

Corpus de données enregistrées – projet voiceHome

Table des matières

1	Généralités	1
1.1	Historique du document	1
1.2	Objet du document	2
2	Arborescence des fichiers	2
3	Nommage des fichiers	3
3.1	Indices des différents champs	3
3.2	Définition des différents champs	3
3.3	Format des fichiers	4
3.3.1	Format des fichiers textes	4
3.3.2	Fichiers de plan de pièce de maison/appartement.....	4
3.3.3	Fichiers de positions	4
3.3.4	Fichiers contenant la phrase à prononcer au format texte.....	6
3.3.5	Fichiers de transcription	6
3.3.6	Fichiers de réponse de salle (rir)	7
3.4	Tableaux récapitulatifs des lieux, locuteurs et types de bruits	7

1 Généralités

1.1 Historique du document

Version	Date	Auteurs	Commentaires
1.0	20/01/2016	Emmanuel Vincent Ewen Camberlein Romain Lebarbenchon	Version initiale
1.1	10/03/2016	Ewen Camberlein	- Précisions apportées sur les informations de positions et d'orientations ; - Modification du fichier de transcription ; - Mention du format UTF-8 pour les fichiers textes ; - Modifications mineures.
1.2	22/03/2016	Ewen Camberlein Romain Lebarbenchon	- Format de fichiers de la partie 3.3.3 : commentaire en début de ligne obligatoire

			<ul style="list-style-type: none"> - Suppression de « EL » dans le tableau récapitulatif des locuteurs ; - Mise a jour des tableaux récapitulatifs (lieu, phrases à prononcer et conditions de bruit) - Modifications mineures
--	--	--	---

1.2 Objet du document

Ce document décrit l'organisation et le contenu des fichiers de la base de données d'enregistrements réels, relative au projet voiceHome.

2 Arborescence des fichiers

L'arborescence retenue pour la présentation du corpus de données du projet est la suivante :

- annotations

|- rooms

|- ex : *home<>_room<>_map.pdf* : schéma d'une pièce

|- ex : *home<>_room<>_arrayPos<>.txt* :

position du centre de la l'antenne de microphones et orientation de celle ci

|- ex : *home<>_room<>_speakerPos<>.txt* :

position du locuteur

|- ex : *home<>_room<>_noiseCond<>.txt* :

types et positions des sources de bruit (si connues et fixes).

|- arrays

|- ex : *arrayGeo<>.txt* :

positions (exprimée par rapport au centre de l'antenne) des microphones

- transcriptions

|- prompts

|- ex : *uttNum<>.txt* : *phrase à prononcer*

|- clean

|- ex : *speaker<><>_uttNum<>.txt*:

instants de début et de fin et transcription de chaque phrase réellement prononcée

|- noisy

|- ex : *home<>_room<>_arrayGeo<>_arrayPos<>_speaker<><>_speakerPos<>*

_noiseCond<>_uttNum<>.txt :

instants de début et de fin et transcription de chaque phrase réellement prononcée

-audio

| - noises

| - ex : `home<>_room<>_arrayGeo<>_arrayPos<>_noiseCond<>.wav` :

bruit enregistré

| - ex : `home<>_room<>_arrayGeo<>_arrayPos<>_noiseCond<>_ref.wav` :

bruit de référence quand il est disponible (fichier audio, bande son de dvd etc...)

| - rirs

| - ex : `home<>_room<>_arrayGeo<>_arrayPos<>_speakerPos<>.wav`

| - clean

| - ex : `speaker<><>_uttNum<>.wav`

| - noisy

| - ex : `home<>_room<>_arrayGeo<>_arrayPos<>_speaker<><>_speakerPos<>
_noiseCond<>_uttNum<>.wav`

3 Nommage des fichiers

3.1 Indices des différents champs

Les indices utilisés se décrivent comme suit :

- `<>`: numéro (les numéros utilisés commencent à l'indice 1 !)
- `<><>`: Deux champs sont à renseigner pour les locuteurs :
 - M/F/C (Male/Female/Child)
 - Numéro

3.2 Définition des différents champs

- home + room
Ces champs représentent les lieux (appartements/ maison) et différentes pièces utilisées.
- arrayPos
Ce champ représente la position du centre de l'antenne de microphones dans la pièce (coordonnées cartésiennes) et son orientation.
- arrayGeo
Ce champ représente la position des microphones par rapport au centre de l'antenne de microphones (coordonnées cartésiennes) ainsi que leur orientation si ces microphones ne sont pas omnidirectionnels.
- speaker
Ce champ représente le locuteur.
- uttNum

Ce champ représente le numéro ou la lettre correspondant à (aux) phrase(s) prononcée(s).

- speakerPos

Ce champ décrit la position de la bouche du locuteur (coordonnées cartésiennes dans le repère de la pièce) et son orientation.

- noiseCond

Ce champ décrit le type de bruits et, lorsque cela a un sens, la position et l'orientation de la source de bruit.

3.3 Format des fichiers

3.3.1 Format des fichiers textes

Tous les fichiers textes sont encodés au format UTF-8.

3.3.2 Fichiers de plan de pièce de maison/appartement

Fichier *home<> room<> map.pdf*

Ce fichier est un plan grossier de chaque pièce comportant au minimum :

- la forme générale de la pièce
- l'origine et les axes du repère utilisé pour toutes les mesures de coordonnées (par convention, les axes xy correspondent au plan horizontal, l'axe z est l'axe vertical orienté du bas vers le haut et z=0 correspond au plancher)
- les positions approximatives de l'antenne de microphones, des locuteurs et des sources de bruit

3.3.3 Fichiers de positions

3.3.3.1 Généralités

Ces fichiers ont toujours le format suivant :

Commentaire [Tabulation] Valeur1 [Tabulation] Valeur2 [Tabulation] Valeur3 ... Valeur K [Retour chariot]
Commentaire [Tabulation] Valeur1 [Tabulation] Valeur2 [Tabulation] Valeur3 ... Valeur K [Retour chariot]
...

Chaque ligne débute par un commentaire obligatoire et une suite de valeurs décrivant, dans l'ordre :

- la position cartésienne en mètres sur l'axe x ;
- la position cartésienne en mètres sur l'axe y ;
- la position cartésienne en mètres sur l'axe z ;
- l'azimut en degrés par rapport à l'axe x, indiquant l'orientation de la source, du microphone ou de l'antenne de microphones ;
- l'élévation en degrés par rapport au plan xy, indiquant l'orientation de la source, du microphone ou de l'antenne de microphones.

3.3.3.2 Précision des données de position et d'orientation

Les positions du centre de l'antenne, de la bouche des locuteurs et des sources de bruits (lorsque cette information est disponible) sont exprimées avec une précision de +/- 5 cm (précision relative de nos outils de mesure, murs de la pièce rarement droits et rarement parallèles/perpendiculaires, etc.).

Les positions des microphones dans le repère de l'antenne sont exprimées avec une précision inférieure à 0.5 cm.

Les orientations de l'antenne, des microphones et des sources de bruits (lorsque cette information est disponible) sont exprimées avec une précision de +/- 5°.

L'orientation du locuteur est exprimée avec une précision de +/- 30°.

3.3.3.3 Détails

- Fichier *home<> room<> speakerPos<>.txt*:

Ce fichier contient un champ textuel décrivant librement la position de la bouche du locuteur, et de l'orientation de sa bouche.

Un exemple est le suivant :

sitting on the sofa	1.15	3.97	0.85	15	-5
---------------------	------	------	------	----	----

- Fichier *home<> room<> noiseCond<>.txt*:

Chaque ligne de ce fichier contient un champ textuel décrivant librement le type de bruit, suivi (s'il est ponctuel et immobile) de ses coordonnées dans le repère de la pièce considérée. Une condition de bruit peut correspondre à plusieurs sources de bruit, par exemple une chaîne HiFi avec deux haut-parleurs. La possibilité et la pertinence d'annoter ces sources individuellement ou non sont laissées à l'appréciation de l'annotateur. Par convention, *noiseCond1* correspond à une situation calme (*quiet condition*) sans aucune source de bruit spécifique identifiée.

Voici quelques exemples :

quiet condition

person walking

TV loudspeaker left	1.15	1.50	1.50	0	0
TV loudspeaker right	1.15	2.25	1.50	0	0

- Fichier *home<> room<> arrayPos<>.txt*:

Ce fichier contient un champ textuel décrivant librement la position de l'antenne de microphones, suivi des coordonnées cartésiennes de son centre dans le repère de la pièce considérée et de l'azimut et l'élévation du repère cette antenne par rapport au repère de la pièce considérée.

Un exemple est le suivant :

table of the kitchen	2.15	2.97	1.25	90	0
----------------------	------	------	------	----	---

- Fichier *home<> room<> arrayGeo<>.txt*:

Chaque ligne de ce fichier contient les coordonnées de chaque microphone dans le repère de l'antenne. Par convention, l'origine du repère est le centre de cette antenne. L'orientation des microphones est indiquée uniquement lorsqu'ils ne sont pas omnidirectionnels.

Un exemple est le suivant :

Side1_A8	0.0370	0.0560	-0.0380
Side1_H1	-0.0340	0.0560	0.0380
Side3_A8	-0.0560	0.0370	-0.0380
Side3_H1	-0.0560	-0.0340	0.0380
Side4_A8	-0.0370	-0.0560	-0.0380
Side4_H1	0.0340	-0.0560	0.0380
Side5_A8	0.0560	-0.0370	-0.0380
Side5_H1	0.0560	0.0340	0.0380

3.3.4 Fichiers contenant la phrase à prononcer au format texte

Fichiers *uttNum<>.txt*

Ce fichier contient la (ou les) phrase(s) à prononcer (une phrase par ligne). L'identifiant est un chiffre pour une phrase unique et une lettre pour une combinaison de plusieurs phrases. Un exemple est le suivant :

OK Vesta, enregistrer TF1 à partir de minuit durant 3 heures. OK Vesta, baisser la lumière.
--

3.3.5 Fichiers de transcription

Ces fichiers décrivent les instants de début et de fin de la phrase prononcée ainsi que la transcription de la phrase réellement prononcée (sans ponctuation pour les besoins de la reconnaissance vocale) :

- *speaker<><> uttNum<>.txt*
- *home<> room<> arrayGeo<> arrayPos<> speaker<><> speakerPos<> noiseCond<> uttNum<>.txt*

Le format concernant les fichiers de transcription est le suivant :

Time_stamp_start_seconds [Tabulation] Time_stamp_end_seconds [Tabulation] Label
Time_stamp_start_seconds [Tabulation] Time_stamp_end_seconds [Tabulation] Label
...

Le label [NO_SPEECH] est utilisé pour indiquer les plages de silence.

Un exemple de fichier de transcription est le suivant :

0.000000	14.994000	[\$NO_SPEECH]
14.994000	15.774938	OK Vesta
15.774938	16.310063	[\$NO_SPEECH]
16.310063	19.253563	enregistrer TF1 durant 3 heures
19.253563	28.999938	[\$NO_SPEECH]
28.999938	29.230140	OK Vesta baisse la lumière

3.3.6 Fichiers de réponse de salle (rir)

Fichier home<> room<> arrayGeo<> arrayPos<> speakerPos<>.wav

Ces fichiers décrivent les filtres de réponse de salle.

Ils sont encodés au format WAV avec une précision de 24 bits par échantillon.

3.4 Tableaux récapitulatifs des lieux, locuteurs et types de bruits

Les tableaux ci-dessous sont des bilans des différentes données à disposition dans le corpus.

Tableau récapitulatif des lieux

home 1	Appartement témoin du LOUSTIC à Rennes
room 1	Pièce centrale (cuisine/salon)
room 2	Salle de réunion

home2	Appartement (EV) à Nancy
room1	Salon / salle a manger
room2	Cuisine
room3	Chambre
room4	Salle de bain

home3	Appartement (D) à Nancy
room1	Salon
room2	Cuisine
room3	Chambre
room4	Salle de bain

home3	Appartement (II) à Nancy
room1	Salon / salle a manger
room2	Cuisine
room3	Chambre
room4	Salle de bain

Tableau récapitulatif des locuteurs

Male	Initials
M1	EC
M2	SP
...	...

Female	
F1	NB
...	...

Child	
...	...

Tableau récapitulatif des types de bruit

noiseCond	comment
1	quiet condition
2	Someone doing the dishes
3	TV
4	vacuum cleaner on (fixed position)
5	Someone reading (without speaking)
6	Someone speaking
7	Footsteps / with or without door opening / with or without stairs usage
8	Someone using the oven
9	Window opened (inside / outside noise) / with or without speech / with or without window opening
10	Someone preparing meals
11	Lunch noise / with or without people speaking
12	Door opened to the patio (noise from outside) / with or without door opening
13	Door opened to the living room (TV is on in the living room) / with or without door opening
14	Someone using roller shutters
15	Radio
16	Someone using bathroom sink (hands washing or teeth brushing) / with or without electrical device usage
17	Shower / bathtub usage
18	Someone using the washing machine
19	Someone using the clothes dryer
20	Playing board games (chess, dice, ...) / with or without someone answering to a phone call in an other room
21	Dialog
22	Opening fridge / doing the dishes
23	Set or clear the table
24	Blender usage
25	Loud speakers (music)
26	Tidying up storage unit / electrical device usage
27	Doors opening and closing / footsteps / items tidying up
28	Someone vacuuming the room
29	Lunch noise in an other room (door opened) with or without people speaking
30	Microwave or boiler usage
31	Someone using a laptop
32	Toilet flushing

Tableau récapitulatif des phrases prononcées

Phrases uniques :

uttNum	sentence
1	OK VESTA, allume la télévision.
2	OK VESTA, coupe le son.
3	OK VESTA, enregistre TF1 à partir de 18 heures 20 durant 2 heures.
4	OK VESTA, et beaucoup plus tard ?
5	OK VESTA, lecture.
6	OK VESTA, mettre le son sur 13.
7	OK VESTA, programme précédent.
8	OK VESTA, redémarre le programme.
9	OK VESTA, un film de Woody Allen sans Samuel Lee Jackson en tant qu'acteur.
10	OK VESTA, zap sur NRJ12.
11	OK VESTA, active la caméra de la chambre.
12	OK VESTA, ai-je plus consommé en électricité cette année par rapport à l'année dernière ?
13	OK VESTA, arrêter les stores.
14	OK VESTA, augmente la température à 21 degrés.
15	OK VESTA, augmenter la température de 3 degrés.
16	OK VESTA, diminue à 50% les luminaires.
17	OK VESTA, est-ce que l' éclairage de la piscine est allumé ?
18	OK VESTA, ferme au trois quart les volets.
19	OK VESTA, ouvrir à la moitié le store.
20	OK VESTA, redémarre le programme.
21	OK VESTA, chaîne suivante.
22	OK VESTA, donne moi son résumé ?
23	OK VESTA, enregistrer TF1 à partir de minuit durant 3 heures.
24	OK VESTA, j'aimerais que tu m' enregistres le prochain Desperate Housewives !
25	OK VESTA, mets le son sur 17.
26	OK VESTA, monte le volume de 10.
27	OK VESTA, qu'est-ce que tu conseilles ?
28	OK VESTA, réveil.
29	OK VESTA, volume à 7.
30	OK VESTA, éteins la télévision.
31	OK VESTA, activer la surveillance extérieure.
32	OK VESTA, allumer le chauffage.
34	OK VESTA, augmente l' éclairage du jardin à 80%.
35	OK VESTA, augmenter de 20% les éclairages de la piscine.
36	OK VESTA, descendre le portail à la moitié.
37	OK VESTA, désactive la surveillance intérieure.
38	OK VESTA, est-ce que la climatisation fonctionne ?
39	OK VESTA, moyenne composés organiques volatiles.
40	OK VESTA, quelle est la consommation par jour en gaz de la climatisation ?
41	OK VESTA, baisse le volume de 15.
42	OK VESTA, enregistre TF1 durant 1 heure 45.
43	OK VESTA, enregistres le prochain Super 8 !
44	OK VESTA, je le regarderai plus tard.
45	OK VESTA, mettre la 2.
46	OK VESTA, pause.
47	OK VESTA, qu'est-ce qu'il y a à la télé ?
48	OK VESTA, trouve-moi un film de Quentin Tarantino.
49	OK VESTA, y'a quoi à la TV ?

50	OK VESTA, éteindre la télé.
51	OK VESTA, ai-je plus consommé en eau ce mois-ci par rapport au mois dernier ?
52	OK VESTA, arrête le volet à la moitié.
53	OK VESTA, augmente la lampe de 50%.
54	OK VESTA, augmenter la lampe.
55	OK VESTA, différence de consommation en gaz entre ce mois-ci et le mois dernier.
56	OK VESTA, est-ce que j'ai plus consommé en chauffage cette année par rapport à l'année dernière ?
57	OK VESTA, est-ce que la caméra est éteinte dans le salon ?
58	OK VESTA, ouvre la porte du garage.
59	OK VESTA, quelle est la consommation électrique annuelle du luminaire ?
60	OK VESTA, éteins la climatisation.

Combinaisons de phrases (exemples – non utilisé pour le moment):

uttNum	Reference sentences	sentences
A	1-2	OK Vesta, enregistrer TF1 à partir de minuit durant 3 heures. OK Vesta, baisser la lumière.
B	1-3	OK Vesta, enregistrer TF1 à partir de minuit durant 3 heures. OK Vesta, zappe sur la 6.
...	...	