Presented at the 136th meeting of the Acoustical Society of America, October 13, 1998

2pMU9. Musical instrument identification: A pattern-recognition approach[*]

Session: Tuesday Afternoon, October 13, 3:00 p.m.

Keith D. Martin and Youngmoo E. Kim
MIT Media Lab Machine Listening Group
Rm. E15-401, 20 Ames St., Cambridge, MA 02139

Abstract

A statistical pattern-recognition technique was applied to the classification of musical instrument tones within a taxonomic hierarchy. Perceptually salient acoustic features—related to the physical properties of source excitation and resonance structure—were measured from the output of an auditory model (the log-lag correlogram) for 1023 isolated tones over the full pitch ranges of 15 orchestral instruments. The data set included examples from the string (bowed and plucked), woodwind (single, double, and air reed), and brass families. Using 70%/30% splits between training and test data, maximum a posteriori classifiers were constructed based on Gaussian models arrived at through Fisher multiple-discriminant analysis. The classifiers distinguished transient from continuant tones with approximately 99% correct performance. Instrument families were identified with approximately 90% performance, and individual instruments were identified with an overall success rate of approximately 70%. These preliminary analyses compare favorably with human performance on the same task and demonstrate the utility of the hierarchical approach to classification.

1. Introduction

There are both scientific and practical reasons for building computer systems that can recognize and identify the instruments in music. More than a century after Helmholtz's groundbreaking research, arguments still abound over the definition of musical "timbre," and over the relative perceptual importance of various acoustic features of musical instrument sounds. There are currently no developed scientific theories about how humans, as listeners, identify sound sources, yet there are many applications in which sound source identification by computer would be useful. For example, we would like to build computer systems that can annotate musical multimedia data (Foote, in press; Wold, Blum, Keislar, & Wheaton, 1996) or transcribe musical performances for purposes of teaching, theoretical study, or structured coding (Vercoe, Gardner, & Scheirer, 1998). With better theories and models, we might even engineer a system that can understand music enough to collaborate with a human in real-time performance (Vercoe, 1984; Vercoe & Puckette, 1985). By building such systems, we stand to learn a great deal about the human system we seek to emulate.

This work is simultaneously a scientific attempt to understand timbre by quantifying the relevance of various acoustic cues for instrument identification and a practical attempt to build a piece of an annotation/transcription system. The vast literature on the production and perception of musical instrument sound suggests many potentially salient acoustic features of musical sound. In this paper, we consider several of these features and demonstrate their extraction from musical-instrument tones. We then apply pattern-recognition techniques both to evaluate the utility of these features in an identification scenario and to build a useful classifier.

2. Acoustic features for instrument identification

The enormous body of literature on musical instrument sound suggests many acoustic features that may be useful for instrument identification. Historically, these may be divided into two categories: spectral and temporal. The spectral characteristics of musical tones were first examined more than a century ago and have remained the foundation of the conventional-wisdom understanding of musical sound. More recent research, however, has shown that the temporal properties of musical-instrument sounds are *at least* as

[*] Portions of Section 3 were reported in (Martin, 1998).

important as spectral properties for sound-source recognition and for timbral quality. Handel's (1995) overview, written from a modern viewpoint, attempts to unify the two viewpoints and is an excellent home base for forays into the timbre literature jungle.

In the first extensive scientific investigation of musical-tone properties, Helmholtz (1954) showed that the relative amplitudes of the harmonic partials that compose a periodic tone—much more so than their relative phases—are the primary determinants of the tone's sound quality. Subsequently, researchers have identified several important sub-features of the harmonic magnitude spectrum. For example, Strong (1963) interpreted the spectra of several orchestral instruments in terms of *formants* and demonstrated that the oboe, clarinet, and bassoon are identified primarily on the basis of their magnitude spectra. Luce (1967) reported that unmuted brass instruments exhibit a single formant, characterized by a single cutoff frequency, and Benade suggested that the cutoff frequency is one of the principal determinants of sound quality in woodwind instruments (Benade, 1990). The spectral centroid—e.g., the "balancing point" of the spectrum—correlates strongly with "brightness," a primary subjective dimension of timbre (Grey, 1977; Lichte, 1941; von Bismarck, 1974), and for many instruments it varies characteristically with sound intensity (Beauchamp, 1981; Beauchamp, 1982). Although most musical tones consist of harmonically related partials (Brown, 1996), deviations from harmonicity are responsible for the "warmth" of low piano tones (Fletcher, Blackham, & Stratton, 1962) and the "bite" of saxophone onsets (Freedman, 1967). Also, bowed string tones are inharmonic during both their attack and decay (Beauchamp, 1974).

The temporal features of musical sounds also provide many identifying characteristics. As early as 1910, Stumpf recognized the importance of onset for musical timbre (Risset & Wessel, 1982). Strong (1963) showed that several orchestral instruments, including trumpet, flute, trombone, and French horn, are primarily identified on the basis of their temporal envelopes (in part because their spectral envelopes are not unique). The relative rates of energy buildup in the harmonic partials are a salient feature of some instrument tones; Luce (1963), for example, noted that the high partials in brass tones rise in energy more slowly than the low. Vibrato and pitch jitter are salient cues both on their own and through interaction with resonances in the source (Luce, 1963; McAdams, 1984; Risset & Mathews, 1969; Robertson, 1961; Saldanha & Corso, 1964; Tenney, 1965).

Although there have been many reports of the onset's preponderant importance for instrument identification, there is compelling evidence that its importance is context dependent. Kendall (1986) demonstrated that, in musical phrases, properties of the "steady state" are at least as important as transient properties. It is likely that the importance of onset transients has been exaggerated by psychophysical studies performed using isolated instrument tones. There are no accepted criteria for separating the onset transient from the "steady state." Indeed, the distinction is often quite arbitrary, and "onset transients" that contain portions of the "steady state" confound interpretations of many studies.

3. Feature extraction

In an artificial auditory recognition system, the signal representation should encode the salient acoustic features simply and robustly and contain a level of detail similar to that preserved in the human auditory system. For the features identified above, and with these considerations in mind, the *log-lag correlogram* appears to be a good choice of representation. Unlike analysis methods based on the short-time Fourier transform (STFT) or equivalent heterodyne filtering, the correlogram is not based on an assumption that the signal is periodic; it may therefore be better suited than previously studied representations to analysis of inharmonic signals.

The correlogram adopted here is based on Ellis's implementation (Ellis, 1996), which corresponds closely to Licklider's original proposal (Licklider, 1951). Processing occurs in three stages. In the first, the raw acoustic waveform passes through a gammatone filterbank (Slaney, 1993), which models the frequency resolution of the cochlea. Each filter channel is half-wave rectified and lightly smoothed as a simplified model of inner hair cell transduction. These operations remove fine timing structure in high-frequency channels, while preserving the signal's envelope. Although this process does not model the adaptation and dynamic range compression performed by the inner hair cells, the intensity of the signal in each channel can be computed trivially from the resulting representation.

In the third stage, each channel is subjected to short-time autocorrelation, implemented by a delay/multiply/smooth architecture with a 25 ms smoothing constant. The running autocorrelation output is computed with logarithmic lag spacing (each lag is implemented with a fractional-delay filter). The zero-lag autocorrelation is also computed, as a measure of the short-time energy in each channel.

The correlogram representation is three-dimensional. The first dimension (cochlear position) yields critical-band frequency resolution, which can resolve the first five or six harmonics of a periodic signal, corresponding to human abilities (Plomp, 1976). The second dimension (autocorrelation lag) is a logarithmic representation of periodicity, corresponding to the nearly logarithmic pitch resolution exhibited by humans. The third dimension is time. The main panels of Figures 1-3 display snapshots of the correlogram output for 555 Hz tones produced, respectively, by violin, trumpet, and flute.

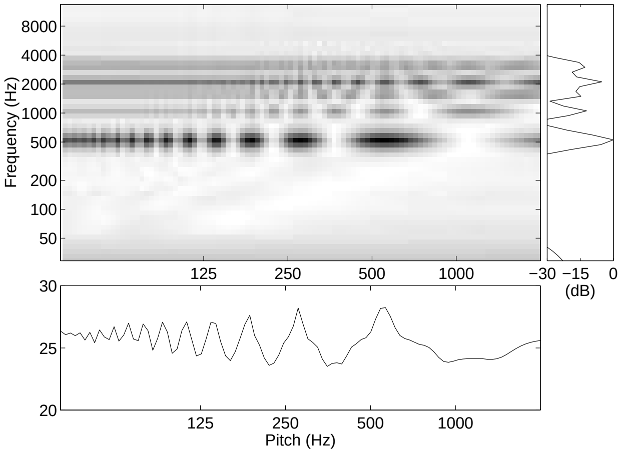

Figure 1: Correlogram snapshot of a violin tone. The horizontal axis, labeled “pitch,” is the inverse of autocorrelation lag. The vertical axis, labeled “frequency,” corresponds to cochlear position. The lower panel displays the summary autocorrelation (the correlogram integrated over the cochlear dimension). The right-hand panel displays the zero-lag energy, which for isolated periodic sources is equal to the spectral envelope.

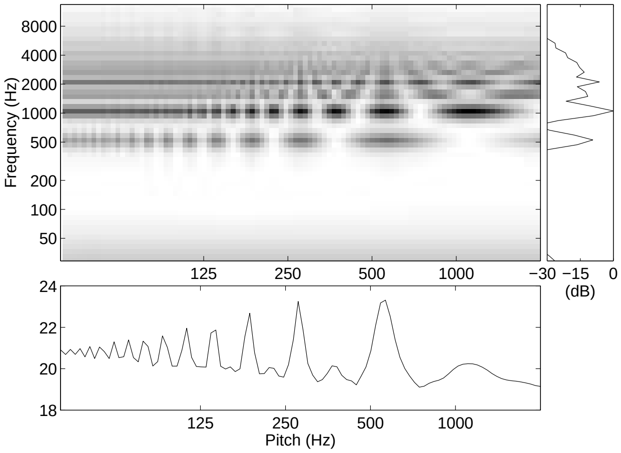

Figure 2: Correlogram snapshot of a trumpet tone. See the caption to Figure 1 for a description of the panels.

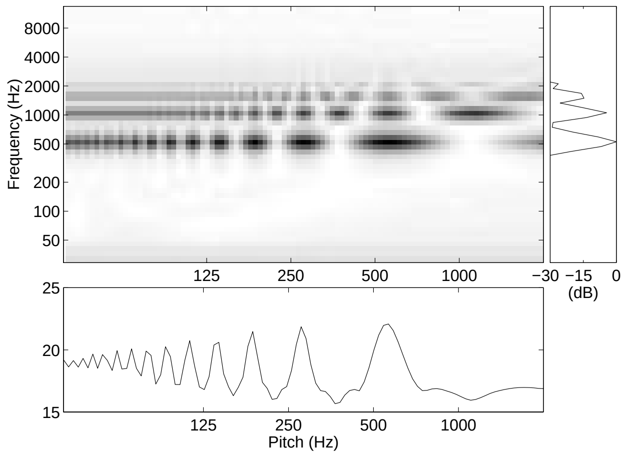

Figure 3: Correlogram snapshot of a flute tone. See the caption to Figure 1 for a description of the panels.

Many of the features mentioned in the previous section are captured vividly in the correlogram representation:

Pitch – Signals yielding a pitch percept exhibit structure in the correlogram; in a two-dimensional snapshot, with lag on the horizontal axis and frequency on the vertical, vertical ridges indicate the period of the signal—and by inversion, the pitch. Vibrato and jitter may be characterized by tracking the pitch over time. Figure 4 displays the pitch over the first 2 seconds of the tones shown in Figures 1-3.

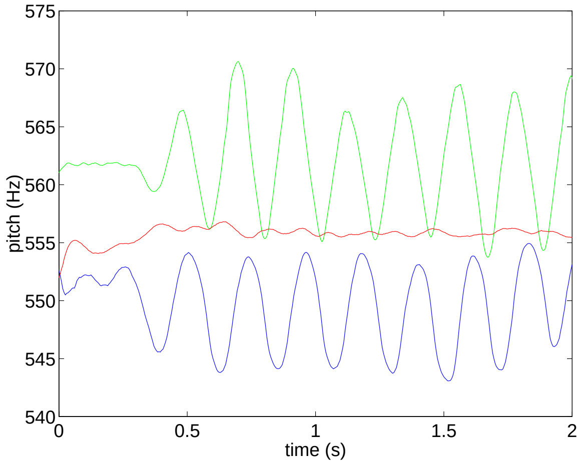

Figure 4: Pitch modulation in tones produced by [bottom to top] violin, trumpet, and flute. (The three tones were performed at the same pitch; for display purposes, the violin's pitch-track has been offset by –5 Hz—the flute's by +5 Hz). The violin and flute tones exhibit periodic frequency modulations consistent with musical *vibrato*. The trumpet tone exhibits random frequency modulations consistent with *jitter*.

Spectral envelope – Once the pitch has been determined, the height of the correlogram's vertical ridge may be measured as a function of frequency to yield an estimate of the spectral envelope with critical-band resolution. The spectral centroid is simply the centroid of the spectral envelope. Figure 5 displays the spectral centroid over the first 2 seconds of the tones shown in Figures 1-3.

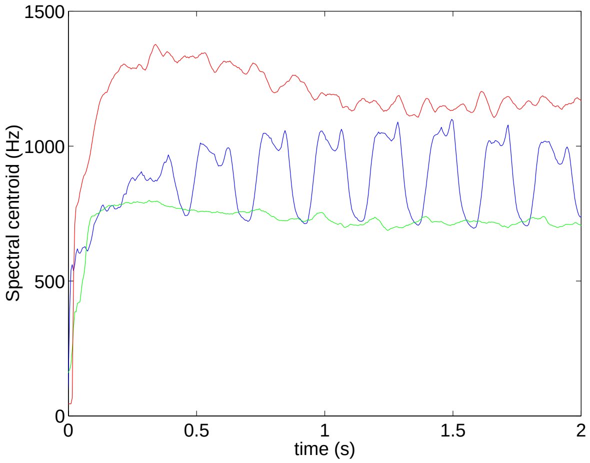

Figure 5: Spectral centroid for tones produced by [bottom-to-top] flute, violin, and trumpet. The trumpet tone is "brighter" than the violin and flute tones. Also note the large degree of variation of the violin tone's spectral centroid during vibrato as compared to the flute's.

Intensity – The sum of the energy in the spectral envelope approximates the instantaneous loudness of the signal. Tracking this over time leads to simple measures of amplitude modulation, which can reveal tremolo—and, by correlation with frequency modulation, resonances. As suggested by Beauchamp (1981; 1982), the relationship between intensity and the spectral centroid may be an important perceptual correlate of timbre. Figure 6 displays the instantaneous loudness estimate over the first 2 seconds of the tones shown in Figures 1-3.

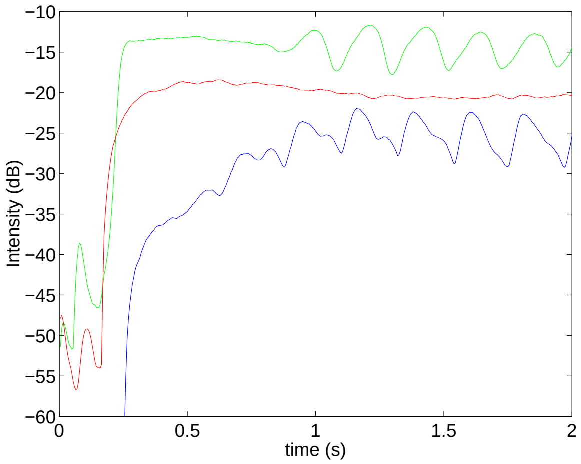

Figure 6: Amplitude envelope for tones produced by [bottom to top] violin, trumpet, and flute. (For display purposes, the violin amplitude has been offset by –5 dB—the flute by +5 dB.) Both the violin and flute tones have clear amplitude modulation. The violin takes much longer (nearly 500 ms!) to reach its steady state energy level than the other two instruments. The flute's onset is nearly instantaneous.

Onset asynchrony – By tracking the spectral envelope over time, with concurrent pitch estimates, it is possible to measure the onset characteristics of a musical tone's harmonics in a psychophysically appropriate manner. Since the first few harmonics are resolved by the filterbank processing stage, they may be monitored individually; upper harmonics may also be monitored, as groups within third-octave bands (this level of detail is consistent with Charbonneau's (1981) study of human sensitivity to the behavior of individual partials).

Inharmonicity – Deviations from harmonicity in the signal will be reflected as deviations from strict vertical structure in the correlogram snapshot. This feature, however, has not been incorporated in the current study.

In this study, we extracted 31 features from each instrument tone, including the pitch, spectral centroid, onset asynchrony (both the relative onset times at various frequencies, and their overall variation), ratio of odd-to-even harmonic energy, and the strength of vibrato and tremolo. Many of the 31 features were variations on other features in the set, measured in a slightly different manner. The feature set was intended to be representative of the many possibilities, but certainly not exhaustive. For example, the *shape* of the spectral envelope was not considered at all. Table 1 contains a list of the features that were extracted. Details of the extraction algorithms have been omitted for space considerations.

Average pitch over steady state	Tremolo frequency
Average pitch Δ ratio †	Tremolo strength
Pitch variance	Tremolo heuristic strength ‡
Pitch variance Δ ratio †	Spectral centroid modulation
Average spectral centroid (Hz)	Spectral centroid modulation strength
Spectral centroid Δ ratio †	Spectral centroid modulation heuristic strength ‡
Variance of spectral centroid	Normalized spectral centroid modulation
Spectral centroid variance Δ ratio †	Normalized spectral centroid modulation strength
Average normalized spectral centroid	Normalized spectral centroid modulation heuristic strength ‡
Normalized spectral centroid Δ ratio †	Slope of onset harmonic skew
Variance of normalized spectral centroid	Intercept of onset harmonic skew
Normalized spectral centroid variance Δ ratio †	Variance of onset harmonic skew
Maximum slope of onset (dB/msec) Onset duration (msec)	Post-onset slope of amplitude decay
Vibrato frequency (Hz) Vibrato amplitude Vibrato heuristic strength ‡	Odd/even harmonic ratio

† The Δ ratio is the ratio of the feature value during the transition period from onset to steady state (~100 msec) to the feature value after the transition period.

‡ The heuristic strength of a feature is the peak height of the feature divided by the average value surrounding the peak.

Table 1: List of features extracted from each tone.

4. Recognition strategies

Human perception is based in part on taxonomic organizations. The most obvious of these are zoological: “Spot” is a Dalmatian, which is a kind of dog, which is a kind of animal. When we recognize an object in the world, we recognize it first at an intermediate level of abstraction, which Rosch has dubbed the “basic” level (Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). After initial recognition, the identification process proceeds either up or down the taxonomic hierarchy, to more or less abstract levels respectively. In the example, we see that Spot is a dog before we identify his breed or particular identity.

This manner of processing has strong computational advantages over direct classification at the lowest level of the taxonomy, because it is often easier to make distinctions at abstract levels in the taxonomy. At more abstract levels, there are fewer classes, and decision processes may be simpler than they would be if classification began at a less abstract level. As recognition proceeds down the branches of the taxonomy, each decision is relatively simple, relying on only a few features to distinguish among the small number of classes at each node.

We speculate that recognition of musical instruments happens in this manner: before the particular instrument is identified, the instrument *family* (or *articulation class*) is recognized. This corresponds well with our subjective experience; it is often easy to tell that a bowed string instrument has produced a particular sound, but much more difficult to determine whether the instrument is a violin or a viola. Although the traditional instrument family taxonomy is based on the form, materials, and history of the instruments rather than the sounds they make, instruments within a family do have many acoustic properties in common. For example, string instruments typically have complex resonant properties and very long attack transients; in contrast, brass instruments have a single resonance and much shorter attacks. Woodwinds have short attacks and resonant properties of intermediate complexity. Figure 7 depicts a possible taxonomic organization of orchestral instrument sounds.

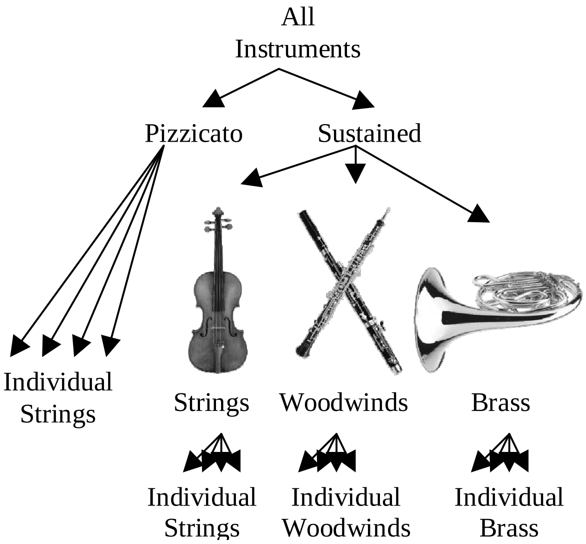

Figure 7: A possible taxonomy of orchestral instrument sounds.

Statistical classifiers require a set of training data whose size grows exponentially with the number of feature dimensions; with 31 features, that data set size would be exorbitantly large. To reduce the training requirements, we employed Fisher multiple discriminant analysis (McLachlan, 1992) at each decision point in the taxonomy. The Fisher technique projects the high-dimensional feature space into a space of fewer dimensions (the number of dimensions is one fewer than the number of data classes) where the classes to be discriminated are maximally separated. The analysis yields the mean feature vector and covariance matrix (in the reduced space) of a single Gaussian density for each class, which can be used to form maximum *a posteriori* (MAP) classifiers by introducing prior probabilities. Figures 8 and 9 show the decision spaces found at two of the nodes of the taxonomy.

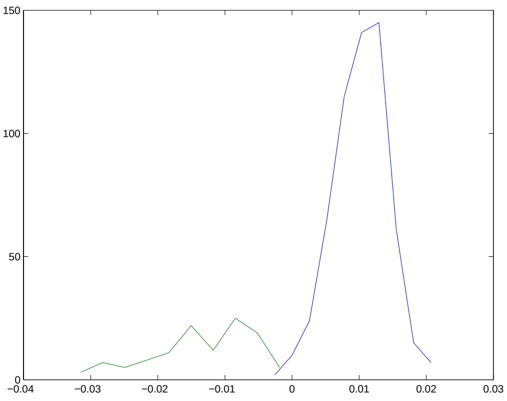

Figure 8: Fisher projection for the Pizzicato vs. Sustained node of the taxonomy. Since there are two classes, the projection is one-dimensional. There are two “modes” in the projection: the one on the left-hand side corresponds to Pizzicato tones; the one on the right to Sustained tones. The Sustained tone distribution is favored by prior probability and therefore appears larger. The axes are not labeled; the abscissa is a linear combination of the 31 features.

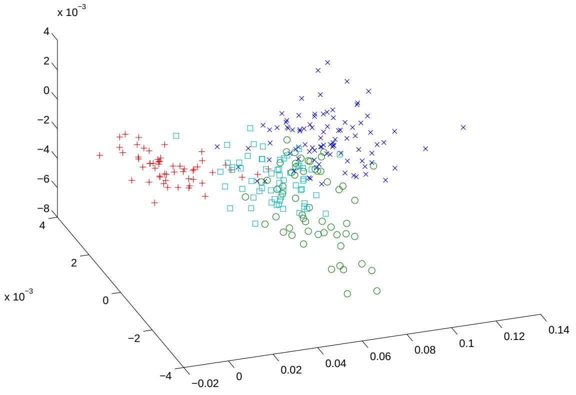

Figure 9: Fisher projection for classifying the individual string instruments. There are four classes and thus three dimensions in the projection. Violin data points are plotted with X's, viola with O's, cello with plus symbols and double bass with squares. The axes are not labeled. Each axis is a linear combination of the 31 features.

In addition to the Fisher projection technique, we tested two varieties of *k*-nearest neighbor (*k*-NN) classifiers. A *k*-NN classifier works by memorizing all of the training samples. When a new sample is to be classified, the system finds the *k* nearest training samples in the feature space, and the new sample is classified by majority rule.

5. Results

We used a database of 1023 solo tones performed over the entire pitch ranges of 14 orchestral instruments (violin, viola, cello, bass, flute, piccolo, clarinet, oboe, English horn, bassoon, trumpet, trombone, French horn, and tuba) with several different articulation styles (e.g., pizzicato, bowed, muted). The tones were recorded from the McGill Master Samples collection (Opolko & Wapnick, 1987). Each classifier was trained with 70% of the tones, leaving 30% as independent test samples. All classifiers were cross-validated with multiple 70%/30% splits.

The Fisher projection method resulted in successful classifiers at all levels of the taxonomy. In our initial implementation, the instrument family was correctly identified for approximately 85% of the test samples, and the particular instrument for approximately 65%. In examining the features for the various instrument families, it was apparent that while the string and brass families have many consistent characteristics, the woodwind family is much more heterogeneous. More precisely, all string instruments and all brass instruments, respectively, are excited in the same manner: string instruments (played non-pizzicato) by the bow, and brass instruments by the periodic lip vibrations of the performer. Woodwind instrument excitation, however, comes in several forms: single reed (clarinet), double reed (oboe, English horn, and bassoon), and air reed (flute, piccolo). By reorganizing the hierarchy to take these differences into consideration, the performance of the classifier was improved. Figure 10 illustrates the revised taxonomy and Table 2 contains a summary of the classification performance of the hierarchical Fisher classifier, a hierarchical *k*-NN classifier, and a non-hierarchical *k*-NN classifier. The results are averaged over 200 test runs with different training/test data splits. The hierarchical Fisher classifier performs best, particularly at the individual instrument level.

	Hierarchical Methods		Non-Hierarchical
	Fisher + MAP	**k-NN**	**k-NN**
Pizzicato vs. continuant	98.8%	97.9%	97.9%
Instrument family	85.3%	79.0%	86.9%
Individual instruments	71.6%	67.5%	61.3%

Table 2: Classification results for the three classifiers tested. Each result was cross-validated with 200 test runs using 70%/30% splits of training/test data.

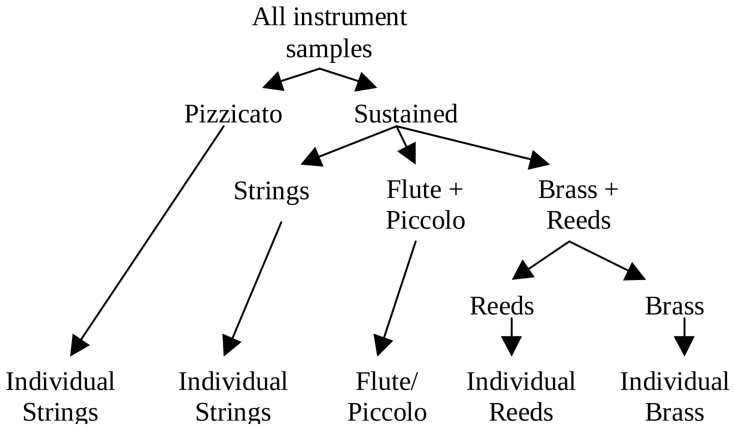

Figure 10: Revised instrument taxonomy, dividing the woodwinds into more homogenous groups.

6. Feature salience

Although Fisher and *k*-NN techniques yield successful classifiers, they provide little insight into the relative importance of the various individual features. It would be valuable to know if particular features are good at characterizing particular instruments or families. To that end, we employed a step-forward algorithm to find the best features for isolating each instrument family. A step-forward algorithm works by testing each feature individually and choosing the best as the “current set”. The algorithm continues by testing all combinations of the current set with each of the remaining features, adding the best of these to the “current set” and repeating. For computational simplicity, we used *k*-NN classifiers in this part of the study. This procedure was followed using three different 70%/30% splits of the training/test data, iterating 10 times to find the 10-feature combination that provided the best average performance over the three different data sets.

By using only the 10 best features at each node, the system’s success rate for instrument family identification increased to 93%. Some of the features were generally salient across many of the instrument families, and some were particularly useful in distinguishing single families. The most common features selected for each subgroup are listed in Table 3.

Subgroup	Selected features
Strings	Vibrato strength Onset harmonic skew Average spectral centroid
Brass	Vibrato strength Variance of spectral centroid Onset harmonic skew
Single reed	Pitch variance Onset duration Vibrato strength Onset harmonic skew
Air reed	Pitch Onset duration Tremolo strength Spectral centroid Vibrato frequency
Double reed	Vibrato strength Average spectral centroid Spectral centroid modulation Onset harmonic skew

Table 3: Features that were particularly useful in distinguishing single instrument families.

Vibrato strength and features related to the onset harmonic skew (roughly, the relative onset times of the various partials) were selected in four of the five instrument subgroups, indicating their relevance across a wide range of instrument sounds. One interesting omission occurs with the clarinet (single reed). One of the original 31 features was the ratio of odd to even harmonic strength. The conventional-wisdom about the clarinet is that its odd partials are much stronger than its even partials, but this is not true over the clarinet's entire range (it is a good approximation only in the lowest register).

7. Conclusions

This study has two important results. First, it demonstrates the utility of a hierarchical organization of musical instrument sounds; humans recognize objects taxonomically, and it is to the benefit of computer systems that they do so as well. Second, it demonstrates that the acoustic properties studied in the literature as components of musical timbre are indeed useful features for musical instrument recognition.

It is dangerous to compare the results in this study with human performance on similar tasks. The system we have demonstrated has learned to identify some of the instruments from the McGill Samples with great success, but so far there has been no demonstration of generalization to different recordings of the same instruments. It seems likely, since the features we have employed are perceptually motivated, that the system will generalize, but this has not been tested.

In everyday life, we almost never hear just a single tone played on a musical instrument. Rather, we hear musical phrases. Recent work, for example by Kendall (1986), suggests that human instrument identification based on phrases is much more robust than identification based on isolated tones. We are currently at work on extracting features from phrases taken directly from commercial recordings, and we will soon see if these ideas generalize to a more realistic context.

8. References

Beauchamp, J. W. (1974) Time-variant spectra of violin tones. *J. Acoust. Soc. Am.,* **56**(3), 995-1004.

Beauchamp, J. W. (1981) Data reduction and resynthesis of connected solo passages using frequency, amplitude, and 'brightness' detection and the nonlinear synthesis technique, *Proc. ICMC* (pp. 316-323).

Beauchamp, J. W. (1982) Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. *J. Audio Eng. Soc.,* **30**(6), 396-406.

Benade, A. H. (1990) *Fundamentals of Musical Acoustics.* (Second ed.). New York: Dover.

Brown, J. C. (1996) Frequency ratios of spectral components of musical sounds. *J. Acoust. Soc. Am.,* **99**(2), 1210-1218.

Charbonneau, G. R. (1981) Timbre and the Perceptual Effects of Three Types of Data Reduction. *Computer Music Journal,* **5**(2), 10-19.

Ellis, D. P. W. (1996) *Prediction-driven computational auditory scene analysis.* Unpublished Ph.D. Thesis, Massachusetts Institute of Technology.

Fletcher, H., Blackham, E. D., & Stratton, R. (1962) Quality of Piano Tones. *J. Acoust. Soc. Am.,* **34**(6), 749-761.

Foote, J. (in press) An overview of audio information retrieval. *ACM Multimedia Systems Journal*.

Freedman, M. D. (1967) Analysis of musical instrument tones. *J. Acoust. Soc. Am.,* **41**, 793-806.

Grey, J. M. (1977) Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.,* **61**(5), 1270-1277.

Handel, S. (1995) Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing* . New York: Academic Press.

Helmholtz, H. (1954) *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (A. J. Ellis, Trans.): Dover.

Kendall, R. A. (1986) The role of acoustic signal partitions in listener categorization of muscial phrase. *Music Perception,* **4**(2), 185-214.

Lichte, W. H. (1941) Attributes of complex tones. *J. Experim. Psychol.,* **28**, 455-481.

Licklider, J. C. R. (1951) A duplex theory of pitch perception. *Experientia,* **7**, 128-133.

Luce, D. (1963) *Physical Correlates of Nonpercussive Musical Instrument Tones.* Unpublished Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Luce, D. & Clark, M. (1967) Physical correlates of brass-instrument tones. *J. Acoust. Soc. Am.,* **42**, 1232-1243.

Martin, K. D. (1998). *Toward automatic sound source recognition: identifying musical instruments.* Paper presented at the NATO Computational Hearing Advanced Study Institute, Il Ciocco IT.

McAdams, S. (1984) *Spectral fusion, spectral parsing and the formation of auditory images.* Unpublished Ph.D. Thesis, Stanford University.

McLachlan, G. J. (1992) *Discriminant Analysis and Statistical Pattern Recognition*. New York, NY: Wiley Interscience.

Opolko, F. & Wapnick, J. (1987) *McGill University Master Samples [Compact disc]*: Montreal, Quebec: McGill Univeristy.

Plomp, R. (1976) *Aspects of Tone Sensation.* London: Academic Press.

Risset, J.-C. & Mathews, M. V. (1969) Analysis of musical-instrument tones. *Physics Today,* **22**(2), 23-30.

Risset, J.-C. & Wessel, D. L. (1982) Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 26-58). New York: Academic.

Robertson, P. T. (1961) *The aurally perceptual basis for the classification of musical instruments by families.* Unpublished Bachelor's Thesis, Massachusetts Institute of Technology.

Rosch, E. (1978) Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* . Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology,* **8**, 382-439.

Saldanha, E. L. & Corso, J. F. (1964) Timbre cues and the identification of musical instruments. *J. Acoust. Soc. Am.,* **36**, 2021-2026.

Slaney, M. (1993) *An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank* (Technical Report 35): Apple Computer.

Strong, W. J. (1963) *Synthesis and recognition characteristics of wind instrument tones.* Unpublished Ph.D. Thesis, Massachusetts Institute of Technology.

Tenney, J. C. (1965) The Physical Correlates of Timbre. *Gravesaner Blaetter,* **26**, 106-109.

Vercoe, B. L. (1984). *The synthetic performer in the context of live performance.* Paper presented at the ICMC, Paris.

Vercoe, B. L., Gardner, W. G., & Scheirer, E. D. (1998) Structured audio: The creation, transmission, and rendering of parametric sound representations. *Proc. IEEE,* **85**(5), 922-940.

Vercoe, B. L. & Puckette, M. S. (1985). *Synthetic rehearsal: Training the synthetic performer.* Paper presented at the ICMC, Burnaby BC, Canada.

von Bismarck, G. (1974) Timbre of steady sounds: a factorial investigation of its verbal attributes. *Acustica,* **30**, 146-159.

Wold, E., Blum, T., Keislar, D., & Wheaton, J. (1996) Content-based classification, search, and retrieval of audio. *IEEE Multimedia*(Fall), 27-36.