DOES THE FINISH MATTER IF NO ONE IS LOOKING? VISUAL BIAS IN TRUMPET TESTING

Michael Oehler[1], Frauke Brandt[1], Christoph Reuter[2]

[1] *Osnabrück University, Institute of Musicology, Osnabrück, Germany*
[2] *Vienna University, Institute of Musicology, Vienna, Austria*

Abstract

Surface finish is frequently discussed by brass players as a determinant of sound and playability, yet such judgments may be influenced not only by acoustical differences but also by visual information and prior expectations. This exploratory pilot study examines whether perceived differences between three nominally comparable trumpets are larger when their surface finishes are visible and whether corresponding differences are reflected in the acoustic signal. Ten trumpet players performed identical material on three B-flat trumpets of the same model with different finishes (raw brass, silver-plated, lacquered) in blind and open conditions. After each trial, participants rated 13 perceptual dimensions. Across 11 of the 13 dimensions, descriptive differences between the instruments were larger in the open condition than in the blind condition. In the acoustic feature analysis, variability was dominated by residual and participant-related components, whereas the finish-associated component was small. These pilot findings are consistent with the possibility that visual information contributes to instrument evaluation. However, because of the small sample, the pairing of one instrument with each finish, and the exploratory nature of the analyses, the results should not be interpreted as demonstrating a causal effect of surface finish or visual expectation.

Keywords: *trumpet, surface finish, timbre perception, visual influence, blind testing, pilot study*

1. Introduction

Brass players routinely describe instruments using perceptual terms such as bright, warm, focused, rich, responsive, or resistant. Yet the scientific study of instrument evaluation shows that such judgments are neither trivial nor purely auditory. Timbre is a multidimensional percept, and performer-centered evaluation of an instrument extends beyond radiated sound to include dimensions such as control, flexibility, resistance, and response [1], [2], [3]. From the player's perspective, instrument assessment is therefore inherently multisensory, combining auditory, haptic, and proprioceptive information [4].

This is particularly important for trumpet research, because several commonly used evaluation terms are not equally reliable. In large-scale blind tests with trumpet players, Bertsch, Waldherr, and Kausel showed that overall instrument classification, brilliance, output power, dynamic range, and some flexibility-related judgments can be assessed comparatively consistently, whereas terms such as general response, resistance, and certain tone-color judgments are far less stable across repetitions [5], [6]. These findings suggest that studies of trumpet quality should not simply collect verbal ratings, but should also consider which perceptual dimensions are robust enough to support meaningful comparisons.

A second consequence is methodological. If some player judgments are already variable under blind conditions, then open evaluations may be influenced not only by acoustic and haptic feedback, but also by expectation and visual knowledge. More generally, music psychology has repeatedly shown that visual information can bias evaluative judgments. Tsay's "sight-over-sound" findings suggested that observers may rely surprisingly strongly on visual cues when judging musical performance [7]. Recent work further shows that such visual dominance is context-dependent and shaped by musical expertise: in brass-band settings, for example, the effect appears in some evaluator groups but not in expert brass-band musicians [8], [9]. For trumpet players, surface finish is an especially salient visual cue, because raw brass, silver plating, and lacquer evoke strong expectations about brightness, projection, resistance, and overall quality. At the same time, the acoustical literature cautions

Corresponding author: *michael.oehler@uos.de.*

against reducing such expectations either to myth or to simple physical determinism. Controlled studies on wind instruments have shown that material and coating effects must be examined under carefully blinded conditions. Coltman's study on flutes found that, once visual and tactile cues were removed, players and listeners could not reliably identify otherwise comparable instruments by material alone [10]. In brass acoustics, earlier work on surface coatings and later studies on trumpet bell vibrations demonstrated that structural properties of the instrument can affect the radiated sound, even if the perceptual relevance of these effects may be subtle and task-dependent [11], [12]. Moreover, psychophysical studies on nominally identical trumpets have shown that measurable acoustical differences do not necessarily lead to clear perceptual discrimination for all players or listeners [13], [14].

For this reason, combining subjective evaluation with objective acoustical analysis is essential. Previous trumpet studies have already linked reliable sensory judgments to features of the instrument's input impedance, and more recent work has renewed interest in data-driven characterization of trumpet timbre and performance descriptors [6], [15], [16], [17]. Acoustic analysis of recorded performances can therefore provide an important complement to player ratings by testing whether perceived differences are mirrored in measurable spectral and temporal properties of the produced sound.

The present exploratory pilot study brings together three elements that have previously been investigated largely separately: (i) the reliability of trumpet evaluations under blind conditions, (ii) perceptual consequences of small acoustical differences between nominally comparable instruments, and (iii) material or coating effects in controlled settings. To our knowledge, few studies have compared player evaluations of trumpets under blind and open conditions while also relating the resulting rating patterns to acoustic features of the produced sound.

To this end, the present study compares three nominally comparable B-flat trumpets of the same model and maker, each with a different surface finish (raw brass, silver-plated, lacquered). Each player evaluated all three instruments under both blind and open conditions. By combining within-subject ratings with a feature-based acoustic analysis of the recorded performances, the study explores whether the availability of visual information is associated with larger perceived differences between the instruments and whether instrument-associated differences are also apparent in the acoustic features.

2. Method

2.1. Participants

Ten trumpet players participated in the study. The mean age was 28.9 years (SD=12.1) and the mean playing experience was 20.2 years (SD=12.3). All participants completed all instrument-by-condition combinations.

2.2. Instruments and conditions

Three THEIN Sibirius B♭ trumpets were used. The instruments were of the same model and maker, with one instrument in each of three surface finishes: raw brass, silver-plated, and lacquered. They were treated as nominally comparable instruments; however, because only one instrument represented each finish, instrument identity and surface finish were not independently manipulated. The same mouthpiece was used across trials. The study used a within-subject design with two visibility conditions. In the blind condition, participants wore a blindfold and did not know which trumpet they were playing. In the open condition, the instrument was visible. The order of the three finishes was counterbalanced across participants and differed between blind and open blocks.

2.3. Musical material and recording

Each participant first warmed up on their own instrument. The test material consisted of sustained natural tones, natural tones with the second valve, a scale played in mezzo-forte with staccato and legato articulation, and two simple melodies. A metronome was presented via headphones. Audio was recorded during the trials using a stand microphone approximately 1.5 m in front of the player and an additional microphone at the bell. The present paper reports the subjective ratings and the acoustic analysis of the recordings.

2.4. Rating scales

After each trumpet trial, participants rated the instrument on 13 seven-point bipolar scales. The scales covered sound-related dimensions and playability-related dimensions. Higher values correspond to the right-hand anchor of each scale. Blowing resistance was included as a player-centered scale, although its midpoint may be interpreted as close to optimal rather than monotonically positive. The rated dimensions were brightness, warmth, focus, sharpness, homogeneity, timbre, sound volume, overtone richness, response, blowing resistance, intonation stability, dynamic control, and overall impression.

2.5. Acoustic analysis

The recorded performances were segmented into individual note events corresponding to the musical material (sustained tones, scale tones, and melodic notes). Each segment was annotated with its nominal pitch, participant, trumpet finish, and experimental condition.

The resulting dataset comprised 1033 note-level observations from 10 participants across 10 nominal pitch categories (A♯3, A♯4, C4, C5, D4, D5, F4, F5, G4, G5). The number of segments per pitch ranged from 57 to 148, reflecting the distribution of notes in the musical material.

For each segment, a large set of 133 acoustic features was extracted using the SInES Tools suite [18]. The feature set included spectral descriptors (e.g., spectral centroid, bandwidth, and spectral contrast), cepstral and formant-related descriptors (e.g., MFCCs and formant-related measures), temporal descriptors (e.g., attack time and spectral flux), and energy-based measures across frequency bands.

Feature extraction was performed consistently across all recordings, and features were retained at the segment level. Only observations with valid feature values were included in the analysis.

The resulting dataset therefore consisted of high-dimensional feature vectors describing each note segment, enabling a feature-based analysis of acoustic variability across pitch, performer, and instrument finish.

2.6. Statistical analysis

For each item, a linear mixed-effects model was fitted with trumpet finish, visibility condition, and their interaction as fixed effects and participant as a random intercept:

$$\text{rating} \sim \text{trumpet} \times \text{condition} + (1|\text{participant}). \quad (1)$$

As a descriptive summary of perceptual differentiation, the mean absolute trumpet difference was computed as the average absolute value of the three pairwise differences among the trumpet cell means.

Planned contrasts examined condition-related and pairwise instrument differences. P-values were Holm-corrected within each rating item. A global Holm correction across all reported contrasts was additionally used when interpreting the item-specific findings. Model-based effect sizes were computed as

$$d_{\text{model}} = \frac{\hat{\beta}_{\text{contrast}}}{\sqrt{\sigma^2_{\text{residual}}}}. \quad (2)$$

Acoustic variability was analyzed at the feature level using linear mixed-effects models:

$$\text{feature} \sim \text{finish} + \text{pitch} + (1|\text{participant}). \quad (3)$$

Pitch was included as a categorical fixed effect to control for systematic differences across fundamental frequency. Descriptive variance contributions were approximated by isolating each fixed effect in the model predictions, while participant and residual variances were obtained from the fitted models. These values are used as descriptive summaries of the fitted model components rather than as exact or causal variance decompositions.

3. Results

3.1. Perceptual results

The descriptive distance analysis showed a consistent directional pattern. For 11 of the 13 rating dimensions, the mean absolute difference between instruments was larger in the open condition than in the blind condition (Table 2). The largest increases occurred for response ($\Delta = 0.60$), brightness and warmth (both $\Delta = 0.47$), and focus, sharpness, and overtone richness (all $\Delta = 0.40$). Only sound volume and intonation stability showed smaller trumpet differences in the open condition than in the blind condition.

The model-based effect-size summaries showed a similar descriptive pattern. Across all pairwise trumpet contrasts, the mean absolute model-based effect size was larger in the open condition (mean $|d_{\text{model}}| = 0.58$, median $= 0.60$) than in the blind condition (mean $|d_{\text{model}}| = 0.33$, median $= 0.23$). Using the conventional reference threshold of $|d_{\text{model}}| \geq 0.8$, 12 contrasts exceeded this threshold in the open condition, compared with 5 in the blind condition

Table 1: Summary of absolute model-based effect sizes.

Analysis	Mean	Median	Max	$n_{\|d\| \geq .8}$
Open vs. blind	0.46	0.49	1.19	6
Trumpet, blind	0.33	0.23	1.09	5
Trumpet, open	0.58	0.60	1.29	12

Exploratory item-level contrasts indicate where the descriptive differences between conditions were most pronounced. For homogeneity, raw brass was rated lower in the open condition than in the blind condition (estimate = -1.20, $d_{\text{model}} = -1.19$, within-item Holm $p = .024$). In the open condition, the silver-plated trumpet was rated higher than the raw-brass trumpet for homogeneity (estimate = 1.30, $d_{\text{model}} = 1.29$, within-item Holm $p = .024$). In the open condition, the lacquered trumpet was rated lower than the silver-plated trumpet for overtone richness (estimate = -1.20, $d_{\text{model}} = -1.22$, within-item Holm $p = .040$) and response (estimate = -1.30, $d_{\text{model}} = -1.28$, within-item Holm $p = .025$). None of these item-specific findings remained significant after global Holm correction across all reported contrasts. They are therefore treated as exploratory observations rather than confirmatory evidence.

3.2. Acoustic analysis

To assess whether perceptual differences between instruments are reflected in the recorded sound, a feature-based acoustic analysis was conducted. A set of 133 spectral, temporal, and perceptual descriptors was extracted, and for each feature a linear mixed-effects model with finish and pitch as fixed effects and participant as a random intercept was fitted. Variance contributions were estimated from model components and predictions.

Across all features, variability was overwhelmingly dominated by residual variance (mean = 94.35%), with substantially smaller contributions from participant (5.22%) and finish (0.43%) (Table 3). Thus, performer-related variance exceeded finish-related variance by approximately an order of magnitude. The feature-level distribution of participant- and finish-associated variance components is shown in Fig. 2.

The finish-associated model components were generally small relative to the residual component and, on average, also smaller than the participant-related component. The acoustic results should therefore be interpreted primarily as descriptive evidence of limited

Table 2: Mean absolute trumpet differences by item and condition. Positive Δ values indicate larger trumpet differences in the open condition.

Item	Blind	Open	Δ	Item	Blind	Open	Δ
Brightness	0.20	0.67	0.47	Sound volume	0.13	0.07	-0.07
Warmth	0.13	0.60	0.47	Overtone rich.	0.40	0.80	0.40
Focus	0.40	0.80	0.40	Response	0.27	0.87	0.60
Sharpness	0.13	0.53	0.40	Blow. resistance	0.20	0.33	0.13
Homogeneity	0.73	0.87	0.13	Intonation stab.	0.73	0.47	-0.27
Timbre	0.60	0.80	0.20	Dynamic control	0.40	0.47	0.07
				Overall impr.	0.13	0.40	0.27

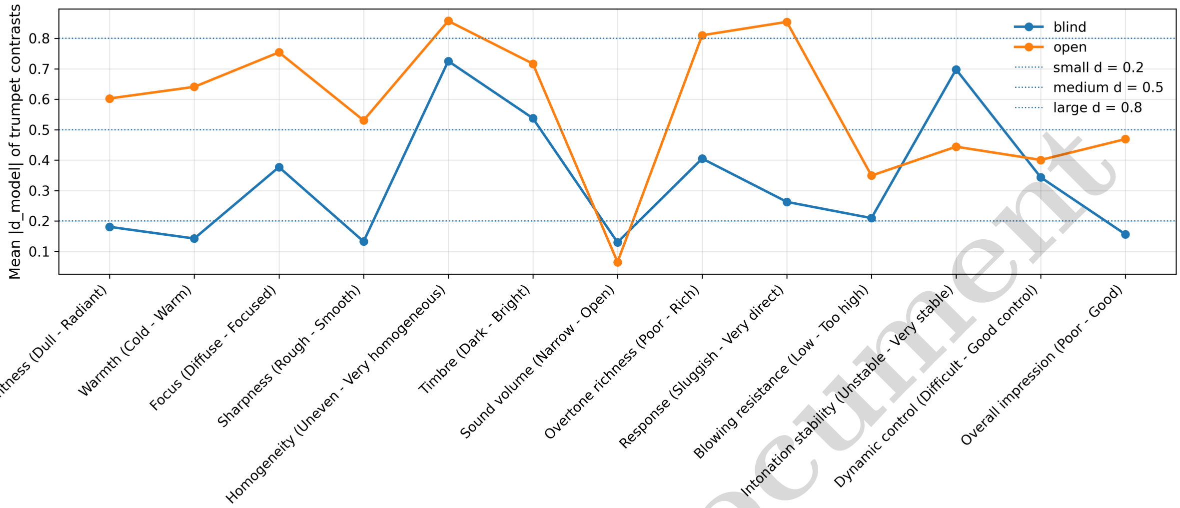

Figure 1: Mean absolute model-based effect sizes for trumpet contrasts by item and condition (blind vs. open). Horizontal reference lines indicate common small, medium, and large thresholds.

instrument-associated variability within the present feature set.

3.2.1. Acoustic effects in relation to perceptual findings

Exploratory feature-level models yielded nominal finish-associated differences for a small subset of features, generally with low effect sizes. Figure 3 presents the relationship between the finish-associated model component and the corresponding model-based effect-size summary.

In the exploratory pairwise comparisons, eight nominally significant contrasts across six features involved the silver-plated instrument, whereas no nominally significant contrasts were observed between the raw-brass and lacquered instruments. Because a large number of acoustic features and pairwise comparisons were examined, these findings are not interpreted as confirmatory evidence for specific finish effects.

Taken together, the acoustic analyses suggest that instrument-associated differences within the present feature set were small relative to residual variability. They do not provide a clear acoustic counterpart to the larger descriptive rating differences observed in the open condition.

Table 3: Mean variance contributions of performer, finish, and residual factors across all acoustic features. Standard deviations are given in parentheses.

Source	Mean (%)	SD (%)
Residual	94.35	(14.39)
Performer (participant)	5.22	(14.44)
Finish	0.43	(0.83)

4. Discussion

The present exploratory pilot study combined player ratings under blind and open conditions with a feature-based analysis of recorded trumpet performances. Across 11 of the 13 rating dimensions, descriptive differences between the three instruments were larger when the instruments were visible. Model-based effect-size summaries showed a similar directional pattern. At the same time, the acoustic feature analysis was dominated by residual variability, while the average instrument-finish-associated component was small.

The perceptual pattern is consistent with the possibility that visual information contributes to the evaluation of musical instruments. Visual appearance may serve as a contextual cue and may activate expectations concerning qualities such as brightness, projection, resistance, or overall quality. This interpretation is compatible with previous research showing that visual information can influence musical evaluation, although the size and direction of such effects depend on the task and the evaluator group [7], [8], [9].

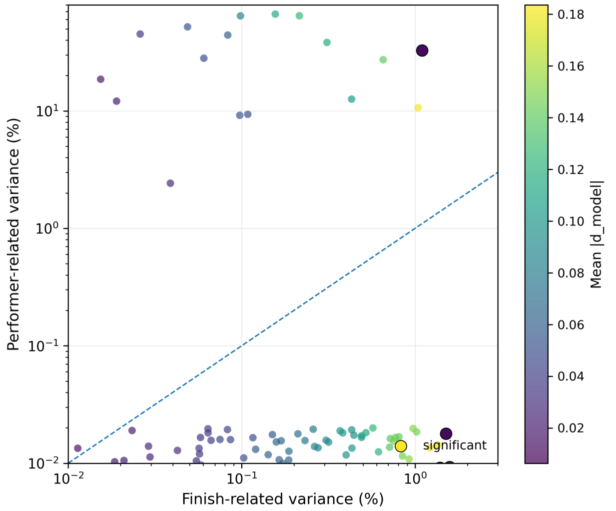

Figure 2: Feature-level comparison of finish- and participant-related variance contributions on log–log scales. Points above the diagonal indicate a larger participant-related component, whereas points below the diagonal indicate a larger finish-related component.

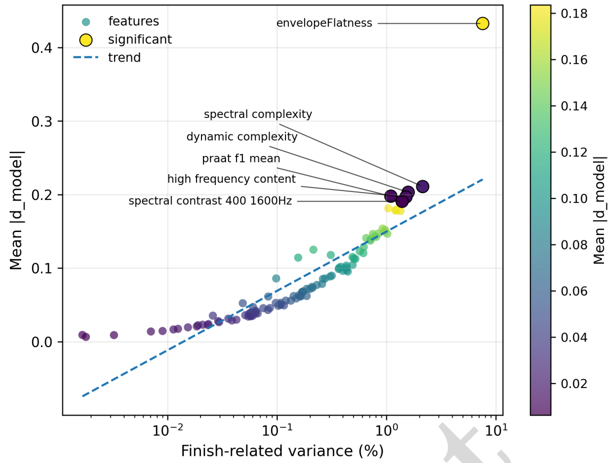

Figure 3: Descriptive relationship between the finish-associated variance component and the mean absolute model-based effect-size summary across acoustic features.

The findings also complement earlier blind-test research on trumpet evaluation. Previous studies have shown that some player judgments are more reliable than others under blind conditions [5], [6]. In the present data, the availability of visual information was associated with larger descriptive differences across most rating dimensions. However, the current design does not establish whether this pattern resulted specifically from expectations concerning surface finish, from recognition of individual instruments, from response tendencies, or from other differences between the blind and open testing situations.

The acoustic analyses showed a different pattern. On average, the model component associated with instrument finish was small relative to residual variability and participant-related variability. The exploratory feature-level comparisons did not provide clear confirmatory evidence for specific finish-related acoustic effects. This general asymmetry is compatible with previous findings that measurable differences between instruments do not necessarily correspond directly to robust perceptual discrimination [10], [13], [14].

The perceptual and acoustic findings should therefore not be interpreted as demonstrating that visual expectation caused the observed rating differences. Rather, the results suggest that instrument evaluations made under open and blind conditions may differ even when the acoustic differences identified by the selected feature analysis are small. This highlights the value of reporting testing conditions explicitly when interpreting player-based instrument evaluations.

Several limitations restrict the conclusions that can be drawn. First, the sample consisted of only ten players and was heterogeneous with respect to age and playing experience. Second, only one instrument represented each surface finish. Instrument identity and finish were therefore confounded, and observed instrument-associated differences cannot be attributed specifically to surface treatment. Third, the blindfold manipulation changed the overall testing situation and not only access to information about finish. Fourth, the acoustic analysis involved a large number of features and repeated note-level observations; the feature-level results should consequently be regarded as exploratory and descriptive. Fifth, some rating dimensions may overlap conceptually, and the present study examined evaluations by players rather than independent listeners.

Future studies should include multiple instruments for each finish, larger and more homogeneous participant groups, and experimental manipulations that separate visual appearance from instrument identity. They could also examine whether prior beliefs about particular finishes predict the magnitude or direction of open-condition ratings.

5. Conclusion

In this exploratory pilot study, descriptive differences between three nominally comparable trumpets were larger across most rating dimensions when the instruments were visible than when they were evaluated under blind conditions. In contrast, the acoustic feature analysis was dominated by residual variability, while the average instrument-finish-associated component was small.

The findings are consistent with the possibility that visual information contributes to player-based instrument evaluation. However, the study does not establish a causal effect of surface finish or visual expectation, because of the small sample and the confounding of finish with individual instrument identity.

References

[1] C. Saitis and K. Siedenburg, "Brightness perception for musical instrument sounds: Relation to timbre dissimilarity and source-cause categories," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2256–2266, 2020. DOI: 10.1121/10.0002275

[2] H. Lee and D. Müllensiefen, "The timbre perception test (tpt): A new interactive musical assessment tool to measure timbre perception ability," *Attention, Perception, & Psychophysics*, vol. 82, no. 7, pp. 3658–3675, 2020. DOI: 10.3758/s13414-020-02058-3

[3] C. Fritz and D. Dubois, "Perceptual evaluation of musical instruments: State of the art and methodology," *Acta Acustica united with Acustica*, vol. 101, no. 2, pp. 369–381, 2015. DOI: 10.3813/AAA.918833

[4] C. Saitis, H. Järveläinen, and C. Fritz, "The role of haptic cues in musical instrument quality perception," in *Musical Haptics*, Cham: Springer, 2018, pp. 73–93. DOI: 10.1007/978-3-319-58316-7_5

[5] M. Bertsch and K. Waldherr, "Reliability of trumpet players' judgments during instrument blind-tests," in *Proceedings of the Second Congress of the Alps-Adria Acoustics Association and the First Congress of the Acoustical Society of Croatia*, M. Horvat and K. Jambrošić, Eds., Opatija, Croatia, 2005, pp. 207–216.

[6] M. Bertsch, K. Waldherr, and W. Kausel, "Sensory evaluation testing of trumpets and correlation with acoustic measurements," in *Proceedings of Forum Acusticum 2005*, Budapest, Hungary, 2005, pp. 575–580.

[7] C.-J. Tsay, "Sight over sound in the judgment of music performance," *Proceedings of the National Academy of Sciences*, vol. 110, no. 36, pp. 14 580–14 585, 2013. DOI: 10.1073/pnas.1221454110

[8] T. Samma, K. Honda, and S. Fujii, "Sight-over-sound effect depends on interaction between evaluators' musical experience and auditory-visual integration: An examination using Japanese brass band competition recordings," *PLOS One*, vol. 20, no. 4, e0321442, 2025. DOI: 10.1371/journal.pone.0321442

[9] K. Vogt, G. Gutzmann, and R. Kopiez, "When eyes outvote ears: Revisiting Tsay's Sight-Over-Sound effect in music performance evaluation," *Frontiers in Psychology*, vol. 17, p. 1 767 475, 2026. DOI: 10.3389/fpsyg.2026.1767475

[10] J. W. Coltman, "Effect of material on flute tone quality," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 520–523, 1971. DOI: 10.1121/1.1912381

[11] R. W. Pyle Jr., "The effect of surface coatings upon brass-instrument tone quality," *The Journal of the Acoustical Society of America*, vol. 69, no. S1, S37, 1981. DOI: 10.1121/1.386293

[12] T. R. Moore, E. T. Shirley, I. E. W. Codrey, and A. E. Daniels, "The effects of bell vibrations on the sound of the modern trumpet," *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 578–589, 2005.

[13] A. Mamou-Mani and D. Sharp, "Evaluating the suitability of acoustical measurement techniques and psychophysical testing for studying the consistency of musical wind instrument manufacturing," *Applied Acoustics*, vol. 71, no. 7, pp. 668–674, 2010. DOI: 10.1016/j.apacoust.2010.01.013

[14] L. Placido, A. Mamou-Mani, and D. Sharp, "Investigating perceptual differences between two trumpets of the same model type," *Applied Acoustics*, vol. 72, no. 12, pp. 907–914, 2011. DOI: 10.1016/j.apacoust.2011.05.014

[15] E. Poirson, J.-F. Petiot, and J. Gilbert, "Study of the brightness of trumpet tones," *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2656–2666, 2005. DOI: 10.1121/1.2006007

[16] A. Acquilino, N. Puranik, I. Fujinaga, and G. Scavone, "A dataset and baseline for automated assessment of timbre quality in trumpet sound," in *Proceedings of the 24th International Society for Music Information Retrieval Conference*, 2023, pp. 684–691.

[17] M. Mohamed, V. Fréour, C. Vergez, K. Arimoto, V. Emiya, and B. Cochelin, "Prediction of trumpet performance descriptors using machine learning," *Acta Acustica*, vol. 8, p. 65, 2024. DOI: 10.1051/aacus/2024042

[18] C. Reuter et al., "SInES tools, helpers, visualizers as your scientific realizers," in *Fortschritte der Akustik – DAGA 2026, 52. Jahrestagung der Deutschen Gesellschaft für Akustik*, Dresden, 2026, pp. 1002–1005. DOI: 10.71568/daga2026.094 [Online]. Available: https://sinestools.univie.ac.at