

التوصية ITU-R BS.1387-2

(2023/05)

السلسلة BS: الخدمة الإذاعية (الصوتية)

طريقة القياسات الموضوعية لجودة الصوت المدركة

تمهيد

يضطلع قطاع الاتصالات الراديوية بدور يتمثل في تأمين الترشيد والإنصاف والفعالية والاقتصاد في استعمال طيف الترددات الراديوية في جميع خدمات الاتصالات الراديوية، بما فيها الخدمات الساتلية، وإجراء دراسات دون تحديد المدى الترددات، تكون أساساً لإعداد التوصيات واعتمادها. ويؤدي قطاع الاتصالات الراديوية وظائفه التنظيمية والسياساتية من خلال المؤتمرات العالمية والإقليمية للاتصالات الراديوية وجمعيات الاتصالات الراديوية بمساعدة لجان الدراسات.

سياسة قطاع الاتصالات الراديوية بشأن حقوق الملكية الفكرية (IPR)

يرد وصف للسياسة التي يتبعها قطاع الاتصالات الراديوية فيما يتعلق بحقوق الملكية الفكرية في سياسة البراءات المشتركة بين قطاع تقييس الاتصالات وقطاع الاتصالات الراديوية والمنظمة الدولية للتوحيد القياسي واللجنة الكهروتقنية الدولية (ITU-T/ITU-R/ISO/IEC) والمشار إليها في القرار ITU-R 1. وترد الاستثمارات التي ينبغي لحاملي البراءات استعمالها لتقديم بيان عن البراءات أو للتصريح عن منح رخص في الموقع الإلكتروني <http://www.itu.int/ITU-R/go/patents/en> حيث يمكن أيضاً الاطلاع على المبادئ التوجيهية الخاصة بتطبيق سياسة البراءات المشتركة وعلى قاعدة بيانات قطاع الاتصالات الراديوية التي تتضمن معلومات عن البراءات.

سلسلة توصيات قطاع الاتصالات الراديوية

(يمكن الاطلاع عليها أيضاً في الموقع الإلكتروني <https://www.itu.int/publ/R-REC/en>)

العنوان

السلسلة

البث الساتلي	BO
التسجيل من أجل الإنتاج والأرشفة والعرض؛ الأفلام التلفزيونية	BR
الخدمة الإذاعية (الصوتية)	BS
الخدمة الإذاعية (التلفزيونية)	BT
الخدمة الثابتة	F
الخدمة المتنقلة وخدمة التحديد الراديوي للموقع وخدمة الهواة والخدمات الساتلية ذات الصلة	M
انتشار الموجات الراديوية	P
علم الفلك الراديوي	RA
أنظمة الاستشعار عن بُعد	RS
الخدمة الثابتة الساتلية	S
التطبيقات الفضائية والأرصاد الجوية	SA
تقاسم الترددات والتنسيق بين أنظمة الخدمة الثابتة الساتلية والخدمة الثابتة	SF
إدارة الطيف	SM
التجميع الساتلي للأخبار	SNG
إرسالات الترددات المعيارية وإشارات التوقيت	TF
المفردات والمواضيع ذات الصلة	V

ملاحظة: تمت الموافقة على النسخة الإنكليزية لهذه التوصية الصادرة عن قطاع الاتصالات الراديوية بموجب الإجراء الموضح في القرار ITU-R 1.

النشر الإلكتروني

جنيف، 2024

© ITU 2024

جميع حقوق النشر محفوظة. لا يمكن استنساخ أي جزء من هذه المنشورة بأي شكل كان ولا بأي وسيلة إلا بإذن خطي من الاتحاد الدولي للاتصالات (ITU).

التوصية 2-1387-ITU-R BS

طريقة القياسات الموضوعية لجودة الصوت المدركة

(1998-2001-2023)

مجال التطبيق

توصف هذه التوصية طريقة للقياس الموضوعي لجودة الصوت المدركة.

مصطلحات أساسية

جودة الصوت المدركة، القياس الموضوعي، تشفير بمعدل بتات منخفض

إن جمعية الاتصالات الراديوية للاتحاد الدولي للاتصالات،

إذ تضع في اعتبارها

- (أ) أن الطرائق الموضوعية التقليدية (مثل قياس النسبة إشارة إلى ضوضاء والتشوه) لم تعد كافية لقياس جودة الصوت المدركة للأنظمة التي تستخدم مخططات تشفير بمعدل بتات منخفض أو التي تستخدم معالجة الإشارات التماثلية أو الرقمية؛
- (ب) أن مخططات التشفير بمعدل بتات منخفض آخذة في الانتشار بسرعة؛
- (ج) أن عمليات التنفيذ التي تمثل مواصفة ما أو لمعيار ما لا تضمن جميعها أفضل جودة يمكن تحقيقها في إطار تلك المواصفة أو ذلك المعيار؛
- (د) أن طرائق التقييم الشخصي الرسمية لا تصلح للمراقبة المستمرة لجودة الصوت، في ظل الظروف التشغيلية مثلاً؛
- (هـ) أن طرائق القياس الموضوعي لجودة الصوت المدركة قد تكمل في نهاية الأمر أو تحل محل طرائق الاختبار الموضوعي التقليدية في جميع مجالات القياسات؛
- (و) أن القياس الموضوعي لجودة الصوت المدركة يمكن أن يكمل بصورة مفيدة طرائق التقييم الشخصي؛
- (ز) أنه بالنسبة لبعض التطبيقات، من الضروري وجود طريقة يمكن تنفيذها في الوقت الفعلي،

توصي

- 1 بأن تستخدم الطريقة الواردة في الملحق 2 للقياس الموضوعي لجودة الصوت المدركة، لكل تطبيق مدرج في الملحق 1.

تمهيد

توصف هذه التوصية طريقة للقياس الموضوعي لجودة الصوت المدركة لجهاز قيد الاختبار، على سبيل المثال، كوديك ذو معدل بت منخفض. وهي مقسمة إلى ملحقين. يقدم الملحق 1 للمستعمل نظرة عامة على الطريقة ويتضمن أربعة مرفقات. حيث يصف المرفق 1 التطبيقات وإشارات الاختبار. ويسرد المرفق 2 متغيرات مخرجات النموذج ويناقش قيود الاستخدام والدقة. ويقدم المرفق 3 الخطوط العريضة للنموذج بينما يصف المرفق 4 مبادئ وخصائص طرائق قياس جودة الصوت الإدراكية الموضوعية بشكل عام. يوفر الملحق 2 للمنفذ وصفاً تفصيلياً للطريقة باستخدام نسختين من النموذج النفسي الصوتي الذي تم تطويره خلال مرحلة التكامل حيث تم دمج ستة نماذج. ويرد في المرفق 1 بالملحق 2 وصف لعملية التحقق من صحة طريقة القياس الموضوعي. ويقدم المرفق 2 بالملحق 2 نظرة عامة على جميع قواعد البيانات التي تم استخدامها في تطوير الطريقة والتحقق من صحتها.

جدول المحتويات

الصفحة

ii	سياسة قطاع الاتصالات الراديوية بشأن حقوق الملكية الفكرية (IPR)
7	الملحق 1 - نظرة عامة
7	1 مقدمة
7	2 التطبيقات
8	3 الصيغ
8	4 الميدان الشخصي
10	5 الاستبانة والدقة
10	6 المتطلبات والقيود
10	المرفق 1 بالملحق 1 - التطبيقات
10	1 عام
11	2 التطبيقات الرئيسية
11	1.2 تقييم عمليات التنفيذ
11	2.2 ترتيب الجودة الإدراكية
11	3.2 المراقبة عبر الإنترنت
11	4.2 حالة المعدات أو التوصيلة
12	5.2 تحديد هوية الكوديك
12	6.2 تطوير الكوديكات
12	7.2 تخطيط الشبكة
12	8.2 المساعدة في التقييم الشخصي
13	9.2 ملخص التطبيقات
13	3 إشارات الاختبار
13	1.3 اختيار إشارات الاختبار الطبيعية
15	2.3 مدة الإشارة
15	4 التزامن
15	5 مسائل حقوق التأليف والنشر
15	المرفق 2 بالملحق 1 - متغيرات الخرج
15	1 مقدمة
16	2 متغيرات خرج النموذج
16	3 جودة الصوت الأساسية

17 هامش التشفير.	4
17 متطلبات المستعمل.	5
17 المرفق 3 بالملحق 1 - الخطوط العريضة للنموذج.	
18 معالجة الصوت	1
18 1.1 الإعدادات المحددة من قبل المستعمل	
18 2.1 النموذج الصوتي النفسي	
19 3.1 النموذج الإدراكي	
19 المرفق 4 بالملحق 1 - مبادئ وخصائص طرائق قياس جودة الصوت الإدراكية الموضوعية.	
19 1 مقدمة والتسلسل التاريخي	
20 2 البنية العامة لطرائق قياس جودة الصوت الإدراكية الموضوعية	
21 3 الأسس النفسية والصوتية والإدراكية.	
21 1.3 خاصية النقل للأذن الخارجية والوسطى.	
21 2.3 سلام التردد الإدراكية.	
22 3.3 الإثارة.	
23 4.3 الكشف	
23 5.3 التقنيع	
24 6.3 الجهاراة والتقنيع الجزئي	
24 7.3 الحدة	
24 8.3 المعالجة الإدراكية	
25 4 النماذج المدججة	
25 1.4 مؤشر التشويش (DIX)	
26 2.4 نسبة الضوضاء إلى الإشارة المقنعة (NMR)	
26 3.4 التقييم الموضوعي للإشارة الصوتية	
27 4.4 قياس جودة الصوت الإدراكية (PAQM)	
27 5.4 التقييم الإدراكي	
28 6.4 القياس الموضوعي الإدراكي	
28 7.4 نهج مجموعة الأدوات	
30 الملحق 2 - وصف النموذج	
30 1 الخطوط العريضة	
31 1.1 الصيغة الأساسية	
31 2.1 الصيغة المتقدمة	

31 نموذج الأذن المحيطية.....	2
31 نموذج الأذن القائم على التحويل FFT.....	1.2
45 نموذج الأذن القائم على بنك المرشحات.....	2.2
53 المعالجة المسبقة لأنماط الإثارة.....	3
53 تكيف المستوى والنمط.....	1.3
55 التشكيل.....	2.3
55 الجهارة.....	3.3
56 حساب إشارة الخطأ.....	4.3
56 حساب متغيرات خرج النموذج.....	4
56 نظرة عامة.....	1.4
57 فرق التشكيل.....	2.4
58 جهارة الضوضاء.....	3.4
60 عرض النطاق.....	4.4
61 نسبة الضوضاء إلى القناع.....	5.4
61 Relative Disturbed FramesB.....	6.4
61 احتمال الكشف.....	7.4
63 البنية التوافقية للخطأ.....	8.4
64 التوسيط.....	5
64 التوسيط الطيفي.....	1.5
64 التوسيط الزمني.....	2.5
66 التوسيط عبر القنوات الصوتية.....	3.5
66 تقدير جودة الصوت الأساسية المدركة.....	6
66 الشبكة العصبية الاصطناعية.....	1.6
67 الصيغة الأساسية.....	2.6
69 الصيغة المتقدمة.....	3.6
70 مطابقة عمليات التنفيذ.....	7
70 عام.....	1.7
70 الاختيار.....	2.7
70 إعدادات اختبار المطابقة.....	3.7
71 فاصل التسامح المقبول.....	4.7
71 عناصر الاختبار.....	5.7
72 المرفق 1 بالملحق 2 - عملية التحقق.....	
72 عام.....	1

73 المرحلة التنافسية.	2
74 المرحلة التعاونية	3
74 التحقق	4
75 المقارنة بين الدرجات ODG وSDG	1.4
75 الترابط	2.4
77 درجة الخطأ المطلق (AES)	3.4
79 مقارنة الدرجات ODG مقابل فاصل الثقة.	4.4
83 مقارنة الدرجات ODG مقابل فاصل التسامح	5.4
84 اختيار صيغ النموذج المثلى	5
84 معايير الاختيار المسبق استناداً إلى الترابط	1.5
85 تحليل عدد القيم الشاذة.	2.5
86 تحليل شدة القيم الشاذة.	3.5
87 الخلاصة	6
87 المرفق 2 بالملحق 2 - أوصاف قواعد البيانات المرجعية.	
87 مقدمة	1
89 العناصر لكل قاعدة بيانات	2
89 الحالات التجريبية	3
90 MPEG90	1.3
90 MPEG91	2.3
90 ITU92DI	3.3
90 ITU92CO	4.3
90 ITU93	5.3
91 MPEG95	6.3
91 EIA95	7.3
91 DB2	8.3
91 DB3	9.3
92 CRC97	10.3
92 العناصر لكل حالة بالنسبة لقاعدتي البيانات DB2 وDB3.	4
92 DB2	1.4
95 قاعدة البيانات DB3	2.4
98 المراجع	
99 بيليوغرافيا	

الملحق 1

نظرة عامة

1 مقدمة

تعد جودة الصوت أحد العوامل الأساسية عند تصميم نظام رقمي للإذاعة. وقد أدى الإدخال السريع لمختلف مخططات خفض معدل البتات إلى بذل جهود كبيرة في إنشاء وتحسين إجراءات التقييمات الشخصية، وذلك ببساطة لأن اختبارات الاستماع الرسمية كانت الطريقة الوحيدة ذات الصلة للحكم على جودة الصوت. وكانت الخبرة المكتسبة هي الأساس الذي قامت عليه التوصية ITU-R BS.1116، التي أصبحت فيما بعد الأساس لمعظم اختبارات الاستماع من هذا النوع.

وبما أن التقييمات الشخصية للجودة تستغرق وقتاً طويلاً وباهظة الثمن، فمن المستحسن تطوير طريقة قياس موضوعي من أجل التوصل إلى تقدير لجودة الصوت. ولم يثبت أبداً أن طرق القياس الموضوعي التقليدية، مثل النسبة إشارة إلى ضوضاء (S/N) أو التشوه الكلي الناجم عن التوافقيات (THD) ترتبط بشكل موثوق بجودة الصوت المدركة. وقد أصبحت المشكلات أكثر وضوحاً عند تطبيق الطرائق على الكوديكات الحديثة غير الخطية وغير الثابتة، على السواء.

لقد تم تطوير عدد من الطرائق لإجراء قياسات إدراكية موضوعية لجودة الصوت المدركة خلال العقد الماضي. ولكن لم يتم التحقق من صحة أي من هذه الطرائق بشكل كامل، وبالتالي لم يتم تقييسها أو قبولها على نطاق واسع. وفي عام 1994، حدد قطاع الاتصالات الراديوية الحاجة الملحة إلى وضع معيار في هذا المجال وبدأ العمل في ذلك. وتم إصدار دعوة مفتوحة لتقديم المقترحات وتم تلقي الطرائق الست التالية كطرائق مرشحة للقياس: مؤشر الاضطراب (DIX)، ونسبة الضوضاء إلى القناع (NMR)، وقياس جودة الصوت الإدراكي (PAQM)، والتقييم الإدراكي (PERCEVAL)، والقياس الموضوعي الإدراكي (POM) ونهج الأدوات. ويرد وصف هذه الطرائق في المرفق 4 بالملحق 1.

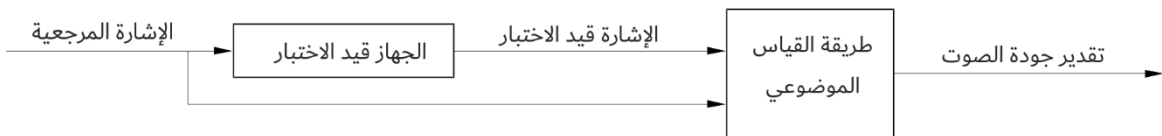
وطريقة القياس الواردة في هذه التوصية هي نتيجة عملية تمت فيها دراسة أداء كل من الطرائق الست المذكورة أعلاه، وتم استخراج أفضل الأدوات الواعدة ودمجها في طريقة واحدة. ولقد تم التحقق من صحة الطريقة الموصى بها بعناية في عدد من مواقع الاختبار. وقد أثبتت أنها تنتج معلومات موثوقة ومفيدة للعديد من التطبيقات. ومع ذلك، لا بد من الأخذ في الاعتبار أن طريقة القياس الموضوعي الواردة في هذه التوصية ليست عموماً بديلاً عن إجراء اختبار استماع رسمي.

2 التطبيقات

يُوضح في الشكل 1 أدناه المفهوم الأساسي لإجراء قياسات موضوعية باستخدام الطريقة الموصى بها.

الشكل 1

المفهوم الأساسي لإجراء قياسات موضوعية



BS.1387-01

تنطبق طريقة القياس الواردة في هذه التوصية على معظم أنواع تجهيزات معالجة الإشارات الصوتية، الرقمية منها والتماثلية. ومع ذلك، من المتوقع أن تركز العديد من التطبيقات على الكوديكات الصوتية.

وقد تم تحديد الفئات الثمانية التالية من التطبيقات:

الجدول 1

التطبيقات

الصيغة	وصف مختصر	التطبيق	
أساسية/متقدمة	إجراء لتحديد خصائص عمليات التنفيذ المختلفة لمعدات معالجة الصوت، وفي كثير من الحالات الكوديكات الصوتية	تقييم عمليات التنفيذ	1
أساسية	إجراء سريع يتم إجراؤه قبل إدخال أي جزء من المعدات أو دارة في الخدمة	ترتيب الجودة الإدراكية	2
أساسية	عملية مستمرة لمراقبة الإرسال الصوتي في الخدمة	المراقبة عبر الإنترنت	3
متقدمة	تحليل مفصل لجزء من المعدات أو دارة	حالة المعدات أو التوصيل	4
متقدمة	إجراء لتحديد نوع كوديك معين وعملية تنفيذه	تحديد الكوديك	5
أساسية/متقدمة	إجراء يحدد خصائص أداء الكوديك بأكثر قدر ممكن من التفصيل	تطوير الكوديك	6
أساسية/متقدمة	إجراء لاستمثال تكلفة وأداء شبكة النقل في ظل قيود معينة	تخطيط الشبكة	7
أساسية/متقدمة	أداة لفرز المواد الهامة لإدراجها في اختبار الاستماع	المساعدة في التقييم الذاتي	8

3 الصيغ

من أجل تحقيق التوافق الأمثل مع مختلف متطلبات التكلفة والأداء، فإن طريقة القياس الموضوعي الموصى بها في هذه التوصية لها صيغتان. وقد تم تصميم الصيغة الأساسية للسماح بتنفيذ فعال من حيث التكلفة في الوقت الفعلي، في حين تركز الصيغة المتقدمة على تحقيق أعلى دقة ممكنة. واعتماداً على التنفيذ، تزيد هذه الدقة الإضافية من التعقيد بمقدار أربعة أضعاف تقريباً مقارنة بالصيغة الأساسية.

ويقدم الجدول 1 بعض الإرشادات حول الصيغة التي يجب الذي تطبيقها لكل تطبيق من التطبيقات.

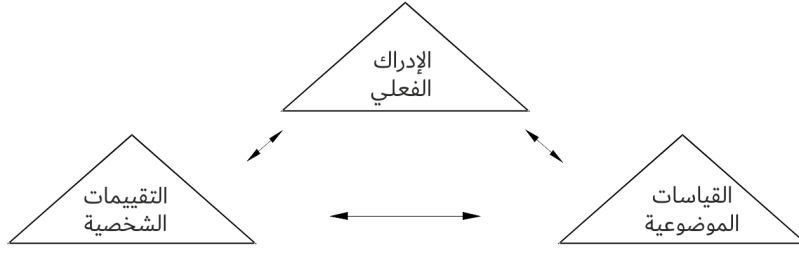
4 الميدان الشخصي

اختبارات الاستماع الشخصية الرسمية، على سبيل المثال، تلك المستندة إلى التوصية ITU-R BS.1116، مصممة بعناية لتقرب قدر الإمكان من تقدير موثوق للحكم على جودة الصوت. ومع ذلك، لا يمكن للمرء أن يتوقع أن تعكس نتيجة اختبار الاستماع الشخصي الإدراك الفعلي بشكل كامل. ويوضح الشكل 2 العيوب الضمنية في كل من الميدان الشخصي والموضوعي.

ومن الواضح أنه لن يتسنى التحقق من صحة طريقة موضوعية مباشرة. وبدلاً من ذلك، يتم التحقق من صحة طرائق القياس الموضوعي مقابل اختبارات الاستماع الشخصية.

الشكل 2

مفاهيم التحقق من الصلاحية



BS.1387-02

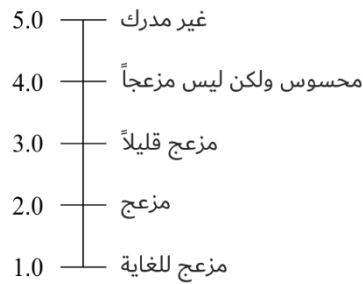
تركز طريقة القياس الموضوعي الواردة في هذه التوصية على التطبيقات التي يتم تقييمها عادة في الميدان الشخصي من خلال تطبيق التوصية ITU R BS.1116. ويمكن وصف المبدأ الأساسي لطريقة الاختبار المحددة هذه بإيجاز على النحو التالي: يمكن للمستمع الاختيار بين ثلاثة مصادر ("A" و "B" و "C"). والإشارة المرجعية المعروفة متاحة دائماً كالمصدر "A". وتتوفر الإشارة المرجعية المخفية والإشارة قيد الاختبار في وقت واحد ولكن يتم تخصيصهما "عشوائياً" للمصدرين "B" و "C"، اعتماداً على التجربة.

ويطلب من المستمع تقييم مظاهر الانحطاط في المصدر "B" مقارنة بالانحطاط في المصدر "A" وفي المصدر "C" مقارنة بالانحطاط في "A" وفقاً لسلم الانحطاط المستمر خماسي الدرجات. وينبغي أن يتعذر تمييز أحد المصدرين "B" أو "C" عن المصدر "A". ويمكن أن يكشف المصدر الآخر عن مظاهر الانحطاط. ويجب تفسير أي اختلافات مُدركة بين المصدر المرجعي والمصدر الآخر باعتبارها انحطاطاً. وتُستخدم عادةً سمة واحدة فقط، وهي "جودة الصوت الأساسية". وتُعرف على أنها سمة شاملة تتضمن أي وجميع الاختلافات المكتشفة بين الإشارة المرجعية والإشارة قيد الاختبار.

ويتم التعامل مع سلم الدرجات باعتباره مستمراً مع وجود "نقاط ارتكاز" مستمدة من سلم الانحطاط خماسي الدرجات لقطاع الاتصالات الراديوية الوارد في التوصية ITU-R BS.1284 والمبين أدناه.

الشكل 3

سلم الانحطاط خماسي الدرجات لقطاع الاتصالات الراديوية



BS.1387-03

يعتمد تحليل نتائج اختبار الاستماع الشخصي بشكل عام على درجة الاختلاف الشخصية (SDG) المحددة على النحو التالي:

$$SDG = \text{GradeSignal Under Test} - \text{GradeReference Signal}$$

وينبغي أن تتراوح قيم الدرجة SDG بشكل مثالي بين 0 و-4، حيث تقابل القيمة 0 انحطاطاً غير محسوس و-4 انحطاطاً يُعتبر مزعجاً للغاية.

5 الاستبانة والدقة

درجة الاختلاف الموضوعية (ODG) هي متغير خرج من طريقة القياس الموضوعي وتقابل الدرجة SDG في الميدان الشخصي. وتفيد استبانة الدرجة ODG برقم عشري واحد. ومع ذلك، ينبغي للمرء أن يكون حذراً وألا يتوقع بشكل عام أن يكون الاختلاف بين أي زوج من الدرجات ODG بمقدار عُشر الدرجة كبيراً. وتسري نفس الملاحظة عند النظر في نتائج اختبار الاستماع الشخصي. ولا يوجد رقم واحد يصف بشكل كامل دقة طريقة القياس الموضوعي. وبدلاً من ذلك، يتعين على المرء أن يأخذ في الاعتبار عدداً من أرقام الجدارة المختلفة. وأحد هذه الأرقام هو العلاقة بين الدرجات SDG و ODG. ومن المهم أن نفهم أنه لا يوجد ضمان بأن الارتباط سيتجاوز قيمة محددة سلفاً. ومن المرجح إلى حد كبير أن يختلف أداء طريقة القياس، على سبيل المثال، باختلاف نوع ومستوى التدهور المُدخل.

وهناك رقم آخر يستحق الاهتمام وهو عدد القيم المتطرفة. وتُعرف القيمة المتطرفة على أنها قيمة مقاسة لا تلي مخطط التفاوت المحدد مسبقاً. ووفقاً لمتطلبات المستخدم، ينبغي أن توفر طريقة القياس أعلى دقة ممكنة للطرف العلوي من سلم الدرجات (أي جودة صوت عالية). وبالتالي، يُسمح للدقة المتحصل أن تكون أقل في النطاق الأوسط والأدنى من سلم الدرجات. وعلى الرغم من أن الارتباط عادة ما يعطي تقديراً جيداً لدقة طريقة القياس الموضوعي، فمن المهم أن نأخذ في الاعتبار أنه حتى رقم الارتباط المرتفع نسبياً يمكن أن يخفي أداء غير مقبول (من منظور القيم المتطرفة) لطريقة القياس. ورقم الجدارة الثالث الذي تم استخدامه أثناء عملية التحقق من الصحة هو درجة الخطأ المطلق (AES)، والتي تعكس متوسط العلاقة بين حجم فاصل الثقة بالنسبة للدرجة SDG والمسافة بين الدرجتين SDG و ODG. ويمكن الاطلاع على مزيد من التفاصيل حول الأداء المتوقع لطريقة القياس وكذلك الأداء أثناء عملية التحقق من الصحة في المرفق 1 بالملحق 2.

6 المتطلبات والقيود

يجب أن تكون الإشارة الصادرة عن الجهاز قيد الاختبار والإشارة المرجعية متزامنتين زمنياً بدقة 24 عينة خلال فترة القياس الكاملة. ولا تشكل آلية التزامن جزءاً من هذه التوصية ومن المتوقع أن تختلف من تنفيذ إلى آخر.

المرفق 1

بالملاحق 1

التطبيقات

1 عام

يقدم هذا المرفق التعاريف والمتطلبات المحددة للتطبيقات الرئيسية التي تستهدفها طريقة القياس الموضوعي الموصى بها لجودة الصوت المدركة.

وتتطلب بعض التطبيقات تنفيذاً في الوقت الفعلي لطريقة القياس الموضوعي، بينما يكون القياس في غير الوقت الفعلي كافياً بالنسبة لتطبيقات أخرى. وبالنسبة للتنفيذ في الوقت الفعلي، يوصى بألا يتجاوز الحد الأقصى للتأخير عبر معدات القياس 200 ms وألا يُقبل التأخير إذا كان أكثر من ثانية واحدة.

وعلاوة على ذلك، لا بد من التمييز بين القياسات عبر الإنترنت وخارجها. ففي القياسات خارج الإنترنت، يتمتع إجراء القياس بإمكانية النفاذ الكامل إلى الجهاز أو التوصيل به بينما يشير القياس عبر الإنترنت إلى أن البرنامج قيد التشغيل، وهو ما يجب عدم قطع تشغيله بواسطة القياس.

2 التطبيقات الرئيسية

1.2 تقييم عمليات التنفيذ

تحتاج الهيئات الإذاعية ومشغلو الشبكات وغيرهم إلى تقييم عمليات التنفيذ المختلفة للمعدات، ولا سيما الكوديكات الصوتية، وذلك عند اختيار هذه المعدات لشرائها أو عند إجراء اختبارات القبول.

وبالنسبة لهذا النوع من التطبيقات، يلزم توفر دقة عالية خاصة لتقييم الانحطاطات الصغيرة وتصنيف التطبيقات المختلفة بشكل صحيح. وفيما يتعلق بمتغيرات الخرج، يعد الخرج البسيط مثل الدرجة ODG كافياً للمستخدمين، ولكن يمكن لمطوري الكوديكات الصوتية إجراء تحليل أكثر شمولاً باستخدام مجموعة مناسبة من متغيرات خرج النماذج (MOV). يمكن استخدام صيغتي النموذج كليهما، ولكن يوصى بالصيغة المتقدمة.

2.2 ترتيب الجودة الإدراكية

هذا إجراء سريع يتم قبل إدخال جزء من المعدات أو الدارة في الخدمة. والهدف هو التحقق من الأداء الوظيفي والجودة. ويتولى موظفو التشغيل التعامل مع معدات القياس. وقد يكون هناك أي نوع من التشوه.

ويلزم إجراء القياس في الوقت الفعلي. ويمكن استخدام إشارات الاختبار أو الإشارات الصوتية المحددة مسبقاً. وينبغي عرض الدرجات ODG بشكل صحيح وينبغي تقديمها مرتين على الأقل في الثانية، أو مباشرة بعد نهاية إشارة الاختبار في حالة استخدام إشارة اختبار خاصة.

ويكفي استخدام الصيغة الأساسية.

3.2 المراقبة عبر الإنترنت

هذه عملية مستمرة تحدث أثناء البث الصوتي المستمر. ويجب ألا يتم قطع تشغيل البرنامج بإجراء القياس. ومن ثم، يجب استخدام إشارة البرنامج نفسها أو جزء صوتي محدد مسبقاً لمن أجل القياس. وقد يكون الأخير إشارة محطة إذاعية أو ترنيمية. ويتولى موظفو التشغيل التعامل مع معدات القياس.

ويلزم إجراء القياس في الوقت الفعلي. ويجب أن يتم عرض الدرجات ODG بشكل صحيح ويجب تقديمها مرتين على الأقل في الثانية الواحدة أو مباشرة بعد نهاية الإشارة المحددة مسبقاً. ولا يُجذب عرض الدرجات MOV.

ويكفي استخدام الصيغة الأساسية.

4.2 حالة المعدات أو التوصيلة

لضمان عمل التوصيلات الصوتية أو المعدات، يلزم إجراء فحص مكثف للجودة من وقت لآخر. وعلى النقيض من المراقبة عبر الإنترنت أو الترتيب الإدراكي، يتطلب هذا التطبيق التحقق من العديد من المعلومات التقنية.

وينبغي لنظام القياس أن يقدم معلومات مفصلة عن تأثير حالة المعدات أو التوصيلة على جودة الصوت المدركة من خلال عرض المجموعة الكاملة من المتغيرات MOV بالإضافة إلى الدرجات ODG. ولا يلزم إجراء القياس في الوقت الفعلي.

ويوصى باستخدام الصيغة المتقدمة.

5.2 تحديد هوية الكوديك

لتحديد هوية الكوديكات (خوارزميات مختلفة أو عمليات تنفيذ مختلفة لنفس الخوارزمية)، يجب أن يكون نظام القياس قادراً على تخزين واسترجاع ومقارنة أنماط الخصائص. ويمكن اعتبار التشابه بين الأنماط بمثابة مقياس للتشابه بين عمليات التنفيذ المختلفة للكوديكات. ويتم استخدام مثل هذا الإجراء لتحديد نوع أي كوديك معين وتنفيذه.

ويجب أن يسجل نظام القياس أكبر قدر ممكن من المعلومات عن الأنماط. والنظر في الدرجات ODG فقط قد لا يوفر معلومات كافية. ويعد استخدام الصيغة الأساسية كافياً، على الرغم من عدم الحاجة إلى القياس في الوقت الفعلي.

ملاحظة - لا يوجد سوى القليل من الخبرة في استخدام الطريقة الموصى بها. علاوة على ذلك، لم يتم بعد تحديد مقياس واحد للتشابه بين الأنماط.

6.2 تطوير الكوديكات

بالنسبة لهذا التطبيق، يجب أن تحدد طريقة القياس خصائص أداء الكوديك قيد الاختبار بأكثر قدر ممكن من الدقة وبأكثر قدر ممكن من التفصيل، ولا سيما بالنسبة للانحطاطات الصغيرة.

وتتطلب اختبارات المراقبة المستمرة معالجة في الوقت الفعلي وهو ما لا تدعمه الصيغة المتقدمة بالضرورة. ومع ذلك، سوف تتطلب حالات التدهور الصغيرة والمعلومات التفصيلية للصيغة المتقدمة. ويجب أن يكون نظام القياس قادراً على عرض المخرجات بنفس المعدل الذي تم حسابها به. ومن المرغوب فيه النفاذ المباشر إلى تاريخ المخرجات خلال فترة مدتها 4 ثوانٍ.

ويُوصى باستخدام الصيغة المتقدمة. ومع ذلك، فإنه للقياس في الوقت الفعلي، تكفي الصيغة الأساسية. ويلزم إجراء تحليل في الوقت الفعلي وفي غير الوقت الفعلي للرتل تلو الآخر. ولا بد من الإشارة إلى أي تشوه شديد، على سبيل المثال، بواسطة عرض الذروة. ومن المرغوب فيه النفاذ إلى المجموعة الكاملة من المتغيرات MOV.

7.2 تخطيط الشبكة

يتطلب تخطيط الشبكات تقييم الجودة المتوقعة في نقاط مختلفة أثناء عملية التخطيط. ويمكن استخدام محاكاة برمجية لمكونات الشبكة، والتي تسمح بدمج مراحل مختلفة لمعالجة الصوت، لفحص التشكيلات المختلفة من أجل تحسين جودة الصوت. وفي مرحلة لاحقة، يمكن اختبار مكونات معالجة الصوت الفعلية في التشكيلة المختارة.

ويتولى مهندسو النظام تخطيط الشبكة حيث ينبغي لهم استرجاع معلومات مفصلة عن تأثير خصائص الشبكة على جودة الصوت. يجب أن يعتمد تصنيف تشكيلات الشبكة المحتملة المختلفة على مجموعة مناسبة من المتغيرات MOV حسب التطبيق المحدد للشبكة. ومن ثم فإن عرض الدرجات ODG فقط لا يكفي. ولا يلزم إجراء القياس في الوقت الفعلي للتقييم في هذا التطبيق.

يمكن استخدام صيغتي النموذج كليهما، ولكن يوصى بالصيغة المتقدمة.

8.2 المساعدة في التقييم الشخصي

توفر طريقة القياس الموضوعي أداة لفرز المواد الصوتية المهمة المقرر استخدامها في اختبارات الاستماع الشخصية. ويمكن استخدام المجموعة الكاملة من المتغيرات MOV لتصنيف المواد الهامة.

ويلزم تحقيق أعلى دقة ممكنة ويوصى باستخدام الصيغة المتقدمة. ومع ذلك، يُجهد القياس في الوقت الفعلي لتقليل الوقت اللازم لاختيار المواد الحرجة.

9.2 ملخص التطبيقات

يلخص الجدول 2 المتطلبات فيما يتعلق بطريقة القياس للتطبيقات الرئيسية.

الجدول 2

المتطلبات فيما يتعلق بطريقة القياس

التطبيق	الفئة	الوقت الفعلي	الحد الأدنى لمعدل قيم الخرج، ROV ⁽¹⁾ (Hz)	عبر/خارج الإنترنت	صيغة النموذج
1	تشخيصي	لا	-	خارج الإنترنت	الصيغتان
2	تشغيلي	نعم/لا	2	خارج الإنترنت	الأساسية
3	تشغيلي	نعم	2	عبر الإنترنت	الأساسية
4	تشخيصي	نعم/لا	-	عبر/خارج الإنترنت	المتقدمة
5	تشخيصي	لا	-	خارج الإنترنت	الصيغتان
6	تطويري	نعم/لا	-	خارج الإنترنت	الصيغتان
7	تطويري	نعم/لا	-	خارج الإنترنت	الصيغتان
8	تطويري	نعم/لا	-	خارج الإنترنت	المتقدمة

(1) معدل قيم الخرج (في الثانية)

3 إشارات الاختبار

تُقسم إشارات الاختبار إلى مجموعتين: طبيعية واصطناعية. وتتألف قائمة إشارات الاختبار الطبيعية المقدمة هنا من تتابعات صوتية حرجة مستخدمة بالفعل في اختبارات الاستماع التي أجراها قطاع الاتصالات الراديوية ومنظمات أخرى لتقييم جودة الصوت. ويجب أن تكون الإشارات متاحة في موقع الإرسال وفي موقع القياس. وبالتالي، لا بد من وجود ذاكرة في جهاز القياس.

يتم تعريف الإشارات الاصطناعية رياضياً ويمكن تغييرها بطريقة يمكن التحكم فيها. ويمكن توليد هذه الإشارات في موقعي الإرسال والقياس. ولا توجد حاجة إلى ذاكرة إضافية في جهاز القياس. ونظراً لطبيعة هذه الإشارات، فمن الصعب، إن لم يكن من المستحيل، استخلاص درجات تقييم شخصية لها. ولذلك، لم يتم التحقق من صحة طريقة القياس مقابل النتائج الشخصية لهذه الإشارات.

1.3 اختيار إشارات الاختبار الطبيعية

يقدم الجدول 3 قائمة بمجموعة فرعية من إشارات الاختبار التي استخدمت أثناء إجراء التحقق الذي أنتج هذه التوصية. ويشار أيضاً إلى نوع الأخطاء البشرية، التي تكشف عنها هذه الإشارات عادةً بسبب التشفير ذي معدل البتات المنخفض.

الجدول 3

قائمة بمجموعة فرعية من إشارات الاختبار

الرقم	العنصر	اسم الملف	ملاحظات
1	صنجات	cas	(1)
2	الكلارينيت	cla	(2)
3	العصا	clv	(1)
4	الفلوت	flu	(2)
5	آلة الاكسيليفون	glo	(1) و(2) و(5)
6	البيان القيثاري	hrp	(1) و(2) و(4)
7	طبلة الغلاية	ket	(1)
8	الماريمبا	mar	(1)
9	بيانو شوبرت	pia	(2)
10	أنبوب النغمات المرجعية	pip	(4)
11	راي كودر	ryc	(2) و(4)
12	الساكسفون	sax	(2)
13	موسيقى القرب	sb1	(2) و(4) و(5)
14	كلام لأنثى باللغة الإنكليزية	sfe	(3)
15	كلام لذكر باللغة الإنكليزية	sme	(3)
16	كلام لذكر باللغة الألمانية	smg	(3)
17	الطبول الجانبية	sna	(1)
18	سوبرانو لموزارت	sop	(4)
19	دف صغير (رق)	tam	(1)
20	بوق	tpt	(2)
21	آلة المثلث الإيقاعية	tri	(1) و(2) و(5)
22	آلة توبا	tub	(2)
23	سوزان فيجا	veg	(3) و(4)
24	إكسيليفون	xyl	(1) و(2)

(1) إشارات عابرة: حساسة لما قبل الصدى، وإثارة الضوضاء في الميدان الزمني.

(2) بنية نغمية: حساسة للضوضاء، خشونة.

(3) كلام طبيعي: (مزيج حرج من الأجزاء النغمية والأصوات الحادة): حساسة للتشوهات وإثارة الصوت الحادة.

(4) صوت مركب: يضغط على الجهاز قيد الاختبار.

(5) عرض نطاق كبير: يضغط على الجهاز قيد الاختبار، فقدان الترددات العالية، ضوضاء ذات تردد عال مشكلة بالبرنامج.

2.3 مدة الإشارة

ينبغي أن تكون مدة إشارة الاختبار الطبيعية هي نفس مدة إشارة الاختبار الطبيعية تقريباً كما لو كانت ستستخدم في اختبار استماع. وتتراوح المدة عادةً من 10 إلى 20 ثانية. ومن المحتمل جداً أن يقتصر الجزء الخرج من إشارة الاختبار، الذي يكشف عن معظم الأخطاء البشرية، على جزء قصير فقط من المدة.

وينبغي أن تكون مدة إشارات الاختبار الاصطناعية طويلة بما يكفي للضغط على الكوديك قيد الاختبار، والذي قد يحتوي على دارئ للإشارة الصوتية المشفرة. وبالنظر إلى أطوال الدارئ هذه والثوابت الزمنية الموجودة في طريقة القياس، يجب أن تكون مدة كل عنصر اختبار واحد في التابع أكثر من 500 ms. ويمكن أن تقتصر المدة على هذه القيمة القصيرة لأنه ليس من المتوقع استخدام هذه الإشارات في اختبارات الاستماع الشخصية.

4 التزامن

بالنسبة لإجراء القياس، يجب أن تتم مزامنة الإشارة قيد الاختبار والإشارة المرجعية من حيث الزمن. وينطبق ذلك على إشارات الاختبار الطبيعية والاصطناعية.

5 مسائل حقوق التأليف والنشر

لا يمكن استعمال إشارات الاختبار الواردة في الجدول 3 دون حقوق التأليف والنشر إلا لأغراض القياس إلى جانب طريقة القياسات الموضوعية الموصوفة في الملحق 2 بهذه التوصية.

ملاحظة - يجب الحصول على إعفاء من حقوق التأليف والنشر لجميع التتابعات، بشكل رئيسي من اتحاد الإذاعات الأوروبية (EBU) (قرص EBU SQAM).

المرفق 2

بالملاحق 1

متغيرات الخرج

1 مقدمة

تقيس طريقة القياس الموضوعي الموصوفة في هذه التوصية جودة الصوت وتخرج قيمة تهدف إلى أن تتوافق مع جودة الصوت المدركة. وتحدد طريقة القياس الخصائص الأساسية للنظام الصوتي. وتحدد عدة مراحل وسيطة التأثيرات الصوتية الفسيولوجية والنفسية.

ويمكن استخدام هذه المخرجات الوسيطة لتحديد خصائص الأخطاء البشرية. وتسمى المعلمات متغيرات خرج النموذج (MOV). وتجمع المرحلة النهائية من نموذج القياس بين قيم المتغيرات MOV لإنتاج قيمة خرج واحدة تقابل بشكل مباشر النتيجة المتوقعة من تقييم شخصي للجودة.

2 متغيرات خرج النموذج

يحتوي الجدول 4 على وصف للمتغيرات MOV المستخدمة للتنبؤ بدرجات الاختلاف الموضوعي. يتم اشتقاق الرموز $Subscripts_A$ من جزء بنك التصفية في النموذج، بينما يتم اشتقاق الرموز $Subscripts_B$ من جزء المحول FFT في النموذج. ويمكن التنبؤ بدرجات الاختلاف الموضوعي إما من جزء المحول FFT فقط (الصيغة الأساسية) أو من مزيج من جزئي المحول FFT وبنك التصفية (الصيغة المتقدمة). يتم إجراء التوسيط دائماً مع مرور الوقت.

3 جودة الصوت الأساسية

المعلمة الأكثر شهرة في اختبارات الاستماع الشخصية هي جودة الصوت الأساسية (BAQ). وتُقاس الجودة BAQ كدرجة اختلاف شخصية (SDG) يتم حسابها على أنها الدرجة المعطاة للمرجع مطروحة من الدرجة المعطاة للإشارة قيد الاختبار في اختبار شخصي¹. عادة ما تكون للدرجة SDG قيمة سالبة. وتسمى معلمة الخرج المقابلة من النموذج بدرجة الاختلاف الموضوعي (ODG). ويعتمد التقابل بين المتغيرات MOV والدرجة ODG على عدد كبير من عناصر الاختبار الموثوقة، انظر المرفق 2 بالملحق 2.

الجدول 4

وصف معلمات خرج النموذج

وصف	معلمة خرج النموذج
متوسط الاختلاف المقسم إلى نوافذ في التشكيل (الأغلفة) بين الإشارة المرجعية والإشارة قيد الاختبار	WinModDiffB
متوسط الاختلاف في التشكيل	AvgModDiff1B
متوسط الاختلاف في التشكيل مع التركيز على التشكيلات المقدمة وتغييرات التشكيل حيث يحتوي المرجع على تعديلات قليلة أو معدومة	AvgModDiff2B
قيمة جذر متوسط التربيع للاختلاف في التشكيل	RmsModDiffA
قيمة جذر متوسط التربيع لجهازة الضوضاء لمكونات التردد المفقودة (المستخدمة في المتغير $RmsNoiseLoudAsym_A$)	RmsMissingComponentsA
قيمة جذر متوسط التربيع لمتوسط جهازة الضوضاء مع التركيز على المكونات المقدمة	RmsNoiseLoudB
المتغير $RmsNoiseLoudA + 0,5$ المتغير $RmsMissingComponentsA$	RmsNoiseLoudAsymA
مقياس لمتوسط التشوهات الخطية	AvgLinDistA
عرض نطاق الإشارة المرجعية	BandwidthRefB
عرض نطاق إشارة الخرج للجهاز قيد الاختبار	BandwidthTestB
لوغاريتم متوسط إجمالي نسبة الضوضاء إلى القناع	TotNMRB
الجزء النسبي من الأرتال التي يحتوي نطاق تردد واحد على الأقل فيها على مكون ضوضاء مهم	RelDistFramesB
اللوغاريتم المتوسط القطاعي لنسبة الضوضاء إلى القناع	AvgSegmNMRB
الحد الأقصى لاحتمال الكشف بعد ترشيح التمير المنخفض	MFPDB
متوسط المجموعة المشوهة (=الرتال)، يتم اعتباره لوغاريتم نسبة التشوه الإجمالي إلى إجمالي عدد الأرتال شديدة التشوه	ADBB
البنية التوافقية للخطأ مع مرور الوقت	EHSB

الدرجة ODG هي المعلمة المقاسة بشكل موضوعي والتي تقابل الجودة المدركة شخصياً. ونظراً لأن مهمة المستمع في اختبار الاستماع هي تقييم الجودة BAQ لعنصر الاختبار، فإن الدرجة ODG تعد أيضاً مقياساً للجودة BAQ.

¹ انظر التوصية ITU-R BS.1116.

4 هامش التشفير

هناك معلمة أخرى قد تكون ذات قيمة في المستقبل وهي هامش التشفير (CM)، وهي طريقة لوصف الأخطاء البشرية غير المسموعة. ويمكن تقييم هامش التشفير الشخصي (SCM) عن طريق تكبير الأخطاء البشرية حتى تصبح مسموعة للشخص القائم بالاختبار. ويصف الهامش SCM الارتفاع إلى عتبة إمكانية سماع الأخطاء البشرية.

ولإيجاد العتبة، يجب تكبير الأخطاء البشرية أو توهينها أثناء اختبار الاستماع. والطريقة المناسبة هي طريقة الاختلاف. ويتم تكبير إشارة الفرق للإشارة الأصلية والمشفرة المتزامنة مع الوقت وإضافتها إلى الإشارة الأصلية. ومن الأفضل إجراء اكتشاف عتبة السمعية باستخدام طريقة الاختيار القسري. ويتم الحصول على هامش SCM عن طريق حساب متوسط قيم العتبة للتكبير أو التوهين التي تم الحصول عليها من الأشخاص القائمين بالاختبار. تمثل قيم الهامش CM السالبة أخطاء مسموعة بينما تمثل قيم الهامش CM الموجبة أخطاء غير مسموعة. وعلى عكس الجودة BAQ، يعد هامش التشفير مقياساً للوقت الذي تصبح فيه الخطأ البشرية (على أي مستوى) مسموعة وليس مدى إزعاج هذه الأخطاء. ويرد وصف التعريف والتحقق من صحة طريقة قياس الهامش SCM في [Feiten، 1997].

وهامش التشفير الموضوعي (OCM) مشتق أيضاً من المتغيرات MOV. وفي الوقت الحاضر، لم يتم تقييم سوى عدد قليل من بنود الاختبار لهامش التشفير الشخصي. ولم تتم بعد دراسة التقابل إلى الهامش OCM من النموذج الوارد في هذه التوصية.

5 متطلبات المستعمل

تختلف متطلبات المستعمل فيما يتعلق بمتغيرات الخرج من طريقة القياس باختلاف التطبيق. وبالنسبة لبعض التطبيقات، فإن الرقمين 2 و3، على سبيل المثال (انظر المرفق 1 بالملحق 1)، يكون القياس جزءاً من إجراء تشغيلي. وفي هاتين الحالتين، من المهم جداً أن تكون نتائج الطريقة سهلة القراءة وسهلة الفهم للأشخاص الذين ليس لديهم معرفة متعمقة بتقنية القياس. ويتحقق ذلك على أفضل وجه إذا قامت الطريقة بإخراج قيمة واحدة فقط تتوافق مع جودة الصوت المدركة.

وقد ينطبق الشيء نفسه أيضاً على التطبيقات الأخرى، على سبيل المثال، التطبيقان 1 و4. ومع ذلك، بالنسبة لهذين التطبيقين، وكذلك للتطبيقات 5-8، قد تكون متغيرات الخرج الأكثر تعقيداً مفيدة للمستعملين الذين لديهم معرفة أعمق عن آليات طريقة القياس.

المرفق 3

بالملاحق 1

الخطوط العريضة للنموذج

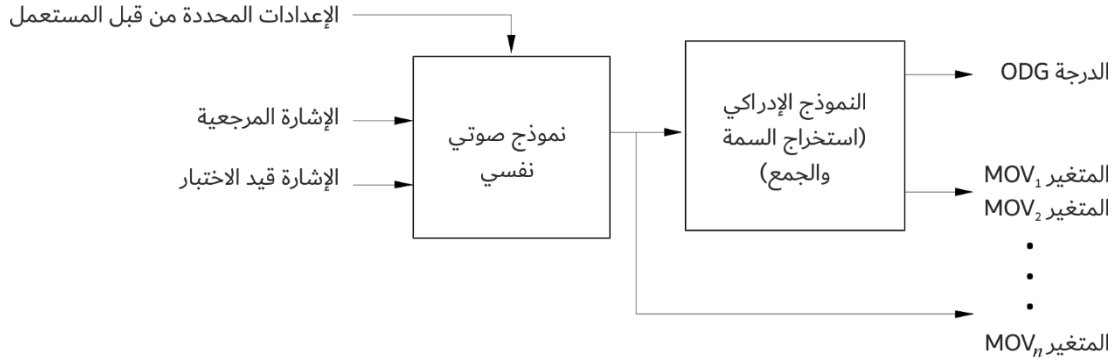
وفقاً للتوصية ITU-R BS.1116، يتم الحصول على الدرجة SDG لعنصر اختبار صوتي في اختبار استماع، ويمثل متوسط الدرجات SDG على عدد من المستمعين الجودة الشخصية للعنصر. وقد يحتوي العنصر على أنواع مختلفة من التشوهات الصوتية، لذلك يتم دمج التغيرات في الجودة بمرور الوقت. لذلك، يتطلب التنبؤ بالدرجات SDG بناءً على القياسات المادية نموذجاً دقيقاً للنظام السمعي المحيطي بالإضافة إلى الجوانب الإدراكية لأحكام جودة الصوت.

وينتج النموذج الموصى به للقياس الموضوعي عدداً من متغيرات خرج النموذج (MOV) بناءً على مقارنات بين الإشارة المرجعية والإشارة قيد الاختبار. ويجرى تقابل بين المتغيرات MOV هذه والدرجة ODG باستخدام تقنية استمثال ي تقلل إلى أدنى حد من الفرق التربيعي بين توزيع الدرجات ODG والتوزيع المقابل لمتوسط الدرجات SDG لمجموعة بيانات كبيرة بما فيه الكفاية.

تم وصف نوعين مختلفين من النموذج - صيغة قائمة على تحويل فورييه المتمايز (DFT) يمكن استخدامها للمراقبة في الوقت الفعلي، وصيغة أخرى، تعتمد على كل من بنك التصفية والتحويل DFT، والتي كان من المتوقع أن تعطي نتائج أكثر دقة. يُطلق على الصيغة المستندة إلى التحويل DFT اسم الصيغة الأساسية، بينما يُطلق على الصيغة المركبة اسم الصيغة المتقدمة. ويعرض الشكل 4 البنية رفيعة المستوى لكل من الصيغة الأساسية والصيغة المتقدمة.

الشكل 4

مراحل المعالجة المنفذة في النموذج



BS.1387-04

1 معالجة الصوت

كما هو الحال في اختبارات الاستماع الشخصية، يتم الحكم على جودة إشارة الاختبار بالنسبة للإشارة المرجعية. يتم تحويل كل من الإشارة المرجعية والإشارة قيد الاختبار (الإشارات غير المجسمة أو المجسمة) إلى تمثيل صوتي نفسي. وتُجرى مقارنة بين هذه التمثيلات من أجل استخلاص الدرجة ODG. ويتم تنفيذ هذه العمليات من خلال مراحل المعالجة الموضحة في الشكل 4.

1.1 الإعدادات المحددة من قبل المستعمل

تتطلب طريقة القياس مستوى الاستماع المفترض كمعلمة. ولذلك، يتعين على المستعمل توفير مستوى ضغط الصوت بوحدة dB SPL الناتج عن موجة جيبية كاملة النطاق بتردد 1019,5 Hz. وفي حالة عدم معرفة مستوى الاستماع الدقيق، يوصى بافتراض مستوى استماع يبلغ 92 dB SPL.

2.1 النموذج الصوتي النفسي

يحول النموذج النفسي الصوتي الأرتال المتعاقبة لإشارة المجال الزمني إلى تمثيل غشاء قاعدي. وتبدأ هذه العملية باستخدام كل من التحويل DFT وبنك التصفية. ويقوم التحويل DFT بتحويل البيانات إلى مجال التردد، وتتم مقابلة النتيجة من مقياس التردد إلى مقياس درجة الصوت، وهو المكافئ النفسي للصوتي للتردد. وفي جزء بنك التصفية من النموذج، يتم أخذ التقابل بين التردد ودرجة الصوت في الاعتبار مباشرةً من خلال عروض النطاق والمباعدة بين مرشحات تمرير النطاق.

يُستخدم مفهومان مختلفان لتحقيق التقنيع المتزامن. وتُحسب بعض المتغيرات MOV باستخدام مفهوم العتبة المقنعة، بينما يعتمد البعض الآخر على مقارنة التمثيلات الداخلية. ويحسب المفهوم الأول العتبة المقنعة مباشرة باستخدام وظائف التقنيع النفسي والمادي. وتعتمد متغيرات خرج النموذج على المسافة بين إشارة الخطأ المادية والعتبة المقنعة تلك. وفي مفهوم مقارنة التمثيلات الداخلية، يتم نشر طاقة كل من الإشارة قيد الاختبار (SUT) والإشارة المرجعية إلى مناطق الصوت المجاورة من أجل الحصول على

أنماط الإثارة. وتعتمد متغيرات خرج النموذج على إجراء مقارنة بين أنماط الإثارة تلك. ويُفقد التقنيع غير المتزامن عن طريق تلوخ تمثيلات الإشارة بمرور الوقت.

تُمدج العتبة المطلقة جزئياً من خلال تطبيق وظيفة للترجيح المعتمد على التردد وجزئياً عن طريق إضافة إزاحة تعتمد على التردد إلى أنماط الإثارة. وهذه العتبة عبارة عن تقدير تقريبي للحد الأدنى من الضغط المسموع [7-389 ISO، الصوتيات - الصفر المرجعي لمعايرة أجهزة قياس الصوت - الجزء 7: العتبة المرجعية للاستماع في ظل ظروف الاستماع للمجال الحر ومجال الانتثار، 1996].
والمخرجات الرئيسية للنموذج الصوتي النفسي هي الإثارة والعتبة المقنعة كدالة في الزمن والتردد. ويُتاح خرج النموذج على عدة مستويات لمزيد من المعالجة.

3.1 النموذج الإدراكي

يقوم النموذج الإدراكي بتكثيف المعلومات المستقاة من تتابع من الأرتال التي ينتجها النموذج الصوتي النفسي. وأهم مصادر المعلومات لإجراء قياسات الجودة هي الاختلافات بين الإشارة المرجعية والإشارة قيد الاختبار في كل من مجال التردد ومجال الصوت. ففي مجال التردد، يتم قياس عرض النطاق الطيفي لكلا الإشارتين، وكذلك البنية التوافقية في الخطأ. وعند قياس الخطأ في مجال الصوت، تُشتق مقاييس الخطأ من كل من تشكيل غلاف الإثارة ومقدار الإثارة.

وتُوزن الميزات المحسوبة، بحيث يؤدي تجميعها إلى درجة ODG قريبة بالقدر الكافي من الدرجة SDG للتشويه الصوتي الخاص محل الاهتمام. وتستخدم الصيغة الأساسية 11 سمة لإنتاج درجة ODG واحدة، بينما تستخدم الصيغة المتقدمة خمس سمات. وأجري الاستمثال باستخدام خوارزمية تعلم الشبكة العصبية ذات الانتشار الخلفي (انظر الملحق 2، الفقرة 6). وتتألف بيانات التدريب من قاعدتي البيانات 1 و2 بالكامل وجزء من قاعدة البيانات 3. وتم الحصول على بيانات اختبار التعميم من بقية قاعدة البيانات 3 وقاعدة البيانات CRC97 بالكامل (انظر المرفق 2 بالملحق 2).

المرفق 4

بالملحق 1

مبادئ وخصائص طرائق قياس جودة الصوت الإدراكية الموضوعية

1 مقدمة والتسلسل التاريخي

يعتمد الإرسال الرقمي للإشارات الصوتية وتخزينها بشكل متزايد على خوارزميات خفض البيانات، والتي تتكيف مع خواص الجهاز السمعي البشري وتعتمد بشكل خاص على تأثيرات التقنيع. ولا تهدف هذه الخوارزميات بشكل أساسي إلى تدنية التشوهات، بل تحاول التعامل مع هذه التشوهات بطريقة بحيث لا تُدرك إلا بأقل قدر ممكن. ولم يعد ممكناً تقييم جودة هذه المشفرات الإدراكية بطرائق القياس التقليدية، التي تُحدد عادة القيمة الإجمالية للتشوه. ومن الأمثلة التي كثيراً ما تُذكر لتوضيح هذه القيود ما يسمى بمعجزة 13 dB: الضوضاء المتراكبة ذات البنية الطيفية المتكيفة مع بنية الإشارة الصوتية والتي تكاد تكون غير مسموعة حتى لو انخفضت نسبة الإشارة إلى الضوضاء (S/N) غير الموزونة الناتجة إلى 13 dB.

ولهذا السبب، تتطلب تقييمات الكوديكات الإدراكية اختبارات استماع من أجل تقييم جودة الصوت. وتتطلب الموثوقية والتكرار الكافيان لاختبارات الاستماع إنفاقاً كبيراً للوقت والعمل.

ويمكن أن تساعد مخططات القياس الموضوعي التي تتضمن خواص الجهاز السمعي البشري في التغلب على هذه المشكلات. وتم نشر هذه الفكرة لأول مرة بواسطة [Schroeder وآخرون، 1979]. وفي هذه الورقة، التي تدور بشكل أساسي حول تشفير الكلام، يتم وصف مخطط قياس "جهاز الضوضاء (NL)".

وفي هذه الورقة، يتم تقدير جهاز الصوت المدركة لإشارة الضوضاء الخاصة بكوديك الكلام، والتي هي الفرق بين إشارة الدخل والخرج له، لكل رتل زمني يبلغ حوالي 20 ms. فإذا كانت إشارة الضوضاء مقنعة تماماً، فإن جهاز الصوت المدركة تكون صفراً. ويعمل التقنيع الجزئي على خفض جهاز إشارة الضوضاء غير المقنعة. وتُستعمل العتبة المقنعة المستخدمة لضوضاء تقنيع النغمات ويتم حساب تدهور الكلام النهائي لكل رتل. ولا يتم حساب أي ملخص للجودة الإجمالية لعينة الكلام.

وفي عام 1985 نشر كارجالاينن مخطط القياس "الفرق الطيفي السمعي (ASD)" [Karjalainen، 1985]. ولقد بدأ بعدة أفكار من شرودر وأتال وهال، لكنه استعاض عن التحليل القائم على الأرتال بينك تصفية به مرشحات متراكبة، وغير طريقة تضمين العتبة المطلقة وأضاف نموذجاً للتقنيع الزمني. وتُعالج كل من إشاراتي الدخل إلى نظام القياس بنفس الطريقة تماماً، مما ينتج عنه نوع من التمثيل الداخلي. وتُقارن هذه التمثيلات الداخلية مع بعضها البعض لشرح الاختلافات المدركة بين إشارتي الدخل والخرج لمخطط تشفير الكلام. ولا يتم حساب أي ملخص للجودة الإجمالية لعينة الكلام. وتُكيف الدقة الزمنية للفرق ASD بشكل أفضل مع خواص النظام السمعي البشري ولكنها تزيد من تعقيد الخوارزمية.

وفي عام 1987، نشر براندنبورغ مخطط القياس "نسبة الضوضاء إلى الفناء (NMR)" [Brandenburg، 1987]، والذي كان من المفترض استخدامه كأداة لتطوير مخططات التشفير الصوتي. تم تقليل تعقيد المخطط مقارنة بالمخطط NL عن طريق حساب الانتشار على النطاقات الإدراكية باستخدام دالة الانتشار التي تم تصميمها على أنها منحى الحالة الأسوأ. وتم استمثال العتبة المقنعة المستخدمة لنغمة تقنيع الضوضاء. وأضيف مخطط بسيط لنمذجة ما بعد التقنيع والعديد من الطرائق لتقييم الجودة المدركة لمقتطفات الصوت الأطول. وكان هذا المخطط هو الأول الذي تم تنفيذه في عتاد الوقت الفعلي.

وفي عام 1989 قدم مور وجلاسبيرغ [Moore، 1989] نموذجاً إدراكياً ولكنهما لم يقدموا طريقة للحكم على الجودة المدركة للإشارات الصوتية الضعيفة.

2 البنية العامة لطرائق قياس جودة الصوت الإدراكية الموضوعية

تعمل جميع مخططات القياس الإدراكي مع إشارتي دخل: إحداهما تسمى الإشارة المرجعية (REF)، والأخرى تسمى الإشارة قيد الاختبار (SUT). وفي الحالات التي لا يمكن فيها إرسال المرجع إلى جهاز القياس، ولكن الإشارة تكون معروفة جيداً، يمكن أن تكون الإشارة المرجعية مرجعاً داخلياً مخزناً في جهاز القياس نفسه. ومن الضروري أن تكون إشاراتي الدخل متراصفتين زمنياً.

ويمكن دمج النماذج الصوتية النفسية في مخططات القياس بطريقتين مختلفتين. الطريقة الأولى تشبه إلى حد كبير بنية مخططات التشفير الصوتي: تُستخدم الإشارة المرجعية لحساب تقدير العتبة المقنعة الفعلية (انظر أدناه). تتم مقارنة الفرق بين الإشارة قيد الاختبار والإشارة المرجعية بهذه العتبة المقنعة. وتسمى هذه الطريقة "مفهوم العتبة المقنعة" وتستخدم في جهاز الضوضاء والنسبة NMR. ويمكن حساب الفرق بين إشاراتي الدخل إما في المجال الزمني أو كالفرق بين أطيايف الطاقة قصيرة الأجل. وتوفر الطريقة الثانية متانة أفضل ضد أخطاء التراصف الزمني ولكنها تقلل من الدقة الزمنية. وعادة ما يكون الفرق في المجال الزمني حساساً جداً لتشوهات الطور، وبالتالي لم يعد يستخدم.

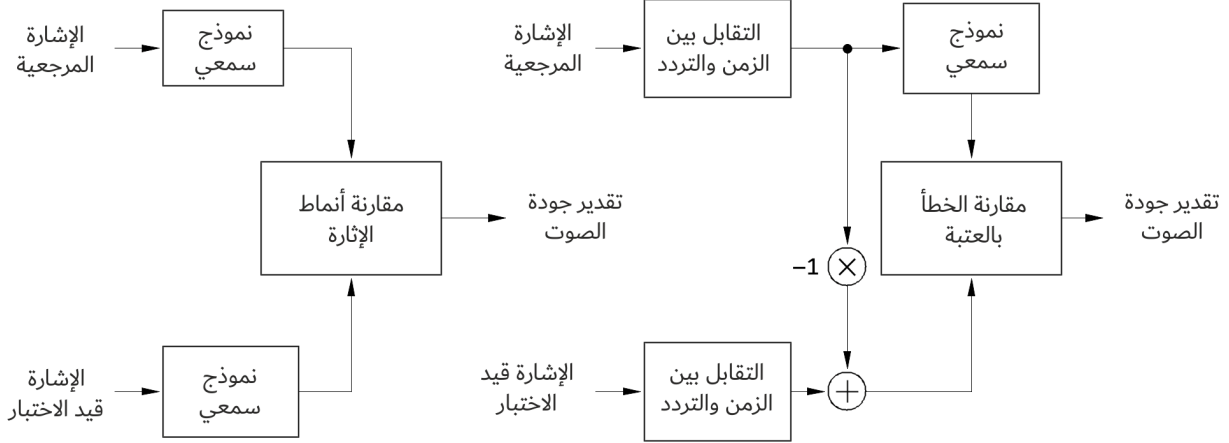
والطريقة الثانية أقرب إلى العمليات الفسيولوجية في الجهاز السمعي البشري: يتم حساب ما يسمى بالتمثيل الداخلي لكل من الإشارة المرجعية والإشارة قيد الاختبار. وهذا التمثيل الداخلي هو تقدير للمعلومات المتوفرة للدماغ البشري لمقارنة الإشارات. وتسمى هذه الطريقة "مقارنة التمثيلات الداخلية" وتستخدم في المخطط ASD.

3 الأسس النفسية والصوتية والإدراكية

يناقش هذا القسم خواص الجهاز السمعي البشري التي تعتبر الأبرز في تقييم الجودة المدركة للإشارات الصوتية. وينصب التركيز الرئيسي على كيفية نمذجة هذه الخواص.

الشكل 5

المفاهيم الصوتية النفسية المستخدمة في النهج المختلفة لمخططات القياس الإدراكية



مقارنة التمثيلات الداخلية

مفهوم العتبة المقنعة

BS.1387-05

1.3 خاصية النقل للأذن الخارجية والوسطى

بوجه عام، يجب أن تمر الإشارات الصوتية عبر الأذن الخارجية والوسطى حتى تصل إلى الأذن الداخلية حيث تُجرى عمليات كشف الصوت وتحليله. وتقوم الأذن الخارجية والوسطى بإجراء عملية مرشح تمرير النطاق على إشارة الدخل. وتُضاف الضوضاء الموجودة في العصب السمعي، إلى جانب الضوضاء الناتجة عن تدفق الدم، إلى إشارة الدخل. ويزداد اتساع هذه الضوضاء مع الترددات المنخفضة. وتحد وظيفة النقل للأذن الخارجية والوسطى إلى جانب الضوضاء الداخلية من القدرة على اكتشاف الإشارات الصوتية الصغيرة، ويكون لها التأثير الأكبر على عتبة السمع المطلقة.

2.3 سلام التردد الإدراكية

مستقبلات ضغط الصوت في الأذن البشرية هي الخلايا الشعرية. وهي تقع في الأذن الداخلية، وبشكل أكثر دقة في القوقعة. وفي القوقعة، يتم إجراء تحويل التردد إلى موضع. ويعتمد موضع الإثارة القصوى على تردد إشارة الدخل. وكل خلية شعرية في موضع معين في القوقعة مسؤولة عن مدى متراكب على سلم التردد. ويرتبط الانطباع الإدراكي للصوت بمسافة ثابتة من الخلايا الشعرية.

واعتماداً على التجربة الصوتية النفسية المستخدمة، يتم العثور على وظائف تحويل مختلفة من التردد إلى الصوت:

وفي [1967، Feldtkeller و Zwicker] يرد جدول يقسم سلم التردد بوحدات Hz إلى 24 نطاقاً غير متراكب، ما يسمى بالنطاقات الحرجة. وترد في الجدول 5 ترددات القطع العليا لهذه النطاقات. ويحتوي الجدول أيضاً على تعريف لسلم بارك (Bark): Bark 1 يقابل 100 Hz، و Bark 24 يقابل 15 500 Hz.

الجدول 5

سلم النطاقات الحرجة كما عرفه زويكر

12	11	10	9	8	7	6	5	4	3	2	1	النطاق الحرج
1 720	1 480	1 270	1 080	920	770	630	510	400	300	200	100	تردد القطع العلوي (Hz)

24	23	22	21	20	19	18	17	16	15	14	13	النطاق الحرج
15 500	12 000	9 500	7 700	6 400	5 300	4 400	3 700	3 150	2 700	2 320	2 000	تردد القطع العلوي (Hz)

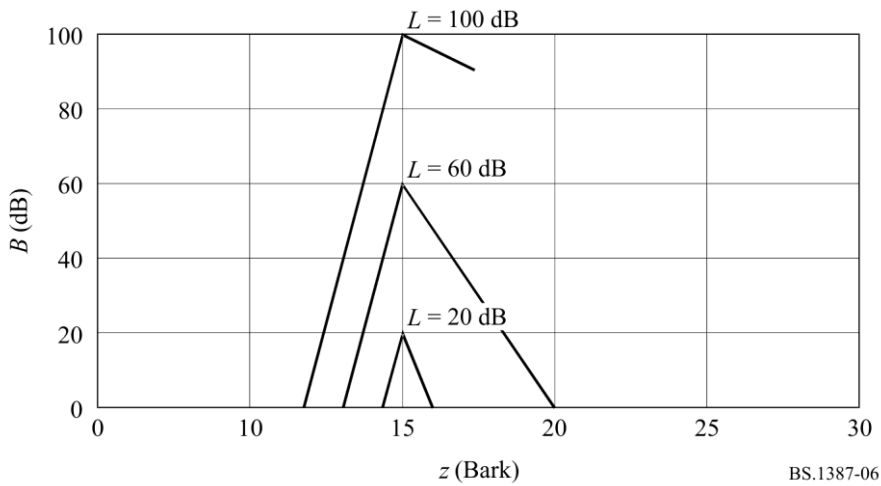
تم في الماضي تحديد عدة تقديرات تقريبية لسلم بارك. ويمكن الاطلاع على مناقشة تفصيلية للسالم المختلفة في [Fielder و Cohen، 1992]. وفي سياق القياس الموضوعي لجودة الصوت المدركة، تم تحقيق أفضل النتائج باستخدام سلم بارك.

3.3 الإثارة

تتفاعل كل خلية شعرية مع مجموعة من الترددات التي يمكن وصفها بخاصية المرشح. ويمكن التعبير عن ميل المرشحات بشكل أفضل على سلم إدراكي كما هو موضح أعلاه. ويكون شكل المرشحات على هذا السلم مستقلاً تقريباً عن التردد المركزي. والميل السفلي للإثارة مستقل عن المستوى L لإشارة الدخل (حوالي 27 dB/Bark). ويكون الميل العلوي أكثر انحداراً بالنسبة للمستويات الأدنى منه بالنسبة للمستويات الأعلى لإشارة الدخل (من -5 إلى -30 dB/Bark). وتنتج هذه الخاصية المنحدرة عن آلية ردود الفعل بين نوعين مختلفين من الخلايا الشعرية وتحتاج إلى بعض الوقت حتى تستقر. ولذلك يتم تحقيق أفضل دقة للتردد السمعي للإشارات الثابتة بعد عدة ميلي ثانية من بداية الإشارة. وتُضاف أنماط الإثارة للإشارات المكونة من عدة مكونات بطريقة غير خطية.

الشكل 6

تبعية مستوى الإثارة وفقاً لتهاردت [1979]



بعد التعرض للإشارة، تحتاج الخلايا الشعرية والمعالجة العصبية إلى بعض الوقت للتعافي حتى يتم الوصول إلى الحساسية الكاملة مرة أخرى. وتعتمد مدة عملية التعافي على مستوى الإشارة ومدتها ويمكن أن تستمر حتى عدة مئات من الملي ثانية. وتُعالج الإشارات

عالية المستوى بشكل أسرع من الإشارات منخفضة المستوى في الطريق بين الخلية الشعرية والدماغ. ولذلك، فإن بداية الإشارة العالية يمكن أن تقَع الإشارة الأضعف السابقة.

وهناك طريقة أخرى لنمذجة الإثارة تعتمد على سلم ERB [Moore، 1986]. وتستخدم هذه الطريقة ما يسمى بمرشحات ROEX [Moore، 1986]. وفي سياق القياس الموضوعي لجودة الصوت المدركة، تحققت نتائج أفضل باستخدام النماذج المستندة إلى [Feldtkeller و Zwicker، 1967] و [Terhardt، 1979].

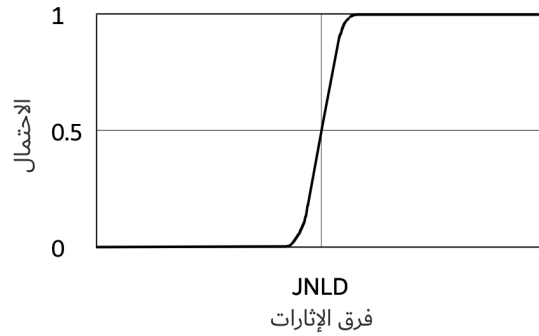
4.3 الكشف

تُنقل إثارة الإشارات الصوتية المختلفة إلى الدماغ البشري. وهناك ثلاثة أنواع مختلفة من الذاكرة تختلف حسب درجة التفاصيل ومدة وجود المعلومات: الذاكرة طويلة الأجل، والذاكرة قصيرة الأجل، والذاكرة قصيرة الأجل للغاية. وفي سياق اختبارات الاستماع، تلعب الذاكرة قصيرة الأجل للغاية الدور الأبرز. ويتم الاحتفاظ بمعظم تفاصيل الإشارة إذا كانت مدة المقتطف الصوتي أقل من خمس إلى ثماني ثوانٍ اعتماداً على المستمع والمقتطف الصوتي. ويؤخذ ذلك في الاعتبار في إجراء التقييم المحدد في التوصية ITU-R BS.1116 حيث يُسمح للأشخاص القائمين بالاختبار باختيار أجزاء قصيرة جداً من مقتطف صوتي للاستماع إليها عن كثب. وعند عتبة الكشف يكون احتمال الكشف 50%. وحول العتبة، يزداد احتمال كشف الفروق بسلسلة من 0% إلى 100%.

إن فرق المستوى الملحوظ بالكاد (JNLD) هو عتبة كشف فروق المستوى. ويتأثر الفرق JNLD بمستوى إشارات الدخل. وبالنسبة للإشارات الصغيرة، يلزم وجود فروق كبيرة للكشف (المستوى: 20 dB SPL، JNLD: 0,75 dB). وبالنسبة للإشارات العالية، تكون الحساسية للفروق الصغيرة أعلى بكثير (المستوى: 80 dB SPL، JNLD: 0,2 dB). وتعتمد هذه الأرقام على تجارب التشكيل بالاتساع.

الشكل 7

المبدأ الخاص باحتمال الكشف



BS.1387-07

5.3 التقنيع

الإشارة التي تكون مسموعة بوضوح إذا تم تقديمها بمفردها يمكن أن تكون غير مسموعة تماماً في وجود إشارة أخرى، وهي الإشارة القائمة بالتقنيع. ويُسمى هذا التأثير بالتقنيع وتسمى الإشارة المقنعة بالمقنَع. ويجب التمييز بين حالتين:

التقنيع المتآون

في هذه الحالة، يتم عرض المقنَع والمقنَع في نفس الوقت ويكونان شبه ثابتين. فإذا كان للمقنَع عرض نطاق منفصل، فسيتم رفع عتبة السمع حتى بالنسبة للترددات الأقل أو الأعلى من المقنَع. وتعتمد كمية التقنيع على بنيتي المقنَع والمقنَع. وفي الحالة التي تقوم فيها إشارة شبيهة بالضوضاء بتقنيع إشارة نغمية، يكون مقدار التقنيع مستقلاً عن التردد تقريباً. وإذا كان مستوى ضغط الصوت للمقنَع أقل بحوالي 5 dB عن مستوى المقنَع، يصبح غير مسموع. وفي الحالة التي تقوم

فيها الإشارة النغمية بتقنيع إشارة شبيهة بالضوضاء، فإن مقدار التقنيع يعتمد على تردد المقنع. ويمكن تقدير هذا المقدار بالمعادلة $dB \left(15.5 + \frac{z}{Bark}\right)$ ، حيث z هي معدل النطاق الحرج للمقنع. بالإضافة إلى ذلك، عند مستويات الإشارة العالية، تؤدي التأثيرات غير الخطية إلى خفض العتبة المقنعة بالقرب من المقنع. وتحدث تأثيرات مماثلة مع الاتجاه نغمة تقنيع نغمة. وتُضاف العتبات المقنعة لعدة إشارات بطريقة غير خطية. وبشكل عام، تكون العتبة المقنعة الناتجة أعلى من العتبة المقنعة التي تنتجها كل إشارة فردية.

التقنيع الموقت

في هذه الحالة، يتم عرض المقنع والمقنع في أوقات مختلفة. وبعد وقت قصير من اضمحلال المقنع، تكون العتبة المقنعة أقرب إلى التقنيع المتأون لهذا المقنع أكثر من العتبة المطلقة. وتبعاً لمدة المقنع، يمكن أن يتراوح وقت اضمحلال العتبة بين 5 ms (المقنع: نبضة غوسية مدتها حوالي 0,05 ms) وأكثر من 150 ms (القناع: ضوضاء وردية مدتها ثانية واحدة). والإشارات الضعيفة تكون قبل تقنيع الإشارات الأعلى جهارة مباشرة. وتبلغ مدة تأثير التقنيع العكسي هذا حوالي 5 ms. وإذا كان المقنع أعلى بقليل من العتبة، فلن يُدرك قبل المقنع ولكن كتغيير للمقنع. ويُظهر التقنيع العكسي انحرافات كبيرة من مستمع إلى آخر.

6.3 الجهارة والتقنيع الجزئي

تعتمد جهارة الصوت المدركة للإشارات الصوتية على ترددها ومدتها ومستوى ضغط الصوت. وبسبب التقنيع التلقائي، تكون جهارة الإشارة المركبة أقل من مجموع جهارة الصوت لجميع مكوناتها. وفي سياق قياس جودة الصوت، يتم تقليل جهارة التشويه غير المرغوب فيه المضاف إلى الإشارة المرجعية، أي جهارة الضوضاء، عن طريق التقنيع الجزئي الناتج عن الإشارة المرجعية.

7.3 الحدة

الحدة، إحدى القيم الأساسية للإحساس، ترتبط بطابع الصوت. ويُنظر إلى الصوت على أنه حاد إذا كان يحتوي بشكل أساسي على مكونات عالية التردد. فعلى سبيل المثال، يقال على نغمة جيبية أو ضوضاء محدودة بالنطاق عند الترددات العالية، أو ضوضاء تمرير مرتفع مع تردد قطع في مدى التردد فوق 3 kHz تقريباً أنها حادة. ومع ذلك، فإن بنية التردد التفصيلية للإشارة الصوتية ليس لها تأثير كبير على الحدة. وأجري البحث الأساسي المتعلق بالحدة بواسطة فون بسمارك [1974].

تم إجراء أبحاث إضافية بخصوص الحدة بواسطة [Aures، 1984]. وكانت نتيجة هذه الأبحاث دالة ترجيح معدلة قليلاً مقارنة بدالة الترجيح التي حددها بسمارك. وهي تساهم بشكل أقل في إسناد الحدة عند معدلات النطاق الحرج المنخفضة جداً والمرتفعة جداً، وأكثر عند معدلات النطاق الحرج بين 14 و 20 Bark. بالإضافة إلى ذلك، أظهرت هذه الأبحاث أن حدة الإشارات الصوتية ذات التباين الكبير في مستوى ضغط صوت الإشارة الصوتية والمحتويات القوية عالية التردد لا يمكن أن تعتمد فقط على جهارة الصوت الإجمالية، بل على دالة الترجيح، التي تعتمد على جهارة الصوت الإجمالية.

8.3 المعالجة الإدراكية

من الواضح أن جودة الصوت المدركة تتأثر بشدة بالتأثيرات الإدراكية. ويمكن إثبات ذلك من خلال تجربة بسيطة.

تُعالج إشارة مرجعية ذات ضوضاء أساسية مسموعة بوضوح بواسطة بعض المعدات الصوتية غير القادرة على نقل الضوضاء الأساسية هذه. وبما أن الضوضاء عبارة عن تشويه غير مرغوب فيه، فسيتم تصنيف الإشارة المرجعية بشكل أسوأ من الإشارة المعالجة في اختبار الاستماع. ومن ناحية أخرى، فإن نفس الإشارة المعالجة ستكون أسوأ في الدرجة إذا كان الجزء الأكثر أهمية في الإشارة المرجعية هو ضوضاء أساسية رخيمة.

إن إدراج جميع التأثيرات الإدراكية المحتملة يقع خارج نطاق هذا المرفق، ولكن فيما يلي بعض الأمثلة:

مثال 1: فصل التشوهات الخطية عن التشوهات غير الخطية

التشوهات الخطية أقل اعتراضاً من التشوهات غير الخطية. ويمكن فصل التشوهات الخطية عن التشوهات غير الخطية بسهولة إلى حد ما باستخدام الترشيح العكسي التكميلي لإشارة الخرج. وتستخدم الطريقة المحددة في هذه التوصية فصل التشوهات الخطية عن التشوهات غير الخطية.

مثال 2: تحليل المشهد الصوتي

تحليل المشهد الصوتي [Bregman، 1990] هو عملية إدراكية تسمح للمستمعين بفصل الأحداث الصوتية المختلفة وتجميعها في كائنات مختلفة. وكان النهج العملي، كما هو مذكور في [Beerends و Stemerding، 1994]، مفيداً في التحديد الكمي لتأثير تحليل المشهد الصوتي. فإذا لم يُشفّر مكون التردد الزمني بواسطة كوديك، فإن الإشارة المتبقية ستظل تشكل مشهداً صوتياً متماسكاً، في حين يؤدي إدخال مكون تردد زمني جديد غير ذي صلة إلى تصورين مختلفين. وبسبب الانقسام إلى تصورين مختلفين، سيكون التشوه أكثر اعتراضاً مما قد يتوقعه المرء على أساس جهازه عنصراً للتشوه الذي تم إدخاله حديثاً. ويؤدي هذا إلى عدم تناظر بين الاضطراب المتصور للتشوه الناجم عن عدم تشفير مكون التردد الزمني مقابل الاضطراب الناجم عن إدخال مكون تردد زمني جديد.

مثال 3: التقنيع المعلوماتي

يمكن نمذجة التقنيع المعلوماتي من خلال تحديد مقياس التعقيد الطيفي الزمني الذي يشبه الإنتروبيا. ويعتمد التأثير على الأرجح على مقدار التدريب الذي يتعرض له الأشخاص القائمين بالاختبار قبل إجراء التقييم الشخصي. وترد المحاولة الأولى لنمذجة هذا التأثير في [Beerends وآخرون، 1996] حيث تم حساب تقدير التعقيد المحلي خلال نافذة زمنية تبلغ حوالي 100 ms. فإذا كان هذا التعقيد المحلي مرتفعاً، فسيكون سماع التشوهات خلال هذه النافذة الزمنية أكثر صعوبة مما لو كان التعقيد المحلي منخفضاً. ويمكن أن يؤدي التدريب إلى تقليل العتبة المقنعة بمقدار عدة عشرات dB [Leek و Watson، 1984].

المثال 4: الترجيح الطيفي الزمني

تحمل بعض المناطق الطيفية الزمنية في الإشارة الصوتية معلومات أكثر، وبالتالي قد تكون أكثر أهمية من غيرها. وقد ثبتت أهمية الترجيح الطيفي الزمني في تقديرات الجودة على كوديكات الكلام. ففي الكلام، من الواضح أن بعض المكونات الطيفية الزمنية، مثل التكوينات، تحمل معلومات أكثر من غيرها [Beerends و Stemerding، 1994]. أما في الموسيقى، فإن جميع المكونات الطيفية الزمنية في الإشارة، حتى فترات الصمت، قد تحمل معلومات.

4 النماذج المدججة

1.4 مؤشر التشويش (DIX)

تعتمد طريقة القياس الإدراكي DIX (مؤشر التشويش) [Kabot و Thiede، 1996] على بنك المرشحات الصوتية الذي ينتج دقة زمنية عالية وبالتالي يسمح (مقارنة بالنهج المعتمدة على محول FFT) بنمذجة أكثر دقة للتأثيرات الزمنية مثل ما قبل وبعد التقنيع. ويتم الحفاظ على البنية الدقيقة الزمنية للأغلفة عند كل مرشح صوتي ويتم استخدامها للحصول على معلومات إضافية حول الإشارات والتشوهات المدخلة.

توزع الترددات المركزية للمرشحات الفردية بالتساوي على سلم درجة الصوت الإدراكي. والجزء العلوي من شكل المرشح مستدير قليلاً لضمان أن العدد المختار من المرشحات يغطي المدى التردد الكامل دون تموجات في استجابة التردد الإجمالية. ومن أجل نمذجة العتبات المقنعة، تنخفض ميول المرشحات بشكل كبير على سلم Bark. ويعتمد انحدار ميول المرشحات على مستوى إشارات الدخل. وتمت تغطية نطاق التردد المسموع بواسطة 80 مرشحاً في الإصدار الأول من النموذج DIX وتم تخفيضه لاحقاً إلى 40 مرشحاً، أي أن دقة التردد تقابل حوالي 0,6 Bark. وتعد خوارزمية بنك المرشحات سريعة إلى حد ما مقارنة بينوك المرشحات

الأخرى ذات المرشحات الفردية، ولكنها لا تزال تستهلك وقتاً أطول بكثير من التحويلات القائمة على المجموعات مثل تحويل FFT وتحويلات حزمة الموجات.

ويقوم النموذج DIX بتكليف المستويات والأطياف دينامياً بين الإشارة قيد الاختبار والإشارة المرجعية من أجل فصل التشوهات الخطية عن التشوهات غير الخطية. ويقوم بتقييم بنية الأغلفة الزمنية عند مخرجات المرشحات من أجل نمذجة الكمية المتزايدة من التقنيع الناتج عن المقنعات المعدلة والشبيهة بالوضوء مقارنة بالنغمات النقية.

وبمقارنة التمثيلات الداخلية للإشارة قيد الاختبار والإشارة المرجعية، يتم حساب العديد من معلمات الخرج، بما في ذلك جهارة الصوت الجزئية للتشوهات غير الخطية، ومؤشرات مقدار التشوهات الخطية، ومقاييس التأثيرات الزمنية والخاصة بالأذنين. ومع ذلك، يمكن تحقيق تقدير جيد لجودة الصوت الأساسية باستخدام اثنين فقط من معلمات الخرج: تُقابل جهارة الصوت الجزئية للتشوهات غير الخطية مع أحد المؤشرات الخاصة بكمية التشوهات الخطية لتقدير جودة الصوت الأساسية المتوقعة للإشارة قيد الاختبار.

2.4 نسبة الضوضاء إلى الإشارة المقنعة (NMR)

يقوم مخطط القياس NMR (نسبة الضوضاء إلى الإشارة المقنعة) [Brandenburg، 1987] بتقييم فرق المستوى بين العتبة المقنعة وإشارة الضوضاء. ويُستخدم التحويل DFT مع نافذة Hann مقدارها 20 ms لتحليل محتوى تردد الإشارة. وتُدمج معاملات التحويل في نطاقات وفقاً لسلم Bark. وتُقدر العتبة المقنعة لكل نطاق. ويتم اشتقاق ميل العتبة المقنعة باستخدام نُهج الحالة الأسوأ مع الأخذ في الاعتبار أن الميول تكون أكثر انحداراً بالنسبة للإشارات الضعيفة ولكنها تصل إلى العتبة المطلقة عند المستويات الأعلى. وتُكيف العتبة المطلقة مع دقة إشارة الدخل (عادةً 16 بتة)، ولكن ليس مع المتطلبات الصوتية النفسية. وبسبب هذه الحقائق يعتبر المخطط NMR قوياً في مواجهة التغيرات في مستوى الاستعادة. وتبلغ دقة سلم درجة الصوت حوالي 1 Bark. ونظراً لأن القدرة الحاسوبية المطلوبة منخفضة، فقد كان من الممكن تنفيذ المخطط NMR كنظام في الوقت الفعلي في مرحلة مبكرة من تطويره.

وتم استخدام النموذج منذ عام 1987 وأثبت اعتماده الأساسية.

وأهم قيم خرج المخطط NMR هي معدل علم التقنيع، الذي يعطي النسبة المثوية للأرتال ذات التشوهات المسموعة، بالإضافة إلى إجمالي ومتوسط نسبة الضوضاء إلى الإشارة المقنعة وهي طرق مختلفة لحساب متوسط المسافة بين طاقة الخطأ والعتبة المقنعة.

3.4 التقييم الموضوعي للإشارة الصوتية

يستخدم مخطط القياس OASE (التقييم الموضوعي للإشارة الصوتية) [Sporer، 1997] بنك مرشحات يحتوي على 241 مرشحاً لتحليل إشارات الدخل. والترددات المركزية متباعدة بشكل متساوٍ على سلم بارك بمسافة 0,1 Bark. وتتراكب المرشحات مع بعضها البعض. ويُكيف كل مرشح مع استجابة التردد لنقطة على الغشاء القاعدي. ويُضمن اعتماد مستوى الميول من خلال نُهج الحالة الأسوأ كما هو الحال في المخطط NMR. وتتطلب المرشحات ذات الترددات المركزية المنخفضة حساباً بمعدل أخذ العينات الكامل بينما يمكن حساب المرشحات ذات الترددات المركزية الأعلى بمعدل أخذ عينات منخفض. وبعد المرشحات، يتم حساب نموذج للتأثيرات الزمنية للجهاز السمعي البشري، كما هو الحال في ASD. وابتاع هذه الخطوة، من الممكن تقليل معدل أخذ العينات في جميع نطاقات المرشحات. ويؤدي هذا إلى دقة زمنية لبنك المرشحات تبلغ 0,66 ms بمعدل أخذ عينات يبلغ 48 kHz. ويُقارن خرج مرشحات المواءمة للإشارة المرجعية والإشارة قيد الاختبار مع احتمال دالة الكشف. وتستخدم هذه الدالة جهارة صوت إشارات الدخل كدخل لحساب الفرق JNLD. ويُشتق الاحتمال الإجمالي للكشف من احتمال الكشف لكل نطاق. وتُجرى هذه العملية لكل من قنوات الدخل وأيضاً لما يسمى بالقناة المركزية. إن احتمال الكشف في القناة المركزية لكل نطاق هو أسوأ حالة لاحتمال الكشف للقناة اليسرى واليمنى. ولكل رتل مدته 0,66 ms يتم حساب مجموع الخطوات فوق العتبة أيضاً.

وتُستخدم عدة أساليب لحساب المتوسط الزمني لاحتمال الكشف والخطوات فوق العتبة:

- المتوسط الزمني لاحتمال الكشف؛
- تردد الأرتال مع احتمال كشف فوق 0,5؛

- الحد الأقصى لاحتمال الكشف مع ترشيح بتمرير منخفض؛
- الحد الأقصى لاحتمال الكشف مع ترشيح بتمرير منخفض مع النسيان؛
- متوسط عدد الخطوات فوق العتبة للأرتال التي يزيد احتمال اكتشافها عن 0,5؛
- متوسط عدد الخطوات فوق العتبة؛
- الحد الأقصى لعدد الخطوات فوق العتبة؛
- متوسط عدد الخطوات فوق العتبة لنسبة 10% من أسوأ الأرتال.

4.4 قياس جودة الصوت الإدراكية (PAQM)

المبدأ الأساسي للمخطط PAQM [Beerends و Stemerding، 1992] هو طرح التمثيلات الداخلية (التمثيلات داخل رأس الشخص القائم بالاختبار) للإشارة المرجعية والإشارة المتدهورة ورسم خريطة للفرق بإجراء تقابل إدراكي لجودة الصوت المدركة شخصياً. ويتم التحول من المجال المادي الخارجي إلى المجال النفسي الجسدي الداخلي عن طريق أربع عمليات:

- إجراء تقابل للتردد الزمني الذي يتم عبر المحول DFT مع نافذة Hann مدتها حوالي 40 ms؛
- تعديل التردد باستخدام سلم بارك؛
- نشر التردد الزمني (تحويل غير خطي)؛
- تعديل الشدة (الانضغاط).

يتيح الجمع بين التحويل والانضغاط نمذجة سلوك التقنيع للجهاز السمعي البشري عند العتبة المقنعة وفوقها. ويُستعمل الانضغاط باستخدام النتائج الشخصية لتقييم كوديك الصوت MPEG الأول [تقرير اختبار 11 /SC 2/WG 11 ISO/IEC JTC 1/MPEG/Audio، الوثيقة MPEG90/N0030، أكتوبر 1990] [تقرير اختبار 11 /SC 2/WG 11 MPEG/Audio، الوثيقة MPEG91/N0010، يونيو 1991]. يتم التعبير عن الفرق في التمثيل الداخلي باضطراب الضوضاء. وفي أحدث إصدارات PAQM، كما قدمت إلى قطاع الاتصالات الراديوية، تم تضمين تأثيرين إدراكيين في التقابل بدءاً من اضطراب الضوضاء وحتى الجودة الشخصية والتدفق الإدراكي [Beerends و Stemerding، 1994] والتقنيع المعلوماتي [Beerends وآخرون، 1996].

تم تطوير إصدار مبسط من PAQM، [Beerends و Stemerding، 1994] باستخدام نموذج إدراكي كما هو معروض في [Beerends و Stemerding، 1994] ولكن تم تديده مع ترجيح الفواصل الصامتة. وأثناء تطوير PSQM، اتضح أنه عند الحكم على جودة الكلام في سياق الاتصال الهاتفي، فإن الضوضاء التي تحدث أثناء الفواصل الصامتة تكون أقل أهمية من الضوضاء التي تحدث أثناء فترات الكلام النشطة. وفي معيار مرجعي أجراه قطاع تقني الاتصالات، أظهر مقترح المخطط PSQM أعلى ارتباط بين الجودة الموضوعية والجودة الشخصية (لجنة الدراسات 12 لقطاع تقني الاتصالات، COM 12 74 - مراجعة اختبارات التحقق من صحة المقاييس الموضوعية لجودة الكلام). وتوصف التوصية ITU-T P.862 التقييم الإدراكي لجودة الكلام (PESQ) حيث تحتوي على خوارزمية محسنة لتقييم جودة الكلام.

5.4 التقييم الإدراكي

ينمذج PERCEVAL (التقييم الإدراكي) [Paillard وآخرون، 1992] خصائص النقل للأذن الوسطى والداخلية لتشكيل تمثيل داخلي للإشارة. تتحلل إشارة الدخل إلى تمثيل تردد زمني باستخدام محول DFT. وعادةً، يتم تطبيق نافذة هان مقدارها 40 ms على بيانات الدخل، مع تراكم بنسبة 50% بين النوافذ المتعاقبة. ويتم ضرب طيف الطاقة في دالة تعتمد على التردد لتمذج تأثير قناة الأذن والأذن الوسطى. وتُقابل قيم الطاقة الطيفية الموهنة من سلم التردد إلى سلم درجة الصوت الأكثر خطية فيما يتعلق بكل من الخواص الفيزيائية للأذن الداخلية والتأثيرات الجسدية النفسية المرصودة. وتُحور بعد ذلك مكونات الطاقة المحولة بدالة انتشار

لمحاكاة تشتت الطاقة على طول الغشاء القاعدي. وأخيراً، تُضاف طاقة متأصلة تعتمد على التردد إلى كل مكون من مكونات درجة الصوت مراعاة العتبة المطلقة للسمع. ويؤدي تحويل الطاقة إلى ديسيبيل إلى تمثيل الغشاء القاعدي للإشارة.

عند محاكاة تجارب التقنيع السمعي، يتم تشكيل تمثيل غشائي قاعدي لكل محفز، والفرق بين التمثيلات هو المعلومات المتاحة لأداء المهمة. وأحد التمثيلين هو للمقنع وحده، والآخر هو للمقنع وإشارة الاختبار مجتمعين. يمثل الفرق بينهما مكون الإشارة غير المقنع. ويحسب المخطط PERCEVAL احتمال كشف هذا الفرق. ويُقدر احتمال عدم كشف الفرق لكل كاشف على طول الغشاء القاعدي المحاكى باستخدام دالة الاحتمال السينية. ومع افتراض أن الكاشفات مستقلة إحصائياً، يُحسب احتمال الكشف العام لمجموعة الكاشفات بأكملها باعتبارها مكملاً لنتائج احتمالات عدم الكشف الفردية. وتمت محاكاة العديد من تجارب التقنيع بنجاح باستخدام هذا النهج، وتم استخدام النموذج لتقييم جدوى نمذجة المستمعين الفرديين [Treurniet، 1996].

وكأداة لتقدير جودة الصوت، يحسب المخطط PERCEVAL الفرق بين تمثيلي الإشارة المرجعية والإشارة قيد الاختبار. ومن خلال تطبيق افتراضات معقولة حول العمليات الإدراكية والمعرفية ذات المستوى الأعلى، يتم حساب عدد من المتغيرات ذات الصلة بالإدراك الحسي ومقابلتها لقياس الجودة الموضوعية للإشارة قيد الاختبار. وتم استمثال التقابل من خلال تدنية الفرق بين توزيع الجودة الموضوعية والتوزيع المقابل لمتوسط تقييمات الجودة الشخصية لمجموعة البيانات المتاحة.

6.4 القياس الموضوعي الإدراكي

الغرض من القياس الموضوعي الإدراكي (POM) [Colomes وآخرون، 1995] هو تحديد مقدار معين من التدهور الذي قد يحدث بين الإشارة المرجعية ونسختها "المتدهورة". ويتحقق ذلك من خلال مقارنة التمثيل القاعدي الداخلي لكلتا الإشارتين، مهما كان التدهور الناتج عنه. وينمذج التمثيل القاعدي العمليات المختلفة التي تخضع لها الإشارة الصوتية عند انتقالها عبر الأذن البشرية. لذلك، فإن المرحلة الأولى من المخطط POM هي حساب التمثيل الداخلي للإشارة الصوتية. وتم اختيار نمط الإثارة (المعطى بالديسيبيل)، المنتشر عبر الغشاء القاعدي، لنمذجة معدل الإطلاق في الخلايا العصبية على طول الغشاء القاعدي.

تسمى عملية حساب نمط الإثارة بالأذن الاصطناعية. وبعد ذلك، بمجرد وجود تمثيلين داخليين للإشارتين المراد مقارنتهما ببعضهما البعض، يجب على المخطط POM أن يشير إلى ما إذا كان الفرق بين تمثيلهما الداخلي مسموعاً أم لا، وإذا كان الأمر كذلك فبأي طريقة. وهذا ما يسمى عملية الكشف.

يستخدم المخطط POM محول DFT بنافذة هان قدرها 40 ms تقريباً (مع تراكب بنسبة 50% بين نافذتي هان). وعدد القنوات القاعدية للتحليل هو 620. الأجزاء المتبقية من النموذج السمعي تكاد تكون مطابقة لتلك المستخدمة في كل من المخططين PAQM وPERCEVAL.

ويتم وصف دالة الانتشار بدقة تامة من خلال تقريب أكثر دقة يأخذ في الاعتبار كلا من تبعية المستوى وفقاً لمؤلف [Terhardt، 1979] والشكل المستدير وفقاً لمؤلف [Schroeder وآخرون، 1979].

وخرج هذا النموذج هو احتمال كشف التشوه بين الإشارتين المقارنتين، بالإضافة إلى ما يسمى بالمسافة القاعدية التي تمثل الفجوة الإدراكية بين الإشارتين المقارنتين.

7.4 نهج مجموعة الأدوات

يستخدم مخطط مجموعة الأدوات (Toolbox) نهجاً من ثلاث خطوات لقياس المسافة المدركة في جودة الصوت لإشارة اختبار الصوت فيما يتعلق بالإشارة المرجعية الصوتية، مما يعطي مؤشراً لمستوى جودة الصوت الشخصية الإجمالية لإشارة الاختبار. وتعتمد الطريقة على نماذج إدراكية معروفة جيداً تُستخدم لوصف التمثيل الإدراكي للفرق بين الإشارتين الصوتيتين. علاوة على ذلك، فهي تتضمن إجراء ترجيح لجودة الصوت المدركة لإشارة اختبار مجسمة، مع الأخذ في الاعتبار نتائج كل من القناتين اليسرى واليمنى. ولا يلزم وجود ارتباط صارم على أساس كل عينة على حدة للإشارة المرجعية والإشارة الصوتية قيد الاختبار.

تعتمد الوظيفة الرئيسية لمخطط مجموعة الأدوات، الخطوة 1، على حساب جهارة الصوت المحددة، المحسوبة وفقاً لمؤلف [Zwicker و Feldtkeller، 1967]، باستخدام تحويل FFT قدره 2 048 نقطة، محاط بنافذة Hann، تقابل مدة قدرها حوالي 40 ms. وتُزاح النافذة بأكملها بزيادات قدرها 10 ms. بالإضافة إلى ذلك، يتم تطبيق تأثيرات التقنيع الزمني، مثل التقنيع المسبق واللاحق، وفقاً لزويكر (Zwicker). ومن هذه القيم الأساسية للإحساس، تُحسب المعلمات الإدراكية الأخرى، مثل جهارة الصوت المدمجة، و جهارة الصوت المقنعة جزئياً، والحدة، وفقاً لمفهوم [von Bismarck، 1974] و [Aures، 1984]، وكمية الأصداء المسبقة كنتيجة لمرحلة ما قبل المعالجة للخطوات اللاحقة.

وتتضمن الخطوة الثانية من مخطط مجموعة الأدوات إجراءات الترجيح التي تعتمد بشكل أساسي على مقدار الفرق المدرك في جهارة الصوت وتغير جهارة الصوت بمرور الوقت.

وتتضمن الخطوة الثالثة من مخطط مجموعة الأدوات إنشاء مجموعة من القيم المتوسطة لمجموعة الأدوات تعتمد على تحليل إحصائي للقيم التي تم الحصول عليها في الخطوتين 1 و 2. ويتضمن خرج هذا التحليل الإحصائي القيم المتوسطة والقصى وجذر متوسط التربيع وكذلك الانحراف المعياري للقيم المتوسطة. ويستخدم مجموع مرجح لقيم خرج مجموعة الأدوات المتوسطة هذه للضبط النهائي للمسافة الإدراكية بين الإشارة قيد الاختبار والإشارة المرجعية. وإذا لزم الأمر، يمكن مطابقة قيمة الخرج الفردية هذه مع درجة الاختلاف الشخصية، والتي يتم الحصول عليها عادةً في اختبارات الاستماع الشخصية عن طريق ضبط بيانات الخرج بشكل تسلسلي لكل زيادة في الوقت باستخدام إما دالة متعددة الحدود خطية أو ذات ترتيب أعلى.

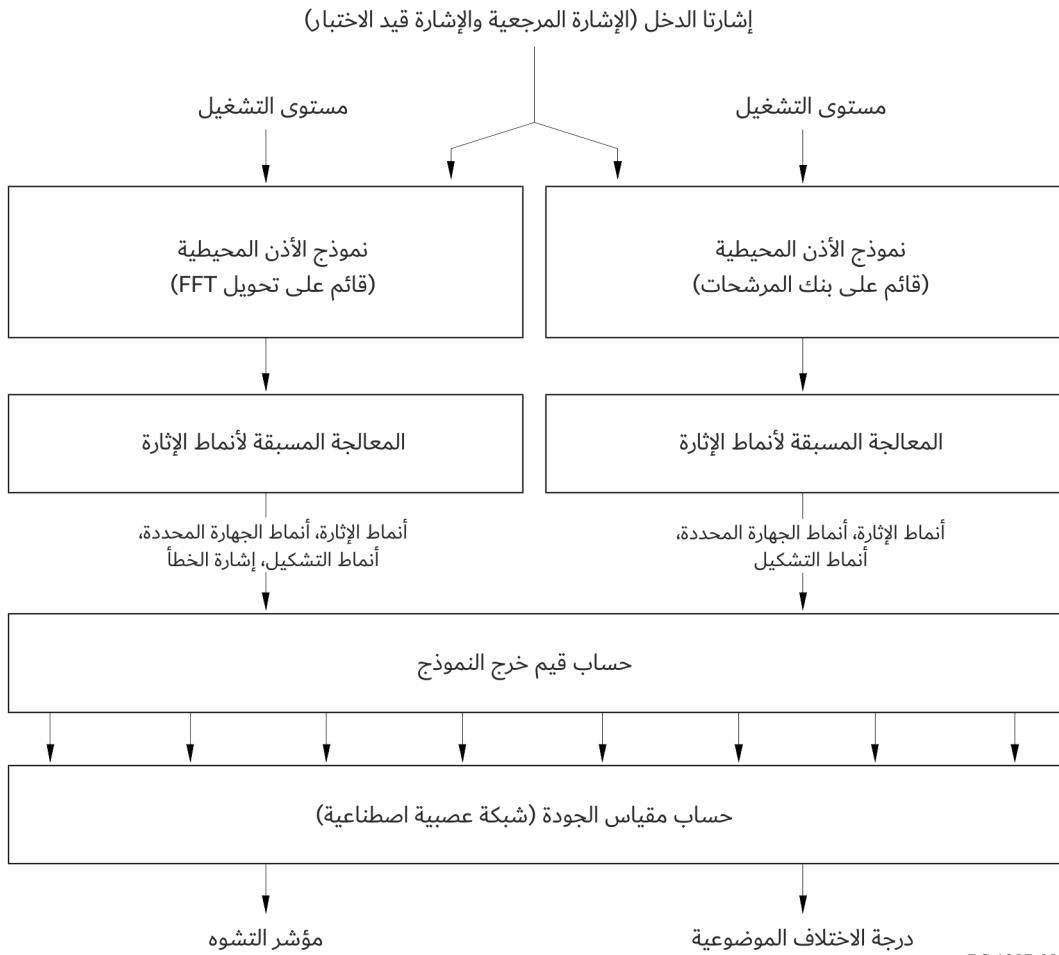
الملحق 2

وصف النموذج²

الخطوط العريضة 1

الشكل 8

مخطط وظيفي عام لنظام القياس



تتكون طريقة القياس الموضوعي لجودة الصوت المدركة المقترحة من نموذج الأذن المحيطية، والعديد من الخطوات المتوسطة (يشار إليها هنا باسم "المعالجة المسبقة لأنماط الإثارة")، وحساب (في الغالب) متغيرات خرج النموذج ("MOV") المستندة إلى الصوت النفسي وإجراء التقابل من مجموعة من متغيرات خرج النموذج إلى قيمة واحدة تمثل جودة الصوت الأساسية للإشارة قيد الاختبار. وتتضمن

² قدم مؤيدو التكنولوجيا الموصوفة في هذه التوصية بيانات براءات اختراع تتوافق مع الملحق 1 بالقرار ITU-R 1. والتكنولوجيا الموصوفة في هذه التوصية محمية ببراءات اختراع دولية، كما هو الحال مع جميع توصيات الاتحاد الدولي للاتصالات الخاضعة لحقوق التأليف والنشر. وتعتبر الموافقة المسبقة للمالكين في شكل ترخيص إلزامية لاستغلال هذه التكنولوجيا. وللحصول على مزيد من المعلومات بشأن ترخيص هذه التكنولوجيا، يرجى الرجوع إلى قاعدة بيانات براءات الاختراع الخاصة بقطاع الاتصالات الراديوية أو إلى أمانة مكتب الاتصالات الراديوية.

نموذجين للأذن المحيطية، أحدهما قائم على التحويل FFT والآخر قائم على بنك المرشحات. وباستثناء حساب إشارة الخطأ (التي تُستخدم فقط مع الجزء القائم على التحويل FFT من نموذج الأذن)، فإن البنية العامة هي نفسها لكلا نموذجي الأذن المحيطية.

وفيما يلي قيم الدخل اللازمة لحساب المتغيرات MOV:

- أنماط الإثارة لكل من إشارة الاختبار والإشارة المرجعية.
- أنماط الإثارة المكيفة طيفياً لكل من إشارة الاختبار والإشارة المرجعية.
- أنماط جهارة الصوت المحددة لكل من إشارة الاختبار والإشارة المرجعية.
- أنماط التشكيل لكل من إشارة الاختبار والإشارة المرجعية.
- تُحسب إشارة الخطأ على أنها الفرق الطيفي بين إشارة الاختبار والإشارة المرجعية (لنموذج الأذن المعتمد على التحويل FFT فقط).

إذا لم يُذكر خلاف ذلك، ففي حالة الإشارات المجسمة، تُجرى جميع الحسابات بشكل مستقل وبنفس الطريقة للقناتين اليسرى واليمنى. ويحدد الوصف إعدادين، أحدهما يسمى "الصيغة الأساسية" والآخر يسمى "الصيغة المتقدمة".

وفي جميع المعادلات المعطاة، المؤشر "Ref." يشير إلى جميع الأنماط المحسوبة من الإشارة المرجعية، والمؤشر "Test" إلى جميع الأنماط المحسوبة من الإشارة قيد الاختبار. ويشير المؤشر "k" إلى متغير التردد المنفصل (أي نطاق التردد) والمؤشر "n" إلى متغير الوقت المنفصل (أي إما عدد الأرتال أو عدد العينات). وإذا لم يتم تعريف قيم k أو n بشكل صريح، فيجب إجراء الحسابات لجميع القيم الممكنة للمؤشرين k و n . يتم شرح جميع الاختصارات الأخرى في مكان ورودها.

وبالنسبة لأسماء متغيرات خرج النموذج، يشير المؤشر "A" إلى جميع المتغيرات المحسوبة باستخدام الجزء القائم على بنك المرشحات من نموذج الأذن ويرمز المؤشر "B" إلى جميع المتغيرات المحسوبة باستخدام الجزء القائم على التحويل FFT من نموذج الأذن.

1.1 الصيغة الأساسية

لا تتضمن الصيغة الأساسية إلا المتغيرات MOV التي يتم حسابها من نموذج الأذن القائم على التحويل FFT. لا يتم استخدام الجزء القائم على بنك المرشحات من النموذج. وتستخدم الصيغة الأساسية إجمالي 11 متغيراً MOV للتنبؤ بجودة الصوت الأساسية المدركة.

2.1 الصيغة المتقدمة

تتضمن الصيغة المتقدمة المتغيرات MOV التي يتم حسابها من نموذج الأذن القائم على بنك المرشحات بالإضافة إلى المتغيرات MOV التي يتم حسابها من نموذج الأذن القائم على التحويل FFT. يتم حساب أنماط الإثارة المكيفة طيفياً وأنماط التشكيل من الجزء القائم على بنك المرشحات من النموذج فقط. وتستخدم الصيغة المتقدمة إجمالي 5 متغيرات MOV للتنبؤ بجودة الصوت الأساسية المدركة.

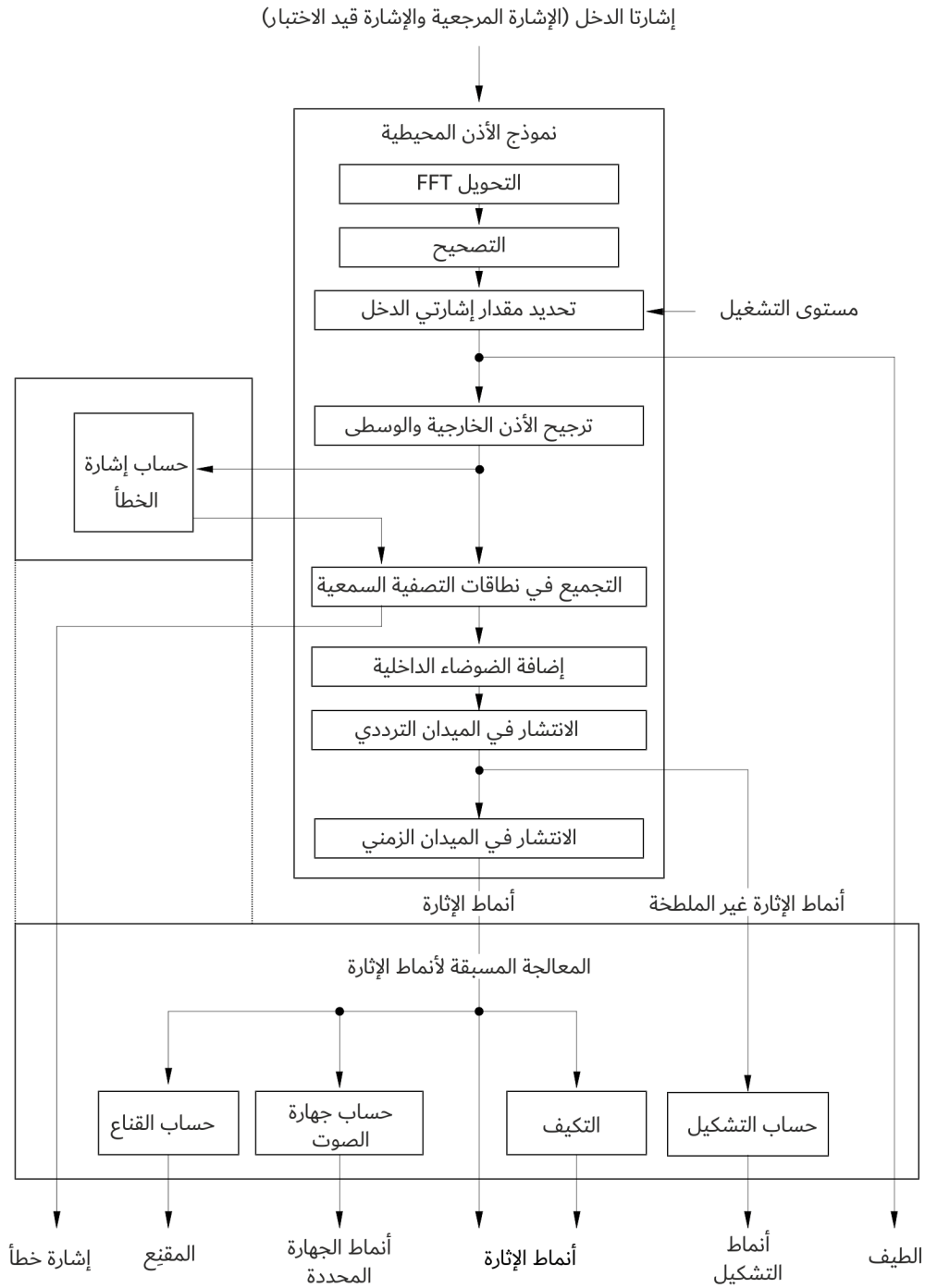
2 نموذج الأذن المحيطية

1.2 نموذج الأذن القائم على التحويل FFT

1.1.2 نظرة عامة

الشكل 9

نموذج الأذن المحيطة والمعالجة المسبقة لأنماط الإثارة لجزء النموذج القائم على التحويل FFT



BS.1387-09

يتم قطع دخل نموذج الأذن القائم على التحويل FFT، الإشارة المرجعية وإشارة الاختبار المتراصفتين زمنياً بعينات 48 kHz، إلى أرتال تبلغ حوالي 0,042 s مع تراكم بنسبة 50%. ويجول كل رتل الميدان الترددي باستخدام نافذة Hann ومحول FFT قصير المدى، ويتم تقييسه إلى مستوى التشغيل. وتُطبق دالة الترجيح على المعاملات الطيفية، والتي تتمدح استجابة تردد الأذن الخارجية والوسطى. ويتم التحويل إلى تمثيل درجة الصوت عن طريق تجميع المعاملات الطيفية المرشحة في نطاقات حرجة. وتتم إضافة إزاحة تعتمد

على التردد لمحاكاة الضوضاء الداخلية في الجهاز السمعي. وتستخدم دالة انتشار قائمة على المستوى لنمذجة المرشحات الصوتية الطيفية في الميدان الترددي. وهي تتبع انتشاراً في الميدان الزمني يراعي تأثيرات التقنيع الأمامي. تُستخدم أنماط الإثارة التي تم الحصول عليها الآن لحساب أنماط جهازة الصوت المحددة وأنماط التقنيع. وتستخدم الأنماط قبل النشر النهائي في الميدان الزمني ("أنماط الإثارة غير المملخة") لحساب أنماط التشكيل. ولنمذجة إشارة الخطأ، يتم دمج أنماط الإشارة المرجعية وإشارة الاختبار لخرج مرشح الأذن الخارجية والوسطى وإجراء تقابل لها مع سلم درجة الصوت من خلال التجميع إلى نطاقات حرجة. وتستخدم هذه المخرجات مع أنماط الإثارة لحساب قيم متغيرات خرج النموذج.

2.1.2 المعالجة الزمنية

يتم قطع دخل نموذج الأذن القائم على التحويل FFT، إشارة الاختبار والإشارة المرجعية، إلى أرتال من 2 048 عينة مع تراكم لعدد 1 024 عينة:

$$(1) \quad t_n[k_t, n] = t[1024 \cdot n + k_t] \quad n=0, 1, 2 \dots k_t = 0..2047$$

حيث n عدد الأرتال الزمنية و k_t العداد الزمني داخل الرتل.

3.1.2 تحويل فورييه السريع (FFT)

يُجرى التقابل من الميدان الزمني للميدان الترددي باستخدام نافذة هان (Hann):

$$(2) \quad h_w[k] = \frac{1}{2} \sqrt{\frac{8}{3}} \left[1 - \cos\left(2\pi \frac{k}{N-1}\right) \right] \quad \left| \quad N = 2048 \right.$$

$$(3) \quad t_w[k_t, n] = h_w[k_t] \cdot t_n[k_t, n]$$

يتبعها تحويل فورييه قصير الأمد:

$$(4) \quad F_f[k_f, n] = \frac{1}{2048} \sum_{k_t=0}^{2047} t_w[k_t, n] e^{-j \frac{2\pi}{2048} k_f k_t}$$

يُحسب عامل المقايسة للتحويل FFT من مستوى ضغط الصوت المفترض، L_p ، لموجة جيبيية كاملة كالتالي:

$$(5) \quad fac = \frac{\frac{L_p}{10^{20}}}{Norm}$$

$$(6) \quad F[k_f, n] = fac \cdot F_f[k_f, n]$$

حيث يتم حساب عامل المقايسة $Norm$ بأخذ موجة جيبيية قدرها 1 019,5 Hz و 0 dB بمقياس كامل كإشارة دخل وحساب القيمة المطلقة القسوى للمعاملات الطيفية على مدى 10 أرتال.

وإذا كان مستوى ضغط الصوت غير معروف، يُوصى بضبط L_p على 92 dB_{SPL}.

4.1.2 الأذن الخارجية والوسطى

تُمدج الاستجابة الترددية للأذن الخارجية والوسطى بدالة ترجيح قائمة على التردد:

$$(7) \quad W[k]/\text{dB} = -0.6 \cdot 3.64 \cdot \left(\frac{f[k]}{\text{kHz}}\right)^{-0.8} + 6.5 \cdot e^{-0.6 \cdot \left(\frac{f[k]}{\text{kHz}} - 3.3\right)^2} - 10^{-3} \cdot \left(\frac{f[k]}{\text{kHz}}\right)^{3.6}$$

حيث:

$$(8) \quad f[k]/\text{Hz} = k \cdot 23.4375$$

التمثيل الترددي عند الخط k المطبق على خرج المحول FFT (المعادلة (9)).

$$(9) \quad F_e[k_f, n] = \left| F[k_f, n] \right| \cdot 10^{\frac{W[k_f]}{20}}$$

$F_e[k_f]$ يُشار إليها على أنها "خرج المحول FFT المرجح للأذن الخارجية".

5.1.2 التجميع في نطاقات حرجة

يُحسب سلم درجة الصوت من تقريب يُعطى بمفهوم [Schroeder وآخرون، 1979].

$$(10) \quad z / \text{Bark} = 7 \cdot \text{arsinh}\left(\frac{f / \text{Hz}}{650}\right)$$

يُطلق على وحدات درجة الصوت اسم $Bark$ (على الرغم من أن هذا السلم لا يمثل بالضبط سلم $Bark$ كما هو محدد بواسطة [Zwicker وFeldtkeller، 1967]).

تتراوح حدود تردد المرشحات من 80 Hz إلى 18 000 Hz. يتوافق عرض نطاقات المرشحات وتباعدها مع دقة قدرها $res = 0,25 \text{ Bark}$ للصيغة الأساسية و $res = 0,5 \text{ Bark}$ للصيغة المتقدمة.

ويؤدي هذا إلى عدد نطاقات التردد $Z = 109$ للصيغة الأساسية و $Z = 55$ للصيغة المتقدمة.

الجدول 6

نطاقات تردد نموذج الأذن القائم على التحويل FFT في الصيغة الأساسية

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	k
23,445	103,445	91,708	80	0
23,577	127,023	115,216	103,445	1
23,739	150,762	138,87	127,023	2
23,932	174,694	162,702	150,762	3
24,155	198,849	186,742	174,694	4
24,408	223,257	211,019	198,849	5
24,693	247,95	235,566	223,257	6

الجدول 6 (تابع)

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	K
25,009	272,959	260,413	247,95	7
25,358	298,317	285,593	272,959	8
25,738	324,055	311,136	298,317	9
26,151	350,207	337,077	324,055	10
26,598	376,805	363,448	350,207	11
27,079	403,884	390,282	376,805	12
27,594	431,478	417,614	403,884	13
28,145	459,622	445,479	431,478	14
28,731	488,353	473,912	459,622	15
29,354	517,707	502,95	488,353	16
30,014	547,721	532,629	517,707	17
30,713	578,434	562,988	547,721	18
31,451	609,885	594,065	578,434	19
32,229	642,114	625,899	609,885	20
33,048	675,161	658,533	642,114	21
33,909	709,071	692,006	675,161	22
34,814	743,884	726,362	709,071	23
35,763	779,647	761,644	743,884	24
36,757	816,404	797,898	779,647	25
37,799	854,203	835,17	816,404	26
38,888	893,091	873,508	854,203	27
40,028	933,119	912,959	893,091	28
41,218	974,336	953,576	933,119	29
42,461	1016,797	995,408	974,336	30
43,758	1060,555	1038,511	1016,797	31
45,111	1105,666	1082,938	1060,555	32
46,521	1152,187	1128,746	1105,666	33
47,991	1200,178	1175,995	1152,187	34
49,522	1249,7	1224,744	1200,178	35

الجدول 6 (تابع)

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	K
51,116	1300,816	1275,055	1249,7	36
52,776	1353,592	1326,992	1300,816	37
54,502	1408,094	1380,623	1353,592	38
56,298	1464,392	1436,014	1408,094	39
58,167	1522,559	1493,237	1464,392	40
60,109	1582,668	1552,366	1522,559	41
62,128	1644,795	1613,474	1582,668	42
64,226	1709,021	1676,641	1644,795	43
66,406	1775,427	1741,946	1709,021	44
68,671	1844,098	1809,474	1775,427	45
71,023	1915,121	1879,31	1844,098	46
73,466	1988,587	1951,543	1915,121	47
76,003	2064,59	2026,266	1988,587	48
78,637	2143,227	2103,573	2064,59	49
81,371	2224,597	2183,564	2143,227	50
84,208	2308,806	2266,34	2224,597	51
87,154	2395,959	2352,008	2308,806	52
90,21	2486,169	2440,675	2395,959	53
93,382	2579,551	2532,456	2486,169	54
96,672	2676,223	2627,468	2579,551	55
100,086	2776,309	2725,832	2676,223	56
103,627	2879,937	2827,672	2776,309	57
107,302	2987,238	2933,12	2879,937	58
111,112	3098,35	3042,309	2987,238	59
115,065	3213,415	3155,379	3098,35	60
119,164	3332,579	3272,475	3213,415	61
123,415	3455,993	3393,745	3332,579	62
127,823	3583,817	3519,344	3455,993	63
132,395	3716,212	3649,432	3583,817	64

الجدول 6 (تابع)

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	K
137,136	3853,348	3784,176	3716,212	65
142,051	3995,399	3923,748	3853,348	66
147,148	4142,547	4068,324	3995,399	67
152,432	4294,979	4218,09	4142,547	68
157,911	4452,89	4373,237	4294,979	69
163,592	4616,482	4533,963	4452,89	70
169,48	4785,962	4700,473	4616,482	71
175,585	4961,548	4872,978	4785,962	72
181,915	5143,463	5051,7	4961,548	73
188,476	5331,939	5236,866	5143,463	74
195,278	5527,217	5428,712	5331,939	75
202,329	5729,545	5627,484	5527,217	76
209,637	5939,183	5833,434	5729,545	77
217,214	6156,396	6046,825	5939,183	78
225,067	6381,463	6267,931	6156,396	79
233,208	6614,671	6497,031	6381,463	80
241,646	6856,316	6734,42	6614,671	81
250,392	7106,708	6980,399	6856,316	82
259,458	7366,166	7235,284	7106,708	83
268,854	7635,02	7499,397	7366,166	84
278,594	7913,614	7773,077	7635,02	85
288,688	8202,302	8056,673	7913,614	86
299,152	8501,454	8350,547	8202,302	87
309,996	8811,45	8655,072	8501,454	88
321,237	9132,688	8970,639	8811,45	89
332,887	9465,574	9297,648	9132,688	90
344,962	9810,536	9636,52	9465,574	91
357,477	10168,013	9987,683	9810,536	92
370,447	10538,46	10351,586	10168,013	93

الجدول 6 (تمة)

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	K
383,891	10922,351	10728,695	10538,46	94
397,824	11320,175	11119,49	10922,351	95
412,264	11732,438	11524,47	11320,175	96
427,231	12159,67	11944,149	11732,438	97
442,742	12602,412	12379,066	12159,67	98
458,817	13061,229	12829,775	12602,412	99
475,48	13536,71	13296,85	13061,229	100
492,748	14029,458	13780,887	13536,71	101
510,645	14540,103	14282,503	14029,458	102
529,192	15069,295	14802,338	14540,103	103
548,415	15617,71	15341,057	15069,295	104
568,339	16186,049	15899,345	15617,71	105
588,986	16775,035	16477,914	16186,049	106
610,385	17385,42	17077,504	16775,035	107
614,58	18000	17690,045	17385,42	108

الجدول 7

نطاقات تردد نموذج الأذن القائم على التحويل FFT في الصيغة المتقدمة

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	k
47,023	127,023	103,445	80	0
47,671	174,694	150,762	127,023	1
48,563	223,257	198,849	174,694	2
49,702	272,959	247,95	223,257	3
51,096	324,055	298,317	272,959	4
52,75	376,805	350,207	324,055	5
54,673	431,478	403,884	376,805	6
56,875	488,353	459,622	431,478	7
59,368	547,721	517,707	488,353	8

الجدول 7 (تابع)

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	k
62,164	609,885	578,434	547,721	9
65,277	675,161	642,114	609,885	10
68,723	743,884	709,071	675,161	11
72,52	816,404	779,647	743,884	12
76,687	893,091	854,203	816,404	13
81,245	974,336	933,119	893,091	14
86,219	1060,555	1016,797	974,336	15
91,632	1152,187	1105,666	1060,555	16
97,513	1249,7	1200,178	1152,187	17
103,892	1353,592	1300,816	1249,7	18
110,801	1464,392	1408,094	1353,592	19
118,275	1582,668	1522,559	1464,392	20
126,354	1709,021	1644,795	1582,668	21
135,077	1844,098	1775,427	1709,021	22
144,489	1988,587	1915,121	1844,098	23
154,64	2143,227	2064,59	1988,587	24
165,579	2308,806	2224,597	2143,227	25
177,364	2486,169	2395,959	2308,806	26
190,054	2676,223	2579,551	2486,169	27
203,713	2879,937	2776,309	2676,223	28
218,414	3098,35	2987,238	2879,937	29
234,229	3332,579	3213,415	3098,35	30
251,238	3583,817	3455,993	3332,579	31
269,531	3853,348	3716,212	3583,817	32
289,199	4142,547	3995,399	3853,348	33
310,343	4452,89	4294,979	4142,547	34
333,072	4785,962	4616,482	4452,89	35
357,5	5143,463	4961,548	4785,962	36
383,754	5527,217	5331,939	5143,463	37

الجدول 7 (تمة)

عرض التردد/Hz	التردد الأعلى/Hz	التردد المركزي/Hz	التردد الأدنى/Hz	المجموعة
$f_w[k]$	$f_u[k]$	$f_c[k]$	$f_l[k]$	k
411,966	5939,183	5729,545	5527,217	38
442,281	6381,463	6156,396	5939,183	39
474,853	6856,316	6614,671	6381,463	40
509,85	7366,166	7106,708	6856,316	41
547,448	7913,614	7635,02	7366,166	42
587,84	8501,454	8202,302	7913,614	43
631,233	9132,688	8811,45	8501,454	44
677,849	9810,536	9465,574	9132,688	45
727,924	10538,46	10168,013	9810,536	46
781,715	11320,175	10922,351	10538,46	47
839,495	12159,67	11732,438	11320,175	48
901,56	13061,229	12602,412	12159,67	49
968,229	14029,458	13536,71	13061,229	50
1039,837	15069,295	14540,103	14029,458	51
1116,754	16186,049	15617,71	15069,295	52
1199,371	17385,42	16775,035	16186,049	53
614,58	18000	17690,045	17385,42	54

يُجرى تقابل التردد إلى درجة الصوت بواسطة الخوارزمية الموضحة في القسم الفرعي التالي، حيث $F_{sp}[k_f]$ هو تمثيل طاقة المخرجات التحويلية FFT المرجحة للأذن الخارجية:

$$(11) \quad F_{sp}[k_f, n] = |F_e[k_f, n]|^2$$

أو تمثيل طاقة إشارة الخطأ:

$$(12) \quad F_{sp}[k_f, n] = |F_{noise}[k_f, n]|^2$$

على التوالي. أنظر الفقرة 4.3 لحساب إشارة الخطأ.

وخرج هذه المرحلة من المعالجة هو طاقات المجموعات الترددية، $P_e[k, n]$.

1.5.1.2 شبه شفيرة

/* inputs */

طاقات الدخل Fsp[]

/* outputs */

Pe[] : طاقات جرى تقابلها مع درجة الصوت

/* intermediate values */

مؤشر للمجموعات الترددية : i
 مؤشر \ لخطط المحول fft : k
 عدد المجموعات الترددية: Z
 109 للصيغة الأساسية
 55 للصيغة المتقدمة
 التردد الأدنى للمجموعة الترددية : fl[]
 التردد الأعلى للمجموعة الترددية : fu[]
 ثابت للدقة الترددية : Fres

```

Fres = 48000/2048;
for(i=0; i<Z; i++ )
{
  Pe[i]=0;
  for(k=0;k<1024;k++)
  {
    /* line inside frequency group */
    if( (( k-0.5)*Fres >= fl[i]) && ((k+0.5)*Fres <= fu[i]))
    {
      Pe[i] += Fsp[k];
    }
    /* frequency group inside*/
    else if( (( k-0.5)*Fres < fl[i]) && ((k+0.5)*Fres > fu[i]))
    {
      Pe[i] += Fsp[k]*(fu[i]-fl[i])/Fres;
    }
    /* left border */
    else if( ((k-0.5)*Fres < fl[i]) && ((k+0.5)*Fres > fl[i]))
    {
      Pe[i] += Fsp[k]*( (k+0.5)*Fres - fl[i])/Fres;
    }
    /* right border
    else if( ((k-0.5)*Fres < fu[i]) && ((k+0.5)*Fres > fu[i]);
  
```

```

{
  Pe[i] += Fsp[k]*(fu[i]- (k-0.5)*Fres)/Fres;
}
/* line outside frequency group */
else
{
  Pe[i] += 0;
}
}

/* limit result */
Pe[i]=max(Pe[i],0.000000000001);
}

```

6.1.2 إضافة الضوضاء الداخلية

تُضاف إزاحة قائمة على التردد، P_{Thres} ، للطاقات في كل مجموعة ترددية:

$$P_{Thres}[k] = 10^{0.4 \cdot 0.364 \cdot \left(\frac{f_c[k]}{\text{kHz}}\right)^{-0.8}} \quad (13)$$

$$P_p[k, n] = P_e[k, n] + P_{Thres}[k] \quad (14)$$

وخرج هذه المرحلة من المعالجة، $P_p[k, n]$ ، يُشار إليه باسم "أنماط درجة الصوت".

7.1.2 الانتشار

يتم تليخ أنماط درجة الصوت $P_p[k, n]$ عبر التردد باستخدام دالة انتشار تعتمد على المستوى. ودالة الانتشار هي دالة أسية ذات وجهين. ويبلغ الميل الأدنى دائماً 27 dB/Bark ويعتمد الميل الأعلى على التردد والطاقة.

وتُحسب قيم الميل وفقاً لما يلي:

$$\frac{S_u[k, L[k, n]]}{\text{dB/Bark}} = -24 - \frac{230 \text{ Hz}}{f_c[k]} + 0.2 \cdot L[k, n] / \text{dB} \quad (15)$$

$$S_l[k, L[k, n]] = 27 \frac{\text{dB}}{\text{Bark}} \quad (16)$$

مع:

$$L[k, n] = 10 \cdot \log_{10}(P_p[k, n])$$

ويُنفذ الانتشار بشكل مستقل لكل مجموعة من مجموعات الترددات k :

$$(17) \quad E_2[k, n] = \frac{1}{\text{Norm}_{SP}[k]} \left(\sum_{j=0}^{Z-1} E_{line}[j, k, n] \right)^{0.4} \frac{1}{0.4}$$

حيث E_{line} تُشتق من:

$$(18) \quad E_{line}[j, k, n] = \begin{cases} \frac{\frac{L[j, n]}{10} \frac{-res \cdot (j-k) \cdot s_l[j, L[j, n]]}{10}}{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot s_l[j, L[j, n]]}{10} + \sum_{\mu=j}^{Z-1} \frac{res \cdot (\mu-j) \cdot s_u[j, L[j, n]]}{10}} & \text{if } k < j \\ \frac{\frac{L[j, n]}{10} \frac{res \cdot (k-j) \cdot s_u[j, L[j, n]]}{10}}{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot s_l[j, L[j, n]]}{10} + \sum_{\mu=j}^{Z-1} \frac{res \cdot (\mu-j) \cdot s_u[j, L[j, n]]}{10}} & \text{if } k \geq j \end{cases}$$

$\text{Norm}_{SP}[k]$ تُحسب وفقاً لما يلي:

$$(19) \quad \text{Norm}_{SP}[k] = \left(\sum_{j=0}^{Z-1} \tilde{E}_{line}[j, k] \right)^{0.4} \frac{1}{0.4}$$

مع:

$$(20) \quad \tilde{E}_{line}[j, k] = \begin{cases} \frac{\frac{-res \cdot (j-k) \cdot s_l[j, 0]}{10}}{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot s_l[j, 0]}{10} + \sum_{\mu=j}^{Z-1} \frac{res \cdot (\mu-j) \cdot s_u[j, 0]}{10}} & \text{if } k < j \\ \frac{\frac{res \cdot (k-j) \cdot s_u[j, 0]}{10}}{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot s_l[j, 0]}{10} + \sum_{\mu=j}^{Z-1} \frac{res \cdot (\mu-j) \cdot s_u[j, 0]}{10}} & \text{if } k \geq j \end{cases}$$

و res هي دقة سلم درجة الصوت بوحدات Bark (0,25 للصيغة الأساسية و0,5 للصيغة المتقدمة).

وتُستخدم الأنماط في هذه المرحلة من المعالجة، $E_2[k, n]$ ، لاحقاً لحساب أنماط التشكيل ويشار إليها باسم "أنماط الإثارة غير الملطخة".

8.1.2 الانتشار في الميدان الزمني

من أجل نمذجة التقنيع الأمامي، يتم تلطيخ الطاقات الموجودة في كل مجموعة تردد بمرور الوقت بواسطة مرشحات التمرير المنخفض من الدرجة الأولى. وتعتمد الثوابت الزمنية على التردد المركزي لكل مجموعة (كما هو مُعطى في المعادلة (10) والجدول 6) ويتم حسابها وفقاً لما يلي:

$$(21) \quad \tau = \tau_{min} + \frac{100 \text{ Hz}}{f_c[k]} \cdot (\tau_{100} - \tau_{min}) \left| \begin{array}{l} \tau_{100} = 0.030 \text{ s} \\ \tau_{min} = 0.008 \text{ s} \end{array} \right.$$

وتُحسب مرشحات التمرير المنخفض من الدرجة الأولى وفقاً لما يلي:

$$(22) \quad E_f[k, n] = a \cdot E_f[k, n-1] + (1-a) \cdot E_2[k, n]$$

$$(23) \quad E[k, n] = \max(E_f(k, n), E_2(k, n))$$

حيث يُحسب a من الثوابت الزمنية أعلاه بالمعادلة:

$$(24) \quad a = e^{-\frac{4}{187.5} \cdot \frac{1}{\tau}}$$

n رقم الرتل الفعلي، و k مؤشر المجموعة و $E_f[k, 0] = 0$.

ويُشار إلى الأنماط في هذه المرحلة من المعالجة، $E[k, n]$ ، باسم أنماط الإثارة".

9.1.2 عتبة التقنيع

تصف عملية التقنيع التأثير الذي تصبح به الإشارة الخافتة ولكن المسموعة بوضوح غير مسموعة عند حدوث إشارة أعلى جهازة. وتُحسب هذه العتبة عن طريق ترجيح أنماط الإثارة بدالة الترجيح $m[k]$.

$$(25) \quad m[k] = \begin{cases} 3.0 & \text{for } k \cdot \text{res} \leq 12 \\ 0.25 \cdot k \cdot \text{res} & \text{for } k \cdot \text{res} > 12 \end{cases}$$

$$(26) \quad M[k, n] = \frac{E[k, n]}{10^{10} \cdot m[k]}$$

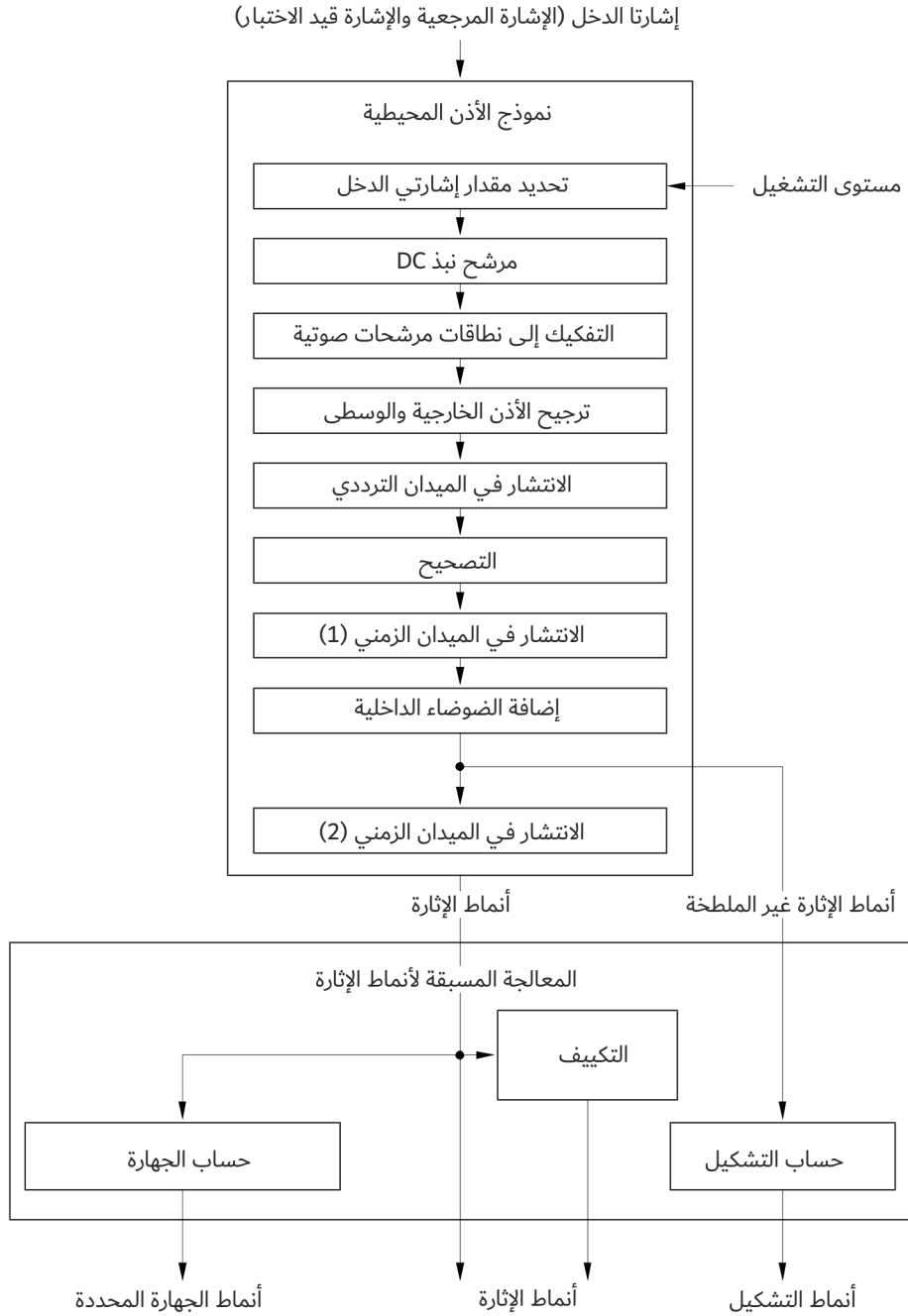
ويُشار إلى الأنماط في هذه المرحلة من المعالجة، $M[k, n]$ ، باسم أنماط القناع".

2.2 نموذج الأذن القائم على بنك المرشحات

1.2.2 نظرة عامة

الشكل 10

نموذج الأذن المحيطية والمعالجة المسبقة لأنماط الإثارة لجزء النموذج القائم على بنك المرشحات



BS.1387-10

عند دخل نموذج الأذن القائم على بنك المرشحات، يتم ضبط الإشارة قيد الاختبار والإشارة المرجعية على مستوى التشغيل المفترض وتُرسل عبر مرشح تمرير عالٍ من أجل إزالة المكونات DC والمكونات دون الصوتية للإشارات. ويتم بعد ذلك تفكيك الإشارات إلى إشارات تمرير النطاق بواسطة مرشحات الطور الخطي التي يتم توزيعها بالتساوي على سلم درجة الصوت الإدراكي. ويُطبق

ترجيح يعتمد على التردد على إشارات تمرير النطاق من أجل نمذجة الخصائص الطيفية للأذن الخارجية والوسطى. وتُمدج الدقة الطيفية المعتمدة على المستوى للمرشحات الصوتية من خلال تحويل الميدان الترددي للمخرجات باستخدام دالة انتشار تعتمد على المستوى.

تُحسب أغلفة الإشارات باستخدام تحويل هيلبرت لإشارات تمرير النطاق ("التصحيح") ويتم تطبيق تحويل الميدان الزمني مع وظيفة نافذة من أجل نمذجة التقنيع العكسي. وبعد ذلك، تُضاف إزاحة تعتمد على التردد تأخذ في الاعتبار الضوضاء الداخلية في الجهاز السمعي وتندمج العتبة في وضع اللانشاط. وأخيراً، تُنفذ عملية تحويل ثانية للميدان الزمني باستخدام دالة الانتشار الأسية التي تراعي التقنيع الأمامي.

تُستخدم أنماط الإثارة التي تم الحصول عليها الآن لحساب أنماط الجهازة المحددة، وتُستخدم الأنماط قبل انتشار الميدان الزمني النهائي ("أنماط الإثارة غير المملخة") لحساب أنماط التشكيل. وهذه، بالإضافة إلى أنماط الإثارة نفسها، هي الأساس الذي يتم على أساسه حساب قيم النموذج. ومن أجل فصل تأثير استجابة تردد الحالة الثابتة للجهاز قيد الاختبار عن التشوهات الأخرى، يتم أيضاً تكيف أنماط الإثارة للإشارة قيد الاختبار والإشارة المرجعية طيفياً مع بعضها البعض ("التكيف"). ويتم حساب أنماط التشكيل وأنماط الجهازة المحددة من أنماط الإثارة المكيفة وغير المكيفة.

2.2.2 الاعتيان الفرعي

عند خرج بنك المرشحات، يُخفّض اعتيان الإشارات بعامل 32 وبعد انتشار الميدان الزمني الأول، يُخفّض اعتيان الإشارات بعامل 6 (انظر الشكل 11).

3.2.2 ضبط مستوى التشغيل

يُحسب عامل المقايسة للدخل من مستوى التشغيل المفترض لإشارة دخل كاملة النطاق بواسطة:

$$fac = \frac{10^{L_{max}/20}}{32767} \quad (27)$$

في حالة معرفة مستوى التشغيل الدقيق، يوصى بضبط L_{max} على 92 dB SPL.

4.2.2 مرشح نبذ DC

نظراً لأن بنك المرشحات حساس للإشارات دون الصوتية في إشارات الدخل، يتم تطبيق مرشح نبذ DC على إشارات الدخل. ويستخدم مرشح التمير العالي من الدرجة الرابعة لبتوروث بتردد قطع قدره 20 Hz. ويُنفذ المرشح كتشلشل من مرشحين IIR من الدرجة الثانية.

$$y_n = x_n - 2x_{n-1} + x_{n-2} + b_1 y_{n-1} + b_2 y_{n-2} \quad (28)$$

حيث معاملات المجموعة الأولى هي:

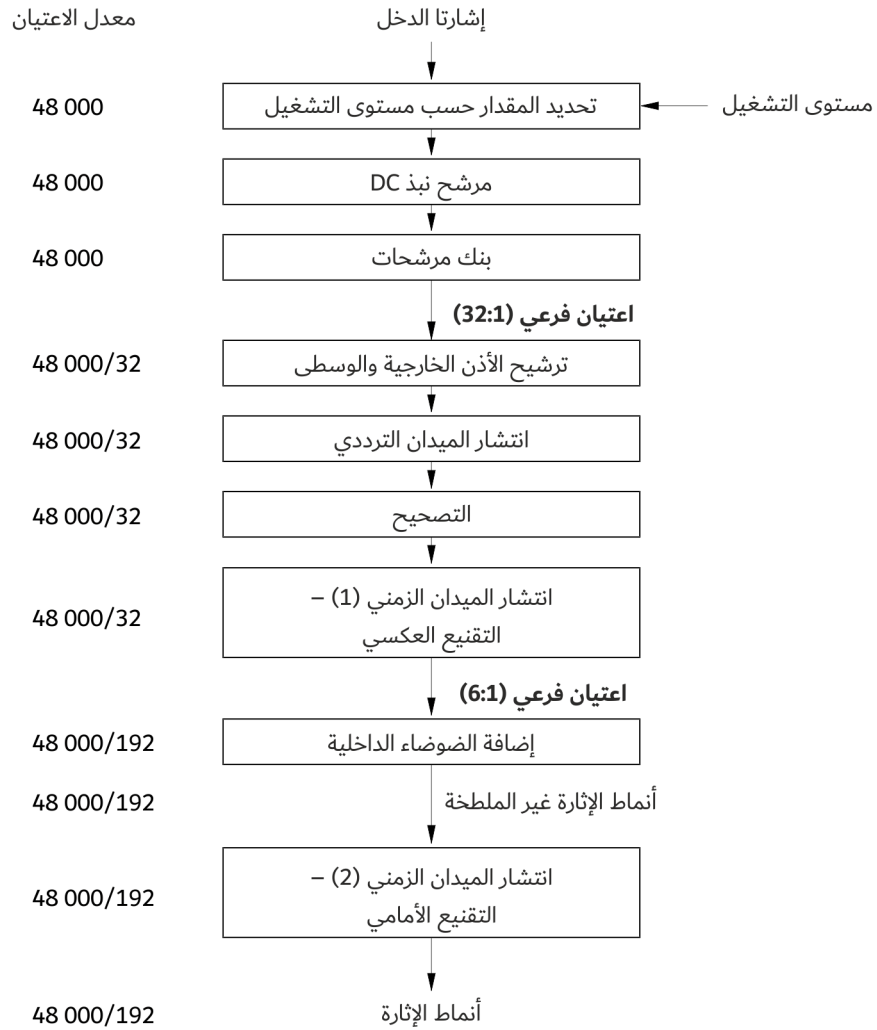
$$b_{1,2} = 1.99517, -0.995174$$

ومعاملات المجموعة الثانية هي:

$$b_{1,2} = 1.99799, -0.997998$$

الشكل 11

الاعتيان الفرعي في نموذج الأذن المحيطية القائم على بنك المرشحات



BS.1387-11

5.2.2 بنك المرشحات

يتألف بنك المرشحات من 40 زوجاً من المرشحات لكل قناة من قنوات الإشارة قيد الاختبار والإشارة المرجعية. والمرشحات متباعدة بشكل متساوٍ ولها عرض نطاق مطلق ثابت عندما يتعلق الأمر بسلم درجة الصوت السمعي. ويتكون كل زوج من المرشحات من مرشحين لهما استجابة تردد متساوية ولكن بفارق 90 درجة في استجابة الطور. وبالتالي، فإن خرج المرشح الثاني يمثل تحويل هيلبرت لخرج المرشح الأول (أو الجزء التخيلي، إذا افترض أن المرشح الأول يمثل الجزء الحقيقي من إشارة مركبة). وأغلقة استجاباتها النبضية لها شكل \cos^2 . وتحدد المرشحات في الجدول 8 والمعادلة (29) (حيث k هو مؤشر المرشح، و n هو مؤشر العينة الزمنية و T هو الوقت بين عينتين: $T = 1/48\ 000$). ويمكن تحقيقها كمرشحات FIR باستخدام القيم $h_{re}(k, n)$ و $h_{im}(k, n)$ كمعاملات. وعندما تكون إشارات الدخل محدودة الوقت، يمكن أيضاً حساب مخرجات المرشحات بواسطة خوارزميات متكررة سريعة جداً.

$$(29) \quad \left. \begin{aligned} h_{re}(k, n) &= \frac{4}{N[k]} \cdot \sin^2\left(\pi \cdot \frac{n}{N[k]}\right) \cdot \cos\left(2\pi \cdot f_c[k] \cdot \left(n - \frac{N[k]}{2}\right) \cdot T\right) \\ h_{im}(k, n) &= \frac{4}{N[k]} \cdot \sin^2\left(\pi \cdot \frac{n}{N[k]}\right) \cdot \sin\left(2\pi \cdot f_c[k] \cdot \left(n - \frac{N[k]}{2}\right) \cdot T\right) \end{aligned} \right| \begin{aligned} &0 \leq n < N[k] \\ &n < 0 \\ &n \geq N[k] \end{aligned}$$

$$h_{re}(k, n) = h_{im}(k, n) = 0$$

الجدول 8

التردد المركزي وطول الاستجابة النبضية والتأخير الإضافي لكل مرشح

التأخير الإضافي/العينات	طول الاستجابة النبضية/العينات	التردد المركزي/Hz	مؤشر المرشح
$D[k]$	$N[k]$	$f_c[k]$	k
1	1456	50,00	0
10	1438	116,19	1
26	1406	183,57	2
48	1362	252,82	3
75	1308	324,64	4
107	1244	399,79	5
141	1176	479,01	6
177	1104	563,11	7
214	1030	652,97	8
251	956	749,48	9
287	884	853,65	10
322	814	966,52	11
355	748	1089,25	12
386	686	1223,10	13
416	626	1369,43	14
444	570	1529,73	15
469	520	1705,64	16
493	472	1898,95	17
514	430	2111,64	18
534	390	2345,88	19
552	354	2604,05	20
569	320	2888,79	21

الجدول 8 (تمة)

مؤشر المرشح	التردد المركزي/Hz	طول الاستجابة النبضية/العينات	التأخير الإضافي/العينات
k	$f_c[k]$	$N[k]$	$D[k]$
22	3203,01	290	584
23	3549,90	262	598
24	3933,02	238	610
25	4356,27	214	622
26	4823,97	194	632
27	5340,88	176	641
28	5912,30	158	650
29	6544,03	144	657
30	7242,54	130	664
31	8014,95	118	670
32	8869,13	106	676
33	9813,82	96	681
34	10858,63	86	686
35	12014,24	78	690
36	13292,44	70	694
37	14706,26	64	697
38	16270,13	58	700
39	18000,02	52	703

تتراوح الترددات المركزية من 50 Hz إلى 18 000 Hz. ويُحسب سلم درجة الصوت السمعي من تقريب مقدم من [Schroeder وآخرون، 1979]:

$$(30) \quad z / \text{Bark} = 7 \cdot \text{arsinh} \left(\frac{f / \text{Hz}}{650} \right)$$

يطلق على وحدات درجة الصوت اسم Bark (على الرغم من أن هذا السلم لا يمثل بالضبط سلم Bark كما حدده [Zwicker وFeldtkeller، 1967]).

من أجل الحصول على تأخيرات متساوية لجميع المرشحات، يتم تأخير دخل كل مرشح بعدد D من العينات، حيث D هو نصف الفرق بين طول الاستجابة النبضية للمرشح وطول الاستجابة النبضية الأطول³:

$$(31) \quad D[k] = 1 + \frac{1}{2} (N[0] - N[k])$$

ويُجرى اعتيان فرعي لخرج المرشحات بعامل مقداره 32، أي أن قيم الخرج تُحسب عند عينة الدخل رقم 32 لجميع المرشحات⁴.

³ التأخير الإضافي لعينة واحدة ليس ضرورياً للتنفيذ. ويلاحظ أن التنفيذ المرجعي المستخدم لاختبار الامتثال يتضمن هذا التأخير الإضافي.

⁴ في الواقع، فإن أغلفة المرشحات في نطاقات التردد العليا لا تحقق بالضرورة نظرية الاعتيان. على الرغم من أن الاستداف لن يحدث إلا في ظل ظروف خاصة جداً (أي مكونات عالية التردد مشكلة بترددات أكبر من 1,5 kHz) وأن المشاكل المتعلقة بهذه التأثيرات لم تحدث أبداً داخل قواعد البيانات المعروفة، إلا أنه ينبغي الإشارة إلى أن مشاكل الاستداف قد تحدث، خاصة عند استخدام إشارات اختبار اصطناعية.

6.2.2 ترشيح الأذن الخارجية والوسطى

تُمدج الاستجابة الترددية للأذن الخارجية والوسطى بواسطة دالة ترجيح تعتمد على التردد تُطبق على خرج المرشحات (المعادلة (32)).

$$(32) \quad W[k]/\text{dB} = -0.6 \cdot 3.64 \cdot \left(\frac{f_c[k]}{\text{kHz}}\right)^{-0.8} + 6.5 \cdot e^{-0.6 \cdot \left(\frac{f_c[k]}{\text{kHz}} - 3.3\right)^2} - 10^{-3} \cdot \left(\frac{f_c[k]}{\text{kHz}}\right)^{3.6}$$

1.6.2.2 شبه شفرة

```

/* inputs */
: out_re,out_im
: W
/* outputs */
: out_re,out_im
/* intermediate values */
: k
: Wt

```

خرج بنك المرشحات
دالة ترجيح (أنظر المعادلة (32))
خرج بنك المرشحات
مؤشر المرشح
عامل الترجيح

```

/* outer and middle ear filtering */
for(k=0..39)
{
    Wt = pow(10,W[k]/20)
    out_re[k] *= Wt;
    out_im[k] *= Wt;
}

```

7.2.2 انتشار الميدان الترددي

تُطخ قيم خرج بنك المرشحات عبر التردد باستخدام دالة انتشار قائمة على المستوى. ودالة الانتشار هي دالة أسية ذات وجهين. يكون الميل الأدنى دائماً 31 dB/Bark ويتراوح الميل الأعلى بين 24 و-4 dB/Bark.

ويُحسب الميل الأعلى، $s[k]$ ، وفقاً لما يلي:

$$(33) \quad \frac{s[k]}{\text{dB/Bark}} = \min\left(-4, -24 - \frac{230 \text{ Hz}}{f_c[k]} + 0.2 \cdot L[k]/\text{dB}\right)$$

يُحسب المستوى $L[k]$ بشكل مستقل لكل قناة مرشح عن طريق أخذ القيمة المطلقة المربعة لخرج المرشح وتحويلها إلى سلم dB. والترددات المركزية، $f_c[k]$ ، مأخوذة من الجدول 8. ويتم تشذيب التمثيلات الخطية للميول مع مرور الوقت بواسطة مرشح تمرير منخفض من الدرجة الأولى له ثابت زمني قدره 100 ms.

ويُنفذ الانتشار بشكل مستقل بالنسبة للمرشحات التي تمثل الجزء الحقيقي من الإشارات والمرشحات التي تمثل الأجزاء التخيلية من الإشارات (المعادلة (29)). ويُنفذ الانتشار الترددي أولاً للميل الأعلى (القائم على المستوى) وبعد ذلك للميل الأدنى باستخدام خوارزمية مرشح IIR من الدرجة الأولى.

شبه شفيرة 1.7.2.2

```

/* inputs */
: out_re,out_im
: z[ ]

/* outputs */
: A_re,A_im
/* intermediate values */
: j,k
: a,b
: dist
: L[ ]
: s[ ]
: d1,d2
/* static */
: cl, cu[ ]

```

خرج بنك المرشحات (الأجزاء الحقيقية والتخيلية)
 معدلات النطاقات الحرجة للترددات المركزية لنطاقات المرشحات بوحدات Bark
 (طبقاً للجدول 8 والمعادلة (30))

أنماط الخرج

مؤشر المرشح

معاملات للتشذيب الزمني

مقدار ثابت لحساب اللغط

المستوى عند خرج كل مرشح

الميل المحلي للانتشار لأعلى

دارثان

(قيم محجوزة من الرتل السابق؛ قيم تُستهل بالصفر عند بدء القياس)
 جزء انتشار الإشارة

```

/* level dependent upward spreading */
dist = pow(0.1,(z[39]-z[0])/(39*20));
/* (z[39]-z[0])/39 is the distance in Bark between two adjacent filter bands */
a = exp(-32/(48000*0.1));
b = 1 - a;
for(k=0..39)
{
  A_re[k] = out_re[k];
  A_im[k] = out_im[k];
}
for(k=0..39)
{
  /* calculate level dependent slope */
  L[k] = 10*log10(out_re[k]* out_re[k] + out_im[k]* out_im[k]);
  s[k] = max(4,(24 + 230/fcentre[k] - 0.2*L[k]));
  /* calculate spread fraction and smooth it over time*/
  cu[k] = a*pow(dist,s[k])+b*cu[k];
  /* spreading of band k */
  d1 = out_re[k]
  d2 = out_im[k]
  for(j=k+1..39)
  {
    d1 *= cu[k];
    d2 *= cu[k];
    A_re[j] += d1;

```

```

        A_im[j] += d2;
    }
}
/* downward spreading */
c1    = pow(dist,31);
d1    = 0;
d2    = 0;
for(k=39..0)
{
    /* spreading of band k */
    d1  = d1 * c1 + A_re[k];
    d2  = d2 * c1 + A_im[k];
    A_re[k] = d1;
    A_im[k] = d2;
}

```

8.2.2 التصحيح

تُحسب الطاقات عند خرج المرشحات عن طريق إضافة القيم التربيعية للمرشح الذي يمثل الجزء الحقيقي من الإشارة وذلك الذي يمثل الجزء التخيلي من الإشارة.

$$(34) \quad E_0[k, n] = A_{re}[k, n]^2 + A_{im}[k, n]^2$$

وتنفذ جميع العمليات التالية على هذه الطاقات.

9.2.2 تلطيح الميدان الزمني (1) - التقنيع العكسي

من أجل نمذجة التقنيع العكسي، تُلطخ الطاقات عند خرج المرشحات بمرور الوقت بواسطة مرشح FIR مع استجابة نبضية على شكل \cos^2 مع 12 تفرعة (والتي تقابل استجابة مرشح مكونة من 384 عينة عند معدل أخذ عينات دخل بنك المرشحات). وبعد التلطيح الزمني، يُخفف اعتيان الخرج بعامل مقداره 6. ويتم ضرب القيم الناتجة بعامل معايرة $cal_1 = 0,9761$ للحصول على مستويات الخرج المناسبة لمستوى التشغيل المحدد.

$$(35) \quad E_1[k, n] = \frac{0.9761}{6} \cdot \sum_{i=0}^{11} E_0[k, 6n-i] \cdot \cos^2\left(\pi \cdot \frac{(i-5)}{12}\right)$$

10.2.2 إضافة الضوضاء الداخلية

بعد الانتشار الأول للميدان الزمني، تتم إضافة إزاحة E_{Thres} قائمة على التردد إلى الطاقات في قناة كل مرشح.

$$(36) \quad E_{Thres}[k] = 10^{0.4 \cdot 0.364 \cdot \left(\frac{f_c[k]}{\text{kHz}}\right)^{-0.8}}$$

$$(37) \quad E_2[k, n] = E_1[k, n] + E_{Thres}[k, n]$$

تُستخدم الأنماط في هذه المرحلة من المعالجة، $E_2[k, n]$ ، لاحقاً لحساب أنماط التشكيل ويشار إليها باسم "أنماط الإثارة غير المملخة".

11.2.2 تلطيخ الميدان الزمني (2) - التقييع الأمامي

من أجل نمذجة التقييع الأمامي، تُلطخ الطاقات الموجودة في قناة كل مرشح بمرور الوقت بواسطة مرشحات التمرير المنخفض من الدرجة الأولى. وتعتمد ثوابت الوقت على التردد المركزي لكل مرشح (كما هو موضح في الجدول 5) ويتم حسابها وفقاً لما يلي:

$$(38) \quad \tau = \tau_{min} + \frac{100 \text{ Hz}}{f_c[k]} \cdot (\tau_{100} - \tau_{min}) \quad \left| \begin{array}{l} \tau_{100} = 0.020 \text{ s} \\ \tau_{min} = 0.004 \text{ s} \end{array} \right.$$

وتُحسب مرشحات التمرير المنخفض من الدرجة الأولى وفقاً لما يلي:

$$(39) \quad E[k, n] = a \cdot E[k, n-1] + (1-a) \cdot E_2[k, n]$$

حيث يُحسب a من الثوابت الزمنية أعلاه بواسطة:

$$(40) \quad a = e^{-\frac{192}{48000 \cdot \tau}}$$

ويُشار إلى الأنماط في هذه المرحلة من المعالجة، $E[k, n]$ ، باسم "أنماط الإثارة".

3 المعالجة المسبقة لأنماط الإثارة

تُستخدم معظم الحسابات الموضحة في هذا القسم مع نموذج الأذن القائم على بنك المرشحات وكذلك مع نموذج الأذن القائم على التحويل FFT. ونظراً لاختلاف عامل الاعتيان الفرعي وعدد نطاقات التردد بين نموذجي الأذن، يتم وصف الثوابت القائمة على هذا العامل باستخدام المتغيرين القائمين على نموذج الأذن $StepSize$ و Z . فبالنسبة لنموذج الأذن القائم على التحويل FFT، تكون قيمة $StepSize$ هي 1024 و Z إما 55 (الصيغة المتقدمة) أو 109 (الصيغة الأساسية). وبالنسبة لنموذج الأذن القائم على بنك المرشحات، تكون قيمة $StepSize$ هي 192 و Z هي 40. وإذا لم يُذكر خلاف ذلك، تتم ضبط كافة المتغيرات والمرشحات التكرارية على الصفر.

1.3 تكييف المستوى والنمط

من أجل التعويض عن فروق المستوى والتشوهات الخطية بين إشارة الاختبار والإشارة المرجعية، يتم تكييف متوسط مستويي إشارة الاختبار والإشارة المرجعية مع بعضهما البعض.

وفي الخطوة الأولى، يتم تشذيب الطاقات الموجودة في قناة كل مرشح بواسطة مرشحات التمرير المنخفض من الدرجة الأولى. وتعتمد الثوابت الزمنية على الترددات المركزية للمرشحات ويتم اختيارها على النحو التالي:

$$(41) \quad \tau = \tau_{min} + \frac{100 \text{ Hz}}{f_c[k]} \cdot (\tau_{100} - \tau_{min}) \quad \left| \begin{array}{l} \tau_{100} = 0.050 \text{ s} \\ \tau_{min} = 0.008 \text{ s} \end{array} \right.$$

وتُحسب مرشحات التمرير المنخفض من الدرجة الأولى وفقاً لما يلي:

$$(42) \quad P_{Ref}[k, n] = a \cdot P_{Ref}[k, n-1] + (1-a) \cdot E_{Ref}[k, n]$$

$$(43) \quad P_{Test}[k, n] = a \cdot P_{Test}[k, n-1] + (1-a) \cdot E_{Test}[k, n]$$

حيث E_{Ref} و E_{Test} هما أنماط الإثارة التي سيتم تكييفها مع بعضها البعض ويتم حساب a من الثوابت الزمنية بواسطة:

$$(44) \quad a = e^{-\frac{StepSize}{48000 \cdot \tau}}$$

1.1.3 تكييف المستوى

من أنماط الإدخال ذات التمرير المنخفض P_{Test} و P_{Ref} ، يُحسب عامل التصحيح اللحظي $LevCorr$ بواسطة:

$$(45) \quad LevCorr[n] = \left(\frac{\sum_{k=0}^{Z-1} \sqrt{P_{Test}[k,n] \cdot P_{Ref}[k,n]}}{\sum_{k=0}^{Z-1} P_{Test}[k,n]} \right)^2$$

فإذا كان عامل التصحيح أكبر من واحد، تتم قسمة الإشارة المرجعية على عامل التصحيح، وإلا يتم ضرب إشارة الاختبار في عامل التصحيح.

$$(46) \quad E_{L,Ref}[k,n] = E_{Ref}[k,n] / LevCorr[n] \quad | \quad LevCorr[n] > 1$$

$$(47) \quad E_{L,Test}[k,n] = E_{Test}[k,n] \cdot LevCorr[n] \quad | \quad LevCorr[n] \leq 1$$

2.1.3 تكييف النمط

تُحسب عوامل التصحيح لكل قناة من خلال مقارنة الأغلفة الزمنية لخرج المرشحات لإشارة الاختبار والإشارة المرجعية:

$$(48) \quad R[k,n] = \frac{\sum_{i=0}^n a[k]^i \cdot E_{L,Test}[k,n-i] \cdot E_{L,Ref}[k,n-i]}{\sum_{i=0}^n a[k]^i \cdot E_{L,Ref}[k,n-i] \cdot E_{L,Ref}[k,n-i]}$$

تُحسب قيم a على النحو الوارد أعلاه (المعادلة (44)) من الثوابت الزمنية المعطاة في المعادلة (41). وإذا كانت $R[k,n]$ أكبر من الواحد، يضبط عامل التصحيح لإشارة الاختبار على $R[k,n]^{-1}$ ويضبط عامل التصحيح للإشارة المرجعية على واحد. وفي الحالة المعاكسة، يضبط عامل التصحيح للإشارة المرجعية على $R[k,n]$ ويضبط عامل التصحيح لإشارة الاختبار على واحد.

$$(49) \quad \begin{array}{l} R_{Test}[k,n] = \frac{1}{R[k,n]}, \quad R_{Ref}[k,n] = 1 \quad | \quad R[k,n] \geq 1 \\ R_{Test}[k,n] = 1, \quad R_{Ref}[k,n] = R[k,n] \quad | \quad R[k,n] < 1 \end{array}$$

إذا كان مقام المعادلة (48) صفراً (وبالتالي ستكون $R[k,n]$ قيمة غير محددة) وكان البسط أكبر من الصفر، فإن $R_{Test}[k,n]$ تُضبط على الصفر وتُضبط $R_{Ref}[k,n]$ على الواحد. وعندما يكون بسط المعادلة (48) صفراً أيضاً، يتم نسخ النسب $R_{Test}[k,n]$ و $R_{Ref}[k,n]$ من نطاق التردد أدناه. وإذا لم يكن هناك نطاق تردد أدناه (أي $k=0$) فإن النسب $R_{Test}[k,n]$ و $R_{Ref}[k,n]$ تُضبط على واحد.

ويُحسب متوسط عوامل التصحيح على عدد M من قنوات المرشحات وتُضبط زمنياً (المعادلة (50)) باستخدام نفس الثوابت الزمنية المذكورة أعلاه (المعادلات من (41) إلى (44)). وعرض نافذة التردد M هو 3 لنموذج الأذن القائم على بنك المرشحات. وبالنسبة لنموذج الأذن القائم على التحويل FFT، فهو 4 (الصيغة المتقدمة) أو 8 (الصيغة الأساسية) على التوالي.

$$(50) \quad \begin{array}{l} PattCorr_{Test}[k,n] = a \cdot PattCorr_{Test}[k,n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Test}[k+i,n] \\ PattCorr_{Ref}[k,n] = a \cdot PattCorr_{Ref}[k,n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Ref}[k+i,n] \end{array}$$

$$\left| \begin{array}{l} M_1 = M_2 = \frac{M-1}{2} \\ M_1 = \frac{M}{2} - 1, \quad M_2 = \frac{M}{2} \end{array} \right. \begin{array}{l} | M \text{ odd} \\ | M \text{ even} \end{array}$$

وعند حدود سلم التردد حيث تتجاوز نافذة التردد مدى نطاقات المرشحات، يُخفّض عرض نافذة التردد وفقاً لذلك:

$$(51) \quad M_1 = \min(M_1, k), \quad M_2 = \min(M_2, z - k - 1), \quad M = M_1 + M_2 + 1$$

وتُرجح أنماط الدخل المكيفة المستوى مع عوامل التصحيح المقابلة $PattCorr_{Test/Ref}[k, n]$ من أجل الحصول على الأنماط المكيفة طيفياً.

$$(52) \quad E_{P,Ref}[k, n] = E_{L,Ref}[k, n] \cdot PattCorr_{Ref}[k, n]$$

$$(53) \quad E_{P,Test}[k, n] = E_{L,Test}[k, n] \cdot PattCorr_{Test}[k, n]$$

2.3 التشكيل

من أنماط الإثارة غير الملطخة، $E_2[k, n]$ ، تُحسب جهازة الصوت المبسطة عن طريق رفع الإثارة إلى أس قدره 0,3. وتُلتخ هذه القيمة والقيمة المطلقة لاشتقاقها الزمني زمنياً.

$$(54) \quad \bar{E}_{der}[k, n] = a \cdot \bar{E}_{der}[k, n-1] + (1-a) \cdot \frac{48000}{StepSize} \cdot \left| E_2[k, n]^{0.3} - E_2[k, n-1]^{0.3} \right|$$

$$(55) \quad \bar{E}[k, n] = a \cdot \bar{E}[k, n-1] + (1-a) \cdot E_2[k, n]^{0.3}$$

وتُحسب قيم a كما في المعادلة (44) من الثوابت الزمنية المعطاة بواسطة:

$$(56) \quad \tau = \tau_0 + \frac{100 \text{ Hz}}{f_c} \cdot (\tau_{100} - \tau_0) \quad \left| \begin{array}{l} \tau_{100} = 0.050 \text{ s} \\ \tau_0 = 0.008 \text{ s} \end{array} \right.$$

ومن القيم الناتجة، \bar{E} و \bar{E}_{der} ، يُحسب مقياس تشكيل الغلاف عند خرج كل مرشح:

$$(57) \quad Mod[k, n] = \frac{\bar{E}_{der}[k, n]}{1 + \bar{E}[k, n]/0.3}$$

وتُستخدم قيم \bar{E} فيما بعد لحساب فرق التشكيل.

3.3 الجهازة

تُحسب أنماط الجهازة المحددة للإشارة قيد الاختبار والإشارة المرجعية طبقاً للمعادلة:

$$(58) \quad N[k, n] = const \cdot \left(\frac{1}{s[k]} \cdot \frac{E_{Thres}[k]}{10^4} \right)^{0.23} \cdot \left[\left(1 - s[k] + \frac{s[k] \cdot E[k, n]}{E_{Thres}[k]} \right)^{0.23} - 1 \right]$$

كما ورد في [1967, Feldtkeller و Zwicker]، تُحسب جهازة الصوت الإجمالية للإشارة قيد الاختبار والإشارة المرجعية كمجموع عبر جميع قنوات المرشحات لجميع قيم جهازة الصوت المحددة فوق الصفر.

$$(59) \quad N_{total}[n] = \frac{24}{Z} \cdot \sum_{k=0}^{Z-1} \max(N[k, n], 0)$$

يتم اختيار ثابت القياس على أنه $const = 1.07664$ لنموذج الأذن المحيطية القائم على التحويل FFT و $const = 1.26539$ لنموذج الأذن المحيطية القائم على بنك المرشحات من أجل إعطاء جهازة صوت إجمالية لصوت واحد لنغمة جيبيية قدرها 40 dB SPL عند 1 kHz. ويُحسب مؤشر العتبة s والإثارة عند العتبة E_{Thres} وفقاً لما يلي:

$$E_{Thres}[k] = 10^{0.364 \cdot \left(\frac{f}{1 \text{ kHz}}\right)^{-0.8}} \quad (60)$$

$$s[k] = 10^{\frac{1}{10} \left(-2 - 2.05 \cdot \text{atm} \left(\frac{f}{4 \text{ kHz}} \right) - 0.75 \cdot \text{atm} \left(\left(\frac{f}{1600 \text{ Hz}} \right)^2 \right) \right)} \quad (61)$$

على التوالي.

ملاحظة - نظراً لاختلاف نماذج الأذن المحيطية، فإن جهازة الصوت المحسوبة هنا لا تتطابق مع جهازة الصوت المحددة في المعيار ISO 532 (الصوتيات - طريقة حساب مستويات جهازة الصوت 1975).

4.3 حساب إشارة الخطأ

تُحسب إشارة الخطأ فقط في النموذج القائم على التحويل FFT. ويُحسب في الميدان الترددي بأخذ الفرق بين أطيايف القدر المرشح في الأذن الخارجية والوسطى للإشارة المرجعية وإشارة الاختبار (انظر الفقرة 4.1.2).

$$F_{noise}[k_f, n] = \left| \left| F_{eref}[k_f, n] \right| - \left| F_{etest}[k_f, n] \right| \right| \quad (62)$$

F_{noise} تُقابل مع ميدان درجة الصوت باستخدام الخوارزمية الموصوفة في الفقرة 5.1.2.

ويُشار إلى خرج هذه الخوارزمية، $P_{noise}[n, k]$ ، باسم "أنماط الضوضاء".

4 حساب متغيرات خرج النموذج

1.4 نظرة عامة

الجدول 9

نظرة عامة على متغيرات خرج النموذج المستخدمة في التنبؤ بجودة الصوت الأساسية

يُستخدم في الصيغة ...		محسوب في نموذج الأذن ...		متغير خرج النموذج (MOV)
المتقدمة	الأساسية	بنك المرشحات	التحويل FFT	
لا	نعم	لا	نعم	WinModDiff1B
لا	نعم	لا	نعم	AvgModDiff1B
لا	نعم	لا	نعم	AvgModDiff2B
نعم	لا	نعم	لا	RmsModDiffA
لا	نعم	لا	نعم	RmsNoiseLoudB
نعم	لا	نعم	لا	RmsNoiseLoudAsymA
نعم	لا	نعم	لا	AvgLinDistA

الجدول 9 (تتمة)

يُستخدم في الصيغة ...		محسوب في نموذج الأذن ...		متغير خرج النموذج (MOV)
المتقدمة	الأساسية	بنك المرشحات	التحويل FFT	
لا	نعم	لا	نعم	BandwidthRefB
لا	نعم	لا	نعم	BandwidthTestB
لا	نعم	لا	نعم	Total NMRB
لا	نعم	لا	نعم	RelDistFramesB
نعم	لا	لا	نعم	Segmental NMRB
لا	نعم	لا	نعم	MFPDB
لا	نعم	لا	نعم	ADB B
نعم	نعم	لا	نعم	EHS B

2.4 فرق التشكيل

تُقاس الفروق في تشكيل الأغلفة الزمنية للإشارة قيد الاختبار والإشارة المرجعية عن طريق حساب فرق التشكيل المحلي لكل قناة مرشح (المعادلة (63))، حيث يتم اشتقاق Mod_{test} و Mod_{Ref} من تطبيق المعادلة (57) على إشارة R_{test} المرجعية.

$$ModDiff[k, n] = w \cdot \frac{|Mod_{test}[k, n] - Mod_{Ref}[k, n]|}{offset + Mod_{Ref}[k, n]}$$

$$(63) \quad \begin{cases} w = 1.0 & |Mod_{test}[k, n] > Mod_{Ref}[k, n] \\ w = negWt & |Mod_{test}[k, n] < Mod_{Ref}[k, n] \end{cases}$$

ويحسب فرق التشكيل اللحظي باعتباره متوسط فروق التشكيل المحلية عبر جميع قنوات المرشحات (المعادلة (64)).

$$(64) \quad ModDiff[n] = \frac{100}{Z} \sum_{k=0}^{Z-1} ModDiff[k, n]$$

تؤخذ عتبة اللانشاط في الاعتبار بواسطة عامل ترجيح قائم على المستوى (المعادلة (65)) محسوب من أنماط الإثارة المعدلة للإشارة المرجعية على النحو الوارد في المعادلة (55) ودالة الضوضاء الداخلية على النحو المحدد في المعادلة (36) لنموذج الأذن القائم على بنك المرشحات والمعادلة (13) لنموذج الأذن القائم على الترشيح FFT.

$$(65) \quad TempWt[n] = \sum_{k=0}^{Z-1} \frac{\bar{E}_{ref}[k, n]}{\bar{E}_{ref}[k, n] + levWt \cdot E_{Thres}[k]} ^{0.3}$$

ويرد وصف التوسيط الزمني لفروق التشكيل اللحظي $ModDiff[n]$ باستخدام عوامل الترجيح $TempWt[n]$ في الفقرة 2.5. وترد في الجدول 10 قيم الثوابت $levWt$ و $offset$ و $negWt$.

الجدول 10

متغيرات خرج النموذج التي تقدر فرق التشكيل الإجمالي

levWt	offset	negWt	MOV (Xxx=Win/Avg/Rms)
100	1	1	$X_{xx}ModDiff1_B$
100	0,01	0,1	$X_{xx}ModDiff2_B$
1	1	1	$X_{xx}ModDiff_A$

RmsModDiff_A 1.2.4

متغير خرج النموذج $RmsModDiff_A$ هو المتوسط المربع لفرق التشكيل المحسوب من نموذج الأذن القائم على بنك المرشحات. انظر الفقرة 2.2.5 للاطلاع على التوسيط الزمني والجدول 10 للاطلاع على الثوابت.

WinModDiff1_B 2.2.4

متغير خرج النموذج $WinModDiff1_B$ هو متوسط مقسم إلى نوافذ لفرق التشكيل المحسوب من نموذج الأذن القائم على التشكيل FFT. انظر الفقرة 3.2.5 للاطلاع على التوسيط الزمني والجدول 10 للاطلاع على الثوابت. ولا يطبق عامل الترجيح الزمني الوارد في المعادلة (65) على هذا المتغير.

AvgModDiff2_B و AvgModDiff1_B 3.2.4

متغيراً خرج النموذج $AvgModDiff1_B$ و $AvgModDiff2_B$ هما المتوسط الخطي لفرق التشكيل المحسوب من نموذج الأذن القائم على التحويل FFT. والفرق بين المتغيرين $AvgModDiff1_B$ و $AvgModDiff2_B$ هو أن الثوابت يتم اختيارها بشكل مختلف. انظر الفقرة 1.2.5 للاطلاع على التوسيط الزمني والجدول 10 للاطلاع على الثوابت.

3.4 جهازة الضوضاء

تقوم متغيرات خرج النموذج هذه بتقدير جهازة الصوت الجزئية للتشوهات الإضافية في وجود الإشارة المرجعية المقنعة. وصُممت معادلة جهازة الصوت الجزئية (المعادلة (66)) للحصول على جهازة الصوت المحددة للضوضاء وفقاً لمؤلف [Zwicker و Feldtkeller، 1967] في حالة عدم وجود مقنن، وإنتاج شيء مثل النسبة بين الضوضاء والقناع إذا كانت الضوضاء صغيرة جداً مقارنة بالمقنن.

وتُحسب جهازة الضوضاء الجزئية وفقاً لما يلي:

$$(66) \quad NL[k, n] = \left(\frac{1}{s_{test}} \cdot \frac{E_{Thres}}{E_0} \right)^{0.23} \cdot \left[\left(1 + \frac{\max(s_{test} \cdot E_{test} - s_{ref} \cdot E_{ref}, 0)}{E_{Thres} + s_{ref} \cdot E_{ref} \cdot \beta} \right)^{0.23} - 1 \right]$$

حيث تكون E_0 دائماً 1 و E_{Thres} هي دالة الضوضاء الداخلية $E_{Thres}[k]$ كما هي محددة في المعادلة (36) وتُحسب s وفقاً لما يلي:

$$(67) \quad s = ThresFac_0 \cdot Mod[k, n] + S_0$$

وما لم توصف بشكل مختلف، تُستخدم أنماط الإثارة المكيفة طيفياً (انظر الفقرة 1.3) كمدخلات: $E_{Test} = E_{P,Test}[k, n]$ و $E_{Ref} = E_{P,Ref}[k, n]$. وتُحسب المعامل β ، الذي يحدد مقدار التقنيع، بواسطة:

$$(68) \quad \beta = \exp\left(-\alpha \cdot \frac{E_{test} - E_{ref}}{E_{ref}}\right)$$

لا تؤخذ قيم جهارة الضوضاء اللحظية في الاعتبار إلا بعد 50 ms من تجاوز جهارة الصوت الإجمالية للقناة السمعية اليسرى أو اليمنى مرة واحدة قيمة $N_T N_{Thres} = 0,1 \text{ sone}$ لكل من إشارة الاختبار والإشارة المرجعية (انظر الفقرة 2.4.2.5). وفي التوسيط الطيفي، تُعابر القيم اللحظية بعدد نطاقات المرشحات لكل نطاق حرج بدلاً من العدد الإجمالي لنطاقات المرشحات، أي يتم ضرب نتيجة التوسيط الطيفي في عامل قيمته 24. وإذا كان جهارة الضوضاء اللحظية أقل من قيمة العتبة N_{Lmin} ، يتم ضبطها على الصفر.

الجدول 11

متغيرات خرج النموذج التي تقدر جهارة الضوضاء الإجمالية

N_{Lmin}	S_0	$ThresFac_0$	α	MOV ($X_{xx}=Win/Avg/Rms$)
0	1	0,15	1,5	$X_{xx}MissingComponents_B$
0	0,5	0,15	1,5	$X_{xx}NoiseLoud_B$
0	1	0,15	1,5	$X_{xx}MissingComponents_A$
0,1	1	0,3	2,5	$X_{xx}NoiseLoud_A$
0	1	0,15	1,5	$X_{xx}LinDist_A$

RmsNoiseLoud_A 1.3.4

متغير خرج النموذج $RmsNoiseLoud_A$ هو المتوسط المربع لجهارة الضوضاء المحسوب من نموذج الأذن القائم على بنك المرشحات. انظر الفقرة 2.2.5 للاطلاع على التوسيط الزمني والجدول 11 للاطلاع على الثوابت.

RmsMissingComponents_A 2.3.4

متغير خرج النموذج $RmsMissingComponents_A$ هو المتوسط المربع لجهارة الضوضاء المحسوب من نموذج الأذن القائم على بنك المرشحات. ويتم حسابه باستخدام أنماط الإثارة المكيفة طيفياً لإشارة الاختبار والإشارة المرجعية المتبادلة من أجل الحصول على جهارة المكونات في الإشارة المرجعية المفقودة في إشارة الاختبار. انظر الفقرة 2.2.5 للاطلاع على التوسيط الزمني والجدول 11 للاطلاع على الثوابت.

RmsNoiseLoudAsym_A 3.3.4

متغير خرج النموذج $RmsNoiseLoudAsym_A$ هو المجموع المرجح للمتوسطات المربعة لجهارة الضوضاء (انظر الفقرة 1.3.4) و جهارة مكونات الإشارة المفقودة (انظر الفقرة 2.3.4)، وكلاهما محسوب من نموذج الأذن القائم على بنك المرشحات.

$$(69) \quad RmsNoiseLoudAsym = RmsNoiseLoud + 0.5 \cdot RmsMissingComponents$$

AvgLinDist_A 4.3.4

يقيس متغير خرج النموذج $AvgLinDist_A$ جهارة مكونات الإشارة المفقودة أثناء التكيف الطيفي للإشارة قيد الاختبار والإشارة المرجعية. ويستخدم الإثارة المكيفة طيفياً للإشارة المرجعية كمرجع والإثارة غير المكيفة للمرجع كإشارة اختبار. ويُحسب هذا المتغير من نموذج الأذن القائم على بنك المرشحات. انظر الفقرة 1.2.5 للاطلاع على التوسيط الزمني والجدول 11 للاطلاع على الثوابت.

RmsNoiseLoudB 5.3.4

متغير خرج النموذج *RmsNoiseLoudB* هو المتوسط المربع لجهازة الضوضاء المحسوب من نموذج الأذن القائم على التحويل FFT. انظر الفقرة 2.2.5 للاطلاع على التوسيط الزمني والجدول 11 للاطلاع على الثوابت.

4.4 عرض النطاق

تقدر قيم خرج النموذج هذه متوسط عرض النطاق للإشارة قيد الاختبار والإشارة المرجعية في خطوط التحويل FFT. وبالنسبة لكل رتل، يُحسب عرض النطاق المحلي $BwRef[n]$ و $BwTest[n]$ وفقاً لشبه الشفرة أدناه.

1.4.4 شبه شفرة

```

/* inputs */
مستوى خرج التحويل FFT بوحدات dB : FLevRef[], FlevelTest[]
/* outputs */
أنماط الخرج : BwRef, BwTest
/* intermediate values */
مؤشر خطوط التحويل FFT : k
عتبة عرض النطاق : ZeroThreshold
    
```

```

ZeroThreshold = FLevelTst(921);
BwRef = BwTst = 0.0;
for(k=921;k<1024;k++)
{
    ZeroThreshold=max(ZeroThreshold,FLevelTst(k));
}

for (k = 920; k>=0; k--)
{
    if (FLevelRef[k] >= 10.0+ZeroThreshold)
    {
        BwRef = k+1;
        break;
    }
}
for (k = BwRef-1; k>=0; k--)
{
    if(FLeveltest[k] >= 5.0+ZeroThreshold)
    {
        BwTest=k+1;
        break;
    }
}
    
```

2.4.4 BandwidthTest_B و BandwidthRef_B

المتغير BandwidthRef_B هو المتوسط الخطي لعرض النطاق BwRef، والمتغير BandwidthTest_B هو المتوسط الخطي لعرض النطاق BwTest. بالنسبة للعملية التوسيط، يتم أخذ الأرتال ذات $BwRef > 346$ فقط بعين الاعتبار. ويتم تجاهل الأرتال ذات الطاقة المنخفضة في بداية ونهاية العناصر (انظر الفقرة 4.4.2.5). انظر الفقرة 1.2.5 للاطلاع على التوسيط الزمني.

5.4 نسبة الضوضاء إلى القناع

تُحسب القيم التالية للنموذج من قيم الضوضاء والتقنيع.

النسبة NMR المحلية للرتل الحالي n تُعرف كالتالي:

$$(70) \quad NMR_{local}[n] = 10 \cdot \log_{10} \frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{noise}[k, n]}{M[k, n]}$$

1.5.4 Total NMR_B

متغير خرج النموذج $Total NMR_B$ هو المتوسط الخطي للنسبة ضوضاء إلى قناع باستخدام

$$(71) \quad NMR_{tot} = 10 \cdot \log_{10} \frac{1}{N} \sum_n \left(\frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{noise}[k, n]}{M[k, n]} \right)$$

ويتم تجاهل الأرتال ذات الطاقة المنخفضة في بداية ونهاية العناصر (انظر الفقرة 4.4.2.5).

2.5.4 Segmental NMR_B

متغير خرج النموذج $Segmental NMR_B$ هو المتوسط الخطي للنسبة ضوضاء إلى قناع المحلية. انظر الفقرة 1.2.5 للاطلاع على التوسيط الزمني.

ويتم تجاهل الأرتال ذات الطاقة المنخفضة في بداية ونهاية العناصر (انظر الفقرة 4.4.2.5).

6.4 Relative Disturbed Frames_B

متغير خرج النموذج $Relative Disturbed Frames_B$ (الاسم المختصر: RelDistFrames_B) يمثل عدد الأرتال ذات:

$$\max_{\forall k} \left(10 \cdot \log \left(\frac{P_{noise}[k, n]}{M[k, n]} \right) \right) \geq 1.5 \text{ dB} \quad k \in [0, Z-1]$$

بالنسبة لعدد الأرتال الإجمالي في العنصر.

ويتم تجاهل الأرتال ذات الطاقة المنخفضة في بداية ونهاية العناصر (انظر الفقرة 4.4.2.5).

7.4 احتمال الكشف

تعتمد المتغيرات MOV المعرفة في هذا القسم على الدالة $\tilde{E}[k, n]$ (النطاق k ، الرتل n)، وهي أنماط الإثارة $E[k, n]$ المعبر عنها بوحدات dB:

$$(72) \quad \tilde{E}[k, n] = 10 \cdot \log_{10}(E[k, n])$$

بالنسبة لكل رتل n :

تُنفذ الخطوات التالية بشكل مستقل لكل قناة c (تكون قيم c يسار ويمين). وأنماط الإثارة اللوغاريتمية هي $\tilde{E}_{ref}[k, n]$ للإشارة المرجعية و $\tilde{E}_{test}[k, n]$ للإشارة قيد الاختبار على التوالي.

ولكل نطاق k :

حساب متوسط الإثارة غير المتماثلة.

$$(73) \quad L[k,n] = 0.3 \cdot \max(\tilde{E}_{ref}[k,n], \tilde{E}_{test}[k,n]) + 0.7 \cdot \tilde{E}_{test}[k,n]$$

حساب حجم خطوة الكشف الفعال s . المعادلة التالية هي تقريب لفرق المستوى الملحوظ فقط كما تم قياسه بواسطة مفهوم [Fastl و Zwicker، 1990].

إذا كانت $L[k,n] > 0$:

$$s[k,n] = 5.95072 \cdot ((6.39468)/L[k,n])^{1.71332} + 9.01033 \times 10^{-11} \cdot L[k,n]^4 + 5.05622 \times 10^{-6} \times L[k,n]^3 - 0.00102438 \cdot L[k,n]^2 + 0.0550197 \cdot L[k,n] - 0.198719$$

وخلاف ذلك

$$(74) \quad s[k,n] = 1.0 \times 10^{30}$$

حساب الخطأ الجبري e :

$$(75) \quad e[k,n] = \tilde{E}_{ref}[k,n] - \tilde{E}_{test}[k,n]$$

إذا كانت $\tilde{E}_{ref}[k,n] > \tilde{E}_{test}[k,n]$ ، يتم ضبط انحدار الميل b على 4,0 وخلاف ذلك يتم ضبطه على 6,0. وهذا يمدج التأثير المتمثل في أن زيادة طاقة الإشارة للإشارة قيد الاختبار مقارنة بالإشارة المرجعية تكون أكثر إثارة للإنتباه من النقصان.

حساب عامل القياس a :

$$(76) \quad a[k,n] = \frac{10^{\frac{\log_{10}(\log_{10}(2.0))}{b}}}{s[k,n]}$$

حساب احتمال الكشف. وتحدد المعادلة (76) عامل القياس a كما لو كانت $e[k,n]$ تساوي $s[k,n]$ ، $p_c[k,n]$ تصبح 0,5:

$$(77) \quad p_c[k,n] = 1 - 10^{(-a[k,n] \cdot e[k,n])^b}$$

حساب العدد الإجمالي للخطوات فوق العتبة:

$$(78) \quad q_c[k,n] = \frac{|\text{INT}(e[k,n])|}{s[k,n]}$$

ويكون احتمال الكشف بكلتا الأذنين كالتالي:

$$(79) \quad p_{bin}[k,n] = \max(p_{left}[k,n], p_{right}[k,n])$$

ويكون عدد الخطوات فوق العتبة لقناة الأذنين كالتالي:

$$(80) \quad q_{bin}[k,n] = \max(q_{left}[k,n], q_{right}[k,n])$$

ويكون الاحتمال الإجمالي لكشف القناة c للرتل n كالتالي:

$$(81) \quad P_c[n] = 1 - \prod_{\forall k} (1 - p_c[k,n])$$

حيث قد تأخذ c القيمة $left$ أو $right$ أو bin . ويكون العدد الإجمالي للخطوات فوق العتبة للقناة c للرتل n كالتالي:

$$(82) \quad Q_c[n] = \sum_{\forall k} q_c[k,n]$$

1.7.4 الحد الأقصى لاحتمال الكشف المصفي (MFPD_B)

تُحسب الصيغة المشدبة لاحتمال الكشف لكل قناة c :

$$(83) \quad \tilde{P}_c[n] = (1 - c_0) \cdot P_c[n] + c_0 \cdot \tilde{P}_c[n - 1]$$

حيث $P_c[-1] = 0$ ويعتمد المقدار الثابت c_0 على حجم الخطوة:

$$(84) \quad c_0 = 0.9 \text{StepSize} / 1024$$

ويقلل الثابت c_0 من الحساسية للتشوهات بالغة القصر.

ويُحسب الحد الأقصى لاحتمال الكشف المصفي (الاسم المختصر: MFPD):

$$(85) \quad PM_c[n] = \max(PM_c[n-1] \cdot c_1, \tilde{P}_c[n])$$

حيث $PM_c[-1]$ تساوي الصفر. ويعتمد المقدار الثابت c_1 على حجم الخطوة:

$$(86) \quad c_1 = 0.99 \text{StepSize} / 1024$$

ينمذج الثابت c_1 التأثير المتمثل في أن التشوهات في بداية مقتطف من الصوت تكون أقل حدة مما تكون عليه في نهاية المقتطف بسبب النسيان. ويُلاحظ أن هذا الثابت مفيد في نمذجة اختبارات الاستماع حيث لا يُسمح للأشخاص القائمين بالاختبار باختيار أجزاء أقصر من المقتطف. وبالنسبة للنموذج الحالي، الذي تتم معيارته باستخدام بيانات من اختبارات الاستماع وفقاً للتوصية ITU-R BS.1116، ينبغي أن تكون قيمة الثابت c_1 تساوي 1,0.

ومتغير خرج النموذج MFPD هو $PM_{bin}[n]$ للرتل الأخير.

2.7.4 متوسط المجموعة المشوهة⁵ (ADB_B)

يُحسب عدد الأرتال الصالحة التي لديها احتمال كشف قناة الأذنين $P_{bin}[n]$ فوق 0,5 ($n_{distorted}$).

وبالنسبة لجميع الأرتال الصالحة، يُحسب العدد الإجمالي للخطوات فوق عتبة قناة الأذنين $Q_{bin}[n]$:

$$Q_{sum} = \sum_{\forall n} Q_{bin}[n]$$

ويُحسب تشوه متوسط المجموعة المشوهة ADB:

- if $n_{distorted}$ is zero then $ADB = 0$ (no distortion audible);
- if $n_{distorted} > 0$ and $Q_{sum} > 0$ then $ADB = \log_{10} ((Q_{sum}) / n_{distorted})$;
- if $n_{distorted} > 0$ and Q_{sum} is zero then $ADB = -0.5$.

8.4 البنية التوافقية للخطأ

الإشارة المرجعية التي تحتوي على توافقيات قوية (مثل الكلارينيت، هاريسكورد) لها طيف يتميز بعدد من القمم المتباعدة بانتظام والتي تفصلها قيعان عميقة. وفي بعض الظروف، قد تترث إشارة الخطأ تلك البنية. فعلى سبيل المثال، من المرجح أن تظل الضوضاء الممزوجة تمثل هذه الإشارة غير مقنعة حيث تكون الإشارة منخفضة في القيعان الطيفية. وسيحتوي طيف الخطأ الناتج بعد ذلك على بنية مشابهة للطيف الأصلي ولكن بإزاحة في التردد ليتوافق مع مواضع القيعان. وقد تؤدي هذه البنية إلى تشويه بالصفات النغمية التي يمكن أن تزيد من بروز الخطأ.

⁵ مصطلح "مجموعة" يعادل "رتل" في هذا السياق.

ويُعرف الخطأ على أنه الفرق في الأطياف اللوغاريتمية للإشارة المرجعية والإشارة المعالجة، كل منها مرجح بالاستجابة الترددية للأذن الخارجية والوسطى (انظر الفقرة 4.1.2، المعادلة (7)). ولا يُستخدم نمط الإثارة من النموذج النفسي الصوتي هنا لأن تحويل التردد غير الخطي إلى Bark من شأنه تلطيخ البنية التوافقية.

1.8.4 EHS_B

يُحصل على مقدار البنية التوافقية من خلال تحديد وقياس أكبر قمة في طيف دالة الترابط الآلي. يتم حساب كل ترابط على أنه جيب تمام الزاوية بين متجهين وفقاً للصيغة التالية، حيث \vec{F}_0 هو متجه الخطأ و \vec{F}_t هو نفس المتجه بتأخير بمقدار معين. وطول الترابط هو نفس الحد الأقصى للتأخر (أي 256 في المثال أدناه).

$$C = \frac{\vec{F}_0 \cdot \vec{F}_t}{|\vec{F}_0| \cdot |\vec{F}_t|} \quad (87)$$

الحد الأقصى للتأخر في الحصول على دالة الترابط الآلي هو القوة الأكبر للرقم 2 ويكون أصغر من نصف رقم مكون تردد التحويل FFT المقابل للتردد 18 kHz.

فعلى سبيل المثال، عند معدل اعتيان قدره 48 kHz وحجم نافذة FFT يبلغ 2 048 عينة، فإن مكون FFT المقابل للتردد 18 kHz هو $1024 \times (18/24) = 768$. وبالتالي، سيكون الحد الأقصى للتأخر 384. وسيكون العدد الفعلي للتأخيرات 256، وهي أكبر قوة للرقم 2 أقل من 384. وسيتم الحصول على القيمة الأولى لدالة الترابط عن طريق تراصف $F_t[0]$ مع $F_0[0]$ والقيمة الأخيرة عن طريق تراصف $F_t[0]$ مع $F_0[255]$.

يوضع متجه الترابطات الناتج في نافذة هان معيارية، وبعد إزالة المكون DC عن طريق طرح القيمة المتوسطة، يتم حساب طيف القدرة باستخدام تحويل FFT. وتُحدد الذروة القصوى في الطيف بعد القاع الأول للتردد السائد في دالة الترابط الآلي. ومتوسط قيمة هذا الحد الأقصى عبر الأرتال مضروباً في 1 000.0 هو متغير البنية التوافقية للخطأ (EHS).

5 التوسيط

1.5 التوسيط الطيفي

ما لم يُذكر خلاف ذلك في أوصاف المتغيرات MOV (انظر الفقرة 4)، تُستخدم الخوارزمية التالية عند توسيط القيم المحلية عبر نطاقات التردد.

1.1.5 المتوسط الخطي

تُحسب قيمة المتوسط الخطي كالتالي:

$$AvgS = \frac{1}{Z} \cdot \sum_{k=0}^{Z-1} S[k] \quad (88)$$

حيث S ترمز إلى اسم متغير خرج النموذج و Z عدد نطاقات التردد.

2.5 التوسيط الزمني

ما لم يُذكر خلاف ذلك في أوصاف المتغيرات MOV (انظر الفقرة 4)، تُستخدم واحدة أو أكثر من الخوارزميات التالية عند توسيط القيم اللحظية مع الزمن. ويُشار إلى عامل الترجيح الزمني (في حالة تطبيقه) بالرمز W ، و Z هو عدد نطاقات التردد.

1.2.5 المتوسط الخطي

تُحسب قيمة المتوسط الخطي (السابقة "Avg") كالتالي:

$$(89) \quad AvgX = \frac{1}{N} \cdot \sum_{n=0}^{N-1} X[n]$$

حيث يشير X إلى اسم المتغير MOV و N هو عدد العينات الزمنية التي تم حساب القيم اللحظية للمتغير X لها. وفي حالة تطبيق الترجيح الزمني (انظر الفقرة 2.4)، يُحسب المتوسط الخطي وفقاً لما يلي:

$$(90) \quad AvgX = \frac{\sum_{n=0}^{N-1} W[n] \cdot X[n]}{\sum_{n=0}^{N-1} W[n]}$$

بدلاً من ذلك.

2.2.5 متوسط التربيع

تُحسب قيمة متوسط التربيع (السابقة "Rms") كالتالي:

$$(91) \quad RmsX = \sqrt{\frac{1}{N} \cdot \sum_{n=0}^{N-1} X[n]^2}$$

حيث يشير X إلى اسم المتغير MOV و N هو عدد العينات الزمنية التي تم حساب القيم اللحظية للمتغير X لها. وعند تطبيق الترجيح الزمني (انظر الفقرة 2.4)، يُحسب متوسط التربيع وفقاً لما يلي:

$$(92) \quad RmsX = \sqrt{Z} \cdot \sqrt{\frac{\sum_{n=0}^{N-1} W[n]^2 \cdot X[n]^2}{\sum_{n=0}^{N-1} W[n]^2}}$$

بدلاً من ذلك.

3.2.5 المتوسط المقسم إلى نوافذ

تُحسب قيمة المتوسط المقسم إلى نوافذ (السابقة "Win") كالتالي:

$$(93) \quad WinX = \sqrt{\frac{1}{N-L+1} \cdot \sum_{n=L-1}^{N-1} \left(\frac{1}{L} \cdot \sum_{i=0}^{L-1} \sqrt{X[n-i]} \right)^4}$$

حيث يشير X إلى اسم المتغير MOV و N هو عدد العينات الزمنية التي تم حساب القيم اللحظية للمتغير X لها و L طول النافذة الزمنية المتدرجة في العينات الزمنية. ويبلغ طول النافذة 100 ms تقريباً، أي أن L تساوي 4 لنموذج الأذن القائم على التحويل FFT و 25 لنموذج الأذن القائم على بنك المرشحات.

4.2.5 اختيار الأرتال

1.4.2.5 التوسيط المتأخر

بالنسبة إلى المتغيرات MOV التي تستخدم هذا المعيار، لا تؤخذ القيم المحسوبة خلال أول 0,5 s من القياس في الاعتبار في التوسيط الزمني. ويُستخدم التوسيط المتأخر لمتغيرات خرج النموذج التالية:

WinModDiff1، AvgModDiff1، AvgModDiff2، RmsNoiseLoudness، RmsNoiseLoudAsym، AvgLinDist، RmsModDiff

2.4.2.5 عتبة الجهازة

بالنسبة إلى المتغيرات MOV التي تستخدم هذا المعيار، فإن جميع القيم اللحظية المحسوبة حتى 50 ms بعد أن تصل الجهازة الإجمالية لإحدى القنوات الصوتية المقابلة إلى قيمة $NThres\ sone$ لكل من إشارة الاختبار والإشارة المرجعية لا تؤخذ في الاعتبار في التوسيط الزمني. ويُستخدم عتبة الجهازة فقط من أجل المتغيرات MOV الموصوفة في الفقرة 3.4.

3.4.2.5 عتبة الطاقة

عندما تكون طاقة النصف الأحداث من الرتل المؤلف من 2048 عينة أقل من 8000*، في القناة الأحادية، أو في كل من القناتين اليسرى واليمنى لكل من البيانات المرجعية وبيانات الاختبار، يتم تجاهل الرتل. وتتسم الأرتال بتراكب بنسبة 50% ويتم تقييم نصف الأرتال التي تحتوي على بيانات جديدة فقط. ويجوز تطبيق هذا المعيار دون معالجة الأرتال ذات الطاقة القليلة جداً. ويستخدم هذا المعيار فقط من أجل المتغير MOV الموصوف في الفقرة 8.4.

4.4.2.5 حدود البيانات

إذا كان الملف المعالج يحتوي على ضوضاء قبل أو بعد بيانات الملف المرجعي الشرعي، فقد يكون الخطأ النسبي كبيراً جداً نظراً لأن المستوى المرجعي يقترب من $-\infty$. وعندما يعتبر هذا الخطأ نقيصة، فقد يتم تجاهله من خلال تطبيق معيار رفض حدود البيانات. وعند فتح الملفات لأول مرة، يتم تحديد مواقع بداية ونهاية البيانات الفعلية في الملف المرجعي. وتُحدد بداية البيانات أو نهايتها على أنها الموقع الأول، الذي يتم المسح فيه من بداية الملف أو نهايته، حيث يتجاوز مجموع القيم المطلقة على مدى خمس عينات متتالية 200، في إحدى القنوات الصوتية المقابلة. ومن ثم تُغفل تجاهل الأرتال التي تقع خارج هذا النطاق بالكامل. ويُستخدم هذا المعيار للمتغيرات MOV.

3.5 التوسيط عبر القنوات الصوتية

ما لم يُذكر خلاف ذلك، في حالة الإشارات المجسمة، يتم توسيط المتغيرات MOV للقناة اليسرى واليمنى خطياً بعد التوسيط الزمني.

6 تقدير جودة الصوت الأساسية المدركة

تُقدر جودة الصوت الأساسية المدركة من خلال إجراء تقابل للعديد من المتغيرات MOV مع رقم واحد باستخدام بنية شبكة عصبية اصطناعية ذات طبقة مخفية واحدة.

1.6 الشبكة العصبية الاصطناعية

وظيفة التنشيط للشبكة العصبية هي دالة سينية غير متماثلة:

* يشير هذا الرقم إلى بيانات الدخل ذات نسق العدد الصحيح الجبري المكون من 16 بته بمدى يتراوح بين -32768 و32767 على النحو المستخدم في القرص المدمج.

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (94)$$

تستخدم الشبكة مدخلات I وعقد J في الطبقة المخفية. ويُحدد التقابل من خلال مجموعة من عوامل قياس المدخلات $a_{min}[i]$ و $a_{max}[i]$ ومجموعة من أوزان المدخلات $w_x[i]$ ومجموعة من أوزان المخرجات $w_y[j]$ وزوج من عوامل قياس المخرجات b_{min} و b_{max} . ويُجرى تقابل للمدخلات مع مؤشر التشوه

$$DI = w_y[J] + \sum_{j=0}^{J-1} \left(w_y[j] \cdot \text{sig} \left(w_x[I, j] + \sum_{i=0}^{I-1} w_x[i, j] \cdot \frac{x[i] - a_{min}[i]}{a_{max}[i] - a_{min}[i]} \right) \right) \quad (95)$$

والذي يرتبط مباشرة بجودة الصوت الأساسية المدركة المقدرة من حيث درجة الاختلاف الموضوعي (ODG). ويُعبر عن العلاقة بين مؤشر التشوه ودرجة الاختلاف الموضوعي من خلال:

$$ODG = b_{min} + (b_{max} - b_{min}) \cdot \text{sig}(DI) \quad (96)$$

2.6 الصيغة الأساسية

تستخدم الصيغة الأساسية نموذج الأذن القائم على التحويل FFT فقط. وتستخدم المتغيرات MOV التالية: $BandwidthRef_B$ و $BandwidthTest_B$ و $Total\ NMR_B$ و $WinModDiff1_B$ و ADB_B و EHS_B و $AvgModDiff1_B$ و $AvgModDiff2_B$ و $RmsNoiseLoud_B$ و $mfpdB$ و $RelDistFrames_B$. ويُجرى تقابل لهذه المتغيرات MOV البالغ عددها 11 متغيراً مع مؤشر جودة وحيد يستخدم شبكة عصبية على النحو الموصوف في الفقرة 1.6 (الشبكة العصبية الاصطناعية) التي تضم ثلاث عقد في الطبقة المخفية. وترد في الجداول من 12 إلى 16 معلمات التقابل.

الجدول 12

المتغيرات MOV المستخدمة في الصيغة الأساسية

الغرض	المتغير MOV
تغيرات في التشكيل (تتعلق بالاضطراب)	WinModDiff1 _B
	AvgModDiff1 _B
	AvgModDiff2 _B
جهازة التشوه	RmsNoiseLoud _B
التشوهات الخطية (استجابة ترددية، إلخ)	BandwidthRef _B
	BandwidthTest _B
تردد التشوهات المسموعة	RelDistFrames _B
النسبة ضوضاء إلى قناع	Total NMR _B
	MFPD _B
احتمال الكشف	ADB _B
	EHS _B
البنية التوافقية للخطأ	

الجدول 13

عوامل القياس لمداخلات الصيغة الأساسية

$a_{\max}[i]$	$a_{\min}[i]$	MOV (x[i])	index (j)
921	393,916656	BandwidthRef _B	0
881,131226	361,965332	BandwidthTest _B	1
16,212030	24,045116-	Total NMR _B	2
107,137772	1,110661	WinModDiff1 _B	3
2,886017	0,206623-	ADB _B	4
13,933351	0,074318	EHS _B	5
63,257874	1,113683	AvgModDiff1 _B	6
1145,018555	0,950345	AvgModDiff2 _B	7
14,819740	0,029985	RmsNoiseLoud _B	8
1	0,000101	MFPD _B	9
1	0	RelDistFrames _B	10

الجدول 14

أوزان عقد الدخل للصيغة الأساسية

node 3 ($w_x[i,2]$)	node 2 ($w_x[i,1]$)	node 1 ($w_x[i,0]$)	MOV (x[i])	index (j)
1,219602	0,436333	0,502657-	BandwidthRef _B	0
1,123743	3,246017	4,307481	BandwidthTest _B	1
0,192096-	2,211189-	4,984241	Total NMR _B	2
4,331315	1,762424-	0,051056	WinModDiff1 _B	3
0,754560-	1,789971	2,321580	ADB _B	4
10,814982-	3,452257-	5,303901-	EHS _B	5
1,519223	6,111805-	2,730991	AvgModDiff1 _B	6
5,955151-	1,331523-	0,624950	AvgModDiff2 _B	7
5,922878-	0,871260	3,102889	RmsNoiseLoud _B	8
0,142913-	0,939882-	1,051468-	MFPD _B	9
0,620456-	0,503610-	1,804679-	RelDistFrames _B	10
2,207228-	0,654841	2,518254-	bias	11

الجدول 15

أوزان عقد الخرج للصيغة الأساسية

bias ($w_y[3]$)	node 3 ($w_y[2]$)	node 2 ($w_y[1]$)	node 1 ($w_y[0]$)
0,307594-	4,629582	4,107138	3,817048-

الجدول 16

عوامل قياس خرج الصيغة الأساسية

b_{max}	b_{min}	
0,22	3,98-	ODG

3.6 الصيغة المتقدمة

تستخدم الصيغة المتقدمة كل من نموذج الأذن القائم على بنك المرشحات ونموذج الأذن القائم على التحويل FFT. وتستخدم المتغيرات MOV: $RmsModDiff_A$ و $RmsNoiseLoudAsym_A$ و $AvgLinDist_A$ و $Segmental\ NMR_B$ و EHS_B . ويُجرى تقابل لهذه المتغيرات MOV الخمسة مع مؤشر جودة وحيد يستخدم شبكة عصبية على النحو الموصوف في الفقرة 1.6 التي تضم خمس عقد في الطبقة المخفية. وترد في الجداول من 17 إلى 21 معلمات التقابل.

الجدول 17

المتغيرات MOV المستخدمة في الصيغة المتقدمة

الغرض	المتغير MOV
جهازة التشوه	$RmsNoiseLoudAsym_A$
تغييرات في التشكيل (تتعلق بالاضطراب)	$RmsModDiff_A$
التشوهات الخطية (الاستجابة الترددية، إلخ)	$AvgLinDist_A$
النسبة ضوضاء إلى قناع	$Segmental\ NMR_B$
البنية التوافقية للخطأ	EHS_B

الجدول 18

عوامل القياس لعقد دخل الصيغة المتقدمة

$a_{max}[i]$	$a_{min}[i]$	MOV ($x[i]$)	index (i)
2166,5	13,298751	$RmsModDiff_A$	0
13,24326	0,041073	$RmsNoiseLoudAsym_A$	1
13,46708	25,018791-	$Segmental\ NMR_B$	2
10,226771	0,061560	EHS_B	3
14,224874	0,024523	$AvgLinDist_A$	4

الجدول 19

أوزان مدخلات الصيغة المتقدمة

node 5 (w ₄ [i,4])	node 4 (w _x [i,3])	node 3 (w _x [i,2])	node 2 (w _x [i,1])	node 1 (w _x [i,0])	MOV (x[i])	index (j)
0,320899-	14,545348-	1,382553-	39,913052-	21,211773	RmsModDiff _A	0
3,238586-	1,686586-	0,935389	19,956049	8,981803-	RmsNoiseLoudAsym _A	1
1,783120-	5,606502	7,442935-	2,877505-	1,633830	Segmental NMR _B	2
0,511314-	1,088213	0,240284-	19,587435	6,103821	EHS _B	3
11,031250-	3,287205-	9,720441	3,892028	11,556344	AvgLinDist _A	4
3,087055	1,327851-	2,096598	2,686103	1,330890	bias	5

الجدول 20

أوزان عقد خرج الصيغة المتقدمة

bias (w _y [5])	node 5 (w ₄ [i,4])	node 4 (w _x [i,3])	node 3 (w _x [i,2])	node 2 (w _x [i,1])	node 1 (w _x [i,0])
1,360308-	4,009144	6,651897	7,004782	3,289959-	4,696996-

الجدول 21

عوامل قياس خرج الصيغة المتقدمة

b _{max}	b _{min}	
0,22	3,98-	ODG

7 مطابقة عمليات التنفيذ

1.7 عام

يعرض هذا القسم مجموعة من عناصر الاختبار للتحقق من التنفيذ الأمثل للطريقة.

2.7 الاختبار

أختبرت عناصر الاختبار من قاعدة البيانات 3 (DB3)، والتي أستخدمت في التحقق من صحة النماذج. ولتبسيط الاختبار، أختبرت مجموعة فرعية من عناصر قاعدة البيانات DB3 البالغ عددها 84 عنصراً. وتتكون هذه المجموعة الفرعية من 16 عنصراً. وكان المعيار الرئيسي للاختيار هو أن تغطي قيم المتغيرات MOV وDI (مؤشر التشويه) الناتجة نطاقاً واسعاً.

3.7 إعدادات اختبار المطابقة

عناصر الاختبار متاحة للتنزيل كملفات WAV (نسق Microsoft RIFF). تم اعتيان من جميع العناصر عند 48 kHz والتشكيل PCM بمعدل 16 بته. والإشارة المرجعية وإشارة الاختبار التي قدمها الاتحاد الدولي للاتصالات هي بالفعل كيفية من

حيث الزمن والمستوى ببعضهما البعض، بحيث لا تكون هناك حاجة إلى كسب إضافي أو تعويض للتأخير. ويجب تعديل خوارزمية القياس إلى مستوى استماع يبلغ 92 dB SPL.

4.7 فاصل التسامح المقبول

من أجل المطابقة مع التوصية، يجب أن تستنسخ قيم DI المحسوبة القيم الواردة في الجدولين 22 و 23، مع تسامح أقل من $\pm 0,02$ لجميع عناصر الاختبار⁶. وإذا لم يسفر التنفيذ عن نتائج ضمن هذا التسامح فإنها لا تتطابق مع هذه التوصية.

5.7 عناصر الاختبار

يعرض الجدولان التاليان أسماء العناصر المرجعية وعناصر الاختبار⁷، وقيم DI الناتجة. ويرتبط الجدول 22 بالصيغة الأساسية، ويحتوي الجدول 23 على قيم الصيغة المتقدمة.

الجدول 22

عناصر الاختبار وقيم DI الناتجة للصيغة الأساسية

العنصر	DI	ODG
acodsna.wav	1,304	0,676-
bcodtri.wav	1,949	0,304-
ccodsax.wav	0,048	1,829-
ecodsmg.wav	1,731	0,412-
fcodsb1.wav	0,677	1,195-
fcodtr1.wav	1,419	0,598-
fcodtr2.wav	0,045-	1,927-
fcodtr3.wav	0,715-	2,601-
gcodcla.wav	1,781	0,386-
icodsna.wav	3,029-	3,786-
kcodsmc.wav	3,093	0,038
lcodhrp.wav	1,041	0,876-
lcodpip.wav	1,973	0,293-
mcodcla.wav	0,436-	2,331-
ncodsf.wav	3,135	0,045
scodclv.wav	1,689	0,435-

⁶ لتحقيق هذه الدقة ينبغي استخدام حساب النقطة الطليقة لمعهد مهندسي الكهرباء والإلكترونيات.

⁷ تُشتق أسماء العناصر المرجعية المقابلة بالاستعاضة عن السلسلة الفرعية "cod" في أسماء عناصر الاختبار بالسلسلة "ref"، على سبيل المثال، العنصر المرجعي "bcodtri.wav" إلى "breftri.wav".

الجدول 23

عناصر الاختبار وقيم DI الناتجة للصيغة المتقدمة

العنصر	DI	ODG
acodsna.wav	1,632	0,467-
bcodtri.wav	2,000	0,281-
ccodsax.wav	0,567	1,300-
ecodsmg.wav	1,594	0,489-
fcodsb1.wav	1,039	0,877-
fcodtr1.wav	1,555	0,512-
fcodtr2.wav	0,162	1,711-
fcodtr3.wav	0,783-	2,662-
gcodcla.wav	1,457	0,573-
icodsna.wav	2,510-	3,664-
kcodsme.wav	2,765	0,029-
lcodhrp.wav	1,538	0,523-
lcodpip.wav	2,149	0,219-
mcodcla.wav	0,430	1,435-
ncodsfe.wav	3,163	0,050
scodclv.wav	1,972	0,293-

المرفق 1

بالملاحق 2

عملية التحقق

1 عام

في عام 1994 اعتمد قطاع الاتصالات الراديوية المسألة ITU-R 210/10 - طرائق التقييم الموضوعية لجودة الإدراكية، وتم إنشاء فريق مهام. وكان من بين الإجراءات الأولى إطلاق دعوة مفتوحة لتقديم المقترحات، وتم تلقي ردود من ستة من مقدمي النماذج. تم بذل الكثير من الجهد لتحديد إجراءات عملية التحقق. وتبين أن من المفيد تجميع قاعدة بيانات أولى، يشار إليها باسم DB1، تتكون من مواد من اختبارات الاستماع التي تم إجراؤها بالفعل. وكان التركيز الرئيسي على الجودة الصوتية المتوسطة والعالية وبالتالي لم تؤخذ في الاعتبار إلا النتائج المستقاة من اختبارات الاستماع المتطابقة مع التوصية ITU-R BS.1116. وتمثل المواد المستقاة من

هذه الاختبارات مواد بث مهمة للكوديكات ذات معدلات البتات المنخفضة مثل MPEG1 Layer II، وMPEG1 Layer III، وDolby AC2، وMini Disc، وNICAM وغيرها. وتم إنشاء قاعدة البيانات 1 لتزويد مقدمي النماذج بمنصة مشتركة تضم المواد التي تغطي نطاقاً كبيراً من الانحطاطات ومجموعة متنوعة من الكوديكات والتدهور الناتج عن الكوديكات المتشعبة. يمكن الاطلاع على وصف تفصيلي للاختبارات المجمعة في قاعدة البيانات 1 في المرفق 2 بالملحق 2.

ومن الواضح أن طريقة القياس الموضوعي لجودة الصوت المدركة التي تحاكي السلوك البشري لا يمكن التحقق من صحتها إلا على قاعدة بيانات تحتوي على نتائج مستقاة من الاختبارات الشخصية. ويتطلب التحقق المناسب قاعدة بيانات تعتمد على مواد غير معروفة. ولهذا السبب كان من الضروري إجراء اختبارات استماع جديدة. وبما أن طريقة القياس ينبغي أن تستهدف بشكل مثالي أي نوع من النقص التي يمكن أن تظهر في تطبيقات البث، فلا ينبغي إدراج نقائص التشفير فحسب، بل يجب أيضاً إدراج النقص التقليدية مثل التشوه والضوضاء. وقد تم إنشاء قاعدة البيانات 2 وقاعدة البيانات 3 في عامي 1996 و1997 على التوالي لتلبية هذه المتطلبات. وإلى جانب الكوديكات المضمنة بالفعل في قاعدة البيانات DB1، تم أيضاً إدراج الكوديكين Dolby AC-3 وAAC. ويمكن الاطلاع على تفاصيل إضافية في المرفق 2 بالملحق 2.

وينبغي أن يُراعى في التحقق أوجه عدم اليقين، التي غالباً ما تُقدم على شكل فواصل ثقة، متأصلة في اختبارات الاستماع الشخصية. ويعتمد حجم فاصل الثقة على عدد من العوامل. وأهمها خبرة الأشخاص القائمين بالاختبار وإجراءات التدريب والسياق الذي تُعرض فيه عناصر الاختبار، فضلاً عن عدد الأشخاص القائمين بالاختبار.

ويستند تكييف الطريقة الموضوعية الواردة في هذه التوصية والتحقق منها إلى "مستمع ذي خبرة متوسطة". وتُستخدم القيم المتوسطة من تقييمات الجودة الشخصية مع فواصل الثقة 95% لتحديد خصائص "المستمع ذي الخبرة المتوسطة".

تعتبر اختبارات الاستماع الشخصية حساسة جداً للعوامل المختلفة التي تؤثر على النتائج وأنتجت الدرجات SDG لكل من قاعدة البيانات 2 وقاعدة البيانات 3 في ثلاثة مواقع اختبار مختلفة وتحقق عدد من الدراسات من إمكانية دمج البيانات بالفعل. وعلى الرغم من عدم توصل جميع الدراسات إلى استنتاجات متطابقة، فقد وجد أن من المعقول دمج البيانات وشكلت قاعدة البيانات المدججة هذه الأساس لعملية التحقق.

وُقِّسَت عملية التحقق إلى ثلاث مراحل:

- المرحلة 1: المرحلة التنافسية
- المرحلة 2: المرحلة التعاونية
- المرحلة 3: الاختبار النهائي

وسيرد وصف هذه المراحل بالتفصيل في الأقسام التالية.

2 المرحلة التنافسية

أُقتَرحت ست طرائق (DIX، NMR، PAQM، PERCEVAL، POM، TTA) للقياس الموضوعي لجودة الصوت المدركة وتقرر مقارنة أداء هذه الطرائق باستخدام قاعدة البيانات 2 ومجموعة فرعية من قاعدة البيانات 1. وتم إنشاء قاعدة البيانات 2 في بداية عام 1996. وكان اختيار مواد الاختبار النهائي بمثابة جهد مشترك بين SR (السويد) وهيئة الإذاعة البريطانية (المملكة المتحدة). وأُجريت اختبارات الاستماع في NRK في النرويج، وDR في الدنمارك، وNHK في اليابان. وأُعد تحليل إحصائي للبيانات المستقاة من الاختبارات من قبل شركة Deutsche Telekom (ألمانيا) وTeracom (السويد). وخلال المرحلة 1، تم إنشاء البيانات الموضوعية في موقع محايد (Swisscom، سويسرا). بعد ذلك، تلقى مقدمو النماذج النصف الأول من قاعدة البيانات 2 للتكييف النهائي للطرائق (المرحلة 2). وأخيراً، تم إنشاء درجات الاختلاف الموضوعي الجديدة في Swisscom.

تم إجراء تحليل أداء للطرائق من قبل شركة Teracom (السويد) وكذلك من قبل مقدمي النماذج أنفسهم. وعلى الرغم من أن نتائج بعض الطرائق المقترحة أظهرت ترابطاً كبيراً بالدرجات SDG، فقد كان هناك إجماع على أن أيّاً من الطرائق لم تلب متطلبات المستعملين. وأظهرت

دراسة منفصلة أن أياً من الطرائق المقترحة لم تكن أفضل بكثير من الطرائق الأخرى. ولذلك تقرر تطوير طريقة قياس محسنة كجهد مشترك بين جميع مقدمي النماذج الحاليين. وينبغي مقارنة أداء الطريقة الجديدة بإحدى الطرائق القائمة بالفعل والمشار إليها باسم النموذج B3.

3 المرحلة التعاونية

ارتكزت المرحلة التعاونية على فكرة الجمع بين أفضل عناصر الطرائق المختلفة في طريقة واحدة جديدة. وتلبية احتياجات المستعملين على أفضل وجه، تقرر تطوير صيغتين من الطريقة. واحدة مناسبة لعمليات التنفيذ في الوقت الفعلي، والتي قد تتطلب قدرة حاسوبية أكبر لتحقيق دقة أعلى.

وَصُمم إجراء التحقق للطرائق الجديدة بطريقة ماثلة لتلك الخاصة بالمرحلة التنافسية. وكان لا بد من إنشاء قاعدة بيانات جديدة (DB3). وُحددت العناصر والشروط في نهاية المطاف في ربيع عام 1997 وتم تجميعها في SR و Swisscom و BBC. ويمكن الاطلاع على وصف كامل لقاعدة البيانات في المرفق 2 بالملحق 2. وأجريت اختبارات الاستماع الشخصية في ثلاثة مواقع اختبار، وهي Deutsche Telekom و NHK و SR. وطبقت جميع المواقع "طريقة الحجب المزدوج الثلاثي الحوافز مع المرجع المحجوب" الموصوفة في التوصية ITU-R BS.1116. وتم جمع النتائج المستقاة من اختبارات الاستماع في السويد. وأجري تحليل إحصائي شامل لنتائج اختبار الاستماع في شركة Teracom وكذلك من قبل أطراف أخرى. ونتيجة لهذا التحليل، تم استبعاد بعض المستمعين من مواصلة التقييم. وُجمعت النتائج من مواقع الاختبار لتكوين قاعدة البيانات 3.

وفي خريف عام 1997، تم إصدار 52 عنصراً من قاعدة البيانات إلى مقدمي النماذج. كُيفت الطرائق الجديدة مع البيانات الجديدة. ونظراً لوجود العديد من قيم ضبط المعلمات التي أعطت نتائج ماثلة، فقد تم تأجيل قرار الاختيار النهائي لأطول فترة ممكنة. وفي النهاية، أُستخدمت في سويسرا العناصر المتبقية البالغ عددها 32 عنصراً في التحقق من الطرائق الجديدة على مجموعة بيانات "غير معروفة". وبالإضافة إلى ذلك، تم استخدام نتائج اختبار استماع جديد أجرته CRC (كندا) في التحقق من الطرائق الجديدة إزاء المواد "غير المعروفة". ويرد وصف عملية الاختيار والتحقق في الأقسام التالية.

4 التحقق

أُجريت اختبارات موسعة للصبغ الثماني عشرة المحددة لطريقة القياس الموضوعي. ويرد في هذا القسم وصف معايير الاختيار، وتعرض نتائج مقارنة الدرجات SDG مع النتائج التي تم الحصول عليها من الصبغ الثماني عشرة لطريقة القياس. وكان الهدف هو اختيار الصبغ المثلى التي سيوصى بها الاتحاد الدولي للاتصالات والتحقق منها.

معايير الاختيار

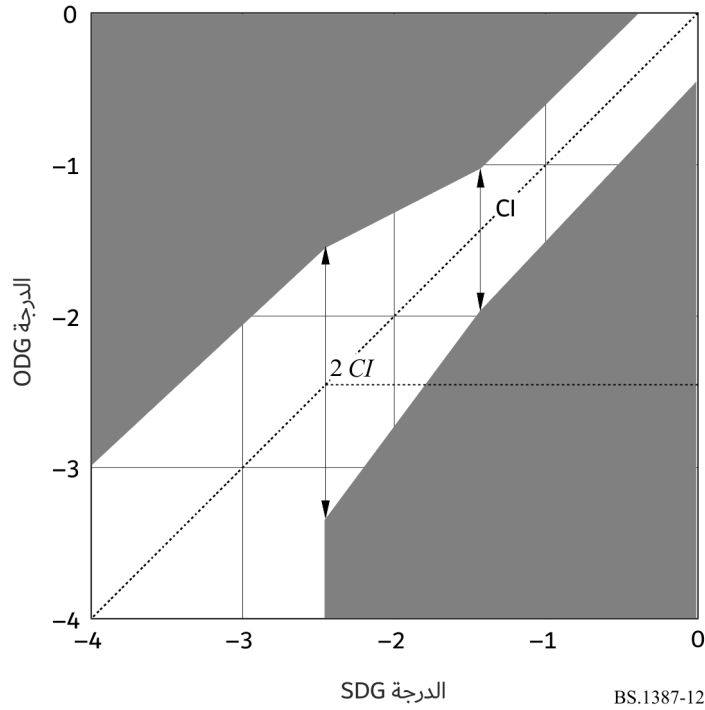
العلاقة بين النتائج الشخصية والموضوعية هي المعيار الأكثر وضوحاً للتحقق من الطريقة الموضوعية. وبالإضافة إلى ذلك، تم طرح معيارين آخرين يأخذان في الاعتبار موثوقية القيم المتوسطة للتحقق - المعيار AES ومخطط التسامح. ويُعرف المعيار AES، الذي طُرح لربط دقة النموذج بدقة اختبار الاستماع، وفقاً للمعادلة التالية:

$$AES = 2 * \sqrt{\frac{\sum ((ODG - SDG) / CI)^2}{N}} \quad \text{if } CI < 0.25 \text{ then } CI = 0.25$$

حيث CI فاصل الثقة.

وَصُمم مخطط التسامح للسماح بانحرافات مختلفة للدرجات ODG عن الدرجات SDG في الطرف العلوي والسفلي من سلم الانحطاط. ويرتبط نطاق التسامح بفواصل الثقة في اختبارات الاستماع. ويقتصر هذا النطاق على قيمة لا تقل عن 0,25 من الدرجات. وأُستخدمت مسافة الدرجات ODG خارج مخطط التسامح إلى مخطط التسامح لتقييم جودة طريقة القياس.

الشكل 12

مخطط التسامح، فاصل الثقة $CI \geq 0,25$ 

1.4 المقارنة بين الدرجات SDG و ODG

قُسمت القياسات الموضوعية إلى ثلاث مراحل مختلفة. ففي المرحلة 1، كانت جميع عناصر الاختبار البالغ عددها 84 غير معروفة للجميع باستثناء فريق الاختيار. وفي المرحلة 2، تم إصدار معلومات عن 52 عنصراً. وتضمنت المعلومات كلاً من قيم الدرجات SDG والمقتطفات الصوتية الفعلية. وفي المرحلة الثالثة، أُستخدمت هذه المعلومات لتحسين أداء صيغ الطريقة. ويرجى العلم أنه تم اختبار أربع صيغ إضافية خلال المرحلة 3، مقارنة بالمرحلة 1. وحُسبت قيم الدرجات SDG المقدمة من البيانات التي تم إصدارها بواسطة 75 شخصاً مؤهلاً للاختبار.

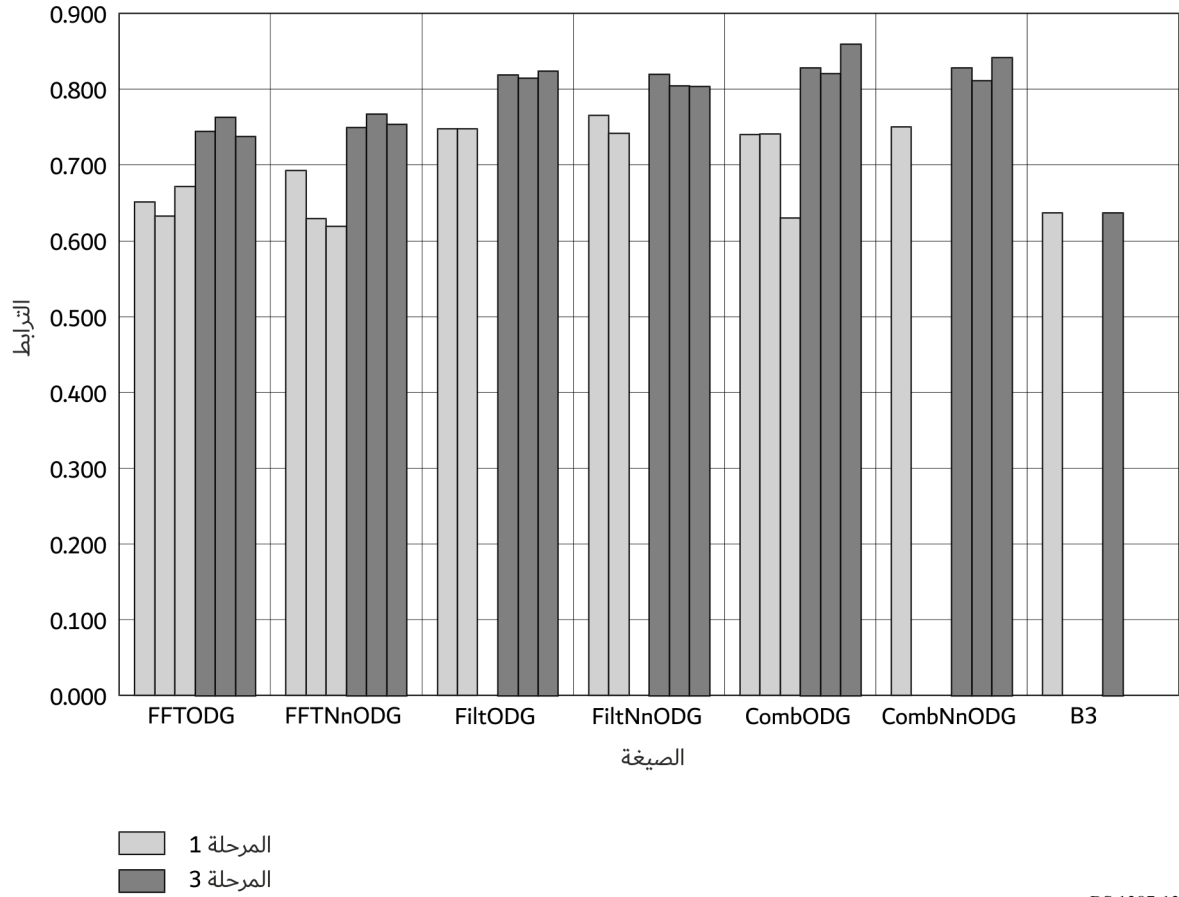
وهناك العديد من الأساليب المختلفة لتقييم مدى الدقة التي تبرز بها الدرجات ODG الدرجات SDG. ولسوء الحظ، لا توجد قيمة واحدة توضح الأداء الكامل بحق. وبدلاً من ذلك، يتعين على المرء أن يبحث في عدد من وجهات النظر. وترد الترابطات في الفقرة 2.4 وتُعرض درجات الخطأ المطلق (AES) في الفقرة 3.4. والنموذج B3 هو أحد النماذج التي اختبرها قطاع الاتصالات الراديوية في عام 1996 وتقرر مقارنة مختلف الصيغ الجديدة مع هذه الصيغة الأقدم.

2.4 الترابط

رُسمت مخططات الترابط من المرحلتين 1 و 3 في الشكل 13 (84 عنصراً) والشكل 14 (32 عنصراً).

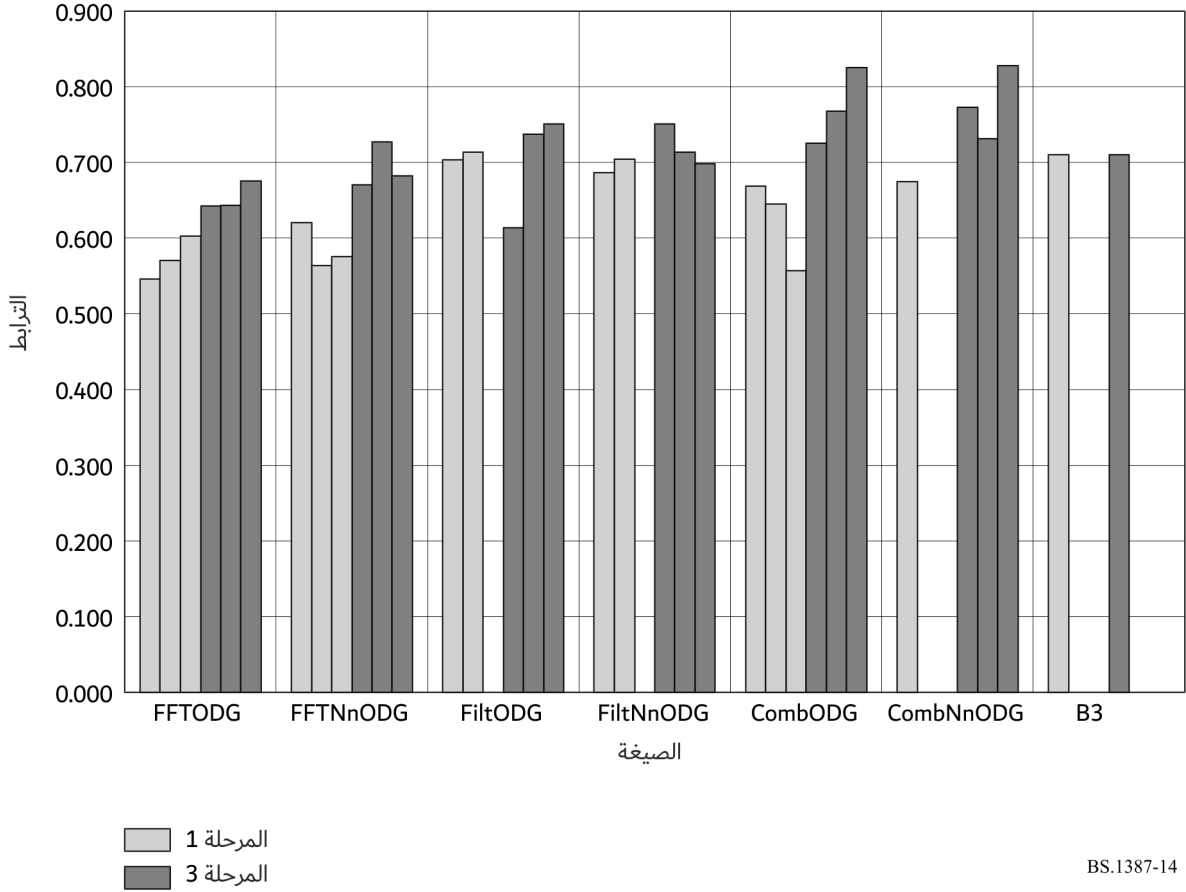
الشكل 13

التربط بين الدرجات SDG و od
تم إدراج جميع العناصر البالغ عددها 84



الشكل 14

الترايط بين الدرجات ODG وSDG
تم إدراج العناصر التي لم تصدر البالغ عددها 32



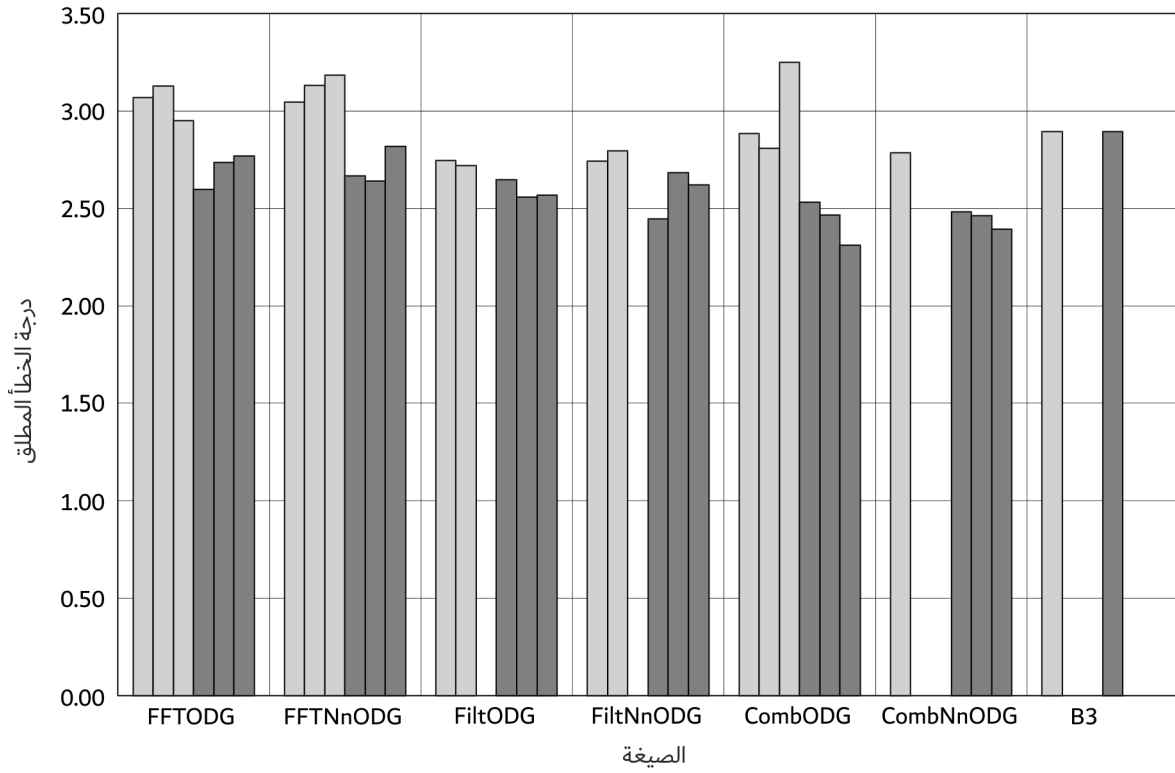
BS.1387-14

3.4 درجة الخطأ المطلق (AES)

أي نموذج ينتج، في المتوسط، قيم للدرجات ODG ضمن فاصل الثقة للدرجات SDG سيحصل على قيمة لدرجة الخطأ المطلق (AES) تقترب من 2. وترد نظرة عامة على قيم الدرجة AES في الشكلين 15 و16.

الشكل 15

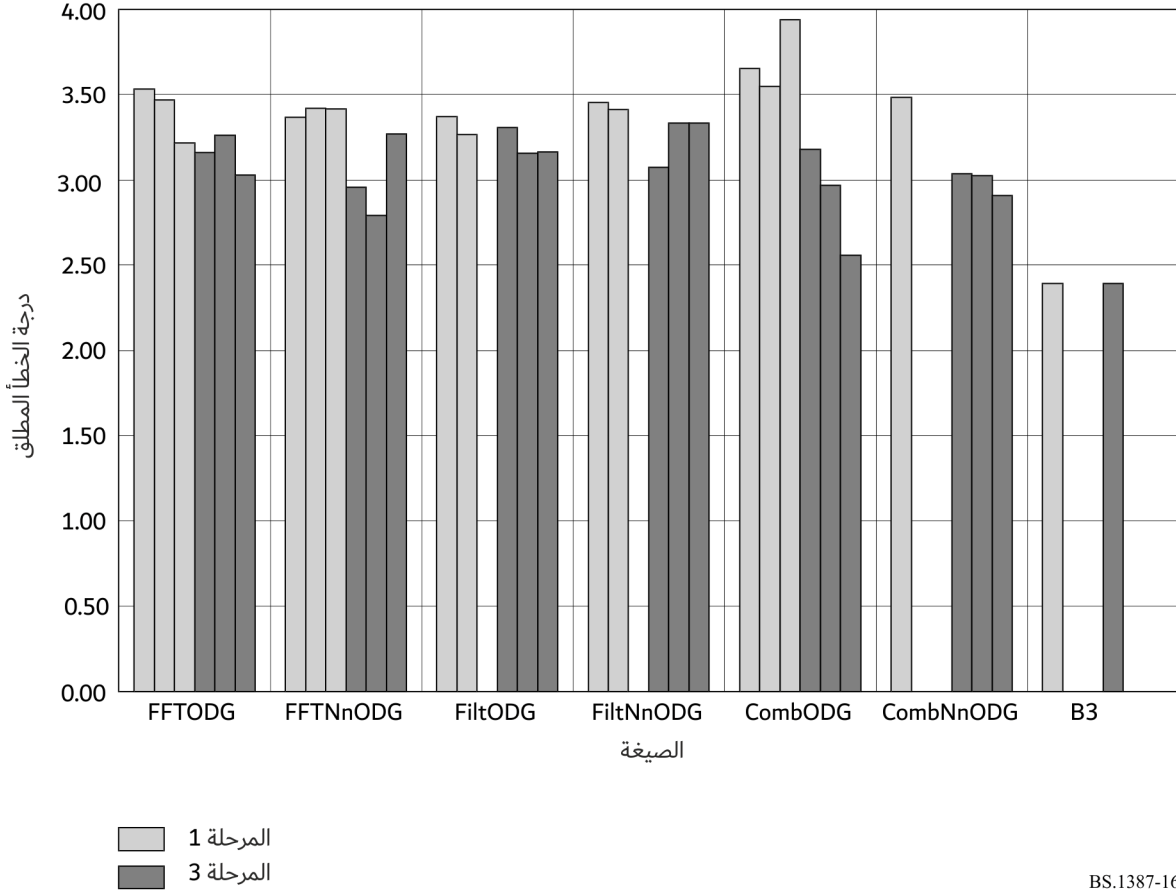
درجة الخطأ المطلق للصيغ المختلفة
تم إدراج جميع العناصر البالغ عددها 84



المرحلة 1
المرحلة 3

الشكل 16

درجة الخطأ المطلق للصيغ المختلفة
تم إدراج العناصر التي لم تصدر البالغ عددها 32



BS.1387-16

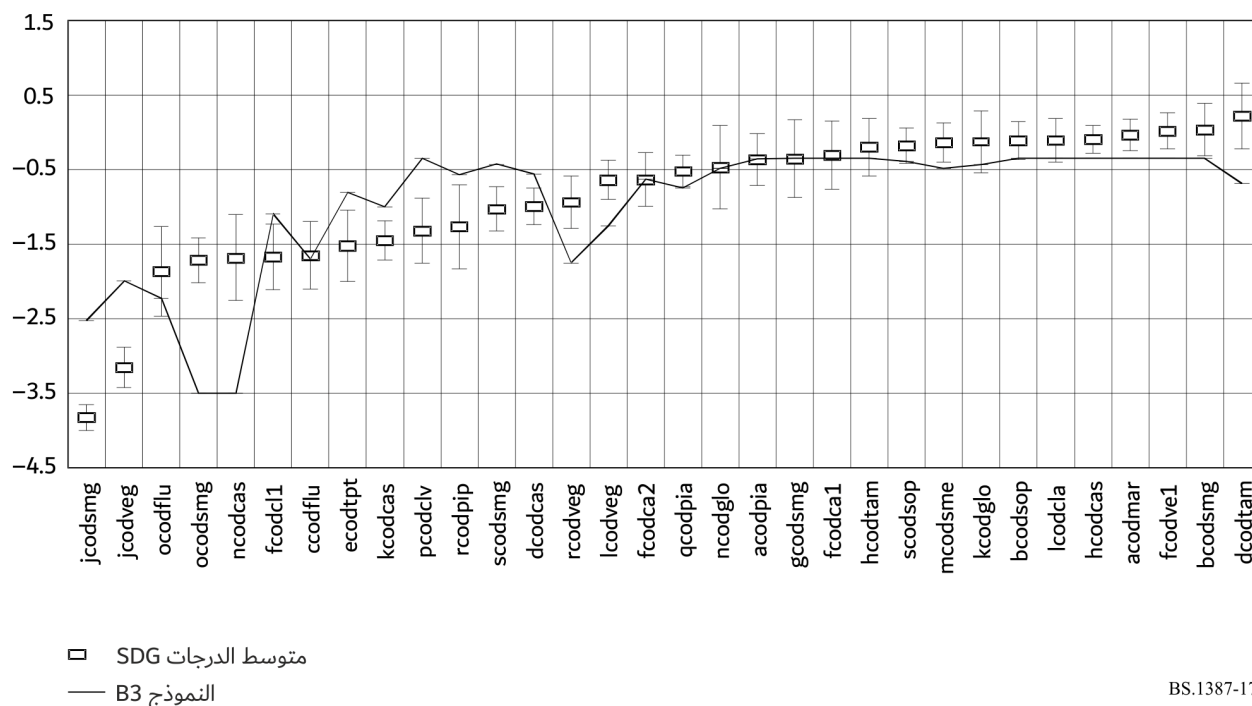
4.4 مقارنة الدرجات ODG مقابل فاصل الثقة

كان أداء معظم الصيغ متشابهاً، كما كان متوقعاً على الأرجح. وعُرض الكثير من المخططات في الاجتماع ولكن في هذه التوصية فإن مستودع المخططات محدود. ولمزيد من التفاصيل، يرجى الرجوع إلى تقرير اختبار التحقق الكامل.

وفي الأشكال من 17 إلى 22، تم رسم متوسط الدرجات SDG وفاصل الثقة والدرجات ODG للعناصر التي لم تصدر البالغ عددها 32 للنموذج B3 وصيغتي النموذج FFTNnODG1 و CombNnODG3.

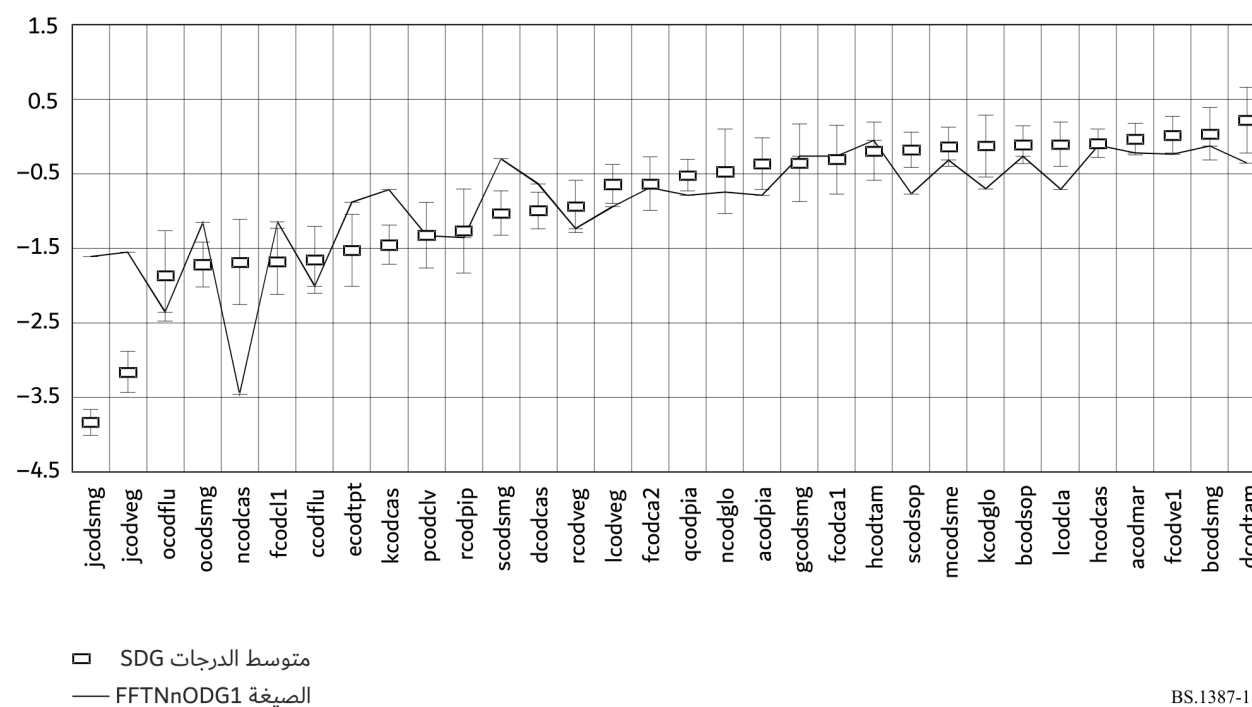
الشكل 17

النموذج B3: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG للعناصر التي لم تصدر البالغ عددها 32



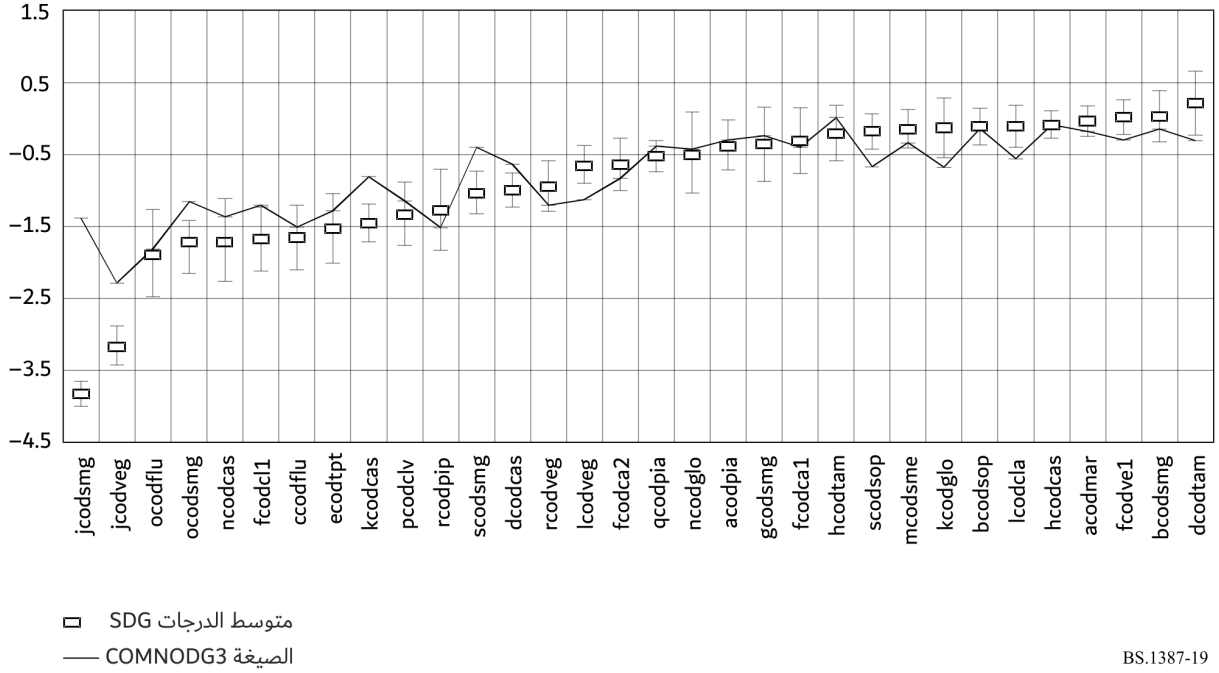
الشكل 18

الصيغة FFTNnODG1: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG بعد المرحلة الثالثة للعناصر التي لم تصدر البالغ عددها 32



الشكل 19

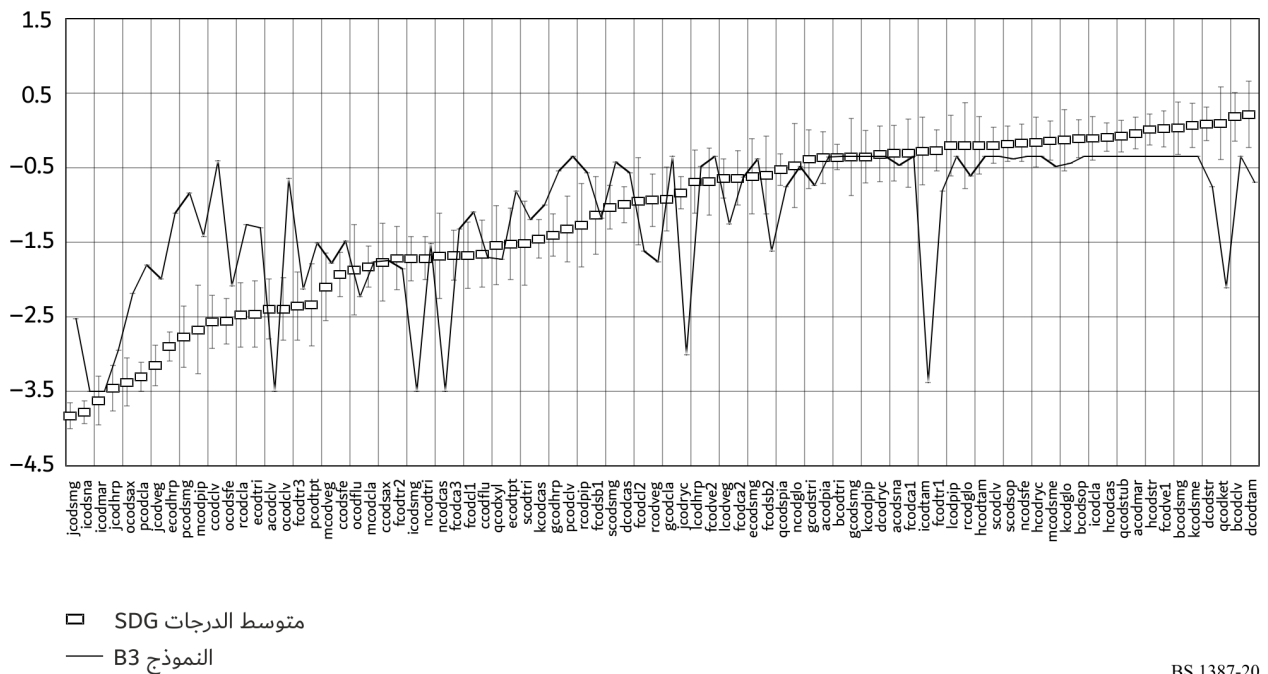
الصيغة CombNnODG3: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG بعد المرحلة الثالثة للعناصر التي لم تصدر البالغ عددها 32



وترد في الشكل 20 والشكل 21 مخططات ماثلة، ولكن في هذه الحالة لجميع العناصر البالغ عددها 84 خلال المرحلة 3. وبالإضافة إلى ذلك، يوضح الشكل 22 أداء الصيغة CombNnODG3.

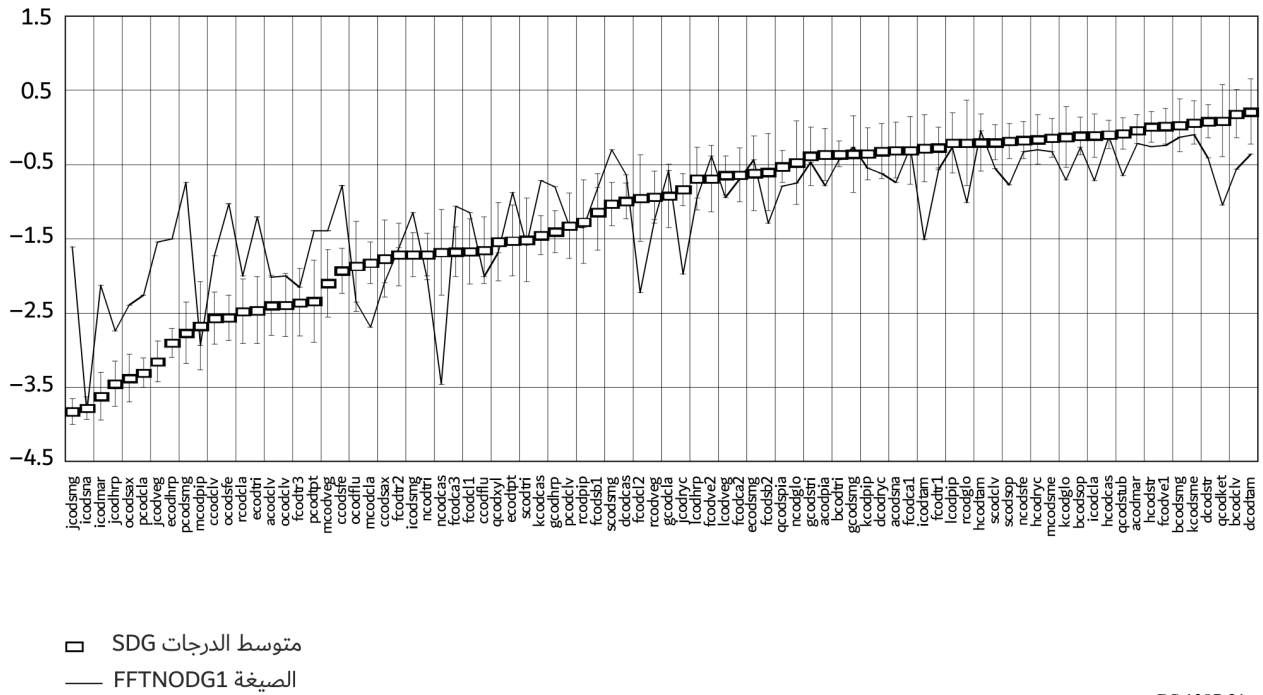
الشكل 20

النموذج B3: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG لجميع العناصر البالغ عددها 84



الشكل 21

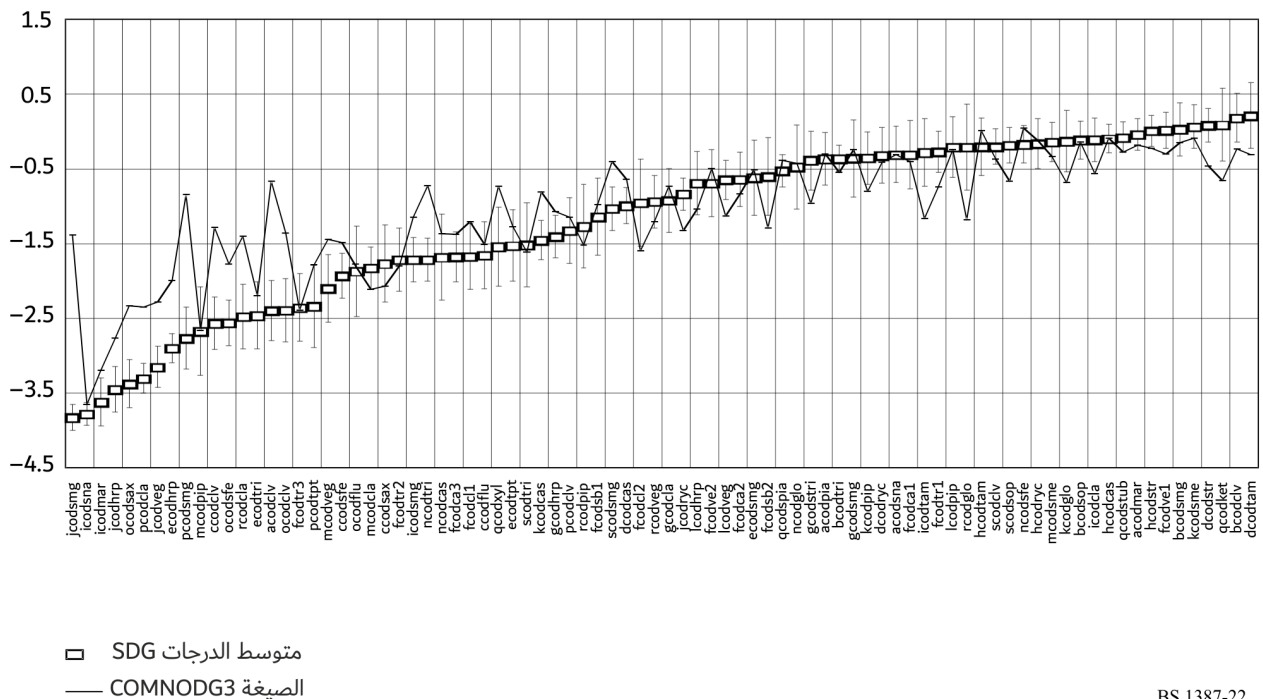
الصيغة FFTNnODG1: مخطط متوسط الدرجات و SDG وفاصل الثقة والدرجات ODG أثناء المرحلة الثالثة لجميع العناصر البالغ عددها 84



BS.1387-21

الشكل 22

الصيغة CombNnODG3: مخطط متوسط الدرجات و SDG وفاصل الثقة والدرجات ODG أثناء المرحلة الثالثة لجميع العناصر البالغ عددها 84



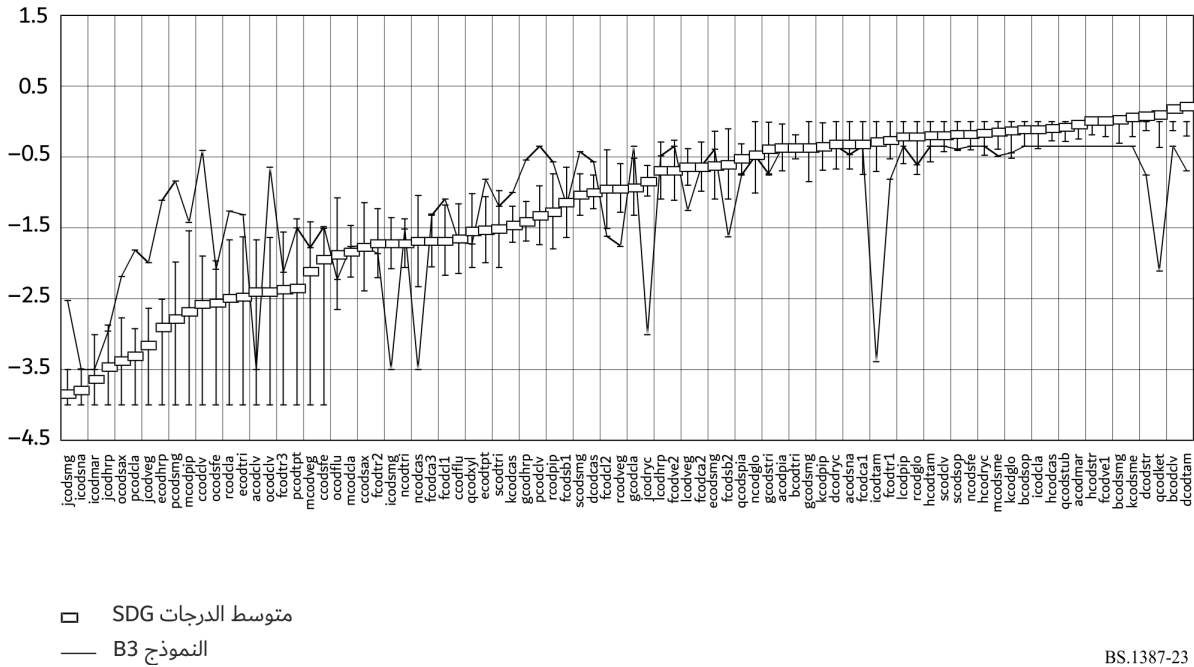
BS.1387-22

5.4 مقارنة الدرجات ODG مقابل فاصل التسامح

حدد قطاع الاتصالات الراديوية متطلباً مستهدفاً للمستعمل يمكن تعاقبه مع فاصل التسامح. وتكون المتطلبات المستهدفة أكثر صرامة بالنسبة للمستويات الأعلى من جودة الصوت وأقل صرامة بالنسبة لجودة الصوت المنخفضة. وتوضح الأشكال أدناه الأداء في هذا البعد للنموذج B3 وصيغتي النموذج FFTNnODG1 و CombNnODG3 لجميع العناصر البالغ عددها 84 خلال المرحلة 3.

الشكل 23

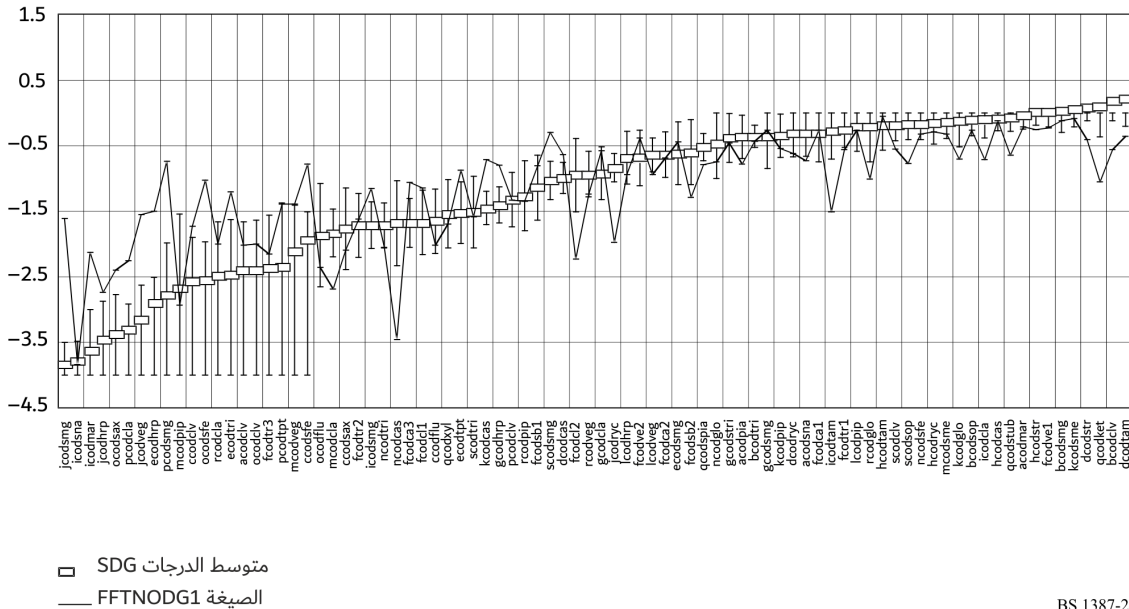
النموذج B3: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG أثناء المرحلة 3 لجميع العناصر البالغ عددها 84



BS.1387-23

الشكل 24

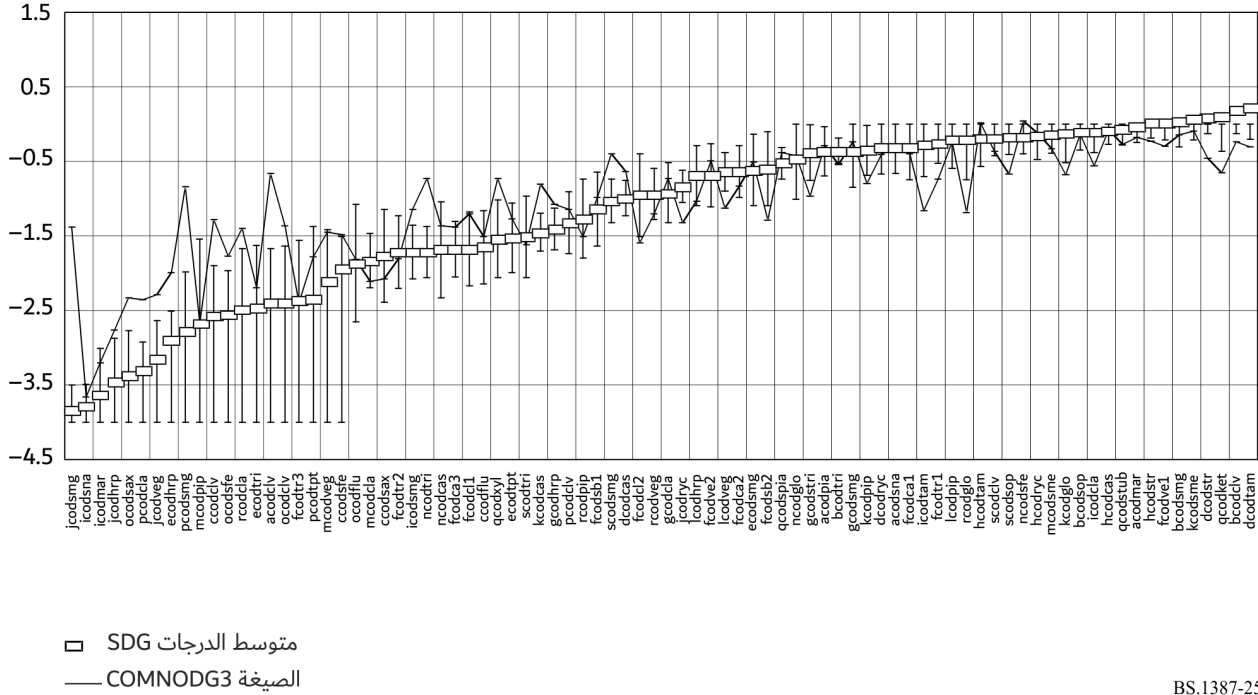
الصيغة FFTNnODG1: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG أثناء المرحلة 3 لجميع العناصر البالغ عددها 84



BS.1387-24

الشكل 25

الصيغة CombNnODG3: مخطط متوسط الدرجات SDG وفاصل الثقة والدرجات ODG أثناء المرحلة 3 لجميع العناصر البالغ عددها 84



BS.1387-25

5 اختيار صيغ النموذج المثلى

قُدمت ثماني عشرة صيغة مختلفة للنموذج إلى موقع الاختبار الموضوعي: 6 صيغ قائمة على التحويل FFT، وست صيغ تستخدم بنك مرشحات مُكَيَّف إدراكياً وست صيغ تستخدم مزيجاً من كل من التحويل FFT وبنك المرشحات. وتهدف مجموعة التحويل FFT إلى استخدامها في مخطط قياس قادر على التنفيذ في الوقت الفعلي، بينما تحتاج المجموعتان المتبقيتان إلى تعقيد حاسوبي أكبر ومن المتوقع أن تحققان دقة أعلى. ويمكن تقسيم صيغ النموذج الست في كل مجموعة إلى مجموعتين فرعيتين: إحداهما تستخدم الشبكات العصبية والأخرى لا تستخدم الشبكات العصبية. وأدرج أيضاً أداء النموذج المرجعي من أبريل 1996 (B3) في قاعدة البيانات 3 لأغراض المقارنة.

1.5 معايير الاختيار المسبق استناداً إلى الترابط

- قُيم أداء 18 صيغة مختلفة للنموذج والنموذج المرجعي B3 باستخدام مجموعتي بيانات لم يتم استخدامهما في الغالب لتدريب النماذج (لم يتم استخدام الجزء الثاني من المرحلتين 1 و3 وقاعدة البيانات CRC للتدريب).
- أنشئت قاعدة البيانات 3 خصيصاً لتقييم النماذج الإدراكية. وأستخدم حوالي نصف قاعدة البيانات هذه لتدريب النماذج. وأستخدم الترابط بين النتائج الشخصية والنتائج الموضوعية على العناصر المتبقية (DB3_2nd) لتقييم النماذج الموضوعية.
- أنشأت قاعدة بيانات CRC لتقييم مخططات التشفير الصوتي الإدراكي. ولم تُستخدم قاعدة البيانات هذه لتدريب النماذج الإدراكية. وأستخدم الترابط بين النتائج الشخصية والنتائج الموضوعية على جميع العناصر لتقييم النماذج الموضوعية. وترد نتائج قاعدة البيانات 3 في الجدولين 24 و25 (الترابطات ودرجة الخطأ AES).

وفي مرحلة الاختيار المسبق، تم أخذ جميع المدخلات التي تحدد أداء جميع صيغ النموذج كمياً بعين الاعتبار مع إعطاء أهمية كبيرة لقاعدة البيانات DB3. وبناءً على المقارنة الشاملة، تقرر إجراء مقارنة إضافية بين صيغتي النموذج في كل مجموعة والتي يبدو أنها الأفضل. ويبين الجدول 24 الترابط بين هذه الصيغ الست (ثلاث مرات لصيغتين من كل نموذج).

الجدول 24

الترابط بين الدرجات ODG وSDG

B3	CombODG3	CombNnODG3	FiltODG3	FiltODG2	FFTNnODG2	FFTNnODG1	
0,710	0,826	0,828	0,751	0,738	0,728	0,671	DB3_2nd
0,656	0,777	0,851	0,839	0,862	0,779	0,837	CRC

الجدول 25

درجة الخطأ المطلق

B3	CombODG3	CombNnODG3	FFTOG3	FFTOG2	FFTNnODG2	FFTNnODG1	
2,39	2,56	2,91	3,16	3,16	2,79	2,96	DB3_2nd
2,78	1,90	1,61	1,67	1,61	1,85	1,55	CRC

تُظهر صيغتنا النموذج القائمتان على التحويل FFT أداءً مختلفاً في قاعدتي البيانات. ويحتاج قرار اختيار إحدى الصيغتين إلى معايير إضافية. ومع الأخذ في الاعتبار أن قاعدة البيانات CRC كانت غير معروفة تماماً، يبدو أن أداء الصيغة FFTNnODG1 أفضل قليلاً. وتعطي صيغ النموذج المدججة نتائج أفضل على قاعدتي البيانات مقارنة بالصيغ القائمة على التحويل FFT. ومع ذلك، فإنها تظهر أداءً أسوأ على قاعدة البيانات CRC مقارنة بصيغ بنك المرشحات. ويمكن اعتبار صيغ بنك المرشحات حالة خاصة من صيغ النموذج المدججة حيث يكون ترجيح متغيرات خرج الصيغة القائمة على التحويل FFT صفراً. ولذلك تُفضل صيغ النموذج المدججة.

2.5 تحليل عدد القيم الشاذة

يُعبّر عن دقة البيانات الشخصية الواردة من اختبارات الاستماع عبر فاصل قدره 95% حول متوسط عدد من المستمعين. ويُقيم أداء صيغ النموذج أيضاً من خلال النظر في القيم الشاذة. ويعتبر العنصر بمثابة قيمة شاذة عندما يكون الفرق بين البيانات الشخصية والبيانات الموضوعية أكبر من ضعف فاصل الثقة.

ويوضح الجدول 26 عدد القيم الشاذة لصيغ النموذج الست لقاعدة البيانات DB3. و"حساس" يعني أن صيغة النموذج تشير إلى جودة صوت أقل من التقييم الشخصي، بينما "غير حساس" هو العكس.

الجدول 26

القيم الشاذة

CombODG3	CombNnODG3	FiltODG3	FiltODG2	FFTNnODG2	FFTNnODG1	
5	3	4	4	4	10	حساس
14	12	13	11	13	13	غير حساس
19	15	17	15	17	23	الإجمالي

وبالنظر إلى عدد القيم الشاذة، يبدو أن صيغة النموذج FFTNnODG2 تعمل بشكل أفضل من صيغة النموذج FFTNnODG1. وتُظهر صيغة النموذج CombNnODG3 أفضل أداءً بين الصيغ الأكثر دقة.

3.5 تحليل شدة القيم الشاذة

الجدول 27

العناصر ذات الانحرافات التي تزيد عن 0,1 من درجات الفرق بين التنبؤ والدرجات SDG

6	9	12	12	4	4	14	عدد القيم الشاذة
FiltODG3	FiltODG2	FFTNnODG2	FFTNnODG1	CombNnODG3	CombODG3	B3	صيغة النموذج
jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	العنصر
rcodpip	rcodpip	qcodket	qcodket	pcodsmg	qcodket	qcodket	
qcodket	qcodket	pcodsmg	pcodsmg	ccodclv	pcodsmg	pcodsmg	
pcodsmg	pcodsmg	pcodcla	ocodsfe	acodclv	icodmar	pcodcla	
ocodsax	ocodsax	ocodsfe	ncodcas			ocodclv	
fcodsb2	kcodpip	ncodglo	mcodcla			ncodcas	
	jcodveg	kcodcas	jcodveg			mcodpip	
	fcodsb2	jcodveg	jcodyrc			jcodyrc	
	fcodcl2	jcodyrc	icodtam			icodtam	
		icodtam	icodmar			icodsmg	
		ecodhrp	fcodcl2			fcodsb2	
		ccodclv	ecodhrp			ecodhrp	
						ccodclv	
						acodclv	

الجدول 28

العناصر ذات الانحرافات التي تزيد عن 1,5 من درجات الفرق بين التنبؤ والدرجات SDG

2	2	2	3	2	2	8	عدد القيم الشاذة
FiltODG3	FiltODG2	FFTNnODG2	FFTNnODG1	CombNnODG3	CombODG3	B3	صيغة النموذج
jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg		العنصر
				pcodsmg		qcodket	
		pcodsmg	pcodsmg		pcodsmg	pcodsmg	
pcodsmg	pcodsmg						
			ncodcas				
						ncodcas	
						jcodyrc	
						icodtam	
						icodsmg	
						ccodclv	
						acodclv	

6 الخلاصة

بعد القرارات المتخذة أثناء التطوير، تم اختيار صيغتين للنموذج من بين الصيغ المقترحة البالغ عددها 18 من أجل تلبية متطلبات التطبيقات المحددة لطرائق القياس الموضوعية. وحُدِدت صيغة قليلة التعقيد، مصممة للسماح بالتنفيذ في الوقت الفعلي بكفاءة من حيث التكلفة، وصيغة أعلى دقة، لا تعمل بالضرورة في الوقت الفعلي. وبالنسبة لعملية الاختيار، تم تطبيق وتحليل المعايير المذكورة أعلاه.

وباعتبارها صيغة الوقت الفعلي، أُختير نموذج قائم على التحويل FFT، يشار إليه باسم "FFTNnODG1" للأسباب التالية:

لم تُظهر أي من الصيغ القائمة على التحويل FFT ميزة كبيرة على الصيغ الأخرى فيما يتعلق بأحد المعايير المذكورة أعلاه. وفيما يتعلق بمعاملات الترابط بين الدرجات SDG و ODG، فإن عدد وشدة القيم الشاذة لكل صيغة من الصيغ التي تم التحقق منها لها مزاياها وعيوبها. ومع ذلك، جدير بالذكر أن النموذج "FFTNnODG1" حقق أفضل ترابط في قاعدة بيانات CRC'97 التي كانت غير معروفة تماماً (0,837).

وبالنسبة للصيغة ذات الدقة الأعلى، تم إعطاء الأفضلية للصيغة المدججة، التي تشتمل على التحويل FFT وبنك المرشحات، لأن هذا النهج يشتمل أيضاً على مجموعة فرعية من نموذج بنك المرشحات الخالص وبالتالي يجب أن يكون له إمكانات أداء أفضل. وكانت الصيغ الست المختلفة للنموذج المدمج متاحة لعملية الاختيار. وأظهرت الصيغة المختارة "CombNnODG3" قيمة شاذة أقل بالإضافة إلى ترابط أعلى ($r=0.851$ for CRC'97) مقارنة بالصيغ الأخرى. وكان ترابط هذه الصيغة بالنسبة لقاعدة البيانات 3 الكاملة بنفس الترتيب من حيث الحجم الموجود الخاص بالصيغة "CombODG3"، ولكنها أظهرت ترابطاً أعلى مقارنة بالصيغ الأخرى.

المرفق 2

بالملاحق 2

أوصاف قواعد البيانات المرجعية

1 مقدمة

أثناء تطوير طريقة القياس الموضوعي لجودة الصوت المدركة، تم استخدام عدد من قواعد البيانات للتدريب والتحقق.

وتحتوي بعض قواعد البيانات المدرجة على بيانات سماعات الرأس ومكبرات الصوت، وبعضها يحتوي على بيانات سماعات الرأس فقط. وبالنسبة لقواعد البيانات التي تحتوي على مجموعات منفصلة من البيانات المتاحة لعرض مكبرات الصوت وسماعات الرأس، أُستخدمت بيانات سماعات الرأس فقط.

ويُعرف العنصر على أنه جزء صوتي يستخدم في التقييم الشخصي. وتشير الحالة إلى حالة تدهور واحدة وأُستخدمت جميع العناصر لجميع الحالات التجريبية باستثناء دراسات قاعدتي البيانات DB2 و DB3. وأُستخدمت قاعدة البيانات DB3 جزئياً للتدريب وجزئياً للتحقق (تم استخدام 52 عنصراً من أصل 84 عنصراً للتدريب في المرحلة الثانية من التحقق).

التدريب

– MPEG90

– يغطي متوسط الدرجات SDG لكل عنصر بشكل منتظم النطاق من 0,0 إلى -4,0.

– تقرير اختبار MPEG/Audio للفريق ISO/IEC JTC 1/SC 2/WG 11، الوثيقة MPEG90/N0030، أكتوبر 1990.

- MPEG91 –
 – كان ما لا يقل عن 88 في المائة من متوسط الدرجات SDG لكل عنصر أعلى من -2,0، وكان النطاق من 0,1 إلى -3,8.
 – تقرير اختبار MPEG/Audio للفريق ISO/IEC JTC 1/SC 2/WG 11، الوثيقة MPEG91/N0010، يونيو 1991.
- ITU92DI –
 – كان ما لا يقل عن 80 في المائة من متوسط الدرجات SDG لكل عنصر أعلى من -2,0، وكان النطاق من 0,1 إلى -3,4.
- ITU92CO –
 – كان ما لا يقل عن 96 في المائة من متوسط الدرجات SDG لكل عنصر أعلى من -2,0، وكان النطاق من 0,2 إلى -2,4.
- ITU93 –
 – كان معظم متوسط الدرجات SDG لكل عنصر أعلى من -2,0، وكان النطاق من -0,1 إلى -2,3. ولم يكن هناك فرق كبير بين البيانات من المختبرين.
 – Grusec وآخرون [1997].
- MPEG95 –
 – كان ما لا يقل عن 63 في المائة من متوسط الدرجات SDG لكل عنصر أعلى من -2,0، وكان النطاق من -0,2 إلى -3,8.
 – Kim و Meares [1995]
- EIA95 –
 – كان ما لا يقل عن 93 في المائة من متوسط الدرجات SDG لكل عنصر أعلى من -2,0، وكان النطاق من 0,1 إلى -3,7.
 – Grusec وآخرون [1997].
- DB2 –
 – لم تُستخدم جميع العناصر لجميع الحالات.
- التحقق
- DB3 –
 – لم تُستخدم جميع العناصر لجميع الحالات.
- CRC97 –
 – يغطي متوسط الدرجات SDG لكل عنصر بشكل منتظم النطاق من 0,1 إلى -3,6.
 – Soulodre وآخرون [1998].
- وتصف الأقسام التالية العناصر التي تم إدراجها في قواعد البيانات المختلفة والحالات التي تم تطبيقها.

العناصر لكل قاعدة بيانات 2

Item	MPEG90	MPEG91	ITU92DI	ITU92CO	ITU93	MPEG95	EIA95	DB2	DB3	CRC97
Accordion/Triangel		*								
Åsa Jinder			*	*	*					
Bag Pipe						*		*	*	
Bag Pipe-2								*		
Bass Clarinet								*		
Bass Guitar	*		*	*						
Bass Synth	*									
Carmen		*								
Castanets	*		*	*	*	*		*	*	
Clarinet					*		*	*	*	*
Clarinet-mono								*		
Clarinet2								*		
Claves									*	
Dalarnas Spelmansförbund "Trettondagsmarschen"			*	*						
Dire Straits "Ride Across the River"			*				*			*
Double Bass										*
Drum								*		
Fireworks	*									
Flute									*	
George Duke		*								
Glockenspiel	*	*				*	*	*	*	
Harpicord			*	*	*	*	*	*	*	*
Horn								*		
Kettle drums									*	
Marimba								*	*	
MPE mono = Speech male engl. mono								*		
Music and rain							*			*
Muted trumpet							*			
Ornette Coleman	*	*	*	*						
Pearl Jam							*			
Percussion		*						*		
Piano Schubert									*	
Pitch Pipe						*			*	*
Ravel "Feria"			*							
Ry Cooder								*	*	
Ry Cooder (mono)										
Saxophon									*	
Snare drum									*	
Soprano Mozart									*	
Speech female engl						*			*	
Speech female germ									*	
Speech male engl	*	*							*	
Speech male germ			*	*	*				*	
Stravinsky "Wind Octet"			*	*						
Strings								*	*	
Strings mono								*		
Suzanne Vega "Toms Diner"	*	*	*	*				*	*	*
Suzanne Vega with breaking glass							*			
Tambourine		*						*	*	
Tracy Chapman	*									
Triangle			*	*				*	*	
Trumpet								*	*	*
Trumpet (Haydn)	*									
Tuba								*	*	
De sålde sina hemman (violin solo)					*					
Water Sound							*			
Wind Ensemble								*		
Xylophone									*	

3 الحالات التجريبية

بالنسبة لجميع معدلات البتات مع بيان بالإشارة المجسمة kbit/s، يُحدد معدل البتات الإجمالي، على سبيل المثال، صوت مجسم kbit/s 256 يعني أنه تم تخصيص ما مجموعه kbit/s 256 لكلتا قناتي إشارة الصوت المجسم. وإذا لم تتم الإشارة إلى أي شيء آخر، يشير الصوت المجسم إلى تشفير قناة مستقلة.

MPEG90 1.3

ثلاثة معدلات للبتات: 64 kbit/s للقناة الأحادية، و 192 kbit/s و 256 kbit/s للصوت المجسم، وجميع المواد غير متاحة لقاعدة البيانات هذه.

.Musicam –

.SB-ADPCM –

MPEG91 2.3

ثلاثة معدلات للبتات: 64 kbit/s للقناة الأحادية، و 192 kbit/s و 256 kbit/s للصوت المجسم.

.MPEG1 Layer I –

.MPEG1 Layer II –

.MPEG1 Layer III –

.MUSICAM –

.ASPEC –

.NICAM –

ITU92DI 3.3

خمسة كوديكات توزيع: 240 kbit/s للصوت المجسم.

خضع كل عنصر للمعالجة بواسطة الكوديك نفسه ثلاث مرات جنباً إلى جنب مع انخفاض في المستوى بمقدار 0,1 dB قبل كل مرة.

.MPEG1 Layer II –

.MPEG1 Layer III –

.Dolby AC-2 –

.Aware –

.NHK –

ITU92CO 4.3

سنة كوديكات مساهمة: 360 kbit/s للصوت المجسم. خضع كل عنصر للمعالجة بواسطة الكوديك نفسه ثلاث مرات جنباً إلى جنب مع انخفاض في المستوى بمقدار 0,1 dB قبل كل مرة.

.MPEG1 Layer II –

.MPEG1 Layer III –

.Dolby AC-2 –

.Dolby Low-Delay –

.Aware –

ITU93 5.3

تشكيلات MPEG1 Layer II مع الكوديكات الترادفية

– كوديك بث فقط مع صوت مجسم 256 kbit/s.

- كوديك بث فقط مع صوت مجسم 192 kbit/s (تشفير مجسم مشترك).
- ثمانية كوديكات مساهمة بمعدل 360 kbit/s متبوعة بكوديك بث واحد بمعدل 256 kbit/s، كلها بصوت مجسم.
- ثمانية كوديكات مساهمة بمعدل 360 kbit/s متبوعة بكوديك بث واحد بمعدل 192 kbit/s، كلها بصوت مجسم.
- خمسة كوديكات مساهمة بمعدل 360 kbit/s متبوعة بثلاثة كوديكات توزيع بمعدل 240 kbit/s وكوديك بث واحد بمعدل 256 kbit/s، كلها بصوت مجسم.
- خمسة كوديكات مساهمة بمعدل 360 kbit/s متبوعة بثلاثة كوديكات توزيع بمعدل 240 kbit/s وكوديك بث واحد بمعدل 192 kbit/s، كلها بصوت مجسم.

6.3 MPEG95

عمليات تنفيذ الكوديكات (64 kbit/s)

- تم اختيار اثنين وعشرين شكلاً مختلفاً للتشفير من مجموعة أكبر من طرائق التشفير المتاحة من 6 كوديكات تنفذ مجموعة فرعية مكونة من 4 نماذج ذات دقة منخفضة و 17 نموذجاً للزمن/التردد عالية الدقة.
- كانت المؤسسات المشاركة هي AT&T، و Fraunhofer، و Sony، و GCL، و RAI/Alcatel، و Philips.
- كانت جميع العناصر عبارة عن تسجيلات أحادية الأذن تم عرضها ثنائية الأذن.

7.3 EIA95

- | | | |
|---|------------------------------|---|
| 224 kbit/s stereo (تشفير مجسم مشترك) | Eureka 147/MPEG1 Layer II #1 | - |
| 192 kbit/s stereo (تشفير مجسم مشترك) | Eureka 147/MPEG1 Layer II #2 | - |
| 160 kbit/s stereo | AT&T/Lucent | - |
| 128 kbit/s stereo | AT&T/Lucent/Amati #1 | - |
| 160 kbit/s stereo | AT&T/Lucent/Amati #2 | - |
| 160 kbit/s stereo | VOA/JPL | - |
| 128-256 kbit/s stereo (معدل بتات متغير) | USADR-FM #1 | - |
| 128-256 kbit/s stereo (معدل بتات متغير) | USADR-FM #2 | - |
| 96 kbit/s stereo | USADR-AM | - |

8.3 DB2

- | | | |
|---|------------------------------|---|
| 256 kbit/s stereo, 1, 3, 5, 7, and 9 stages | MPEG1 Layer II | - |
| 256 kbit/s stereo, 1, 3, 5, 7, and 9 stages | Dolby AC2 | - |
| 192 kbit/s stereo (joint stereo coding) | MPEG1 Layer II | - |
| 64 kbit/s mono | MPEG1 Layer II | - |
| 64 kbit/s mono | MPEG2 Layer II | - |
| 384 kbit/s stereo | MPEG1 Layer II | - |
| 128, 160, 192 kbit/s, all stereo. | MPEG1 Layer III | - |
| 256 and 384 kbit/s both stereo. | APT-X | - |
| | التقدير الكمي للتشوه | - |
| | تسجيل تماثلي 1 و 2 و 3 مراحل | - |
| | تقليم البيانات | - |

9.3 DB3

- | | | |
|--|----------------------------------|---|
| | NICAM | - |
| 192 kbit/s, stereo (joint stereo coding) | MiniDisc and MiniDisc + Layer II | - |

256 kbit/s stereo, 1, 3, 5, 7, and 9 stages	Dolby AC2	–
selection from Swisscom database, >192 kbit/s stereo	MPEG1 Layer II	–
128 and 160 kbit/s both stereo (joint stereo coding)	MPEG1 Layer III	–
128 kbit/s stereo (joint stereo coding)	MPEG AAC	–
128 + Layer II, 384 + Layer II, 224 kbit/s, all stereo	MPEG Layer III	–
256 kbit/s stereo	Dolby AC3	–
256 + MPEG Layer II, 224 kbit/s, both stereo	Dolby AC3	–
	التقدير الكمي للتشوه	–
	التشوه الكلي الناجم عن التوافقيات	–
	ضوضاء	–
	CRC97	10.3
64, 96, 128, and 160 kbit/s, all stereo	AT&T PAC	–
128, 160, and 192 kbit/s, all stereo	Dolby AC3	–
128, 160, and 192 kbit/s, all stereo	MPEG1 Layer II software	–
96, 128, 160, 192 kbit/s, all stereo	MPEG1 Layer II hardware (ITIS)	–
96 and 128 kbit/s, both stereo	MPEG4 AAC	–
128 kbit/s stereo	MPEG1 Layer III	–

العناصر لكل حالة بالنسبة لقاعدتي البيانات DB3 و DB2 4

DB2 1.4

العناصر	رقم الحالة	موقع الاختبار I، NHK، اليابان
CLA,RYC,SB1,STR	CO13	1 stage Layer II, 256 kbit/s
CLA,RYC,SB1,STR	CO11	3 stages
CLA,RYC,SB1,STR	CO19	5 stages
CLA,RYC,SB1,STR	CO18	7 stages
CLA,RYC,SB1,STR	CO15	9 stages
CAS,RYC,STR,WIN	CO1A	1 stage NBC (Dolby AC2)
CAS,RYC,STR,WIN	CO12	3 stages
CAS,RYC,STR,WIN	CO17	5 stages
CAS,RYC,STR,WIN	CO16	7 stages
CAS,RYC,STR,WIN	CO14	9 stages
العناصر	رقم الحالة	موقع الاختبار II، DR، الدانمارك
CLA,RYC,SB1,STR	CO2B	1 stage Layer II, 256 kbit/s
CLA,RYC,SB1,STR	CO25	Layer II, 192 kbit/s js
MLA,MPE,MTR,MYC	CO27	Layer II, 64 kbit/s mono
CAS,RYC,STR,WIN	CO29	5 stages NBC (Dolby AC2)
MLA,MPE,MTR,MYC	CO22	MPEG2/L2 LSF
PER	CO23	Analogue 1
PER	CO2A	Analogue 2
PER	CO28	Analogue 3
GLO,HRN,TRI	CO24	Errors 1
GLO,HRN,TRI	CO21	Errors 2
BAS,CL2,TUB	CO26	Clipping

موقع الاختبار III، NRK، الترويج

CLA,RYC,SB1,STR	CO34	Layer II, 384 kbit/s
CLA,RYC,SB1,STR	CO31	1 stage Layer II, 256 kbit/s
CAS,RYC,STR,WIN	CO3B	5 stages NBC (Dolby AC2)
CLA,STR,TAM,VEG		CO32 Layer III (ASPEC3), 192 kbit/s
CLA,STR,TAM,VEG		CO39 Layer III (ASPEC3), 128 kbit/s
CLA,STR,TAM,VEG		CO3A Layer III (ASPEC3), 160 kbit/s
HAR,SB2,STR,TPT	CO33	APT-X, 256 kbit/s
HAR,SB2,STR,TPT	CO36	APT-X, 384 kbit/s
DRU	CO35	Quantizing dist. 1
DRU	CO37	Quantizing dist. 2
DRU	CO38	Quantizing dist. 3

عناصر الاختبار

الموسيقى الشعبية السويدية، تسجيل SR، تم استخدامه سابقاً	STR
مزمار القرية، تسجيل SR	SB1
مزمار القرية، تسجيل SR	SB2
الكلارينيت، SQUAM 16/2	CLA
الدف الصغير، تسجيل SR، تم استخدامه سابقاً	TAM
سترافينسكي، مجموعة الرياح، تم استخدامه سابقاً	WIN
البوق، SQUAM 21/2	TPT
مجموعة المهرج، تسجيل بي بي سي G 49/17	HAR
سوزان فيجا، المعلم القديم، تم استخدامه سابقاً	VEG
صنجات، SQUAM 27	CAS
كلام بالألمانية، SQUAM 54	SPE
راي كودر، الأسطوانة CD: JAZZ tr 11 (0.25 – 0.47)	RYC
قرع طبول، باس ماريمبا الياباني، الأسطوانة CD: Sony/CBS 32DC 5027	PER
النفير، SQUAM 54	HRN
الإكسيليفون، SQUAM 35/1، تم استخدامه سابقاً	GLO
المثلث الرنان، SQUAM 32/2	TRI
الطبول، SQUAM 28	DRU
الكلارينيت، SQUAM 16/2	CL2
باس كلابريت، SQUAM 17	BAS
توبا، SQUAM 24	TUB
مزيج أحادي القناة من SPE	MPE
مزيج أحادي القناة من STR	MTR
مزيج أحادي القناة من CLA	MLA
مزيج أحادي القناة من RYC	MYC

Test/Item	Clarinet	Clarinet mono	Ry Cooder	Ry Cooder mono	Bag Pipes 1	Strings	Strings mono	Castanettes	Wind Ensemble	MPE mono	Tambourine	Suzanne Vega	Hapsichord	Bag Pipes 2	Trumpet	Drum	Percussion	Glockenspiel	Hom	Triangle	BAS	Clarinet 2	Tuba
MPEG1 Layer 2, 256 kbit/s, 1 stage	X		X		X	X																	
MPEG1 Layer 2, 256 kbit/s, 3 stages	X		X		X	X																	
MPEG1 Layer 2, 256 kbit/s, 5 stages	X		X		X	X																	
MPEG1 Layer 2, 256 kbit/s, 7 stages	X		X		X	X																	
MPEG1 Layer 2, 256 kbit/s, 9 stages	X		X		X	X																	
Dolby AC2, 256 kbit/s, 1 stage			X			X		X	X														
Dolby AC2, 256 kbit/s, 3 stages			X			X		X	X														
Dolby AC2, 256 kbit/s, 5 stages			X			X		X	X														
Dolby AC2, 256 kbit/s, 7 stages			X			X		X	X														
Dolby AC2, 256 kbit/s, 9 stages			X			X		X	X														
MPEG1 Layer 2, 192 kbit/s joint stereo	X		X		X	X																	
MPEG1 Layer 2, 64 kbit/s mono		X		X			X			X													
MPEG2 Layer 2, 64 kbit/s mono		X		X			X			X													
MPEG1 Layer 2, 384 kbit/s	X		X		X	X																	
MPEG1 Layer 3 (ASPEC 3), 192 kbit/s	X					X					X	X											
MPEG1 Layer 3 (ASPEC 3), 128 kbit/s	X					X					X	X											
MPEG1 Layer 3 (ASPEC 3), 160 kbit/s	X					X					X	X											
APT-X, 256 kbit/s						X							X	X	X								
APT-X, 384 kbit/s						X							X	X	X								
Quantizing distorsion 1																	X						
Quantizing distorsion 2																	X						
Quantizing distorsion 3																	X						
Analogue recording 1 stage																	X						
Analogue recording 2 stages																	X						
Analogue recording 3 stages																	X						
Bit errors 1																		X	X	X			
Bit errors 2																		X	X	X			
Clipping																					X	X	X

قاعدة البيانات DB3 2.4

Item/Test	Name	1	2MD	2MDL2	3 1Step	3 3Step	3 5Step	3 7Step	3 9Step	4	5	6 low	6 high	7	8	9	10	11	12	13	
																					1.) NICAM
																					2.) MD and MD + L2 (192 kbit/s)
																					3.) AC2 (256 kbit/s), 1, 3, 5, 7, 9
																					4.) L2 (ST d-b, *192 kbit/s)
13(1) Flute	flu							X	X												5.) Layer2 (256 kbit/s), 8 stages
16(2) Clarinet	cia				X		X			X	X			X	X						6.) Layer3, 128 and 160 kbit/s
20(1) Saxophon	sax							X	X												7.) AAC, 128 kbit/s
21(2) Trumpet	tpt					X	X														8.) L3 (128) + L2 (384) + L2 (224)
24(2) Tuba	tub																		X		9.) AC3 (256)
26(1) Claves	civ		X	X			X	X	X												10.) AC3 (256) + L2 (224)
27 Castanets	cas									X		X			X	X					11.) Quantizing distortion
28 Snare drum	sna	X																			12.) THD
30 Kettle drums	ket																			X	13.) Noise
32(1/2) Triangle	tri		X	X	X	X				X											
35(1/2) Glockenspiel	glo										X	X									
36(1) Xylophone	xyl																				X
40(1) Harpsicord	hrp				X	X				X				X							
49 Speech female engl	sfe							X	X												X
54 Speech male germ	smg	X	X	X	X	X	X			X											
60 Piano Schubert	pia																			X	X
61 Soprano Mozart	sop		X	X																	
53 Speech female germ	sfg																				
50 Speech male engl	sme												X		X						
Ref_tam	tam	X														X	X				
Ref_str	str															X	X				
Ref_har	har																				
Ry Cooder	ryc									X						X	X				
Susanne Vega	veg									X	X	X			X	X					
Pitch Pipe	pip											X	X	X	X						
Marimba	mar	X																			X
Bag Pipe	sb1									X											
Name		i	b	s	g	e	p	c	o	f	j	r	k	l	m	h	d	q	n	a	

Name Examples:	Reference:	irefflu
	Test:	icodflu

مسرد المصطلحات

درجة الخطأ المطلق (AES)

تُشتق درجة الخطأ المطلق من معادلة وُضعت خصيصاً لتقييم جودة النتائج المتحصل عليها من طريقة قياس إدراكية موضوعية. وهي تأخذ في الاعتبار فواصل الثقة للتقييم المتوسطة لاختبارات الاستماع الشخصية.

جودة الصوت الأساسية

تُعرف جودة الصوت الأساسية بأنها خاصية شخصية عالمية تتضمن أي من وجميع الفروق المكتشفة بين الإشارة المرجعية والصيغة المعالجة منها.

هامش التشفير

هامش التشفير هو معلمة جودة تقيس المساحة العالية لنواقص التشفير غير المسموعة حتى العتبة عندما تصبح هذه النواقص مسموعة.

متغيرات خرج النموذج (MOV)

متغيرات خرج النموذج هي قيم الخرج المتوسطة لطريقة القياس الإدراكي. وتستند هذه المتغيرات إلى النتائج النفسية الصوتية الأساسية، وبالتالي يمكن استخدامها لتحديد خصائص نواقص التشفير بشكل أكبر.

درجة الاختلاف الموضوعية (ODG)

درجة الاختلاف الموضوعية هي معلمة الخرج الرئيسية لطريقة القياس الإدراكي. وهي تقابل الدرجات SDG وهي معلمة القياس تعطي جودة الصوت الأساسية العالمية. ويتراوح نطاق الدرجات ODG بين 0 و 4.

القياس خارج الخط

إجراء القياس الذي لا يتفاعل مع إرسال البرنامج الجاري.

القياس على الخط

إجراء القياس الذي يعتمد على إرسال البرنامج الجاري، أو أجزاء منه.

درجة الاختلاف الشخصية (SDG)

في اختبار الاستماع وفقاً للتوصية ITU-R BS.1116، يتم تصنيف الجودة الصوتية الأساسية للمرجع المخفي والنسخة المعالجة من المرجع على سلم الخطاط من خمس درجات. وتُعرف درجة الاختلاف بأنها الدرجة المعطاة للإشارة قيد الاختبار مطروحاً منها الدرجة المعطاة للإشارة المرجعية. وينبغي لنطاق نطاق الدرجات SDG المثالي أن يتراوح بين 0 و-4. إذا لم يُحدد المرجع بشكل صحيح، تكون الدرجة SDG موجبة.

المختصرات

متوسط المجموعة المشوهة (<i>Average Distorted Block</i>)	ADB
درجة الخطأ المطلق (<i>Absolute Error Score</i>)	AES
الفرق الطيفي السمعي (<i>Auditory Spectral Difference</i>)	ASD
متوسط (خطي) (<i>Average (linear)</i>)	Avg
جودة الصوت الأساسية (<i>Basic Audio Quality</i>)	BAQ
عرض النطاق (<i>Bandwidth</i>)	Bw
فاصل الثقة (<i>Confidence Interval</i>)	CI
هامش التشفير (<i>Coding Margin</i>)	CM
قاعدة البيانات رقم n (1 أو 2 أو 3) (<i>Database n (1, 2 or 3)</i>)	DBn
تيار مستمر (<i>Direct Current</i>)	DC
تحويل فورييه المتمايز (<i>Discrete Fourier Transform</i>)	DFT
مؤشر الاضطراب (<i>Disturbance Index</i>)	DIX
الجهاز قيد الاختبار (<i>Device Under Test</i>)	DUT
البنية التوافقية للخطأ (<i>Error Harmonic Structure</i>)	EHS
عرض النطاق المستطيل المكافئ (<i>Equivalent Rectangular Bandwidth</i>)	ERB
عامل (<i>factor</i>)	fac
تحويل فورييه السريع (<i>Fast Fourier Transform</i>)	FFT
الاستجابة النبضية المحدودة (<i>Finite Impulse Response</i>)	FIR
استجابة نبضية غير محدودة (<i>Infinite Impulse Response</i>)	IIR
المنظمة الدولية للتوحيد القياسي (<i>International Standards Organization</i>)	ISO

فرق المستوى الملحوظ بالكاد (<i>Just Noticeable Level Difference</i>)	JNLD
الحد الأقصى لاحتمال الكشف المصفى (<i>Maximum Filtered Probability of Detection</i>)	MFPD
متغير خرج النظام (<i>Model Output Variable</i>)	MOV
فريق خبراء الصور المتحركة (<i>Moving Picture Expert Group</i>)	MPEG
جهازة الضوضاء (<i>Noise Loudness</i>)	NL
النسبة ضوضاء إلى قناع (<i>Noise-To-Mask Ratio</i>)	NMR
التقييم الموضوعي للإشارة الصوتية (<i>Objective Audio Signal Evaluation</i>)	OASE
هامش التشفير الموضوعي (<i>Objective Coding Margin</i>)	OCM
درجة الاختلاف الموضوعية (<i>Objective Difference Grade</i>)	ODG
قياس جودة الصوت الإدراكية (<i>Perceptual Audio Quality Measure</i>)	PAQM
القياسات الموضوعية لجودة الصوت المدركة (<i>Objective Measurements of Perceived Audio Quality</i>)	PEAQ
التقييم الإدراكي (<i>Perceptual Evaluation</i>)	PERCEVAL
القياس الموضوعي الإدراكي (<i>Perceptual Objective Measure</i>)	POM
الإشارة المرجعية (<i>Reference Signal</i>)	REF
الدقة (<i>Resolution</i>)	res
جذر متوسط التربيع (<i>Root Mean Squared</i>)	r.m.s.
أسي مقرب (<i>Rounded Exponential</i>)	ROEX
معدل قيم الخرج (<i>Rate of Output Values</i>)	ROV
هامش التشفير الشخصي (<i>Subjective Coding Margin</i>)	SCM
درجة الاختلاف الشخصية (<i>Subjective Difference Grade</i>)	SDG
النسبة إشارة إلى ضوضاء (<i>Signal-to-Noise Ratio</i>)	SNR
مستوى ضغط الصوت (<i>Sound Pressure Level</i>)	SPL
الإشارة قيد الاختبار (<i>Signal under Test</i>)	SUT
التشوه الكلي الناجم عن التوافقيات (<i>Total Harmonic Distortion</i>)	THD
متوسط مقسم إلى نوافذ (<i>Windowed Average</i>)	Win

المراجع

- AURAS, W. [September 1984] *Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse*. Dissertation an der Fakultät für Elektrotechnik der Technischen Universität München, Federal Republic of Germany.
- BEERENDS, J. G. and STEMERDINK, J. A. [December, 1992] A perceptual audio quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, Vol. 40, p. 963-978.
- BEERENDS, J. G. and STEMERDINK, J. A. [February, 1994] Modeling a cognitive aspect in the measurement of the quality of music codecs. Contribution to the 96th AES Convention, preprint 3800. Amsterdam, Netherlands.
- BEERENDS, J. G. and STEMERDINK, J. A. [March, 1994] A perceptual speech quality measure based on a psycho-acoustic sound representation. *J. Audio Eng. Soc.*, Vol. 42, p. 115-123.
- BEERENDS, J. G., van den BRINK, W. A. C. and RODGER, B. [May, 1996] The role of informational masking and perceptual streaming in the measurement of music codec quality. Contribution to the 100th AES Convention, preprint 4176. Copenhagen, Denmark.
- BRANDENBURG, K. [1987] Evaluation of quality for audio encoding at low bit rates. Contribution to the 82nd AES Convention, preprint 2433. London, United Kingdom.
- BREGMAN, A. S. [1990] *Auditory Scene Analysis: The Perceptual Organisation of Sound*. MIT Press, Cambridge MA, United States of America.
- COHEN, E. A. and FIELDER, L. D. [May, 1992] Determining noise criteria for recording environments. *J. Audio Eng. Soc.*, Vol. 40, p. 384-402.
- COLOMES, C., LEVER, M., RAULT, J. B. and DEHERY, Y. F. [April, 1995] A perceptual model applied to audio bit-rate reduction. *J. Audio Eng. Soc.*, Vol. 43, p. 233-240.
- FEITEN, B. [March, 1997] Measuring the Coding Margin of Perceptual Codecs with the Difference Signal. 102nd AES-Convention, preprint 4417. Munich, Federal Republic of Germany.
- GRUSEC, T., THIBAUT, L. and SOULODRE, G. [September, 1997] EIA/NRSC DAR systems subjective tests. Part 1: Audio codec quality. *IEEE Transactions on Broadcasting*, Vol. 43, 3.
- KARJALAINEN, J. [March, 1985] A new auditory model for the evaluation of sound quality of audio system. Proceedings of the ICASSP, p. 608-611. Tampa, Florida, United States of America.
- LEEK, M. R. and WATSON, C. S. [1984] Learning to detect auditory pattern components. *J. Acoust. Soc. Am.*, Vol. 76, p. 1037-1044.
- MEARES, D. J. and KIM, S. W. [July, 1995] "NBC time/frequency module subjective tests: overall results", ISO/IEC JTC 1/SC 29/WG 11 N0973 MPEG95/208.
- MOORE, B. C. [1986] *Frequency Selectivity in Hearing*. Academic Press, London, United Kingdom.
- MOORE, B. C. [1989] *An introduction to the psychology of hearing*. Academic Press, London, United Kingdom.
- PAILLARD, B., MABILIEAU, P., MORISETTE, S. and SOUMAGNE, J. [1992] Perceval: Perceptual evaluation of the quality of audio signals. *J. Audio Eng. Soc.*, Vol. 40, p. 21-31.
- SCHROEDER, M. R., ATAL, B. S. and HALL, J. L. [December 1979] Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am.*, Vol. 66, p. 1647-1652.
- SOULODRE, G., GRUSEC, T., LAVOIE, M. and THIBAUT, L. [March 1998] Subjective evaluation of state-of-the-art 2-channel audio codecs. *J. Audio Eng. Society*.
- SPORER, T. [October 1997] Objective audio signal evaluation – applied psychoacoustics for modeling the perceived quality of digital audio. 103rd AES-Convention, preprint 4512. New York, United States of America.

- TERHARDT, E. [1979] Calculating Virtual Pitch, Hearing Research. Vol. 1, p. 155-182.
- THIEDE, T. and KABOT, E. [1996] A New Perceptual Quality Measure for Bit Rate Reduced Audio. Contribution to the 100th AES Convention, preprint 4280. Copenhagen, Denmark.
- TREURNIET, W. C. [1996] Simulation of individual listeners with an auditory model. Proceedings of the Audio Engineering Society, Reprint Number 4154. Copenhagen, Denmark.
- von BISMARCK, G. [1974] Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30, p. 159-172.
- ZWICKER, E. and FASTL, H. [1990] *Psycho-acoustics, Facts and Models*. Berlin; Heidelberg: Springer Verlag, Federal Republic of Germany.
- ZWICKER, E. and FELDTKELLER, R. [1967] *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel Verlag, Federal Republic of Germany.

بيبلوغرافيا

- GRUSEC T., THIBAUT L. and SOULODRE, G. [1995] Subjective evaluation of high quality audio coding systems: methods and results in the two-channel case. Preprint 4065 (F-5), Proceedings of the AES. New York, United States of America.
-