Creation of a corpus of realistic urban sound scenes with controlled acoustic properties

Jean-Rémy Gloaguen[1], Arnaud Can[1],
Mathieu Lagrange[2], Jean-François Petiot[2]

[1]French Institute of Science and Technology for Transport, Development and Networks,
Environmental Acoustics Laboratory
Allée des Ponts et Chaussées, Route de Bouaye, - CS4, 44344 Bouguenais Cedex, FR

[2] LS2N,
1, rue de la Noë, BP 92101, 44321 Nantes Cedex 3, FR

Acoustics' 17 Boston
06/29/2017

Introduction

Urban sound environments

- Subject of many researches both on perceptual and physical aspects
- Based on real sound environments with *soundwalks* or recordings listened on laboratory
  $\Rightarrow$ good ecological validity
  $\Rightarrow$ no controlled framework
  $\Rightarrow$ no repeatable experience
- Based on sound mixtures based on simulation process
  $\Rightarrow$ controlled framework
  $\Rightarrow$ creation of specific urban sound environments
  $\Rightarrow$ low ecological validity

How can we simulate sound mixtures of urban sound environments that are realistic enough ?

Introduction

Different approaches

- Simulate an entire neighborhood (buildings, sound emissions and propagation)
    - too heavy to implement
    - still too artificial
- Design sound mixtures by superposing sound events with sound backgrounds
    - build a representative and clean sound database
    - structure correctly the disposition of the events

Proposition

- Transcribe real recordings
- use the transcription to simulate controled stimuli
- Validate the realism of the simulated sound mixtures obtained with a perceptual test

Table of contents

SimScene: a simulator to design sound environments [1]

Principe

- Mix sound events and a background taken from an isolated sound database
- Control a set of high level parameters
  - onset time and sound level of each sound class
  - number of occurrence of a sound class
- Standard deviation on each parameter to bring random behavior
- Generate a separate audio file for each sound class
  ⇒ enable to know their exact contribution

2 modes to design the sound mixtures

- *abstract*: the user sets abstract information (mean / variance)
- *replicate*: the sound classes and the temporal position of each events come from an annotated text file

---

[1] https://bitbucket.org/mlagrange/simscene

SimScene: a simulation tool to design sound environments

Creation of the isolated sound database
- Sounds found online (*fressound.org*)
- Based on *UrbanSound8k* database (REF)

Recordings of car passages on Ifsttar-Nantes runway
- 4 cars (Renault Senic, Megane, Clio and Dacia Sandero)
- 104 car passages recorded for constant speeds (from 20 to 90 km/h), acceleration and braking phase for multiple gear ratios and stopped car

To sum up
- 154 samples of sound background (traffic, crowds, birds, rain . . .)
- 245 samples of sound events (car passages, birds, car horns . . .)

Study of urban sound recordings

Corpus of urban recordings

- Recordings made in the 13th district in Paris (FR) (REF)
- 19 recorded points on a define path
- Recordings made on 2 days (03/23/2015 and 03/30/2015), twice a day
- Among the 76 audio files available, 74 are used
- Multiple sound environments
  - *Park*, *quiet street*, *noisy street* and *very noisy street*

Figure: Map of the soundwalk with the 19 stop points

Study of urban sound recordings

- Annotation of each audio file
  - sound classes, begin and end times

From the observations and the annotations

- Synthesis of the 74 real scenes with *simScene* with the *replicate* mode
- 15 seconds example of a simulated scene

Play

$\Rightarrow$ Are the simulated scenes realistic enough ?

Design of the perceptual test

Objective

- Evaluation of the realism of the simulated scene among real recordings by a panel of listeners

- 40 audio tested
  - 30 seconds duration
  - Normalization to the same the sound level
  - 20 audio files from the real recordings and the same 20 replicated
- 50 participants
  $\Rightarrow$ Each judge cannot evaluate all the scene

Partially Balanced Incomplete Block Design

- 20 audio files listen by each judge
- Design the order of listening for each judge with a fair statistical distribution
- Each judge listen a mix of 10 real and 10 replicated scenes

Design of the perceptual test

- Web-page administered online on the 8 February 2017 and closed 12 days later
- For each scene listened
  - answer to the question 'Is the scene you listen seems, to you, realistic ?'
  - evaluation on a 7 points scale
  - possibility to re-listen the scene and to put a comment
- The gender (H/F), the age and the experience in the listening of urban sound environments are then asked

The panel

- 18 females and 31 males (1 not documented)
- Average age : 36 ($\pm$ 12) years old
- 62 % of the participant declared having NO experience in the listening of urban sound mixtures

Results

According to the 'type'
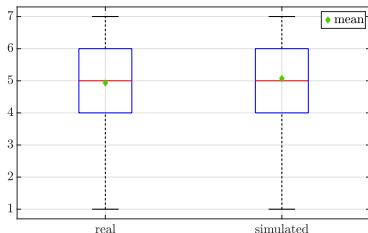
Figure: Box-and-whiskers plot of the rating of realism according to the type of scene

- Hypothesis test H_0 : the distribution of the scores between the real and simulated scenes are similar
- Significance level $\alpha = 5$ % to reject H_0

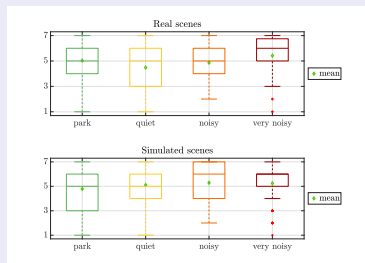|	DOF	$|t|$	p-value
type	49	1.37	0.17

Table: Paired sample t-test

- T-test performed between the mean scores of the real and simulated scenes for each judge
- Real and simulated scenes evaluated in a similar way (p-value $> \alpha = 0.05$)

$\Rightarrow$ Validation of the urban sound mixtures realism

Results

According to the 'type', judge' and 'sound env.' factors

Figure: Data distribution for the 'scene type' and 'sound environment' factors

	DOF	F-stat.	p-value
type	1	1.38	0.24
sound env.	3	4.69	3.60×10^{-3}
judge	49	5.00	0.18×10^{-9}
type/sound env.	3	6.80	0.20×10^{-3}
type/judge	49	1.07	0.35
sound env./judge	147	1.43	1.60×10^{-3}

Table: Three-ways ANOVA

- Distribution between real and simulated scenes non significant again
- Interaction between sound env./judge and type/sound env.
  $\Rightarrow$ Evolution of the distributions different according to the levels

Results

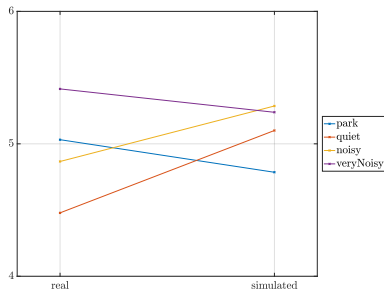## Interaction between the 'type' and 'sound environment' factors

Figure: Mean scores for the 4 sound environments according to the type

- We cannot conclude on the influence of the factors on the score
- Better mean scores for the 'noisy' and 'veryNoisy' simulated scenes
- Comments let by the panelists reveal that some noises are too loud (foot steps and birds) on the park and quiet simulated scenes

Conclusion

- Design of a corpus of realistic urban scenes
- Based on the annotation of 74 real recordings
- Use of the simulation software *simScene*
- Representative sound database of isolated sounds with car passages recorded specifically
- The realism of the simulated scene is validated by a perceptual test
  - No notable difference between the evaluation of the real and the simulated scenes
  - The 'park' and 'quiet street' atmospheres are both perceived as less realistic than the 'noisy' and 'very noisy street' atmosphere
  - some improvement can be made on the sound level of some sound classes (bird, foot step)
- Urban sound scene corpus that can be use to test sound recognition, classification or separation tools or to realize specific perceptual tests

Creation of a corpus of realistic urban sound scenes with controlled acoustic properties

Jean-Rémy Gloaguen[1], Arnaud Can[1],
Mathieu Lagrange[2], Jean-François Petiot[2]

[1]French Institute of Science and Technology for Transport, Development and Networks,
Environmental Acoustics Laboratory
Allée des Ponts et Chaussées, Route de Bouaye, - CS4, 44344 Bouguenais Cedex, FR

[2] LS2N,
1, rue de la Noë, BP 92101, 44321 Nantes Cedex 3, FR

Acoustics' 17 Boston
06/29/2017