A Gold Standard Heart Sound Annotation Tool

Alyssa Ong
University of Auckland
22 Princes St
alyssaong@gmail.com

ABSTRACT
Cardiac auscultation, which is the ability to conduct diagnosis by listening to heart sounds, is an inexpensive but effective method for detection of heart diseases. To ensure that the practice of cardiac auscultation remains widespread in the medical field, there needs to be a well-annotated public database of heart sounds for comparing automated heart diagnosis algorithms and for medical training purposes. However, this has been a challenge to establish because there are very few existing heart sound annotation tools that users are willing to adopt. Through reviewing current literature, this paper presents guidelines crucial to creating a gold standard heart sound annotation tool, and discusses methods in which this can be achieved.

INTRODUCTION
Cardiovascular Diseases (CVDs) are the top cause for deaths around the world every year. In 2012, 31% of all global deaths were due to CVDs. For those with CVDs, early detection and control is crucial to avoid its progression, with the first point of screening often being at primary care. [1] Despite this, the mastery of cardiac auscultation, which is the ability to conduct diagnosis by listening to heart sounds, is rapidly being lost amongst health professionals due to lack of continuous training. [2, 3] At present, only a doctor with years of experience with listening to heart sounds will be able to conduct an accurate diagnosis. For this reason, there has been a rise of necessity in automated auscultations, made possible by performing machine learning on large datasets of heart sounds. In addition, learning resources for those undergoing auscultation training in the medical field are currently lacking.

This naturally leads to the need for a large, publicly accessible gold standard database of heart sounds, so that the performance of diagnostics algorithms around the world can be properly compared to each other, and also for auscultation practice. [4, 5, 8] Contextual information about each heart sound recording plays an extremely important part in the auscultation. This could include information like patient weight and precise location on the body that the heart sound was taken. Hence, this database of heart sounds must be annotated with useful information as a reference for machine learning and medical training. Annotations are thus mostly done by either experienced clinicians or less experienced users, such as those under medical training. While many generic audio processing and annotation tools exist, there are very few audio annotation tools out there specially for heart sound annotation targeted at those in the medical field. Thus, a gold standard heart sound annotation tool is necessary to guide annotators in collecting useful contextual data needed for accurate auscultations.

This paper mainly discusses the characteristics of effective heart sound annotation for the purpose of this public heart sounds database. This includes the type of information collected during annotation, how to make accurate annotations, methods of annotation and how to create a user interface for heart sound specific annotation software with good usability.

The questions being answered are:

1. What key information needs to be annotated to the heart sounds for gold-standard documentation?

2. What kind of annotation tools and methods should be used in the heart sounds annotation software?

3. How can target users be encouraged to adopt the heart sounds annotation software?

To answer these questions, we will analyse some audio annotation methods and tools being proposed or used already. Heart sound annotation tools would be ideal and preferable to analyse, but as not many of these actually exist out there, we also expanded our analysis to lung sound annotation tools (as similar methods are employed in respiratory auscultations and cardiac auscultations), as well as ordinary audio annotation tools. By looking at current research, we can gather some good practices for the collection and annotation of heart sounds in building a well-documented, expert annotated public database.

PROBLEMS AND MOTIVATIONS
Drawing from existing studies in heart sound annotation, we’ve identified some key problem areas faced by researchers.

Performance comparison of an algorithm against another should be run against the same database of heart sounds as a common reference. [4, 5] This “common reference” would consist of agreements between as many qualified health professionals’ annotations as possible regarding key features
in the lung sounds. Dinis et al. state that the effectiveness of algorithms can be measured by comparing the algorithm’s results against these expert annotations using the “confusion” matrix. [6] Another reason for “crowdsourcing” annotations from experts is because individual auscultations can be subjective depending on the individual physician’s interpretation. [3] All this demonstrates the need for an effective annotation system for heart sounds, but currently there are little appropriate “gold standard” heart sound annotation tools and methods for health professionals out there.

A similar theme resonates with Pinho et al.’s study. They note that current tools out there are hardly appropriate for collecting lung sounds, as they have quite a complicated interface, poor usability and are not intuitive. Also, the way important information was presented was poor (bad visual and text summaries) and did not collect necessary data to do proper assessments on patients, making clinicians disinclined to adopt them in their daily practices. [7] While these are mentioned in context of respiratory sounds, the auscultation process between heart and respiratory sounds are quite similar. These tools don’t cater specifically for heart sounds in terms of data gathered that is needed for diagnosis and signal processing features.

We also considered the main target audience who will be directly using the heart sound annotation tool, and the problems they may face. While auscultations may only be performed by cardiologists and physicians, it is highly possible that annotations can be done by less specialised individuals like nurses as well. Naturally, common places where a high concentration of heart auscultations are performed include day-to-day health clinics (primary care), hospitals and emergency rooms, and current tools aren’t so suitable in some of these busy situations. [8] The annotation process and method is too slow and intrusive with their routine work. Hence, we also took a look at some existing annotation methods in similar environments, and how the factors in these environments may shape the annotation methods and tools based on the needs of our target users.

**APPROACH**

In analysing the literature, we pinpointed 4 essential aspects to cover when approaching the design of a heart sound annotation tool:

1. Type of annotations made when recording heart sounds
2. Use cases for the annotation tool
3. Data visualization and annotation input tools
4. Layout and structure of GUI

We will now discuss each of these aspects in greater depth.

**TYPE OF ANNOTATIONS MADE**

The annotations to be collected by the tool should be based on all the data required to perform an accurate and thorough auscultation on patients. [7] However, when constrained on time in busy environments like primary care, this may not be feasible as it would take a while. In a study done on collecting and annotating heart sounds in the primary care of a hospital, named DigiScope, Pereira et al. suggest that data which is critical in distinguishing a normal heart sound from an abnormal one should be taken at the very least. In describing their data model, they stressed on the type of data collected that might be helpful for machine learning and cardiac diagnosis, particularly attributes from the second heart sound. (Figure 1)

![Figure 1: DigiScope data model](Source: Pereira et al., Unobtrusive collection and annotating of auscultations in real hospital environments. *Annual International Conference of the IEEE*, (2011), 1193-1196.)

The auscultation exam is mainly of interest to us, and these are the types of data that could reveal new knowledge like multiple diseases for a single patient, differences between normal and abnormal heart conditions and more. Some of these attributes are useful products of signal processing and not direct annotation. [8] In another study done by Ramos et al. on the representation of cardiac signals, although they talked about the segmentation, extraction and importance of many types of sounds, they acknowledged that a simpler visualisation for the heart sound components was necessary to reduce cognitive load for the physician when viewing the heart sound signals. (Figure 2) [3]
It should be noted that S1, S2, A2 and P2 are the main sounds taken for this simpler visualisation, which draws similarities from the data taken in DigiScope and affirms that second heart sounds should be annotated at the very least.

Pereira et al. further studied the annotated data to discover classifiers in predicting CVDs, and it was also found that murmur is an extremely important attribute to record in predicting pathologies. [9] Many of the heart and lung sound auscultation tools also record the exact position that the sounds were taken from, with studies agreeing on the 4 main positions (M: mitral, P: pulmonary, A: aortic, T: tricuspid) [3, 4]. Other commonly recorded values include the patient’s posture (sitting or lying down). A table is shown with recorded values across the different annotation tools. (Table 1)

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>Patient posture</th>
<th>Physical activity</th>
<th>S1 &amp; S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DigiScope</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Exertion)</td>
<td></td>
</tr>
<tr>
<td>IntelliSensor</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Extraction from HSCT-11, an open heart sounds database [5]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 1: Examination data in heart sound annotation tools

Some of the tools investigated also have automated annotations. This is where useful products of signal processing from the heart sound are computed and automatically presented to physicians. A usability study done by Pinho et al. on their interface to collect and store respiratory data, called RIBS@UA, found that automatic computing of some clinical parameters was very helpful. [7] This, in combination with data interpretation by the interface (labelling heart rate as normal, below average, etc.) could do wonders in saving time and mitigating interpretation errors by clinicians. The automated signal processing and annotation present in IntelliSensor and DigiScope that extracted S2, A2 and P2 separately reflect real potential merits in computation. (Figure 3) This will be further discussed in future work.

USE CASES FOR THE ANNOTATION TOOL

Use cases for the tool need to be reflected on so that environmental factors and target users are considered in determining requirements for annotation input methods and user flow in the GUI during the annotation process. It is argued that one of the greatest factors to creating a successful system for collecting heart sounds is not just focusing on sound quality, but also communicating with the clinical community to address the needs of cardiologists. [9] Many providers of annotation tools don’t do this.

Several existing studies try to account for this issue and tackle it. DigiScope points out that clinicians have hectic schedules and the collection of sounds must not interrupt their routine work. To fulfil this requirement, Pereira et al. have made a clear separation between the sound collection and data annotation stages. Clinicians can make audio recordings one after the other, then return to annotating later when they have time. This is so that the collection system integrates more easily into their routine work, and the separation also helps to avoid annotation errors. [8] Dinis et al. mention similarly that the annotation tool is going to mostly be used by health professionals with tight schedules, and as a result have made user-friendliness the top requirement of their annotation tool, where annotation must be quick and easy to learn. [6] Gomes et al.’s Vital Analysis framework collect and annotate biological data in first responder situations. They found that getting their users to annotate during stressful and dangerous events was very hard because users did not have time and were not in the right mindset to do it at all. Hence, like Dinis et al., they
highlighted the need for usability and usefulness in Vital Analysis. [10] Their solution for achieving this involves input methods, which will be discussed in the next section. On top of clinicians, heart sounds could also be annotated by those undergoing medical training. Pinho et al. point this out and present a notably interesting way of tackling it. In their “respiratory sound toolkit” window, there are three main options for annotations, being:

1. Automatic annotation (from the automatic analysis of sounds)
2. Manual annotation by experts
3. Manual annotation testing by inexperienced users or those who are still undergoing medical training

By separating annotations in this way, machine learning algorithms can ignore or put less weight on annotations by less experienced users, while still retaining all annotation data. Less experienced users can also compare themselves with expert annotations as a benchmark for training purposes.

DATA VISUALIZATION AND ANNOTATION INPUT TOOLS

After defining the annotation requirements, the next step is to evaluate how to present the data in a visual, intuitive way and also to define the input tools for accurate annotation. The visualisation of waveforms appears to be a relatively common theme across all tools, with some also allowing the addition of annotation layers on top of these waveforms. However, not all of the tools provide extra segmentation of the audio files. In SonicVisualizer, a general purpose audio editing and annotation software, separate key shortcuts can be used to call on different types of spectrograms. [11] A similar concept is in IntelliSensor, the phonocardiogram (PCG) analysis tool by Ramos et al.. In IntelliSensor, different waveforms may be visualized on top of one another, but can be shown or hidden based on “toggle buttons”. (Figure 4) These methods could be used to simplify visualisation and reduce information overload for users, achieving better usability. SonicVisualizer also employs a “pane and layer paradigm”, where different spectrograms are stacked on top of one another. (Figure 5) This layout appears to be adopted in the Respiratory Sound Annotation Software (RSAS) by Dinis et al., as well as IntelliSense, meaning it is likely to be the best way to present segments of waveforms.

SonicVisualizer allows editable annotation layers on top of these waveforms, and these annotation input types include labels (essentially text), images, time-value plots and time instants (like beat locations). These annotation types quite generic and are not specific to heart sounds, unlike with RSAS where specific color-coded lung sound annotations can be made on top of the waveforms, being wheezes, crackles, expirations and more. In fact, the usability study done on RSAS found that crackle annotations were more dependent on the graphic representation, whereas wheeze annotations were more dependent on hearing the sound. These practices and findings can be useful when applied to a heart sound annotation tool. Properly defining the type of annotations that are able to be made in context of heart sounds can give users a good guide on how to annotate a heart sound, for instance having special annotation layers for S1 and S2, with colour coding it much easier to visualize and understand. More research would need to be done to see if certain components of heart sounds are more visual and others more auditory. In IntelliSensor, annotations are automated so they are non-editable, but it is noted that they
are also layered on top of the waveform as well for easy visualisation purposes.

Other input methods include voice annotations. In Vital Analysis, voice annotations are used to avoid letting annotation heavily interfere while first responders are out on the job, and at the same time retrieving valuable contextual information with the biological signals. [10] This could be a useful option to explore for heart sound annotation input methods offered, for use cases like in the ER as mentioned by Pereira et al.. Their results from over 3300 field hours showed that voice annotations were being used at events when the firefighters were feeling the most stressed, which indicates that voice annotations can be effective for gathering annotations in more intense situations.

Another kind of annotation that was sometimes visually represented (but still under a set of options) was the position on the patient’s body where the sound measurement was taken, which can be seen in RIBS@UA (Figure 6) and IntelliSensor. Patient posture was also visually represented in IntelliSensor. In these specialised annotation tools, attributes are represented visually where possible (and appropriate) for quick and easy understanding.

![Image](Figure 6: Measurement position visualization with RIBS@UA)

(Source: Pinho et al., RIBS@UA: Interface to collect and store respiratory data, a preliminary study. Computers in biology and medicine 47, (2014), 44-57.)

Different tools used different methods for making precise annotations at certain time instants. With SonicVisualizer, one would add a time instant layer on top of the waveform, followed by either clicking or tapping a key to the beat. However, this is more appropriate for music purposes where a rhythm can be kept, rather than for the purposes of heart sounds. RSAS on the other hand uses zoom, pan, speed and playback tools to make precise sound selections easy for annotations to be made. With regards to the change in speed, they used a phase vocoder to extend the audio file without affecting the pitch significantly. Results from their usability study show that when annotating a sound, users generally listened to the whole file at regular speed at least once, make an initial selection with the zoom tool then slowly go through the rest of the sound with the pan tool. Based on this common behaviour, we can gather that zoom and pan are especially important in the process of annotation and should be present in annotation software. It was also found that selection and playback tools (like playing, pausing and stopping from any position) were vital to making accurate annotations, and these tools should be present as well. A table is shown with the different annotation software and their features. (Table 2)

Lastly, RSAS uses the interesting concept of a “guide plot” under the main plot to show the locations of the annotations being made in relation to the entire audio file, even if the main plot is zoomed into one part of the sound file. This way, the user never loses where they are relative to the whole sound file and can easily visualise annotated areas as a whole and possibly patterns in revealing pathologies.

![Image](Table 2: Annotation features in heart sound annotation tools)

**FEATURES AND LAYOUT OF GUI**

By comparing SonicVisualiser’s GUI to the other tools specially for heart and lung sounds, we can see that using an annotation tool not catered to heart sound annotation can be problematic due to the complex presentation of data and lack of intuitiveness. In contrast, users were quick to adapt to the RSAS interface according to their usability tests. User feedback reflected effectiveness of use for Vital Analysis, RIBS@UA and DigiScope GUIs.

There exists a pattern where the visualisation of waveforms appears underneath or to the right of patient information and other annotations, apparent in RSAS (Figure 7), RIBS@UA (Figure 8) and IntelliSensor. This may be the natural order that clinicians do analysis, by assessing patient data and determining position of auscultation before taking the sound. Set options are given where possible instead of text fields during annotation to avoid error and reduce the amount of recalling needed to be done.
Figure 7: RSAS annotation window
(Source: Dinis et al., Respiratory Sound Annotation Software. HEALTHINF, (2012), 183-188.)

Figure 8: RIBS@UA waveform visualisation window
(Source: Pinho et al., RIBS@ UA: Interface to collect and store respiratory data, a preliminary study. Computers in biology and medicine 47, (2014), 44-57.)

Patient details not directly related to events during the auscultation (this excludes acquisition location and patient position) are usually in their own separate pane, the exception to this being DigiScope which has patient position in the patient details pane. This provides effective data grouping and is a good pattern to adopt.

Across all the GUIs, there does not seem to be any convention for the relative position of navigation tools (like zoom, pan) on the overall window. Some of the methods for panning also differ - RSAS uses click to drag and Intellisensor uses a horizontal scrollbar. More usability studies will have to be done to assess the effectiveness of these different methods.

DigiScope’s software was on a tablet-type PC, and tabs could be an method to organise annotations in a constrained space for annotation tools on portable devices. (Figure 9) In fact, one of the disadvantages mentioned in the interviews done in RIBS@UA was the lack of a mobile version. DigiScope’s early results showed that physicians could use the mobile interface effectively and learn it quickly.

Figure 9: DigiScope examination information window
(Source: Pereira et al., Unobtrusive collection and annotating of auscultations in real hospital environments. Annual International Conference of the IEEE, (2011), 1193-1196.)

One aspect that could be taken from SonicVisualiser, is the hotkey reference shown in the help menu. Having available hotkeys for heart sound annotation tools would cater for experienced users to speed up annotation even more in the long run. Lack of shortcuts was also pointed out in the RIBS@UA heuristic evaluation, strengthening this point.

SUMMARY
Attention is growing in developing usable auscultation annotation tools to collect a central repository of well-annotated biological sounds for machine learning and medical training purposes. Many studies agree that a large gold standard database of well-documented heart sounds is needed for this, but there are some important benchmarks that annotation schemes need to meet before being adopted by target users. From our analysis of current research done on heart sound annotation tools, we find that these tools must:

- Collect and present useful data in context of heart sound annotation
- Have high usability
- Be easily understood at first sight
- Adapt to different use cases and target users
- Have tools that make annotation quick and easy but effective and accurate

We discuss many methods in which these main outcomes for a heart sound annotation application can be met. Future work still needs to be done towards satisfying these goals to achieve a gold standard for annotation needed for a rich database of heart sounds that professionals and the public can turn to.

FUTURE WORK
Automated annotation of data targeted at clinicians has not been explored much in depth. One way in getting clinicians to adopt the annotation tool is if it presents useful data to them, and further analysis into computing heart sound information on the fly that’s useful for clinicians in performing their auscultations will encourage them to adopt the tool.

More work could be done on optimising annotation speed while preserving accuracy. The behaviour for the more common annotation actions could be investigated in detail, and the layout of the GUI, annotation facilitation tools and even hotkeys could be modified based on this.
Lastly, based on the findings from graphical and auditory representations in lung sounds, we can try seeing whether it is more useful to have graphical or audio representations for certain types of heart sounds. This can help us determine which waveform annotations are most helpful to display on the GUI, and which may not be as useful to display.

REFERENCES