



## **Sonie des signaux complexes : influence de la bande passante et des codecs utilisés dans les télécommunications - test subjectif**

I. Edjekouane<sup>a</sup>, C. Plapous<sup>a</sup>, C. Quinquis<sup>a</sup> et S. Meunier<sup>b</sup>

<sup>a</sup>Orange Labs, 2 Avenue Pierre Marzin, 22300 Lannion, France

<sup>b</sup>LMA-CNRS, UPR 7051, Aix-Marseille Univ, Centrale Marseille, 31 chemin Joseph-Aiguier, F-13402

Marseille Cedex 20, France

idir.edjekouane@orange.com

La téléphonométrie est l'étude de l'ensemble des techniques utilisées pour la mesure électroacoustique de la qualité vocale d'une communication téléphonique. La qualité vocale peut se caractériser par plusieurs paramètres : la réponse en fréquence, la sonie en réception, le niveau de bruit en réception et en émission, etc. La sonie est considérée comme un des paramètres principaux de la qualité vocale. En téléphonométrie, on utilise un indicateur appelé "loudness rating" pour étalonner la sonie. Le "loudness rating" a été développé dans les années 1950 pour la bande passante téléphonique traditionnelle (bande étroite, *i.e.* 300-3400 Hz). Les avancées récentes en télécommunication ont poussé à étendre la bande passante de la parole transmise. L'utilisation du "loudness rating" pour les bandes plus larges a montré une divergence entre le niveau sonore étalonné et le niveau sonore ressenti par l'utilisateur. Pour remédier à ce problème, nous proposons d'utiliser les modèles de sonie de l'état de l'art afin d'estimer directement la sonie ressentie par l'utilisateur. Afin d'évaluer ces modèles, il faut comparer les valeurs calculées par rapport aux résultats subjectifs. Nous proposons un test subjectif permettant de mesurer la sonie de signaux complexes (parole ou musique) sur l'échelle des phones. Nous pouvons ainsi étudier l'influence de la bande passante et du codage/décodage sur la sonie de la parole. Dans une première étape, la fonction de sonie individuelle de chaque sujet est mesurée en utilisant une échelle spécifique de 100 points. Puis, le sujet évalue la sonie des signaux complexes après traitement (filtrage ou codage/décodage), en utilisant la même échelle. Les estimations données en "points" sont ensuite converties en phones en utilisant la fonction de sonie individuelle. Les résultats montrent que la sonie augmente avec l'extension de la bande passante jusqu'à la bande super élargie (50 Hz – 14 kHz). Le même comportement est constaté lorsqu'on applique un codage/décodage.

## 1 Introduction et contexte de l'étude

Les systèmes téléphoniques ont été créés pour remplacer la conversation en face à face ; c'est pourquoi cette situation est prise comme référence dans la conception des services de télécommunication. Ainsi, l'information provenant de la bouche du locuteur et allant jusqu'à l'oreille de l'auditeur doit être conservée au mieux. La sonie, *i.e.* le niveau sonore perçu, contribue largement à la qualité globale de la parole transmise [1], c'est un facteur perceptif indispensable pour que la parole soit transmise de façon correcte *i.e.*, il faut que la sonie ne soit ni trop forte ni trop faible. Dans le domaine de la téléphonométrie [2], la chaîne de transmission peut engendrer des pertes de sonie, ces pertes sont alors exprimées par un indicateur appelé en anglais le "loudness rating" (LR). Cet indicateur est directement relié à la sonie (loudness en anglais). Le principe de LR est fondé sur les travaux de H. Fletcher et Munson sur la sonie et l'effet de masquage [3-4]. Une description détaillée du LR est donnée dans [5]. Le modèle de LR utilisé en téléphonométrie a été publié sous forme d'une recommandation de l'Union Internationale des Télécommunications ITU-T P.79 [6]. Il consiste à appliquer une pondération à la sensibilité électroacoustique de l'équipement par bande de tiers d'octave avant de sommer les valeurs de chaque bande. Initialement le LR a été défini pour la bande étroite (Narrowband, noté NB) [6, Annexe A], puis une extension a été proposée pour la bande élargie (Wideband, noté WB) en utilisant de nouveaux coefficients de pondération [6, Annexe G]. Cependant, des études expérimentales [7] ont montré une divergence dans l'ajustement du LR lorsqu'un terminal NB communique avec un terminal WB. En effet, l'utilisateur en WB perçoit le niveau sonore nettement plus faible (environ 6 dB) qu'un utilisateur en NB pour la même valeur de LR calculée. A ce jour, il n'y a aucune intention à l'ITU d'adapter le LR pour la bande super élargie (Super-wideband, noté SWB) et la bande large (Fullband, noté FB). Malgré tout, aujourd'hui, nous avons un véritable besoin d'un modèle objectif pour prédire le niveau sonore perçu dans la chaîne de transmission du NB jusqu'au FB. Ce modèle doit être cohérent pour les communications inter-bandes afin d'assurer un niveau sonore perçu constant.

Comme la sonie est la grandeur que l'on souhaite connaître du point de vue de l'utilisateur, nous pensons que les modèles de sonie (*cf.* [8-9]) sont des candidats naturels pour remplacer le LR. En effet, ces modèles fonctionnent pour toutes les bandes de fréquence du NB jusqu'au FB. Pour évaluer le comportement de ces modèles de sonie lorsque la parole est transmise en utilisant différentes bandes de fréquence (y compris lorsque des codecs sont appliqués), les prédictions de ces modèles doivent être comparées aux résultats d'un test subjectif. La première étape d'une telle approche est donc de mesurer subjectivement la sonie des signaux utilisés dans l'évaluation des systèmes téléphoniques, notamment le signal appelé P.501 (séquence de parole en anglais britannique décrite dans la clause 7.3.2 de la recommandation de l'ITU-T P.501 [10]). Ainsi, le but du travail présenté dans cet article est de concevoir un tel test subjectif. Jusqu'à présent, il n'existe pas de standard pour mesurer la sonie des signaux complexes. Cependant, quelques études [11-12] existent sur la mesure de la sonie du signal de parole et elles sont basées sur des méthodes dérivées essentiellement des méthodes d'ajustement de sonie ("loudness matching") ou de répartition catégorielle de sonie ("categorical loudness scaling", CLS) [13]. La limitation de la méthode d'ajustement de sonie est qu'elle est difficilement applicable aux signaux de longues durées telles que les signaux de test généralement utilisés en téléphonométrie. La limitation du CLS est que nous avons peu d'informations sur le lien entre les "categorical units" et les autres échelles de sonie déjà établies telles que les sones et les phones. De plus, cette méthode est conçue pour étudier toute la plage de dynamique de l'oreille ce qui est trop large pour notre besoin. Pour les raisons citées, dans cet article nous proposons une nouvelle méthode qui permet de mesurer la sonie des signaux complexes en utilisant une échelle spécifique. Le niveau de sonie du signal complexe est ainsi obtenu à partir de la connaissance de la sonie d'une bande de bruit étroite (de largeur inférieure à une bande critique).

## 2 Méthodologie

La procédure du test subjectif inclut deux étapes. La première étape consiste à estimer la fonction de sonie individuelle de chaque sujet. Dans la deuxième étape, les sujets évaluent la sonie des échantillons sélectionnés (signaux de test). Toutes les évaluations sont faites en utilisant une échelle spécifique composée de 100 points. Les stimuli sont présentés en monaural afin de simuler l'utilisation d'un terminal en mode combiné. Les résultats de sonie obtenus en termes de point sont ensuite convertis en phones grâce aux fonctions de sonie individuelles.

### 2.1 Signaux utilisés

Des échantillons audio avec différents contenus ont été sélectionnés (cf. tableau 1). Les échantillons peuvent être de la parole dans des contextes variés et avec des langues différentes (français et anglais), de la musique ou un mélange de parole et de musique. Le signal P.501 qui est largement utilisé en téléphonométrie a une importance majeure dans notre étude parce que c'est actuellement le signal utilisé dans la mesure du LR. A cause de la durée relativement longue du signal P.501 (36,5s, 12 phrases prononcées par 12 locuteurs différents, 6 femmes et 6 hommes, nous avons choisi de le découper en 4 parties; chaque partie contient ainsi 3 locuteurs ou 3 locutrices. Ces 4 parties forment les échantillons 5 à 8 que l'on retrouve dans le tableau 1.

Les échantillons ont été traités selon le diagramme présenté à la Figure 1. Ainsi, pour chaque bande de fréquence (FB, SWB, WB ou NB), les échantillons filtrés ont été codés/décodés en utilisant deux familles de codecs (cf. tableau 2). La première famille est constituée de codecs principalement utilisés pour le signal de parole (référéncée "codecs de parole"), tandis que la deuxième famille est constituée de codecs fonctionnant indépendamment du type de contenu (référéncée "codecs génériques"). Les signaux obtenus après "filtrage" ou "filtrage+codage/décodage" correspondent à ce qu'on appelle le niveau "Nominal" (gain à 0 dB sur la Figure 1). Ces signaux sont ensuite amplifiés de 5 dB ce qui correspond au niveau "Nominal +5 dB" ou atténués de 10 dB ce qui correspond au niveau "Nominal -10 dB". Ces deux dernières conditions ont été ajoutées afin d'obtenir une dynamique globale plus large pour les niveaux. Finalement, nous obtenons 36 conditions

appliquées aux 9 échantillons ce qui donne un total de  $[(8+4) \times 3] \times 9 = 324$  signaux de test.

TABLEAU 1: Description des échantillons

	Description du contenu	durée (secondes)	langage
Échantillon 1	Musique Rock	7,8	X
Échantillon 2	Musique suivie de parole mixée avec de la musique	12,4	Français
Échantillon 3	Parole (annonce vocale)	7,6	Français
Échantillon 4	Parole mixée avec de la musique	10,2	Français
Échantillon 5	Parole (P.501) Partie 1	8,3	Anglais britannique
Échantillon 6	Parole (P.501) Partie 2	9	Anglais britannique
Échantillon 7	Parole (P.501) Partie 3	9,2	Anglais britannique
Échantillon 8	Parole (P.501) Partie 4	10	Anglais britannique
Échantillon 9	Parole suivie de parole mixée avec de la musique	8,5	Français

TABLEAU 2: Description des codecs

Bande passante	Codecs (débit)	
	Codecs de parole	Codecs génériques
Codecs FB, échantillonnés à 48kHz	OPUS (64 kb/s) [14]	G.719 (64 kb/s) [15]
Codecs SWB, décimation à 32kHz	G.729.1 (32 kb/s) [16]	G.722.1 C (48 kb/s) [17]
Codecs WB, décimation à 16kHz	AMR-WB (12,65 kb/s) [18]	G.722 (64 kb/s) [19]
Codecs NB, décimation à 8kHz	AMR (12,2 kb/s) [20]	G.711 (64 kb/s) [21]

### 2.2 Description de l'échelle de réponse

Après avoir écouté chaque stimulus, le sujet évalue la sonie perçue sur une échelle de 100 points (Figure 2). Il a 5 secondes pour faire son évaluation, le passage au stimulus suivant est automatique afin de pousser le sujet à donner une réponse spontanée. La valeur numérique choisie par le sujet est affichée sur l'échelle (à côté du curseur). Les trois labels "Très fort", "Moyennement fort" et "Pas fort" sont utilisés uniquement afin de donner une référence au sujet. Le terme "fort" est utilisé dans les trois labels étant donné que la sonie des signaux du test est relativement élevée.

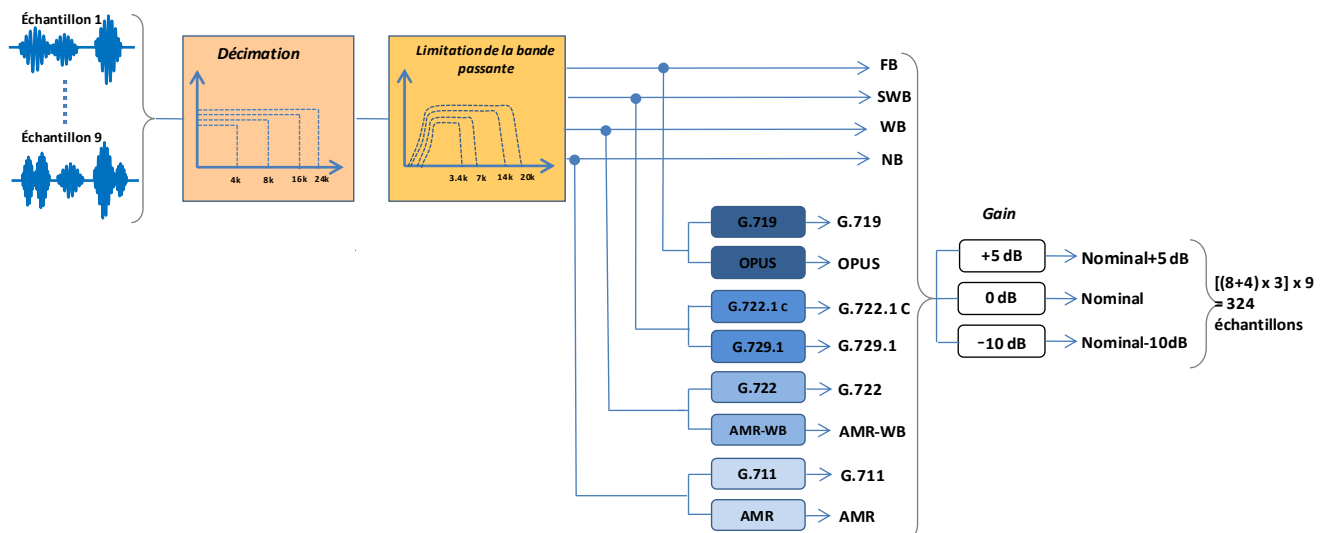


FIGURE 1: Schéma décrivant la préparation des signaux de test pour le test subjectif

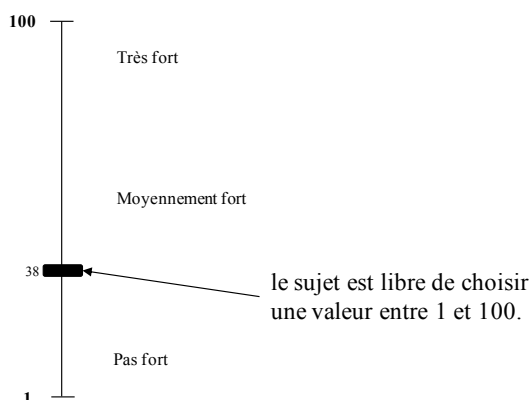


FIGURE 2: Illustration de l'échelle de réponse de 100 points utilisée dans le test.

## 2.3 Sujets et équipement

Dix-huit sujets avec une audition normale ont participé au test subjectif. Préalablement au test, il a été demandé au sujet d'indiquer son oreille préférée (gauche ou droite) lorsqu'il passe des appels téléphoniques. Les stimuli ont ensuite été présentés en monaural (gauche ou droite) au sujet via un casque (Stax SR-404) avec une courbe de réponse en fréquence très proche de la courbe de réponse en "diffuse field" décrite dans la clause 6.1.3 de la recommandation de l'ITU-T P.58 [22]. Les stimuli qui sont ré-échantillonnés à 48 kHz (si nécessaire) sont ensuite convertis en analogique (SPL 2489) puis amplifiés (Stax SRM-006t) avant d'être joués sur le casque. La chaîne de mesure a été calibrée de sorte à avoir un niveau acoustique de 77 dB SPL (Sound Pressure Level) au niveau du tympan pour un signal FB présenté au niveau "Nominal".

## 2.4 Première étape : mesure de la fonction de sonie individuelle

Les stimuli sont constitués d'une bande de bruit d'une largeur de 160 Hz, centrée sur la fréquence 1 kHz, et d'une durée de 1 seconde. Ces stimuli sont présentés dans un ordre pseudo-aléatoire avec des niveaux acoustiques différents. La plage des niveaux acoustiques des bandes de bruits doit bien entendu correspondre à la plage de sonie des signaux de test. Pour ce faire, un test subjectif préalable a été mis en œuvre afin de déterminer cette plage de niveaux acoustiques.

### 2.4.1. Détermination de la dynamique

Parmi les 324 signaux de test, les signaux avec la sonie la plus élevées sont à priori les signaux relatifs à la condition "FB et Nominal+5 dB". De même, ceux avec la sonie la plus faible sont à priori les signaux relatifs à la condition "NB et Nominal-10 dB". Le test préalable consiste alors à faire des tests de comparaison de sonie entre ces signaux sélectionnés et la bande de bruit émise avec une large gamme de pressions acoustiques, dans notre cas de 58 jusqu'à 91 dB SPL avec un pas de 3 dB. Les résultats de ce test montrent qu'en moyenne les signaux de test relatifs à la condition "FB et Nominal+5 dB" ont un niveau sonore équivalent à celui de la bande de bruit émise à une pression acoustique de 85 dB SPL, et les signaux du test relatifs à la condition "NB et Nominal-10 dB" ont un

niveau sonore équivalent à celui de la bande de bruit émise à une pression acoustique de 73 dB SPL. Pour être certains de contenir toute la plage dynamique de sonie, nous avons élargi l'intervalle jusqu'à [61 dB SPL; 88 dB SPL]. L'intervalle est couvert par un pas de 3 dB, ce qui nous donne un total de 10 stimuli. L'élargissement de l'intervalle n'est pas symétrique car nous avons constaté qu'au-delà de 88 dB SPL l'écoute de la bande de bruit était trop forte et devenait gênante. Ce test a été réalisé une seule fois et avant le test subjectif principal, par des collègues travaillant à Orange Labs.

### 2.4.2. Procédure pour la mesure de la fonction de sonie individuelle

La mesure de la fonction de sonie individuelle est composée de deux phases (*cf.* Figure 3) pendant lesquelles le sujet évalue la sonie en utilisant l'échelle décrite dans la section 2.2. Dans la première phase (apprentissage), le sujet entend une sélection de stimuli couvrant toute la gamme des niveaux acoustiques. Cette étape permet d'éviter le biais causé par les premiers essais tirés au hasard, qui peuvent ne pas couvrir toute la dynamique des niveaux acoustiques des stimuli. Dans la deuxième phase, les 10 stimuli sont présentés 6 fois chacun suivant 6 ordres pseudo-aléatoires. Pour éviter que le jugement du sujet sur le stimulus actuel soit influencé par le stimulus précédent, nous avons pris soin de conserver la différence de niveau entre deux stimuli successifs inférieure à la moitié de la dynamique totale des signaux.

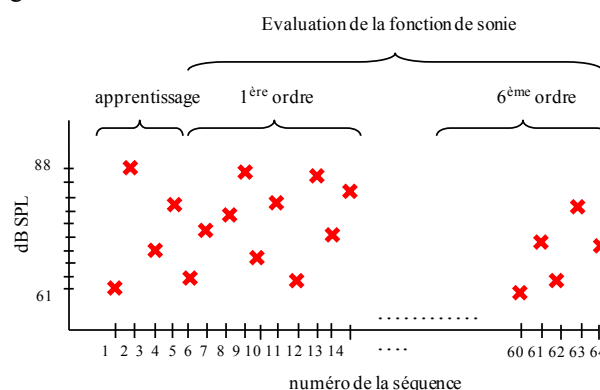


FIGURE 3: Séquences de détermination de la fonction de sonie individuelle. L'apprentissage est suivi de 6 ordres aléatoires.

### 2.4.3. Résultats obtenus pour la fonction de sonie individuelle

La Figure 4 représente les fonctions de sonie individuelles des 18 sujets en termes de points. En général, les courbes sont en forme de "S" du fait des deux parties saturées : la partie supérieure [85 dB SPL; 88 dB SPL] et la partie inférieure [61 dB SPL; 70 dB SPL]. Ces deux parties sont dues à la saturation de l'échelle de sonie. En effet, le sujet juge toujours le stimulus comme "Trop fort" à partir du moment où son niveau acoustique dépasse les 85 dB SPL. De la même manière, le sujet juge toujours le stimulus comme "Pas fort" à partir du moment où son niveau acoustique descend sous les 70 dB SPL. Dans la partie centrale [70 dB SPL; 85 dB SPL], les sujets utilisent efficacement l'échelle de sonie. La conversion de la sonie

en points vers la sonie en phones sera donc effectuée en utilisant cette partie centrale (cf. section 2.6).

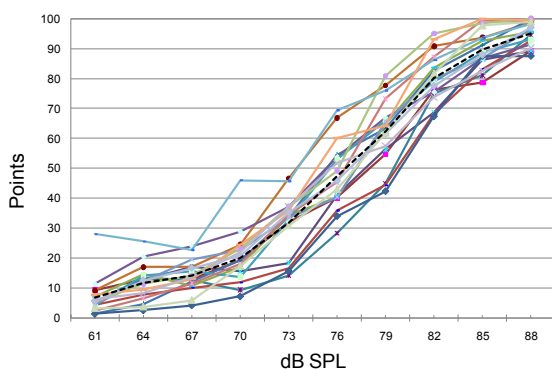


FIGURE 4: Fonctions de sonie individuelle (exprimées en termes des points) obtenues pour les 18 sujets. La moyenne des courbes est représentée en pointillé.

## 2.5 Deuxième étape : Evaluation de la sonie des signaux de test

L'évaluation de la sonie des signaux de test se fait en deux phases et toujours en utilisant l'échelle de sonie illustrée par la Figure 2. La première phase est une phase d'apprentissage. Elle consiste à présenter au sujet une sélection de signaux de test qui couvre la dynamique totale de sonie. Cette sélection contient les conditions relatives à la sonie a priori la plus forte et les conditions relatives à la sonie a priori la plus faible. Tous les 9 échantillons sont utilisés dans cette première phase afin que le sujet soit familiarisé avec ces sons avant de commencer la deuxième phase. Dans la deuxième phase le sujet évalue la sonie des 324 signaux de test présentés aléatoirement, en utilisant l'échelle décrite dans la section 2.2.

## 2.6 Conversion des points en phones

La fonction de sonie individuelle estimée donne la relation entre le niveau acoustique en dB SPL et les points pour chaque sujet (Figure 4). Il est important de rappeler que cette relation est valide pour une présentation monaurale des signaux. Elle montre une tendance linéaire dans sa partie centrale [70 dB SPL; 85 dB SPL]. Un modèle linéaire d'ajustement par la méthode des moindres carrés a donc été adopté pour les données expérimentales de chaque sujet :

$$L_{dB\ SPL} = \alpha_i \times N_{po\ int.s} + \beta_i \quad (1)$$

Tel que  $\alpha_i$  et  $\beta_i$  sont les paramètres d'ajustement déterminés pour chaque sujet  $i$  ( $i = 1, 2, \dots, 18$ ).  $L_{dB\ SPL}$  est le niveau de pression acoustique sur une seule oreille exprimée en dB SPL.  $N_{po\ int.s}$  est le niveau de sonie exprimé en points.

D'après la définition du phone [23, 24], l'échelle des phones est égale à celle des dB SPL pour une bande de bruit centrée sur 1 kHz et de largeur d'une bande critique, présentée en binaural. Cela correspond à la description des stimuli utilisés dans notre test à l'exception que les stimuli étaient présentés en monaural. Il existe un rapport de  $A$  entre la sonie en binaural et la sonie en monaural ce qui correspond alors à un écart de  $\Delta N = 10 \log_2(A)$  phones.

Si on suppose que  $N_{phones}$  est la sonie perçue par le sujet exprimée en phones, alors  $N_{phones}$  peut s'écrire en fonction de l'écart  $\Delta N$  :

$$N_{phones} = L_{dB\ SPL} - \Delta N \quad (2)$$

Ainsi,

$$N_{phones} = \alpha_i \times N_{po\ int.s} + \beta_i - \Delta N \quad (3)$$

Dans les premières études sur la sonie menées par Fletcher et Munson [3-4] le rapport  $A$  a été estimé à une valeur de 2. D'autres études *i.e.* [25-27] ont confirmé cette valeur. Cependant, des études plus récentes [28] montrent que la valeur de  $A$  est inférieure à 2 et qu'elle se rapproche de la valeur de 1,5. D'autre part, l'étude [29] confirme que la valeur de  $A$  généralement déterminée pour des signaux purs est en réalité inférieure à 2 pour les signaux complexes comme la parole. Dans le futur, nous avons prévu de déterminer la valeur de  $A$  grâce à un nouveau test subjectif. Dans la suite de cet article, afin d'obtenir des résultats chiffrés, nous avons supposé que  $A = 1,5$  ce qui correspond à  $\Delta N = 5,85$  phones.

Bien entendu, pour chaque sujet la conversion des points vers les phones est basée sur sa propre fonction de sonie; cela est dû au fait que chaque sujet utilise l'échelle de sonie à sa propre manière. En effet, le sujet crée son propre système interne de référence qui peut varier largement d'un sujet à un autre. Cependant, tant que le sujet conserve le même système interne de référence tout au long du test subjectif, les points peuvent être convertis en phones en utilisant la fonction de sonie individuelle estimée ainsi que l'équation (3).

## 3 Résultats et discussion

La Figure 5 rassemble les résultats des niveaux de sonie moyennés sur les 9 échantillons sélectionnés. Toutes les conditions sont représentées sur cette figure, *i.e.* "bande passante", "codecs de parole" et "codecs génériques" ainsi que les trois niveaux, *i.e.* "Nominal+5dB", "Nominal" et "Nominal-10dB". Ces résultats sont présentés en phones avec un intervalle de confiance à 95%. Les résultats corroborent la théorie selon laquelle la sonie augmente avec l'extension de la largeur de la bande fréquentielle du signal. Après avoir vérifié que les données expérimentales pour chaque condition sont normalement distribuées (test de normalité de Kolmogorov-Smirnov), une analyse de la variance (ANOVA) à un facteur (les facteurs sont les conditions, niveau de signification à 5%) a été conduite séparément pour chaque niveau : "Nominal+5dB", "Nominal" et "Nominal-10dB", dans le but de déterminer si nous avons une différence significative de sonie dans chaque condition *i.e.* "bande passante", "codecs de parole" et "codecs génériques". Ainsi, pour chaque niveau, les résultats d'analyse de variance montrent une différence significative de sonie. On appelle un élément les composantes de chaque condition (*e.g.* les éléments de la condition "bande passante" sont : NB, WB, SWB et FB). Des tests Post-hoc ANOVA montrent que la différence de sonie entre les éléments relatifs au NB (*i.e.* NB, G.711, AMR), et les éléments relatifs au WB (*i.e.* WB, G.722, AMR-WB) est significative, ainsi que la différence de sonie entre les éléments relatifs au WB et les éléments relatifs au

SWB (*i.e.* SWB, G.722.1 C, G.729.1). Cependant, la différence de sonie entre les éléments relatifs au SWB et les éléments relatifs au FB (*i.e.* FB, G.719, OPUS) ne l'est pas. On constate les mêmes résultats pour les "codecs de parole" et les "codecs génériques" bien que la méthode de codage soit différente.

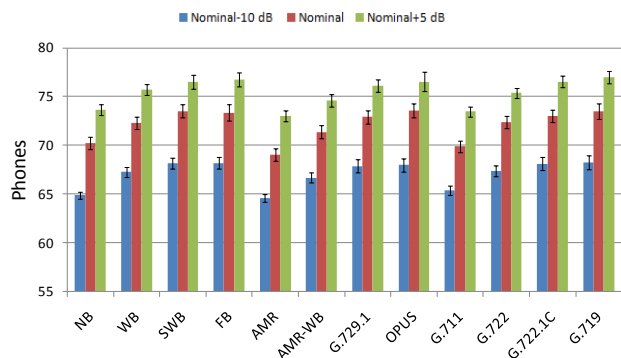


FIGURE 5: Illustration des résultats moyens sur les échantillons. Toutes les conditions sont présentées: "bande passante", "codecs de parole", "codecs génériques", "Nominal+5dB", "Nominal" et "Nominal-10dB".

## 4 Conclusion et travaux futurs

Nous avons proposé un test subjectif permettant d'étudier le comportement de la sonie vis-à-vis de l'élargissement de la bande fréquentielle et des codecs disponibles sur le marché. Les résultats montrent que la sonie augmente avec l'élargissement de la bande fréquentielle et ce jusqu'à la bande super élargie (SWB), y compris lorsque les codecs sont appliqués. La prochaine étape sera d'appliquer des algorithmes objectifs de sonie sur les mêmes signaux de test, afin de confronter le comportement des modèles de sonie déjà existants aux résultats de notre test subjectif.

## Références

- [1] N. Côté, V. Gautier-Turbin, and S. Möller, Influence of Loudness Level on the Overall Quality of Transmitted Speech, 123rd Audio Eng. Soc. Conv., New York (2007).
- [2] ITU-T, Handbook on telephonometry, ITU-T, Genève (1992).
- [3] H. Fletcher and W.A. Munson, Relation between Loudness and Masking, J. Acoust. Soc. Am., 9(1), 1-10, New York (1933).
- [4] H. Fletcher and W.A. Munson, Loudness, Its Definition, Measurement and Calculation, J. Acoust. Soc. Am., 5(2), 82-108, New York (1937).
- [5] S. Möller, Assessment and Prediction of Speech Quality in Telecommunications, 18-24, Springer (2000).
- [6] ITU-T Recommendation P.79: Calculation of loudness ratings for telephone sets, ITU-T, Genève (2000).
- [7] K. Allen Woo, Wide-band loudness ratings confusion (ref ITU-T P.79), ETSI STQ(12), (2007).
- [8] ISO 532 B, Method for calculating loudness, International standard (1975).
- [9] ANSI S3, 4-2007, American National Standard Procedure for the Computation of Loudness of Steady Sound, (2007).
- [10] ITU-T Recommendation P.501: Test signals for use in telephonometry, (2012).
- [11] H. Fastl, Loudness of running speech, J. Audiol. Technique 16, 2-13, (1977).
- [12] J. Rannies, J. L. Verhey, J. E. Appell, and B. Kollmeier, Loudness of complex time-varying sounds. A challenge for current loudness models, J. Acoust. Soc. Am., 19(1), 050189, (2013).
- [13] ISO 16832. Acoustics – Loudness scaling by means of categories, (2006).
- [14] JM. Valin, K. Vos, and T. Terriberry, Internet Engineering Task Force (IETF) RFC6716: Definition of the Opus Audio Codec, (2012).
- [15] ITU-T Recommendation G.719: Low-complexity, full-band audio coding for high-quality, conversational applications, (2008).
- [16] ITU-T Recommendation G.729.1: G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729, (2006).
- [17] ITU-T Recommendation G.722.1 Annex C: Low Complexity Coding at 24 and 32 kb/s for Hands-Free Operation in Systems with Low Frame Loss Annex C 14 kHz Mode at 24, 32, and 48 kb/s, (2005).
- [18] 3GPP TS 26 290 V10.0.0, Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions, (2011).
- [19] ITU-T Recommendation G.722: 7 kHz audio-coding within 64 kbit/s, Genève (2012).
- [20] 3GPP TS 126 071 V3.0.1, Mandatory speech CODEC speech processing functions; AMR speech Codec; General description, (2000).
- [21] ITU-T Recommendation G.711: Pulse code modulation (PCM) of voice frequencies, Genève (1988).
- [22] ITU-T Recommendation P.58: Head and torso simulator for telephonometry, ITU-T, Genève (2013).
- [23] E. Zwicker and H. Fastl, Psychoacoustics: Facts and models, 2nd Edition, Springer-Verlag, pp. 203, (1999).
- [24] ISO 226. Normal equal-loudness-level contours, (2003).
- [25] Hellman, R.P., and Zwislocki, J.J., Monaural loudness function at 1000 cps and interaural summation, J. Acoust. Soc. Am., 35(6), 856-865, (1963).

[26] Marks, L.E., Binaural summation of the loudness of pure tones, *J. Acoust. Soc. Am.*, 64(1), 107-113, (1978).

[27] Stevens, Stanley S., On the psychophysical law. *Psychological review*, 64(3), 153, (1957).

[28] Moore, Brian CJ et Glasberg, Brian R, Modeling binaural loudness. *J. Acoust. Soc. Am.*, 121(3), 1604-1612, (2007).

[29] Pollack, Irwin, On the measurement of the loudness of speech. *J. Acoust. Soc. Am.*, 24(3), 323-324, (1952).