Automatic Hybrid Following in Real-Time Mixed Music: A Case Study with Antescofo and ipt˜ for Flute Playing Techniques

Nicolas Brochec
Tokyo University of the Arts
nicolas.brochec@pm.me

Jean-Louis Giavitto
STMS, CNRS, IRCAM, Sorbonne Université
jean-louis.giavitto@ircam.fr

ABSTRACT

This paper investigates how real-time recognition of instrumental playing techniques can extend automatic score following beyond the limits of pitch-based alignment. While systems such as Antescofo provide robust and largely plug-and-play score following, their listening model is primarily designed for stable, pitched events aligned with a fixed symbolic score. This makes them difficult to adapt to extended techniques, unpitched sounds, and musical forms involving partial improvisation or open notation.

To address these limitations, we explore a hybrid approach that combines multiple listening machines with complementary capabilities and allows dynamic switching between them during performance according to the musical context. Specifically, we integrate Antescofo with `ipt˜`*, a real-time playing technique recognition system based on lightweight machine learning models. We focus on the integration of real-time instrumental playing technique recognition as a means to enrich the listening process and support technique-aware navigation of the score.*

We evaluate this approach on the case of extended flute techniques, assessing both the feasibility of technique-aware following and the trade-off between system generality and performance. Results suggest that learning-based listening modules provide a practical compromise: they improve robustness for specific techniques while preserving much of the plug-and-play character supporting multiple works and performers. The results highlight a promising balance between generality, specificity, and performative robustness.

1. INTRODUCTION

1.1 Mixed Music

During the 1950s, experiments with electronic devices led to the emergence of electronic and electroacoustic music and to their blending. Unlike electronic music, which uses only electronic devices, or electroacoustic music, composed of pre-recorded sounds, mixed music (from French *musique mixte*) combines one or more instruments with an electroacoustic apparatus in a single musical work. The electroacoustic apparatus can be represented by one or more performers manipulating the tape or directly transforming acoustic sounds with electronic devices. Almost all the 20th-century leading mixed music composers were primarily contemporary composers, with complex musical languages and extensive use of instrumental gestures and techniques.

One of the main challenges in performing mixed music is synchronizing the instrumental and electronic parts. Unlike orchestra and ensemble music, mixed music combines electronic and instrumental parts performed through different media. The electronic part is executed by a machine operating on absolute time, in seconds, while the instrumental part is performed by a musician playing in relative time, the *tempo*. These two parts follow different time paradigms and cannot be synchronized as an orchestra would, where only the relative time paradigm is used. As for instrumental music, synchronization between the instrumental and electronic parts is necessary, as the quality of its performance relies on it [1].

Antescofo (Anticipatory Score Follower) is a system that enables real-time alignment of the electronic part with the live instrumental performance [2]. To our knowledge, Antescofo is one of the most robust following systems. However, it may have difficulties to follow complex musical gestures, such as extended playing techniques which are often used in contemporary mixed music. A second drawback is that the Antescofo's listening mechanism relies on alignment paradigm that is relevant for written music, but limits the system's use to entirely written forms, leaving out more open forms.

1.2 Instrumental Playing Techniques

Instrumental Playing Techniques (IPTs) comprise a diverse family of performance techniques related to "musical gestures" [3] and encompass the actions the performer applies to the instrument to produce different timbres. These actions can be types of articulations (e.g., *staccato*), ornaments (e.g., *trill*), preparation such as with mutes (e.g. straight mute), and sound-producing methods (e.g. *harmonic*) depending on physical features of the instrument. They have been practiced throughout the history of music worldwide, shaping local music identities [4].

The performance of playing techniques in Western written music is closely tied to their notation. As instrumental music gained independence from the 15th to the 17th centuries, treatises such as Praetorius's Syntagma Musicum (1618) [5] described instruments with minimal indications of playing technique. The 19th century introduced virtuosic performance, prompting the need to standardize the notation of playing techniques. In the 20th century, explorations of timbre through *Klangfarbenmelodie* and later

avant-garde movements led composers such as Lachenmann, Ferneyhough, and Grisey to expand instrumental expression through extended playing techniques, establishing sound and timbre as core compositional elements [6].

1.3 Alignment versus Recognition

In score following, the alignment paradigm refers to the continuous matching of a live audio stream against a predefined symbolic representation of the musical work (or, in the audio-to-audio alignment, to another audio stream). Antescofo is a representative and successful instance of this paradigm: its listening machine estimates pitch and tempo to align performance events with a priori written symbolic score. While this approach has proven robust for a wide range of notated repertoires, it implicitly presupposes the existence of a fixed, determinate form against which the performance can be aligned. As a consequence, alignment-based following becomes problematic in musical situations where the form is partially open, non-deterministic, or only loosely specified. This includes scores that offer branching choices to the performer, as well as sections involving guided or free improvisation, where the notion of a unique reference trajectory through the score no longer applies. In such contexts, recognition-oriented approaches—focusing on identifying salient musical states, gestures, or playing techniques rather than precise score positions—appear more appropriate.

Conversely, approaches based solely on recognition do not fully address the requirements of mixed-music performance. While recognition systems can identify musical features such as playing techniques, timbral states, or gesture classes without relying on a fixed score position, they generally lack an explicit temporal and structural model of the musical form. As a result, they provide limited support for anticipation, synchronization, and the precise scheduling of electronic processes over extended time spans. In the context of mixed music, where the electronic part often unfolds according to carefully designed temporal relations, the absence of a reliable notion of progression through the piece can lead to fragile or musically unsatisfactory interactions. Recognition-based systems are therefore better suited to describing local musical states than to managing long-range temporal dependencies.

1.4 Hybrid Following: Extending alignment with specific recognition

This tension between alignment and recognition suggests that neither alignment nor recognition alone is sufficient, and that a productive approach lies in their combination: alignment provides a global temporal and formal framework, while recognition supplies context-sensitive information that can guide, constrain, or disambiguate the alignment process, especially in presence of IPT. In the other hand, recent studies in the field of real-time instrumental playing technique recognition have led to the development of reliable systems, such as `ipt~` Max external object [7], which are applied in the context of interactive co-creative systems [8].

These considerations motivate the present work and frame our exploration of hybrid listening strategies that can extend the applicability of score following beyond pitched sound and strictly pre-composed, fully specified forms. For the needs of our study, we focused on flute playing techniques as recent studies [8, 9] provide sufficient elements to implement a robust system and a comprehensive evaluation. Note, however, that some extended techniques for the flute still preserve sufficient spectral information for the limitations of current pitch-based listening machines to remain moderate. Our aim is not to address the performance of the two approaches to flute IPT, but rather to assess the feasibility of a hybrid listening approach within the framework of Antescofo, the robustness (*i.e.*, determinism and stability) of the system, and to examine how such an approach impacts the specification and writing of the electronic parts in mixed-music works.

1.5 Organization of the paper

In the first section, we present various flute-playing techniques, introduce Antescofo's score-following system, and examine existing methodologies for real-time recognition of playing techniques. In the second section, we detail our proposed system that merges Antescofo and `ipt~`. In the third section, we design an evaluation to observe how our proposed method performs in a real-world, similar setting for mixed music performance. Finally, we discuss the results of this evaluation and draw our conclusions.

2. BACKGROUND

2.1 The Diversity of Flute Playing Techniques

Each type of musical instrument has a wide range of instrumental playing techniques. In the case of the flute, various playing techniques can be produced by different blowing, tonguing, and ornamentation techniques, which can be performed separately or simultaneously. The sound of the flute is produced by the friction of the blown air into the mouthpiece [10]. Differences in blowing pressure, the degree of mouthpiece coverage, and blowing direction can produce distinct playing techniques [11]. The flute register extends from medium (C3 $\approx$ 261Hz) to treble pitch (C6 $\approx$ 2093Hz) with some exceptions [11]. Thanks to a unique key, professional flutes can produce B2 ($\approx$ 247 Hz) [12]. Professional flutists can reach pitches up to G6 ($\approx$ 3136 Hz) [11].

We provide a brief overview of flute playing techniques to illustrate the range of sounds the performer can produce. The *ordinario* is the ordinary sound of the flute; it is stable in frequencies. The *aeolian* sound involves blowing into the mouthpiece to make a noisy sound similar to the sound of air, often with a pitch component. The *flatterzunge* involves blowing while rolling the tongue, creating a sound of successive attacks. The *key click*, also known as *key percussion*, involves using the finger to strike flute keys and producing a percussive sound. The *multiphonics* consist of using a specific fingering to produce chord-like sounds [12], which is rarely stable over time and involves a lot of fluctuation in frequencies. The *pizzicato*, also known as *slap*, consists of making a powerful attack with the tongue. The *play and sing* technique involves singing while blowing into the flute. The *staccato* is a short sound made with a sharp attack. The *tongue ram* technique involves forcefully propelling the tongue into the mouthpiece of the

flute, producing a percussive, pitched sound. The *trill* involves alternating two adjacent notes rapidly, typically a main note, and a minor or major second above or below this main note. The *whistle tone* technique involves blowing a very soft, focused stream of air across the mouthpiece to produce high-pitched, whistle-like tones. Some of the playing techniques above are contemporary, as they extend the flute's sonic palette by using unusual gestures on the instrument. These techniques are the *key click*, *multiphonics*, *play and sing*, *pizzicato*, *tongue ram*, and *whistle tone*, and most of them have emerged in 20th-century contemporary music.

In the music of the 20th and 21st centuries, the use of playing techniques is often combined; some composers, such as Lachenmann and Ferneyhough, specialize in the extensive use of intricate playing techniques [13]. Combining playing techniques can be achieved through simultaneous actions on the instrument (blowing, tonguing, and ornamentation). For example, blowing and producing an *aeolian* sound with a *flatterzunge* tongue and trilling a key simultaneously are ergonomically possible.

2.2 Automatic Score Following with Antescofo

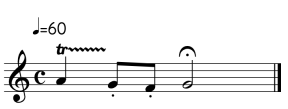
Score following is the process of tracking live players as they play through a composer's written score. The resulting information can be used, for example, to provide automatic computer accompaniment. The process involves entering a musical score representation into the computer, playing it as live MIDI or audio input, and comparing the computer's stored score to the live input. The early score-following systems of Dannenberg [14] and Vercoe [15] were based on MIDI.

Antecofo couples two different systems, a listening machine and a domain-specific programming language [16]. Composers use it to create an augmented score that describes musical scenarios in which electronic processes are scheduled to interact with a live performance.

An augmented score is a script composed of three main elements: *events*, which are recognized by the listening machine to describe the performance; *actions*, which are triggered when corresponding events are detected, when logical conditions are met, or simply as time elapses; and *synchronization strategies*, which link the performer's musical flow to the electronic processes.

Events can be grouped into three main categories. Abstract events (EVENT) do not correspond to a musical sound and are typically triggered manually. They were originally designed to enable an operator to manually trigger an event, causing the listening machine to wait for this notification and resume execution after a predefined delay. Atomic events (NOTE) represent elementary musical units, such as a single pitch. Finally, compound events act as containers for other events and describe more complex musical structures, such as simultaneous pitches (CHORD), rapid alternations (TRILL), or ordered sequences of heterogeneous events (MULTI).

In addition to basic events and actions, Antescofo also includes other functionalities, such as event attributes notated with a `@` sign. Event attributes enable the specification of additional information about the live performance. For example, the event attribute `@staccato` indicates that the event has a short temporal morphology, helping the listening machine to detect the corresponding event. The same applies to attribute `@fermata`, which indicates to the listening machine that the related event may be out of tempo.




```
BPM 60
TRILL (A4 B4) 1
NOTE G4 1/2 @staccato
NOTE F4 1/2 @staccato
NOTE G4 2 @fermata
```

Figure 1. Example of a conventional music score alongside its score for Antescofo, which includes *staccato* technique and *fermata* sign.

By combining the different attributes, Antescofo can cover a fair range of standard and extended flute playing techniques. Table 1 shows playing techniques and how to note them in the augmented score.

Playing Techniques	Antescofo Notation
ordinario	`NOTE p d`
staccato	`NOTE p d @staccato`
trill	`TRILL (p1 p2) d`
flatterzunge	`TRILL (p1 p1) d`
multiphonics	`CHORD (p1 p2 ...) d`
glissando	`MULTI (p1) -> (p2) d`

Table 1. Antescofo score notation corresponding to playing techniques, with `p` the pitch and `d` the duration.

Nevertheless, contemporary flute performance is far from being limited to the six techniques described in Table 1 as described in Section 2.1. Even though they can be notated in the augmented score, this does not guarantee that the listening machine will recognize them accurately during the performance. Some of them are out of the range of Table 1, such as unpitched playing techniques, *i.e*, *aeolian*, *key-click*, and *pizzicato*, which makes Antescofo difficult to use for these playing techniques. We think that using an external listening machine specialized in playing techniques in parallel of Antescofo would benefit the automatic following of specific IPT used in contemporary mixed music. In the following section, we examine the existing real-time recognition systems for playing techniques.

2.3 Real-Time Playing Techniques Recognition

Playing technique recognition is part of the Musical Information Retrieval (MIR) domain. Studies on the subject encompass a wide range of instruments. For Western instruments, studies mainly focus on acoustic guitar [17, 18], electric guitar and bass guitar playing techniques [19, 20], piano playing techniques [21], violin [22], cello [23], and flute [8]. Studies also include ones on Eastern instruments, as traditional music includes a wide range of playing techniques. Recent studies focus on the Chinese bamboo flute [24], and the bonang barung, an Indonesian Javanese gamelan instrument [25].

Yet, only a few studies focus on real-time recognition. These studies are about the cello [23], the guitar [18], the electric guitar [7], and the flute [8]. From these four studies, we can highlight several key points necessary for robust real-time playing recognition: sufficient data diversity and quantity, a playing technique taxonomy when applicable, a meaningful audio signal representation, a lightweight

classification algorithm, and a fast inference routine via dedicated buffers.

3. PROPOSED SYSTEM

In this section, we detail the proposed system that enables the automatic following of flute playing techniques. The first subsection describes the design of our recognition system, drawing on existing methodologies. The second subsection details how the IPT recognition system is linked to Antescofo.

3.1 Real-time Flute IPT Recognition System Design

The methodology for real-time flute IPT recognition is divided into the following steps, each with its own dedicated subsection: data selection, taxonomy, dataset setup and preparation, system architecture, training, and results.

3.1.1 Data Selection

A limited number of sound banks that include contemporary flute playing techniques exist. We identified four sound banks to ensure diversity and quantity: conTimbre [26], FullSOL [27], xSample [1] , and Geidai Flute Database (GFD) [28]. We selected these sound banks because, to our knowledge, they are the most exhaustive sound banks of flute IPTs, which include variable performance of the same IPT [26], recordings of different acoustic conditions, and with various performers, named as "P1" and "P2" in GFD [28].

3.1.2 From Data To Taxonomy

Unlike the cello and the guitar [23, 18], the performance of the flute IPTs is not only based on limb actions, but is also based on the quality of the blown air, e.g., *ordinario* and *aeolian*. These characteristics make gesture-based or hierarchical taxonomies harder to apply to the flute. Considering these constraints, we adopted a non-hierarchical taxonomy based on the 11 IPTs available in GFD [28]. Our task is then an 11-class classification problem using *aeolian*, *flatterzunge*, *key-click*, *multiphonics*, *ordinario*, *pizzicato*, *play-and-sing*, *staccato*, *tongue-ram*, *trill*, *whistletone* playing technique.

3.1.3 Dataset Setup and Preparation

We designed our dataset by combining the most populated datasets we have, "P1" from GFD, conTimbre, and xSample, by carefully moving the audio files in the corresponding playing technique category in GFD. The rest have been discarded. "P2" from GFD serves as the validation set, and FullSOL as the test set. To prepare the audio files for each subset, we first downsampled audio from 96 kHz and 48 kHz, depending on the audio file's origin, to 44.1 kHz, as flute harmonics can exceed 10 kHz [11]. Silence is trimmed with a threshold of -60 dBFS, as it is irrelevant. Using both short and long playing techniques, audio files are segmented into 1/3-second contiguous sequences and padded with zeros if shorter. To increase the duration of our training dataset, we applied domain-specific data augmentations, such as in [18]. Playback speed was randomly varied by ± 0.1 to emulate variability in IPT execution speed. Recordings were detuned by ± 100 Hz around A440 to reflect natural tuning fluctuations. Gaussian noise was added to replicate the noise introduced by signal amplification. Thanks to data augmentation, the training dataset length increased, totaling 16 hours.

[1] Available on the web: `https://www.xsample.de/xsample_flute.htm`.

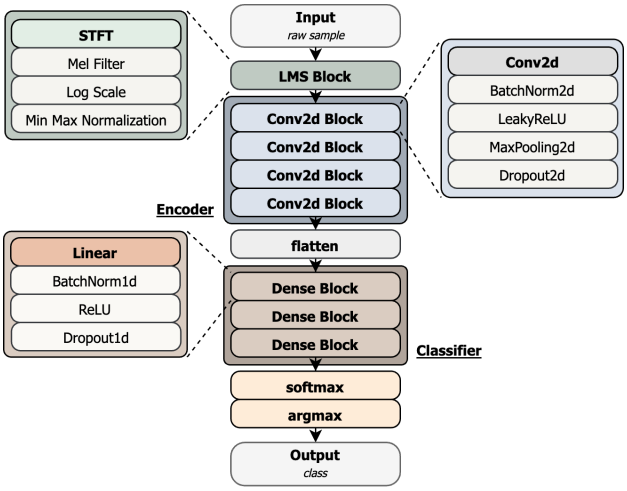


Figure 2. System architecture using a Logarithmic Mel Spectrogram, an encoder composed of four convolutional layers, and a classifier comprising three fully connected (dense) layers.

3.1.4 System Architecture

For the system architecture, we adopted the architecture from [29] and removed redundancy in the deep layers to make the model lighter. Our architecture consists of a Logarithmic Mel Spectrogram (LMS) block that first transforms raw audio into a grayscale spectrum. Long-duration IPTs benefit from long-range features, whereas short IPTs require high-resolution features [30]. As our dataset includes long and short IPTs, a compromise must be found between long-range and high-resolution features. Inspired by the study [31], in which concatenating spectrograms from different FFT window sizes to improve onset detection is suggested, we adopted a similar strategy to enhance short IPT classification (*e.g.*, *key-click*, *pizzicato*, and *staccato*). Additionally, a previous study showed that high-resolution features improve classification performance [23]. Based on these insights, we defined an LMS configuration that stacks three spectrums with FFT sizes of 512, 1024, and 2048, utilizing a hop size of 128 samples and 384 mel bands, which is three times that used for the cello [23]. The minimum frequency is set to 120 Hz based on the lowest reachable pitch (through *tongue-ram*).

The LMS block is followed by an encoder composed of convolutional layers and a classifier comprising three fully connected (dense) layers. Batch normalization and dropout layers are introduced to stabilize training and prevent overfitting. A flattened layer between the encoder and the classifier ensures dimension compatibility. The channels from the convolutional layers are 30, 60, 120, and 120, and the hidden sizes of the dense layers are 60 and 30. Kernel sizes for convolutional layers were set to `4x8`, `3x6`, `2x4`, and `2x4` to increase the receptive field along the time axis.

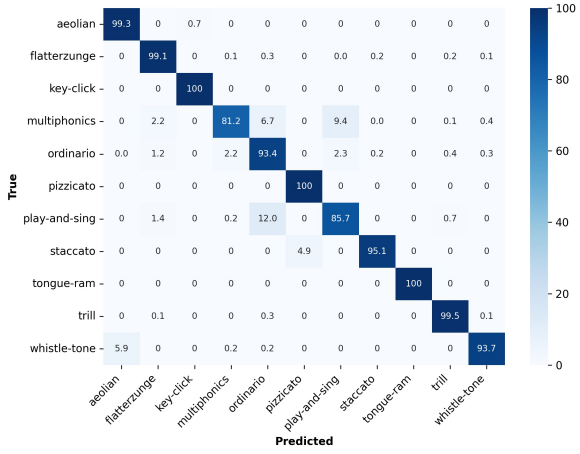


Figure 3. Normalized composite confusion matrix across the five tests showing true over predicted classes (%).

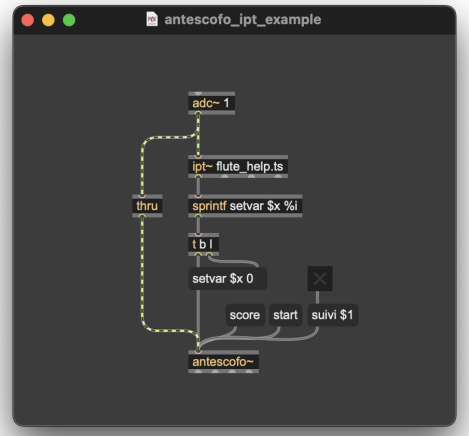


Figure 4. Max patch implementing `antescofo~` alongside `ipt~` object. IPT class index is sent to `antescofo~` using `setvar` message.

3.1.5 Training

We trained our configurations with a batch size of 128, cross-entropy loss, and the ADAM optimizer with a weight decay of 1×10^{-5}. The architecture is trained for up to 100 epochs. Training is stopped when the validation loss stops improving after 10 epochs. Five independent training runs are performed to ensure robust measurements.

3.1.6 Results

Following prior studies, we used macro F1 as our primary metric, as it provides a more accurate assessment of performance on unbalanced datasets [32]. Results show a strong macro F1 score of 93.4%±0.1, indicating robust generalization across separate training and test sources. The normalized composite confusion matrix across the five tests, as shown in Figure 3, indicates that models are accurate for almost all classes, except for *multiphonics* and *play-and-sing* (81.2% and 85.7%).

3.2 Crossbreeding Antescofo with `ipt~`

Our approach builds on existing mechanisms in Antescofo to integrate IPT recognition, as processed by `ipt~` into the score-following process without altering its core architecture. In particular, we take advantage of abstract events (EVENT), which were originally introduced to allow an operator to manually signal the occurrence of a musical event. In the standard workflow, the listening machine suspends progression until such a notification is received, after which the system resumes following.

In our hybrid approach, these abstract events are repurposed to represent musical events associated with instrumental playing techniques. Rather than relying on a human operator to notify their occurrence, the detection of these events is delegated to `ipt~`. The interaction between `ipt~` and Antescofo is achieved through explicit message passing, allowing the two systems to remain loosely coupled (they can be upgraded independantly) while exchanging minimal but sufficient information.

More specifically, `ipt~` continuously outputs the index of the currently recognized playing technique. This information is transmitted to Antescofo using the `setvar` message, which updates a global internal variable. Inside the Antescofo score, this variable is monitored using the construction `whenever` which evaluates logical conditions on variable updates [33]. When the expected playing technique is recognized, the corresponding condition becomes true and triggers an action. In our case, this action simply advances the system to the next event, effectively replacing the manual notification that would otherwise be performed by an operator.

To prevent unwanted interference with the standard score-following mechanism, particular care is taken to control the behavior of the listening machine during these passages. While IPT-based detection is active, Antescofo's following is temporarily disabled to avoid erroneous jumps to subsequent events that could be mistakenly detected by the pitch-based listening module. At the exit of the IPT passages, the score following is reactivated. This ensures a clear separation between the alignment-driven and the ecognition-driven phases while preserving global coherence.

This design allows extended playing techniques to be handled in the same event-based framework as other musical events. As a result, the electronic part specification remains uniform, and all the advantages of Antescofo's programming model are preserved, including the organization of electronic processes around events and tempo-based synchronization. Importantly, Antescofo's tempo inference is based solely on inter-onset intervals, which allows tempo estimation to remain active even during passages where alignment is suspended and event progression is driven by IPT recognition.

In this first study, communication remains unidirectional: `ipt~` informs Antescofo, but Antescofo does not influence the behavior of `ipt~`. However, the same messaging infrastructure would allow actions defined in the augmented score to dynamically adapt recognition parameters according to the musical context. Such bidirectional interaction opens promising perspectives for future work, in which listening strategies could be actively shaped by the structure and semantics of the score itself.

For the validation presented in this work, we deliberately restricted the interaction to the mere occurrence of events detected by `ipt~`, without exploiting additional information provided by the recognition process. However, the

proposed scheme readily supports such extensions. making possible to design error-recovery strategies, for instance to handle missed events, delayed detections, or uncertain classifications during performance.

4. EVALUATION

The proposed hybrid approach proved straightforward to implement and to integrate within existing Antescofo-based workflows. Its evaluation was conducted at two complementary levels. First, we designed ad hoc scores specifically written to stress and probe the behavior of the listening machines, allowing controlled comparisons between different following strategies. Second, the approach was deployed in the production of full-scale musical works that were rehearsed and performed in concert conditions.

These real-world pieces enabled validation of the transparency of the hybrid approach from the perspective of electronic writing, including the synchronization between the human performer and the electronic voices. This transparency largely stems from the fact that extended playing techniques can be treated within the same "event-with-duration" paradigm as other musical events. As a result, the specification of the electronic part remains unchanged: electronic processes are still organized, triggered, and synchronized through events and tempo, exactly as in standard Antescofo practice. This confirms that the hybrid approach does not introduce a conceptual rupture in the compositional workflow.

In the remainder of this section, we focus on a quantitative comparison between Antescofo used alone and the proposed hybrid system. The evaluation targets listening performance along two main criteria: *robustness*, understood here as determinism and consistency of event detection across performances, and *precision*, defined as the temporal accuracy of detected events with respect to their reference onset times. This comparison is meaningful because most flute playing techniques preserve enough spectral information for Antescofo's listening machine to achieve alignment. In addition to pitch, Antescofo exploits temporal cues such as notated durations and real-time tempo inference, allowing progression when pitch information is unreliable. However, in this case, the alignment-based approach is expected to show greater variability across performances, which motivates comparison with the more deterministic behavior targeted by technique-aware listening.

4.1 Scores Setups

The first score of our evaluation, top of Figure 5, aims to test two situations: the alternation of timbres on the same pitch, and the chaining of *aeolian* and *ordinario* across different pitches. The second score, bottom of Figure 5, includes all playing techniques available in our recognition model at the exception of *ordinario*, *staccato*, and *trill* whose recognition can be safely left to Antescofo. It aims to test a situation in which playing techniques are chained without repetition across different pitches.

As previously explained in Section 3.2, a `whenever` statement monitors the global variable that reflects the index of the detected IPT class. These indices are attached to playing techniques starting with *aeolian* (0) and ending with *whistle-tone* (10) in alphabetical order. Each event

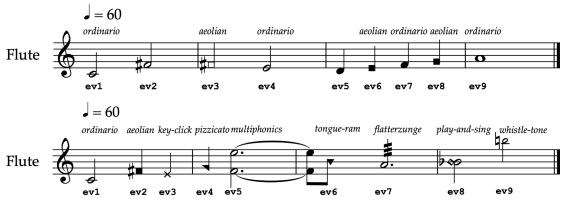


Figure 5. Top: Score 1 including *aeolian* and *ordinario* techniques. Bottom: Score 2 including *aeolian*, *key-click*, *pizzicato*, *multiphonics*, *tongue-ram*, *flatterzunge*, *play-and-sing* and *whistle-tone* techniques.

is labeled, and each score totals nine events. In the case where Antescofo is used alone, the corresponding augmented scores utilize as many notations as possible from Table 1. The parameters of `ipt~` are set to facilitate fast, rather than stable, recognition.

4.2 Metrics

Following a previous methodology [34], we define a reference onset time for each score event by mapping event timings onto a reference grid, thereby defining the ground truth between the score and its performance. We define the **error** $e_i = t_i^e - t_i^r$, the time lapse between the estimated alignment time t_i^d, the time at which the system reports the event, and the reference time t_i^r, the reference onset time according to the reference grid, for each event i.

Misaligned notes are notes that are correctly detected, but are too far from the reference onset time to be considered correct. We define a **misalign rate** such defined in [34], which is the percentage of correctly aligned notes p_e with absolute error $|e_i|$ greater than a threshold θ_e. We consider this alignment rate as the second metric for our experiments and use 300, 250, and 200 ms as our thresholds.

4.3 Results

Line charts Figures 6 and 7 represent the mean latency for each event across the six performance of scores in Figure 5.

Line chart Figure 6 shows that Antescofo performs well alone on events 3, 6 and 8, that are all *aeolian* techniques, see Figure 5. Line chart Figure 7 shows that Antescofo has difficulties in following techniques on events 5 and 9, which are respectively *multiphonics* and *whistle-tone* playing techniques, see Figure 5.

This line chart also shows that standard deviation for each event represented by the error bars is smaller for Antescofo and `ipt~` than Antescofo alone on events 2, 5, 6, 7, 8, and 9. Despite being slower than Antescofo on all event except on the fifth and ninth, this result shows that the use of `ipt~` led to a more robust following across different performances.

Table 2 shows that only when the threshold θ_e is set to 300 ms, the setup with Antescofo and `ipt~` performs better than Antescofo alone.

5. DISCUSSION

In the first experiment (Score 1), which alternates *aeolian* and *ordinario* techniques, Antescofo alone performs better than the hybrid approach. *Aeolian* sounds combine air noise with a pitched component, which appears sufficient for Antescofo's pitch- and time-based listening to operate

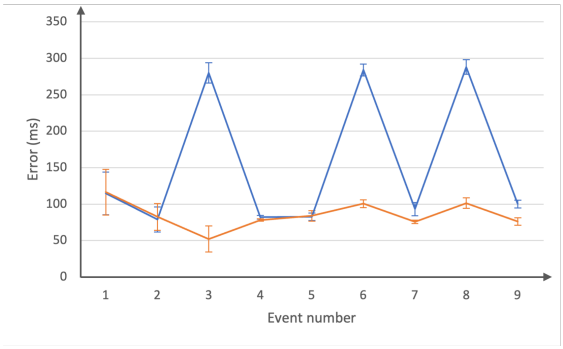


Figure 6. Mean error for each event across the six performances for the score 1, with error bars representing the standard deviation (ms). Antescofo performance (orange), Antescofo with `ipt~` performance (blue).

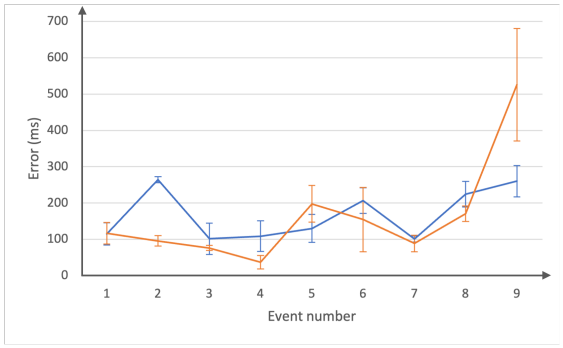


Figure 7. Mean error for each event across the six performances for the score 2, with error bars representing the standard deviation (ms). Antescofo performance (orange), Antescofo with `ipt~` performance (blue).

	Setup	Threshold θ_e		
		300 ms	250 ms	200 ms
Score 1	Antescofo	0.00	0.00	0.00
	Ant. and `ipt~`	7.41	31.48	33.33
Score 2	Antescofo	11.11	16.67	22.22
	Ant. and `ipt~`	**3.70**	20.37	40.74

Table 2. Misalignment rates for each setup across 6 performances (%).

reliably. In this case, the additional latency introduced by playing technique recognition leads to slower alignment without improving robustness. Misalignment rates confirm that Antescofo alone is fully adequate for such situations.

The second experiment (Score 2), which involves a broader range of extended techniques, reveals a different behavior. When used alone, Antescofo exhibits difficulties with techniques such as *multiphonics* and *whistle tones*, which are characterized by unstable or very low-energy spectral content. These difficulties result in greater variability across performances, as reflected by higher standard deviations in alignment time. In such cases, the progression of the listening machine relies primarily on temporal information to approximate the occurrence of musical events, which in turn leads to less reliable tempo inference. These issues are expected to worsen as the duration of sequences dominated by instrumental playing techniques increases. In contrast, the hybrid approach produces more consistent alignment for these events, even though it remains slower overall. This increased robustness comes at the cost of additional latency introduced by the recognition process.

These two experiments also reveal some limitations of the recognition model and `ipt~`, especially in the time it takes to recognize specific techniques. The line charts show that for the *aeolian* techniques, it takes more than 250 ms to be recognized. *Aeolian* techniques are events 3, 6, 8 on score 1, and event 2 on score 2. Line chart on Figure 7, shows that *tongue-ram*, *play-and-sing*, and *whistle-tone* techniques are recognized in more than 200 ms. Further improvements are required to enable faster recognition on the playing techniques recognition model side, which would make it more competitive against Antescofo's listening machine alone.

The error bars, representing the standard deviation in alignment time across the six performances, show that, for Antescofo, alignment time varies widely. When using `ipt~`, the error bars for events 2, 5, 6, 7, 8, and 9 show that playing technique recognition led to minor variations of alignment time, even though it scores a slower one than Antescofo alone. The misalignment rates in Table 2 show that for the performance on score 2 with a threshold of 300 ms, the use of `ipt~` yielded a lower misalignment rate (3.70%). However, when the threshold is shorter, misalignment rates show that Antescofo alone performed better, 16.67% against 20.37% for a threshold of 250 ms, and 22.22% against 40.74% for a threshold of 200 ms. The results suggest that, if robustness across performance is preferred to alignment time, or if the score is composed of many *multiphonics* and *whistle-tone* techniques, then our proposed hybrid approach is recommended. In any case other than these, Antescofo is recommended alone, and the usage of `ipt~` in the context of following flute playing techniques is case-specific.

Misalignment rates further highlight this trade-off. When using a relatively tolerant threshold (300 ms), the hybrid system outperforms Antescofo alone on the more complex score. For stricter thresholds, however, Antescofo alone yields better results. These observations suggest that the hybrid approach is particularly relevant when robustness and consistency across performances are prioritized over fast alignment, especially in scores dominated by unstable or weakly pitched techniques.

6. CONCLUSIONS

In this work, we investigated how integrating real-time playing-technique recognition can enhance automatic score following in mixed music contexts. By combining Antescofo with `ipt~`, we introduced a technique-aware, hybrid, score-listening workflow that handles a wide range of contemporary flute techniques. Our recognition model, trained on diverse sound banks, achieved strong generalization and provided reliable IPT information to drive conditional navigation in Antescofo scores.

Evaluation with professional flutists showed that the hybrid system improves robustness for unstable techniques, even at the cost of additional alignment time. These results suggest that technique-aware following is beneficial when stability across performances that involve many playing techniques is more important than fast alignment. The experiments also expose limitations of the current recognition model, notably the time required to reliably identify specific techniques, such as *aeolian*, *tongue-ram*, *play-and-sing*, and *whistle tones*. Reducing recognition latency would be a key factor in improving the competitiveness of the hybrid approach. The evaluation remains limited

to short excerpts and does not address failure cases such as missed detections or performer errors. Assessing these situations and testing longer, structurally richer pieces, as well as ones with open notation or partial improvisation, and exploring adaptation with other instruments constitutes an important direction for future work.

In addition to these observations, the proposed approach was successfully validated in a real-world mixed music setting. It was implemented in an original mixed music composition of approximately eight minutes, in which 20% of the events relied on flute playing technique recognition to drive the electronic part. This work was submitted by the first author to the music track of this conference.

Acknowledgments

This research was supported by the ERC REACH Project (GA #883313) and a MEXT scholarship from the Japanese Government awarded to Nicolas Brochec.

7. REFERENCES

[1] A. Cont and M. Rhéaume, "Synchronisme musical et musiques mixtes: du temps écrit au temps produit," *Circuit*, vol. 22, no. 1, pp. 9–24, 2012.

[2] A. Cont, "ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music." in *International Computer Music Conference (ICMC)*, 2008, pp. 33–40.

[3] C. Cadoz and M. M. Wanderley, "Gesture-music," *Trends in gestural control of music*, 2000.

[4] P. V. Bohlman, *The Cambridge history of world music*. Cambridge University Press, 2013.

[5] M. Praetorius *et al.*, *Syntagma Musicum II: De Organographia, Parts III–V with Index*. Zea Books, 2014.

[6] M. Solomos, *From music to sound: The emergence of sound in 20th- and 21st-century music*. Routledge, 2019, pp. 28–30.

[7] M. Fiorini, N. Brochec, J. Borg, and R. Pasini, "Introducing EG-IPT and ipt˜: a novel electric guitar dataset and a new Max/MSP object for real-time classification of instrumental playing techniques," in *New Interfaces for Musical Expression Conference (NIME)*, Canberra, Australia, Jun. 2025.

[8] N. Brochec, M. Fiorini, M. Malt, and G. Assayag, "Interactive Music Co-Creation with an Instrumental Technique-Aware System: A Case Study with Flute and Somax2," in *International Computer Music Conference (ICMC 2025)*, Boston (MA), United States, Jun. 2025.

[9] N. Brochec, T. Tanaka, and W. Howie, "Microphone-based Data Augmentation for Automatic Recognition of Instrumental Playing Techniques," in *International Computer Music Conference (ICMC)*, 2024.

[10] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*. Springer Science & Business Media, 2012, pp. 503–537.

[11] J. Meyer, *Acoustics and the performance of music: Manual for acousticians, audio engineers, musicians, architects and musical instrument makers*. Springer Science & Business Media, 2009, pp. 64–70.

[12] C. Levine and C. Mitropoulos-Bott, *The Techniques of Flute Playing I/Die Spieltechnik der Flöte I*. Bärenreiter-Verlag, 2019.

[13] R. Feller, "Resistant Strains of Postmodernism: The Music of Helmut Lachenmann and Brian Ferneyhough," in *Postmodern music/postmodern thought*. Routledge, 2013, pp. 249–262.

[14] R. B. Dannenberg, "An on-line algorithm for real-time accompaniment," in *International Computer Music Conference (ICMC)*, vol. 84, 1984, pp. 193–198.

[15] B. Vercoe, "The synthetic performer in the context of live performance," in *Proceedings of International Computer Music Conference*, 1984, pp. 199–200.

[16] J.-L. Giavitto, J.-M. Echeveste, A. Cont, and P. Cuvillier, "Time, timelines and temporal scopes in the antescofo dsl v1. 0," in *International Computer Music Conference (ICMC)*, 2017.

[17] L. Su, L.-F. Yu, and Y.-H. Yang, "Sparse Cepstral, Phase Codes for Guitar Playing Technique Classification." in *International Society for Music Information Retrieval Conference (ISMIR)*, 2014, pp. 9–14.

[18] A. Martelloni, A. P. McPherson, and M. Barthet, "Real-time Percussive Technique Recognition and Embedding Learning for the Acoustic Guitar," *arXiv preprint arXiv:2307.07426*, 2023.

[19] Y.-P. Chen, L. Su, Y.-H. Yang *et al.*, "Electric Guitar Playing Technique Detection in Real-World Recording Based on F0 Sequence Pattern Recognition." in *ISMIR*, 2015, pp. 708–714.

[20] J. Abeßer and G. Schuller, "Instrument-centered music transcription of solo bass guitar recordings," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 25, no. 9, pp. 1741–1750, 2017.

[21] B. Liang, G. Fazekas, A. McPherson, and M. Sandler, "Piano pedaller: a measurement system for classification and visualisation of piano pedalling techniques," 2017.

[22] L. Su, H.-M. Lin, and Y.-H. Yang, "Sparse modeling of magnitude and phase-derived spectra for playing technique classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2122–2132, 2014.

[23] J.-F. Ducher and P. Esling, "Folded CQT RCNN for real-time recognition of instrument playing techniques," in *International Society for Music Information Retrieval*, 2019.

[24] C. Wang, E. Benetos, and E. Chew, "CBFdataset: A Dataset of Chinese Bamboo Flute Performances," Dec. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5744336

[25] V. L. Hardjanto and W. Wahyono, "Recognising Bonang Barung Gamelan Instrument Playing Technique Using Convolutional Neural Networks," *Applied Ethnomusicology*, vol. 1, no. 1, pp. 71–86, 2025.

[26] T. A. Hummel, "Algorithmic orchestration with contimbre," in *Journées d'Informatique Musicale*, 2014.

[27] C. E. Cella, D. Ghisi, V. Lostanlen, F. Lévy, J. Fineberg, and Y. Maresz, "OrchideaSOL: a dataset of extended instrumental techniques for computer-aided orchestration," *arXiv preprint arXiv:2007.00763*, 2020.

[28] N. Brochec and W. Howie. (2025, Jan.) GFDatabase: A Database of Flute Playing Techniques. Accessed: 2025-12-16. [Online]. Available: https://doi.org/10.5281/zenodo.14712391

[29] M. Fiorini and N. Brochec, "Guiding Co-Creative Musical Agents through Real-Time Flute Instrumental Playing Technique Recognition," in *Sound and Music Computing Conference (SMC)*, Porto, Portugal, Jul. 2024.

[30] D. Li, M. Che, W. Meng, Y. Wu, Y. Yu, F. Xia, and W. Li, "Frame-level multi-label playing technique detection using multi-scale network and self-attention mechanism," in *Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[31] J. Schlüter and S. Böck, "Improved musical onset detection with convolutional neural networks," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 6979–6983.

[32] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv preprint arXiv:2008.05756*, 2020.

[33] J.-L. Giavitto. (v1.1, 2055) The Antescofo Documentation. Accessed: 2025-12-15. [Online]. Available: https://antescofo-doc.ircam.fr/

[34] A. Cont, D. Schwarz, N. Schnell, and C. Raphael, "Evaluation of real-time audio-to-score alignment," in *International Symposium on Music Information Retrieval (ISMIR)*, 2007.