**Guidelines for producing a database of continuous acoustic environment recordings in a neonatal intensive care unit**

Authors: Ganna Raboshchuk, Climent Nadeu  
TALP Research Center, Dept. TSC, UPC  
January 2017

**Index**
1. Recording setup for database acquisition  
2. Labelling protocol  
   2.1. Equipment alarms  
   2.2. Vocalization sounds  
3. Database structure  
4. Data preparation  
Annex 1. Checklist for recordings

**1. Recording setup for database acquisition**

This section provides a step-by-step guide of the recording setup for the database acquisition. The highlighted words correspond to the highlighted parts of the figures (same color is used).

**Follow the exact steps given below to prepare the recording setup**

I. **Setting up Audiobox** (see Figure 1 and Figure 2)
   A. Connect the microphone with the **green label** to the first input (1 Mic) as on Figure 1. Connect the other microphone to the **2 Instrument** input.  
   B. Switch on 48V button. The **red light** must appear.  
   C. Set the knobs to the following positions:  
      1. **INST MIC 1 and INST MIC 2**: to +25 (5 grades back from +30 max)  
      2. **Main**: max (according to last session, this may have to be changed at each recording session)  
      3. **Mixer**: down to “INPUTS”  
   D. Connect **green** and **yellow** cables to Audiobox as shown on Figure 2.  
   E. Connect Audiobox to the computer via **USB**.

---

*Also a set of child physiological variables was simultaneously recorded with polysomnography equipment. These signals and a frontal video signal were included in the database. Work done in collaboration with two clinical teams of Hospital Materno-Infantil Sant Joan de Déu: 1) Neonatal care, led by Ana Riverola de Veciana, and 2) Sleep disorders, led by Oscar Sans Capdevila.*
Figure 1. Audiobox setup: frontal view

Figure 2. Audiobox setup: rear view
II. Setting up mixer (See Figure 3)
   A. Connect the green wire to the left mixer input and the yellow wire to the right mixer input - LINE IN 4/5.
   B. Set the knobs to following the positions:
      1. Level of the 4/5 line - to 0 (middle)
      2. Level of the other lines - to -∞ (minimum)
      3. Level of MAIN MIX - to 0 (middle)

III. The earphones should be prepared, here or later Setting up Audacity (see Figure 4)
   A. Choose Audiobox microphones for recording.
   B. Make sure there is enough space on the computer and the sampling frequency is 44100 Hz.
   C. Press Record button to start the recording.
   D. Slightly knock each microphone and observe picks in Audacity.
   E. Pause the recording.

Figure 3. Mixer setup

Figure 4. Audacity setup
IV. Setting up the microphone positions and camera
   A. Attach the microphone with the green label as the microphone used with the Brain Quick station† (to the camera holder). Make sure this microphone is pointing to the center of the room.
   B. Place another microphone close to infant’s ear. If possible, attach to the cap rubber. Make sure it doesn’t touch baby or cloth.
   C. If possible, camera should be pointed to the baby’s face.

V. Starting the recording session (the order of these steps must be followed)
   A. Press the pause button in Audacity (see Figure 4) to re-start the recording.
   B. Start the recording with BrainQuick. Make sure the video recording is switched on.
   C. Reproduce the tonal sound. HOW

VI. During the recording session
   A. Make notes about:
      1. Recording date and start time.
      2. Incubator position number (see Figure 5).
      3. The name of the preterm baby.
      4. The environment: quiet or noisy. If noisy, note down activities happening (e.g. pediatric observation, surgical intervention, parents visit, etc.)

---

† BRAINQUICK is the videopolysomnography equipment used in the recordings
VII. **After the recording session** *(the order of these steps must be followed)*

A. Finish the recording with BrainQuick.

B. Finish the recording with Audacity:
   1. Press the **stop button** (see Figure 4).
   2. Export the audio file: Archivo → Exportar → Guardar.

   ![Export Audio File Dialog](image)

C. Copy the Brain Quick files: .AVI, .TRC and .WAV to a shared location.

D. Set the mixer **level of line 1** to 0 (see Figure 3) to be ready for other types of recordings with Brain Quick.
2. Labelling protocol

This section outlines the protocol for producing annotations for the two relevant types of sounds from the NICU environment: equipment alarms and vocalizations. The format of produced labelling, the particular sound classes to be annotated as well as the labelling guidelines and the tools to be used are described.

Basic rules

1. Only the defined labels must be used for annotating the relevant acoustic events.
2. For each recording, a single annotation is produced which corresponds to the audio acquired using the microphones placed both inside and outside the incubator (44.1 kHz sampling frequency). The annotation should be primarily based on the audio acquired using the microphone placed inside the incubator. (In previous databases, the one outside was used due to the presence of a strong ventilation noise.) The audio obtained with the other microphone should be always consulted in case of ambiguity or doubt as well as used for verification of the timestamps.
3. The spectrogram must be checked during the annotation process (using Adobe Audition\(^\dagger\)).
4. The annotations must be saved to the <name>.csv file, where <name> is the corresponding audio file name.
5. A simple text editor can be used for producing annotations. The timestamps from the observations in Adobe Audition (Selection) should be copied to .csv file according to the annotations format (see further).

6. All the labels (first timestamps) must be put in chronological order.
7. Annotation for both classes should be produced consecutively, e.g. first vocalizations and then alarms.
8. For blurred or very low sounds, the decision of whether to label or not should be based on whether the sound is heard on the audio recording acquired inside the incubator.
9. It is preferable that the difference in timestamps from several annotators does not exceed 10ms.

\(^\dagger\) Version 3 is available for free, e.g.: [http://www.techspot.com/downloads/5733-adobe-audition.html](http://www.techspot.com/downloads/5733-adobe-audition.html) (see publisher’s description for a valid serial code).
2.1 Equipment alarm sounds

Figure 1. Graphical description of terms used to denote particular alarm properties. Only the fundamental (the lowest) frequency is depicted for clarity of presentation. A sequence of two periods is depicted. An extra period (or a sequence of them) may appear after a sequence break.

The general annotation format for alarm sounds is the following:

\[
<\text{start\_time}>,<\text{end\_time}>,<\text{periods}>,<\text{label}>
\]

where \(<\text{start\_time}>\) and \(<\text{end\_time}>\) are the alarm sequence timestamps (see Figure 1), which are the start timestamps of the first and the last periods of the sequence, correspondingly, \(<\text{periods}>\) is the number of alarm periods in the sequence and \(<\text{label}>\) is the corresponding alarm class. Note, that there are no spaces in the annotations format.

For example, for the first alarm sequence given in Figure 1 the annotation will be: \(T0,T1,2,a16\).

Particular rules

1. The number of periods in a sequence should be counted twice to avoid errors.
2. In case of long sequences (usually >20 periods) or overlaps with other alarms, the periods should be counted (and re-counted) in chunks. The chunk is defined as a temporal interval where an alarm sequence up to 20 periods is overlapped or not with another one. For example, the calculations could look like \(6(overlap)+2+4(overlap)+17+20(\text{till 25:08})+5+2(overlap)\)..., where 6, 2, 4, 17, 20, 5, 2 are the number of periods in each chunk. Note that it is better to split the chuck that is longer than 20 periods and write down the corresponding time.
3. If an alarm signal was stopped abruptly (e.g. by nurse or at the end of audio file), the corresponding period is labelled separately and the “c” is added to the class labels, e.g. \(T0c,T1c,1,a16c\).
4. For facilitation, the decision about the alarm class could be based on the following scheme:
a. There is only one tone in the signal interval: a1, a6, a7, a8, a9, a10, a16
b. Several tones in the signal interval are same: a4, a5, a11, a12, a14
c. Several tones in the signal interval are different: a2, a3, a13, a15

To narrow search in each group, the major frequencies of alarm (see Table 1) should be checked. In each case, the samples only from one of the 3 groups require listening.

5. If an alarm sample corresponds to a new class that was not considered before, an exemplary sample of that class should be extracted and a new alarm class should be documented as in Table 1.

6. In case of strong doubts about the specific class of an alarm, a more generic label al should be used.

7. The start timestamp of the alarm period should be based either on the strongest visible frequency of alarm or the frequency that is also strong and clearly visible but starts earlier.

Annotation time estimate:

30s of audio → 10-12m, that is about 20-24 hours of annotation per 1 hour of audio

Table 1. List of relevant alarm sound classes and their description

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
<th>Major frequencies (kHz)</th>
<th>Signal duration (s)</th>
<th>Silence duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>1 tone</td>
<td>1) 0.495, 1.465, 2.435</td>
<td>0.698</td>
<td>1) 1.352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) 0.515, 2.455, 3.445, 4.415</td>
<td></td>
<td>2) 1.548</td>
</tr>
<tr>
<td>a2</td>
<td>4 higher tones and 1 lower. There is a longer pause between each two alarm periods.</td>
<td>1 / 0.830</td>
<td>1.232</td>
<td>0.803</td>
</tr>
<tr>
<td>a3</td>
<td>1 higher and 1 lower tone</td>
<td>1) 0.665, 1.330, 1.990, 2.660 / 0.540, 1.600, 3.150</td>
<td>0.634</td>
<td>14.666</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) 0.520 / 0.420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a4</td>
<td>3 shorter and 1 longer tone, short pause, 1 longer tone</td>
<td>2.350, 4.700</td>
<td>2.675</td>
<td>0.785</td>
</tr>
<tr>
<td>a5</td>
<td>3 tones</td>
<td>0.530, 1.060, 1.590, 2.120</td>
<td>0.970</td>
<td>3.280</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Frequency 1</td>
<td>Frequency 2</td>
<td>Frequency 3</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>a6</td>
<td>1 tone</td>
<td>2.410</td>
<td>0.374</td>
<td>0.073</td>
</tr>
<tr>
<td>a7</td>
<td>1 tone</td>
<td>1) 0.980, 2.935</td>
<td>0.836</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) 2.880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a8</td>
<td>1 tone</td>
<td>0.490, 1.480, 2.460, 3.440, 4.420</td>
<td>0.280</td>
<td>1.965</td>
</tr>
<tr>
<td>a9</td>
<td>1 tone</td>
<td>1) 5.320</td>
<td>0.525</td>
<td>0.545</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) 5.190</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) 6.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a10</td>
<td>1 tone</td>
<td>1) 1.140, 2.280, 3.425</td>
<td>0.675</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) 0.880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a11</td>
<td>3 tones</td>
<td>0.880, 1.740</td>
<td>0.376</td>
<td>2.574</td>
</tr>
<tr>
<td>a12</td>
<td>3 tones, short pause, 2 tones</td>
<td>2.305, 4.610, 6.915</td>
<td>1.820</td>
<td>0.680</td>
</tr>
<tr>
<td>a13</td>
<td>2 lower tones and 1 higher</td>
<td>0.475, 1.335, 3.100, 3.985, 5.750 / 0.540, 1.590, 2.650, 3.680, 5.770</td>
<td>1.105</td>
<td>10.905</td>
</tr>
<tr>
<td>a14</td>
<td>3 tones, short pause, 2 tones. There is a longer pause between each two alarm periods.</td>
<td>3.075, 6.115, 9.195</td>
<td>1.730</td>
<td>0.370</td>
</tr>
<tr>
<td>a15</td>
<td>2 higher tones and 1 lower</td>
<td>1.270, 3.810, 6.330, 8.870 / 1.015, 3.015, 5.015, 7.020</td>
<td>0.560</td>
<td>0.300</td>
</tr>
<tr>
<td>a16</td>
<td>1 tone</td>
<td>0.495</td>
<td>0.307</td>
<td>1.746</td>
</tr>
<tr>
<td>a17</td>
<td>3 tones</td>
<td>0.71, 4.9</td>
<td>1.055</td>
<td>7.985</td>
</tr>
<tr>
<td>a18</td>
<td>4 tones</td>
<td>0.6</td>
<td>0.385</td>
<td></td>
</tr>
<tr>
<td>a19</td>
<td>3 tones, short pause, 2 tones.</td>
<td>0.815, 3.985</td>
<td>1.550</td>
<td>0.850, 6</td>
</tr>
</tbody>
</table>
There is a longer pause between each two alarm periods.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a20</td>
<td>3 tones</td>
<td>3.860, 4.975</td>
<td>0.850</td>
<td>16.525</td>
</tr>
<tr>
<td>a21</td>
<td>1 tone</td>
<td>0.99</td>
<td>0.250</td>
<td>1.8</td>
</tr>
<tr>
<td>a22</td>
<td>3 tones</td>
<td>2.780, 3.880, 4.980, 6.060</td>
<td>0.817</td>
<td>-</td>
</tr>
</tbody>
</table>

2.2 Vocalization sounds

The general annotation format for vocalization sounds is the following:

\[<\text{start\_time}>,<\text{end\_time}>,<\text{label}>\]

where \(<\text{start\_time}>\) and \(<\text{end\_time}>\) are the vocalization sound timestamps, and \(<\text{label}>\) is the corresponding vocalization class. For example, \(T0,T1,fs\) or \(T0,T1,bcs\).

<table>
<thead>
<tr>
<th>Label</th>
<th>Sound class</th>
<th>Description</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>fs</td>
<td>Foreground voices</td>
<td>Any kind of speech whose source is perceived as being close to the microphone. A distinct voice over background voices or babble having a high intensity. Short loud speech produced either close or far from the microphone.</td>
<td>The short pauses of less than (1)1.5 s should be labelled as a part of the vocalization</td>
</tr>
<tr>
<td>bci</td>
<td>Baby crying (from recorded preterm)</td>
<td>Cry or other vocalization sounds coming from preterm infants.</td>
<td>Should be guided by the spectrogram of the audio acquired with the inside microphone and by the video recording*.</td>
</tr>
<tr>
<td>bco</td>
<td>Baby crying (from other preterms)</td>
<td></td>
<td>Should be guided by the spectrogram of the audio acquired with the outside microphone.</td>
</tr>
</tbody>
</table>
Particular rules

1. Principle: go fast and do as simple as possible.
2. The decision for labeling or not can be based on listening or/and how strongly vocalizations are seen on the spectrogram (see examples below - vocalizations in green were labelled, in blue - not).
3. In case of doubts about specific class of baby crying, a more generic label $bc$ should be used.
4. It is preferable that the timestamps are extended instead of being too narrow and possibly causing the acoustic event boundaries to be cut (“better to label more than less”).
5. To decide on the end timestamp, make an estimation based on the spectrogram and listen if something is heard after the timestamp, refine. Similarly for the start timestamp.

*For the old databases where the video information is not available, a $bc$ label should be used when the baby crying is only heard inside the incubator.
3. Database structure

This section describes the database directory structure and file organization, and discusses some preparation steps (i.e. synchronization) that have to be performed before the inclusion of data to the database.

The database directory structure is the following (the folders only for internal use are marked in blue):

```
/readme.txt
/notes.pdf       // detailed description of the recording sessions
/<session_name>
/audio_44100    // audio recordings, 44100 Hz sampling rate
/in             // recordings made with the microphone close to infant’s ear
    <session_name>_01_in.wav
    <session_name>_02_in.wav
    ...
    A readme file may be needed for each directory that includes relevant information of how to use the files, though it is already in the pdf notes file. A possible example? if there are two files in the session, which is the beginning time stamp of each one
/out            // recordings made with the outside microphone
    <session_name>_01_out.wav
    <session_name>_02_out.wav
    ...
/stereo
    <session_name>_01.wav
    <session_name>_02.wav
    ...
/raw            // original (unsynchronized) audio recordings
    <session_name>_01.wav
    <session_name>_02.wav
    ...
/audio_16000    // audio recordings, 16000 Hz sampling rate
/in
    ...
/out
    ...
/stereo
    ...
/labelling_al   // annotation files for equipment alarms
    <session_name>_01.csv
    <session_name>_02.csv
```
4. Data preparation

1. Save at least two copies of all the data from the recording session. One should go to the general folder that contains all the recording session carried out so far.

2. Organize the folders and files according to the directory structure outlined above, e.g. rename the .wav, .avi, .trc files using session codes and move to corresponding folders. Note that if one audio recording corresponds to several video recordings, include the ordinal number of the video file to it’s filename (e.g. RS26_01-1.AVI and RS26_01-2.AVI), and vice versa.

3. Extract the audio from the video recording and place in the session /video folder with the corresponding name (e.g. RS25_01_vid.wav):
   ```bash
   avconv -i filename.avi -vn filename.wav
   ```

4. Synchronize the audio recording by running the Synchronize.m. MATLAB code with proper input and output filename variables.

5. Separate the synchronized audio to inside and outside (in Adobe Audition put cursor on the left or right channel, CTRL + A, Save selection...). The baby crying will usually be stronger on the recording inside the incubator. The video recording could be used for checking.
Annex 1. Checklist for recordings

Audiobox setup controls

• Both microphones amplification - 5 grades from max
• Main (output) - max
• Mixer (inputs & playback) - to “inputs”
• 48V switched on! (electric supply for microphones)
• Check with the phones that both microphones are working (touch each microphone).
• Observe with Audacity that the recording levels are ok. - levels can seem rather low if the recording cursor is moving fast

Output connections

• Check that left Audiobox output is connected to left mixer input, and correspondingly for the right output.

Microphone and camera positions

• 1st mic taped to the camera holder pointing to the center of the unit room
• 2nd mic is close to infant’s ear, but avoiding it is touched (e.g. if possible attach to the rubber of the EEG cap; or if an open bed, hand on the upper part).
• Camera should be pointed to the baby’s face.

Mixer setup

• Level of the 2/3 line - middle
• Level of the other lines - to min
• Main output mix - middle
• Check the recording level lights at the mixer

PC recording with Audacity

• Choose Audiobox microphones for recording
• Make sure there is storage enough (bottom line of Audacity screen)
• At the end, export to .wav file with 32 bits

BQ recording

• Make sure video recording is switched in BQ.

At the beginning of the session:

1. Start audio recording with audiobox and Audacity.
2. Start recording with BQ.
3. Play the rectangular tone.
4. Produce Claquette sound, making sure it is visible on video (may not be necessary)

At the end of the session:

- Copy the BRAINQUICK video and MED files.

Extraction of audio from video (Ubuntu, command line)

```
avconv -i filename.avi - get information about file’s audio and video streams
avconv -i filename.avi -vn filename.wav - extract audio from video file
```