Who's Playing? Towards Machine-Assisted Identification of Jazz Trumpeters by Timbre

Janet G. Lazar, Michael Lesk
Rutgers University

Abstract
The purpose of our proposed study is to contribute to the growing research in machine-assisted identification of jazz performers. In particular, we seek to identify unknown jazz trumpeters. We plan to take an approach that has not received recent attention; namely, using human observation to compare spectrograms and other data representing musical timbre. We believe that human observation, when combined with machine learning, will improve accuracy of timbre recognition and performer identification. We will collect 100 music samples: five each from 20 trumpeters. We will manually sort spectrograms and other data in order to distinguish the most salient timbre characteristics. Once we choose those features, we will use a computer to filter for them. If our approach is successful, we will develop a larger database of trumpet solos.
Keywords: Jazz information retrieval; music information retrieval; performer identification; timbre recognition; individuality in music performance
doi: 10.9776/16481

Contact: janet.lazar@rutgers.edu, lesk@acm.org

1 Introduction

The purpose of our proposed study is to contribute to the growing research in machine-assisted identification of jazz performers. In particular, we are seeking to identify unknown jazz trumpet players in small ensembles. The trumpet is an appropriate instrument for this study because it often plays the melody and is easily distinguished from the rhythm instruments that accompany it.

We propose to return to an approach that has not received recent attention; namely, using human observation to compare spectrograms and other data representing the timbre of different jazz trumpeters. Current technology allows these spectrograms and other data to be filtered in many different ways, and we plan to use this technology to help us distinguish timbre patterns unique to the individual performers.

Sound identification by machine has a range of potential uses in addition to performer identification. In the field of Music Information Retrieval (MIR), sound identification is used to distinguish genres, composers, individual works, and types of instruments. In other fields, sound identification is involved in achieving a variety of goals, including the identification of background noises by law enforcement and assisting the hearing impaired with voice recognition.

Jazz information retrieval (JIR) is a subfield of MIR that is still in its infancy. Historically, jazz scholarship has been neglected. Recently, however, that trend has reversed. An act of Congress has designated jazz as a national treasure and scholars worldwide are turning to the study of JIR (Farley, 2011).

JIR has as its goal improving the availability of jazz information resources to academics, performers, students, and audiences. To all these users, the correct identification of recorded jazz instrumentalists is important. For example, the Institute of Jazz Studies in Newark, New Jersey, a world-renowned jazz archive and research facility associated with Rutgers University (Rutgers, 2015), has in its collection many unpublished recordings for which some or all of the musicians are unnamed. The staff of the IJS has told the authors that a means of identifying these performers would be invaluable to them and to the users of the archive.

Note that because jazz is largely improvisational, the performers are co-creators, contributing as much as – or sometimes more than – the composers of the songs on which the improvisations are based. Identifying the performers is significant from the perspective of musicology, historiography, and pedagogy, as well as aesthetics.

2 Context

"Who's playing?" is a question that every music-lover has asked upon hearing an unidentified rendition of a tune. This widespread wish to identify performers is one of the raisons d'être of smartphone applications like Shazam. Sufficiently precise metadata might be able to identify all the performers on a given recording. But often that metadata doesn't exist. See, for example, the disc pictured below (Figure 1), from 1915.

Figure 1. Image of 1915 disc of St. Louis Blues played by Prince's Band

The label names the band but none of the participating musicians. Often, for a recording of this age, the issuing company is out of business and any paperwork from the recording session has been lost. In addition, there are many unpublished recordings for which insufficient documentation exists (Berman, 2002). Jazz scholars can make educated guesses as to the identity of the performers, using discographies, their knowledge of jazz history, and their own trained ears. But reliable machine identification would be a boon to JIR.

One approach to automatic sonic identification is visualizing the sound patterns. Scientists have long been interested in visualizing patterns of sound, and have done so using water, sand, and other media (Wood, 1913, 1944). By the 1940's, oscilloscopes were employed for the same purpose (Bartholomew, 1942, 1950).

Computers were used to model sound at least as early as the 1960's (Hewlett & Selfridge-Field, 1991). Such a model is sometimes called a spectrogram. The term refers to a visual representation of the spectrum of sound frequencies as they vary with time or intensity. Aleksandr Solzhenitsyn describes "reading" from sound spectrograms in his 1968 novel, *The First Circle*.

By the late 20th century, software had been developed to generate spectrograms of performed music. In one relevant study, these spectrograms were manually compared in order to identify jazz performers (Smith & Westbrook, 2001). More recently, scholars have turned to machine learning to identify performers in various genres, including jazz (see, for example, Ramirez, et al., 2007; Chudy & Dixon, 2010).

Different acoustic features are relevant to performer identification, Timbre has proven to be a robust feature for that purpose (Tzanetakis, 2012) and, as such, is the focus of this proposed study. We believe that the accuracy of timbre recognition can be improved by human observation of spectrograms, along the lines of Smith & Westbrook (2001), in combination with the findings from the various machine learning studies.

It has already been shown that humans can perceive auditory differences in trumpet tone quality, including differences in timbre, and that these perceptions can serve as an intermediate step to identifying the relevant spectral characteristics (Knight, Upham, & Fujinaga, 2011). We wish to use a related approach to identify the timbral features that distinguish one trumpeter from another.

3 Preliminary work

To begin our study, we used basic time-frequency spectrograms generated by Audacity, the open-source acoustic modeling software. Sound samples are widely available, including from the source we selected, the Internet Archive. We chose to start with excerpts from performances by three trumpeters, Louis Armstrong, Harry James, and Wynton Marsalis, of W. C. Handy's *St. Louis Blues*.

In order to simplify the first test, we narrowed the spectrograms to samples of one note only. Examples of the spectrograms generated are shown in Figure 2, below.

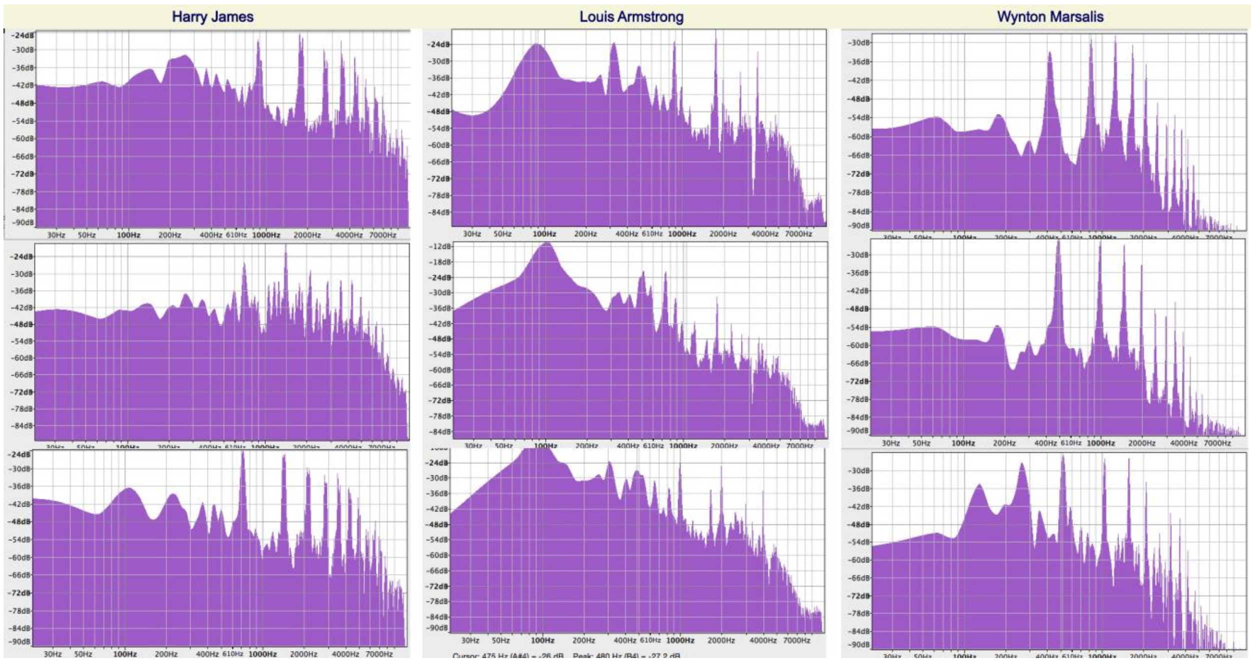

Figure 2. Spectrograms of three trumpeters playing a single note

The spectrograms of these three performers are distinctive, with Marsalis's showing a clear note, Armstrong's showing a note with somewhat fuzzy edges, and James's showing a note clearer than Armstrong's but fuzzier than Marsalis's. These visual judgments as to fuzziness and clarity accord with our auditory judgments of the playbacks.

For each note, we calculated the maximum energy at the note's peak frequency divided by the energy at 100 Hz (taken as a background). The data for the three examples are plotted below (Figure 3).

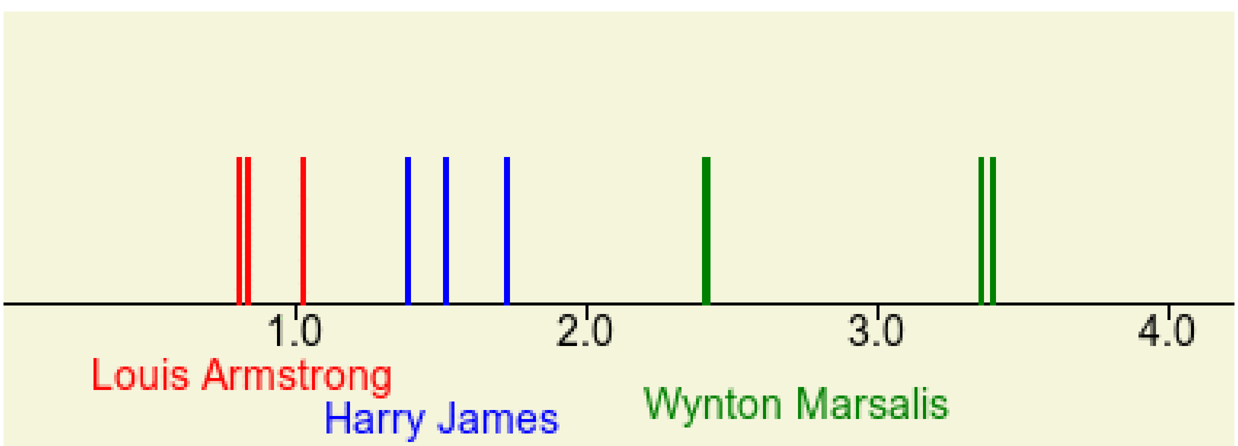

Figure 3. Graph of energy at frequency peak

This trivial example shows a clear distinction as to the amount of energy at the frequency peak compared with energy in other ranges. Armstrong has as strong a frequency output below the note as at the note; Marsalis has relatively little energy outside of the specific note being played.

Next (Figure 4) are examples of sound spectrograms for a slightly longer (10 second) performance by Louis Armstrong, Harry James and Wynton Marsalis.

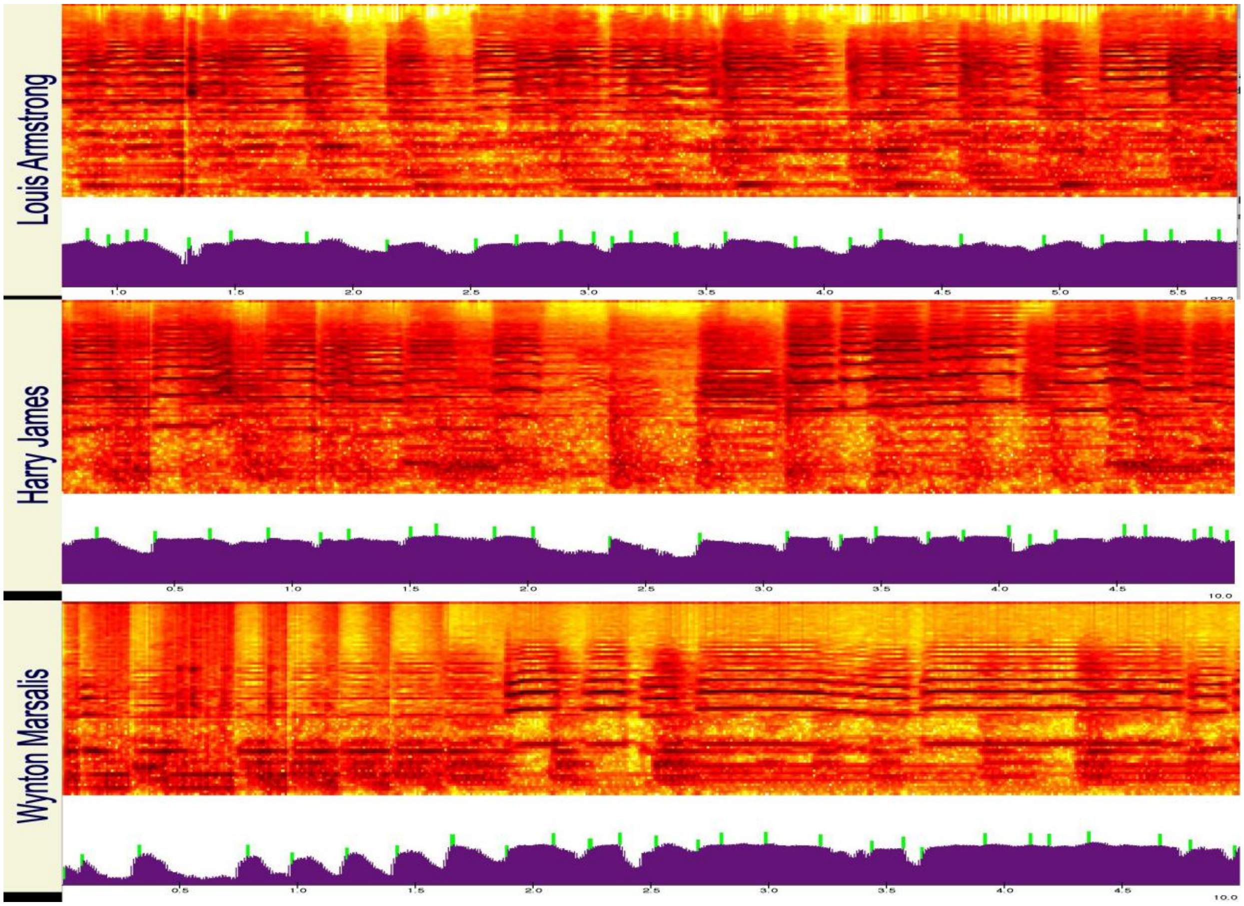

Figure 4. Spectrograms of 10-second samples for three trumpeters

A cursory inspection of these sequences shows Armstrong with the most complex sound (least dominated by the main note frequency) and Marsalis with the most straightforward sound. Marsalis's playing is the most staccato, with evident separation between individual notes; Armstrong and James play more smoothly. Marsalis plays with the most stable notes (i.e., with the least vibrato) while James's playing is a bit more variable and Armstrong's the most variable.. The rise time is fastest for Marsalis's notes and slowest for Armstrong's. None of the musicians follows the beat mechanically, but Marsalis's playing is the most regular.

4 Future work

We plan to collect five examples each of playing by 20 trumpeters. We will then manually sort the spectrograms and other data and attempt to distinguish the most salient timbre characteristics by observation. As the late Yogi Berra famously said, "You can observe a lot just by watching." Why not rely solely on machine learning? The risks range from accidental overinclusion – of microphone characteristics or other irrelevant properties – to accidental underinclusion – of minor but salient properties. We intend our technique to supplement machine learning, not to substitute for it.

Once we choose timbre characteristics for this set of samples, we will use the Timbre Toolbox (Peeters, et al., 2011) utilities to filter for the chosen features, and in turn test for these features and identify the performers. We will then evaluate on a different set of five performances each by the same 20 players.

If our approach is successful on this limited sample, we will develop a larger database of trumpet solos. This larger database will be used for our own further study and will also be made available for study by other jazz scholars. Ultimately, we hope our approach can be used to help develop a reliable tool for automatic performer identification, which in turn can be adapted for broader purposes that also involve sonic identification.

5 Table of figures

6 References

Audacity (Version 2.1.1, 2015) [software]. Available from http://audacityteam.org/

Bartholomew, W. T. (1942, 1950). *Acoustics of Music.* New York: Prentice Hall. Retrieved from http://babel.hathitrust.org/cgi/pt?id=mdp.49015000730417

Berman, J. (2002). The jazz detectives: Identifying unknown or mislabeled musicians. *Jazz Times, 32.* Retrieved from http://jazztimes.com/articles/19770-the-jazz-detectives-identifying-unkown-or-mislabeled-musicians

Chudy, M., & Dixon, S. (2010). Towards music performer recognition using timbre features. *Proceedings of the 3rd International Conference of Students of Systematic Musicology* [SysMus 10], 45-50. Retrieved from https://www.eecs.qmul.ac.uk/~simond/pub/2010/Chudy-Dixon-SysMus-2010.pdf

Farley, J. (2011). Jazz as a Black American art form: Definitions of the Jazz Preservation Act. *Journal of American Studies, 45*(1), 113-129. doi:10.1017/S002187581000127

Hewlett, W. B., & Selfridge-Field, E. (1991). Computing in musicology, 1966-91. *Computers and the Humanities, 25*(6), 381-392. doi:10.2307/30208121

Internet Archive (n.d.). The St. Louis Blues [web page]. Retrieved from https://archive.org/details/TheSt.LouisBlues

Knight, T., Upham, F., & Fujinaga, I. (2011). The potential for automatic assessment of trumpet tone quality. *Proceedings of the 12th International Society for Music Information Retrieval Conference* [ISMIR 2011], 573–578. Retrieved from http://ismir2011.ismir.net/papers/PS4-17.pdf

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America, 130*(5), 2902-2916. Retrieved from http://www.music.mcgill.ca/~jason/mumt621/Peeters_2011_JASA.pdf

Ramirez, R., Maestre, E., Pertusa, A., Gomez, E., & Serra, X. (2007). Performance-based interpreter identification in saxophone audio recordings. *IEEE Transactions on Circuits and Systems for Video Technology, 17*(3), 356-364. doi:10.1109/TCSVT.2007.890862

Rutgers, The State University of New Jersey. (2015). Institute of Jazz Studies – Mission. Retrieved from http://newarkwww.rutgers.edu/IJS/jazz1aa_about.html

Shazam Entertainment Limited (n.d.). Shazam Apps [web page]. Retrieved from http://www.shazam.com/apps

Smith, T., & Westbrook, G. (2001). Acoustic technology for the identification of mystery jazz recordings. *Jazz Research Proceedings Yearbook,* 31, 61-66.

Tzanetakis, G. (2012). Audio feature extraction. In *Music Data Mining*, pp. 43-74. Boca Raton, Florida: CRC Press.

Wood, A. (1913, 1944). *The Physics of Music*. London: Methuen. Retrieved from https://archive.org/details/physicsofmusic006900mbp