EXTENDING BRASS & WOODWIND INSTRUMENTS WITH ACOUSTIC-AGGREGATE-SYNTHESIS

Paul Clift
IRCAM
paul.clift@ircam.fr

Adrien Mamou-Mani
IRCAM
adrien.mamou-mani@ircam.fr

René Caussé
IRCAM
rene.causse@ircam.fr

ABSTRACT

This article discusses progress and advances in the Acoustic-Aggregate-Synthesis project (first described in Proceedings from ICMC, 2012).

Acoustic-Aggregate-Synthesis *is a real-time performance-tool which fuses synthetic and acoustic sound sources in order to achieve semi-acoustic re-synthesis of a pre-defined acoustic model. This technology functions most effectively, from a cognitive stand-point, when it is used with an instrument which has been modified in order to allow synthesis to emanate via the same channels as those from which that instrument's acoustic signal emanates (i.e. with electronic diffusion of synthesis taking place inside the instrument itself). Thus, the project comprises elements of both software development and instrument modification. At the heart of this initiative is the desire to maintain the acoustic amplification & diffusion patterns, attack/sustain/release characteristics, etc. of a given instrument whilst overriding its timbral characteristics in favour of a contrasting, secondary tone.*

1. BACKGROUND

The synthesis aspect of this project is essentially a specialised application of additive timbre-frame concatenation. The nature of additive-synthesis is well-known and its use (at least since the mid-1990's when computers offering the processing-power required for such operations became widely available) is commonplace. Nonetheless, while this type of synthesis has proved effective in producing complex timbres in a variety of contexts, the notion of *unifying* acoustic & additive-synthetic sound-sources in the creation of hybrid timbres with historically-defined semantic characteristics (*e.g.* recognisable as a flute) remains a fecund territory for creative exploration.

1.1 Challenges

In order to effectively *fuse* two contrasting timbres into a unified, hybrid timbre, two key problems must be addressed:

1.1.1 Perceiving two sounds as a singular entity

Acoustic instruments have idiosyncratic, *perceptible* patterns of sound-diffusion. It is often observed that the sound of a loudspeaker replaying, for example, a recording of a violin, is clearly distinguishable in most acoustic-environments from that of a violin being played live.[1] No loud-speaker, regardless of shape, size or quality, is capable of diffusing sound in quite the same way as the body of a violin; indeed, juxtaposing sound-sources such as a violin and a violin being played back through a loud-speaker will, in most situations, result in the listener perceiving two distinct sound-sources, regardless of the placement of one sound-source relative to the other. In this case, assuming a good-quality loud-speaker is being used, acoustic diffusion alone distinguishes the two. As such, in order to achieve a genuine fusion of two sound sources -one acoustic and one synthetic- the latter must be diffused not only from, as much as possible, the same physical space as the former, but *via the same channels*. In the case of wind instruments, for example, the synthetic sound-source must therefore pass through the instrument's tubing (subjecting it to the instrument's internal reverberations and latent amplifications/attenuations) before being diffused to the exterior through the same channels as those used by the acoustic signal (subjecting it to the instrument's modes of acoustic diffusion).

Whilst modifying instruments permanently allows for highly tailored, even drastic alterations, it is, for two reasons in particular, undesirable: beyond the obvious question of cost, there is the fact that many professional musicians feel uncomfortable performing on an instrument with which they are unfamiliar. Therefore, we have opted rather for the redesign *only of removable components* (of which most instruments have several); in this way, performers are able to use their preferred instruments, simply adding or removing these ad-hoc parts as required; to do so takes little more than a few seconds.

Our project in its present phase is focused on the extended bass-clarinet and the extended tenor trombone. The modification of each instrument, as we shall see, called for a unique approach. Indeed, the process of extending one instrument-type is seldom applicable to another; regardless of

[1] The term 'live' here implies that the instrument is being performed acoustically. It is also implicit that the sound reaches the listener *directly*, i.e. not primarily after reverberating on other surfaces.

which instrument one sets out to modify, the task inevitably calls for refined and creative engineering.

1.1.2 Dual timbres with a propensity to merge

The second major challenge was in the re-synthesis of the target timbre itself; for two timbres to merge perceptually, their fundamentals (as well as a number of other components) must be uniform. The approach that was pursued in addressing this issue is discussed in section 2.1 of this article.

2. ACOUSTIC-AGGREGATE-SYNTHESIS

Acoustic-aggregate-synthesis (henceforth 'AAS') makes a real-time comparison between an incoming signal and an *instrumental-template* (in actuality a list of data describing the 64 most prominent sinusoidal components of a given timbre, created in deferred-time).

2.1 Instrument templates

The terms 'instrument-template' and 'acoustic-model' are used somewhat interchangeably in this piece of writing. The data contained in the instrument-template is, in fact, a multitude of acoustic-models created in the analysis of sounds from a single acoustic sound-source. Typically, each and every note of a given instrument's register[2] is recorded at three different dynamic levels (*pp, mf* & *ff*) and described in a list of frequencies/intensities. This data is indexed and compiled into an external file, or instrument-template.

2.1.1 Managing pitch

Given that the f^0 of the incoming signal will invariably differ to some extent from that of the corresponding note in the acoustic-model, it is necessary to apply, in real-time, a process of multiplication to the frequencies of all acoustic-model components. The acoustic-model data is therefore subject to constant micro-transpositions in order to follow the contour of minor deviations in pitch (as with vibrato) in the incoming signal. If the incoming signal's f^0 deviates significantly enough from that of the acoustic-model, the system will recognise it as a new and different note, and a new acoustic model will be loaded. This process of frequency-matching is essential if acoustic and synthetic sound sources are to become unified. A common, uniform movement in pitch between two sound-sources encourages the ear to merge the two entities into one.

2.1.2 Managing intensity

If we are to re-synthesise an acoustic-model in such a way that it is recognisable and convincing, then that model must emulate not only the frequency, but also the intensity of the incoming signal. This does not merely imply that a static acoustic model should simply be played louder. To do so would emphasise the synthetic nature of the sound. Rather, we must take into account changes in timbre which occur in an instrument as a consequence of changes in volume.

As explained in the description of instrument-templates, data is gathered for a given acoustic sound-source at multiple intensities. The global-average intensity of the incoming acoustic signal is measured, and this value is sent to an algorithm which interpolates between acoustic-models of different intensities. As such, the re-synthesised sound is, in actuality, constantly being calculated from points of intensity in-between those at which acoustic models were originally generated.

2.1.3 Process of comparison

In response to each component detected in the incoming sound-source, one of the following three outcomes occurs:

1.) in the event that the incoming signal contains a component which is present in the instrumental-template (i.e. is deemed to be within sufficient proximity to a component, as determined by the *margin-of-frequency-deviation* variable) but of lower intensity, that difference in intensity is calculated and subsequently used to determine intensity of the electronic diffusion of that frequency;

2.) in the event that the incoming signal contains a component which is absent in the instrumental-template, nothing is diffused; in contrast, if a harmonic component which is present in the template and absent altogether in the incoming signal, that component is diffused at its full, original value.

3.) in the event that the incoming signal contains a component which is present (or is deemed to be within sufficient proximity to a component) in the instrumental-template but of greater intensity, nothing is diffused;

Thus, in the re-synthesis of the instrumental-template model, a variable proportion is generated acoustically (i.e. by the instrument providing the acoustic signal), and the remainder, synthetically. The process is summarised in the following FIGURE:

[2] For 'every note', read each chromatic pitch; microtonal alterations are not taken into account.

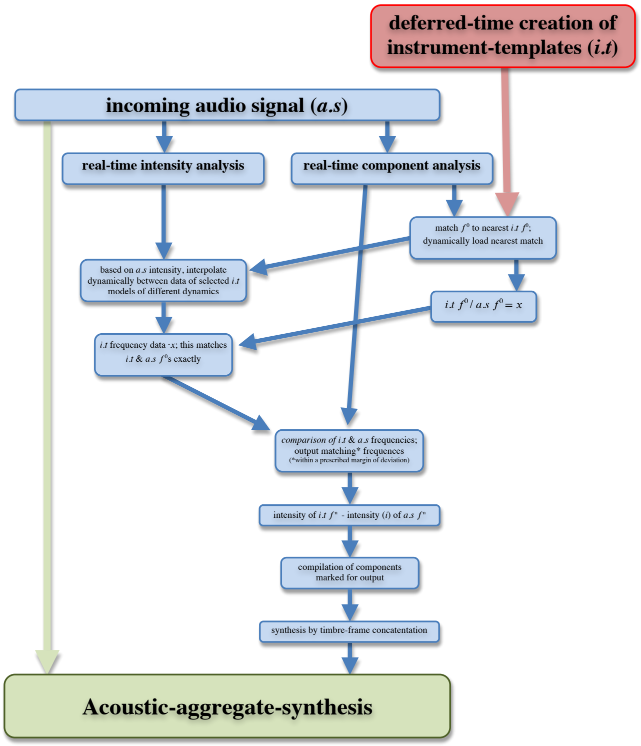

FIG 1. A simplified flowchart describing the process

2.2 From simple to complex timbres

As the reader might intuit, situations in which an incoming timbre is globally 'less-complex' than the timbre one is attempting to re-synthesise yield the most effective results. It is, of course, far more practicable to add sinusoidal components to a sound (as, for example, with adding the components necessary to render the timbre of a trombone like that of a bassoon) than to subtract them (as would be necessary to perform this operation in reverse, assuming that one does not significantly distort the global intensity[3]).

Through both a.) an amplification of sinusoidal components which are present in the acoustic signal but at a lower intensity than corresponding components in the acoustic-model, and b.) the addition of components which are absent in the acoustic signal but present in the acoustic-model, we may, in most situations, succeed in synthesising a timbre which is recognisable as that acoustic model, provided the rule of 'simple-to-complex' is respected. To give an example, if we use an acoustic-model of an oboe [as represented in FIGURE 3], and a flute as acoustic sound-source [FIGURE 2], by enhancing harmonics 2-10 and 'inserting' harmonics 12, 13, 14, 18, & 21 etc., we will readily achieve an aggregate-timbre which is immediately recognisable as an oboe [represented in FIGURE 4].

[3] See **sub-section 2.3** for a discussion of this principle

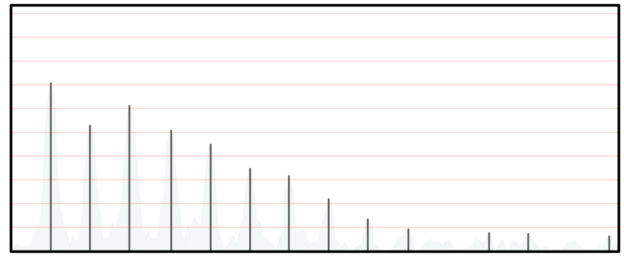

FIG 2. A simplified timbral profile of a **flute (C5, *mezzo forte*)**; in the example described here, this timbre constitutes the incoming-signal

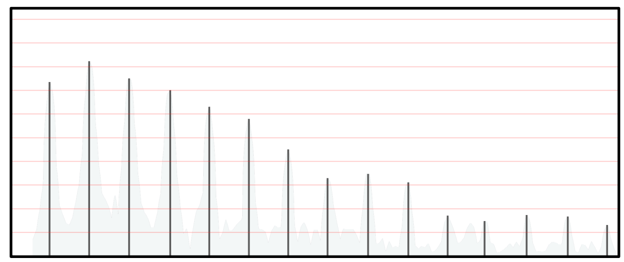

FIG 3. A simplified timbral profile of an **oboe (C5, *mezzo forte*)**; in the example described here, this timbre serves as the acoustic-model; the global average-intensity is therefore adjusted to correspond with that of the incoming, acoustic signal.

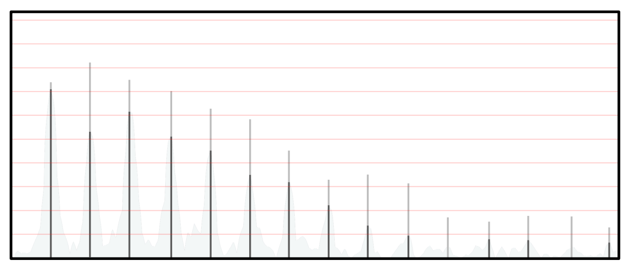

FIG 4. Illustration of the aggregate-timbre. The incoming, acoustic signal is presented in dark grey; component reinforcement/insertion is shown in light-grey.

2.3 From complex to simple timbres

Of course, we cannot effectively render sounds produced by the acoustic sound-source inaudible;[4] therefore, the transformation of a sound with a complex timbre, such as that of an oboe, into that of a flute (with a relatively less complex timbre) poses a real problem. The solution we have pursued in seeking to address this issue lies in the question of relative-intensities. If we define a value which describes the intensity of the harmonic from the incoming oboe which deviates *most*, in terms of intensity, from the corresponding harmonic in the flute (as one may see by comparing FIGURES 5 & 6, the variance between flute's idiosyncratically

[4] Attempts to attenuate sinusoidal components through phase-opposition have demonstrated results which were too inconsistent to be useful.

weak 6th harmonic and the oboe's relatively prominent corresponding harmonic, stands out as the greatest point of divergence between the two instruments) and multiply all data describing intensities in our acoustic model by that value -such that the sixth harmonic of the flute is amplified by such a factor that it matches the intensity of the corresponding component in the oboe- we arrive at 'reproduction' of the flutes timbral identity, albeit at a significantly distorted global intensity (see FIGURE 7).

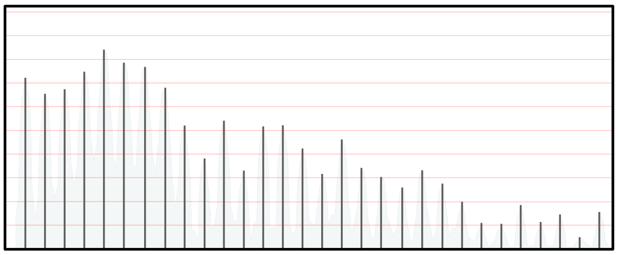

FIG 5. Timbral profile of the **oboe (C4, *mezzo forte*)**; in the example described here, this timbre constitutes the incoming-signal

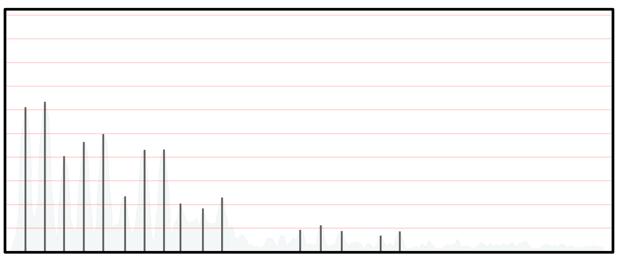

FIG 6. Timbral profile of the **flute (C4, *mezzo forte*)**; in the example described here, this timbre serves as the acoustic-model; the global average-intensity is therefore adjusted to correspond with that of the incoming, acoustic signal.

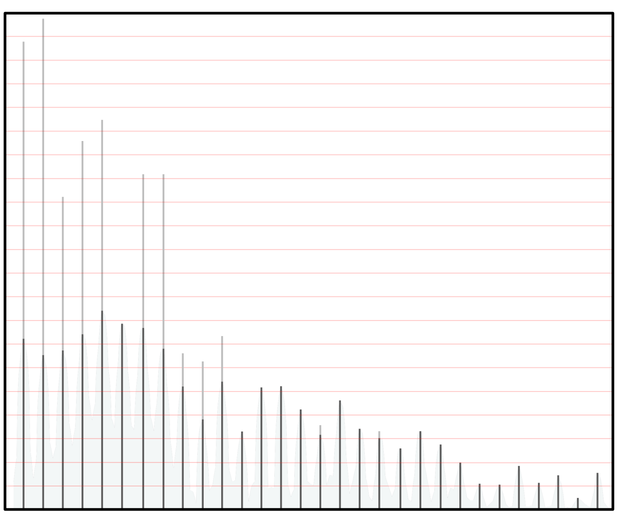

FIG 7. 'Amplified' flute spectrum superimposed over the original oboe, whereby re-synthesis of the flute is achieved through a combination of components generated acoustically and components which are 'reinforced' via the transducer. Here we see the intensities for *each* component which are required to create the desired relationships between harmonics. Components which are reinforced electronically are therefore generated partially by the acoustic instrument (relevant section shown in dark grey), and partially by electronic diffusion (shown in light grey).

As we have seen, in order to respect the ratios among key harmonic-components necessary to produce a timbre which is recognisable as that of flute playing *mezzo forte*, an 'amplification' must be applied to harmonics 1 to 7 (thus also significantly augmenting the intensity globally). The amplification of these components depends upon a diffusion of frequencies which are already present in the oboe spectrum at a greater intensity, thus increasing their total intensity to such a point that they are in proportion to harmonic n°6 of the flute. Obviously, given the limitations of the integrated transducer, this process becomes ineffective once a certain threshold, in terms of the intensity of the incoming acoustic signal, is reached (i.e. once the transducer reaches its peak output before harmonic distortion occurs). Nonetheless, at lower intensities, the process functions effectively, albeit with the necessary consequence of (often drastically) augmenting the intensity of the aggregate timbre (in FIGURE 7 for example, the second harmonic is amplified to approximately 3.3 times the intensity at which it is present in the acoustic signal alone). The result of this phenomenon is that the listener perceives re-synthesis of a flute whose intensity and timbre are not in proportion (since we are conditioned to associate certain instrumental timbres with certain intensities; when these ratios are not respected, the effect, cognitively, is akin to that of artificial amplification).

3. AUGMENTED INSTRUMENTS

As mentioned in the first part of this article, the success of this project was dependent, in no small part, upon the listener perceiving aggregate timbres as homogenous, and not as two separate and distinct sound-sources.

3.1 Diffusion of sounds inside instruments

Certain instruments, owing purely to their construction and physical dimensions, lend themselves more readily to this undertaking.

3.1.1 Piano

The piano is an obvious choice; even with little to no response from the piano's sound-board, a modest-sized transducer is sufficient to produce strong sympathetic resonances which evoke the instrument's own natural sound-decay. Furthermore, sounds diffused inside a piano are, to some extent, diffused outwardly through the same channels as those the instrument produces itself, acoustically (i.e. rebounded outwards from the lid of the instrument). Finally, a transducer may easily be placed inside the piano without any need to physically modify the instrument.

3.1.2 Bowed strings

In the past, many attempts have been made with the bowed strings, with varying degrees of success. The stumbling block is most often that in order to introduce sufficient energy into the rigid wooden face of the instrument so that it resonates (at least in any way which is comparable to the modes of resonance which occur when the instrument is played 'traditionally'), a point of direct contact between the vibrating entity and the wood of the instrument is required. This has the potential to damage the wood, or at least remove the instrument's varnish. Simply directing sound towards the instrument-body, even at considerable intensity, is not sufficient to produce complex nodal vibrations, the very thing which evokes the sought-after idiosyncratic timbre and acoustic diffusion.

3.1.3 Woodwinds

For wind instruments, various possibilities exist. As a general rule, wide-bore tubes which are open at both ends are most apposite. Instruments in this category include the saxophones, clarinets, bassoons and the bass & contrabass flutes. For these instruments, a transducer may be placed at the bottom end of the instrument and directed-inwards; in this case, the sounds diffused from that transducer, whilst entering the tube from the opposite end to those produced at the embouchure, are diffused via the finger-holes, and as such are subject, for the most part, to the same spectral-envelope-filtering as sounds produced at the embouchure. Of course, in partially covering the ending of the instrument's tube, the lowest note in the tessitura will be weakened, if not rendered unusable altogether. However, given that it is with this lowest pitch alone (bass clarinet B♭1, bass flute C3 or B3, etc.) that sound diffuses predominantly from the end of the tube (and not through open finger-holes), and taking into account the potential creative benefits, the authors consider this to be a reasonable trade-off. Furthermore, one can fashion a transducer setup with relative ease which is simply inserted into the instrument itself, much in the same way that a brass player might insert a mute into the bell, thus avoiding the need to modify this part of the instrument.

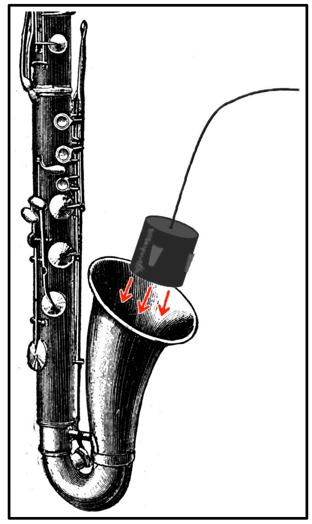

FIG 8. Sketch for an ad-hoc device fashioned to suit the bell of a bass-clarinet; the outside of the transducer casing may make use of pieces of cork to ensure it stays in place, much in the same way a mute does with a brass instrument. In this way, the transducer may easily be inserted & removed in a performance situation. [NB. PHOTO OF DEVICE WILL BE AVAILABLE AT THE TIME OF PUBLICATION]

3.1.4 Brass

Brass instruments are more challenging, given that a.) in the case of trumpets, horns and trombones, the gauge of the tubing is relatively narrow - certainly too narrow to insert a transducer of any consequence without completely blocking the tube (which would, of course, render the instrument inoperable for anything save the diffusion of electronic sounds), and b.) there are no openings of sufficient size to allow an injection of sound.

In the past, attempts have been made to diffuse sounds via a relatively small transducer placed inside the bell and directed outwards. Whilst this does address the issue of directionality (i.e. acoustic/instrumental sounds and electronically-diffused sounds appear to emanate from the same source), the transducer can not be positioned very deeply inside the instrument; the synthetic signal, therefore, does not pass through a sufficient length of the instrument's tubing for it adopt significant characteristic acoustic diffusion. Furthermore, the presence of the loud-speaker in this position attenuates the instrument's acoustic signal by simple virtue of the fact that it blocks a significant part of the opening.

As such, the diffusion of sounds inside brass instruments via a transducer necessitates that part of the instrument be physically, and permanently, modified. As mentioned, it is desirable to apply any modifications only to small sections of the instrument, and at that, those which may be removed easily and/or substituted in a performance setting. In the case of trumpets and trombones, the most logical component to modify is the tuning slide; this part may be purchased independently, and provided that the instrument is not a custom design, constitutes an 'interchangeable' part (i.e. two examples from the same manufacturer should be of precisely the same dimensions).

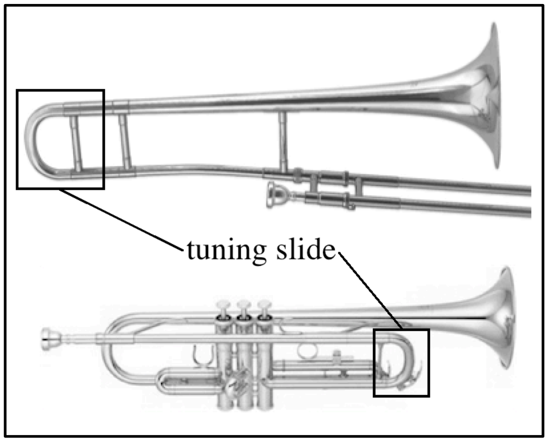

FIG 9. Tunings slides shown on a tenor ('straight') trombone and a trumpet in B♭

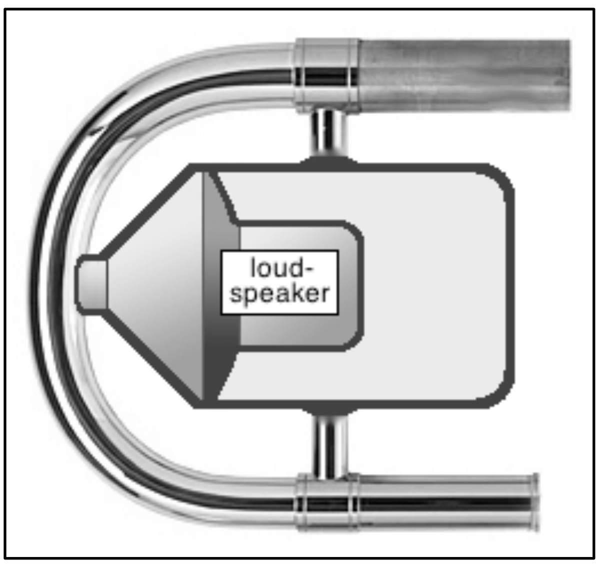

FIG 10. Illustration of an encased loud-speaker which has been permanently integrated into a project-dedicated trombone tuning-slide [NB. PHOTO OF DEVICE WILL BE AVAILABLE AT THE TIME OF PUBLICATION]

The ad-hoc extension to the trombone illustrated in FIGURE 10 directs output from a transducer into the instrument-tube. Since the apparatus is entirely sealed, the only affect on the instrument is a small increase in the volume of the tube (not more than 3-4cm^3), and therefore, a lowering in intonation (in our tests, this has not caused any perceptible impairment to the operation of the instrument).

3.2 Capturing sounds inside instruments

Given that sounds are being diffused inside instrument-tubes, the problem of microphone placement is not insignificant if we are to obtain an 'unpolluted signal'; beyond obvious problems that would arise from feedback, the signal which is sent into the system for analyses must be that of the instrument alone (and free from the presence of sounds emanating from the integrated transducer). The problem, once again, becomes one of engineering and the creation of ad-hoc instrumental components.

3.2.1 Pre-existing solutions

Compared to the diffusion of sounds inside instruments via a transducer, capturing the signal of a wind-instrument by integrating a microphone into the mouthpiece has fairly broad applications and a number of systems which integrate a piezo microphone into the component closest to the embouchure (notably for bassoon, clarinet in B♭, trumpet & trombone) are sold commercially, albeit usually on a small scale.

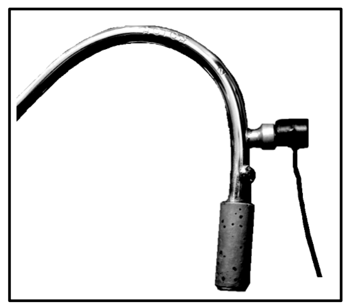

FIG 11. A commercially-available bassoon bocal which has been adapted to allow a piezo microphone to be attached.

3.2.2 Optimising signal quality

The extent to which a clean signal may be obtained varies from one instrument-type to the next. In the case of the bass clarinet, the vast majority of sound-energy emanating from the loud-speaker has diffused *out* of the instrument before reaching the neck-piece.[5] With brass instruments, given that the output of the loudspeaker is funnelled *straight* in to the tube (i.e. the signal is directed neither towards the bell *nor* towards the mouthpiece), when the instrument is not being played, roughly an equal proportion of the sound-energy will travel in either direction. Whilst the absence of holes in the instrument-tube assures that sounds diffused from the loud-speaker reach the microphone with sufficient intensity as to be problematic, we must remember that, when the instrument is being played, the movement of air through the tube will incite the loud-speaker signal to favour movement

[5] Unless of course all fingering-holes are closed, thus leaving the sound nowhere to go but directly towards the embouchure.

towards the bell. As a consequence, a relatively small proportion of sound-energy from the transducer arrives at the mouthpiece.[6]

The intensity of sound to which a microphone placed very close to the embouchure is subject further facilitates the task. Any sounds from the transducer spilling 'upwards' towards the microphone are of a vastly inferior volume to those being produced locally. Given that the sounds detected from the microphone are sent into the system and analysed in order to determine the 64 most prominent sinusoidal components, the presence of contaminant-sound at a far lower level is, in most cases, inconsequential. Finally, the signal received from the microphone must be filtered and balanced to some extent (predominantly to minimise distortion, but also to replicate component amplification/attenuation which occurs within the instrument itself). These factors ensure that a clean, useful input signal may be reliably obtained.

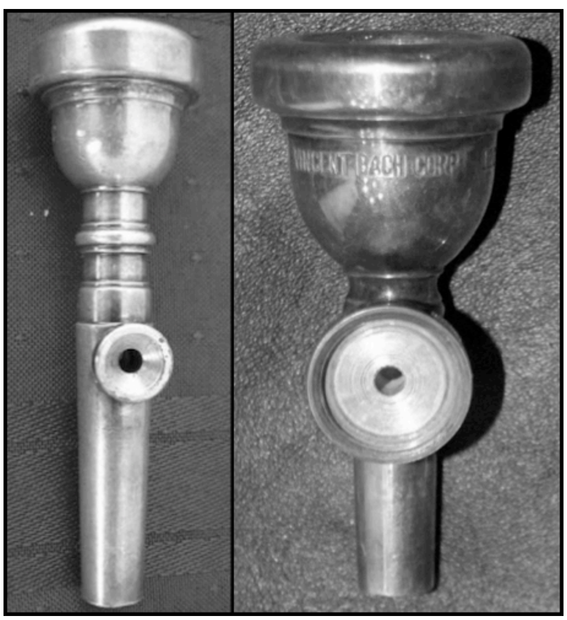

FIG 12. Some commercially-available mouthpieces with an integrated opening to which a piezo microphone is attached, for trumpet [left] & trombone [right].

4. OTHER CONSIDERATIONS

4.1 Filtering

Now that the essential notions of the project have been discussed, there are a couple of additional aspects which were necessary to consider in order to achieve satisfactory results. Since, as we have seen, synthetic components are diffused inside acoustic instruments, they are inevitably subject to that instrument's tendency, depending upon the fingering/tube-length being used, to reinforce certain frequencies whilst attenuating others (i.e. to alter a sound's spectral envelope). Thus, with anechoic-chamber tests comparing sounds diffused via a loud-speaker *inside* and *outside* of the instrument, we may predict which components may, if desired, be reinforced or attenuated, in order that the resulting output is as faithful a reproduction of the acoustic-model timbre as possible.

5. CONCLUSIONS

Having achieved extremely promising results at the time of writing this article, the next step is to encourage composers to make use of the technology in new works of music.

Given the nature of this technology, the optimal setting for its use is with instrumental solos or works for small chamber groups. Within the context of a large ensemble setting, with the exception of concerto-style works, the technology is unlikely to yield very consequential results. In order to maximise the efficacy of the technology, the listener should be conscious which instrument is playing. In this way, modifications to that instrument's behaviour are most striking.

Both the software/synthesis and the instrument-augmentations aspects of project are in a continual process of evolution and refinement. It should be said that each of these two dimensions has the potential to be used independently of the other if results other than those intended by the authors are sought. Misuse of these tools is encouraged!

Acknowledgments

The authors gratefully acknowledge the support of Benny Sluchin, Alain Billard, Arshia Cont, Axel Röbel, Sean O'Leary & Alain Terrier.

---

[6] Interestingly, when using the system the trombonist clearly felt vibrations on his lips from sounds diffused via the loud-speaker, even though the integrated microphone did not pick them up in any significant sense.