test MOS 1) et tentera de retrouver les valeurs d' « Impairment Factor » *Ie* de l'ITU. Puis, la seconde partie sera consacrée aux résultats du test alliant les conditions bande-étroite et bande élargie (test MOS 2).

4.1 Test MOS 1

4.1.1 Comparaison entre Locuteur, moyenne globale :

La figure 4.1 suivante montre les résultats pour les 9 conditions utilisées dans le test MOS bande-étroite. Les résultats de chaque locuteur sont tracés et permettent de comprendre l'influence du locuteur sur la qualité globale du stimulus.

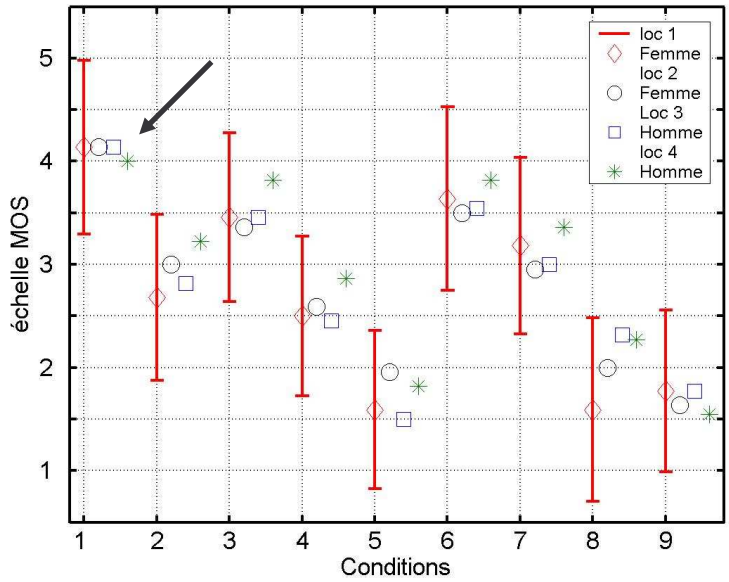

Figure 4.1 : Résultats MOS bande-étroite pour chaque locuteur.

Les résultats montrent que pour certaines conditions, comme la condition 1 (G.711), les sujets ont évalué de même qualité les stimuli quelque soit le locuteur. Un calcul d'analyse de variance (ANOVA), entre les locuteurs, montre que sur la globalité du test, il n'existe pas de différence significative entre les résultats des quatre locuteurs (p=0,131). Pourtant, pour d'autres conditions, comme la condition 8, les résultats sont assez différents suivant le locuteur. Un calcul d'ANOVA pour chaque condition montre qu'il existe une différence significative entre les quatre locuteurs pour la condition 2 (G.726 à 32 kbits/s, p=0,033) et 8 (300-2000 Hz) (p=0,04). Ceci est vérifié sur la figure 4.1.

Un autre calcul d'ANOVA entre les sujets, montre qu'il existe toujours des différences significatives dans les réponses des sujets à part pour la condition 1 (p=0,381), et dans une moindre mesure les conditions 3 (G.729) (p=0,053) et 8 (p=0,059). Cela nous indique que les différences entre les locuteurs pour la condition 8 sont voulues par l'ensemble des sujets. Il est intéressant de remarquer que pour cette condition 8, sur les 4 locuteurs, les 2 ayant la moins bonne qualité globale sont les locutrices. Cela montre que la fréquence de coupure haute à 2000 Hz touche principalement les locutrices.

L'ensemble de ces résultats permet de calculer une valeur pour chaque condition regroupant les quatre locuteurs.

Condition	Moyenne (MOS)	Ecart type (MOS)
G.711	4,10	,695
G.726 (32 kbits/s)	2,93	,657
G.729	3,52	,711
IS-54	2,60	,635
2*G.726 (24 kbits/s)	1,72	,642
300-3400 Hz	3,63	,848
600-3400 Hz	3,13	,800
300-2000 Hz	2,05	,757
600-2000Hz	1,68	,653
Total :	2,82	1,086

Tableau 4.1 : Résultats du test MOS 1.

Les résultats en valeur MOS du tableau 4.1 montrent que la condition G.711, référence de la téléphonie bande-étroite, est bien évaluée comme la condition ayant la meilleure qualité. En effet en l'absence de compression de CoDec, la qualité sonore est meilleure que pour les autres conditions de référence. De plus, la différence entre la condition G.726 avec un seul ou deux codages montre que le codage successif du même CoDec induit une dégradation plus importante.

Il faut remarquer également que la différence entre la condition G.711 et la bande passante 300-3400 Hz et simplement due à l'utilisation du filtre IRS. Ce filtre qui permet une meilleure intelligibilité de la voix. La figure 4.2 montre que l'écart type est assez important pour chaque condition, mais dans les limites typiques de ce type de test ; voir par exemple [Möller 2000].

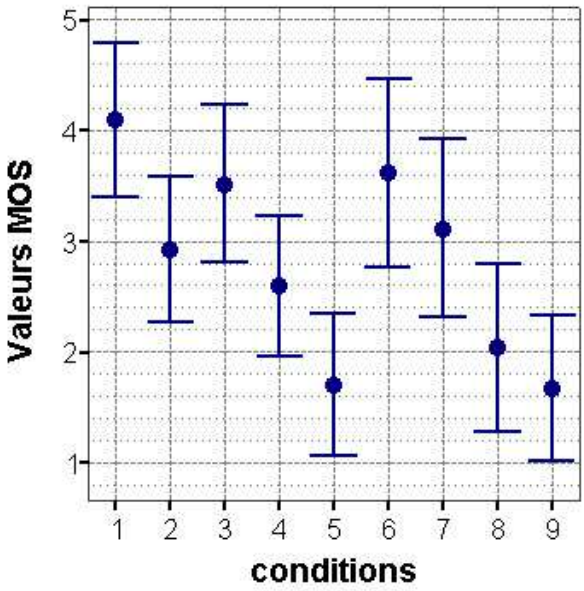

Figure 4.2 : Moyenne pour chaque condition avec écart type.

Le prochain paragraphe consistera à transformer les valeurs MOS obtenues en valeurs sur l'échelle R, puis à retrouver les *Ie* des cinq références utilisées dans ce test et à les comparer aux valeurs théoriques.

4.1.2 Valeurs sur l'échelle R, *Ie*

La première étape consiste à utiliser la formule 2.5 pour obtenir des valeurs en échelle R, puis à d'utiliser [ITU-T Rec. P.833, Appendice C, 2001] pour obtenir les *Ie*. Les valeurs obtenues *Ie,sub* (pour « subject »), sont calculées en fonction de la valeur en échelle R de la condition de référence G.711.

$$Ie,sub = R(\text{G.711}) - R(\text{autres conditions}) \tag{4.1}$$

Il est nécessaire par la suite de tracer les points *Ie,sub* en fonction des valeurs attendues *Ie,exp* (pour « expected »), puis d'obtenir la droite passant au plus proche des points tracés, au sens des moindres carrés (figure 4.3) pour ensuite transformer toutes nos valeurs (références et bandes passante) en valeurs normalisées sur l'échelle R. La courbe linéaire de la figure 4.2 a été calculée avec seulement quatre références. En effet, les valeurs de bruit de ligne (*Nfor*) utilisées pour la simulation du système RNIS étaient inférieures aux bruits normalisés par l'ITU (-90 dB au lieu de -64 dB), et par conséquent le CoDec G.726 (24 kbits/s)*G.726 (24 kbits/s) qui produit un bruit de quantification important a été évalué de moins bonne qualité que la valeur théorique. A part le niveau de bruit de ligne moins élevé, d'autres tests sur la qualité du CoDec G.726 avait aussi montré une influence plus importante du CoDec [Möller, 2000]. De plus, une interface large bande a été utilisée, ce qui a augmenté les dégradations perçues du CoDec G.726.

Le calcul des nouvelles valeurs R utilise le coefficient de la courbe obtenu (a=1,0836, RMSE[1]=128) pour trouver de nouvelles valeurs *Ie*.

$$Ie,sub = a \cdot Ie,exp$$

[1] RMSE : root mean squared errors, correspond à l'erreur introduite par la courbe au sens des moindres carrés.

Puis calculer les nouvelles valeurs normalisées sur l'échelle R, en fonction de la valeur pour la référence G.711.

$$R\_833 = 93.2 - Ie\_833 \tag{4.3}$$

Condition	Moyenne (R)	*Ie,sub* (R)	*Ie,exp* (R)	*Ie*_833 (R)	R_833 (R)
G.711	82.1404	0	0	0	93,2000
G.726 (32 kbits/s)	56.7768	25.3636	7	23,4075	69,7925
G.729	68.4341	13.7063	10	12,6493	80,5507
IS-54	50.5191	31.6213	20	29,1826	64,0174
2*G.726 (32 kbits/s)	32.5162	49.6242	50	45,7971	47,4029
300-3400 Hz	70.5997	11.5407	-	10,6506	82,5494
600-3400 Hz	60.4833	21.6571	-	19,9869	73,2131
300-2000 Hz	39.6207	42.5197	-	39,2405	53,9595
600-2000Hz	31.7286	50.4118	-	46,5240	46,6760

Tableau 4.2 : Calcul des valeurs normalisées sur l'échelle R des conditions du test MOS 1.

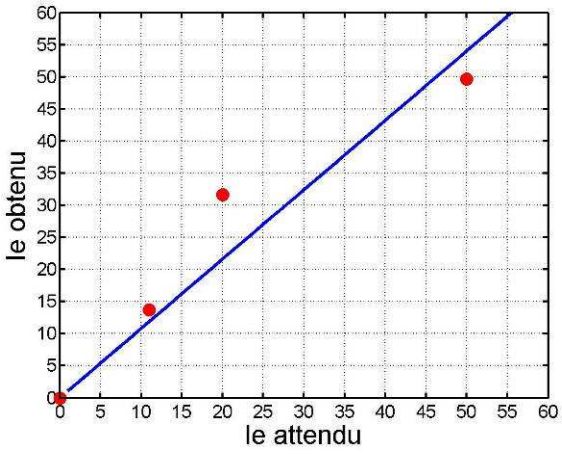

Figure 4.3 : Courbe d'interpolation des *Ie*, obtenus pour le test MOS 1, suivant [ITU-T Rec. P.833, 2001].

Le tableau 4.2 montre que les références sont évaluées comme ayant une plus grande dégradation que les valeurs théorique. Le bruit de ligne étant plus faible que sur une vraie ligne téléphonique, les dégradations dues aux CoDecs sont perceptivement plus fortes.

4.2 Test MOS 2

4.2.1 Résultats

Le deuxième test MOS alliant condition bande-étroite et large-bande va permettre de comparer les références dans ces deux tests et ainsi de trouver une première valeur sur l'échelle R d'une condition de référence large-bande. Comme pour le test MOS 1, une comparaison entre les locuteurs est d'abord effectuée pour connaître l'influence du sexe du locuteur sur les différentes conditions. Par la suite une valeur globale pour l'ensemble des locuteurs est calculée.

Valeur MOS	1	2	3	4	5	Total
% test MOS 1	11,9	28,3	31,9	22,1	5,8	100,0
% test MOS 2	10,3	27,9	31,5	19,9	10,4	100,0

Tableau 4.3 : Répartition des réponses sur l'échelle MOS.

Ce tableau montre que la répartition des réponses sur l'échelle MOS entre les deux tests est différente. En effet, on obtient le double pour la valeur 5. Cela montre une augmentation globale de la qualité sur l'ensemble des stimuli.

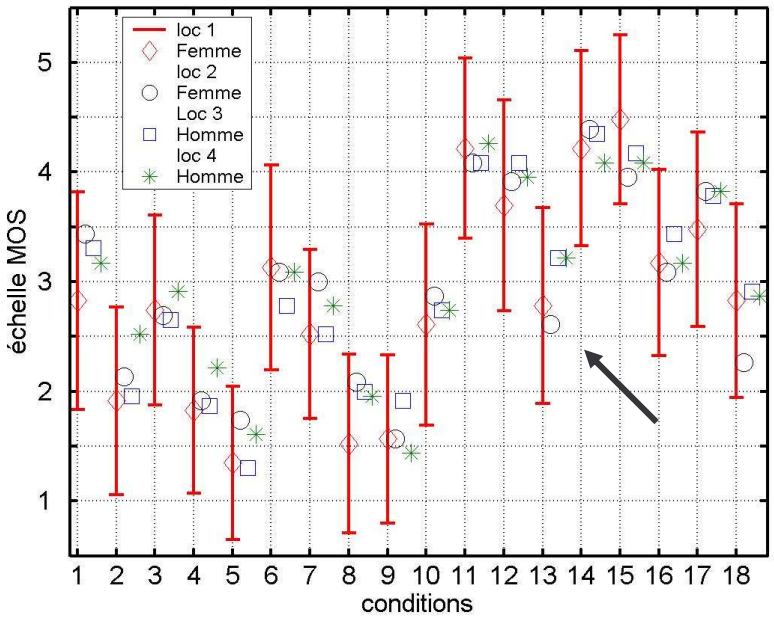

Figure 4.4 : Résultats du test MOS 2 pour chaque locuteur.

Ici les mêmes stimuli bande-étroite sont utilisés. Or ceux-ci sont bien évalués comme ayant une moins bonne qualité par rapport au test précédent. De plus, on remarque qu'il existe une plus grande différence entre les locuteurs sur les conditions de référence. Un calcul d'ANOVA montre que les conditions large-bande ont peu de différences entre les locuteurs. Par exemple un calcul pour la condition 11 (G.722.2 à 23,05 kbits/s), montre que la moyenne est identique quelque soit le locuteur ($F=0,37$; $p=0,7768$), et qu'il y a peu de différences entre les réponses des sujets ($F=1,34$; $p=0,1818$). De plus on remarque que pour la condition 13 (G.722.2 à 6,06 kbits/s), il existe des différences significatives entre les locuteurs ($F=4,49$; $p=0,063$), mais que celle-ci est très défavorables aux locutrices.

Cependant un calcul sur l'ensemble des conditions montre qu'il n'existe pas de différence significative entre les locuteurs (F=1,623 ; p=0,182). Le tableau montre donc la moyenne pour chaque condition en valeur MOS et, par transformation de celle-ci avec la formule (2.5), en valeur sur l'échelle R.

condition	Moyenne (MOS)	Ecart type (MOS)	Moyenne (R)	Moyenne (MOS 1)
G.711	3,18	,876	61,6446	4,10
G.726 (32 kbits/s)	2,13	,744	41,3458	2,93
G.729	2,75	,673	53,3235	3,52
IS-54	1,96	,610	37,7768	2,60
2*G.726 (24 kbits/s)	1,50	,584	27,2688	1,72
300-3400 Hz	3,02	,825	58,4948	3,63
600-3400 Hz	2,71	,704	52,4989	3,13
300-2000 Hz	1,89	,703	36,3946	2,05
600-2000Hz	1,62	,608	30,2550	1,68
G.722	2,74	,797	53,1174	
G.722.2 (23,05 kbits/s)	4,16	,730	83,9168	
G.722.2 (12,65 kbits/s)	3,91	,847	77,1784	
G.722.2 (6,6 kbits/s)	2,96	,876	57,2478	
50-7000 Hz	4,26	,797	87,0731	
200-7000 Hz	4,17	,765	84,2474	
600-7000 Hz	3,22	,782	62,2819	
100-5000 Hz	3,73	,813	72,8634	
50-3400 Hz	2,72	,816	52,7051	
Total :	2,92	1,141	57,2019	2,82

Tableau 4.4 : Résultats du test MOS 2.

Une des premières informations données par la figure 4.5 est la préférence de presque toutes les conditions large-bande sauf la 10, 13 et 18. Les valeurs des conditions 10 et 13 montrent qu'une dégradation trop importante du signal liée à un CoDec est plus importante dans l'évaluation globale du stimulus que l'augmentation de la bande passante. Les trois conditions du CoDec AMR-WB montrent effectivement que la qualité de la transmission diminue quand le débit diminue. Enfin, nous observons que l'augmentation de la bande passante dans les fréquences basses sans changer la fréquence de coupure haute, induit une qualité inférieure du stimulus que la référence bande-étroite G.711. Ces fréquences peuvent donc jouer un rôle négatif dans la compréhension du message malgré une sensation de la voix plus proche de la réalité. Il faut aussi remarquer qu'une légère différence, non significative (F=0,571 ; p=0,451) de 0,1 point MOS, existe entre les deux conditions 50-7000 Hz et 200-7000 Hz, la première étant préféré sur la deuxième, comme dans [krebber, 1995]. L'inverse est obtenu dans [Raake, 2005], où l'écart entre ces conditions est de 0,4 point MOS (3,8 pour 50-7000 Hz et 4,2 pour 200-7000 Hz). Il faut noter également qu'il trouve pour la condition 100-5000 Hz 3,5 point MOS, or ici cette même condition obtient 3,7. Ce test a peut être un défaut de saturation de la qualité du à l'échelle MOS. Cependant, dans [Raake, 2005], les valeurs pour la condition de référence bande-étroite G.711, dans deux test l'un en bande-étroite et le second comprenant des conditions bande-étroite et bande élargie, sont de 4,14

points MOS pour le premier test et 3,21 pour le second test. Ce qui est assez proche des valeurs obtenues dans les tests de ce mémoire (respectivement 4,10 et 3,18). De même, pour la condition 200-7000 Hz, le test MOS 2 donne un résultat de 4,17 points MOS, ce qui est très proche de la valeur trouvée dans [Raake, 2005], qui est de 4,16.

Un calcul d'amélioration entre certaines conditions large-bande et la référence G.711, nous donne :

- Une amélioration de 1,08 points MOS pour la condition 50-7000 Hz (soit 34 %).
- Une amélioration de 0,99 points MOS pour la condition 200-7000 Hz (soit 31 %).

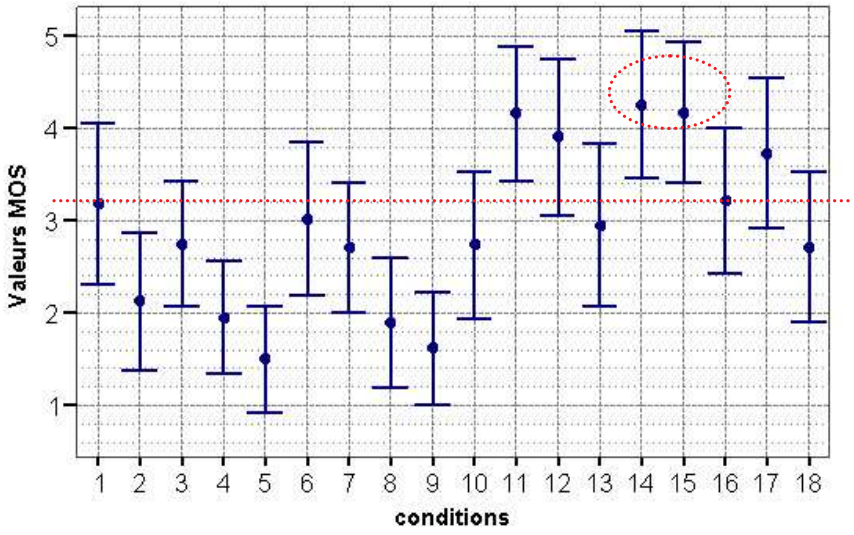

Figure 4.5 : Résultats avec écart type pour le test MOS 2.

4.2.2 Comparaison

Dans un premier temps, il est utile de trouver la relation entre les deux tests. Celle-ci peut être linéaire ou non. La figure 4.6, montre la relation entre les valeurs bande-étroite du test MOS 1 et MOS 2. Deux courbes sont tracées en plus des points. La première propose une interpolation linéaire et la seconde, une interpolation non linéaire. Les valeurs MOS du test 1 sont utilisées avant la normalisation pour obtenir les *Ie*.

Nous observons que la courbe non linéaire, du type « $y = a * x^2 + b * x$ », semble mieux correspondre à la relation entre les deux tests, que la fonction linéaire « $y = c * x$ ».

Les valeurs obtenues sont : a=-0,0411, b=0,9422 et c=0,8089. Mais une comparaison plus approfondie des deux tests MOS permettra dans le chapitre 8 de trouver une extension de l'échelle R du modèle E. En effet ce mapping entre les deux tests nous donne une idée de la relation permettant de comparer les résultats de l'un et de l'autre. Cependant il est nécessaire d'utiliser l'échelle R pour relier cette transformation au modèle E en s'écartant de l'échelle d'un test de jugement. C'est dans cette optique que le chapitre suivant montre les résultats d'un test de perception utilisant l'échelle R comme échelle de qualité.

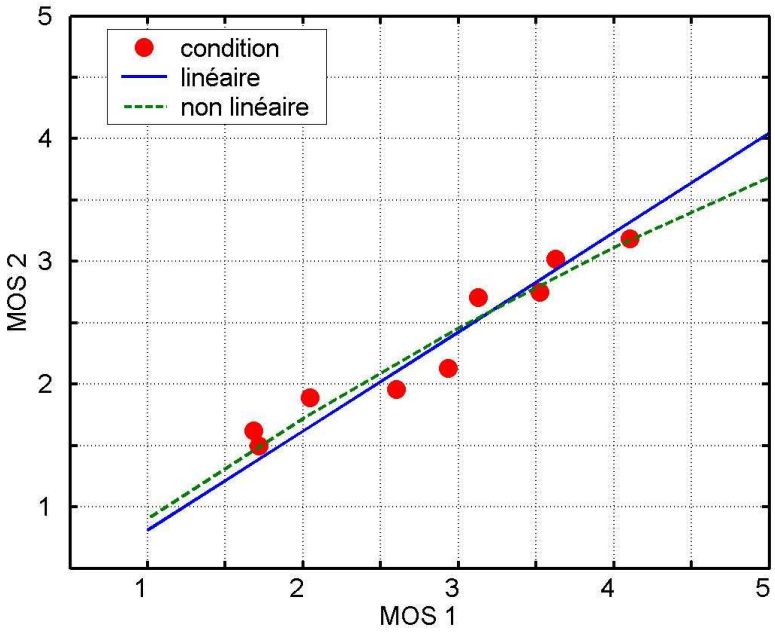

Figure 4.6 : Fonction de mapping entre MOS 1 et MOS 2.

En conclusion de ce chapitre, la comparaison entre les deux tests MOS, on montré une nette amélioration de la qualité de la parole lors d'une transmission téléphonique, grâce à l'utilisation d'une bande passante élargie (d'environ 1 points MOS ou 30 %, entre la référence bande-étroite G.711 et les meilleures conditions large-bande).

Chapitre 5 Résultats du test avec échelle R

Ce test utilise les 18 conditions bande-étroite et bande élargie. L'interface de jugement correspond à une échelle R du modèle agrandie au dessus de 100 (100 correspond au maximum de qualité pour la téléphonie bande-étroite) et sur laquelle quatre boutons permettaient d'écouter le stimulus utilisant les quatre CoDecs de références (Cf. paragraphe 3.2.2). En effet, la même phrase était jouée quelque soit la condition à évaluer par le sujet. Ceci permettait au sujet de mémoriser la qualité des références afin de les comparer plus facilement aux stimuli. Ces quatre références étaient :

- La référence bande-étroite, le CoDec G.711, positionné à 100 sur l'échelle R.
- Le CoDec G.729, positionné à 90.
- Le CoDec IS-54, positionné à 80.
- Enfin le tandem de CoDecs G.726 à 24 kbits/s, positionné à 50.

Ce test R a été effectué dans les deux pays France et Allemagne avec à chaque fois deux locuteurs, un masculin et un féminin. Certains sujets ayant passé le test en France ont suivi des séances d'écoute précédant le passage du test, afin de leur donner une idée de ce qu'était un service téléphonique large-bande. Il est important de préciser également que tous les stimuli correspondaient à la même phrase processée avec des conditions différentes. Ainsi l'évaluation de la qualité ne prend pas en compte le sens de la phrase prononcée.

Ce chapitre se compose de la description de manière globale des résultats pour les 4 locuteurs, puis, d'une manière plus détaillée, d'une interprétation des résultats en fonction du pays où fut passé le test et en fonction du locuteur.

Enfin, une dernière partie sera consacrée à la comparaison des résultats en fonction de l'entraînement ou non des sujets.

5.1 Résultats

La figure 5.1 montre les résultats pour l'ensemble des locuteurs, les deux Français et les deux Allemands. On remarque ici des variations plus importantes entre les locuteurs pour chaque condition. Pour certaines conditions, il existe deux groupes distincts entre les deux locuteurs Allemands (diamant et rond) et les deux Français (carré et étoile).

Pour les conditions 10 à 18 (celles en bande élargie), on observe qu'une proportion plus importante de points se situe au dessus de la barre des 100. Cette ligne représente la qualité théorique de la référence bande-étroite, le G.711. Cette figure montre aussi que les écart - types sont moins importants pour les quatre conditions de référence qui sont présentes sur l'échelle. Les sujets sont donc cohérents dans leurs réponses, ils semblent avoir réussi à comparer les références présentes sur l'échelle et les stimuli qu'ils devaient écouter.

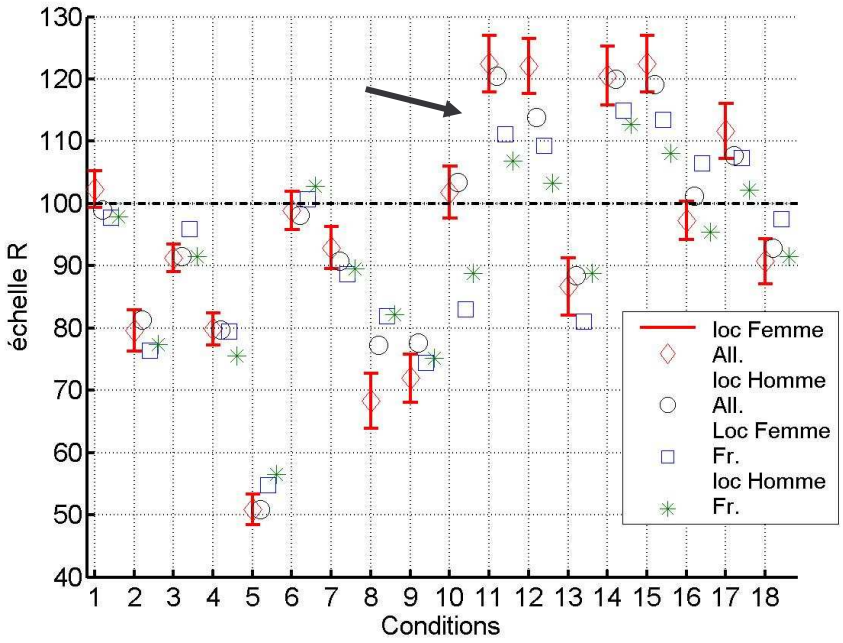

Figure 5.1 : Résultats pour l'ensemble du test utilisant l'échelle R.

5.2 Comparaison entre locuteurs

5.2.1 Test Allemagne :

Un test d'ANOVA entre les locuteurs, sur l'ensemble des conditions montre qu'il n'existe pas de différence significative entre les locuteurs (F=,005 ; p=0,946). Le tableau 5.1 montre les résultats pour les deux locuteurs. Un second calcul d'ANOVA pour chaque condition montre qu'il n'y a pas de différence entre les deux locuteurs quelle que soit la condition. Pourtant, la condition G.722.2 à 12,65 kbits/s (F=2,921 ; p=0,095) semble plutôt désavantageux pour le locuteur homme. De même, la condition 300-2000 Hz, montre une différence entre les deux locuteurs (F=2,127 : p=0,153), la fréquence de coupure haute à 2000 Hz semble toucher principalement la locutrice. De plus, la condition G.726 à 32 kbits/s, semble également être évaluée de moins bonne qualité que sa valeur théorique (qui est normalement de 93). Les résultats semblent donc cohérents avec le test MOS 2.

condition	Moyenne Homme	Ecart type	Moyenne Femme	Ecart type
G.711	99,05	5,084	102,29	8,615
G.726 (32 kbits/s)	81,33	8,639	79,62	10,925
G.729	91,52	3,544	91,24	4,816
IS-54	79,76	1,513	79,86	6,560
2*G.726 (24 kbits/s)	50,90	7,739	50,86	5,833
300-3400 Hz	98,14	11,573	98,86	9,345
600-3400 Hz	90,81	6,161	92,90	11,322
300-2000 Hz	77,29	14,217	68,29	19,499
600-2000Hz	77,67	14,434	71,95	14,746
G.722	103,38	17,834	101,81	17,509
G.722.2 (23,05 kbits/s)	120,48	18,597	122,43	20,289
G.722.2 (12,65 kbits/s)	113,90	17,216	122,10	19,131
G.722.2 (6,6 kbits/s)	88,48	19,884	86,67	21,355
50-7000 Hz	120,00	21,881	120,52	22,200
200-7000 Hz	119,19	20,532	122,48	20,868
600-7000 Hz	101,19	13,512	97,29	9,514
100-5000 Hz	107,67	12,959	111,67	19,417
50-3400 Hz	92,90	15,016	90,76	13,126
Total :	95,20	22,370	95,09	24,832

Tableau 5.1 : Résultats du test avec l'échelle R pour les locuteurs Allemand.

5.2.2 Test France

Pour les résultats du test R effectué en France, dans un premier temps, seuls les sujets non entraînés sont pris en compte.

Les mêmes tests d'ANOVA que pour le test effectué en Allemagne montrent que pour ce test ci il n'y a pas de différence significative entre les locuteurs sur l'ensemble des conditions (F=0,888 ; p=0,346) ni pour une condition particulière. Cependant certaines conditions montrent des différences entre les locuteurs, conditions différentes du test effectué en Allemagne. Par exemple la condition G.722.2 à 6,6 kbits/s (F=1,859 ; p=0,184), se révèle ici désavantageuse pour la locutrice. La bande passante 600-7000 Hz, montre également une différence entre les deux locuteurs (F=3,210 ; p=0,084). En effet, une fréquence de coupure basse semble très désavantageuse pour le locuteur. La condition 50-3400 Hz, montre à l'inverse un désavantage pour le locuteur en raison de l'accentuation trop prononcée des fréquences graves (F=1,859 ; p=0,184). Enfin, La condition G.726 à 32 kbits/s semble avoir le même problème que pour le test effectué en Allemagne.

condition	Moyenne Homme	Ecart type	Moyenne Femme	Ecart type
G.711	97,93	7,005	97,73	4,713
G.726 (32 kbits/s)	77,33	13,200	76,40	11,667
G.729	91,47	5,489	95,93	9,270
IS-54	75,60	11,969	79,47	7,615
2*G.726 (24 kbits/s)	56,47	12,194	54,80	10,557
300-3400 Hz	102,80	11,839	100,67	10,913
600-3400 Hz	89,53	8,271	88,73	13,285
300-2000 Hz	82,13	19,324	81,93	15,281
600-2000Hz	75,13	13,092	74,40	16,561
G.722	88,80	13,203	83,07	17,714
G.722.2 (23,05 kbits/s)	106,80	13,311	111,20	13,246
G.722.2 (12,65 kbits/s)	103,27	10,032	109,33	9,225
G.722.2 (6,6 kbits/s)	88,80	16,372	81,07	14,646
50-7000 Hz	112,80	16,032	114,93	13,177
200-7000 Hz	108,13	9,234	113,47	11,432
600-7000 Hz	95,47	11,103	106,53	21,189
100-5000 Hz	102,13	15,108	107,40	11,012
50-3400 Hz	91,47	13,851	97,53	14,667
Total :	91,45	18,544	93,03	20,496

Tableau 5.2 : Résultats du test avec l'échelle R pour les locuteurs Français.

Ce paragraphe n'a pas réussi à donner des résultats cohérents en prenant en compte une différenciation entre les locuteurs. Nous pourrons donc par la suite prendre en compte une moyenne des deux locuteurs. Le prochain paragraphe en faisant une comparaison entre les conditions, permettra de comprendre les préférences des sujets entre la bande passante ou le bruit induit par un CoDec.

5.3 Comparaison entre conditions

Comme dans le paragraphe précédent, les résultats pour le test France ne prennent en compte que les sujets n'ayant pas suivi l'entraînement. Cela nous permet d'effectuer une comparaison non biaisée entre les pays.

Dans un premier temps, une comparaison entre les conditions permet d'observer que la qualité des stimuli, pour les conditions G.722.2, baisse avec le débit. Puis, un calcul d'ANOVA entre les conditions, en comparant chaque condition à la référence G.711 montre que :

- Pour le test effectué en Allemagne, il existe une différence significative entre la référence bande-étroite et toutes les conditions large-bande sauf pour les conditions 300-3400 Hz, G.722 et 600-7000 Hz.

- De même, pour le test en France, il existe également une différence significative sauf pour les conditions G.729, 300-3400 Hz, 600-7000 Hz, 100-5000 Hz et 50-3400 Hz.

Nous pouvons en conclure que la fréquence de coupure basse à 50 Hz, est considérée comme une dégradation pour les sujets Allemands mais pas pour les Français. En reprenant le tableau 3.1, la description des conditions montre la fréquence centrale et la taille de la bande passante en Barks. Ici, une augmentation de 2 Barks de la bande passante, pour cette condition, comme l'augmentation de 1,5 Barks de la condition 600-7000 Hz, ne semble pas être suffisante pour améliorer la qualité de la transmission. La fréquence centrale, qui double ou diminue de moitié pour ces conditions, semble donc avoir sur la qualité, une influence plus importante que la grandeur spectrale [voir Raake, 2005].

Enfin, la différence significative la plus grande, pour les deux pays, est la condition 200-7000 Hz.

- Pour la France : $F=32,55$; $p<0,001$.
- Pour l'Allemagne : $F=29,10$; $p<0,001$.

Cependant, il n'existe pas de différence significative entre les conditions 200-7000 Hz et 50-7000 Hz. Nous ne pouvons donc pas conclure quant à l'utilité de la transmission des fréquences inférieures à 200 Hz.

condition	Moyenne Allemagne	Ecart type	Moyenne France	Ecart type
G.711	100,67	7,176	97,83	5,867
G.726 (32 kbits/s)	80,48	9,766	76,87	12,250
G.729	91,38	4,179	93,70	7,822
IS-54	79,81	4,702	77,53	10,051
2*G.726 (24 kbits/s)	50,88	6,769	55,63	11,239
300-3400 Hz	98,50	10,395	101,73	11,240
600-3400 Hz	91,86	9,065	89,13	10,881
300-2000 Hz	72,79	17,459	82,03	17,117
600-2000Hz	74,81	14,699	74,77	14,673
G.722	102,60	17,473	85,93	15,625
G.722.2 (23,05 kbits/s)	121,45	19,248	109,00	13,238
G.722.2 (12,65 kbits/s)	118,00	18,447	106,30	9,959
G.722.2 (6,6 kbits/s)	87,57	20,400	84,93	15,761
50-7000 Hz	120,26	21,773	113,87	14,460
200-7000 Hz	120,83	20,514	110,80	10,565
600-7000 Hz	99,24	11,710	101,00	17,548
100-5000 Hz	109,67	16,430	104,77	13,263
50-3400 Hz	91,83	13,972	94,50	14,352
Total :	95,15	23,618	92,24	19,542

Tableau 5.3 : Résultats pour les deux pays

Une indication peut nous être fournie par [Moore et Tan, 2003], car il existe une différence assez importante entre ces conditions lors d'une écoute au casque de manière diotic. L'écoute de ces fréquences semble donc ne pas être importante lors d'une présentation sur une seule oreille. Cela pourrait avoir un lien avec une écoute de basse fréquence en champ libre, qu'un individu ne perçoit pas comme spatialisée.

Enfin, le tableau 4.5 montre également des différences parfois importantes entre les deux pays pour certaines conditions, ce qui est confirmé par un calcul d'ANOVA entre les 2 tests. Les conditions considérées comme les meilleurs en Allemagne sont en France évaluées avec en moyenne 10 points de moins sur l'échelle R. La diminution la plus importante étant de 16,67 points pour la condition G.722 ($F=14,96$; $p=0,0017$).

5.4 Résultats des personnes entraînées

Ce paragraphe met l'accent sur la différence entre les sujets qui ont suivi un entraînement ou non, pour le test effectué en France. La but de se paragraphe est de montrer que la référence interne de la téléphonie d'un utilisateur peut évoluer, en fonction de ses expériences.

Nous pouvons dans un premier temps regarder la différence entre la moyenne des deux locuteurs pour les sujets entraînés et ceux non entraînés (Colonne ΔR du tableau). La qualité des conditions large-bande est pour la plupart améliorée de 5 à 10 points sur l'échelle R. Seules les conditions 600-7000 Hz et G.722.2 à 6,6 kbits/s sont atténuées. L'entraînement semble donc avoir un effet sur l'évaluation de la qualité de transmission large-bande. Un calcul d'ANOVA sur l'ensemble des conditions, montre qu'il existe effectivement des différences significatives entre les deux locuteurs ($F=5,190$; $p=0,024$).

Un second calcul sur chaque condition, montre que seules les conditions G.722, 50-7000 Hz et 50-3400 Hz, obtiennent une différence entre les deux locuteurs. La condition G.722 semble induire une dégradation plus importante pour une locutrice ($F=29.94$; $p=0.0028$). Cela, confirme les résultats des sujets non entraînés, mais ici de manière réellement significative. Les deux autres conditions montrent que la transmission des fréquences basses inférieures à 200 Hz induit une amélioration assez importante de la qualité vocale de l'homme. Ces résultats sont à mettre en relation avec le protocole de l'entraînement que les sujets ont suivi. En effet, le débat écouté par les sujets ne comprenait que des locuteurs hommes. Il semblerait donc que l'augmentation de la bande passante dans le bas du spectre ait plus d'influence sur la qualité vocale que l'augmentation dans le haut du spectre. En effet, un calcul d'ANOVA entre les sujets entraînés ou non donne un résultat très différent suivant le locuteur :

- Pour la locutrice : $F=1,458$; $p=0,228$.
- Pour le locuteur : $F=5,839$; $p=0,016$.

L'entraînement a donc un effet beaucoup plus accentué sur le locuteur que sur la locutrice.

condition	Moyenne Homme	Ecart type	Moyenne Femme	Ecart type	Total :	Ecart : ΔR
G.711	99,83	,408	95,83	5,601	97,83	0
G.726 (32 kbits/s)	79,67	7,062	72,17	11,583	75,92	-0,9500
G.729	88,83	9,968	97,83	7,600	93,33	-0,3700
IS-54	77,50	6,411	80,00	,000	78,75	1,2200
2*G.726 (24 kbits/s)	51,67	4,082	54,17	13,776	52,92	-2,7100
300-3400 Hz	102,50	9,503	92,67	5,428	97,58	-4,1500
600-3400 Hz	83,50	15,083	79,50	14,419	81,50	-7,6300
300-2000 Hz	91,83	13,586	80,83	13,136	86,33	4,0300
600-2000Hz	73,33	16,476	72,17	6,616	72,75	-2,0200
G.722	100,33	7,916	92,17	10,265	96,25	10,3200
G.722.2 (23,05 kbits/s)	124,83	15,303	114,83	22,329	119,83	10,8300
G.722.2 (12,65 kbits/s)	121,00	14,792	100,50	19,665	110,75	4,4500
G.722.2 (6,6 kbits/s)	93,17	26,210	72,33	12,628	82,75	-2,1800
50-7000 Hz	131,33	19,582	113,67	9,048	122,50	8,6300
200-7000 Hz	116,67	10,857	113,50	6,411	115,08	5,0000
600-7000 Hz	82,83	15,052	94,00	9,778	88,42	-12,5800
100-5000 Hz	112,83	11,339	106,50	13,795	109,67	4,9000
50-3400 Hz	114,17	14,972	91,83	8,448	103,00	8,5000
Total :	96,99	23,702	90,25	19,591	93,62	1,3800

Tableau 5.4 : Résultats du test avec l'échelle R en France, pour les sujets ayant suivi l'entraînement.

Nous retrouvons de manière générale les mêmes résultats que pour le test MOS 2 :

- La condition G.726 à 32 kbits/s est toujours évaluée de moins bonne qualité que ce que prévoit l'ITU.
- Les conditions large-bande semblent avoir une qualité meilleure que la référence G.711, sauf la condition G.722.2 à 6,6 kbits/s, comme pour le test MOS.

Ceci montre qu'une dégradation trop importante du signal liée à un CoDec ne permet pas une amélioration grâce à la bande élargie. De plus, suivant le pays où fut effectué le test, d'autres conditions large-bande sont évaluées de moins bonne qualité que le G.711 ; En France le G.722 et en Allemagne la bande passante 50-3400 Hz.

En conclusion, nous pouvons quantifier l'amélioration d'une transmission large-bande par l'échelle R. Pour cela nous pouvons utiliser comme référence large-bande la bande passante 50-7000 Hz. Ceci donne une différence de 19,59 points sur l'échelle R pour les résultats du test en Allemagne et 16,04 pour ceux effectués en France.

Cette différence pour les sujets entraînés augmente à 24,67. Ceci montre l'effet des attentes des auditeurs sur les tests de perception. De plus, le protocole de l'entraînement, induit une différence de l'amélioration suivant le locuteur ; Une différence de 31,5 points est obtenue pour le locuteur et seulement 17,84 pour la locutrice. Cependant ces deux valeurs

sont supérieures à celles des sujets n'ayant pas suivi l'entraînement. Ce chapitre a donc permis de mettre en évidence l'amélioration évidente de la qualité de la bande élargie lorsque l'utilisateur est habitué à une écoute d'un signal large-bande. En effet, le dernier paragraphe a montré que la transmission de la fréquence fondamentale F_0, améliore la qualité du signal téléphonique.

Chapitre 6 Résultats du test de comparaison par paires

Le test de comparaison par paires a permis de montrer des préférences entre certaines conditions qui ne sont pas mises en relief dans les deux précédents tests. Un seul locuteur a été utilisé pour les stimuli, la locutrice française du test utilisant l'échelle R.

6.1 Résultats par condition

Le tableau 3.2 montre que seule une partie des conditions a été utilisée pour ce test. En effet, le but étant de comparer chaque condition entre elles, il fut nécessaire de réduire le nombre de conditions à 13 : le nombre de paires de stimuli étant alors de 78 pour chaque sujet. Ce type de test permet au sujet de choisir la transmission ayant la meilleure qualité globale entre deux stimuli sans être influencé par une échelle de qualité. Ainsi, le sujet peut analyser les différentes caractéristiques perceptives présentes dans le stimuli et ainsi pondérer chaque dimension suivant ses préférences pour chacune d'elles.

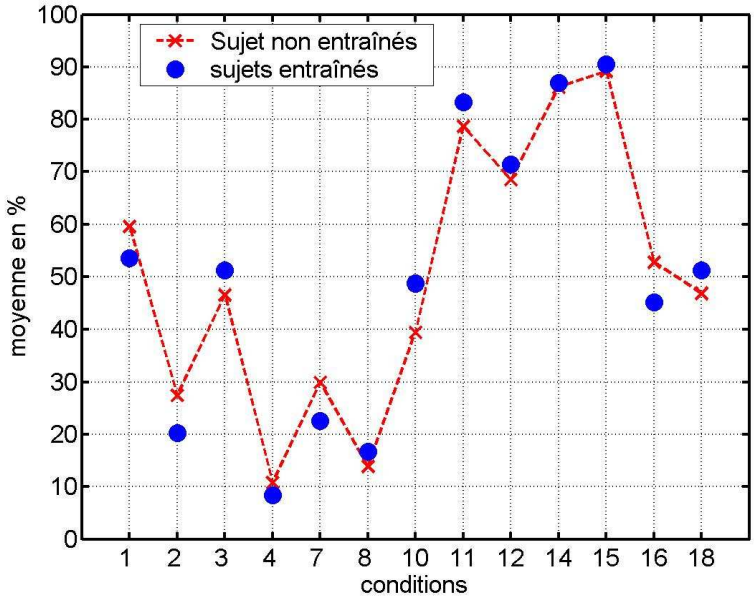

Figure 6.1 : Résultats du test de comparaison par paires pour les sujets non entraînés et entraînés.

La figure 6.1 montre la préférence de chaque condition sur les 12 autres.

Tout d'abord, la courbe « en pointillé » représente le choix des sujets n'ayant pas suivi l'entraînement. Nous remarquons que la forme de la courbe respecte les résultats obtenus dans les deux chapitres précédents. En effet, on remarque que les conditions G.722.2 à 23,05 et 12,65 kbits/s (respectivement 11 et 12), ainsi que les bandes passantes 50-7000 Hz et 200-

7000 Hz (respectivement 14 et 15) sont préférées plus souvent sur l'ensemble des conditions que la référence G.711 (1).

6.2 Résultats par comparaison

Les deux tableaux 6.1 et 6.2 montre les pourcentages de préférence de chaque condition par rapport à toutes les autres, pour les sujets non entraînés (6.1) et entraînés (6.2). On peut voir ainsi l'évolution des choix des sujets entre les différents stimuli avant et après entraînement. Les valeurs au dessus de la diagonale sont arbitrairement remplacées par « - », par désir de lisibilité du tableau. Les valeurs peuvent être retrouvées facilement par $X\{i, j\} = 100 - X\{j, i\}$.

	Stimulus de référence												
Stimulus comparé à la référence	G.711	G.726 (32 kbits/s)	G.729	IS-54	600-3400 Hz	300-2000 Hz	G.722	G.722.2 (23,05 kbits/s)	G.722.2 (12,65 kbits/s)	50-7000 Hz	200-7000 Hz	600-7000 Hz	50-3400 Hz
G.711		-	-	-	-	-	-	-	-	-	-	-	-
G.726 (32 kbits/s)	100		-	-	-	-	-	-	-	-	-	-	-
G.729	93,3	6,7		-	-	-	-	-	-	-	-	-	-
IS-54	100	86,7	100		-	-	-	-	-	-	-	-	-
600-3400 Hz	100	33,3	53,3	13,3		-	-	-	-	-	-	-	-
300-2000 Hz	93,3	80	80	60	86,7		-	-	-	-	-	-	-
G.722	66,7	53,3	53,3	26,7	53,3	26,7		-	-	-	-	-	-
G.722.2 (23,05 kbits/s)	13,3	6,7	13,3	0	6,7	0	0		-	-	-	-	-
G.722.2 (12,65 kbits/s)	13,3	6,7	6,7	0	6,7	0	26,7	60		-	-	-	-
50-7000 Hz	6,7	0	6,7	0	6,7	0	0	46,7	13,3		-	-	-
200-7000 Hz	6,7	0	6,7	0	0	0	0	40	0	60		-	-
600-7000 Hz	46,7	20	40	6,7	13,3	0	40	66,7	66,7	73,3	86,7		-
50-3400 Hz	80	40	40	0	53,3	6,7	33,3	100	93,3	100	93,3	66,7	

Tableau 6.1 : Pourcentage de préférence de la référence sur le comparé, pour chaque condition, pour les personnes sans entraînement.

Stimulus comparé à la référence	*Stimulus de référence*												
	G.711	G.726 (32 kbits/s)	G.729	IS-54	600-3400 Hz	300-2000 Hz	G.722	G.722.2 (23,05 kbits/s)	G.722.2 (12,65 kbits/s)	50-7000 Hz	200-7000 Hz	600-7000 Hz	50-3400 Hz
G.711		-	-	-	-	-	-	-	-	-	-	-	-
G.726 (32 kbits/s)	100		-	-	-	-	-	-	-	-	-	-	-
G.729	85,7	0		-	-	-	-	-	-	-	-	-	-
IS-54	100	71,4	85,7		-	-	-	-	-	-	-	-	-
600-3400 Hz	100	42,9	100	28,6		-	-	-	-	-	-	-	-
300-2000 Hz	71,4	71,4	100	28,6	85,7		-	-	-	-	-	-	-
G.722	71,4	14,3	57,1	0	0	42,9		-	-	-	-	-	-
G.722.2 (23,05 kbits/s)	0	14,3	0	0	0	0	0		-	-	-	-	-
G.722.2 (12,65 kbits/s)	28,6	0	14,3	0	0	0	42,9	71,4		-	-	-	-
50-7000 Hz	0	0	0	0	14,3	0	14,3	28,6	14,3		-	-	-
200-7000 Hz	0	0	0	0	0	0	14,3	28,6	28,6	28,6		-	-
600-7000 Hz	42,9	28,6	71,4	0	28,6	14,3	42,9	100	85,7	100	85,7		-
50-3400 Hz	42,9	0	71,4	0	14,3	0	57,1	85,7	85,7	85,7	100	42,9	

Tableau 6.2 : Pourcentage de préférence de la référence sur le comparé, pour chaque condition, pour les personnes avec entraînement.

Dans un premier temps, nous observons que la condition de référence G.711 n'est pas moins préférée aux conditions large-bande dégradées comme la condition G.722 et G.722.2 à 12,65 kbits/s. Au contraire, nous observons une augmentation de la préférence du G.711 sur le G.722.2 à 12,65 kbits/s. Mais nous constatons également une augmentation de la préférence des autres conditions large-bande non bruitées sur la référence G.711. Les trois conditions qui étaient perçues comme les meilleures pour les deux test précédents (G.722.2 à 23,05 kbits/s et les bandes passantes 50-7000 Hz et 200-7000 Hz), ne sont pas une seule fois choisies comme étant de moins bonne qualité que le G.711. Enfin, une augmentation importante (de 37 %) concerne la bande passante 50-3400 Hz qui est préférée par 57 % des sujets sur la référence G.711. Nous ne pouvons donc pas conclure quant à la préférence entre ces deux conditions pour une locutrice.

Cependant une comparant deux CoDecs bruitées, comme le G.726 à 32 kbits/s et le G.722, nous observons qu'avant l'entraînement il n'existe pas de préférence entre l'un ou l'autre (53,3 % pour le G.726), mais après l'entraînement, le G.722 est clairement préféré sur le G.726 (85,4 % pour le G.722).

Une autre comparaison peut être faite entre un CoDec bruité large-bande (le G.722) et une bande passante bande-étroite 600-3400 Hz ne transmettant pas le premier formant : Nous observons qu'avant l'entraînement il n'existe pas non plus de préférence (53,3 % pour la bande passante), or après l'entraînement tous les sujets ont préféré le CoDec large-bande. La

même observation est faite sur le CoDec bande-étroite G.729 qui passe de 53,3 % à 100 % de préférence par rapport à la bande passante 600-3400 Hz. Il semble que l'entraînement ait changé les attentes vis-à-vis de la perception des fréquences basses. Cette information semble cohérente avec les résultats obtenus dans les deux tests précédents.

Enfin, la comparaison des deux conditions large-bande 50-7000 Hz et 200-7000 Hz, montre que l'entraînement a avantagé la condition ayant la fréquence de coupure basse à 200 Hz. Le choix des sujets allant à 60 % pour la condition 50-7000 Hz avant l'entraînement et seulement à 28,6 % après l'entraînement.

Cette information montre que le choix des sujets est grandement influencé par le locuteur. Il est donc nécessaire de reproduire ce même test avec plusieurs locuteurs.

Nous avons vu dans le chapitre 2 que l'échelle R du modèle est actuellement définie pour une utilisation en bande-étroite. Sa dynamique étant de 0 à 100, où 100 représente une référence en bande-étroite, le G.711 vu précédemment. Le chapitre suivant va nous permettre d'étendre l'échelle de la bande-étroite à la large-bande. Cela nous permettra d'évaluer la qualité d'une transmission large-bande sur la même échelle qu'une transmission bande-étroite, notamment l'échelle R.

Chapitre 7 Extension du modèle E à la large-bande

Le quatrième chapitre, montrant les résultats des deux tests MOS, a permis d'obtenir les valeurs du test MOS 1 sur l'échelle R après normalisation suivant [ITU-T Rec. P.833, 2001]. Les valeurs R obtenues à partir du test MOS 2 n'ont elles pas subi cette normalisation. Le dernier paragraphe de ce quatrième chapitre avait pour but de comparer les valeurs des conditions bandes étroites, commune aux deux tests MOS. Cette comparaison sur échelle MOS nous donne l'impression que la relation entre les deux tests est non linéaire. Cependant, cette comparaison doit être faite également sur une échelle R pour nous permettre de l'étendre aux conditions large-bande. L'échelle R*lb* obtenue pourra alors faire l'objet de comparaison avec les résultats des tests décrits dans le cinquième chapitre, portant sur le test utilisant échelle R.

Le but de cette partie est de déterminer une valeur interpolée R*lb* des tests MOS du quatrième chapitre. Pour obtenir l'extension de l'échelle R, nous procéderons dans ce paragraphe à un protocole provenant de [Raake, ITU-T, COM- D.028] qui se déroule en quatre étapes :

- Normalisation des valeurs MOS.
- Transformation, de ces valeurs normalisées, de l'échelle MOS vers l'échelle R.
- Tracé des valeurs en bande-étroite obtenues avec le test MOS 1 en fonction de valeurs du test MOS 2 de ces mêmes conditions.
- Obtention d'un maximum de la nouvelle échelle R*lb* pour bande élargie, à partir des valeurs des conditions large-bande du test MOS 2, par extrapolation linéaire ou non linéaire.

Deux paragraphes vont permettre de déterminer cette valeur, le premier utilisant les valeurs R provenant directement des tests MOS (trois dernières étapes), le second utilisant la normalisation des valeurs MOS avant transformation vers l'échelle R.

7.1 Extrapolation sans normalisation :

Les valeurs R, sont reprises du quatrième chapitre.

condition	Valeur échelle R, MOS 1	Valeur échelle R, MOS 2
G.711	93,2000	61,6446
G.726 (32 kbits/s)	69,7925	41,3458
G.729	80,5507	53,3235
IS-54	64,0174	37,7768
2*G.726 (24 kbits/s)	47,4029	27,2688
300-3400 Hz	82,5494	58,4948
600-3400 Hz	73,2131	52,4989
300-2000 Hz	53,9595	36,3946
600-2000Hz	46,6760	30,2550

Tableau 7.1 : Résultats des conditions bande-étroite en échelle R, pour les deux tests MOS.

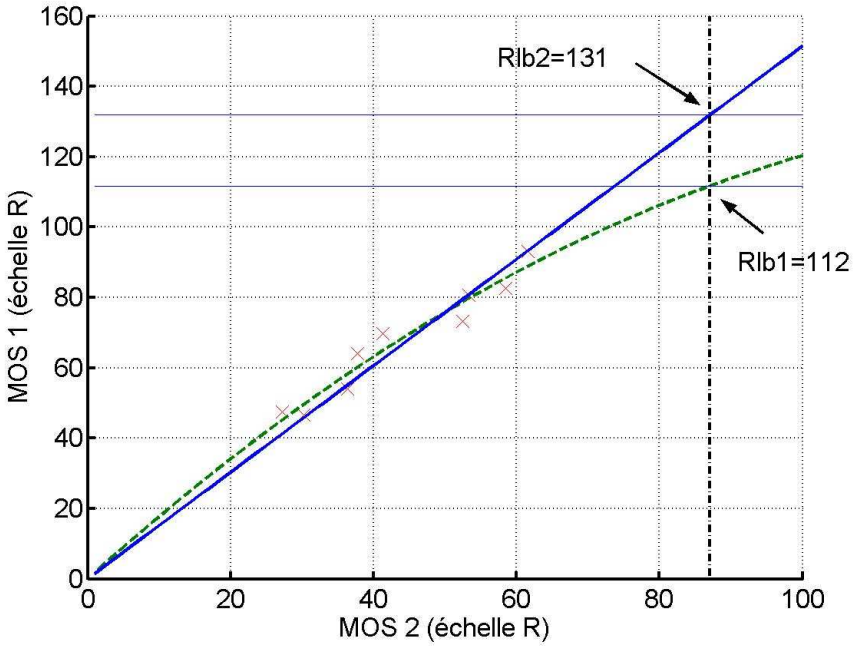

Figure 7.1 : Interpolation des résultats des conditions communes aux tests MOS 1 et MOS 2, sans normalisation.

Deux courbes sont interpolées des points correspondant aux conditions bande-étroite : Une non linéaire, du type $y = a * x^2 + b * x$, avec a=-0,0063 et b=1,8296 (RMSE=129,2). Une seconde linéaire du type $y = c * x$ avec c=1,5144 (RMSE=213,3). Celles-ci nous permettent d'extrapoler les valeurs de nouvelles échelles R.

condition	Valeur échelle R, MOS 2	Extrapolation linéaire (R*lb2*)	Extrapolation non linéaire (R*lb1*)
G.711	61,6446	93,3571	88,9438
G.726 (32 kbits/s)	41,3458	62,6157	64,9211
G.729	53,3235	80,7552	79,7214
IS-54	37,7768	57,2106	60,1628
2*G.726 (24 kbits/s)	27,2688	41,2970	45,2257
300-3400 Hz	58,4948	88,5869	85,5551
600-3400 Hz	52,4989	79,5064	78,7602
300-2000 Hz	36,3946	55,1175	58,2772
600-2000Hz	30,2550	45,8194	49,6115
G.722	53,1174	80,4430	79,4819
G.722.2 (23,05 kbits/s)	83,9168	127,0869	109,3536
G.722.2 (12,65 kbits/s)	77,1784	116,8820	103,8353
G.722.2 (6,6 kbits/s)	57,2478	86,6983	84,1790
50-7000 Hz	87,0731	131,8669	111,7424
200-7000 Hz	84,2474	127,5876	109,6096
600-7000 Hz	62,2819	94,3222	89,6143
100-5000 Hz	72,8634	110,3472	100,0025
50-3400 Hz	52,7051	79,8187	79,0014

Tableau 7.2 : Résultats de l'extrapolation, sans normalisation des résultats MOS.

La condition 50-7000 Hz, en vue des résultats des différents tests de ce mémoire, semble correspondre à la bande passante de référence d'une condition large-bande. Cette condition est donc utilisée pour obtenir les deux valeurs de R*lb*, arrondies à l'entier près, pour chaque extrapolation :

- $Rlb_1 = 131$, pour l'extrapolation linéaire.
- $Rlb_2 = 112$, pour l'extrapolation non linéaire.

Cette valeur est égale à celle trouvée dans [Raake, ITU-T, COM- D.028], qui correspond également à une interpolation sans normalisation et avec une extrapolation linéaire. Cependant cette valeur 112, est obtenue pour une bande passante 200-7000 Hz, qui est dans le cas du test MOS évaluée de moins bonne qualité que la condition 50-7000 Hz.

7.2 Extrapolation avec normalisation

Les valeurs MOS obtenues pour les deux tests MOS, ont un maximum qui est assez loin du 4,41 attendu pour un test bande-étroite et 4,5 pour un test large-bande. En effet, la valeur maximale pour le test MOS 1 correspond à la condition G.711 qui est évaluée à 4,10 et celle pour le test MOS 2 est de 4,26 pour la condition 50-7000 Hz. Ce second paragraphe reprend donc les mêmes étapes que précédemment mais en normalisant d'abord les valeurs MOS, avec la formule suivante :

$$MOS_1 = \frac{MOS_t - 1}{MOS_t(ref) - 1} \cdot (MOS_r - 1) + 1 \qquad (7.1)$$

Où MOS_r est la valeur maximale attendu, donc 4,41 pour le test bande-étroite et 4,5 pour le test large-bande. MOS_t (ref) est la valeur MOS trouvée pour la condition G.711 (référence). MOS_t sont les valeurs obtenues pour les différentes conditions. Enfin MOS_l sont les nouvelles valeurs ayant subi la transformation linéaire.

Suite à cette normalisation, les valeurs MOS sont comprises entre 4,41 et 2 points MOS pour le test MOS 1, et entre 4,5 et 2,02 pour le test MOS 2 (cette normalisation pour le test MOS 1 n'ayant pas affecté les valeurs due à la précédente normalisation [ITU-T Rec. P.833, Appendice C, 2001] du chapitre 4). Nous pouvons alors utiliser la formule 2.5 pour obtenir les résultats normalisés sur l'échelle R. Enfin, ces valeurs sur l'échelle R des neuf conditions bande-étroite du test MOS 1 sont tracées en fonction des résultats au test MOS 2. Nous obtenons la courbe de la figure 7.2. Une seule interpolation a été effectuée ici, car l'interpolation non linéaire ne donnait pas de différence significative avec celle linéaire. En effet les valeurs des coefficients sont (en reprenant les mêmes notations que dans le paragraphe précédant) ; Pour l'interpolation linéaire, c=1,4494 (RMSE=232,8), et pour l'interpolation non linéaire, a=-0,0063, b=1,78 (RMSE=129,2).

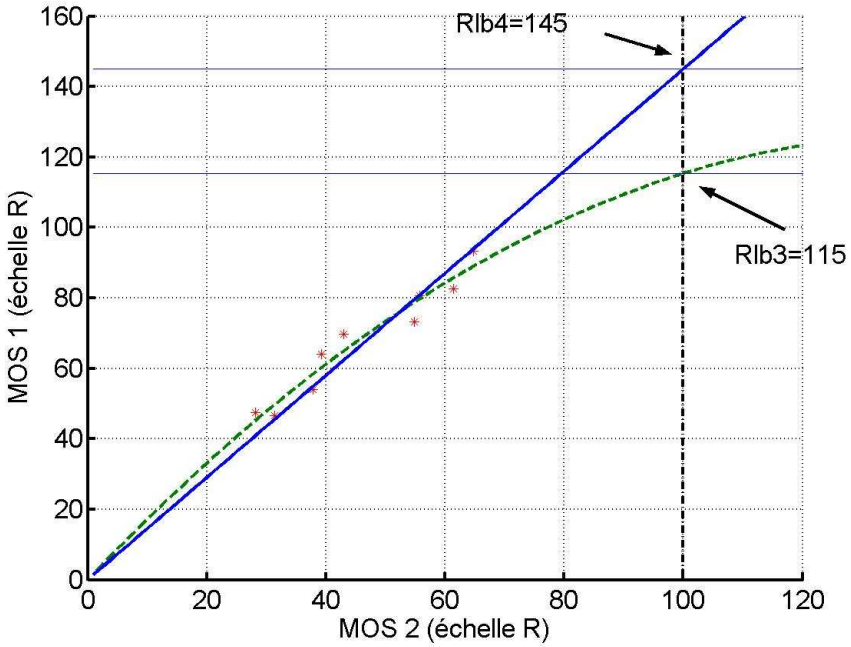

Figure 7.2 : Interpolation des résultats des conditions communes aux tests MOS 1 et 2, après normalisation.

Cette extrapolation avec une normalisation avant transformation en valeur R, permet d'obtenir deux valeurs du maximum d'une nouvelle échelle R pour la large-bande, R*lb4*=145 et R*lb3*=115. Il est intéressant de noter que dans [Raake, ITU-T, COM- D.028], une normalisation des données MOS permet d'obtenir une valeur de R*lb* de 138, pour une interpolation linéaire. Cependant, dans cette même communication, l'utilisation de résultats d'un test décrit dans [Barriac et al., 2004] permet d'obtenir une valeur de R*lb* de 129 avec une interpolation non linéaire.

condition	MOS 2 normalisé (échelle R)	Extrapolation linéaire (R*lb4*)	Extrapolation non linéaire (R*lb3*)
G.711	64,8072	93,9345	88,9991
G.726 (32 kbits/s)	43,0000	62,3262	64,9278
G.729	55,7590	80,8196	79,7350
IS-54	39,2347	56,8685	60,1681
2*G.726 (24 kbits/s)	28,2096	40,8882	45,2096
300-3400 Hz	61,3567	88,9331	85,5873
600-3400 Hz	54,8727	79,5349	78,7723
300-2000 Hz	37,7798	54,7597	58,2814
600-2000Hz	31,3352	45,4186	49,6051
G.722	55,5373	80,4983	79,4951
G.722.2 (23,05 kbits/s)	92,4840	134,0505	110,9669
G.722.2 (12,65 kbits/s)	82,8398	120,0717	104,4019
G.722.2 (6,6 kbits/s)	60,0000	86,9667	84,2049
50-7000 Hz	100,0000	144,9445	115,2750
200-7000 Hz	93,0655	134,8933	111,3255
600-7000 Hz	65,5101	94,9532	89,6758
100-5000 Hz	77,5550	112,4117	100,3100
50-3400 Hz	55,0941	79,8559	79,0137

Tableau 7.3 : Résultats de l'extrapolation avec normalisation des résultats MOS.

En conclusion, la valeur obtenue R*lb4* correspond aux valeurs décrites dans la littérature pour une interpolation linéaire, mais une augmentation de l'ordre de 54 % de la qualité semble être légèrement surélevée. En effet, le RMSE semble monter qu'une interpolation non linéaire est mieux adaptée à l'obtention d'une nouvelle échelle R et nous permet d'obtenir une amélioration de 30 % (*Rlb3*=115).

7.3 Impairment factor, en large-bande

Nous avons obtenu dans le paragraphe précédant une valeur du maximum de la nouvelle échelle R*lb* étendue à la large-bande. Nous pouvons maintenant définir deux valeurs d' « Equipment Impairment Factor » *Ie* ; l'un en bande-étroite *Ie*, et un deuxième large-bande, *Ie,lb*. En effet, les valeurs de *Ie*s, peuvent être obtenue par la différence entre la valeur donnée pour la condition G.711 sur l'échelle R et les autres conditions, la condition G.711 restant la référence des échantillons bande-étroite. La formule 4.1 permet de calculer cette valeur *Ie* lors d'un test de jugement.

La calcul de *Ie,lb* utilise d'abord un calcul de l'augmentation de la nouvelle échelle Rlb par :

$$\Delta Ie = Rlb - (G.711) \tag{7.2}$$

Soit ΔIe=115,3 - 89=26,3

Puis, pour trouver *Ie,lb* :

$$Ie,lb = Ie + \Delta Ie \tag{7.3}$$

[ref com d.29]Nous avons alors une nouvelle formule de calcul de *Ie,lb* :

$$Ie,lb(test) = R_{[0:Rlb]}(direct) - R_{[0:Rlb]}(test) \tag{7.4}$$

Où la condition directe correspond à la bande passante 50-7000 Hz, qui est une condition non codée, contrairement au G.711 qui utilise un codage logarithmique. Cependant aucune condition de référence large-bande n'est définie aujourd'hui par l'ITU.

Nous pouvons alors donner les valeurs *Ie,lb* des différents CoDecs utilisés dans les tests de ce mémoire pour l'échelle R*lb4*.

Nom du CoDec	Impairment Factor (*Ie*)	*Ie,lb*
G.711	0	26,3
G.726	24	50,3
G.729	9,3	35,5
IS-54	28,8	55,1
G.726	43,8	70,1
G.722	9,5	35,8
AMR-WB (23,05 kbits/s)	-22	4,3
AMR-WB (12,65 kbits/s)	-15,4	10,9
AMR-WB (6,6 kbits/s)	4,8	31,1

Tableau 7.4 : Valeur des *Ie* et *Ie,lb* pour les CoDecs utilisés comme conditions dans les tests de jugement.

Ces chiffres sont obtenus à partir des valeurs du tableau 7.3 et reflètent le problème lié à la référence large-bande. En effet, on remarque que la nouvelle échelle R*lb4* obtenue dans le paragraphe précédent introduit une dégradation de 26,3 points entre la référence large-bande 50-7000 Hz et le CoDec G.711. On obtient donc un avantage de 26,3 points, ce qui semble plus réaliste. Des tests supplémentaires pour stabiliser les statistiques devront être effectués.

Enfin, il est possible de modifier la fonction de transformation des données MOS en R, pour obtenir une règle de transformation directe des valeurs MOS en valeurs de la nouvelle échelle R*lb*.

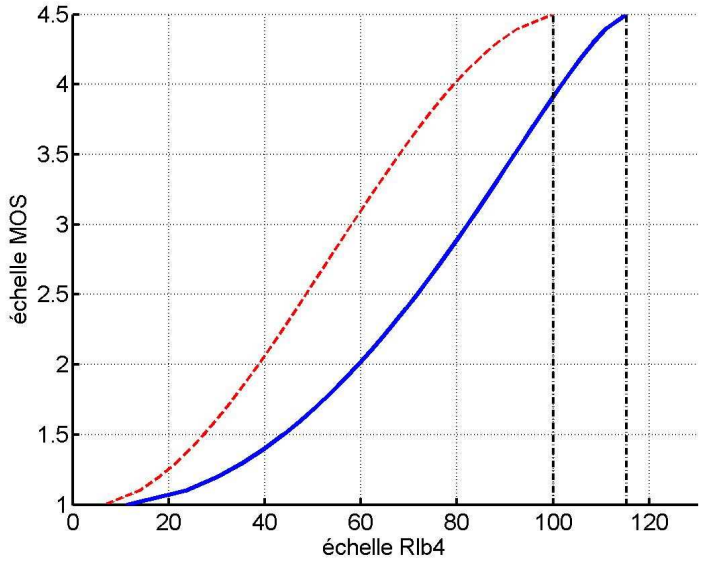

Figure 7.3 : Règle de transformation entre l'échelle MOS et les deux échelles R et R*lb4*.

7.4 Comparaison avec les résultats du test utilisant l'échelle R.

Nous pouvons alors, comparer les résultats obtenus dans le cinquième chapitre où le test de jugement utilisé l'échelle R du modèle E, avec les résultats des interpolations précédentes. Pour cela il est nécessaire d'obtenir la même valeur pour la référence G.711. En effet, le test utilisant l'échelle R positionnait cette référence à la valeur 100, pour permettre une comparaison plus facile entre les conditions. Cependant les résultats des interpolations précédentes, utilise la référence G.711 du modèle E positionnée à 93,2 (ou proche de cette valeur dus aux erreurs de l'interpolation).

Ainsi, pour les interpolations non linéaires (R*lb1* et R*lb3*), nous ajoutons 11 points aux valeurs obtenues et 6 points aux interpolations linéaires (R*lb2* et R*lb4*).

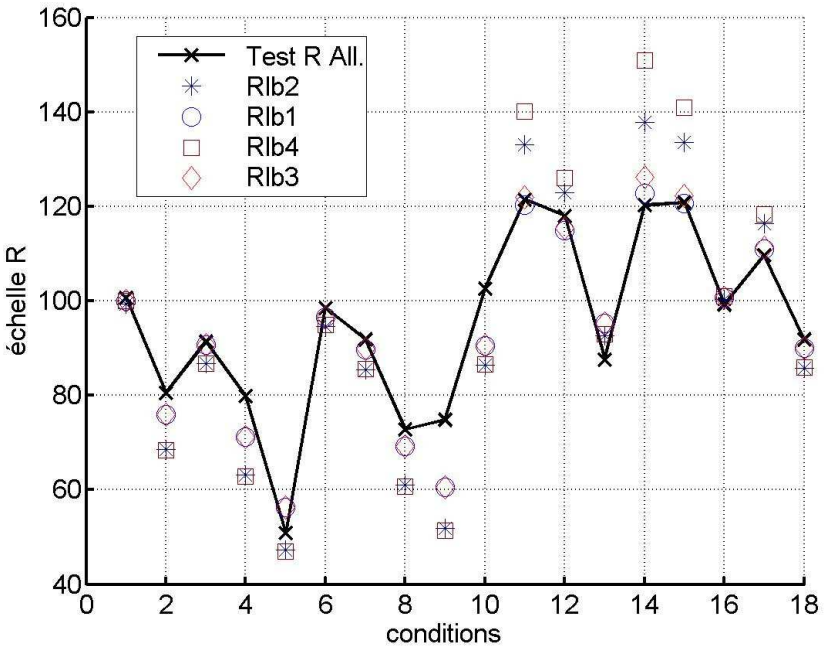

Figure 7.4 : Comparaison entre les test R et MOS.

Nous observons ici que les interpolations non linéaires (rond, R*lb1* et diamant, R*lb3*) sont assez proches des résultats du test de jugement utilisant l'échelle R, contrairement aux interpolations linéaires. Le protocole utilisant l'échelle R semble donc permettre une évaluation de stimuli large bande sans être obligé de recourir la mise en place de deux tests MOS différents, l'un en bande-étroite et le second mixant les conditions bande-étroite et large-bande.

Chapitre 8 Conclusion

Nous avons observé qu'il existe de grandes différences de qualité de la voix transmise par une système de communication suivant le CoDec utilisé, et que l'utilisation d'une transmission large-bande permettait d'améliorer la qualité de la parole perçue par l'auditeur. Nous avons observé également que certains CoDec dégradent le signal de parole si nous le comparons à la référence de la téléphonie bande-étroite, le G.711 [ITU-T Rec. G.712, 1996], et cela même pour certains codeurs large-bande, comme le G.722.2 à 6,6 kbits/s. Mais, les résultats des tests de perception ont montré que cette référence bande-étroite peut être vue comme une dégradation en la comparaison à la transmission téléphonique large-bande.

Le dernier chapitre a permis l'extension de l'échelle R du modèle E qui permet l'évaluation, sur une échelle commune R*lb* des conditions bande-étroite et large-bande. Nous obtenons ainsi un modèle paramétrique, où il suffit de varier la nouvelle valeur de dégradation du au CoDec *Ie,lb* pour déterminer la qualité globale d'une transmission. De plus, les différents tests de perception décrits dans ce mémoire ont permis de quantifier l'amélioration de la qualité de la voix transmisse grâce à l'apport de la large-bande. Nous obtenons ainsi une valeur de 1,08 points MOS (« Mean Opinion Score », [ITU-T Rec. P.800, 1996]) d'amélioration entre la référence G.711 et la bande passante 50-7000 Hz, soit 34 %. Cette valeur après transformation sur la nouvelle échelle R*lb*, nous donne une amélioration de 26 points (soit 30 %), ce qui correspond à l'amélioration la plus réaliste.

Cette valeur est à mettre en relation avec les résultats obtenus dans le cinquième chapitre où nous utilisons l'échelle R du modèle E [ITU-T Rec. G.107, 2005] pour évaluer les stimuli. Ce test donne une amélioration moyenne, entre les deux conditions G.711 et 50-7000 Hz, de l'ordre de 20 points sur l'échelle R. Le paragraphe 7.4 montre que ce protocole permet d'obtenir des valeurs proches des résultats du test utilisant l'échelle MOS. Les trois tests décrits dans ce mémoire ont donc donnés des résultats cohérents entre eux avec parfois certaines différences entre les résultats. Ceci nous montre l'importance du protocole dans la mise en place de test de jugement [Jekosch, 2005].

Cependant nous obtenons des résultats différents suivant le pays où a été effectué le test, ou lorsque les sujets ont une expérience dans la téléphonie large-bande. En effet, nous obtenons grâce à la large-bande une valeur d'amélioration de 20 points pour l'Allemagne et de 16 points pour la France. De plus, les sujets avec expérience donne un résultat de 25 points. Ainsi, nous observons qu'un changement dans les attentes du locuteur face à la téléphonie permet une amélioration des résultats.

Enfin, les résultats des sujets entraînés montre une différence suivant le locuteur. En effet, nous obtenons une amélioration de 31 points pour le locuteur et seulement 18 points pour la locutrice. Pour autant, les résultats ont montrés que c'est l'augmentation de la bande passante dans les fréquences basses et également dans les hautes fréquences qui permet une amélioration de la qualité.

Les résultats ayant montré certains problèmes lors de leurs analyses, il sera nécessaire de procéder à de nouveaux tests par la suite. En effet, l'utilisation de l'échelle R comme interface de jugement à parfois créé des problèmes pour les sujets dans le choix des références positionnés sur l'échelle. De plus, le test de comparaison par paires n'utilisait qu'une locutrice. Les différences entre les sujets entraînés et non entraînés ayant montré des résultats positifs, un nouveau test avec plusieurs locuteurs pourra confirmer les résultats obtenus ici.

De plus, il sera nécessaire de procéder à de nouveaux tests où le choix du matériel d'entraînement sera mieux égalisé afin de ne pas avantager un locuteur homme ou femme.

Enfin, il serait souhaitable d'étudier les dimensions perceptives liées aux transmissions large-bande par l'utilisation de tests analytiques (Cf. paragraphe 2.4.2.3). Cela nous permettra de mieux définir la référence à choisir pour les transmissions en bande élargie.

Néanmoins, les résultats obtenus dans ce mémoire montre que la voix d'un auditeur est perçue comme étant de meilleure qualité lors de l'augmentation de la bande passante. Celle-ci semble plus naturelle pour les auditeurs. La bande élargie pourra donc être utilisé dans certaines applications de l'industrie ; La téléphonie classique mais aussi les conversations multi-auditeurs, grâce à une meilleure identification de l'interlocuteur, ou encore les applications multimodales.

Bibliographie

ITU-T Rec.G.107 (2005). *The E-model, a Computational Model for Use in Transmission Planning*. International Telecommunication Union, CH-Geneva.

Raake, A., (2005). *Assessment and Parametric Modelling of Speech Quality in Voice-over-IP Networks*. Doctoral dissertation, Institut für Kommunikationsakustik, Ruhr-Universität, DE-Bochum.

Möller, S. (2000). *Assesment and Prediction of Speech Quality in Telecommunications*. Kluwer Academic Publishers, USA-Boston.

Jekosch, U. (2000). *Spache Hören und Beurteilen: Ein Ansatz zur Grundlegung der Sparchqualitätsbeurteilung*. Habilitation thesis, UniversitÄt/Gesamthochschule, DE-Essen.

Jekosch, U. (2005). *Voice and Speech Quality Perception – Assessment and Evaluation*. Springer, DE-Berlin.

ITU-T Rec. E.800 (1994). *Terms and Definition Related to Quality of Service and Network Performance Including Dependability.* CH-Geneva.

Duncanson, J. P. (1969). *The Average Telephone Call Is Better than the Average Telephone Call*. The Public Opinion Quarterly, 33(1), 112-116.

Zwicker, E. (1961). *Subdivision of the Audible Frequency Range into Critical Bands*. J. Acoust. Soc. Am., Vol. 33 (B, 248).

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and Models*. Springer, DE-Berlin.

Gleiss, N. (1989). *Desirable Sending Frequency Response of Telephone Sets*. TELE (edition anglaise), 1/89, 18-23, Swedish Telecommunications Administration, SU-Stockholm.

O'Shaughnessy, D. (2000). *Speech Communication – Human and Machine*. 2ème edition. IEEE Press, USA-Piscataway.

Gleiss, N. (1970). *The effect of Bandwidth Restriction on Speech transmission Quality in Telephony*. Proc. 4th Int. Symp. On Human Factors in Telephony, 1-6, VDE-Verlag, DE-Berlin.

Moore, B. C. J. and Tan, C-T. (2003). *Perceived naturalness of spectrally distorted speech and music*. J. Acoust. Soc. Am. Vol. 114 (A, 408-19).

Tanenbaum, A. S. (2003). *Computer Networks*. Prentice Hall, USA-Upper Saddle River.

ITU-T Rec. G.712 (1992). *Transmission Performance Characteristics of Pulse Code Modulation.* International Telecommunication Union, CH-Geneva.

ITU-T Rec. G.711 (1988). *Pulse Code Modulation (PCM) of Voice Frequencies.* International Telecommunication Union, CH-Geneva.

ITU-T Rec. G.107 (2005). *The E-Model, a Computational Model for Use in Transmission Planning.* International Telecommunication Union, CH-Geneva.

ITU-T Rec. G.722 (1988). *7 kHz Audio-Coding Within 64 kbits/s.* International Telecommunication Union, CH-Geneva.

ITU-T Rec. G.722.2 (2002). *Wideband Coding of Speech at Around 16 kbits/s Using Adaptative Multi-Rate Wideband (AMR-WB).* International Telecommunication Union, CH-Geneva.

ETSI TS 126 173 (2004). *Universal Mobile telecommunications System (UMTS); ANSI-C code for the Adaptative Multi-Rate-Wideband (AMR-W) speech codec (3GPP TS 26.173).* European Telecommunications Standards Institute, FR-Sophia Antipolis.

ITU-T Rec. P.56 (1993). *Objective Measurement of Active Speech Level*. International Telecommunication Union, CH-Geneva.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, USA-Cambridge MA.

Jekosch, U. (2004). *Basic Concepts and Terms of "Quality", Reconsidered in the Context of Product Sound Quality*. Acta Acoustica united w. Acoustica.

Mariani, J. (2002) (Dir.). *Analyse, Synthèse et Codage de la Parole: Traitement Automatique du Langage Parlé 1.* Hermes Sciences Publications ; Lavoisier, FR-Paris.

Letowski, T. (1989). *Sound Quality Assessment: Concepts and Criteria*. Preprint 87th Audio Engineering Society (AES) Convention, (Paper D-8, Preprint 2825), USA-New York, October.

Möller, S. et Raake, A. (2002). *Telephone Speech Quality Prediction : Towards Network Planning and Monitoring Models for Modern Network Scenarios*. Speech Communication, 38(1-2), 47-75.

ITU-T Rec. P.862 (2001). *Perceptual Evaluation of Speech Qualiy (PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrowbnad Telephone Networks and Speech Codecs.* International Telecommunication Union, CH-Geneva.

ITU-T Contribution COM-12 D.046 (2005). *Discussion on unified objective methodologies for the comparison of voice quality of narrowband and wideband scenarios.* France-Telecom R&D. International Telecommunication Union, CH-Geneva.

ETSI Technical Report ETR 250 (1996). *Transmission and Multiplexing (TM); Speech Communication Quality Mouth to Ear for 3,1 kHz Handset Telephony across Networks*. European Telecommunications Standards Institute, FR-Sophia Antipolis.

ITU-T Rec. G.113 (2002). *Provisional planning values for the Equipment Impairment Factor Ie and Packet-Loss Robutness Factor Bpl.* International Telecommunication Union, CH-Geneva.

ITU-T Rec. P.833 (2001). *Methodology for Derivation of Aquipement Impairment Factors from Subjective Listening Tests.* International Telecommunication Union, CH-Geneva.

ITU-T Contribution COM-12 D.049 (2005). *Methods for assessing requirement mixing different bandwidth*. France-Telecom R&D. International Telecommunication Union, CH-Geneva.

ITU-T Rec. P.830 (1996). *Subjective Performance Evaluation of Network Echo Cancellers.* International Telecommunication Union, CH-Geneva.

Borg, J.-C. (1998). *Borg G. Borg's perceived exertion and pain scales*. Human Kinetics, UK.

Krebber, W. (1995). *Spachübertragungsqualität von Fernsprech-Handapparaten,* Vol. 357. Series 10, VDI-Verlag GmbH, DE-Düsseldorf.

ITU-T Contribution COM-12 D.033 (2005). *Subjective assesment result for Widebnad speech coding*. Kitawaki, N. Univ. of Tsukuba, JP.

Pascal, D. (1988). *Comparative Performances of Two Subjective Methods for Improving the Fidelity of Speech Signals*. Preprint 84th Audio Engineering Society (AES) Convention, (Paper L-3, Preprint 2639), FR-Paris, March.

ITU-T Rec. P.810 (1996). *Modulated Noise Reference Unit (MNRU).* International Telecommunication Union, CH-Geneva.

ITU-T Rec. P.800 (1996). *Methods for Subjective Determination of Transmission Quality.* International Telecommunication Union, CH-Geneva.

Gibbon, D. (1992). *EUROM.1 German Speech Database*. ESPRIT project 2589 report (SAM, Multi-Lingual Speech Input/Output Assessment, Methodology and Standardisation), Universität Bielefeld, DE-Bielefeld.

ITU-T Contribution COM-12 D.028 (2005). *How Much Better can Wideband Telephony be? – Estimating the necessary R-scale extension*. Raake, A. DE-Bochum.

Krebber, J., Möller, S., Raake, A., Rehmann, S., Berger, J., et Johannsen, W. (2002). *Ein Simulationsystem zur Untersuchung des Einflusses von Übertragungskanälen bei Smart-Home-Anwendungen*. 13. Konferenz Elektronische Sprachsignalverarbeitung, DE-Dresden, 75-82.

ITU-T Rec. P.79 (1993). *Measurements related to speech loudness. Calculation of loudness ratings for telephone sets.* International Telecommunication Union, CH-Geneva.

Dimolitsas, S; Corcoran, F.L.; Ravishankar, C (1995). *Dependence of Opinion Scores on Listening Sets Used in Degradation Category Rating Assessments*. IEEE Speech and Audio Processing.

Raake, A. (2000). *Perceptual Dimensions of Speech Sound Quality in Modern Transmission Systems*. ICSLP, 2000.