Fall 9-19-2019

# A method for automatic detection of tongued and slurred note transitions in clarinet playing

Whitney L. Coyle
*Rollins College*, wcoyle@rollins.edu

Jack D. Gabriel
*Rollins College*, jgabriel@rollins.edu



## Published In

Whitney L. Coyle, and Jack D. Gabriel. "A method for automatic detection of tongued and slurred note transitions in clarinet playing." The Journal of the Acoustical Society of America 146, EL238 (2019): DOI 10.1121/1.5126025.

# A method for automatic detection of tongued and slurred note transitions in clarinet playing

**Whitney L. Coyle[a] and Jack D. Gabriel**
*Physics Department, Rollins College, 1000 Holt Avenue, Winter Park, Florida 32789, USA*
*wcoyle@rollins.edu, jgabriel@rollins.edu*

**Abstract:** This study offers a simple method to characterize two transition types in passages of music in order to automatically distinguish slurred transitions from tongued transitions in musical settings. Data were recorded from musicians playing a clarinet with a sensor-equipped mouthpiece measuring blowing pressure in the mouth and pressure in the mouthpiece. This method allows for comparing transitions in different musical contexts, playing regimes, and between players. The method is highly reliable in automatically detecting transition types in recorded clarinet playing in both simple and more complex passages.



[DMC]
**Date Received:** July 18, 2019 **Date Accepted:** August 27, 2019

## 1. Introduction and background

Musical transitions are note-to-note separations that can allow for identification and comparison of musical pieces, distinction between players, and classification of transitions—all of which can inform methods for compression, delivery, indexing, and retrieval.[1] In the past, transitions have been studied by considering the concepts such as onset, attack, or transient region as described by Bello *et al.*[1] While many studies are interested in the analysis of external sound pressure, this work focuses on a different method of transition detection that utilizes the internal pressure signal, inside the clarinet mouthpiece, as is the standard in clarinet acoustics research.[2–5]

As a way to better understand note-to-note transitions, musical acoustics research over the past decade has focused on studying transients.[2–6] The transient portion of a musical tone is what aids in the recognition of instrument timbre.[7] The length of the initial transient of a particular instrument varies from musician to musician and is dependent on musical context.[3] Most research on clarinet transients has focused on separated tones both with artificial playing machines[4,5] and musician data.[3,6] The selections being performed are not generally excerpts of music. This means that, unfortunately, most transient studies are being done in non-musical contexts resulting in analysis and conclusions which lack applicability for musicians.

During performance, with written rests aside, the majority of musical tones in a passage are connected in some way—slurred (connected) or articulated tones (with a variety of methods for tongue or breath attacks as described by Li *et al.*[3]). The most current methods used to analyze measured clarinet transients come from Bergeot *et al.*[4] and Li *et al.*[3,6] Both of these studies used either an artificial mouth with a well-regulated input pressure ramp, or a mouthpiece equipped with sensors to measure the blowing pressure signals as well as the oscillating pressure inside the mouthpiece or barrel when played. However, both studies focused on individual, isolated tones. To study the transient portion of a pressure signal, Bergeot *et al.* analyzed the difference between specific peaks and valleys of the second time derivative of the root-mean-square (RMS) mouthpiece pressure ($\ddot{P}_{\mathrm{RMS}}$) to signify the beginning of the transient. However, this method still incorporated some means of manual delimitation to mark the beginning and end of a transient. Other authors analyze the difference in time between 10% and 90% of the steady state pressure, and sometimes 20% and 80%, or they rely on the portion of the initial signal that is increasing exponentially until saturation.[6] With these techniques there is still a high degree of manual delimitation in marking the start and stop times of this transient portion. While the transient of a note is of interest to musical acousticians, for signals where there is not a decrease of blowing pressure back to zero, during a slurred transition, for example, the transient is not well defined. On the other hand, the portion of the signal between notes, called throughout this paper the transition, varies between players and articulation styles and

[a]Author to whom correspondence should be addressed.

 

is of great interest. The transient and transition are related, but very different. This manuscript will focus on the transitions from note-to-note in clarinet playing based on two general articulation types. To the authors' knowledge, this is the first study on note-to-note transitions in a selection of music.

The current work provides a method for automatic detection of musician articulation as an objective means to mark transition beginnings and ends. After visual inspection of players' mouthpiece pressure signals, the beginning and ends of transitions were not visibly different enough for successful automatic detection of transition type. However, while comparing to transient detection methods and calculating the shape of the second derivative of the RMS mouthpiece pressure as described by Bergeot *et al.*,[4] the shape of this curve was noticeably different between articulation types such as tongued notes and slurred notes. Using this mathematical tool as an analysis technique succeeds in revealing these otherwise subtle features in the pressure signals. The resulting characteristic shape of the second derivative can be analyzed by an algorithm which searches for thresholds of signal peaks compared to surrounding peaks and can therefore determine the articulation type automatically. This detection algorithm is suitable for passages of music for which the articulation is known but also where it was perhaps unclear—such as a long passage of music with many varying transition types. While this method is not suitable to detect or measure the onset, nor transient portion (as defined in literature), it succeeds in providing an objective means of marking the beginning and end of a transition between notes.

Section 2 will describe the procedure for tests performed in this study and detail the analysis technique and algorithm used to differentiate between two articulation styles. Section 3 will then detail the results of the tests after algorithm application. Finally, we offer an example of automatic detection applied to a performance of a portion of the Mozart clarinet concerto by various players of different ability levels.

## 2. Experimental methodology

To measure playing parameters, a sensor-equipped mouthpiece (SEM) similar to that used by Li *et al.*[3,6] for the clarinet and Munoz *et al.*[8] for the saxophone was used. The SEM was built using a Vandoren M30 mouthpiece and played on a Buffet (Mantes-la-Ville, France) R13 Festival B♭ soprano clarinet for each of the tests. All signals were read to the computer via an acquisition box (NI-6212, Austin, TX) at a sampling frequency of 40 kHz. A Vandoren reed (strength 1.5) was used and a Vandoren (Paris, France) M30 mouthpiece was fitted with two Honeywell (Charlotte, NC) SCX05DNC microstructure pressure sensors that were connected with small tubes to the mouthpiece measuring $P_b$ (blowing pressure) and $P_{mp}$ (pressure inside of the mouthpiece). Given that a softer reed was used, the blowing pressures and thresholds of oscillation reported throughout this study will be lower than reported in other literature.

Six musicians participated in this series of measurements and were allowed a 5 min training period with the SEM before tests began. The musicians ranged in experience from novice (with 10 years of sporadic playing and no formal training) to professional (a university clarinet professor). The musicians were placed in order of their self-reported experience level and will be referred to as musicians A–F, with musician A having the most experience. There is a large gap in performance ability between players A thru F; however, all tasks were performed by all six musicians as requested.

In this set of tests, five full measurements sets were taken with eight individual iterations of each. A summary of these tests can be found in Table 1. Test 1 required the musician to play concert D3 to B♭3 chromatically, about 1 s per note at an *mf* dynamic, with the notes slurred from one to the next. Test 2 involved the same instructions and content but the notes were tongued. In Test 3, the musician would play the same fingerings and articulation style as in Test 1 but in the second register (same fingerings, apply the register key). Test 4 then required the musician to play the same fingerings and

Table 1. Playing tests performed by musician. Test number, description of tests, and articulation type detailed. All note names refer to concert pitch.

| Test Number | Description | Articulation Type |
|---|---|---|
| 1 | Low 1st register: D3 - B♭3 chromatic (9 notes) | Slur |
| 2 | Low 1st register: D3 - B♭3 chromatic (9 notes) | Tongue |
| 3 | Low 2nd register: A4 - F5 chromatic (9 notes) | Slur |
| 4 | Low 2nd register: A4 - F5 chromatic (9 notes) | Tongue |
| 5 | Mozart Adagio Selection | As written |

articulation style as in Test 2 but in the second register (same fingerings, apply the register key). Test 5 completed the series and required the musician to play a well-known, short passage of the Mozart clarinet concerto that included several tongued and slurred transitions, and was transcribed in both the first and second registers of the clarinet.

Figure 1 contains plots of typical data from this study, the transition between concert E3 and F3 played (from Tests 1 and 2). The left column contains three plots, all analysis from a slurred transition (Test 1) and the right column contains three plots, all analysis from a tongued transition (Test 2). The top row, for each transition type, shows three signals: (1) the blowing pressure, $P_b$ (dashed black line), the mouthpiece pressure, $P_{mp}$ (dotted black line), and the RMS pressure, $P_{\mathrm{RMS}}$ (thick gray line). The middle row of Fig. 1 shows plots of $P_{\mathrm{RMS}}$ (thick gray line) which is the same as in the top row, and $\dot{P}_{\mathrm{RMS}}$ (thin black line), the first derivative of the RMS pressure signal. Finally, the bottom row shows $P_{\mathrm{RMS}}$ (thick gray line) again, and $\ddot{P}_{\mathrm{RMS}}$ (black line), the second derivative of the RMS pressure signal. The first local maximum of $\ddot{P}_{\mathrm{RMS}}$ is marked with a vertical dashed-dotted black line for both Tests 1 and 2 (in the bottom row). The values on the $y$ axis refer to the $P_{\mathrm{RMS}}$ curve only.

Throughout this study, the shape of the $\ddot{P}_{\mathrm{RMS}}$ curve was used to classify the transitions as slurred or tongued. While all transitions exhibit a drop in an otherwise stable $P_{\mathrm{RMS}}$, in the case of a slurred transition this drop is short in duration and of small amplitude before the rise back to a stable pressure for the next note. The mechanism of tonguing causes a sharp drop followed by a fairly stable, nearly-zero RMS pressure for a significant time interval before the reemergence of sound. These three separate stages, characteristic of a tongued transition, are reliably identified by finding two successive and distinct positive extrema in $\ddot{P}_{\mathrm{RMS}}$, a few milliseconds apart, which reflect the curvature change at the near extinction and re-emergence of the sound. A count of the maxima found in $\ddot{P}_{\mathrm{RMS}}$ during a transition was used as a way to classify it as slurred or tongued since the trend of one maximum for a slurred transition and two maxima for a tongued transition was present in nearly all cases (further discussed in Sec. 3).

With this in mind, the automatic detection algorithm, written in MATLAB, begins by searching $\ddot{P}_{\mathrm{RMS}}$ for all potential transitions, and then classifies each as a tongue, slur, or not a transition according to the number of maxima detected. In order for a segment to be classified first as a potential transition, a local maximum of $\ddot{P}_{\mathrm{RMS}}$ must be larger than a particular threshold (somewhat arbitrary but specific for each measurement), for example, 12% of $P_{\mathrm{RMS}}$ of the smaller amplitude of the two surrounding notes. The idea behind this threshold value is that it will not be constant for the whole measurement as the extent to which the pressure will drop during the transition will vary from one note to the next and from player to player. The choice of threshold at 12% will be discussed in Sec. 3.

For potential transitions, shown in Fig. 1 (bottom row, vertical lines), the immediate area surrounding the peak was analyzed (for time values $\pm 0.25$ s). If the shape of the $\ddot{P}_{\mathrm{RMS}}$ that surrounded this peak was determined to be bimodal (two local maxima), the potential transition was classified as a tongued transition (Fig. 1, bottom row, right). If the shape was determined to be unimodal (one local maximum), the potential

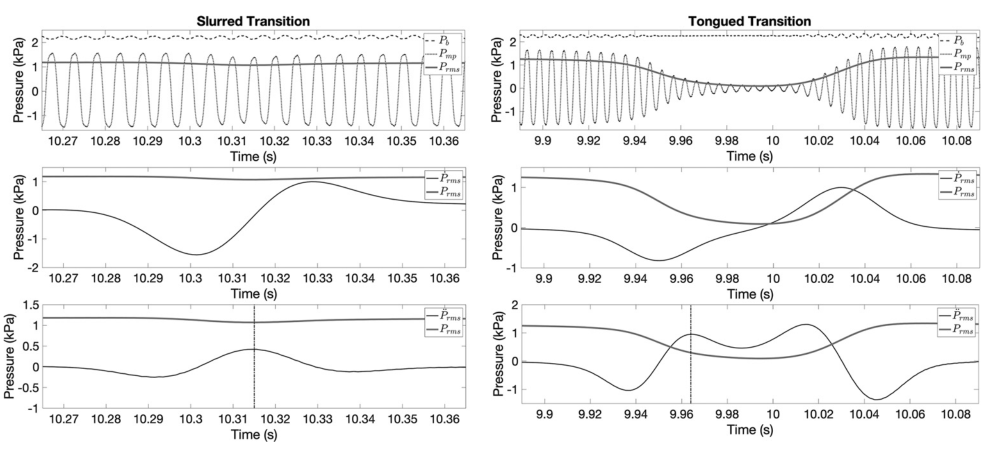


Fig. 1. Typical data from study. From Tests 1 (slurred) and 2 (tongued), the transition between concert E3 and F3: Left column is slurred, right column is tongued. Top row: $P_b$ (dashed black line), $P_{mp}$ (dotted black line), and $P_{\mathrm{RMS}}$ (thick gray line). Middle row: $P_{\mathrm{RMS}}$ (thick gray line), $\dot{P}_{\mathrm{RMS}}$ (thin black line). Bottom row: $P_{\mathrm{RMS}}$ (thick gray line), $\ddot{P}_{\mathrm{RMS}}$ (black line), the first maximum of $\ddot{P}_{\mathrm{RMS}}$ is marked with a vertical dashed-dotted black line for both Tests 1 and 2 (in the bottom row). The values on the $y$ axis refer to the $P_{\mathrm{RMS}}$ curve only.

Table 2. Algorithm performance for all player data, Tests 1–4. Slur performance on the left and tongue performance on the right.

| | Found | Expected | | Found | Expected |
|---|---|---|---|---|---|
| **Slur** | 625 | 640 | **Tongue** | 450 | 640 |
| False Positive Slur | 6 | 0 | False Positive Slur | 17 | 0 |
| False Positive Tongue | 3 | 0 | False Positive Tongue | 5 | 0 |
| False Negative Slur | 12 | 0 | False Negative Tongue | 15 | 0 |
| Misclassification | 3 | 0 | Misclassification | 175 | 0 |
| % Correct | **97.65**% | | % Correct | **70.31**% | |

transition was classified as a slurred transition (Fig. 1, bottom row, left). If the potential transition did not match any of these cases, it was classified as not a transition.

## 3. Experimental results

The main outcome from this study is the new automatic classification method for note-to-note transitions based on the signature shape of the second derivative of the internal mouthpiece RMS pressure signal ($\ddot{P}_{\mathrm{RMS}}$). For all tests where the musician was instructed to play either slurred or tongued, however, an exact method for tonguing was not specified and the six players incorporated a variety of techniques.

### *3.1 Results from tests 1–4*

The algorithm detected the majority of slurred and tongued transitions as confirmed by visual and aural inspection. The visual difference, even in the raw internal pressure, was obvious between the two transition types and therefore for the simple passages, accurate manual classification could be achieved. For the suite of measurements included in Tests 1–4 there were an expected 640 possible slurred transitions and 640 tongued transitions. After the algorithm was applied, the number of automatically detected slurred and tongued transitions were noted and compared to the expected number. However, as there were often discrepancies in playing technique from player to player, the algorithm would also register false positives, false negatives, or misclassifications. All of these classifications are detailed in Table 2. A false positive is when the algorithm detected a transition where there was no transition. A false negative is when the algorithm detected no transition where there is in fact a transition. And a misclassification is simply where the algorithm correctly detected there is a transition but incorrectly classified it as a different type of transition. The number of correct classifications, misclassifications, and false negatives should sum to the total number of expected transitions. With a threshold percentage set to 12%, the algorithm was able to detect 625 of 640 slurred transitions and 450 out of 640 tongued transitions for 97% and 70% accuracy, respectively, as presented in Table 2.

The choice of a 12% threshold was made after a study of this parameter's effect was complete. As described in Sec. 2, in order to be classified as a transition, the first, second-derivative peak must be larger than a given threshold, for example, 12% of the RMS pressure envelope of the preceding note. Figure 2 shows the percentages

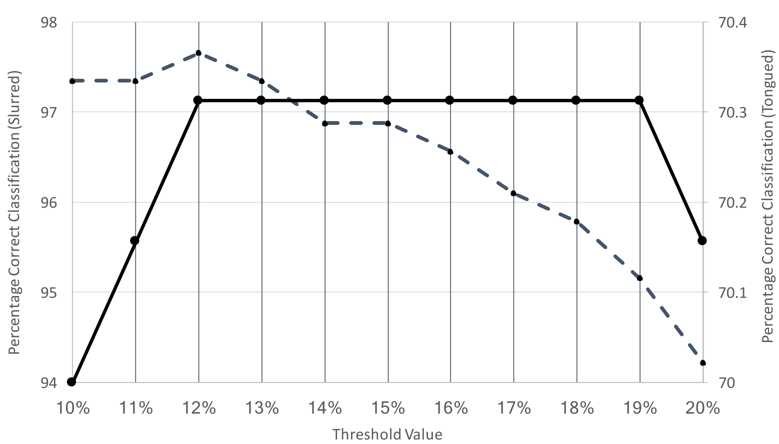


Fig. 2. (Color online) Algorithm performance based on threshold value. A 12% threshold represented the highest level of identification of articulation type based on experimental protocol. The black lines (left axes labels) are for slurred transitions and the dashed lines (right axes labels) are for tongued transitions.

Fig. 3. Passage of Mozart clarinet concerto played by musicians. The top line would be played first to represent the second register and the bottom line would be played directly after the measurement to represent the first register.

of correct classifications (tongue and slurs) by the algorithm given different threshold percentages. The value chosen (12%) represented the best case for each transition type, for this set of data. Given other players, or other playing tasks, this value could likely change and should be monitored.

Another difficulty in classifying the tongued transitions was that musicians used a variety of tonguing techniques. By aural inspection of the measurements, a few players did not in fact tongue all transitions when instructed to do so (either by accident or by misunderstanding the task). When running the analysis without including players C, D, and F the percentage of correct tongue classifications increased from around 70% to nearly 90% and the misclassifications decreased from 27% to 9%. As the algorithm was meant to classify tongued signals (for Tests 2 and 4), not including the data from these players in the final performance analysis of the algorithm is justified. The correct classification percentage for slurred transitions is always greater than for tongued transitions. This is also due to the difference in tonguing style and duration throughout the tests. Plans for future studies include varying tempos and tonguing style to test the classification ability in these different situations, like those described by Pàmies-Vilà *et al.*[2]

### *3.2 Results from test 5*

The next result of this study was the application of this algorithm to detect transition types in a more difficult classification test–a passage of music. The musicians concluded the measurement sessions by playing a short, well known passage from the Mozart clarinet concerto in order to measure transition signatures in a more musical context (Fig. 3). The method and metric were successful for this passage as well, showing the second derivative curve following the patterns outlined in Fig. 1 where the musicians first played the top line (second register), then the bottom line (first register) with the articulations shown, *mf* dynamic level at an adagio tempo. Shown in Fig. 4 is a small portion of the algorithm output from the nearly 60 s full measurement signal—the $P_{\mathrm{RMS}}$ signal and $\ddot{P}_{\mathrm{RMS}}$. The dotted black vertical line indicates a slurred transition and a dashed black vertical line indicates a tongued transition.

To determine the algorithm's accuracy, the transition types in this file were determined based on the shape of $\ddot{P}_{\mathrm{RMS}}$ at the transition and compared to the expected transitions as depicted in the passage of music with the majority of the transitions

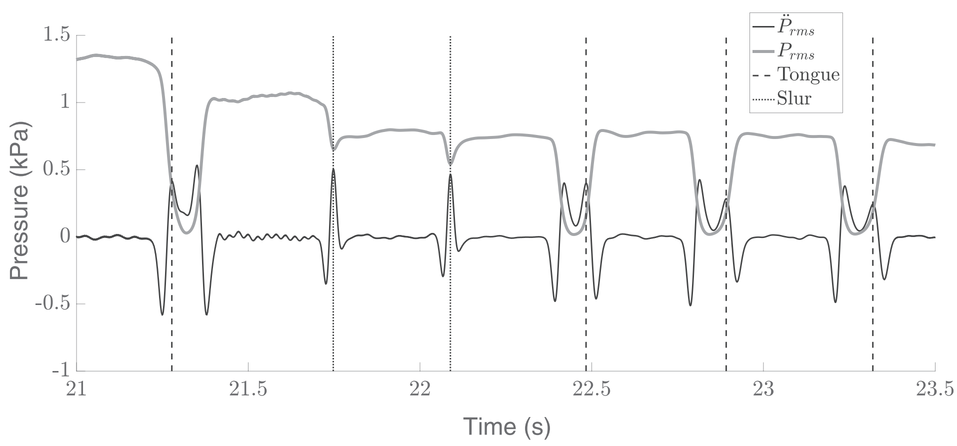


Fig. 4. Analysis of a portion of the Mozart clarinet concerto selection. Dark gray line—$P_{\mathrm{RMS}}$, thin black line—$\ddot{P}_{\mathrm{RMS}}$, black vertical dashed—a tongued transition, black dotted vertical line—a slurred transition. All transitions here were automatically detected given the criteria detailed in Sec. 2. The coordinates on the $y$ axis refer to the $P_{\mathrm{RMS}}$ curve only.

Table 3. Correct classifications and false negatives for three different configurations of players for slurred transitions and tongued transitions. For each section of the table, the column represents the player configuration: Left—all players, middle—player D removed from analysis, right—players D and F removed from the analysis. The rows offer the percentage correct classifications for each type and the number false negatives detected in the Mozart excerpt.

| | All Players All/Expected = %Correct | No Player D All/Expected = %Correct | No Player D or F All/Expected = %Correct |
|---|---|---|---|
| Slurred | 100/110 = 90.9% | 71/80 = 88.7% | 55/60 = 91.6% |
| False Negative Slur | 5 | 5 | 4 |
| Tongued | 296/399 = 64.1% | 234/320 = 73.1% | 192/240 = 80% |
| False Negative Tongue | 7 | 6 | 3 |

being tongued transitions. Then, as before, the number of correct classifications and false or misclassifications were noted. Table 3 shows the correct classifications and false negatives for three different configurations of players for slurred transitions and tongued transitions. For each section of the table, the column represents the player configuration: Left—all players, middle—player D removed from analysis, Right—players D and F removed from the analysis. The rows in Table 3 offer the percentage correct classifications for each type and the false negatives detected. By listening to the audio signal (and live tests) it was not clear that players D and F were correctly performing the articulations as requested. The choice of removing players from the analysis serves to show that the algorithm performance in classifying the tongued transitions drastically increases (row three of Table 3) when these players are not included in analysis (from 64% to 80% correct), yet the slurred classifications (row one of Table 3) were stable whatever the combination of players. This is likely due to the fact that there are a variety of tonguing methods employed by players and the current analysis does not make a clear distinction in tonguing type. Nevertheless, for more experienced players, the far right column of Table 3 represents the best case detection scenario for this data set, a correct classification of 80% of the tongued transitions and 90% of the slurred transitions in the Mozart excerpt.

## 4. Conclusion

This paper presented a computational method for detecting slurred and tongued note-to-note transitions in clarinet playing. Authors of past studies focusing on transients mention the lack of musical context in their research, choosing to study single, isolated notes. How this method can be expanded to pinpoint the exact start and stop of a transient is left for future work. However, the results of this work confirm that the method is a reliable way to determine the types of transitions and when they occur. An extension of the method could be to then determine the transition time between two notes, allowing for comparisons of various note fingerings in chromatic passages—perhaps seeking the smallest transition time. Finally, especially for more experienced players, it will also allow for further comparison of clarinetists, instruments, or instrument accessories as a measure of quality for listeners, manufacturers, and musicians alike.

## Acknowledgments

This work was supported by Grant No. PHY-160749 from the U.S. National Science Foundation and the Rollins College Student-Faculty Collaborative Scholarship Program. The authors thank P. Sanchez, A. David-Sivelle, and the SONS group at CNRS-Marseille for the SEM, and both K. Koons and T. Moore for helpful discussions.

## References and links

[1]J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in music signals," IEEE Trans. Speech Audio Process. **13**, 1035–1047 (2014).
[2]M. Pàmies-Vilà, A. Hofmann, and V. Chatziioannou, "Analysis of tonguing and blowing actions during clarinet performance," Front. Psychol. **9**, 617 (2018).
[3]W. Li, A. Almeida, J. Smith, and J. Wolfe, "The effect of blowing pressure, lip force and tonguing on transients: A study using a clarinet-playing machine," J. Acoust. Soc. Am. **140**(2), 1089–1100 (2016).
[4]B. Bergeot, A. Almeida, and B. Gazengel, "Response of an artificially blown clarinet to different blowing pressure profiles," J. Acoust. Soc. Am. **135**(1), 479–490 (2014).
[5]A. Almeida, W. Li, J. R. Smith, and J. Wolfe, "The mechanism producing initial transients on the clarinet," J. Acoust. Soc. Am. **142**(6), 3376–3386 (2017).
[6]W. Li, A. Almeida, J. Smith, and J. Wolfe, "How clarinettists articulate: The effect of blowing pressure and tonguing on initial and final transients," J. Acoust. Soc. Am. **139**(2), 825–838 (2016).

[7]E. L. Saldanha and J. F. Corso, "Timbre cues and the identification of musical instruments," J. Acoust. Soc. Am. **36**(11), 2021–2026 (1964).

[8]A. Munoz, B. Gazengel, J. Dalmont, and G. Plantier, "Comparative study of different physical models describing the reed behaviour in real and artificial playing conditions," in *Proceedings for the International Symposium on Musical Acoustics*, Le Mans, France (2014), pp. 131–136.