

قاعدة بيانات صوتية للغة العربية

Dr. Ghania Droua-Hamdani, Khoudir Benbellil, Khadidja Nesrine Boubakeur,
Dr. Youssef Bentrchia, Dr. Abdelkader Delmadji, Dr. Ben Aissa Kebir, Nawel Behloul, Dr. Aissa
Amrouche, Dr. Abed Ahcen, Dr. Khaled Baazi, Haddad Amel, Redouane Reggai

قاعدة بيانات للكلام باللغة العربية الفصحى، مسجلة من قبل متكلمين جزائريين مخصصة ORAL من 139 جملة متوازنة ORAL لأبحاث و تطبيقات التعرف الآلي على الكلام و المتكلم. تتكون صوتيا مسجلة من 55 متكلم (ذكور وإناث) مع مراعاة الخصائص الاجتماعية و الثقافية (الفئة العمرية والمستوى التعليمي واللكنة...الخ)

(مع بت العلامة)، أحادي القناة وتردد 16 كيلوهرتز PCM تمت التسجيلات بترميز 16 بت باستخدام معدات عالية الجودة في بيئة هادئة

في القطاعات الآتية ORAL يمكن استعمال القاعدة

- قطاع التعليم العالي والبحث العلمي
- قطاع الصحة
- قطاع الأمن الوطني
- قطاع الاتصالات السلكية واللاسلكية
- القطاعات التي تستخدم تكنولوجيات اللغة

ORAL est une base de données de signaux de parole pour l'Arabe standard, enregistrés par des locuteurs algériens, dédiée à la recherche et au développement des applications de la reconnaissance automatique de la parole et du locuteur.

ORAL est composé de 139 phrases phonétiquement équilibrées disposant de tous les contextes phonétiques de la langue. Ces phrases sont prononcées par 55 locuteurs (hommes et femmes) en tenant compte des spécificités sociales et culturelles (catégorie d'âge, niveau d'éducation, accent régional, etc.). Les enregistrements sont réalisés dans un environnement calme, avec un codage PCM 16 bits signés, mono, en utilisant un matériel d'enregistrement de haute qualité. La fréquence d'échantillonnage est de 16 kHz.

La base ORAL peut être utilisée dans les secteurs suivants:

- Secteur de l'enseignement supérieur et de la recherche scientifique
- Secteur de la santé
- Secteur de la sécurité nationale
- Secteur des télécommunications
- Secteurs utilisant les technologies linguistiques

ORAL is a speech database for Modern Standard Arabic sentences, recorded by Algerian speakers, dedicated to automatic speech and speaker recognition research and applications.

ORAL is composed of 139 phonetically balanced sentences. These sentences are pronounced by 55 speakers (male and female) taking into account several social and cultural characteristics (age category, education level, regional accent, etc.).

Recordings are made in a clean environment, with signed 16 bits PCM mono encoding, using high-quality recording equipment. The sampling frequency is 16 kHz.

The ORAL database can be used in the following sectors:

- Higher education and scientific research sector
- Health sector
- National security sector
- Telecommunications sector
- Sectors that use language technologies