Doctoral Research Prospectus

Hans Fugal

June 25, 2008

Abstract

I discuss the task of registration identification, the approach I plan to take, the assumptions and principles involved, and the planned details of methodology, software, etc. The appendix is a primer on the pipe organ, its concepts, construction, terminology, etc.

1 Task

The task at hand is to identify the registration (set of pulled stops) of a pipe organ recording. The recording will be on a single division (so every note played has the same registration) but polyphonic (multiple notes played simultaneously, i.e. "real-world music").

The motivation for this task is first because it is related to instrument identification but with some inherent simplifications and complications that make it an interesting problem. It is simpler than normal instrument identification because of the nature of the organ: stops sound simulaneously and with the same volume and the same timbre every time. It is more complex than normal instrument identification to date because it is tackling fully polyphonic music head-on.

Second, the identification of stops has pedagogical value for organists. They learn the art of registration by listening to samples and then analyzing the chosen stops.

Third, there is potential practical value for the organist wishing to choose a registration on his organ that gives a similar tone to that chosen by the artist on a professional recording. This algorithm, when trained on the smaller organ but applied to the professional recording, may give a reasonable approximation of the professional's registration using the stops actually available.

2 Hypothesis

I will describe the principles and assumptions that I believe will alow this algorithm to work well.

We distinguish organ stops primarily on the basis of attack and timbre, and to a lesser degree using amplitude and position cues.

The attack phase consists of the initial evolution of the harmonics and the transient inharmonics, usually referred to by organists as "chiff." Organists have a word for chiff because of its contrast with the constant nature of the sustained sound.

The timbre of the sustained sound is to a great extent due to the harmonics and partials of the fundamental that are present. I have made a recording of different stops in Aeolus, a pipe organ simulator [And]. Figure 1 is the log-frequency spectrogram of this recording. The harmonic partials are easy to detect and the harmonic spectra are easily distinguishable.

It looks like an organ stop can be identified by looking at its steady-state spectrum. In particular, the

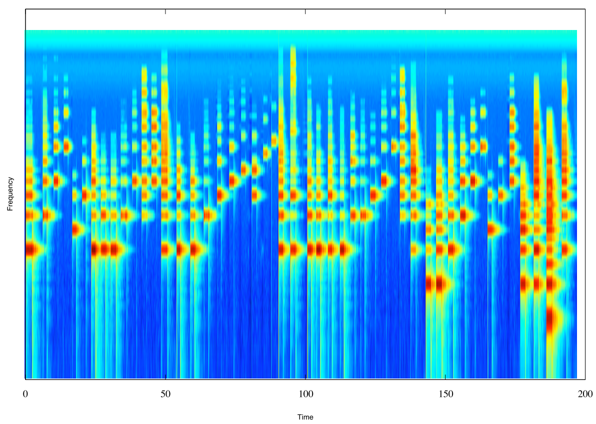

Figure 1: Log-frequency spectrogram of middle C played on each stop sequentially of the Aeolus organ simulator.

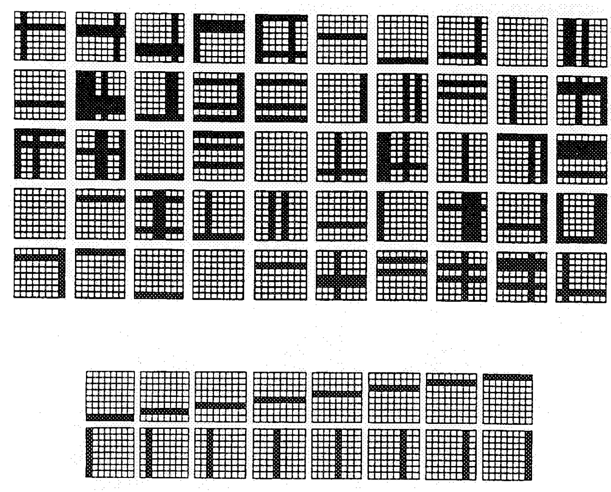

Figure 2: The bars problem, designed by Földiák [Föl90]. The bottom group are the independent bases, and the top group are possible by combining bases. Figure from [Sau95].

harmonics (integer multiples of the fundamental) appear to be a sufficient basis. Based on preliminary examination it will probably not be necessary to examine the attack and transients.

Each pipe in a stop is a distinct instrument, and the actual spectra can vary even within the same stop. While not large, this variation may be sufficient to derail classification, so I will develop a statistical model of the stop (e.g. perhaps using a Gaussian Mixture Model).

The harmonic partials are related to the fundamental frequency of the note, so the fundamental frequency needs to be identified in order to use harmonics as a feature.

Polyphony means multiple fundamental frequencies. The notes and/or their harmonics may overlap the harmonics of other notes, complicating the spectrum. However, since the first overtone (which is the second harmonic—the first harmonic is the fundamental) is an octave above the fundamental, and music in up to 4 voices (as is much organ music) tends to double only one or two tones at most at any time, the patterns of interference are likely to change enough to get a clear picture over time if we accept and work around some uncertainty.

Combining stops is just causing multiple pipes to sound simultaneously when a key is pressed, which is an additive combination of sound. The spectra of the stops combine linearly. So, given a recording of a note played on the organ with an unknown registration, we want to determine which combination of stops accounts for the observed spectrum.

This is similar to the bars problem (see Figure 2). Saund [Sau95], Klinseisen and Plumbley [KP00], Fyfe [Fyf97], and others use artificial neural networks (ANN) as a multiple-cause mixture model for problems like the bars problem, including, in the case of Klingseisen and Plumbley, instrument separation. These approaches are unsupervised, so they find bases without supervision but then have no labels. I will use a traditional supervised ANN variation.

I feel confident that an algorithm can be devised with good performance on monophonic music with multiple stops. Polyphony is a considerable challenge and the one where the bulk of the research will probably lie. I believe reasonable results will be achievable for music with at least two voices. At that point I will explore separate divisions (two

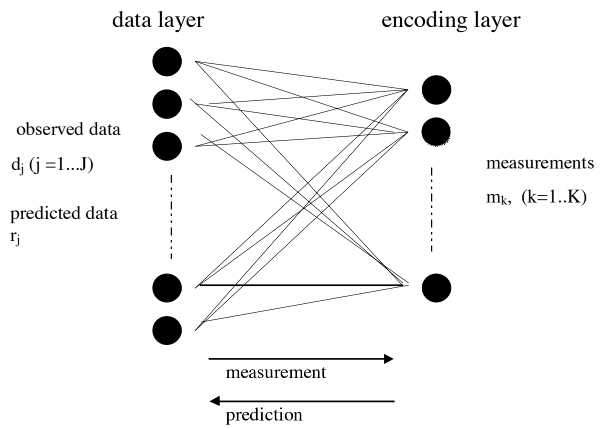

Figure 3: Multiple-cause mixture model architecture, from [KP00]. My network will be similar, except in the way the $c_{i,j}$ are updated in a supervised manner instead of unsupervised.

manuals or manual and pedal) with distinct registrations. I will explore additional complexity (three or more voices, swell, tremolo, etc.) briefly, to see what the algorithm is capable of and where its limitations are.

In addition I wish to explore a more general classifier that identifies not specific ranks of a specific organ, but the class of organ stops present. For example, principal stops, flute stops, stopped flute stops, reed stops, etc. This is dependent on getting a wide enough variety of recorded data.

3 Algorithm

3.1 Training

There are two stages to training the algorithm for a specific organ. First, stops will be examined in isolation and a statistical model of each stop will be constructed. This model will provide a vector of harmonics for a given note. Second, a neural network will be trained with randomly-chosen stop combinations at randomly-chosen notes.

The statistical models of stops will account for variation in spectra through the different pipes, probably representing the harmonic partials with relative strength as Gaussians. The model will produce a vector of harmonic partial strengths, probably with some noise added.

Although the neural network could be trained with recordings of stop combinations, the number of stop combinations is 2^n for n stops, which is prohibitive for live recording even for small organs. The isolated stops could be combined by mixing audio recordings, but the same effect can be achieved by using the statistical models of the stops to generate the harmonics for random stop combinations. This should provide representative training data until the network is fully trained. This method will be compared to training on mixed isolated stop recordings for validation.

3.2 Identification

Once training is complete, the network will take a vector of harmonics and output the set of stops responsible. In order to get the harmonics, the spectrum needs to be analyzed to find the fundamental frequency and pick out its harmonics. In polyphonic music, overlapping notes and their harmonics need to be considered.

In monophonic organ music, the fundamental frequency is easy to find. It is almost always the peak with the most power and always the peak with the lowest frequency. The harmonics are integer multiples of the fundamental, but because of slight error the harmonic must be found by finding the peak in a small window around the harmonic frequency.

In polyphonic music, there won't be one dominating peak in the spectrum because there may be multiple fundamentals. Not all fundamentals may have the same power, either, as some may coincide with the harmonics of others. Still, the lowest large peak will be the fundamental of the lowest note, which gives a starting place. The harmonics will be picked as before, but some may be corrupted by the presence of other notes. This possibility will be ignored and the identification run as normal. Once the harmonics are picked, it will move to the next peak which hasn't already been accounted for as a harmonic. This will be treated as a fundamental, the harmonics picked, and the identification run. This repeats for all major peaks that aren't harmonics of previously-seen notes.

The identified stops at each step are tallied. As more notes are seen the confidence for the true stops should increase faster than that of false stops. In this manner the interference due to overlapping harmonics should come out in the wash.

3.3 Limitations

There are variables in the real world which add complexity.

Registration can change at any time, although it happens relatively infrequently and often not at all. This will be ignored; each segment could be examined independently as needed. Future work could detect a registration change and adjust the registration hypothesis (and make a note of when the change ocurred).

An organ usually has more than one division, each with its own registration, which can be played simultaneously and/or switched between at any moment. Divisions can sometimes be coupled together, essentially making one division a subset of the other. This is the biggest practical limitation, because multi-division organ music is very common indeed. I am developing some ideas for adapting the algorithm to output stops of multiple divisions, which I plan to implement for this dissertation.

Usually some ranks are enclosed in a swell box, which can be opened or closed with the swell pedal to control volume. This changes not only the amplitude but the timbre as well because the swell box acts as a filter. I do not know how the volume change or filtering will affect the algorithm—it may adapt just fine or it may be less effective. I will test this and report on it.

There is often a tremolo which can be applied to some stops, which modulates the fundamental frequency and in the process probably the timbre of the harmonics. Celeste stops are slightly out of tune relative to the other stops, which may give difficulty. I will test these effects and report on how the algorithm handles them.

Some electronic organs can be set to different temperaments. I don't forsee this as a problem because regardless of temperament the pipes have the same harmonic partials. The temperament may slightly increase or decrease the harmonic intereference due to polyphony, but this should not be a problem. As a changed temperament is in some sense a completely different organ, and as I am recording primarily real pipe ranks, I will ignore temperament changes. I will use the same temperament at all times on any given organ, but I may use well temperament on one and equal temperament on another if the opportunity arises. Likewise, there is a small chance that one of the all-pipe organs I am able to record is well tempered, but this is unlikely.

4 Methodology

4.1 Data

As is always the case with machine learning, it is paramount to have sufficient data of good quality. Preliminary data has been gathered using Aeolus. I will record training and testing data on a handful of organs in Las Cruces as well as a few in Utah (where I will be visiting in July) and perhaps some in El Paso.

In discussions with Fons Andriaesen (Aeolus author), I have decided that a few seconds recording of every fourth semitone (C, E♭, F♯, and A in each octave) in each rank should provide sufficient characterization to train on.

The recordings will be made with a mid-grade USB condenser microphone at about CD quality.

As even a small organ may have 5–20 stops, and any given registration may have as few as one or as many as all of the stops pulled, recording every stop combination is impractical, even if one ignores unlikely stop combinations. Instead, I will record some representative short pieces in several sub-genres (baroque, romantic, hymnal, etc.) with "typical" registration for each piece on each organ. Where several typical registrations present themselves I will record as many as time allows. These pieces will be test cases to validate the network trained on the random combinations of the statistical models of stops.

4.2 Learning and Evaluation

This is a machine learning problem, which means the problem formulation and evaluation are important factors. The task under consideration is to identify the set of pulled stops. The exact task will vary from organ to organ. There will be a predefined set of stop labels for each organ (as each organ has a unique set of pipes). Each stop in this set is either on or off in a registration. Training is supervised, i.e. we know the registration of stops in the training data.

There are actually two machine learning problems in the algorithm, first to learn a representative model of the harmonics of a stop and second to learn the multiple-cause network which identifies the stops given the harmonics.

The performance measure still needs to be finalized, but wouldn't be a simple pass/fail because some stops could be identified but not all, or harmonically related stops mistakenly identified.

The training experience is recordings as described above. A single audio analysis frame will be the unit of experience, and so each recording segment will produce a great number of experience units. However, not every frame will be suitable for training. Data will need to be prefiltered to separate out silence, transients, etc., and labeled with the fundamental(s) and any other preprocessing labels that are found to be necessary.

Data generated from the statistical models will include added noise to better simulate the real world and avoid overfitting, as is common.

In evaluating a machine learning algorithm, it is important to properly partition the data into training and validation sets. The precise method of validation is yet to be determined but will be determined before evaluation and will follow machine learning best practice.

4.3 Software and Hardware

I expect that I will write the bulk of the code implementing the algorithms explored in this dissertation. I will probably use Octave/Matlab at least initially, possibly moving to other languages as convenient or necessary. The primary consideration (after correctness, of course) will be the balance of speed of implementation and execution, available signal processing libraries, and my own experience and preferences. Where possible I will leverage libraries made available to me by others.

I will implement primarily on my laptop in OS X and/or Linux.

4.4 Evaluation

There are several natural stages in developing the algorithm which will be evaluated. They are, in rough chronological order:

1. Single note, single stop;
2. Single note, multiple stops;
3. Monophonic melody, single stop;
4. Monophonic melody, multiple stops;
5. Single chord, single stop;
6. Single chord, multiple stops;
7. Polyphonic music, single stop; and

8. Polyphonic music, multiple stops.

At each multiple stop stage, first two stops, then three, and so forth to a "typical" registration. Several different kinds of registrations will be considered. At every musical stage, several different pieces with their respective registration choices. As mentioned before, the exponential possibilities cannot all be exhausted, so I am trying to decide on a characteristic set of music and stops to record.

4.5 Aeolus

The organ simulator Aeolus is aimed at being an organ that organists will enjoy playing. It doesn't strive for perfect simulation, but if it achieves that goal it can be argued that it probably captures the essence of the organ sound. Personally I have found it very satisfactory indeed, and I in fact prefer to practice with Aeolus to the electronic organ at my church. I intend to continue to make heavy use of this excellent software, but recognizing that it is a simulation and not a real organ. I will compare the algorithm's performance on Aeolus and real organs, and may gain insight into its strengths and weaknesses as a result.

5 Timeline

This is a tentative timeline. My goal is to graduate in spring 2009, which gives a hard deadline of the beginning of April for the dissertation and defense. I have estimated the time it will take to accomplish each stage, but there will no doubt be unforseen difficulties that may push things back. There is some room for this, and also some possibility to get ahead of schedule early on.

1. Recording of data: July 26
2. Single note, single stop: August 2
3. Single note, multiple stops: August 16
4. Monophonic melody, single stop: August 30
5. Monophonic melody, multiple stops: September 6
6. Single chord, single stop: September 20
7. Single chord, multiple stops: September 27
8. Polyphonic music, single stop: October 25
9. Polyphonic music, multiple stops: November 15
10. Complete Dissertation/Defense: April 1

A The Pipe Organ

> "Simply stated, the pipe organ is a big box of whistles." [pip]

A.1 General

When an organist presses a key, pressurized air (wind) is allowed to enter the pipe corresponding to that key in zero or more ranks of pipes. This action is accomplished either in a fully mechanical fashion (tracker action) or electronically (direct electric action or electro-pneumatic action).

Each rank of pipes consists of a series of similar pipes—one for each note on the keyboard—with a tone quality that is similar within the rank and more or less different from other ranks.

The organist may choose which ranks to sound by pulling stops. A stop prevents wind from reaching a rank of pipes, or allows it when pulled. Stop action may also be mechanical or electric, and is controlled by drawknobs or tablets. Each drawknob/tablet (also called a register, or simply a stop) indicates its category, tone color, pitch level, and the number of ranks in the stop. Stops that control multiple ranks are called mixtures or compound stops.

The pitch level of a stop is expressed in terms of the (approximate) length of the longest pipe in the rank (the low C). An $8'$ stop sounds at the same pitch level as a piano, and is called unison pitch. A $4'$ stop sounds an octave above unison, a $2'$ stop two octaves above, a $16'$ stop an octave below, a $32'$ stop two octaves below, etc. Some stops sound not at unison or an octave pitch, but at a harmonic. Such stops are called mutation stops. For example, a $2\frac{2}{3}'$ stop is an octave plus a perfect fifth above unison.

The manual and pedal keyboards each control a division of the organ. Each division is more or less physically separate from the others, with its own ranks of pipes, stops, wind supply, etc. Most organs have two or three manual divisions and a pedal division. Each division has a name, originating centuries ago in Europe. The manual divisions are usually named the Great, the Swell, and the Positive or Choir. The swell is so named because it is usually enclosed in a box with shutters that can be opened or closed by the organist using the swell pedal. In this way the organist can control the overall volume of the swell division.

Often divisions can be coupled together; e.g. by activating the Great to Swell coupler the Great and Swell divisions are both played by the Great manual.

The choice of pulled stops is referred to as registration.

A.2 Stops

Organ pipes fall into two categories: flue (also called labial) and reed (also called lingual). Flue pipes generate sound in the same manner as a flute, by vibrating a column of air. Reed pipes work similar to a reed instrument, by vibrating a reed (also called the tongue), held between a wedge of wood and the shallot, which is amplified by the cylindrical or conical "pipe" (also called a resonator).

A.2.1 Flue

The tone quality of a flue pipe is determined by a number of interrelated factors, the most important of which are the material, scale (diameter to length ratio), shape, whether it is stopped, the mouth design, and the wind pressure.

Flue pipes are usually classified in three general categories: principals (also called diapasons), strings, and flutes.

Principals are open cylindrical metal pipes producing the distinctive organ tone. Principal tone is rich and full with a wide, even distribution of harmonics [Dav85]. Principal stops include: Principal, Diapason, Prestant, Montre, Octave, Twelfth, Super Octave, Fifteenth, Quint, and Mixture.

Strings are also open cylindrical pipes, but their small diameter gives a softer and thinner tone. Strings have a wide even spectrum of harmonics [Dav85]. Some string stops are: Salicional, Voix Céleste, Gamba, Viola da Gamba, Violin, Cello, and Salicet.

Flutes may be made of wood or metal, and are distinguished by a relatively pure tone (primarily fundamental with few upper harmonics). There are many varieties of flutes, which can be classified as open flutes, tapered (or conical) flutes, harmonic flutes, stopped flutes, and half-stopped flutes.

A.2.2 Reed

The tone quality of a reed pipe is influenced by the length and shape of the resonator, the length of the tongue, and the size of the shallot.

All reeds have considerable high harmonics and a bright and sometimes brassy sound. Reed stops are classified as chorus reeds and solo reeds. Chorus reeds mix well with the ensemble (like the brass in an orchestra), while solo reeds don't mix well and are used to solo the melody (supported by softer stops on another manual).

Nomenclature

Choir See Positive

Coupler Mechanism for playing additional divisions from one keyboard.

Diapason See Principal.

Division Separate set of ranks, stops, etc. controlled by one keyboard.

Fundamental Frequency The nominal and generally lowest frequency produced by a pipe.

Great The primary manual division.

Harmonic Series Consisting of partials at $1\times$, $2\times$, $3\times$, etc. the fundamental frequency.

Manual Organ keyboard for the hands.

Mixture A stop that sounds multiple ranks, usually octaves and quints.

Mutation An off-octave harmonic stop, e.g. $2\frac{2}{3}$.

Octave Principal stops on the octave above the $8'$ pitch level.

Partial An integer multiple of the fundamental frequency. Part of the harmonic series.

Pedalboard Organ keyboard for the feet.

Positive Third division, usually with a softer foundation and solo stops.

Principal Foundational organ stops with the distinctive organ sound.

Rank Series of pipes with similar tone, one for each note on the keyboard.

Stop Mechanism for putting a rank of pipes into play. Also refers to a generic class of pipes, the actual rank of pipes in an organ, or the drawknob or tablet for controlling the stop.

Swell Secondary manual division, enclosed (at least partially) in a box with shutters that can be opened or closed with the swell pedal for dynamic expression.

Temperament The adjustment of intervals in tuning the scale. Equal temperament is most common now. Well-tempered, mean-tempered, and other temperaments were common previously.

Unison The $8'$ pitch level, corresponding to the same key on the piano.

References

[And] F. Andriaensen. Aeolus. [Online]. Available: http://www.kokkinizita.net/linuxaudio/aeolus/index.html

[Dav85] R. E. Davis, *The Organists' Manual*. W. W. Norton & Company, 1985.

[Föl90] P. Földiák, "Forming sparse representations by local anti-Hebbian learning," *Formal Aspects of Computing*, vol. 64, no. 2, pp. 165–170, 1990.

[Fyf97] C. Fyfe, "A Neural Network for PCA and Beyond," *Neural Processing Letters*, vol. 6, no. 1, pp. 33–41, 1997.

[KP00] J. Klingseisen and M. D. Plumbley, "Towards musical instrument separation using multiple-cause neural networks," in *Proc. ICA*, 2000, pp. 447–452.

[pip] How a pipe organ works. [Online]. Available: http://pipedreams.publicradio.org/articles/how_a_pipe_organ_works/howapipeorganworks.shtml

[Sau95] E. Saund, "A multiple cause mixture model for unsupervised learning," *Neural Computation*, vol. 7, no. 1, pp. 51–71, 1995.