الاتحـــــاد الـــدولـــــي للاتصـــــــالات

ITU-T السلسلة P

الإضافة 24

(2005/10)

قطـــاع تقييس الاتصـــالات
في الاتحاد الدولي للاتصالات

السلسلة P: جودة الإرسال الهاتفي والمنشآت الهاتفية
وشبكات الخطوط المحلية

معلمات وصف التفاعل مع أنظمة الحوار الأوتوماتية

توصيات السلسلة P – ITU-T – الإضافة 24

توصيات السلسلة P الصادرة عن قطاع تقييس الاتصالات

جودة الإرسال الهاتفي والمنشآت الهاتفية وشبكات الخطوط المحلية

مفردات وتأثيرات معلمات الإرسال على رأي الزبائن في جودة الإرسال	السلسلة P.10
خطوط المشتركين وأجهزتهم	السلسلة P.30
	P.300
معايير الإرسال	السلسلة P.40
أجهزة القياسات الموضوعية	السلسلة P.50
	P.500
قياسات كهرسمعية موضوعية	السلسلة P.60
قياسات متصلة بالجهارة الصوتية	السلسلة P.70
طرائق التقدير الموضوعي والشخصي لنوعية الأداء	السلسلة P.80
	السلسلة P.800
نوعية الأداء السمعي البصري في الخدمات متعددة الوسائط	السلسلة P.900
أداء الإرسال وجوانب نوعية الخدمة في النقاط الانتهائية للشبكات القائمة على بروتوكول الإنترنت	السلسلة P.1000

لمزيد من التفاصيل يرجى الرجوع إلى قائمة التوصيات الصادرة عن قطاع تقييس الاتصالات.

الإضافة 24 إلى توصيات السلسلة ITU-T P

معلمات وصف التفاعل مع أنظمة الحوار الأوتوماتية

الملخص

يرد في هذه الإضافة تعاريف لمجموعة معلمات يمكن استنباطها من خدمات تعتمد على أنظمة حوار صوتية. ويمكن استخلاص المعلمات من التفاعلات بين مستعمل مسجل (يشارك في اختبار معين) والخدمة قيد البحث. وتسمح المعلمات بتقدير كمية تدفق التفاعلات وسلوك المستعمل والنظام، ونوعية أداء أجهزة التكنولوجيا الصوتية المشاركة في التفاعل. وتوفر هذه المعلمات معلومات مفيدة فيما يتعلق بتطوير الأنظمة واستخدامها على النحو الأمثل وصيانتها، وهي مكملة للتقديرات الشخصية بشأن الجودة والمجمعة وفقاً للتوصية ITU-T P.851.

المصدر

وافقت لجنة الدراسات 12 (2005-2008) لقطاع تقييس الاتصالات بتاريخ 21 أكتوبر 2005 على الإضافة 24 لسلسلة التوصيات ITU-T P.

عبارات أساسية

التقييم، التعرف على الصوت أوتوماتياً، فهم الكلام أوتوماتياً، إدارة الحوار، معلمة تفاعل، توليد الكلام، تكنولوجيا الكلام، نظام حوار صوتي.

تمهيـــد

الاتحاد الدولي للاتصالات وكالة متخصصة للأمم المتحدة في ميدان الاتصالات. وقطاع تقييس الاتصالات (ITU-T) هو هيئة دائمة في الاتحاد الدولي للاتصالات. وهو مسؤول عن دراسة المسائل التقنية والمسائل المتعلقة بالتشغيل والتعريفة، وإصدار التوصيات بشأنها بغرض تقييس الاتصالات على الصعيد العالمي.

وتحدد الجمعية العالمية لتقييس الاتصالات (WTSA)، التي تجتمع مرة كل أربع سنوات، المواضيع التي يجب أن تدرسها لجان الدراسات التابعة لقطاع تقييس الاتصالات وأن تُصدر توصيات بشأنها.

وتتم الموافقة على هذه التوصيات وفقاً للإجراء الموضح في القرار رقم 1 الصادر عن الجمعية العالمية لتقييس الاتصالات.

وفي بعض مجالات تكنولوجيا المعلومات التي تقع ضمن اختصاص قطاع تقييس الاتصالات، تعد المعايير اللازمة على أساس التعاون مع المنظمة الدولية للتوحيد القياسي (ISO) واللجنة الكهرتقنية الدولية (IEC).

ملاحظة

تستخدم كلمة "الإدارة" في هذه التوصية لتدل بصورة موجزة سواء على إدارة اتصالات أو على وكالة تشغيل معترف بها.

والتقيد بهذه التوصية اختياري. غير أنها قد تضم بعض الأحكام الإلزامية (بهدف تأمين قابلية التشغيل البيني والتطبيق مثلاً). ويعتبر التقيّد بهذه التوصية حاصلاً عندما يتم التقيّد بجميع هذه الأحكام الإلزامية. ويستخدم فعل "يجب" وصيغ ملزمة أخرى مثل فعل "ينبغي" وصيغها النافية للتعبير عن متطلبات معينة، ولا يعني استعمال هذه الصيغ أن التقيّد بهذه التوصية إلزامي.

حقوق الملكية الفكرية

يسترعي الاتحاد الانتباه إلى أن تطبيق هذه التوصية أو تنفيذها قد يستلزم استعمال حق من حقوق الملكية الفكرية. ولا يتخذ الاتحاد أي موقف من القرائن المتعلقة بحقوق الملكية الفكرية أو صلاحيتها أو نطاق تطبيقها سواء طالب بها عضو من أعضاء الاتحاد أو طرف آخر لا تشمله عملية إعداد التوصيات.

وعند الموافقة على هذه التوصية، لم يكن الاتحاد قد تلقى إخطاراً بملكية فكرية تحميها براءات الاختراع يمكن المطالبة بها لتنفيذ هذه التوصية. ومع ذلك، ونظراً إلى أن هذه المعلومات قد لا تكون هي الأحدث، يوصى المسؤولون عن تنفيذ هذه التوصية بالاطلاع على قاعدة المعطيات الخاصة ببراءات الاختراع في مكتب تقييس الاتصالات (TSB).

© ITU 2005

جميع الحقوق محفوظة. لا يجوز استنساخ أي جزء من هذه المنشورة بأي وسيلة كانت إلا بإذن خطي مسبق من الاتحاد الدولي للاتصالات.

المحتويات

الصفحة

1 مجال التطبيق 1
2 المراجع..... 1
3 التعاريف..... 1
4 المختصرات 2
5 مقدمة 3
6 خصائص معلمات التفاعل..... 4
7 استعراض معلمات التفاعل..... 5
8 تفسير قيم معلمات التفاعل 15
بيبليوغرافيا..... 16

الإضافة 24 إلى توصيات السلسلة ITU-T P

معلمات وصف التفاعل مع أنظمة الحوار الأوتوماتية

1 مجال التطبيق

تصف هذه الإضافة معلمات تقدم معلومات عن التفاعل مع خدمات تستعمل أنظمة الحوار الأوتوماتية، وذلك من وجهة نظر مُطور النظام ومشغل الخدمة. وتتيح أنظمة الحوار الأوتوماتية التي تتناولها هذه الإضافة التفاعل بلغة منطوقة مع مستعمل بشري على أساس التناوب في الدور بواسطة الشبكة الهاتفية، وامتلاك قدرات التعرف على الصوت أوتوماتياً، وفهم الكلام، وإدارة الحوار، وتوليد الاستجابة، وإرسال الكلام. يمكنها أن تسهل النفاذ إلى المعلومات المخزنة في قواعد البيانات، أو تسمح بإجراء مختلف أنواع المعاملات.

وتقدر المعلمات المحددة في هذه الإضافة كمية تدفق التفاعلات وسلوك المستعمل والنظام، ونوعية أداء أجهزة التكنولوجيا الصوتية المشاركة في التفاعل. ومن أجل استخراج جميع هذه المعلمات، ينبغي أن يتسنى النفاذ إلى أنظمة الحوار الأوتوماتية بوصفها صندوقاً زجاجياً؛ ومع ذلك، يمكن استخراج بعض المعلمات باتباع نهج الصندوق الأسود، أي، دون النفاذ إلى مختلف مكونات النظام. ويمكن استخراج جزء من المعلمات أوتوماتياً، ويعتمد الأمر جزئياً على الخبرة البشرية في ميدان نسخ ملفات تسجيل التفاعلات واستنباط حاشية تفسيرية. وتتناول المعلمات نوعية أداء النظام من وجهة نظر الجهة التي قامت بتطويره؛ وبالتالي، فهي تقدم معلومات تكمل تجارب التقييم الشخصي بالتلازم مع أنظمة الحوار الأوتوماتية التي ترد بشأنها توجيهات في التوصية P.851 الصادرة عن قطاع تقييس الاتصالات في الاتحاد (ITU-T). ويتيسر في التوصيتين P.800 وP.85 الصادرتين عن القطاع ITU-T وفي كتيب المهاتفة، المزيد من التوجيهات بشأن طرائق التقييم الشخصي عموماً وبشأن تقييم أجهزة إرسال الكلام. والمعلمات الواردة في هذه الإضافة لا تشير تحديداً إلى حالات الانحطاط التي يحتمل أن تحدثها قناة الإرسال. وهذه الآثار بند يخضع للمزيد من البحث من جانب لجنة الدراسات 12 التابعة لقطاع تقييس الاتصالات في الاتحاد (ITU-T).

2 المراجع

- التوصية ITU-T P.85 (1994)، *طريقة تقييم شخصي لجودة كلام أجهزة وحدات الخدمة الصوتية*.
- التوصية ITU-T P.800 (1996)، *طرق التقدير الشخصي لنوعية الإرسال*.
- التوصية ITU-T P.851 (2003)، *التقييم الشخصي لجودة الخدمات الهاتفية القائمة على أنظمة الحوار الأوتوماتية*.
- *كتيب المهاتفة* ITU-T (1992).

3 التعاريف

يُرجى الرجوع إلى التوصية ITU-T P.10 للاطلاع على التعاريف غير الواردة هنا.

1.3 التدخل: قدرة شخص على التحدث من خلال ملقن نظام أو خرج نظام ما [10].

2.3 الحوار: هو محادثة أو تبادل للمعلومات. وبالنظر إليه بوصفه وحدة تقييم: فهو مسير من عدة مسيرات ممكنه تمر ببنية الحوار.

3.3 **الكفاءة**: قياسات دقة واكتمال مهام النظام المتعلقة بالموارد (مثل الزمن، الجهد البشري) المستعملة لتحقيق مهام النظام المحددة.

4.3 **التبادل**: هو زوج من الأدوار المتعاقبة والمترابطة، يأخذهما طرفا الحوار [8].

5.3 **العنصر الوظيفي**: هي قدرة النظام على توفير وظائف تلبي الاحتياجات المذكورة صراحة وضمناً عند استعمال النظام في ظل ظروف معينة.

6.3 **الاتصالات الوصفية**: هو الاتصال من أجل توصيل رسالة، أي من أجل إزالة سوء الفهم مثل ("هل فهمتك تماما؟") أو من أجل التوصل إلى اتفاق بشأن استعمال اللغة.

7.3 **الأداء**: قدرة وحدة معينة على أداء الوظيفة التي صُممت لأجلها.

8.3 **التكنولوجيا الصوتية**: نظام يُعنى بالبحث والتطوير في ميدان أنظمة دخل وخرج لغة الحديث، وذلك باستعمال المساهمات المقدمة من الأنظمة المجاورة مثل الصوتيات، والهندسة الكهربائية، والإحصاءات، وعلم الأصوات، والمعالجة بلغة طبيعية، وبتحديد متطلبات النظام، والتصميم والتنفيذ والتقييم، ومعالجة مجموعات الوثائق والموارد اللغوية، وتقييم النواتج الموجهة إلى المستهلك [10].

9.3 **نظام الحوار الأوتوماتي**: هو نظام حاسوبي يتفاعل معه مستعملون بشريون على أساس التناوب في الأدوار بواسطة لغة الحديث.

10.3 **المهمة**: هي جميع الأنشطة التي يجب أن يضطلع بها المستعمل من أجل بلوغ هدف ثابت في مجال معين.

11.3 **الحوار الموجه إلى المهمة**: هو حوار يتعلق بموضوع معين يرمي إلى تحقيق هدف واضح (من قبيل حل مشكلة ما أو الحصول على معلومات معينة) [8].

12.3 **المعاملة**: جزء من حوار يُكرس لأداء مهمة وحيدة رفيعة المستوى (مثل الحجز للسفر أو التحقق من رصيد حساب مصرفي). وقد تتزامن المعاملة مع حوار معين، أو قد يتألف الحوار من أكثر من معاملة واحدة [8].

13.3 **الدور**: التفوه. مدة متواصلة من كلام ينطق به أحد طرفي الحوار اعتباراً من اللحظة التي يبدأ فيها هذا الطرف بالحديث حتى اللحظة التي يأخذ فيها قطعاً الطرف الآخر دوره في الحديث [1].

14.3 **التفوه**: انظر الدور.

4 المختصرات

ASR التعرف على الصوت أوتوماتياً *(Automatic Speech Recognition)*

AVM مصفوفة نعوت–قيم *(Attribute-Value Matrix)*

AVP زوج نعوت– قيم *(Attribute-Value Pair)*

DARPA وكالة تطوير مشاريع البحوث المتقدمة التابعة لوزارة الدفاع الأمريكية

DP برمجة دينامية *(Dynamic Programming)*

DTMF تردد متعدد بنغمة مزدوجة *(Dual Tone Multiple Frequency)*

IVR استجابة صوتية تفاعلية *(Interactive Voice Response)*

MOS متوسط علامة الرأي *(Mean Opinion Score)*

SDS نظام حوار أوتوماتي *(Spoken Dialogue System)*

WoZ ساحر Oz *(Wizard-of-Oz)*

5 مقدمة

أنظمة الحوار الأوتوماتية (SDS)، أي، الأنظمة الحاسوبية التي يتفاعل معها مستعملون بشريون على أساس التناوب في الأدوار بواسطة لغة الحديث، هي أنظمة يمكن أن تشكل جزءاً من الشبكات الهاتفية الحديثة. وهي تتيح النفاذ إلى قواعد البيانات والمعاملات بواسطة الهاتف، وذلك من أجل الحصول مثلاً على معلومات عن جدول مواعيد رحلات القطارات والطائرات، أو عن أسعار صرف العملات، أو عن السياحة، أو من أجل إجراء عمليات حسابية ومصرفية أو الحجز في الفنادق. وعلى النقيض من أنظمة الاستجابة الصوتية التفاعلية (IVR) بدخل التردد المتعدد بنغمة مزدوجة (DTMF)، توفر أنظمة الحوار الأوتوماتية (SDS) طائفة كاملة من قدرات التفاعل بواسطة الكلام التي تشمل التعرف على صوت المستعمل، وإعطاء معاني للكلمات المتعرف عليها، واتخاذ قرار بشأن كيفية مواصلة الحوار، وتشكيل إجابة لغوية، وتوليد ناتج تكلم لاسترعاء انتباه المستعمل. وبذلك يمكن التفاعل بواسطة التكلم بين المستعمل والنظام بشكل "طبيعي" إلى حد ما.

ولتقييم نوعية الخدمات التي تستعمل أنظمة SDS من وجهة نظر المستعمل، قامت لجنة الدراسات 12 (SG 12) التابعة لقطاع تقييس الاتصالات في الاتحاد (ITU-T) بإعداد التوصية ITU-T P.851 في عام 2003. وتصف هذه التوصية طرائق لإجراء تجارب التقييم الشخصي من أجل تحديد النوعية من *منظور المستعمل*، على اعتبار أنظمة SDS صندوقاً أسود. وبفضل إجراء تجارب وفقاً للتوصية ITU-T P.851، يمكن الحصول على معلومات قيمة عن النوعية حسب ما يراها المستعمل. ومع ذلك، فقد يصعب تحديد الكيفية التي يساهم بها كل مكون من مكونات النظام في النوعية بشكل عام التي يدركها المستعمل، مثلاً، في تحديد المكونات التي تحتاج إلى تحسين في حالة حصول مشاكل في التفاعل. وعليه، ينبغي استكمال التقييم بمعلومات تتناول أداء النظام *من منظور مصممه ومن منظور مشغل الخدمة*.

ويمكن وصف المعلومات المتعلقة بالنظام على أساس ما يُسمى *معلمات التفاعل*. وتساعد هذه المعلمات في تقدير كمية تدفق التفاعلات وسلوك المستعمل والنظام، وفي تقييم أداء أجهزة التكنولوجيا الصوتية المشاركة في التفاعل. وتتطرق المعلمات إلى أداء النظام من وجهة نظر الجهة المستحدثة للنظام ومشغل الخدمة، وتوفر بالتالي معلومات تكمل بيانات التقييم الشخصي. ومن أجل استخلاص بعض هذه المعلمات، ينبغي أن يتسنى النفاذ إلى أنظمة الحوار الأوتوماتية بوصفها صندوقاً زجاجياً؛ كما يمكن استنباط بعضها الآخر باتباع نهج الصندوق الأسود، أي، دون النفاذ إلى مختلف مكونات النظام.

وتوفر هذه الإضافة مجموعة من معلمات التفاعل استُعملت خلال السنوات الخمسة عشر الماضية في تقييم أنظمة (SDS). وتتعلق المعلمات المذكورة بتبادل المعلومات بشكل عام بين المستعمل والنظام، والاتصالات الوصفية في حالات سوء الفهم، ومدى تعاون النظام، والمهمة التي يمكن أداؤها بمساعدة النظام، وقدرات دخل الكلام الخاصة بالنظام. ومع ذلك، لا يوجد وصف معلمي بعد لنوعية خرج الكلام (مثلاً، فيما يخص نوعية الكلام التركيبـــي). وتستند مجموعة المعلمات إلى الأعمال النظرية التي يرد وصف لها في المرجع [17].

ولا ترتبط جميع معلمات التفاعل بصلة مباشرة مع الجودة المحسوسة للخدمات القائمة على أنظمة SDS. وفي الواقع، فإن علاقات الارتباط بين فرادى المعلمات والأحكام التي يصدرها المستعمل بشأن النوعية هي عموماً علاقات معتدلة إلى حدٍّ ما. ومع ذلك، فهي مفيدة في تنظيم مجموعة كبيرة من المعلمات التي تصف التفاعل بين المستعمل والنظام، للحصول بهذه الطريقة على معظم المعلومات التي يُحتمل أن تكون ذات صلة بالجودة المحسوسة من منظور مصمم النظام. وتقدم هذه المعلمات معلومات مفيدة عن تطوير الأنظمة وتحقيق الاستعمال الأمثل وصيانتها.

وفور تحديد هذه المعلمات وتطبيقها في تجارب التقييم بمختلف مواقع الاختبار، فإنها يمكن أن تسهل تقدير الآثار التي تخلفها على النوعية المدركة لطائفة واسعة من الأنظمة والخدمات. وبهذه الطريقة، فقد يتسنى إعداد خوارزميات للتكهن بالنوعية على أساس معلمات التفاعل. ولا يزال العمل في هذا الاتجاه جارياً في لجنة الدراسات 12 (SG 12) التابعة لقطاع تقييس الاتصالات بالاتحاد (ITU-T) وفي هيئات أخرى.

6 خصائص معلمات التفاعل

يمكن استخلاص معلمات التفاعل عندما يتفاعل المستعملون الفعليون أو الاختباريون مع الخدمة. ويمكن إنجاز جزء من الاستخلاص بواسطة الأجهزة وإنجاز الجزء الآخر بالاستعانة بملفات التسجيل التي يتعين على أحد الخبراء أن ينسخها ويستنبط حاشية تفسيرية. والمعلمات البسيطة، من قبيل مدة التفاعل أو التفوه المنفرد، هي معلمات يمكن عادة توخي الكمال في قياسها بواسطة الأجهزة بالتلازم مع تطبيق الخوارزميات الملائمة. ومن جهة أخرى، فإن نسخ الملفات واستخلاص الحواشي التفسيرية من جانب الإنسان أمر ضروري لا يقتصر على حالات التطرق إلى الشكل السطحي (لإشارات الكلام) فحسب، بل يشمل أيضاً محتوى ومعنى تفوه النظام أو المستعمل (مثلاً، تحديد دقة كلمة أو فكرة معينة).

وأنظمة SDS غاية في التعقيد بحيث إن وصف سلوك النظام وإجراء مقارنة بين الأنظمة أو صيغ الأنظمة يحتاج إلى الاستناد إلى عدة معلمات مختلفة متعددة [24]. ونتيجة لذلك، ومن أجل الحصول على أكبر قدر ممكن من المعلومات، ينبغي تطبيق كلتا طريقتي جمع معلمات التفاعل (القائمتين على استعمال الأجهزة والاستعانة بالخبراء). وبناء على المعلومات المجمعة، يمكن تحقيق الاستعمال الأمثل لخدمات الحوار الأوتوماتي وصيانتها بأقصى فعالية.

ونظراً إلى أن معلمات التفاعل تستند إلى بيانات تُجمع في أحد التفاعلات بين المستعمل والنظام، فإنها تتأثر بخصائص النظام، وخصائص المستعمل، والتفاعل بينهما على حد سواء. ويتعذر عادة الفصل بين هذه التأثيرات، لأن سلوك المستعمل يتأثر بشدة بسلوك النظام (مثل الأسئلة التي يطرحها النظام)، والعكس بالعكس (تؤثر مثلاً المفردات التي يستخدمها المستعمل وأسلوب الحديث على مدى قدرة النظام في التعرف وفهم الحديث بدقة). وعليه، فإن معلمات التفاعل تعكس بشدة خصائص مجموعة المستعملين التي جُمعت معها.

وتُحدد معلمات التفاعل إما في بيئة تجريبية بالمختبر في ظل ظروف محددة، أو بإجراء اختبار ميداني. وقد يتعذر في الحالة الأخيرة استخلاص جميع المعلمات، بسبب تعذر جمع المعلومات اللازمة كافة. وإذا تعين مثلاً تحديد مدى نجاح أحد التفاعلات الموجهة إلى أداء مهمة معينة (مثل الحصول على جدول مواعيد القطارات)، فإن من الضروري عندئذ معرفة أهداف المستعمل بدقة. ولا يمكن جمع هذه المعلومات سوى في بيئة مختبرية، مثلاً، باتباع الطريقة التي يرد وصف لها في التوصية ITU-T P.851. وفي حال عدم تيسر النظام المتكامل تماماً بعد، يمكن جمع المعلمات بواسطة المحاكاة المسماة ساحر Oz (‘Wizard-of-Oz’)، حيث يقوم أحد المختبرين باستبدال الأجزاء المفقودة من النظام قيد الاختبار. وينبغي مراعاة خصائص هذه المحاكاة عند تفسير المعلمات التي حُصل عليها.

ويمكن حساب معلمات التفاعل على مستوى الكلمة أو مستوى الجملة أو مستوى التفوه، أو على مستوى تفاعل أو حوار كامل. وغالباً ما تُحسب متوسطات القيم لكل حوار في حالة حساب المعلمات على مستوى الكلمة أو التفوه. وبتطبيق طرائق إحصائية معيارية، يمكن تحليل المعلمات المجمعة مع مجموعة مستعملين معينة على أساس أثر النظام (صيغته)، ومجموعة المستعملين، والبيئة الاختبارية (السيناريوهات، بيئة الاختبار، وما إلى ذلك). ويمكن الاطلاع على تحديد خصائص هذه التأثيرات في التوصية ITU-T P.851.

7 استعراض معلمات التفاعل

وبناء على إجراء مسح واسع للأعمال المنشورة، حُددت معلمات استُعملت في مختلف تجارب التقييم والتقدير خلال السنوات الخمس عشرة الماضية. ويمكن الاطلاع على الأعمال المعنية في الوثائق المرجعية [2]، [3]، [4]، [6]، [7]، [8]، [9]، [11]، [12]، [14]، [16]، [21]، [22]، [23]، [24]، [25]، [26]، [27]، [28]، [30]، [31]، [32]، ويرد ملخص بالمعلمات في المرجع [17]. ويمكن تصنيف المعلمات إجمالاً كما يلي:

- معلمات متصلة بالحوار والاتصال؛
- معلمات متصلة بالاتصالات الوصفية؛
- معلمات متصلة بمدى التعاون؛
- معلمات متصلة بالمهام؛
- معلمات متصلة بالدخل الصوتي.

وستُناقش هذه الفئات بإيجاز في البنود الواردة أدناه. وستُدرج المعلمات المعنية في كل فئة، إلى جانب التعريف، ومستوى التفاعل الذي تتناوله المعلمة (كلمة أو تفوه أو حوار)، فضلاً عن طريقة القياس (سواء كانت قائمة على استعمال الأجهزة أو على استخلاص الحاشية التفسيرية من جانب الخبراء).

1.7 المعلمات المتصلة بالحوار والاتصال

هي معلمات تشير إلى الحوار ككل وإلى توصيل المعلومات التي تعطي دلالة تقريبية جداً لكيفية حصول التفاعل. ولا تحدد هذه المعلمات بالتفصيل الوظيفة الإبلاغية لكل تفوه منفرد. وترد في الجدول 1 قائمة بالمعلمات المنتمية إلى هذه الفئة، وتتضمن المعلمات ذات الصلة بالمدة (مدة الحوار ككل، المدة التي يستغرقها دورا النظام والمستعمل، مهلة تأخر استجابة النظام والمستعمل)، والمعلمات المتصلة بالكلمات وأدوار الكلام (متوسط عدد أدوار النظام والمستعمل)، ومتوسط عدد كلمات كل دور في النظام والمستعمل، وعدد الأسئلة التي يطرحها كل من النظام والمستعمل).

وتجدر الإشارة إلى معلمتين اقتُرحتا في المرجع [11]، وهما: تعطي معلمة *كثافة الاستجوابات* دلالة على كيفية تمكن مستعمل ما من القيام بفعالية بتقديم معلومات جديدة إلى أحد الأنظمة، وتصف معلمة *كفاءة المفاهيم* كيفية تمكن النظام من استيعاب هذه المعلومات من المستعمل بفعالية. وتشير أيضاً هاتان المعلمتان إلى قدرة النظام على فهم اللغة، ولكنهما أدرجتا في هذا البند لأنهما ناشئتان عن قدرات تفاعل النظام ككل، وليس عن قدرات فهم اللغة فقط.

وتتسم جميع المعلمات الواردة في هذه الفئة بطابع عام وتشير إلى الحوار ككل، رغم أنها تُحسب جزئياً على مستوى التفوه. وتسبب المعلمات العامة مشاكل أحياناً، لأن الاختلافات الفردية في المهارات الإدراكية قد تكون كبيرة مقارنة بالاختلافات الناشئة عن النظام، ولأن المواضيع قد تتعلم استراتيجيات لحل المهام وتؤثر تأثيراً كبيراً على المعلمات العامة.

الجدول 1 – معلمات التفاعل المتصلة بالحوار والاتصال

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
DD	مدة الحوار	مدة الحوار الكلية محسوبة بالوحدة [ms]، انظر مثلاً المراجع [8] [6] [12] [21].	حوار	أجهزة
STD	مدة دور النظام	متوسط مدة دور نظام معين محسوباً بالوحدة [ms]، اعتباراً من لحظة بدء النظام الكلام وحتى توقفه عن ذلك. والدور هو تفوه، أي، امتداد كلام ينطقه أحد الطرفين في الحوار. [8]	تفوه	أجهزة
UTD	مدة دور المستعمل	متوسط مدة دور مستعمل ما محسوباً بالوحدة [ms]، اعتباراً من لحظة بدء المستعمل الكلام وحتى توقفه عن ذلك. [8]	تفوه	أجهزة

الجدول 1 – معلمات التفاعل المتصلة بالحوار والاتصال

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
SRD	مهلة تأخر استجابة النظام	متوسط مهلة تأخر استجابة النظام محسوباً بالوحدة [ms]، اعتباراً من لحظة توقف المستعمل عن الكلام لحين بدء النظام الكلام. [22]	تفوه	أجهزة
URD	مهلة تأخر استجابة المستعمل	متوسط مهلة تأخر استجابة المستعمل محسوباً بالوحدة [ms]، اعتباراً من لحظة توقف النظام عن الكلام لحين بدء المستعمل الكلام. [22]	تفوه	أجهزة
# turns	عدد الأدوار	مجموع عدد أدوار الكلام في حوار ما. [30]	حوار	أجهزة/خبراء
# system turns	عدد أدوار النظام	مجموع عدد أدوار كلام النظام في حوار ما. [30]	حوار	أجهزة/خبراء
# user turns	عدد أدوار المستعمل	مجموع عدد أدوار الكلام التي يأخذها النظام في حوار ما. [30]	حوار	أجهزة/خبراء
WPST	الكلمات بحسب دور النظام	متوسط عدد كلمات بحسب دور النظام في حوار ما. [6]	تفوه	أجهزة/خبراء
WPUT	عدد كلمات بحسب دور مستعمل	متوسط عدد كلمات بحسب دور مستعمل في حوار ما. [6]	تفوه	أجهزة/خبراء
# system questions	عدد الأسئلة التي يطرحها النظام	مجموع عدد الأسئلة التي يطرحها النظام في كل حوار.	حوار	خبراء
# user questions	عدد الأسئلة التي يطرحها المستعمل	مجموع عدد الأسئلة التي يطرحها المستعمل في كل حوار. [12] [21]	حوار	خبراء
QD	كثافة الاستجوابات	متوسط عدد المفاهيم الجديدة (الفجوات، انظر البند 4.7) الواردة في كل استجواب يطرحه المستعمل. وعلى فرض أن n_d عدد الحوارات، و$n_q(i)$ مجموع عدد استجوابات المستعمل في الحوار i^{th}، و$n_u(i)$ عدد المفاهيم الوحيدة التي "يفهمها" النظام فهماً صحيحاً في الحوار i^{th}، نحصل حينئذ على المعادلة التالية: $$QD = \frac{1}{n_d}\sum_{i=1}^{n_d}\frac{n_u(i)}{n_q(i)}$$ لا يُحسب المفهوم في العدد $n_u(i)$ إذا فهمه النظام بالفعل في واحد من التفوهات السابقة. [11]	مجموعة حوارات	خبراء
CE	كفاءة المفاهيم	متوسط عدد الأدوار اللازمة لكل مفهوم يتعين أن "يفهمه" النظام. وعلى فرض أن n_d عدد الحوارات، و$n_u(i)$ عدد المفاهيم الوحيدة التي "يفهمها" النظام فهماً صحيحاً في الحوار i^{th}، و$n_c(i)$ مجموع عدد المفاهيم في الحوار i^{th}، نحصل حينئذ على المعادلة التالية: $$CE = \frac{1}{n_d}\sum_{i=1}^{N_d}\frac{n_u(i)}{n_c(i)}$$ يُحسب المفهوم كلما تفوهه المستعمل ولم يفهمه النظام بالفعل. [11]	مجموعة حوارات	خبراء

2.7 المعلمات المتصلة بالاتصالات الوصفية

الاتصالات الوصفية، أي الاتصال من أجل توصيل رسالة، هو جانب مهم بشكل خاص في التفاعل بالكلام مع الأنظمة التي لديها قدرات محدودة للتعرف والفهم والتعليل. ومن الضروري في هذه الحالة استرداد تفوهات التصحيح والتوضيح أو حتى الحوارات الفرعية، من حالات سوء الفهم.

وتقدر المعلمات المنتمية إلى هذه المجموعة عدد تفوهات النظام والمستعمل التي تشكل جزءاً من جانب الاتصالات الوصفية. وتُحسب معظم المعلمات بوصفها العدد المطلق للتفوهات في حوار يتعلق بمشكلة تفاعل معينة، ومن ثم يُقدر متوسطها على أساس مجموعة حوارات. وتضم المعلمات عدد طلبات المساعدة المقدمة من المستعمل، وعدد الإعلانات الصادرة من النظام والمتعلقة بانتهاء الفترة الزمنية، وعدد تفوهات المستعمل التي يرفضها النظام في الحالات التي يتعذر فيها استخلاص محتوى ذو معنى منها (حالات الرفض في إطار التعرف على الكلام أوتوماتياً (ASR))، وعدد رسائل أخطاء التشخيص الواردة من النظام، وعدد محاولات تدخل المستعمل، وعدد محاولات المستعمل الرامية إلى إلغاء إجراء سابق.

ويمكن وصف قدرة النظام (والمستعمل) على الاسترداد من مشاكل التفاعل، بطريقتين: إما طريقة واضحة بواسطة معدلات التصحيح، أي، النسبة المئوية لجميع أدوار (النظام أو المستعمل) المعنية أساساً بتقويم إحدى مشاكل التفاعل، أو بطريقة ضمنية بواسطة معلمة *الاسترداد الضمني* التي تقيم قدرة النظام على استعادة التفوهات التي يُعجز جزئياً عن التعرف عليها أو فهمها.

وعلى النقيض من القياسات العامة، تصف معظم المعلمات المتصلة بالاتصالات الوصفية وظيفة تفوهات النظام والمستعمل في عملية الاتصال. وعليه، يتعين تحديد معظم المعلمات بمساعدة أحد الخبراء المعنيين باستخلاص الحواشي التفسيرية. وترد في الجدول 2 قائمة بالمعلمات.

الجدول 2 – معلمات التفاعل المتصلة بالاتصالات الوصفية

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
# help request	عدد طلبات المساعدة المقدمة من المستعمل	مجموع عدد طلبات المساعدة المقدمة من المستعمل في حوار ما. ويُوسم طلب المساعدة المقدم من المستعمل بالخبير المعني باستخلاص بالحواشي التفسيرية إذا طلب المستعمل المساعدة صراحة. ويوضع هذا الطلب في صيغة سؤال (من قبيل، "ما هي الخيارات المتيسرة؟") أو في صيغة عبارة ("أعطني الخيارات المتيسرة!"). [30]	تفوه	خبراء
# system help	عدد رسائل المساعدة المتعلقة بالتشخيص الواردة من النظام	مجموع عدد رسائل المساعدة الصادرة عن النظام في حوار ما. ورسالة المساعدة عبارة عن تفوه نظام يبلغ المستعمل بالخيارات المتيسرة في مرحلة معينة من الحوار.	تفوه	أجهزة/خبراء
# time-out	عدد الإعلانات عن انتهاء الفترة الزمنية	مجموع عدد الإعلانات عن انتهاء الفترة الزمنية في حوار معين، بسبب عدم ورود استجابة من المستعمل. [30]	تفوه	أجهزة
# ASR rejection	عدد حالات الرفض ASR	مجموع عدد حالات الرفض ASR في حوار ما. ويُعرف الرفض ASR على أنه أحد إعلانات النظام التي تبين عجز النظام عن "سماع" المستعمل أو "فهمه"، أي، عجز النظام عن استخلاص أي معنى من تفوه المستعمل. [30]	تفوه	أجهزة

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
# system error	عدد رسائل خطأ التشخيص الواردة من النظام	مجموع عدد رسائل خطأ التشخيص يبثها النظام في حوار ما. وتُعرف رسالة خطأ التشخيص على أنها تفوه نظام يبين فيه النظام أنه عاجز عن أداء مهمة محددة أو تقديم معلومات معينة. [22]	تفوه	أجهزة/خبراء
# barge-in	عدد محاولات تدخل المستعمل	مجموع عدد محاولات تدخل المستعمل في حوار ما. وتُحسب محاولة تدخل المستعمل عندما يتعمد مخاطبة النظام بينما لا يزال النظام يتحدث. وفي هذا التعريف، لا يُنظر إلى تفوهات المستعمل التي لا يُقصد بها أن تؤثر على سير الحوار (الضحك، تعابير الغضب أو الكياسة)، على أنها تفوهات تدخل. [30]	تفوه	خبراء
# cancel	عدد محاولات المستعمل الرامية إلى الإلغاء	مجموع عدد محاولات الإلغاء التي يقوم بها المستعمل في حوار ما. ويُصنف دور المستعمل على أنه محاولة إلغاء إذا حاول المستعمل بدء الحوار مجددا من البداية، أو إذا أراد/أرادت صراحة أن تتجاوز مستوى واحد أو عدة مستويات في تراتب الحوار باتجاه الخلف. [23] [16]	تفوه	خبراء
SCT, SCR	عدد أدوار كلام نظام المعنية بالتصحيح، معدل تصحيح النظام	العدد الكلي (SCT) أو النسبة المئوية (SCR) لجميع أدوار النظام في حوار ما والمعنية أساساً بتقويم "مشكلة" معينة، والتي تقطع بالتالي تدفق الحوار دون أن تضيف إليه محتويات جديدة تنطوي على اقتراحات. وقد تكون "المشكلة" ناجمة عن أخطاء في التعرف على الصوت أو فهمه، أو عن تفوهات المستعمل غير المنطقية، أو المتناقضة، أو غير المحددة. وفي حال عدم إجابة المستعمل على سؤال النظام، يُوسم جواب النظام المقابل بوصفه دور تصحيح النظام، فيما عدا الحالات التي يطلب فيها المستعمل معلومات أو إجراءات لا تدعمها وظيفة النظام الحالية. [8] [24] [9] [7]	تفوه	خبراء
UCT, UCR	عدد أدوار كلام المستعمل المعنية بالتصحيح، معدل تصحيح المستعمل	العدد الكلي (UCT) أو النسبة المئوية (UCR) لجميع أدوار المستعمل في حوار ما والمعنية أساسا بتقويم "مشكلة" معينة، والتي تقطع بالتالي تدفق الحوار دون أن تضيف إليه محتويات جديدة تنطوي على اقتراحات. (انظر SCT، SCR). [8] [24] [9] [7]	تفوه	خبراء
IR	الاسترداد الضمني	قدرة النظام على الاسترداد من تفوهات المستعمل التي تفشل بشأنها عملية التعرف إلى الكلام أو فهمه فشلاً جزئياً. وتُحدد بتوسيم التفوهات المحللة جزئياً (انظر تعريف PA:PA الوارد في البند 5.7) فيما يتعلق بما إذا كانت إجابة النظام "مناسبة" أم لا، كما يلي: $IR = \frac{\#\, utterances\ with\ appropriate\ system\ answer}{PA:PA}$ انظر البند 3.7 للاطلاع على تعريف "مدى الملاءمة". [7]	تفوه	خبراء

3.7 المعلمات المتصلة بمدى التعاون

يُعرف مدى التعاون على أنه جانب أساسي لإنجاح التفاعل مع أحد أنظمة الحوار الأوتوماتية [1]. ومما يؤسف له أن من الصعب إجراء تقدير لمسألة ما إذا كان النظام يسلك سلوكاً تعاونياً أم لا. ويرتبط العديد من المعلمات المتصلة بالحوار والمعلمات المتصلة بالاتصالات الوصفية بطريقة أو أخرى، بمدى تعاون النظام، ولكنها لا تسعى إلى تقدير هذا الجانب.

ويُحصل على القياسات المباشرة لمدى التعاون بواسطة معلمات مدى الملاءمة في السياقية التي استحدثها كل من Simpson وFraser [24]. ويتعين تقدير كل تفوه من تفوهات النظام من جانب عدد معين من الخبراء من حيث ما إذا كان التفوه ينتهك قاعدة واحدة أو أكثر من قواعد Grice المتعلقة بمدى التعاون، انظر المرجع [13]:

- *كمية* المعلومات: اجعل مساهمتك إبلاغية حسب المطلوب (تحقيقاً للغرض الحالي من التبادل)؛ ولا تجعلها إبلاغية أكثر من اللازم.
- *النوعية*: حاول أن تجعل مساهمتك صحيحة؛ ولا تقل ما تعتقد أنه خطأ؛ ولا تقل شيئاً تعجز عن إثباته بأدلة وافية.
- *العلاقة*: كن على صلة بالموضوع.
- *الأسلوب*: كن واضحاً؛ وتجنب الغموض في التعبير؛ وتفادى الالتباس؛ وتوخى الإيجاز (تجنب الإطناب بدون داعي)؛ وكن مرتباً.

وذُكرت هذه المبادئ على نحو أدق من جانب Bernsen وDybkjær [1] في حالة أنظمة الحوار الأوتوماتية.

وتُصنف التفوهات إلى فئات تفوه ملائم (لا ينتهك قواعد Grice)، أو تفوه غير ملائم (ينتهك قاعدة واحدة أو أكثر منها)، أو تفوه ملائم/غير ملائم (يتعذر على الخبراء التوصل إلى اتفاق في تصنيف هذه التفوهات)، أو تفوه غير مفهوم (محتوى التفوه لا يمكن تمييزه في سياق الحوار)، أو تفوه فاشل تماماً (لا توجد استجابة لغوية من النظام). وتجدر الإشارة إلى أن التصنيف ليس سهلاً دوماً، وأن مبادئ التفسير قد تكون ضرورية.

الجدول 3 – معلمات التفاعل المتصلة بمدى التعاون

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
CA:AP, *CA:IA, CA:TF, CA:IC, %CA:AP, %CA:IA, %CA:TF, %CA:IC*	مدى الملاءمة في السياق	العدد الكلي أو النسبة المئوية لتفوهات النظام التي تُعتبر ملائمة في السياق المباشر للحوار. وتُحدد هذه المعلمات بتوسيم التفوهات طبقاً لما إذا كانت تنتهك قاعدة واحدة أو أكثر من قواعد Grice المتعلقة بمدى التعاون: • *CA:AP*: تفوه ملائم لا ينتهك القواعد المذكورة، وليس مبهماً أو معلماً بطريقة معينة على نحو غير متوقع. • *CA:IA*: تفوه غير ملائم ينتهك قاعدة واحدة أو أكثر من قواعد Grice. • *CA:TF*: تفوه فاشل تماماً، لا يوجد استجابة لغوية. • *CA:IC*: تفوه غير مفهوم، يتعذر تمييز محتواه على الخبير المعني باستخلاص الحواشي التفسيرية. انظر المراجع [24] [8] [9] للاطلاع على المزيد من التفاصيل؛ التصنيف مشابه لذلك المعتمد في المرجع [14].	تفوه	خبراء

4.7 المعلمات المتصلة بالمهام

تخوّل الخدمات الحالية الحديثة للغاية التفاعلات الموجهة إلى المهام بين النظام والمستعمل، ويعد نجاح المهمة عاملاً أساسياً للاستفادة من خدمة معينة. وأفضل طريقة لتقييم هذا النجاح على أكمل وجه في إحدى البيئات المختبرية حيث تُخصص مهام واضحة لمواضيع الاختبار، انظر التوصية ITU-T P.851. ومع ذلك، ينبغي أن تراعي القياسات الحقيقية لنجاح المهمة احتمالات انحراف المستعمل عن السيناريو، إما بسبب عدم مراعاته/مراعاتها للتعليمات المبينة في السيناريو، أو نتيجة عدم انتباهه/انتباهها إلى تفوهات النظام، أو لأن المهمة يتعذر حلها وينبغي تعديلها في مجرى الحوار.

ويراعي تعديل المهمة التجريبية في معظم تعاريف نجاح المهمة المبينة في الأعمال المنشورة. ويمكن بلوغ النجاح من خلال القيام ببساطة بتقديم الإجابة الصحيحة على القيود المفروضة في التعليمات، وذلك بتخفيف القيود على النظام أو المستعمل (أو كليهما)، أو عن طريق ملاحظة عدم وجود حل للمهمة المحددة. وقد يُعزى فشل المهمة على سبيل التجربة إلى سلوك النظام أو سلوك المستعمل، بيد أن المستعمل يتأثر بسلوك النظام.

وثمة نهج آخر لتحديد نجاح المهمة هو معامل κ. وتفترض هذه الطريقة نهجاً لفهم الكلام يرتكز إلى نعوت (مفاهيم، فجوات) يتعين أن تُخصص لها قيم مسموح بها خلال سير الحوار بين النظام والمستعمل. وتُسمى أزواج النعوت والقيم المخصصة لها، بأزواج النعوت – القيم (AVP). ومجموعة النعوت المتيسرة ككل، إلى جانب القيم التي تخصصها لها المهمة (المعروفة باسم مصفوفة النعوت – القيم (AVM))، هي مجموعة تصف وصفاً كاملاً مهمة يمكن أداؤها بمساعدة النظام. ومن أجل تحديد معامل κ، تُحدد مصفوفة التباس *M(i,j)* للنعوت الموجودة في المفتاح (تعريف السيناريو) وفي الحل المبين (ملف تسجيل الحوار). ومن ثم يمكن القيام بالاستناد إلى هذه المصفوفة، بحساب الاتفاق بين المفتاح والحل *P(A)* واتفاق الصدفة *P(E)*، انظر الجدول 4. ويمكن حساب المصفوفة *M(i,j)* لكل حوار منفرد، أو لمجموعة حوارات تنتمي إلى نظام معين أو تشكيلة نظام بعينها.

ويعتمد معامل κ على مدى تيسر نظام بسيط لتشفير المهمة، أي مصفوفة AVM. ومع ذلك، لا يمكن تحديد خصائص بعض المهام بسهولة. وثمة حاجة في هذه الحالة إلى وضع نهج أكثر تفصيلاً لتقييم نجاح المهمة، وعادة ما تعتمد هذه النهج على نوع المهمة قيد النظر.

الجدول 4 – معلمات التفاعل المتصلة بالمهام

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
TS	نجاح المهمة	وسم نجاح المهمة طبقاً لما إذا كان المستعمل قد بلغ/بلغت الهدف بنهاية الحوار، شريطة أن يتسنى بلوغه بمساعدة النظام. وتبين الوسوم ما إذا كان الهدف قد بُلغ أم لا، وتحدد مصدر المشاكل المحتمل: • *TS:S*: مهمة ناجحة (يوجد لها حلول) • *TS:SCs*: مهمة ناجحة مع تخفيف القيود من جانب النظام • *TS:SCu*: مهمة ناجحة مع تخفيف القيود من جانب المستعمل • *TS:SCsCu*: مهمة ناجحة مع تخفيف القيود من جانب النظام والمستعمل على حد سواء • *TS:SN*: مهمة ناجحة بعد ملاحظة عدم وجود حل لها • *TS:Fs*: مهمة فاشلة سببها سلوك النظام، الناجم عن أوجه قصور تشوب النظام • *TS:Fu*: مهمة فاشلة سببها السلوك غير التعاوني للمستعمل انظر أيضاً المراجع [8] [7] [24].	حوار	خبراء

κ	معامل kappa	النسبة المئوية للمهام المكتملة وفقا لإحصاءات kappa. وتُحدد هذه النسبة على أساس صحة نتيجة المصفوفة AVM التي تُحقق في نهاية الحوار فيما يخص المصفوفة AVM التابعة للسيناريو (المفتاح). وتُحدد مصفوفة التباس *M(i,j)* للنعوت الموجودة في النتيجة وفي المفتاح، حيث T عدد حالات العد في M، وt_i مجموع حالات العد في العمود i من M. ومن ثم نحصل على المعادلة الآتية: $$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$ حيث *P(A)* عدد المرات التي تتفق فيها المصفوفة AVM التابعة للحوار الفعلي والمفتاح، $P(A) = \sum_{i=1}^{n} \frac{M(i,i)}{T}$. يمكن تقدير *P(E)* من عدد المرات التي يُتوقع فيها أن يتم الاتفاق بالصدفة، $P(E) = \sum_{i=1}^{n} (\frac{t_i}{T})^2$. [4] [31]	حوار أو مجموعة حوارات	خبراء

5.7 المعلمات المتصلة بالدخل الصوتي

تحدد قدرة الدخل الصوتي لنظام حوار أوتوماتي بموجب قدرته على التعرف على الكلمات والتفوهات، وعلى استخلاص المعنى من السلسلة المُتعرف عليها (المُسماة "فهم الكلام"). وينبغي التمييز بين نهجين بشأن التعرف على الصوت أوتوماتياً، وهما: أجهزة التعرف على الكلمات القادرة على استخلاص كلمات منفصلة من كلام المستعمل عند نطقها بصورة منفصلة (تعرف على كلمات منفصلة) أو باستمرار (كشف عن كلمات مفتاحية). ومن جهة أخرى، فإن أجهزة التعرف على الكلام المستمر قادرة على التعرف على جمل أو تفوهات بأكملها. وغالباً ما يُفهم الكلام على أساس أزواج النعوت – القيم، انظر البند 4.7. وتتطرق المعلمات الموصوفة في الفقرة الواردة أدناه إلى التعرف على الكلام وفهمه على حد سواء.

ويرد عموماً في أجهزة التعرف على الكلام المستمر سلسلة كلمات افتراضية بوصفها خرجاً معيناً. ومن أجل الحكم على ما إذا كانت السلسلة تمثل ما يُقال تمثيلاً صحيحاً، ينبغي أن يوفر الخبير المعني بالنسخ، نسخاً مرجعياً. وتُراصف أولاً السلسلة الافتراضية والسلسلة المرجعية في كل تفوه على مستوى الكلمات بواسطة برمجة دينامية (DP) قائمة على خوارزمية مواءمة [19] [20]. ويُحسب على أساس هذا التراصف عدد الكلمات c_w المحددة بشكل صحيح، وعدد حالات الاستبدال s_w، وحالات الإدراج i_w، وحالات الحذف d_w . ويمكن أن ترتبط حالات العد هذه بمجموع عدد الكلمات في المرجع n_w، مما يؤدي إلى الحصول على قياسين بديلين لنوعية أداء التعرف، وهما معدل الأخطاء في الكلمات *(WER)* ودقة الكلمات *(WA)*، انظر الجدول 5.

وبالإمكان تحديد القياسات التكميلية لنوعية الأداء على مستوى الجملة، وذلك على أساس دقة الجملة، *SA*، أو معدل الأخطاء فيها، *SER*، انظر الجدول 5. وعموماً، فإن الدقة *SA* أدنى من الدقة *WA*، لأن أي كلمة مفردة في الجملة يُخطأ في التعرف عليها تؤثر على معلمة *SA*. ومع ذلك، يمكن أن تكون أعلى من دقة الكلمة، وخاصة عندما يتم التعرف بصورة صحيحة على الكثير من الجمل المكونة من كلمة واحدة. وأشار Strik وآخرون إلى الحقيقة القائلة إن المعدل *SER* والدقة *SA* يؤثران على التفوه برمته عند حصول خطأ في التعرف على كلمة واحدة [26] [27]؛ ويمكن تلافي المشكلة بواسطة المعلمتين *NES* و*WES*، انظر الجدول 5. وفي حال عدم تجزئة التفوهات إلى جمل، يمكن أيضاً حساب جميع المعلمات المتصلة بالجملة على مستوى التفوه عوضاً عن حسابها على مستوى الجملة.

وتنتج أجهزة التعرف على الكلمات المنفصلة كلمة أو تفوه افتراضي لكل كلمة أو تفوه دخل. وبالإمكان مقارنة كلمات الدخل والخرج مباشرة، ويمكن تحديد قياسات لنوعية الأداء مماثلة لتلك المحددة في حالة التعرف المستمر، وذلك بحذف حالات الإدراج. وبالإمكان حساب عدد الإنذارات الخاطئة في أثناء فترة زمنية معينة، بدلاً من حالات الإدراج، انظر van Leeuwen وSteeneken [28]. ويمكن أيضاً تحديد المعلمتين *WA* و*WER* للكلمات المفتاحية حصراً عندما يعمل جهاز التعرف بأسلوب الكشف عن هذه الكلمات.

وينبغي تمييز نهجين عاديين فيما يتعلق بتقييم فهم الكلام. ويستند الأول إلى تصنيف إجابات النظام على أسئلة المستعمل في فئات كل من الإجابات الصحيحة، أو الإجابات الصحيحة جزئياً، أو الإجابات غير الصحيحة، أو الإجابات الفاشلة. ويمكن دمج الفئات المختلفة للإجابات في قياسات يستعملها برنامج وكالة تطوير مشاريع البحوث المتقدمة التابعة لوزارة الدفاع الأمريكية (US DARPA)، انظر الجدول 5. أما النهج الثاني فيتمثل في تصنيف قدرات النظام على التحليل، إما على أساس التفوهات المحللة بشكل صحيح، أو أزواج AVP المحددة تحديداً صحيحاً. وبناء على تحديد أزواج AVP تحديداً صحيحاً، يمكن حساب قياسات عامة من قبيل دقة المفهوم، *CA*، أو معدل الخطأ في المفهوم، *CER*، أو دقة الفهم، *UA*. وترد جميع هذه المعلمات في الجدول 5.

الجدول 5 – معلمات التفاعل المتصلة بالدخل الصوتي

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
WER, WA	معدل الأخطاء في الكلمات، دقة الكلمات	النسبة المئوية للكلمات التي يتم التعرف عليها بصورة صحيحة على أساس الشكل الإملائي للتفوه الافتراضي والتفوه المرجعي (المنسوخ)، ويُنفذ تراصف بمساعدة خوارزمية "sclite"، انظر المرجع [18]. وبتعيين n_w على أنه مجموع عدد الكلمات الواردة في جميع تفوهات مستعمل حوار معين، وs_w وd_w وi_w عدد الكلمات المستبدلة والمحذوفة والمدرجة على التوالي، يمكن عندئذ تحديد معدل الأخطاء في الكلمات ودقة الكلمات كما يلي: $$WER = \frac{s_w + i_w + d_w}{n_w}$$ $$WA = 1 - \frac{s_w + i_w + d_w}{n_w} = 1 - WER$$ انظر المرجع [24]؛ يرد في المرجع [28] تفاصيل عن كيفية حساب هذه المعلمات في حالة التعرف على الكلمات المنفصلة.	كلمة	أجهزة/خبراء

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
SER, SA	معدل الأخطاء في الجملة، دقة الجملة	النسبة المئوية لمجموع الجمل التي يتم التعرف عليها بصورة صحيحة. وبتعيين n_s على أنه مجموع عدد الجمل، وs_s وi_s وd_s عدد الجمل المستبدلة والمدرجة والمحذوفة على التوالي، يمكن حينئذ تحديد معدل الأخطاء في الجملة ودقة الجملة كما يلي: $$SER = \frac{s_s + i_s + d_s}{n_s}$$ $$SA = 1 - \frac{s_s + i_s + d_s}{n_s} = 1 - SER$$ [24]	تفوه	أجهزة/خبراء
NES	عدد الأخطاء في الجملة	متوسط عدد الأخطاء التي يتم التعرف عليها في الجملة. وعلى فرض أن $s_w(k)$ و$i_w(k)$ و$d_w(k)$ عدد الكلمات المستبدلة والمدرجة والمحذوفة على التوالي في الجملة k، عندئذ تكون لدينا المعادلة التالية: $$NES(k) = s_w(k) + i_w(k) + d_w(k)$$ ويمكن حساب المتوسط *NES* من المعادلة التالية: $$NES = \frac{\sum_{k=1}^{\#\ user\ turns} NES(k)}{\#\ user\ turns} = \frac{WER \cdot \#\ user\ words}{\#\ user\ turns}$$ [26]	تفوه	أجهزة/خبراء
WES	الخطأ في كلمات كل جملة	يرتبط هذا المتوسط بالمتوسط *NES*، ولكنه مُقيس وفقاً لعدد الكلمات في الجملة k، $w(k)$: $$WES(k) = \frac{NES(k)}{w(k)}$$ ويمكن حساب المتوسط *WES* من المعادلة التالية: $$WES = \frac{\sum_{k=1}^{\#\ user\ turns} WES(k)}{\#\ user\ turns}$$ [26]	كلمة	أجهزة/خبراء
AN:CO, *AN:IN*, *AN:PA*, *AN:FA*, *%AN:CO*, *%AN:IN*, *%AN:PA*, *%AN:FA*	عدد إجابات النظام الصحيحة/غير الصحيحة/ الصحيحة جزئياً/الفاشلة أو النسبة المئوية لهذه الإجابات	العدد الكلي أو إجمالي النسبة المئوية للأسئلة المطروحة من جانب المستعمل التي يجيبها النظام في الحوار، وهي كما يلي: • صحيحة (*AN:CO*)؛ • غير صحيحة (*AN:IC*)؛ • صحيحة جزئياً (*AN:PA*)؛ • فاشلة (*AN:FA*). انظر المراجع [21] [12] [14].	تفوه	أجهزة/خبراء

المختصر	الاسم	التعريف	مستوى التفاعل	طريقة القياس
$DARPA_s$, $DARPA_{me}$	علامة الوكالة DARPA، الخطأ المعدل لوكالة DARPA	قياسات مأخوذة وفقاً لمبادرة فهم الكلام لوكالة تطوير مشاريع البحوث المتقدمة التابعة لوزارة الدفاع الأمريكية (DARPA)، والتي عدلها Skowronek [25] [17] لحساب الإجابات الصحيحة جزئياً كما يلي: $$DARPA_s = \frac{AN:CO - AN:IC}{\#\ user\ questions}$$ $$DARPA_{me} = \frac{AN:FA + 2\cdot(AN:IC + AN:PA)}{\#\ user\ questions}$$ [21] [12] [25]	تفوه	خبراء
PA:CO, *PA:PA*, *PA:IC*, *%PA:CO*, *%PA:PA*, *%PA:IC*	عدد تفوهات المستعمل المحللة بصورة صحيحة/ صحيحة جزئياً/غير صحيحة	تقييم عدد المفاهيم (أزواج النعوت– القيم، AVPs) في تفوه ما، والتي يستنبطها النظام كما يلي: • *PA:CO*: يفهم النظام جميع مفاهيم تفوه المستعمل فهماً صحيحاً. • *PA:PA*: لا يفهم النظام جميع مفاهيم تفوه المستعمل فهماً صحيحاً، ولكنه يفهم واحداً منها على الأقل. • *PA:IC*: لا يفهم النظام أي مفهوم من مفاهيم تفوه المستعمل فهماً صحيحاً. ويُعبر عن هذا التقييم بالعدد الكلي أو النسبة المئوية الإجمالية لتفوهات المستعمل التي تُحلل تحليلاً صحيحاً/صحيح جزئياً/غير صحيح في حوار ما. [7]	تفوه	خبراء
CA, CER	دقة المفهوم، معدل الخطأ في المفهوم	النسبة المئوية لوحدات المعاني المفهومة بشكل صحيح في الحوار. وتُعرف المفاهيم على أنها أزواج نعوت – قيم (AVPs)، حيث إن n_{AVP} مجموع عدد أزواج AVP، وs_{AVP} وi_{AV} وd_{AVP} عدد أزواج AVP المستبدلة والمدرجة والمحذوفة. ومن ثم يمكن تحديد دقة المفهوم ومعدل الخطأ في المفهوم كما يلي: $$CA = 1 - \frac{s_{AVP} + i_{AVP} + d_{AVP}}{n_{AVP}}$$ $$CER = \frac{s_{AVP} + i_{AVP} + d_{AVP}}{n_{AVP}}$$ [2][3][24][9]	تفوه	خبراء
UA	دقة الفهم	النسبة المئوية لتفوهات المستعمل التي تُستنبط فيها جميع وحدات المعاني (AVPs) استنباطاً صحيحاً كالتالي: $$UA = \frac{PA:CO}{\#\ user\ turns}$$ [32]	تفوه	خبراء

6.7 معلمات أخرى

تصف معظم معلمات التفاعل الواردة في الجداول سلوك النظام، وهو سلوك واضح بسبب أهمية نوعية النظام والخدمة. ويمكن تحديد معلمات ذات صلة بالمستعمل بالإضافة إلى المعلمات المذكورة. وتخص هذه المعلمات تحديداً مجموعة مستعملين يشاركون في اختبار معين، ولكنها مع ذلك، قد ترتبط ارتباطاً وثيقاً بخصائص النوعية من منظور المستعمل.

وعند فصل الجوانب المختلفة لنوعية خدمة معينة قائمة على أنظمة الحوار الأوتوماتية (SDS) على غرار ما يبينه البند 3.5/التوصية P.851، يمكن ملاحظة أن معلمات التفاعل لا تتطرق إلى العديد من هذه الجوانب. ولا توجد معلمات ترتبط مباشرة بمدى صلاحية الاستعمال، أو ارتياح المستعمل، أو مدى المقبولية، أو نوعية خرج الكلام. ولا يوجد حتى الآن سوى نهج قليلة للغاية تتناول نوعية خرج الكلام بطريقة معلمية (سواء كان متسلسلاً أم تركيبياً). والقياسات المأخوذة بواسطة الأجهزة والمتصلة بمدى فهم الكلام هي قياسات محددة مثلاً في المعيار [15] 16-60268 الصادر عن اللجنة الكهرتقنية الدولية (IEC)، ولكنها لا تنطبق على البيئة الهاتفية. واقتُرحت قياسات تقدير تكاليف التسلسل التي يمكن حسابها من نص الدخل وقاعدة البيانات الصوتية لنظام تركيبـــي متسلسل [5]. وعلى الرغم من أن هذه القياسات تظهر أحياناً علاقات ترابط رصينة مع متوسطات علامات الرأي (MOS) التي يُحصل عليها في الاختبارات السمعية، إلا أنها تخص تحديداً مركبات الكلام ومجموعات وثائق التسلسل التابعة لها.

8 تفسير قيم معلمات التفاعل

على الرغم من أهمية معلمات التفاعل المحددة في هذه الإضافة بالنسبة لتصميم النظام وتحقيق الحد الأمثل من أدائه وصيانته، فإنها لا ترتبط مباشرة بالنوعية من منظور المستعمل. وعليه، ينبغي إكمال مجموعة معلمات التفاعل بمجموعة أحكام المستعمل بشأن مختلف جوانب النوعية، بحسب الوصف الوارد في التوصية ITU-T P.851. وبهذه الطريقة فقط يمكن الحصول على معلومات صحيحة عن نوعية الخدمات القائمة على أنظمة الحوار الأوتوماتية.

ويمكن أن يستند تفسير قيم معلمات التفاعل إلى نتائج الاختبارات، ولكن هذه النتائج غالباً ما تكون خاصة بالنظام أو الخدمة قيد النظر. فمثلاً، يمكن أن تبين زيادة عدد تفوهات انتهاء الفترة الزمنية، أن المستعمل لا يعرف ما يقوله في مراحل معينة من الحوار، أو أنه/أنها مرتبكة بشأن فهم بعض إجراءات النظام [29]. وقد لا تعكس زيادة محاولات التدخل سوى أن المستعمل على بينة من إمكانية انقطاع النظام. وعلى نقيض ذلك، فإن تقليل عدد هذه المحاولات يمكن أن يبين أيضاً أن المستعمل لا يدري ما يقوله للنظام. وقد تنشأ تفوهات المستعمل الطويلة عن عدد التدخلات الكبير من جانب المستعمل. وتخفيض قيم المعلمات المتصلة بالاتصالات الوصفية (وخاصة المتعلقة منها بالاتصالات الوصفية التي يستهلها المستعمل) هو تخفيض يمكن أن يُتوقع منه زيادة متانة النظام، وسلاسة الحوار، وكفاءة الاتصال [1].

بيبليوغرافيا

[1] BERNSEN, N.O., DYBKJÆR, H., DYBKJÆR, L.: *Designing interactive speech systems: From first ideas to user testing,* Springer, DE-Berlin, 1998.

[2] BILLI, R., CASTAGNERI, G., DANIELI, M.: Field trial evaluations of two different information inquiry systems, *Proc. 3rd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA'96)*, US-Basking Ridge NJ, pp. 129-134, 1996.

[3] BOROS, M., ECKERT, W., GALLWITZ, F., GORZ, G., HANRIEDER, G., NIEMANN, H.: Towards understanding spontaneous speech: Word accuracy vs. concept accuracy, *Proc. 4th Int. Conf. on Spoken Language Processing (ICSLP'96)*, IEEE, US-Piscataway NJ, 2, pp. 1009-1012, 1996.

[4] CARLETTA, J.: Assessing agreement of classification tasks: The kappa statistics, *Computational Linguistics*, Vol. 22(2), pp. 249-254, 1996.

[5] CHU, M., PENG, H.: An objective measure for estimating MOS of synthesized speech, *Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia*), DK-Aalborg, 3, pp. 2087-2090, 2001.

[6] COOKSON, S.: Final evaluation of VODIS – Voice operated data inquiry system, *Proc. of Speech'88, 7th FASE Symposium*, UK-Edinburgh, 4, pp. 1311-1320, 1988.

[7] DANIELI, M., GERBINO, E.: Metrics for evaluating dialogue strategies in a spoken language system, *Empirical Methods in Discourse Interpretation and Generation. Papers from the 1995 AAAI Symposium*, US-Stanford CA, AAAI Press, US-Menlo Park CA, pp. 34-39, 1995.

[8] FRASER, N.: Assessment of interactive systems, *Handbook on Standards and Resources for Spoken Language Systems* (D. Gibbon, R. Moore and R. Winski, eds.), Mouton de Gruyter, DE-Berlin, pp. 564-615, 1997.

[9] GERBINO, E., BAGGIA, P., CIARAMELLA, A., RULLENT, C.: Test and evaluation of a spoken dialogue system, *Proc. Int. Conf. Acoustics Speech and Signal Processing (ICASSP'93)*, IEEE, US-Piscataway NJ, 2, pp. 135-138, 1993.

[10] GIBBON, D., MOORE, R., WINSKI, R., Eds.: *Handbook on Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, DE-Berlin, 2000.

[11] GLASS, J., POLIFRONI, J., SENEFF, S., ZUE, V.: Data collection and performance evaluation of spoken dialogue systems: The MIT experience, *Proc. 6th Int. Conf. on Spoken Language Processing (ICSLP 2000)*, CN-Beijing, 4, pp. 1-4., 2000.

[12] GOODINE, D., HIRSCHMAN, L., POLIFRONI, J., SENEFF, S., ZUE, V.: Evaluating interactive spoken language systems, *Proc. 2nd Int. Conf. on Spoken Language Processing (ICSLP'92)*, CA-Banff, 1, pp. 201-204, 1992.

[13] GRICE, H.P.: Logic and conversation, *Syntax and Semantics, Vol. 3: Speech Acts* (P. Cole and J.L. Morgan, eds.), Academic Press, US-New York NY, pp. 41-58, 1975.

[14] HIRSCHMAN, L., PAO, C.: The cost of errors in a spoken language system, *Proc. 3rd Europ. Conf. on Speech Communication and Technology (Eurospeech'93)*, DE-Berlin, 2, pp. 1419-1422, 1993.

[15] IEC 60268-16 (2003), *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*. International Electrotechnical Commission, CH-Geneva.

[16] KAMM, C.A., LITMAN, D.J., WALKER, M.A.: From novice to expert: The effect of tutorials on user expertise with spoken dialogue systems, *Proc. 5th Int. Conf. on Spoken Language Processing (ICSLP'98)*, AU-Sydney, 4, pp. 1211-1214, 1998.

[17] MÖLLER, S.: *Quality of telephone-based spoken dialogue systems*. Springer, US-New York NY, 2005.

[18] NIST Speech Recognition Scoring Toolkit, *Speech recognition scoring toolkit,* National Institute of Standards and technology, http://www.nist.gov/speech/tools, US-Gaithersburg MD, 2001.

[19] PICONE, J., DODDINGTON, G.R., PALLETT, D.S.: Phone-mediated word alignment for speech recognition evaluation, *IEEE Trans. Acoustics Speech and Signal Processing*, Vol. 38(3), pp. 559-562, 1990.

[20] PICONE, J., GOUDIE-MARSHALL, K.M., DODDINGTON, G.R., FISHER, W.: Automatic text alignment for speech system evaluation, *IEEE Trans. Acoustics Speech and Signal Processing*, Vol. 34(4), pp. 780-784, 1986.

[21] POLIFRONI, J., HIRSCHMAN, L., SENEFF, S., ZUE, V.: Experiments in evaluating interactive spoken language systems, *Proc. DARPA Speech and Natural Language Workshop*, US-Harriman CA, pp. 28-33, 1992.

[22] PRICE, P.J., HIRSCHMAN, L., SHRIBERG, E., WADE, E.: Subject-based evaluation measures for interactive spoken language systems, *Proc. DARPA Speech and Natural Language Workshop*, US-Harriman CA, pp. 34-39, 1992.

[23] SAN-SEGUNDO, R., MONTERO, J.M., COLÁS, J., GUTIÉRREZ, J., RAMOS, J.M., PARDO, J.M.; Methodology for dialogue design in telephone-based spoken dialogue systems: A Spanish train information system, *Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia)*, DK-Aalborg, 3, pp. 2165-2168, 2001.

[24] SIMPSON, A., FRASER, N.M.: Black box and glass box evaluation of the SUNDIAL system, *Proc. 3rd Europ. Conf. on Speech Communication and Technology (Eurospeech'93)*, DE-Berlin, 2, pp. 1423-1426, 1993.

[25] SKOWRONEK, J.: *Entwicklung von Modellierungsansätzen zur Vorhersage der Dienstequalität bei der Interaktion mit einem natürlichsprachlichen Dialogsystem*. Diploma thesis (unpublished), Institut für Kommunikationsakustik, Ruhr-Universität, DE-Bochum, 2002.

[26] STRIK, H., CUCCHIARINI, C., KESSENS, J.M.: Comparing the performance of two CSRs: How to determine the significance level of the differences, *Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia)*, DK-Aalborg, 3, pp. 2091-2094, 2001.

[27] STRIK, H., CUCCHIARINI, C., KESSENS, J.M.: Comparing the recognition performance of CSRs: In search of an adequate metric and statistical significance test, *Proc. 6th Int. Conf. on Spoken Language Processing (ICSLP 2000)*, CN-Beijing, 4, pp. 740-743, 2000.

[28] VAN LEEUWEN, D., STEENEKEN, H.: Assessment of recognition systems, *Handbook on Standards and Resources for Spoken Language Systems* (D. Gibbon, R. Moore and R. Winski, eds.), Mouton de Gruyter, DE-Berlin, pp. 381-407, 1997.

[29] WALKER, M.A., FROMER, J., DI FABBRIZIO, G., MESTEL, C., HINDLE, D.: What can I say?: Evaluating a spoken language interface to email, *Human Factors in Computing Systems. CHI'98 Conference Proc.*, US-Los Angeles CA, ACM, US-New York NY, pp. 582-589, 1998.

[30] WALKER, M.A., LITMAN, D.J., KAMM, C.A., ABELLA, A.: Evaluating spoken dialogue agents with PARADISE: Two case studies, *Computer Speech and Language*, Vol. 12(3), pp. 317-347, 1998.

[31] WALKER, M.A., LITMAN, D.J., KAMM, C.A., ABELLA, A.: PARADISE: A framework for evaluating spoken dialogue agents, *Proc. of the 35th Ann. Meeting of the Assoc. for Computational Linguistics*, ES-Madrid, pp. 271-280, 1997.

[32] ZUE, V., SENEFF, S., GLASS, J.R., POLIFRONI, J., PAO, C., HAZEN, T.J., HETHERINGTON, L.: JUPITER: A telephone-based conversational interface for weather information, *IEEE Trans. Speech and Audio Processing*, Vol. 8(1), pp. 85-96, 2000.

سلاسل التوصيات الصادرة عن قطاع تقييس الاتصالات

السلسلة A	تنظيم العمل في قطاع تقييس الاتصالات
السلسلة D	المبادئ العامة للتعريفة
السلسلة E	التشغيل العام للشبكة والخدمة الهاتفية وتشغيل الخدمات والعوامل البشرية
السلسلة F	خدمات الاتصالات غير الهاتفية
السلسلة G	أنظمة الإرسال ووسائطه والأنظمة والشبكات الرقمية
السلسلة H	الأنظمة السمعية المرئية وتعدد الوسائط
السلسلة I	الشبكة الرقمية متكاملة الخدمات
السلسلة J	الشبكات الكبلية وإرسال إشارات البرامج الإذاعية الصوتية والتلفزيونية وإشارات أخرى متعددة الوسائط
السلسلة K	الحماية من التداخلات
السلسلة L	إنشاء الكبلات وغيرها من عناصر المنشآت الخارجية وتركيبها وحمايتها
السلسلة M	إدارة الاتصالات بما في ذلك شبكة إدارة الاتصالات (TMN) وصيانة الشبكات
السلسلة N	الصيانة: الدارات الدولية لإرسال البرامج الإذاعية الصوتية والتلفزيونية
السلسلة O	مواصفات تجهيزات القياس
السلسلة P	**جودة الإرسال الهاتفي والمنشآت الهاتفية وشبكات الخطوط المحلية**
السلسلة Q	التبديل والتشوير
السلسلة R	الإرسال البرقي
السلسلة S	التجهيزات المطرافية للخدمات البرقية
السلسلة T	المطاريف الخاصة بالخدمات التلماتية
السلسلة U	التبديل البرقي
السلسلة V	اتصالات المعطيات على الشبكة الهاتفية
السلسلة X	شبكات المعطيات والاتصالات بين الأنظمة المفتوحة والأمن
السلسلة Y	البنية التحتية العالمية للمعلومات وملامح بروتوكول الإنترنت وشبكات الجيل التالي
السلسلة Z	لغات البرمجة والخصائص العامة للبرمجيات في أنظمة الاتصالات

طبعت في سويسرا
جنيف، 2005