Audio Files Repository

CONSAT
Telematics

Table of Contents

1. Introduction

This document explains the audio files repository structure for the CTS master audio repository (Consat Telematics Solution).

The repository is used by CTS to import sounds used for calling stops, destinations, lines, etc. in the vehicles, at bus stops and in other output media. During the data import process, sound files are fetched from this repository when needed to match the traffic data, and then recoded to the output format of the media and finally packaged for download by the vehicles, stop signs, etc.

This document explains how to name the files as well as providing examples, a list of miscellaneous sound files and a phonetic map.

2. Audio Files Repository Structure

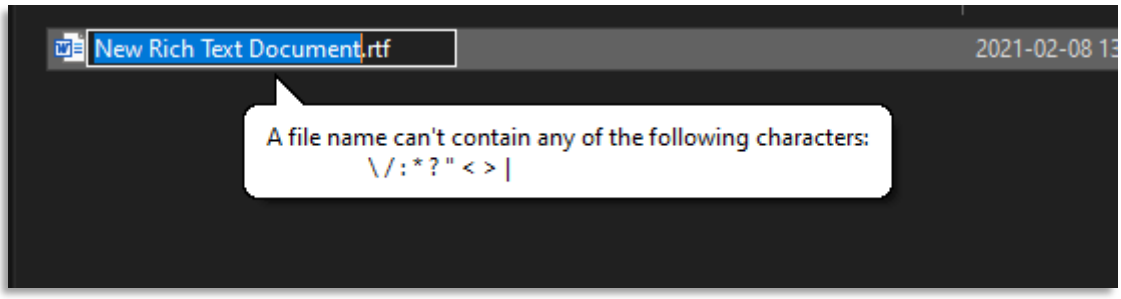
Audio files should be placed into a repository based on one of two naming conventions:

- External Id
    - The reference Id of the entity found in the data import, i.e. stop point id, destination id, etc.
- Name
    - The text representation of the entity found in the traffic data import, i.e. the stop point name, destination text, etc.

Each category must use one of the above naming conventions, but they can vary in different categories.

It is required that the names/ids used exactly match a corresponding entity in the traffic data.

If the name is used, remember that depending on the file system used, some characters are not allowed. For example, in the case of a standard NTFS windows file system (default), the following characters are not allowed (see figure).



In such cases, external Id naming must be used or these characters must be replaced with a space.

The voice sound file format is WAV 44.1 kHz stereo. The files can be FLAC encoded to use less space. Other formats can be used, but they are not included in the standard package due to possible licensing issues and/or loss of raw quality. Support for additional formats such as mp3, ogg, spx and opus, might incur additional costs. Please contact Consat Telematics for more information.

CTS also supports synthetic sound generation, i.e. text to speech conversion. In such cases, there is no need for raw audio files. However, a phonetic map might be required to supply rules about how to pronounce certain texts. A license is required to be able to use

text to speech in CTS, which incur costs per second of generated sound. Note that the cost is only incurred once for each text entity, and not for every call. Changes to the text is treated as a new entity and the same costs are incurred.

It is also possible to mix voice audio files with synthetic text to speech, where CTS supplements all missing audio entities with an automatic generated sound from the text representation in the traffic data. If a phonetic map is supplied, it will be used during this process.

The full tree for the repository looks like this:

```
├──common
│   ├──config
│   │      clock.csv
│   │      filename_map.csv
│   │      misc.csv
│   │      options.props
│   │      phonetic_map.csv
│   │
│   ├──destinations
│   │   ├──common
│   │   │   ├──speaker-default
│   │   │   ├──speaker-id-XX
│   │   │   └──speaker-id-XY
│   │   ├──sign
│   │   │   ├──speaker-default
│   │   │   ├──speaker-id-XX
│   │   │   └──speaker-id-XY
│   │   └──vehicle
│   │       ├──speaker-default
│   │       ├──speaker-id-XX
│   │       └──speaker-id-XY
│   ├──lines
│   │   ├──common
│   │   │   ├──speaker-default
│   │   │   ├──speaker-id-XX
│   │   │   └──speaker-id-XY
│   │   ├──sign
│   │   │   ├──speaker-default
│   │   │   ├──speaker-id-XX
│   │   │   └──speaker-id-XY
```

```
│   │   └──vehicle
│   │       ├──speaker-default
│   │       ├──speaker-id-XX
│   │       └──speaker-id-XY
│   ├──misc
│   │   ├──clock
│   │   │   ├──common
│   │   │   │   ├──speaker-default
│   │   │   │   ├──speaker-id-XX
│   │   │   │   └──speaker-id-XY
│   │   │   ├──sign
│   │   │   │   ├──speaker-default
│   │   │   │   ├──speaker-id-XX
│   │   │   │   └──speaker-id-XY
│   │   │   └──vehicle
│   │   │       ├──speaker-default
│   │   │       ├──speaker-id-XX
│   │   │       └──speaker-id-XY
│   │   ├──phrases
│   │   │   ├──common
│   │   │   │   ├──speaker-default
│   │   │   │   ├──speaker-id-XX
│   │   │   │   └──speaker-id-XY
│   │   │   ├──sign
│   │   │   │   ├──speaker-default
│   │   │   │   ├──speaker-id-XX
│   │   │   │   └──speaker-id-XY
│   │   │   └──vehicle
│   │   │       ├──speaker-default
│   │   │       ├──speaker-id-XX
│   │   │       └──speaker-id-XY
│   │   └──stopPointNames
│   │       ├──common
│   │       │   ├──speaker-default
│   │       │   ├──speaker-id-XX
│   │       │   └──speaker-id-XY
│   │       ├──sign
│   │       │   ├──speaker-default
```

```
│   │       │   ├──speaker-id-XX
│   │       │   └──speaker-id-XY
│   │       └──vehicle
│   │           ├──speaker-default
│   │           ├──speaker-id-XX
│   │           └──speaker-id-XY
│   └──stops
│       ├──stopAreas
│       │   ├──common
│       │   │   ├──speaker-default
│       │   │   ├──speaker-id-XX
│       │   │   └──speaker-id-XY
│       │   ├──sign
│       │   │   ├──speaker-default
│       │   │   ├──speaker-id-XX
│       │   │   └──speaker-id-XY
│       │   └──vehicle
│       │       ├──speaker-default
│       │       ├──speaker-id-XX
│       │       └──speaker-id-XY
│       └──stopPoints
│           ├──common
│           │   ├──speaker-default
│           │   ├──speaker-id-XX
│           │   └──speaker-id-XY
│           ├──sign
│           │   ├──speaker-default
│           │   ├──speaker-id-XX
│           │   └──speaker-id-XY
│           └──vehicle
│               ├──speaker-default
│               ├──speaker-id-XX
│               └──speaker-id-XY
└──company-xx
    ├──config
    │   filename_map.csv
    │   misc.csv
    │   options.props
```

```
│       phonetic_map.csv
│
├───destinations
│   ├───common
│   │   ├───speaker-default
│   │   ├───speaker-id-XX
│   │   └───speaker-id-XY
│   ├───sign
│   │   ├───speaker-default
│   │   ├───speaker-id-XX
│   │   └───speaker-id-XY
│   └───vehicle
│       ├───speaker-default
│       ├───speaker-id-XX
│       └───speaker-id-XY
├───lines
│   ├───common
│   │   ├───speaker-default
│   │   ├───speaker-id-XX
│   │   └───speaker-id-XY
│   ├───sign
│   │   ├───speaker-default
│   │   ├───speaker-id-XX
│   │   └───speaker-id-XY
│   └───vehicle
│       ├───speaker-default
│       ├───speaker-id-XX
│       └───speaker-id-XY
├───misc
│   ├───clock
│   │   ├───common
│   │   │   ├───speaker-default
│   │   │   ├───speaker-id-XX
│   │   │   └───speaker-id-XY
│   │   ├───sign
│   │   │   ├───speaker-default
│   │   │   ├───speaker-id-XX
│   │   │   └───speaker-id-XY
```

```
│   │   └──vehicle
│   │       ├──speaker-default
│   │       ├──speaker-id-XX
│   │       └──speaker-id-XY
│   ├──phrases
│   │   ├──common
│   │   │   ├──speaker-default
│   │   │   ├──speaker-id-XX
│   │   │   └──speaker-id-XY
│   │   ├──sign
│   │   │   ├──speaker-default
│   │   │   ├──speaker-id-XX
│   │   │   └──speaker-id-XY
│   │   └──vehicle
│   │       ├──speaker-default
│   │       ├──speaker-id-XX
│   │       └──speaker-id-XY
│   └──stopPointNames
│       ├──common
│       │   ├──speaker-default
│       │   ├──speaker-id-XX
│       │   └──speaker-id-XY
│       ├──sign
│       │   ├──speaker-default
│       │   ├──speaker-id-XX
│       │   └──speaker-id-XY
│       └──vehicle
│           ├──speaker-default
│           ├──speaker-id-XX
│           └──speaker-id-XY
└──stops
    ├──stopAreas
    │   ├──common
    │   │   ├──speaker-default
    │   │   ├──speaker-id-XX
    │   │   └──speaker-id-XY
    │   ├──sign
    │   │   ├──speaker-default
```

```
│   │   ├──speaker-id-XX
│   │   └──speaker-id-XY
│   └──vehicle
│       ├──speaker-default
│       ├──speaker-id-XX
│       └──speaker-id-XY
└──stopPoints
    ├──common
    │   ├──speaker-default
    │   ├──speaker-id-XX
    │   └──speaker-id-XY
    ├──sign
    │   ├──speaker-default
    │   ├──speaker-id-XX
    │   └──speaker-id-XY
    └──vehicle
        ├──speaker-default
        ├──speaker-id-XX
        └──speaker-id-XY
```

There are several folders at the top level. The “**common**” folder is used when there is no “company override folder”. If no company overrides are needed, only the common folder is used.

Each folder, such as destination, lines, etc., has the possibility to have a customization for a specific target, such as special sound for the sign and another one for a vehicle(bus).

Each sub section, such as destination, lines, etc., can be given a speaker selection by adding directories with the name of the speaker.

There are several folders to separate the sounds for different categories.

- Destinations
    - A sound for each destination.
- Lines
    - A sound for each line.
- Misc
    - clock
        - A sound for each number 0-59, used for call “14:01”.
    - stopPointNames
        - The stop point designation, such as “A”, “1”, used normally at bus stop signs.

    - phrases
        - Different phrases used when calling, such as “next”, “line”, “towards”, “via”. See appendix list for required sounds for different output media.

- Stops
    - Stopareas
        - A sound for each stop area.
    - Stoppoints
        - A specific/optional sound to override the stop area sound for a certain stop point.

- Config
    - Filename_map.csv
        - A map file to map traffic database name to filenames if the text name in the traffic database does not match the filename for the audio file (when using named audio files)
    - Misc.csv
        - A map file to produce sounds for misc if synthetic sounds are used. The phonetic map is referenced. Requires a text to speech license
    - Clock.csv
        - A map file to produce additional sounds for clock sounds if synthetic sounds are used. The phonetic map is referenced. Requires a text to speech license. By default all “numbers” are generated by default.
    - Options.props
        - A file with different options to control the generation system.
            - createSyntheticSoundsWhenMissing=false
                - If the generation system should automatically generate synthetic files for missing sounds in the traffic data. Requires a text to speech license.
            - useFilenameMap=filename_map.csv
                - Which filename map to use to map traffic database name to filename when implicit name match does not work.
            - usePhoneticMap=phonetic_map.csv
                - Which phonetic map to use if synthetic sound creation is used.
            - tempo=<0.10..10.0>
                - If set to anything but 1.0, will change the tempo of the resulting sound. 0.5 will halve the speed, 2.0 will double the speed.
    - phonetic_map.csv

- A text to phonetic conversion file is used if synthetic sounds should be automatically created for all missing sounds. The map will override a traffic database text if found, and will use the text in the map instead. Useful if the name is not pronounced as it is spelled, or if the text in the traffic is shortened for some reason.

Each category has a few subcategories:

- Common
  - If a file is placed here, it is used for all output Medias.
- Sign
  - A file here overrides a file in **common** and will be used on bus stop sign.
- Vehicle
  - A file here overrides a file in **common** and will be used for next stop calls in vehicles.

If no customizations are necessary, all files can be placed in **common**. However, in the instance where synthetic sounds are used for signs, but are recorded for buses, the “sign and vehicle directories” can be used to separate them.

Each subcategory has a few subsections to make it possible to have different speakers for different fleets.

- speaker-default
  - If a file is placed here, it is used if no speaker is set on the fleet, or the file does not exist in the set speaker.
- speaker-xx
  - Files for speaker xx, will be used for fleet with speaker xx set.

The selection of which speaker to be used is configured on the fleet in the depot configuration UI. If no speaker is given, it will fall back to speaker-default. If only one speaker is ever used, all sounds should be placed in speaker-default.

Please have a look at the attached zip-fil if you would like to get a starting directory structure from which you can work.

repobase.zip

2.1 Simple Example

This is a very simple example of how it can look with only one customization for a specific stop point on the stop area “Main Street”. The stop point’s id is 1010.

This example only shows one of all the 60 clock sounds and does not include all the misc sounds (see appendix).

```
└─────common
    ├─────config
    │       clock.csv
    │       filename_map.csv
    │       misc.csv
    │       options.props
    │       phonetic_map.csv
    │
    ├─────destinations
    │    └─────common
    │         └─────speaker-default
    │                 Central Station.wav
    │
    ├─────lines
    │    └─────common
    │         └─────speaker-default
    │                 13.wav
    │
    ├─────misc
    │    ├─────clock
    │    │    └─────common
    │    │         └─────speaker-default
    │    │                 13.wav
    │    │
    │    ├─────phrases
    │    │    ├─────common
    │    │    │    └─────speaker-default
    │    │    │             line.wav
    │    │    │
    │    │    ├─────sign
```

```
│   │   │   └──speaker-default
│   │   │           departsIn.wav
│   │   │
│   │   └──vehicle
│   │       └──speaker-default
│   │               next.wav
│   │
│   └──stopPointNames
│       └──common
│           └──speaker-default
│                   A.wav
│
└──stops
    ├──stopAreas
    │   └──common
    │       └──speaker-default
    │               Main street terminal.wav
    │
    └──stopPoints
        ├──common
        │   └──speaker-default
        │           1010.wav
        │
        └──sign
            └──speaker-default
                    1010.wav
```

3. Miscellaneous Sounds

- Vehicle
  - line.wav – played before calling a stop externally
  - next.wav – played before calling a stop
  - test.wav – test sound played during preop checks
  - touch.wav – played if the stop button has been pressed, before calling the next stop
  - via.wav
  - volume_test.wav – played when the test volume button pressed in the vehicle

- Sign
  - about.wav – played before announcing the departure time
  - allStopsCancelled.wav – called if all departures
  - cancelled.wav – called if a departure is cancelled
  - ding.wav – played before an announcement
  - line.wav – called before announcing the line number
  - noDepartures.wav – called if there are no departures at the moment
  - trafficJam.wav – called if the vehicle for the departures is in congestion.
- Clock
  - timeIs.wav – played before announcing the time
  - now.wav –played if the departure is within 30s
  - minute.wav – played after announcing 1 minute left before departure
  - minutes.wav – played after announcing minutes left before departure

The misc sounds for bus stop signs can contain arbitrary call phrases depending on how the signs are configured, but the sounds shown above are the default phrases required.

4. Phonetic Map

A CSV file containing 2 fields, i.e. the text to match and the substitution text.

```
Text;Phonetic
20th main street & cross road;twentieth main street at cross road
```

It will map any text in any category and replace it with the replacement text.

5. Filename Map

A CSV file containing 3 fields, i.e. the type of name, line, stop, destination or all; the text to that matches the traffic database name; and the filename to use for that text.

```
Wave File;Spoken text;File name;codec;segment time
line;10;Line10
stop;Term 5;Terminal 5
dest;Central Stn;Central station
all;Park avn;Park avenue
```

6. Misc and Clock Map

A CSV file containing 3 - 5 fields.

```
Wave File;Spoken text;File name;codec;segment time
;minute;minute;
;minutes;minutes;
;now;now;
stopat;Stopping at;touch
;midnight;midnight;
;noon;noon;
;9.05am;am;spx;0.938-
;9.05pm;pm;spx;0.931-
;o'clock;aclock;spx;
```

“Wave File” is the wave file for the sound if there is one. If none is specified, and synthetics are enabled, it will use the “File Name” column and generate an audio file named after the “File Name”.

Text is the text to synthetize, if no wave file exists

File Name is the name of the converted file. This name must map to the phrase needed by the software for the different calls/phrases to be used.

Codec is the codec to use after transcoding, i.e. spx. This can be omitted.

Segment time is used to split out a certain sequence from a file, specified as a range in seconds such as 0.387-1.243, which will split out the sound from 0.387 to 1.243, or 0.485-, which will skip the first 0.485 seconds to the end of the sound. This can be omitted.