RECOMMANDATION UIT-R BS.1693

Procédure de test des systèmes automatiques de requête par fredonnement

(Question UIT-R 8/6)

(2004)

L'Assemblée des radiocommunications de l'UIT,

considérant

a) que, à terme, des métadonnées accompagneront la plupart des systèmes de diffusion audio;

b) que la production automatique de métadonnées sera nécessaire pour offrir dans l'avenir un service complet présentant un bon rapport coût-efficacité;

c) que les systèmes de requête par fredonnement offrent un moyen naturel d'interrogation des bases de données audio;

d) que différents systèmes d'extraction de métadonnées sont mis au point actuellement;

e) que la Recommandation UIT-R BS.1657 – Procédure de test des systèmes automatiques d'identification audio, décrit une procédure de test des systèmes automatiques d'identification audio;

f) que le Groupe de Travail 11 de l'ISO/CEI JTC 1/SC 29 élabore actuellement, sous la forme définitive, des systèmes de codage de métadonnées pour les données multimédias;

g) que, jusqu'à présent, aucune procédure d'évaluation de la qualité des systèmes d'extraction de métadonnées audio relatifs à la reconnaissance des mélodies n'a été normalisée,

recommande

1 d'utiliser la procédure décrite dans l'Annexe 1 pour évaluer la qualité de fonctionnement des systèmes automatiques de requête par fredonnement.

Annexe 1

Procédure de test des systèmes automatiques de requête par fredonnement

1 Introduction

A l'heure d'un accroissement toujours plus grand des bases de données à contenu musical, qu'elles contiennent de véritables données audio ou des métadonnées associées («données sur les données»), l'exigence d'outils permettant de conserver ces masses de données devient également chaque jour plus urgente. Ce souhait n'est pas seulement exprimé par des professionnels, mais également par le simple amateur de musique utilisateur de l'Internet qui navigue fréquemment sur la Toile à la recherche de son style musical préféré. Pour faciliter l'extraction des données souhaitées, on peut distinguer ici deux niveaux d'abstraction:

– La recherche de métadonnées de haut niveau telles que les décrivait un auditeur (mélodie, rythme, timbre, instrumentation ou genre par exemple). On peut citer comme exemple d'application les systèmes de requête par fredonnement, qui peuvent être utilisés par les moteurs de recherche.

– L'extraction de métadonnées de niveau intermédiaire pour l'identification automatique de certaines interprétations de contenus musicaux. Des caractéristiques techniques décrivant les données audio (contenu spectral, etc.) sont générées puis comparées à une base de données connue, créant ainsi un lien vers des métadonnées pertinentes telles qu'un nom d'artiste, le titre d'une chanson, etc.

Pour un aperçu de l'état actuel des techniques en matière de systèmes de requête par fredonnement, on se référera au Document de l'ISMIR 2002 (3rd International Conference on Music Information Retrieval, IRCAM – Centre Pompidou Paris, France, octobre 2002).

2 Objet

Pour répondre aux exigences de l'industrie musicale, le taux de reconnaissance des techniques appliquées de requête par fredonnement doit être élevé et ne pas être dégradé par les altérations courantes subies par les représentations stockées dans la base de données musicales.

Ce problème est traité par un certain nombre de solutions différentes, souvent propriétaires, qui sont apparues récemment ([Clarisse et autres, 2002], [Ghias et autres, 1995], [Haus et Pollastri, 2001], [Heinz et Brückmann, 2003]). Pour toutes ces méthodes, cependant, les mêmes problèmes se posent quant à leur robustesse vis-à-vis de modifications ou de détériorations des données d'origine. Il convient donc de proposer que les systèmes de requête par fredonnement soient idéalement aussi précis et robustes vis-à-vis de modifications apportées aux signaux que le sont la perception et la reconnaissance humaine. Par conséquent, un système de requête par fredonnement sophistiqué doit être robuste vis-à-vis des différentes distorsions de qualité du signal et variations par rapport à la mélodie idéale. Par ailleurs, le maniement fiable de grandes bases de données musicales comprenant plusieurs milliers de chansons devrait également être assuré.

En conséquence, pour évaluer la qualité d'un système automatique de requête par fredonnement, il faut définir un environnement de test couvrant différents types de modification des signaux et décrivant la façon de déterminer d'autres paramètres essentiels du système. Une procédure de test unifiée est nécessaire pour parvenir à une évaluation objective de ces systèmes de requête.

3 Paramètres de qualité

Il convient de considérer les paramètres de qualité ci-après pour évaluer les systèmes de requête par fredonnement:

Données audio d'entrée requises:

– Doit-on fredonner une partie précise de la chanson ou peut-on en chanter une partie quelconque?

– Quelle est la taille minimale des données audio d'entrée pour obtenir un résultat fiable?

Taille de la représentation des données:

– Combien de données (octets) par chanson doivent être stockées dans la base de données musicales?

Taille de la base de données musicales:

– Combien de chansons peuvent être traitées dans la base de données musicales?

Mode d'identification:

– De quelle façon le type de données d'entrée (chant dans la langue maternelle, ou chant fredonnement d'un air tel que na na na, etc., utilisation d'un instrument de musique quelconque) a-t-il une incidence sur le taux de reconnaissance et sur la qualité de fonctionnement?

Vitesse de reconnaissance de la mélodie:

– Quel est le temps nécessaire pour identifier une mélodie?

– Comment varie cette durée avec le nombre de chansons figurant dans la base de données?

– Comment varie cette durée avec la qualité des données d'entrée?

Pour évaluer ces propriétés d'une manière réaliste et donc pour déterminer si un système est adapté à des applications réelles, un environnement de test doit présenter des conditions aux limites constantes en ce qui concerne les caractéristiques testées.

Les conditions de test doivent porter sur:

– la taille et le contenu de la base de données musicales (voir le § 4);

– la taille des données de requête (en termes de durée d'enregistrement) et le nombre d'enregistrements de test (voir le § 4);

– les règles exactes de modification des enregistrements de test (voir le § 5 et le § 6); et

– la plate-forme de calcul (spécification de l'unité centrale, de la mémoire et du système d'exploitation) (voir le § 7).

4 Sélection des données de test et taille de la base de données musicales

Une base de données contenant des échantillons musicaux de référence et à laquelle tous les systèmes pourraient adresser leurs requêtes devrait être définie. Elle devrait comprendre un mélange des différents styles musicaux (musiques populaires de différents pays, musique classique, …) avec une priorité donnée à l'échelle mondiale aux chansons les plus connues. Il faudra particulièrement veiller à éviter la duplication des enregistrements dans la base de données (reprises, etc.).

Une base de données musicales comprenant 500 à 1 000 chansons est suggérée pour une évaluation statistiquement fiable et pertinente.

Etant donné qu'il est difficile et coûteux d'élaborer des représentations abstraites de qualité élevée des chansons à mesure que celles-ci sont nécessaires pour la recherche dans la base de données, la création de la base de données musicales de référence est laissée à l'initiative des participants. Cette procédure induira un critère de qualité implicite qui trouvera son sens à la lecture des résultats de test obtenus. Tous les participants sont libres de choisir le format interne de la base de données puisque celui-ci dépend de l'algorithme de recherche.

Un ensemble d'enregistrements de test (base de données d'échantillons de requête) devrait être défini afin de satisfaire aux prescriptions suivantes: pour éviter tout étalonnage d'un ensemble spécifique de requêtes, chaque participant devrait fournir un nombre total de 200 mélodies de requête. L'adaptation des paramètres d'un système de requête par fredonnement à une base de données de requêtes fournies subjectivement peut ainsi être exclue. Les enregistrements de test devraient présenter une bonne qualité audio et être idéalement sans distorsion. Les données d'entrée devraient comprendre une diversité de types, telles que des paroles chantées, des mélodies fredonnées (da, na, ta, la, …) ou des compositions instrumentales. Elles devraient provenir d'un ensemble représentatif de chanteurs et de musiciens.

Tous les enregistrements de test doivent représenter des mélodies contenues dans la base de données de référence. Le rejet d'un enregistrement de test n'est pas une éventualité envisageable en raison du degré continu de similarité entre les mélodies.

Le nombre de systèmes de requête par fredonnement nouvellement testés devenant toujours plus grand, la taille de la base de données des échantillons de requête s'accroîtra sans cesse. Le test répété de ces systèmes sera donc nécessaire pour comparer les qualités de fonctionnement mesurées en utilisant une base de données des requêtes statistiquement toujours plus représentative. Une procédure de test automatique est recommandée.

5 Modifications

Pour représenter de façon plus réaliste les applications réelles, les échantillons de test de qualité élevée (voir le § 4) devraient être modifiés via l'application des sources habituelles de pollution sonore:

– compression audio (mp3, aac, ...);

– limitation de la largeur de bande (téléphonie, ...);

– quantification (modulation par impulsion et codage, loi A, ...);

– distorsion GSM (à plein débit, ...);

– bruit de fond (dû à un public, dans un restaurant, dans un magasin de musique, ...).

La liste des règles exactes est donnée dans le § 6.

6 Méthode de test

Le principal paramètre permettant d'évaluer la qualité des systèmes considérés sera le pourcentage de mélodies classées correctement. On peut distinguer deux cas:

– l'enregistrement recherché est classé en première position sur la liste des résultats présentés;

– l'enregistrement recherché figure parmi les dix mélodies considérées comme les plus ressemblantes par le système.

Ces chiffres ainsi que la vitesse du processus d'extraction et de recherche (classification) doivent être mesurés test par test.

6.1 Test 1

Lors du premier test, aucun titre de la base de données des échantillons de requête ne doit être modifié et tous les titres doivent être identifiés. Les conditions optimales en termes de qualité audio sont donc respectées et les résultats devraient présenter un taux d'identifications correctes très élevé.

6.2 Test 2

Il s'agit de tester la robustesse du système lorsque diverses modifications sont appliquées aux enregistrements de la base de données des échantillons de requête. Les modifications à choisir doivent correspondre à des distorsions acoustiques observées dans la vie courante.

– Distortion GSM:

  Les échantillons de test doivent faire l'objet de trois techniques distinctes de codage vocal utilisées en téléphonie mobile (GSM «plein débit», à «plein débit amélioré» et «à mi-débit»).

– Compression audio:

  Les exemples utilisés doivent être compressés/décompressés en utilisant des codecs audio MPEG-1/2 de couche 3 appliquant des débits de codage de 64, 96 ou 128 kbit/s. Le codec Fraunhofer classique est recommandé.

– Quantification:

  Les échantillons de test doivent faire l'objet d'une quantification de loi A non linéaire (8 kHz, 8 bits).

– Limitation de la largeur de bande:

  Le flux d'entrée est limité par une bande passante correspondant à une qualité de téléphonie courante, c'est-à-dire par la bande 300-3 400 Hz. Les caractéristiques de filtrage de la bande passante utilisée doivent satisfaire à une spécification d'inclinaison de pente d'au minimum 12 dB/oct.

– Bruit de fond:

  Pour avoir une base de données de distorsion quasi normalisée des signaux vocaux et du bruit de murmures confus (brouhaha) observés dans la vie quotidienne, il convient d'utiliser les indications du Document [Dreschler et autres]. Deux types distincts de signaux de bruit doivent être associés aux données de requête d'origine, à savoir respectivement, en utilisant les pistes de bruit et les versions atténuées (–6 dB):

  – murmures confus de 2 personnes (effort normal, piste 6);

  – murmures confus de 6 personnes (effort soutenu, piste 8).

7 Plate-forme de test

Il convient d'utiliser comme plate-forme de calcul et système d'exploitation des équipements adaptés à l'état d'avancement des techniques offertes à l'utilisateur courant. En 2004, on peut citer comme plate-forme appropriée et facilement disponible un ordinateur Pentium 4/Athlon XP fonctionnant à 2,4 GHz avec 512 mégaoctets de mémoire vive et utilisant Windows XP™ ou Linux.

8 Rapport de test

Les rapports de test devraient indiquer, aussi clairement que possible, la logique de l'étude, les méthodes utilisées et les conclusions auxquelles on a abouti. Ils devraient être assez détaillés pour qu'une personne raisonnablement compétente puisse, en principe, reproduire l'étude afin d'en vérifier empiriquement les résultats. Un lecteur informé devrait être capable de comprendre les principaux détails du test et d'en développer un point de vue critique, concernant par exemple les raisons sous-jacentes ayant motivé l'étude, les méthodes de conception expérimentales et leur mise en oeuvre, ainsi que les analyses et les conclusions.

Une attention particulière devrait être portée aux points suivants:

– la spécification et la sélection de la base de données musicales et de la base de données des échantillons audio;

– la description détaillée des systèmes testés;

– la description détaillée de l'ensemble des conclusions tirées.

Références bibliographiques

CLARISSE, L. P., MARTENS, J. P., LESAFFRE, M., DE BAETS, B., DE MEYER, H. et LEMAN, M. [octobre 2002] An Auditory Model Based Transcriber of Singing Sequences. ISMIR 2002, 3rd International Conference on Music Information Retrieval, IRCAM – Centre Pompidou Paris, France, p. 116-123.

GHIAS, A., LOGAN, J., CHAMERLIN, D. et SMITH, B. C. [1995] Query By Humming. Musical Information Retrieval in an Audio Database. Procs ACM Multimedia 1995, p. 231-236.

HAUS, G. et POLLASTRI, E. [2001] An Audio Front End for Systèmes de requête par fredonnement. Procs ISMIR 2001, p. 65-72.

HEINZ, Th. et BRÜCKMANN, A. [mars 2003] Using a Physiological Ear Model for Automatic Melody Transcription and Sound Source Recognition. AES 114th Convention. Amsterdam, Pays-Bas.

DRESCHLER, W. A., VERSCHUURE, H., LUDVIGSEN, C. et WESTERMANN S. ICRA Noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment. *Audiology*, 40, p. 148-157.