JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Audibilization: Data Analysis by Ear

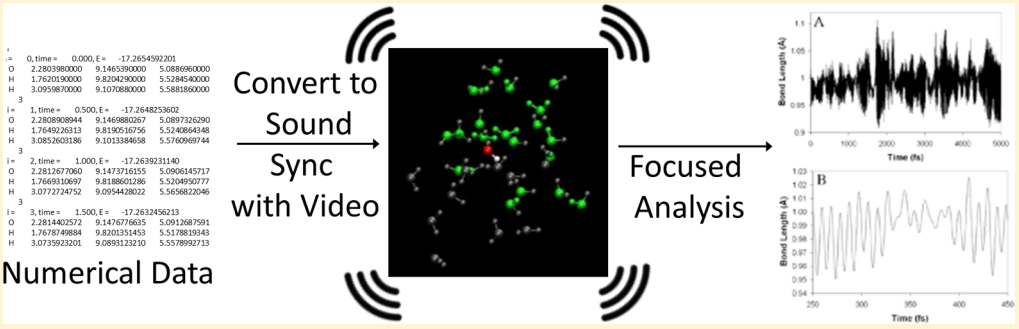
Andrew P. Jallouk[†] and Peter T. Cummings*[,‡,§]

[†]Medical Scientist Training Program, Washington University in St. Louis, St. Louis, Missouri 63110-1093, United States

[‡]Department of Chemical & Biomolecular Engineering, Vanderbilt University, Nashville, Tennesee 37235-1604, United States

[§]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6494, United States

Supporting Information



ABSTRACT: As molecular dynamics simulations continue to grow in size and complexity, new techniques are needed to rapidly identify regions of data likely to benefit from further analysis. Audibilization, the conversion of data to sound, facilitates this task by taking advantage of the user's innate ability to identify anomalies in patterns of sound. Audibilization also complements visualization of a molecular simulation by allowing the user to easily correlate changes in numerical quantities with changes in the overall structure of the molecular system. Here we present three examples highlighting the utility of audibilization in the analysis of three different molecular simulations. First, we present a simulation of liquid water in which the lengths of the O−H bonds are calculated at each time step and audibilized. Interestingly, we find that anomalies in the pattern of bond vibration are due to intermolecular interactions but do not correlate with the formation of hydrogen bonds. Next, we present a simulation of the rupture of a gold nanowire. Here we audibilize the nanowire potential energy and illustrate that sharp changes in this value coincide with important structural events such as the formation of monatomic chains and dislocations. Finally, we present a simulation of single-stranded DNA passing through a nanogap. Here the bond angle is audibilized and used to illustrate the conformational changes of each base as it passes through the nanogap. This simulation also illustrates the use of more advanced audibilization techniques such as the multiplexing of audibilized signals and the weighting of certain segments of data relative to others.

INTRODUCTION

With the rise of petascale computing and with exascale computing looming on the horizon, the ability to generate and compile enormous amounts of scientific data has greatly outpaced the ability to meaningfully analyze the results.[1−3] The need for more efficient data analysis techniques is especially evident in the field of molecular simulation, where both the size and complexity of simulated systems continue to grow at an exponential rate.[4−6] A typical protein simulation, for instance, may consist of the individual trajectories of thousands to millions of atoms as well as a number of quantities characterizing the system being modeled, such as entropy and free energy.[7,8] Furthermore, simply identifying the value of each of these quantities at every time step may not provide useful information regarding the process at hand. Rather, one must analyze the pattern in which these values change over time in order to truly gain insight into the phenomena being modeled.

Many physical phenomena are characterized by quantities that vary in a predictable, regular pattern. Anomalies in these patterns of variation, in turn, often represent interactions that are unusual or require further investigation. For instance, the length of a vibrating bond is expected to oscillate regularly as a function of time. Interactions with other molecules, however, have been shown to alter the frequency of this bond vibration and induce anomalies in the pattern of bond length variation.[9] By determining the times at which these anomalies occur, one may identify small sections of data representing intermolecular interactions that may be studied in more detail. Similar methodologies may be used to analyze phenomena that are not oscillatory in nature. During the process of protein folding, for example, the free energy of the system is expected to decrease steadily as the protein assumes a more stable conformation.[10] A sharp decrease in free energy, however, may represent the formation of a particularly strong stabilizing interaction that merits further study. In these types of simulations, the anomalies

Received: December 24, 2013
Published: March 4, 2014

   dx.doi.org/10.1021/ct401105x | J. Chem. Theory Comput. 2014, 10, 1387−1394

in patterns are often keys to discovering the presence of interesting phenomena. However, the location and characteristics of these anomalies may not be known a priori, making identification using classical signal processing techniques difficult. Data analysis tools that emphasize these anomalies and facilitate their discovery can thus play a critical role in the identification and characterization of interesting phenomena in large molecular simulations.

Most current data analysis techniques involve the visualization of large data sets through the use of graphs, charts, and advanced visualization software. While these techniques are useful in many situations, they suffer from some significant limitations. For instance, in the analysis of a graph of a certain quantity as a function of time, it becomes very difficult to relate changes in this quantity with corresponding events that take place in the molecular simulation. Conversely, when a molecular simulation is viewed, a rapid but significant change in a certain quantity may go unnoticed because of the large number of atoms and molecules present in the simulation. The ability to easily correlate large-scale simulation events with changes in physical parameters could lead to a deeper understanding of the events taking place in molecular simulations. In these situations, it is important to realize that the ear can be just as useful a tool as the eye in data analysis.

Even more so than the visual system, the human auditory system has evolved to identify patterns in input sounds and detect anomalies in these patterns. Whereas a visual signal passes through only a single subcortical structure before reaching the primary visual cortex, an auditory signal initially passes through a number of highly interconnected subcortical structures, all of which perform complex processing tasks on the auditory signal. Furthermore, the ratio of cortical neurons to input nerve fibers is approximately 6-fold larger for the primary auditory cortex than for the primary visual cortex.[11] The large number of subcortical structures and cortical neurons present in the auditory system allows humans to readily compare segments of an audible signal with previous segments of the same or a different signal.[11] In so doing, the auditory system is perfectly suited to detect patterns and anomalies in a sequence of input sounds even with no prior knowledge of the types of patterns and anomalies to be identified. From this perspective, the "audibilization" of large data sets could prove to be a useful complement to the visualization of these data sets. Audibilization involves the conversion of numerical values corresponding to a quantity of interest into a soundtrack that can then be played and analyzed. This technique represents one example of data sonification, a broad area of research that includes technical analyses of data processing and presentation as well as complex studies of human perception and understanding.[12] Data sonification strategies have been proposed for many physical phenomena, including experimentally obtained vibrational spectra and simulated molecular interactions.[13,14] In this article, we propose to use audibilization as a data analysis tool to facilitate the rapid interpretation of large simulation data sets and the identification of anomalies in patterns that merit further analysis. Additionally, when auditory and visual representations of a simulation are combined, the correlation between changes in numerical parameters and changes in the overall structure of a system may be analyzed in ways not possible using either technique alone. Here we present the results of audibilization of three different types of molecular simulation data and the ways in which audibilization of these data improved our understanding of the phenomena taking place in the simulation.

We first present a simulation of liquid water in which the lengths of the O−H bonds are calculated at each time step and audibilized. As described previously, anomalies in the pattern of bond vibration are expected to represent interactions between different water molecules in the simulation. Interestingly, we found that while these anomalies are caused by intermolecular interactions, they do not correspond to the formation of hydrogen bonds between molecules. This data set clearly illustrates that audibilization serves as an excellent tool for identifying anomalies in the pattern of variation for a specific quantity. Additionally, by combining the bond length soundtrack with an appropriately colored visual representation of the simulation, we illustrate that audibilization can serve as a useful complement to visualization that may be used to emphasize the value of a certain quantity as it changes over time.

The second data set we present involves a simulation of the rupture of a gold nanowire. In this simulation, we audibilize the nanowire potential energy in order to analyze the mechanisms of energy dissipation during nanowire rupture. Sharp changes in the potential energy of the nanowire are then correlated to events occurring during the simulation that are clearly visible in the visualization, such as the formation of monatomic chains and dislocations in the nanowire structure. In this way, we again illustrate how audibilization may be used to complement visualization. Furthermore, we illustrate the use of audibilization to analyze data sets characterized by nonoscillatory phenomena.

Finally, we present a simulation of single-stranded DNA passing through a nanogap. As the individual nucleobases enter the nanogap, they are forced into various conformations in order to fit within the nanogap. Since these conformations are dependent on the size and structure of each base, several groups have proposed that strands of DNA may be sequenced by measuring base-specific parameters, such as electrical conductance, as the strand traverses the nanopgap.[15−17] Here we utilize bond angle as an easily calculated measure of DNA deformation and demonstrate how audibilization of this parameter can be used to illustrate the conformational changes that occur as DNA passes through the nanogap. Furthermore, we illustrate higher-level audibilization techniques such as the multiplexing of audibilized signals and the weighting of certain segments of data relative to others.

METHODS

Audibilization Process. For each data set to be audibilized, a transfer function that maps the data values to MIDI pitches in an appropriate range was first intended. This mapping was designed to allow a rapid and intuitive interpretation of the audio data based on the context of the simulation. For instance, when the bond lengths of water molecules were analyzed, long bond lengths indicating stretched bonds were assigned to high pitch values. The raw data were then processed to generate a text file describing all of the notes to be included in the audibilization. This file used the following format:

Note On Frame	Note Off Frame	Instrument	Pitch	Volume
0	1	70	46	60
1	2	70	43	60
2	3	70	42	60
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

A detailed example of the steps involved in data processing is provided in the next section. A Java program was then written to generate a MIDI audio file based on this list of notes using the

Java Sound API. The TiMidity++ software synthesizer was then used to convert this MIDI file into a WAV file.[18] Special care was taken during this conversion to ensure that the resulting audio played at the same rate as the corresponding video file in order to simplify the synchronization process. Video files were generated using Visual Molecular Dynamics (VMD) and colored so as to facilitate interpretation of the audibilized data.[19] The Avidemux video editor was then used to synchronize and combine the audio and video files in order to generate the final audibilized product.[20] Figure 1 shows a schematic of this process. Although the audibilization pipeline described above utilizes software freely available online, a variety of other software packages could be used to achieve the same results, thereby providing a great deal of freedom in the application of this technique.

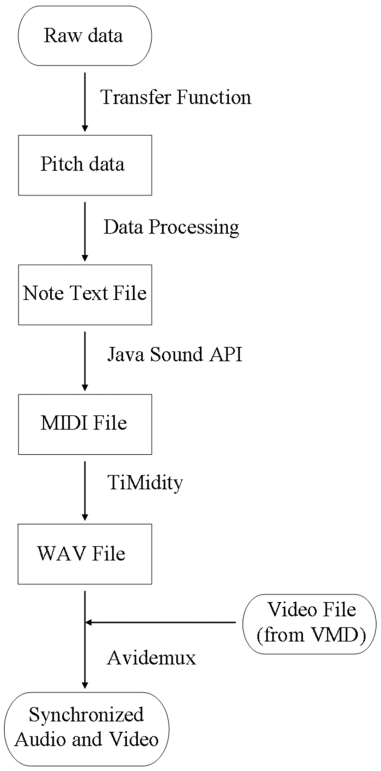


Figure 1. Schematic diagram of the audibilization process.

Data Processing Example. As part of our analysis of a liquid water simulation, we calculated and audibilized the length of each O−H bond over the course of the simulation. These bonds oscillated in length between 0.88 and 1.1 Å. In order to map these values linearly to MIDI pitches ranging from 12 (C0, 16.35 Hz) to 108 (C8, 4160 Hz), the following transfer function was used:

$$\text{pitchValue} = \text{round}\{(\text{bondLength} - 0.88) \times [96/(1.1 - 0.88)] + 12\}$$

We then iterated through these values in order to generate the text file describing all of the notes to be included in the audibilization. The pitch of the initial note corresponded to the length of the bond in the first frame of the simulation. Each time the pitch changed between two consecutive frames, the previous note was terminated and a new note with the appropriate pitch was started. For ease of listening, a series of frames with bond lengths corresponding to the same pitch was represented by a single extended note rather than a series of repeated notes. In this example, all of the notes were assigned to be played on a bassoon (instrument 70) with a volume of 60. The list of notes was then processed in a standard fashion as illustrated in Figure 1.

Liquid Water Simulation. The single water molecule and liquid water trajectories audibilized in this work were obtained from ab initio molecular dynamics simulations based on the Born−Oppenheimer approach performed by M. S. Yeom.[21] Technical aspects of the calculation are as follows: All of the calculations were performed in Quickstep[22] using the Gaussian and plane wave (GPW) method and its augmented extension, the Gaussian and augmented plane-wave (GAPW) method. The GAPW method using DZVP basis sets and an energy cutoff of 300 Ry was used. The liquid water system consisted of 32 water molecules contained in a cubic box of length 9.8528 Å, which corresponds to a density of 1 g cm^{-3}. The PBE exchange−correlation energy functional and the orbital transformation were employed. Periodic boundary conditions were applied in all directions, and time steps of 0.5 and 1.0 fs were used to solve the equations of motion. In order to identify anomalies in the pattern of water molecule bond vibration, the dynamics of single water molecules and groups of water molecules were analyzed over 200 000 time steps, with the lengths of all relevant O−H bonds being calculated at each time step. These lengths were then mapped linearly to MIDI pitches ranging from 12 (C0, 16.35 Hz) to 108 (C8, 4160 Hz). A MIDI file was generated for each bond to illustrate the changes in bond length occurring over the course of the simulation. Corresponding video files were generated using VMD and colored so as to highlight the bond whose length would be represented in each soundtrack. Videos containing multiple molecules were also colored to indicate the presence of hydrogen bonds between two water molecules. Hydrogen bonding was determined using VMD's geometrical definition of a hydrogen bond with an interoxygen cutoff distance of 3.3 Å and an angular cutoff of 20°. In this definition, a donor and acceptor molecule are considered to be hydrogen-bonded if they are within the cutoff distance from each other and the angle formed by the donor, hydrogen, and acceptor atoms is less than the angular cutoff from 180°. The MIDI files were then converted into WAV files and synchronized with the appropriate video files.

Gold Nanowire Simulation. Simulations of gold nanowire elongation were performed using a "stretch and relax" molecular dynamics method[23] in which two layers of rigid atoms at one end of the wire were displaced 0.1 Å in the direction of pulling every 10 ps. Two additional layers of rigid atoms were placed at the opposite end of the wire, while atoms within the core of the wire were dynamic. Each wire was elongated in vacuum at a temperature of 0.01 K. The interactions between gold atoms were calculated using the second-moment approximation to the tight-binding potential (TB-SMA), which is a semiempirical many-body potential designed to capture metallic bonding effects.[24] The total potential energy of each nanowire was recorded every 1 ps during elongation and then mapped linearly to MIDI pitches ranging from 12 (C0, 16.35 Hz) to 108 (C8, 4160 Hz). A MIDI file was generated for each nanowire to illustrate the changes in potential energy occurring over the course of the simulation. These MIDI files were then converted into WAV files and synchronized with a video of the appropriate nanowire simulation generated using VMD.

DNA Simulation. Simulations of single-stranded DNA (ssDNA) translocating through a nanoscale sequencing device were performed. The sequencing device consisted of two mica plates separated by approximately 3 nm. The solvated DNA and water

were enclosed between the plates, each measuring 20.7 nm × 14.4 nm. The detection nodes were constructed of two copper fcc rectangular nanoscale blocks measuring approximately 2 nm × 5 nm × 3 nm and separated by a 2.87 nm gap as measured from center to center of the surface copper atoms. The ss-DNA translocated through the gap between the copper nodes in response to an applied electric field ranging in magnitude from 5×10^{-4} to 5×10^{-1} V/Å. The details of the simulations, including the force fields used, are given by Payne et al.[25] The angle between each base and the DNA backbone was then mapped linearly to MIDI pitches ranging from 36 (C2, 65.4 Hz) to 72 (C5, 523 Hz). In order to emphasize the effect of the nanogap on these angles, the volumes of the pitches at each time step were scaled linearly using the distance between the 1′ carbons of the bases and the beginning or end of the nanogap. For instance, the volume corresponding to each base that had not yet entered the nanogap was assigned a value between 20 and 110 on the basis of the proximity of the base to the entrance of the nanogap. Bases currently within the nanogap were always assigned the maximum volume value of 110. Additionally, each base in the simulation was represented by a different instrument. As a result, MIDI files representing different bases could be combined into a single WAV file and used to track the movements of multiple bases during the simulation. As before, these WAV files were synchronized with videos generated using VMD and colored so as to highlight the bases represented in the soundtrack.

RESULTS AND DISCUSSION

Liquid Water Simulation. Liquid water is one of the most important substances on Earth and has been the subject of numerous molecular dynamics simulations over the last few decades.[26−28] Simulations involving liquid water are rendered especially complex by the strong electrostatic and hydrogen-bonding interactions that take place between individual water molecules. Indeed, a great deal of effort has been dedicated to characterizing and modeling these intermolecular interactions.[29−31] With many computational techniques now available for simulating liquid water, we have the ability to generate large amounts of data regarding this important system.[32,33] Because of the transient nature of many intermolecular interactions, however, new methodologies are needed to help sort through these data and identify regions most likely to represent interesting interactions. These interactions often take the form of disturbances in the normal pattern of bond vibration. Here we demonstrate how audibilization of bond lengths can serve as a useful tool for processing large amounts of simulation data and identifying anomalies in patterns that merit further analysis.

One technique to identify the interactions between water molecules is to analyze the pattern of bond vibration in these molecules. In an isolated water molecule, both O−H bonds are expected to behave as near-harmonic oscillators with lengths that vary in a regular, predictable pattern. Our simulations of individual water molecules support this hypothesis. In our initial simulation of a single water molecule (energy cutoff 300 Ry, time step 0.5 fs), the lengths of both bonds were found to oscillate between 0.918 and 1.020 Å. Audibilization of these bond lengths revealed the oscillation to be quite regular without any significant anomalies (single05_bond0.mp3 and the associated video file in the Supporting Information). Similar results were obtained for single-molecule simulations of water using energy cutoffs of 600 Ry, time steps of 1.0 fs, and more restrictive convergence criteria (single10_bond0.mp3, singledo05_bond0.mp3, and the associated video files in the Supporting Information). Single-molecule simulations of water with the oxygen atom fixed at a particular position also exhibited a regular pattern of bond length oscillation (singlefr05_bond0.mp3 and the associated video file in the Supporting Information). In the presence of other water molecules, however, intermolecular interactions were found to disrupt this pattern. Audibilization of bond lengths in a 32-molecule simulation of liquid water revealed numerous anomalies in the pattern of bond length variation. For instance, starting at approximately 0:33 in the audibilization of bond 43 (water3205_bond43.mp3 in the Supporting Information), the soundtrack abruptly shifts from a regular oscillatory pattern to one characterized by extended tones and erratic pitch changes. The soundtrack reverts to the original pattern after approximately 6 s. Similar anomalies were found to occur from 1:49 to 2:27 in the audibilization of bond 32 (water3205_bond32.mp3 in the Supporting Information), from 4:15 to 4:21 in the audibilization of bond 47 (water3205_bond47.mp3 in the Supporting Information), from 7:51 to 7:55 in the audibilization of bond 51 (water3205_bond51.mp3 in the Supporting Information), and in the audibilizations of many other bonds.

In an attempt to determine the cause of these irregularities, we compared the timing of these anomalies with the occurrence of hydrogen bonds between water molecules. To this end, the videos representing each simulation were colored in order to highlight the presence of hydrogen bonds at each time step (Figure 2). The formation of these hydrogen bonds was then compared with the irregularities in the audibilized signal to determine whether there was any correlation between these events. All of these videos clearly demonstrated that there was no such correlation (water3205_bond32.mp4, water3205_bond43.mp4, water3205_bond47.mp4, and water3205_bond51.mp4 in the Supporting Information). As a result, although it appears that these irregularities are due to some

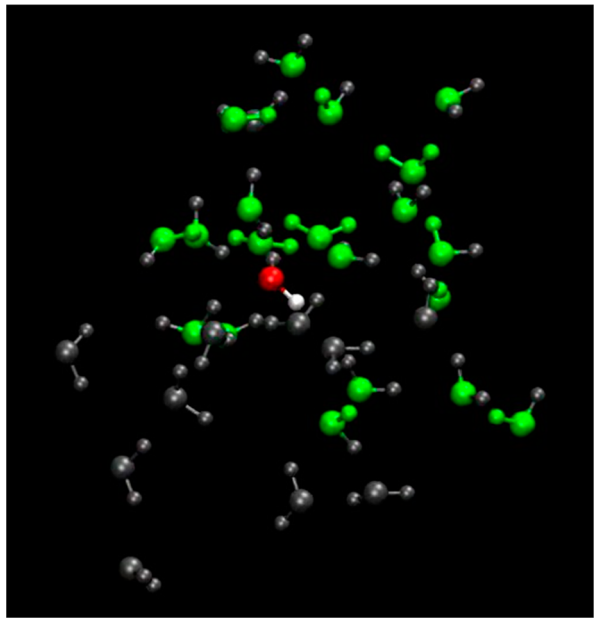


Figure 2. Snapshot from a 32-molecule simulation of liquid water modified to illustrate hydrogen bonding. Hydrogen-bonded atoms are shown in green, while atoms involved in the audibilized bond are shown in red and white. All of the other atoms are shown in gray. This setup allows the user to identify relationships between the formation of hydrogen bonds and anomalies in the audibilized soundtrack.

sort of intermolecular interaction, they do not coincide with the formation of hydrogen bonds. A deeper analysis is thus necessary to understand the nature of these anomalies, which promise to serve as interesting topics for future research.

This water simulation illustrates many of the advantages that audibilization provides in the analysis of large amounts of data. Most importantly, audibilization greatly simplifies the process of finding irregularities in a data set, in large part because of the exceptional ability of the ear to detect anomalies in patterns. The importance of this characteristic is most clearly illustrated by attempting to find the same anomalies in these data sets without audibilization. Figure 3A shows the length of bond 43 over the course of the simulation. Clearly, the noise in the signal and the large amount of data present make it difficult to quickly identify regions representing significant anomalies. Similarly, the Fourier transform of the length of bond 43, shown in Figure 3B, demonstrates a noisy peak at 0.1 fs^{-1} corresponding to the main stretching frequency of the bond but does not provide any obvious information regarding the nature or location of individual vibrational anomalies. This analysis is greatly simplified by the use of audibilization, as illustrated in Figure 3C.

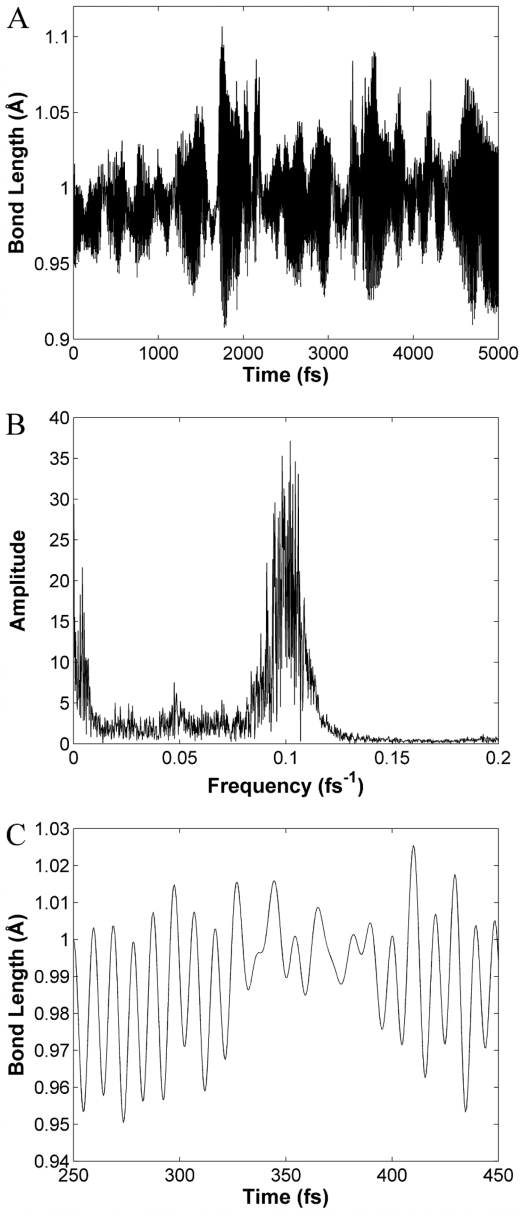


Figure 3. Identification of vibrational anomalies in bond 43 of a 32-molecule simulation of liquid water. (A) Length of bond 43 over the course of the entire simulation. (B) Fourier transform of the length of bond 43. (C) Length of bond 43 during a period determined by audibilization to contain an anomaly. The clearly anomalous pattern of bond vibration between 330 and 390 fs is evident. This anomaly corresponds to extended tones and erratic pitch changes occurring from 0:33 to 0:39 in the audibilized soundtrack.

Although audibilization alone may be useful for identifying anomalies in certain patterns, the true strength of audibilization lies in its combination with other data analysis tools, such as visualization and signal processing algorithms. To take advantage of both the auditory system and the visual system of the user, audibilization and visualization may be combined, allowing simultaneous analysis and comparison of multiple data sets that would not be possible using either technique alone. For instance, when analyzing the importance of hydrogen bonding in the water simulation, we were able to see the movements of the molecules, hear changes in the length of the bonds, and observe the formation of hydrogen bonds all at once. This type of analysis would not have been possible without the use of audibilization. Additionally, simulations containing large numbers of molecules may benefit from the combination of audibilization with more automated signal processing techniques. For instance, in a simulation containing thousands of water molecules, audibilization of a few select bonds may be performed initially to identify vibrational abnormalities of interest. These anomalies may then be characterized mathematically and detected in the remaining bonds using classical signal processing techniques. This approach combines the sensitivity of audibilization with the efficiency of mathematical signal analysis to identify and characterize pattern anomalies in a way that would be difficult using either technique alone.

While audibilization has been used to identify anomalies in oscillating signals up to this point, the technique works well in the analysis of nonoscillatory signals as well. Simulations of the rupture of a gold nanowire illustrate this additional application of audibilization and further explore the relationship between an audibilized signal and the video that it represents.

Gold Nanowire Simulation. The study of gold nanowire behavior under mechanical stress has implications both in fundamental physics and in applications such as molecular electronics.[34−36] In particular, the rupture of gold nanowires and the mechanisms by which energy is released during this process have drawn considerable interest.[23,37] Nanowire rupture is often characterized by the formation of monatomic chains and dislocations that are associated with distinct changes in the potential energy of the system. Here we demonstrate how audibilization of the nanowire potential energy can complement visualization of the rupture process in identifying important events leading to dissipation of energy during nanowire rupture. In order to monitor energy release following nanowire rupture, we audibilized the potential energy corresponding to three gold nanowire simulations in which monatomic chains of different lengths were formed during rupture. These included a 256-atom simulation with a monatomic chain consisting of two atoms, a 256-atom simulation with a monatomic chain consisting of five atoms, and a 576-atom simulation with a monatomic chain consisting of three atoms (Figure 4; also see: PE_256atoms_run2.mp3, PE_256atoms_run12.mp3, PE_576atoms_run1.mp3, and the associated video files in the Supporting Information).

dx.doi.org/10.1021/ct401105x | J. Chem. Theory Comput. 2014, 10, 1387−1394

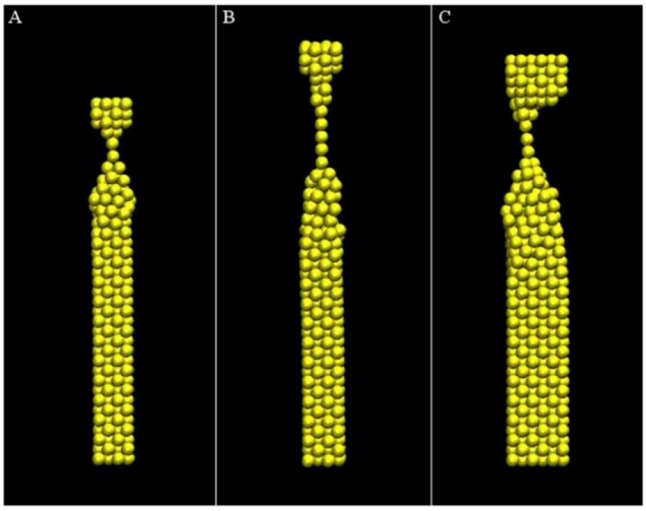


Figure 4. Gold nanowires in which monatomic chains of different lengths are formed during rupture. (A) A 256-atom simulation with a monatomic chain consisting of two atoms. (B) A 256-atom simulation with a monatomic chain consisting of five atoms. (C) A 576-atom simulation with a monatomic chain consisting of three atoms.

As expected, the initial stretching of the nanowire in all three simulations resulted in a sharp increase in potential energy. Afterward, the movement of atoms involved in narrowing of the nanowire (i.e., necking) resulted in a marked decrease in potential energy. The pattern of energy release following this narrowing depended heavily on the formation of monatomic chains. In general, the growth of a monatomic chain was characterized by a slight decrease in potential energy corresponding to the addition of each extra atom to the chain. However, this drop was followed by an increase in potential energy resulting from the continued stretching of the nanowire. Thus, with short monatomic chains, rupture occurred soon after narrowing of the nanowire, accompanied by sharp drops in potential energy corresponding to the visible collapse of the nanowire and the formation of dislocations. Longer monatomic chains, in contrast, caused the potential energy of the nanowire to remain high for a longer period of time. The eventual rupture of these chains was also accompanied by sharp drops in potential energy due to collapse of the monatomic segment and the formation of dislocations within the nanowire. These mechanisms of energy buildup and release are shown graphically in Figure 5. Whereas the water simulation illustrates the use of audibilization to detect anomalies in a repeating signal, the gold nanowire simulation demonstrates that audibilization may be just as useful when analyzing nonoscillatory signals. Indeed, the sensitivity of the ear to changes in patterns allows it to identify even slightly anomalous shifts in a signal. This ability is especially useful when attempting to detect events that are known to influence a quantity characteristic of the simulated system, such as potential energy. Furthermore, the ability to simultaneously analyze both audio and visual representations of a simulation greatly augments the utility of this detection strategy. By synchronizing the soundtrack representing a certain quantity to the video representing the simulation, one may easily correlate changes in the signal to physical events that take place during the simulation. These events, such as the formation of dislocations, are instantaneously recognizable to the average viewer but may be difficult to identify and characterize in a purely mathematical analysis. In this way, the combination of audibilization and

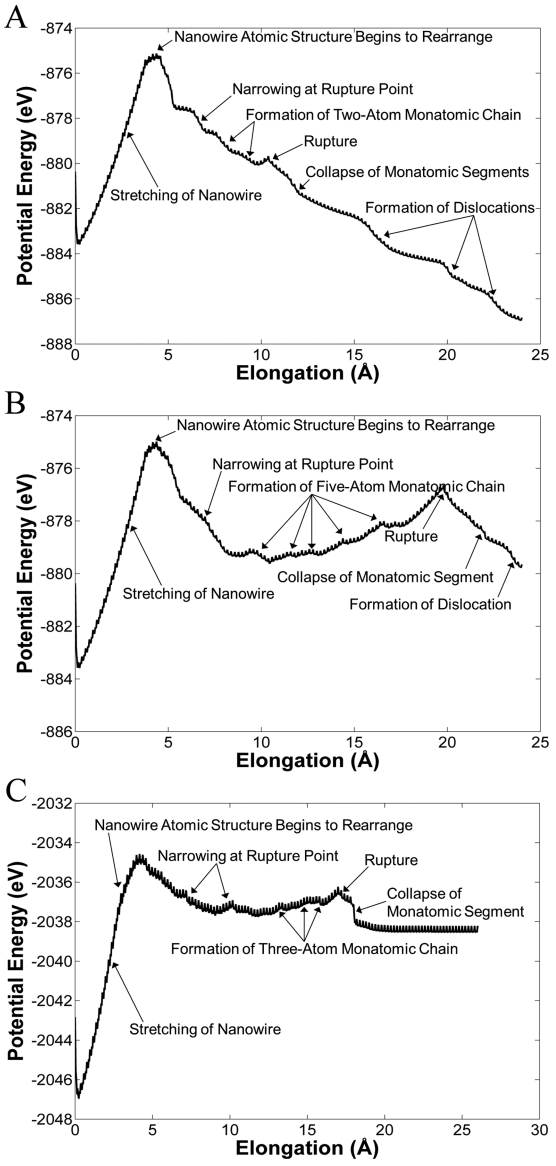


Figure 5. Potential energy of gold nanowires in which monatomic chains of different lengths are formed during rupture. (A) A 256-atom simulation with a monatomic chain consisting of two atoms. (B) A 256-atom simulation with a monatomic chain consisting of five atoms. (C) A 576-atom simulation with a monatomic chain consisting of three atoms. Labels indicating the mechanisms of energy buildup and release were obtained by analyzing video of the simulation synchronized with the audibilized potential energy soundtrack.

visualization takes advantage of the synchronization and unique pattern recognition properties of the ear and eye to facilitate a deeper understanding of the simulation at hand.

Up to this point, we have only discussed audibilizations of a single signal whose value is characterized by pitch. Using other aspects of the MIDI sound creation system, however, we can generate more complex soundtracks that rely on different instruments and volume modulation to illustrate various aspects of the audibilized signal.

DNA Simulation. Several groups have proposed that the controlled passage of ss-DNA through a nanogap could be used as a rapid, cost-effective way of physically sequencing DNA.[15−17] In this paradigm, conformational changes in the DNA as it passes through the nanogap would alter parameters such as electrical

dx.doi.org/10.1021/ct401105x | J. Chem. Theory Comput. 2014, 10, 1387−1394

conductance in a base-specific manner. Measurements of these parameters over time could then be analyzed to determine the sequence of the DNA strand. This system also presents an opportunity to demonstrate some of the more advanced features of audibilization, such as signal weighting and multiplexing, which facilitate the simultaneous analysis of multiple data sets. Here we use the angle between each base and the DNA backbone as a measure of the base's conformational status. We then audibilize and combine the signals from all of the bases to illustrate the changes in conformation that take place as the DNA strand passes through the nanogap.

In order to demonstrate the effectiveness of audibilization in illustrating DNA conformational changes, we utilized a simulation of single-stranded poly(cytosine) DNA traversing a nanogap separating gold and platinum electrodes. Each base in this simulation was represented by a different instrument, and the volume of each instrument in turn was scaled by the distance between the corresponding base and the nanogap. In this way, the angles of bases passing through the nanogap at each time point were emphasized. The resulting soundtracks indicate that angles corresponding to all of the bases experience a shift as the bases pass through the nanogap (dnaSound1_Bassoon.mp3, dnaSound10_Trumpet.mp3, dnaSound17_Piano.mp3, and dnaSoundBases1_10_17.mp3 in the Supporting Information). Synchronization of these soundtracks with videos of the simulation clearly demonstrates that these shifts in angle are due to tilting of the bases as they pass through the nanogap (Figure 6; also see dnaBases1_10_17.mp4 and dnaBases1_10_17_ReducedWall.mp4 in the Supporting Information).

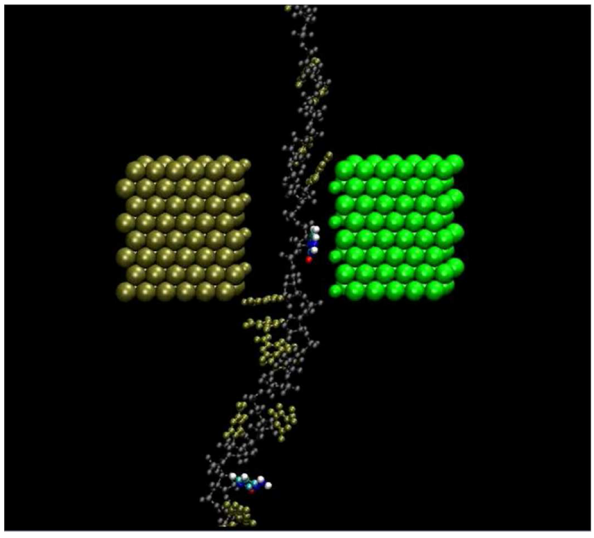


Figure 6. Simulation of single-stranded poly(cytosine) DNA traversing a nanogap. Bases 10 and 17 are shown in red, white, and blue, while all of the other bases are shown in gold. The DNA backbone is shown in gray. It can be clearly seen that bases currently inside the nanogap are tilted relative to the DNA backbone.

This simulation illustrates several advanced audibilization techniques that facilitate the analysis of multiple large data sets. Chief among these techniques is the use of different instruments to create multiplexed soundtracks and allow for the simultaneous analysis of multiple related signals. In addition to recognizing irregularities in a pattern of sounds, the human ear is adept at distinguishing between sounds of different timbre. Thus, simultaneously playing multiple soundtracks synchronized to video permits the ready comparison of these signals both with events occurring during the simulation and with each other. This method of signal multiplexing may be augmented through the use of volume modulation to emphasize one or more of the audibilized signals at different times during the simulation. Furthermore, the degree to which each signal is emphasized may be computed at each time step on the basis of conditions that occur during the simulation or input from the user, providing yet another degree of freedom in the creation and manipulation of audibilized signals. The techniques utilized in this simulation provide just a few examples of how the process of audibilization may be tailored to facilitate the analysis of specific sets of data. Many other modifications are possible on the basis of the nature of the data sets available and the analyses that must be performed. For instance, nonlinear transfer functions may be used when converting raw data to MIDI notes in order to take advantage of the nonlinear response properties of the human ear.[12] These transfer functions could also provide another way of emphasizing certain segments of data and would serve as interesting topics for future research.

CONCLUSIONS

The conversion of data into sound serves as a useful tool for data analysis, especially when utilized in conjunction with data visualization techniques. Because of the ear's exceptional ability to recognize patterns and anomalies in patterns, audibilization can be used to identify irregularities in both oscillating and nonoscillating signals. These irregularities may then be correlated to physical events occurring during the simulation, which in turn may serve as topics for future research. Furthermore, the use of more elaborate audibilization techniques, such as signal multiplexing and volume modulation, provides even greater opportunities for the analysis of complex signals.

ASSOCIATED CONTENT

Supporting Information

Zip files containing all of the sound and video files referenced in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: peter.cummings@vanderbilt.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Min Sun Yeom, Christina Payne, Will French, Chris Iacovella, and Amulya Pervaje for providing datasets and helpful discussions.

REFERENCES

(1) LaBaer, J. Mining the Literature and Large Datasets. *Nat. Biotechnol.* **2003**, *21*, 976−977.

(2) Chervenak, A.; Foster, I.; Kesselman, C.; Salisbury, C.; Tuecke, S. The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. *J. Network Comput. Appl.* **2000**, *23*, 187−200.

(3) Beynon, M. D.; Kurc, T.; Catalyurek, U.; Chang, C.; Sussman, A.; Saltz, J. Distributed Processing of Very Large Datasets with DataCutter. *Parallel Comput.* **2001**, *27*, 1457−1478.

(4) Zagrovic, B.; Sorin, E. J.; Pande, V. B-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model. *J. Mol. Biol.* **2001**, *313*, 151−169.
(5) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Method for Large-Scale Modeling of Chemistry in Protein Environments. *J. Comput. Chem.* **2000**, *21*, 1442−1457.
(6) Bowman, G. R.; Ensign, D. L.; Pande, V. S. Enhanced Modeling via Network Theory: Adaptive Sampling of Markov State Models. *J. Chem. Theory Comput.* **2010**, *6*, 787−794.
(7) Boczko, E. M.; Brooks, C. L. First-Principles Calculation of the Folding Free Energy of a Three-Helix Bundle Protein. *Science* **1995**, *269*, 393−396.
(8) Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289−6292.
(9) Scheiner, S.; Kar, T. Red- versus Blue-Shifting Hydrogen Bonds: Are There Fundamental Distinctions? *J. Phys. Chem. A* **2002**, *106*, 1784−1789.
(10) Duan, Y.; Kollman, P. A. Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science* **1998**, *282*, 740−744.
(11) Worden, F. G. Hearing and the Neural Detection of Acoustic Patterns. *Behav. Sci.* **1971**, *16*, 20−30.
(12) Kramer, G.; Walker, B.; Bonebright, T.; Cook, P.; Flowers, J.; Miner, N.; Neuhoff, J. *Sonification Report: Status of the Field and Research Agenda*, 1997; available at http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1443&context=psychfacpub.
(13) Delatour, T. Molecular Songs. In *Molecular Aesthetics*; Weibel, P., Fruk, L., Eds.; MIT Press: Cambridge, MA, 2013; pp 293−311.
(14) Glowacki, D. R.; Tew, P.; Mitchell, T.; Kriefman, L.; Hyde, J.; Malcolm, L. J.; Price, J.; McIntosh-Smith, S. Sculpting Molecular Dynamics in Real-Time Using Human Energy Fields. In *Molecular Aesthetics*; Weibel, P., Fruk, L., Eds.; MIT Press: Cambridge, MA, 2013; pp 248−257.
(15) Liang, X.; Chou, S. Y. Nanogap Detector Inside Nanofluidic Channel for Fast Real-Time Label-Free DNA Analysis. *Nano Lett.* **2008**, *8*, 1472−1476.
(16) Postma, H. W. C. Rapid Sequencing of Individual DNA Molecules in Graphene Nanogaps. *Nano Lett.* **2010**, *10*, 420−425.
(17) Zhao, X.; Payne, C. M.; Cummings, P. T. Controlled Translocation of DNA Segments through Nanoelectrode Gaps from Molecular Dynamics. *J. Phys. Chem. C* **2008**, *112*, 8−12.
(18) Izumo, M. *TiMidity++*, version 2.13.0; SourceForge: Mountain View, CA, 2004.
(19) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(20) Mean; Gruntster; Fahr. *Avidemux*, version 2.4.1; SourceForge: Mountain View, CA, 2008.
(21) Yeom, M. S. Unpublished results.
(22) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.
(23) Pu, Q.; Leng, Y.; Cummings, P. T. Rate-Dependent Energy Release Mechanism of Gold Nanowires under Elongation. *J. Am. Chem. Soc.* **2008**, *130*, 17907−17912.
(24) Cleri, F.; Rosato, V. Tight-Binding Potentials for Transition Metals and Alloys. *Phys. Rev. B* **1993**, *48*, 22−33.
(25) Payne, C. M.; Zhao, X.; Vlcek, L.; Cummings, P. T. Molecular Dynamics Simulation of ss-DNA Translocation between Copper Nanoelectrodes Incorporating Electrode Charge Dynamics. *J. Phys. Chem. B* **2008**, *112*, 1712−1717.
(26) Kumar, R.; Skinner, J. L. Water Simulation Model with Explicit Three-Molecule Interactions. *J. Phys. Chem. B* **2008**, *112*, 8311−8318.
(27) Rahman, A.; Stillinger, F. H. Molecular Dynamics Study of Liquid Water. *J. Chem. Phys.* **1971**, *55*, 3336−3359.
(28) Sedlmeier, F.; Horinek, D.; Netz, R. R. Spatial Correlations of Density and Structural Fluctuations in Liquid Water: A Comparative Simulation Study. *J. Am. Chem. Soc.* **2011**, *133*, 1391−1398.
(29) Eaves, J. D.; Loparo, J. J.; Fecko, C. J.; Roberts, S. T.; Tokmakoff, A.; Geissler, P. L. Hydrogen Bonds in Liquid Water Are Broken Only Fleetingly. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13019−13022.
(30) Lin, I.-C.; Seitsonen, A. P.; Coutinho-Neto, M. D.; Tavernelli, I.; Rothlisberger, U. Importance of van der Waals Interactions in Liquid Water. *J. Phys. Chem. B* **2009**, *113*, 1127−1131.
(31) Smith, J. D.; Cappa, C. D.; Wilson, K. R.; Messer, B. M.; Cohen, R. C.; Saykally, R. J. Energetics of Hydrogen Bond Network Rearrangements in Liquid Water. *Science* **2004**, *306*, 851−853.
(32) Kühne, T. D.; Krack, M.; Parrinello, M. Static and Dynamical Properties of Liquid Water from First Principles by a Novel Car−Parrinello-like Approach. *J. Chem. Theory Comput.* **2009**, *5*, 235−241.
(33) Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. Molecular Dynamics Simulation of Liquid Water: Hybrid Density Functionals. *J. Phys. Chem. B* **2006**, *110*, 3685−3691.
(34) Coura, P. Z.; Legoas, S. B.; Moreira, A. S.; Sato, F.; Rodrigues, V.; Dantas, S. O.; Ugarte, D.; Galvão, D. S. On the Structural and Stability Features of Linear Atomic Suspended Chains Formed from Gold Nanowires Stretching. *Nano Lett.* **2004**, *4*, 1187−1191.
(35) Da Silva, E. Z.; Novaes, F. D.; da Silva, A. J. R.; Fazzio, A. Theoretical Study of the Formation, Evolution, and Breaking of Gold Nanowires. *Phys. Rev. B* **2004**, *69*, No. 115411.
(36) Tosatti, E.; Prestipino, S.; Kostlmeier, S.; Corso, A. D.; Tolla, F. D. D. String Tension and Stability of Magic Tip-Suspended Nanowires. *Science* **2001**, *291*, 288−290.
(37) Da Silva, E. Z.; da Silva, A. J. R.; Fazzio, A. How Do Gold Nanowires Break? *Phys. Rev. Lett.* **2001**, *87*, No. 256102.