Sensory Evaluation Testing of Trumpets and Correlation with Acoustic Measurements

Matthias Bertsch
University of Music, Vienna , Institute for Music acoustics,
Singerstrasse 26a, A-1010 Vienna, Austria, **bertsch@mdw.ac.at**

Karin Waldherr
University of Vienna, Department of Psychological Basic Research,
Liebiggasse 5, A-1010 Vienna, Austria, **karin.waldherr@univie.ac.at**

Wilfried Kausel
University of Music, Vienna, Institute for Music acoustics,
Singerstrasse 26a / A-1010 Vienna, Austria, **kausel@mdw.ac.at**

The quality and playability of trumpets depends on several parameters. The aim of the trumpet research project in Vienna (www.bias.at/TRP) was to find objective criteria for brass instruments in sensory evaluation playing tests and their bridging with physical measurements. Therefore more than 250 playing tests have been performed by 55 international, mostly professional trumpet players, and terms describing trumpet response have been correlated with measured physical characteristics of those instruments. To rule out terms inconsistently used by players, repeatability of judgments was investigated using a blind test approach. Interestingly, such commonly used terms like response, resistance or certain tone colours turned out to be completely inconsistent. Fortunately some of the more reliable terms seem to actually describe aspects related to trumpet response (detailed results have been published at this year's AAAA conference in Opatija). In the present paper correlation results between the terms “promptness of feedback”, “repetition rate”, “lip up/down capability”, “dynamic range mid register”, “brilliance”, “output power” and “overall quality”, and certain features of the measured input impedance curve of the instruments are discussed. The input impedance curves have been measured using the BIAS system, software release 6.0, which provides additional features extracted from the measured data. From the input impedance magnitude curves size, 3dB bandwidth and radius of curvature are extracted for each played resonance. The impedance phase curves are investigated for remaining phase and group delay at the peak centre frequencies as well as for the distance of the closest phase zero.

1 Introduction

Acoustical research studies on brass wind instruments have revealed knowledge about many aspects of the playing characteristics of trumpets, trombones and all other kinds of brass instruments [1][2]. Today there is hard- and software available to measure the intonation of such instruments as well as some other quality related factors. But, computers require parameters that can be quantified [3] while musicians usually play their instruments without being aware of all the underlying physics. They describe quality parameters of their instruments verbally and expect measurement systems to be consistent with their judgements to some extent.

The trumpet research project was initiated in 2002 in order to find new fundamental knowledge requested by makers, researchers and players of brass wind instruments. Aims and methods were presented at the SMAC (Stockholm Music Acoustics Conference) in 2003 [4]. Five Bb-trumpets with one extra bell and ten extra lead pipes have been donated for this study by instrument makers from Europe, America and Asia but brand names and models where not allowed to be disclosed. Since the time the project was started about 55 professional and semi-professional trumpet players from three different countries have been workinged in more than 250 test sessions, designed in order to identify playability and quality parameters of trumpets which are used by the brass community in a consistent way and which can be measured and quantified.

Thomas Moore of the ITG (International Trumpet Guild) described the project as “the arduous task of trying to quantify the vocabulary of trumpeters … that has been ignored for far too long”.

Currently we have no way to measure the “feel” of a horn. Therefore statements of sellers about playability, response and sound quality are almost meaningless and empirical research in this field is rare.

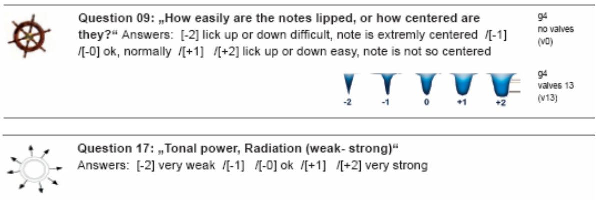
Because of the heavy and critical influence of psychological expectations concerning brands, making, and other visual clues all playing tests were made as blind performing tests.

2 Playing tests and questionnaire

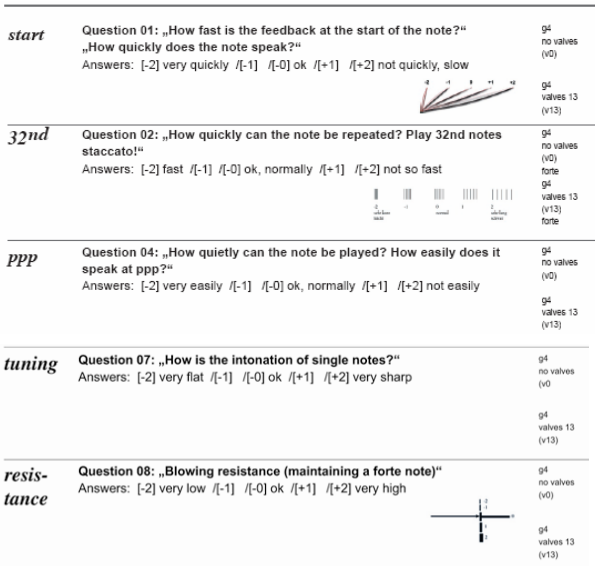
Test subjects got specific test instruments (piston valve and rotary trumpets) in a dark room without being able to identify anything optically. In almost complete darkness they had to handle the instruments, get acquainted to and test them, while they were asked orally about their judgements.

It took from 10 to 60 minutes to answer the 40 questions of one single test. Testers were asked to repeat the test several times with other instruments without knowing that sometimes they actually repeated the test with an instrument they have already tested before. This way it was possible to assess the ability of testers to judge consistently and repeatably.

The complete questionnaire can be downloaded from the project's website http://www.bias.at/TRP. Questions which have been answered consistently by the musicians and which therefore have been considered for further investigations are shown below:



Some questions have been asked twice to investigate playability differences between the notes G4 and G5 played without any valve and the same notes played pressing valve one and three. These answers are especially valuable because the corresponding measured input impedance peaks exhibit specific and distinct features which can be correlated with the judgements. These questions are:



As consistency and repeatability of given answers are already covered by another publication in full statistical detail [5] this study focuses on the correlation between judgements and features of the measured input impedance curves.

3 Subjects and Objects

The sample of players consists of 55 musicians from three countries: Austria (*n=13*), Finland (*n*=9), and USA (*n=33).*

All musicians from Finland are professional players. Participants in the USA and Austria are professional as well as semi-professional players or advanced trumpet students. There are no beginners in the sample (see Figure 1). Therefore, the Finnish musicians have also longer playing experience than the players from the USA and Austria (Kruskal-Wallis Test, $\chi^2 = 20.74$, df = 2, $p < .001$; see Table 1).

Most players in the sample are male and just two musicians play only Pop or Jazz (see Figure 2). Especially US-students are all-rounder concerning the literature.

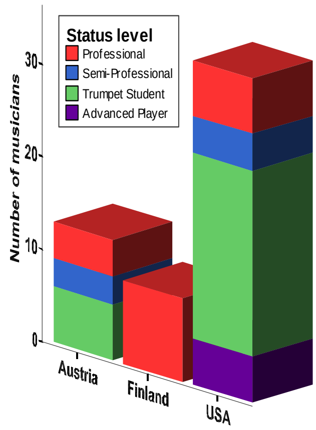


Figure 1: Status level of musicians

Table 1: Musician's experience in Years

	Austria	Finland	USA
Median	13	40	11
Modus	13	40	9
Minimum	8	27	4
Maximum	30	44	40
Percentile 75% .. 25%	16 .. 13	42.5 .. 32.5	14.5 .. 8.5

For 252 playing tests 21 different instruments were available. With 18 additional lead pipes and 2 extra bells many more configurations were possible. Trumpets labelled #14, #19 and #21 were the main Perinet (piston valve) trumpets used. Instruments labelled #24, #27 and #38 was the main rotary trumpets in use.

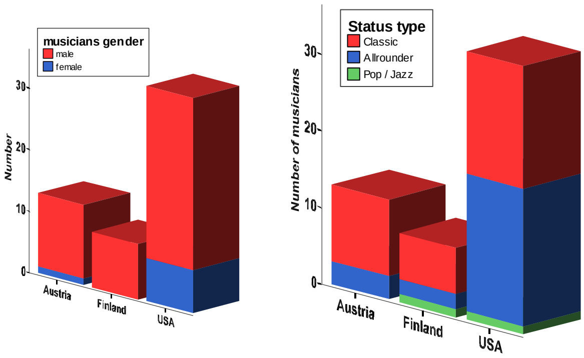


Figure 2: Gender and status type

Trumpets #21 and #27 were used most frequently with different lead pipes, mouthpieces and bells, while trumpets #12, #15, #20, #40, #43, #44 have been used only once or twice (see Figure 3).

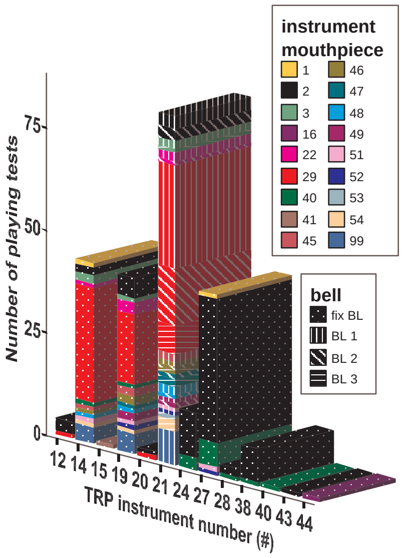


Figure 3: Number of tests

32 out of 252 playing tests were repetitions (same player, same instrument). The time spans between the repetitions varied between a few hours and some months, in further 5 cases the musicians tested a particular instrument configuration a third time without knowing that.

Players were asked to perform the test using one of two provided mouthpieces, if possible: A Bach 1½C for piston trumpets labelled #29 and a Breslmayr G2 for rotary trumpets labelled #2. In case players did not feel comfortable with the default mouthpiece they were allowed to use their own mouthpiece. Figure 4 shows for which mouthpieces and instrument configurations available test repetitions do allow to assess the ability of certain subjects or the repeatability of certain judgements. The standard mouthpieces #2 and #29 are shown in black and red (grey).

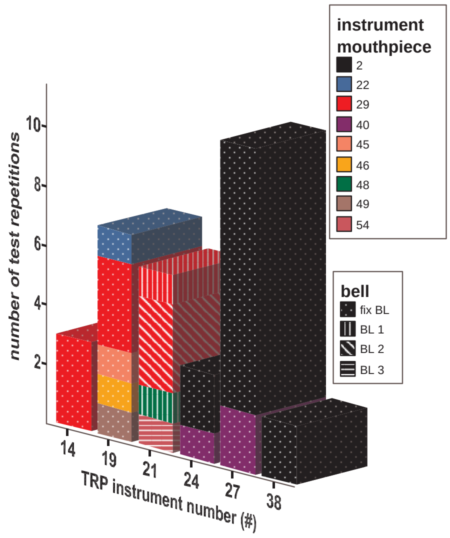


Figure 4: Repeated tests

4 Method

As a measure for the reliability of judgments between repeated tests Kendall's Tau-b (τ) was used. This is a coefficient of association for ordinal-level variables based on the number of concordances and discordances in paired observations. Possible values lie between –1 and +1. +1 corresponds to the case of only concordant pairs, 0 corresponds to statistical independence which means no relationship between two groups of answers.

Results of this test applied to repeated judgements of the same testers about the same instruments which have been published in [5] are not repeated here.

The difference of judgments about playing the same note with standard (V0) or extended (V13) fingering has been analysed using Wilcoxon's "Matched Pairs" test, often also called the Wilcoxon Signed-Ranks test. This is a non-parametric test of location in the case of two related samples (e.g. a before-and-after study). The null hypothesis is that two samples arise from exactly the same distribution, with the alternative that the two underlying distributions differ only in location.

Given a collection of N pairs of data items, the differences between the pairs are found and ranked according to absolute magnitude. The sum of the ranks is then formed for the negative and positive differences separately. W is the smaller of these two rank sums. For sufficiently large samples ($N > 25$) the following value Z is approximately normally distributed:

$$Z = \frac{W - \frac{1}{2} - N(N+1)/4}{\sqrt{N(N+1)(2N+1)/24}}$$

Statistical analyses were performed using SPSS (Statistical Package for the Social Sciences), Version 12.0.

5 Results

Significant correlations with physical parameters were mainly obtained for reliable input data. Question 8 about the blowing resistance of the trumpet was not answered reliably but revealed significant correlations with almost all physical parameters.

5.1 Reliability of difference judgements

Questions Q1 (feedback at the start of a note?), Q2 (repetition rate?), Q4 (how quietly can the note be played?), Q7 (intonation?), Q8 (blowing resistance?) and Q9 (how easy to lip up/down?) have been answered for two identical notes played in two different ways - using no valve and valve combination 1+3. In addition to that, questions Q7 and Q9 have been answered for two different notes; G4 (middle register) and G5 (high register).

Table 2: Reliability of differental judgments

	Kendall τ	p
Q1	.448	.007
Q2	.149	.367
Q4	-.160	1.0
Q7 (G4)	.277	.118
Q7 (G5)	.256	.183
Q8	-.141	1.0
Q9 (G4)	.497	.003
Q9 (G5)	.112	.540

The difference between V0 and V13 regarding start of feedback and lip up/down at least in the middle register was judged consistently between test repetitions.

5.2 Playability differences between V0 and V13

The Wilcoxon Matched-Pairs test was applied when comparing the answers for valve 0 and valve 13 playing a specific trumpet combination the first time. Table 3 shows the medians, 25th and 75th percentiles of the judgements given to the notes played with no valve and with valves 1 and 3. The column labelled with *p* shows the resulting significance level.

Table 3: Playability differences between V0 and V13

	Valve 0		*Valve 13*		*Wilcoxon*	
	Med.	Percentile	Med.	Percentile.	*Z*	*p*
Q1	-.25	-.94/.51	-.11	-.86/.71	-2.446	.014
Q2	-.29	-1.03/.52	.33	-.57/1.12	-7.328	.000
Q4	-.58	-1.42/.29	-.29	-1.13/.61	-3.970	.000
Q7 (G4)	-.03	-.63/.59	1.14	.41/1.75	-11.308	.000
Q7 (G5)	.09	-.60/.73	1.33	.59/1.92	-11.032	.000
Q8	.12	-.72/.84	.31	-.58/1.14	-2.937	.003
Q9 (G4)	.29	-.55/1.01	.18	-.75/1.05	-1.462	.144
Q9 (G5)	.08	-.67/.79	-.31	-1.29/.69	-4.313	.000

Only the judgment difference for lip up/down (Q9) for note G4 was not significant. Valve 13 was judged slower regarding start of feedback (Q1) and repetition rate (Q2), harder regarding response (Q4), sharper regarding intonation of both played notes (Q7), higher regarding resistance (Q8) and more slotted regarding lip up/down (Q9) mainly in the high register (note G5).

5.3 Input impedance measurements

In order to identify those acoustical parameters which musicians base their judgements on concerning the playability of instruments, input impedance measurements have been made and certain features have been extracted from these curves.

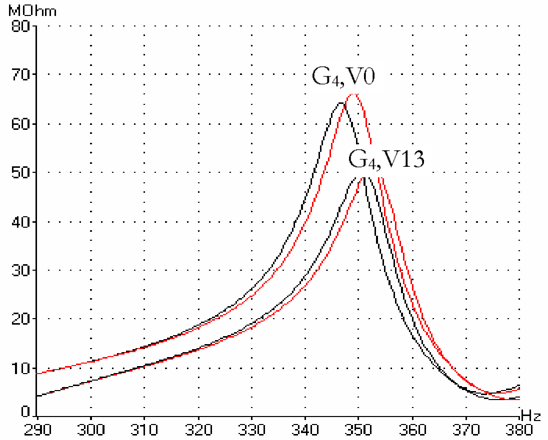


Figure 5: Input impedance of G4 (V0 and V13)

Comparing the resonance peaks (Figure 5) corresponding to the judged notes visually it is easy to see that pressing valves 1 and 3 does

- sharpen the pitch represented by the centre frequency of the resonance and
- lower the absolute impedance magnitude.

Other features, especially some phase related ones, are not so obvious and have been extracted using BIAS version 6. Going from V0 to V13 all extracted features moved in the same direction on all instruments except for the pitch which went slightly down instead of up on one single instrument.

Table 4 shows the differences between V0 and V13 for the features f_0 = pitch [cent], abs(Z) = impedance magnitude [MΩ], curv = curvature [kΩ/Hz^2], $3dB_{L/R}$= left/right limit of 3dB-range [cent], Φ_0= remaining phase at f_0 [deg], GD = group delay [msec], $f_{\Phi=0}$ = pitch at phase zero crossing [cent]. All intonation values are related to A4 = 440Hz.

Table 4: Significance of feature differences of Z_{in}

	Valve 0		Valve 13		Wilcoxon	
	Med.	Percentile	Med.	Percentile	Z	*p*
f_0	-13.8	-21.0/-1.47	3.69	-1.51/15.4	-5.511	.000
abs(Z)	82.2	66.8/114	66.9	51.4/93.1	-5.484	.000
curv	1089	913/1712	895	730/1029	-5.497	.000
$3dB_L$	46.7	40.56/52.6	46.8	40.7/52.8	-5.511	.000
$3dB_R$	47.2	41.1/53.1	47.3	41.3/53.3	-5.014	.000
Φ_0	10.3	5.93/18.8	10.8	7.47/15.9	-3.307	.001
GD	-22.3	-28.6/-15.9	-20.8	-27.9/-16.0	-4.691	.000
$f_{\Phi=0}$	17.6	-4.45/51.7	36.2	23.3/77.6	-4.973	.000

All differences are highly significant and can be explained by the increased length of the resonant air column when valve slides are activated in the instrument. However most of the features are strongly correlated.

Due to the non-normality of the distribution of these features, Spearman correlation coefficients were computed separately for both notes and fingerings which were investigated. For the note G4 all impedance features show significant correlations, for the higher note G5 some correlations vanish.

Table 5 illustrates these inner correlations between all investigated features of the input impedance curve for two selected notes (G4/V0 in the right upper half, G5/V0 below the main diagonal).

Table 5: Inner correlations for G4/V0 and G5/V0

	f_0	abs(Z)	curv	$3dB_L$	$3dB_R$	Φ_0	GD	$f_{\Phi=0}$
f_0		-.75/ .000	-.74/ .000	.81/ .000	.98/ .000	.81/ .000	.68/ .000	.91/ .000
abs(Z)	.41/ .077		.82/ .000	-.44/ .055	-.76/ .000	-.77/ .000	-.64/ .000	-.72/ .000
curv	.48/ .032	.74/ .000		-.43/ .06	-.79/ .000	-.80/ .000	-.90/ .000	-.84/ .000
$3dB_L$	.96/ .000	.23/ .325	.42/ .068		.81/ .000	.65/ .002	.48/ .031	.722/ .000
$3dB_R$	.99/ .000	.45/ .048	.48/ .033	.93/ .000		.82/ .000	.77/ .000	.93/ .000
Φ_0	.16/ .490	.61/ .005	.08/ .743	.07/ .783	.23/ .331		.76/ .000	.94/ .000
GD	.39/ .09	.67/ .001	.11/ .656	.25/ .292	.45/ .045	.77/ .000		.82/ .000
$f_{\Phi=0}$	.73/ .000	.75/ .000	.43/ .062	.62/ .004	.78/ .000	.76/ .000	.84/ .000	

Due to these strong inner correlations external correlations must be interpreted with much care. Due to their strong interdependency it is not possible to identify any impedance feature as a single cause for subjective judgements.

5.4 Correlations between musician's judgments and BIAS measurements

Correlation coefficients between the ranks of musician's judgments and the ranks of measured features were again computed separately for all note-fingering combinations (see Table 6). The quality of these linear regression models was assessed by means of residual analysis.

The answers to Q17 (tonal power), which was judged consistently by musicians, show correlations with 3dB-range, impedance magnitude, curvature and pitch (calculated as position of the impedance maximum as

Table 6: Correlations between judgements and measured features

played note	*judgments*	*measured features*							
		f_0	abs(Z)	curv	$3dB_L$	$3dB_R$	Φ_0	GD	$f_{\Phi=0}$
G4/V0	Q7 (pitch)	-.139/.558	.020/.934	.020/.934	-.100/.675	-.160/.500	-.139/.558	-.219/.354	-.139/.558
	Q9 (slotted)	***-.478/.033***	.159/.504	.002/.992	-.301/.197	-.366/.112	-.141/.552	.096/.688	-.235/.318
	Q10 (dyn.r.)	.324/.164	-.293/.209	-.365/.113	.393/.086	.420/.065	.424/.062	***.468/.037***	.409/.073
	Q17 (power)	***.599/.005***	***-.515/.020***	***-.47/.036***	.385/.093	***.633/.003***	.442/.051	.397/.083	***.521/.019***
G4/V13	Q7 (pitch)	-.338/.145	.147/.536	.264/.260	*-.430/.059*	*-.387/.092*	-.322/.166	-.359/.120	-.402/.079
	Q9 (slotted)	***-.683/.001***	***.50/.025***	.363/.116	***-.647/.002***	***-.648/.002***	-.261/.267	-.243/.303	-.414/.070
	Q10 (dyn.r.)	.279/.233	-.324/.164	-.413/.071	.338/.145	.422/.064	***.620/.004***	***.663/.001***	***.584/.007***
	Q17 (power)	***.491/.029***	-.429/.059	-.354/.126	***.504/.023***	*.427/.061*	.217/.358	.247/.294	.363/.115
G5/V0	Q7 (pitch)	***-.597/.005***	-.248/.292	-.179/.451	***-.535/.015***	***-.603/.005***	-.265/.259	-.284/.224	***-.525/.018***
	Q9 (slotted)	***-.524/.018***	-.271/.247	-.407/.075	-.438/.053	***-.506/.023***	-.039/.871	-.184/.437	-.368/.110
	Q10 (dyn.r.)	.050/.835	.396/.084	.129/.587	-.131/.581	.131/.581	.431/.058	.385/.094	.352/.128
	Q17 (power)	.111/.642	***.475/.034***	***.542/.014***	.038/.874	.138/.563	.100/.674	.113/.635	.219/.353
G5/V13	Q7 (pitch)	-.059/.804	.079/.741	.272/.246	-.034/.887	-.034/.887	***-.477/.033***	-.264/.261	-.108/.650
	Q9 (slotted)	-.149/.531	.311/.183	.191/.419	-.157/.509	-.046/.849	.265/.259	.215/.363	.136/.566
	Q10 (dyn.r.)	.007/.978	.369/.109	.003/.989	-.163/.493	.155/.515	***.632/.003***	***.510/.022***	.299/.200
	Q17 (power)	.144/.545	.441/.052	.358/.121	.067/.780	.259/.270	.169/.477	.156/.510	.302/.195

well as of the phase zero crossing) at least as far as the middle register is concerned. Tonal power is judged stronger when impedance magnitude and curvature are lower, and pitch and 3dB-range higher. In the higher register (G5), however, the correlations either vanish or change the direction.

Ratings for "lip up/down", for which judgments were reliable, too (except for G5/V13), correlate with pitch and 3dB-range (again except for G5/V13). For the note G4/V13, for instance, 3dB-range explains 42% of the variance of the musician's ratings regarding "lip up/down" (r = -.65 $\Rightarrow r^2 = 0.42$). Pitch f_0 explains between 23% and 47% of the variance of the musician's ratings regarding "lip up/down". Lip up/down is judged more slotted if pitch and 3dB-range are lower and impedance magnitude and pitch at phase zero are higher.

Ratings for Q10 (dynamic range mid register) correlate with group delay, Φ_0 and $f_{\Phi=0}$, if all these parameters are higher dynamic range is judged larger.

Responses to questions Q7 (intonation), Q1 (feedback), Q2 (repetition) and Q4 (response) do not correlate with features of the input impedance curve; however, those judgments proved to be unreliable when no direct reference was available. Furthermore there was no correlation between the reliably judged questions Q13 (brilliance) and Q31 (overall classification of the instrument).

The correlation of judgement differences between V0 and V13 with corresponding impedance feature differences were expected to yield the most reliable results. Differences for Q9 (lip up/down) which were judged reliably yielded no significant correlation with impedance feature differences, which means that there is no correlation between the magnitude of the differences in physical parameters and the magnitude of differences in judgements. Differences for Q1 (start of feedback) do correlate with differences in curvature and group delay. The smaller the difference regarding group delay, and the greater the difference in curvature, the greater is the change of judgements for first feedback towards the slower direction. The not consistently rated feature resistance (Q8), on the other hand, correlates well with all impedance feature differences except $f_{\Phi=0}$.

6 Summary

Those qualities of trumpets, which could be judged consistently by musicians, usually show significant correlations with BIAS measurements. Using multiple correlation analyses it can be shown that different combinations of acoustical features or even single such features explain around 50% of the variance of musician's judgments regarding lip up/down and sound output power.

The differences in judgements of notes played using V0 and V13 where highly significant. The notes G4 as well as G5 were generally judged slower in first feedback and repetition rate, worse in *ppp* response, higher in *f* playing resistance, sharper in pitch and more difficult to lip up/down when played with V13. On the other hand, V13 corresponds to a lower peak magnitude of its associated air column resonance, an upward shift of its centre frequency in conjunction with smaller curvature, wider 3dB range, smaller group delay and slightly more residual phase.

A formal correlation between the magnitudes of the judgement differences and the physical characteristics of the corresponding air resonances was only significant for Q1 (feedback). The missing correlation of the other judgements is probably due to the coarse resolution of the used grading scale (very poor, poor, normal, good, very good) and the inability of subjects to judge any more accurately.

Acknowledgements

The authors would like to thank Klaus Wogram and Dale Olson for sharing their experiences concerning organisation and performance of such blind tests.

References

[1] Bertsch, M., 'Studien zur Tonerzeugung auf der Trompete: Variablen der Tonerzeugung'. *Schriftenreihe des Instituts für Wiener Klangstil (Musikalische Akustik) an der Universität für Musik und darstellende Kunst Wien*, Vol. 4. Wien, (2002)

[2] Campbell, M., 'Brass Instruments As We Know Them Today'**.** *Acta Acustica united with Acustica*, Stuttgart: Hirzel, Vol. 90. No. 4. pp. 600-610. (2004).

[3] Widholm, G., '*BIAS 5.1 Manual*'. Vienna: Institut für Wiener Klangstil, (2001).

[4] Bertsch, M., 'Bridging instrument control aspects of brass instruments with physics-based parameters**',** *Proc. of the SMAC03 (Stockholm Music Acoustics Conference 2003)*, Stockholm: KTH Speech, Music and Hearing, pp. 193-196. (2003)

[5] Bertsch, M., Waldherr, K., 'Reliability of Trumpet Players' Judgments during Instrument Blind-Tests', *2nd Congress of AAAA (Alps-Adria Acoustics Association)*, EAA Symposium, Opatija Croatia, (2005)