

Between producers and consumers: Critics' role in guiding listeners' choices

Elena Alessandri^{1,3}, Hubert Eiholzer², and Aaron Williamon^{2,3}

¹ Lucerne University of Applied Sciences and Arts, Switzerland

² Conservatory of Southern Switzerland, Lugano, Switzerland

³ Centre for Performance Science, Royal College of Music, London, UK

This study investigated how professional critics' judgments of recorded performances relate to other listeners' preferences. Music students (n=10) and music professionals (n=7) were asked to rate their liking of five interpretations of the opening of Beethoven's Piano Sonata Op. 111. Listeners' likings were compared with judgments given in reviews published in the *Gramophone*. Correlation between critics' judgments and music professionals' preferences was moderate, while no correlation was found between critics' evaluation and music students' likings. The results suggest that preferences for given interpretations of a piece vary between listeners and may be influenced by the listeners' prior experience of detailed listening and study of repertoire and its renditions.

Keywords: performance preferences; music criticism; musical expertise; listeners' choices; Beethoven

Professional critics' reviews of recorded performances are published regularly in newspapers and specialized music magazines. Arguably, the aim of these reviews is to guide consumers' choices when it comes to deciding which recording to buy (Frith 2009, Pollard 1998). Yet no research so far has investigated the actual role that music reviews play in listeners' choices or their efficacy in guiding listeners toward purchases of recordings that promise long-term satisfaction.

In the classical music market, listeners are exposed to an ever-increasing number of interpretations of canon repertoire from which they can choose to listen. Previous research by the authors found that in the magazine *Gramophone* alone 845 recordings of Beethoven's Piano Sonatas, produced by 217 different pianists, were reviewed between 1923 and 2010. This amount of

material seems impressively large. Findings in decision-making research suggest that this increase in options (quantity of different versions of an item to choose from) may paradoxically lead to paralysis of choice and dissatisfaction (Schwartz 2008): in this scenario the critics' guidance—working as a filter of choice—seems to be particularly significant.

However, critics are seasoned listeners, with an extraordinarily rich experience in listening to music and in comparing high-level professional performances. When it comes to canon repertoire, critics have most likely listened to and evaluated plenty of different interpretations of the same piece. It could be reasonably expected that this level of familiarity with the piece, and with various interpretations of it, may color their attitudes and preferences toward certain performances in such a way that what may be considered a good performance by critics may not be considered thus by a listener who has a different level of musical expertise and listening history. Exploring this hypothesis was the aim of this study.

METHOD

Phase 1: Establishing the valence of critics' judgments

Participants

A total of 28 music undergraduates at the University of Leuven took part in the first phase of the study in the context of a seminar on music criticism that was part of their study programme.

Materials

From the material published in the *Gramophone*, five reviews of recordings of Beethoven's Piano Sonatas Op. 111 were chosen. From the reviews, sentences were extracted that discussed the Maestoso section at the beginning of the sonata. These review excerpts (length=86-169 words) were used as stimuli.

Procedure

Students were asked to read all five review excerpts and to evaluate each on a 7-point scale answering the question: "What is the critic's opinion of this performance?" from -3 (Not at all worth listening to) to +3 (Absolutely to be listened to!).

Phase 2: Critics' versus listeners' likings

Participants

For Phase 2, 17 musicians were recruited: 10 students with a major in music performance and 7 “experts” who studied music at professional level and are currently involved in different music-related jobs that require large amounts of regular musical exposure (performing, teaching, researching).

Materials

The Maestoso parts of the recordings corresponding to the five reviews evaluated in Phase 1 were used. Recordings were cut at the beginning of the *Allegro con brio ed appassionato* (at the end of bar 25) using Audacity 1.2.6, a 1 s fade out was added at the end.

Procedure

Participants listened to the five recordings and rated them on three 7-point scales: (1) their liking of the performance (1=not at all, 7=very much), (2) the expressiveness of the performance (1=not expressive at all, 7=very expressive), and (c) how well the performer managed to keep the tension throughout the passage (1=not well at all, 7=very well).

RESULTS

Phase 1: Establishing the valence of critics' judgments

Agreement among participants on the valence of the single reviews was strong; Kendall's coefficient of concordance (W) was 0.65 ($p < 0.001$). Friedman's test showed that participants were able to discriminate between the different valences of reviews ($\chi^2_4 = 67.19$, $p < 0.001$). Post-hoc Wilcoxon tests were significant for three levels of valence: Good for Taub, Barenboim, and Michelangeli; not so good for Pogorelich; very bad for Ugorski (see Table 1).

Phase 2: Critics' versus Listeners' likings

Results for the liking ratings are shown in Table 1. A multivariate analysis showed a significant main effect of level of expertise on the liking of one or the other interpretation: Wilk's $\lambda = 0.10$, $F_{5,11} = 22.40$, $p < 0.001$. Individual t-tests were significant for one of the five recordings (Ugorski). To test the level of agreement between listeners and critics, Spearman's rank-order correlation coefficient was computed between each participant liking ratings and the

Table 1. The valence of critics' judgements (rated -3 to +3), students' and experts' liking ratings for the five recordings (rated 1 to 7), and t-tests between students' and experts' ratings.

	<i>Barenboim</i>	<i>Taub</i>	<i>Michelangeli</i>	<i>Pogorelich</i>	<i>Ugorski</i>
<i>Critics</i>					
Mean	2.44	2.39	2.25	1.25	-2.22
SD	0.74	0.63	1.11	1.08	0.92
<i>Students</i>					
Mean	4.60	4.20	4.00	4.10	5.60
SD	1.78	1.40	2.26	2.18	1.35
<i>Experts</i>					
Mean	5.29	4.86	4.86	5.71	2.57
SD	1.38	1.46	1.46	1.60	0.98
t	-0.85	-0.94	-0.95	-1.66	5.06
p	0.41	0.36	0.36	0.12	<0.01

valence of critics' reviews (correlations were converted to z scores for the analyses, and means reported here were transformed back into correlation coefficients). The mean correlation was significantly higher for expert listeners' likings ($r_s=0.47$) than for students' likings ($r_s=-0.16$) ($t_{14,10}=-2.73$, $p<0.05$). The correlation between expert listeners and critics was moderate to strong, and significantly different from zero ($t_6=3.87$, $p<0.01$), while that between students and critics was not ($t_9=-0.81$, $p=0.44$).

The largest difference between students' and experts' ratings was in the evaluation of Ugorski's performance (cf. t-tests in Table 1). This performance was harshly criticised by the *Gramophone* reviewer David J. Fanning for the "ultra-spacious" tempo employed by the pianist that in the reviewers' opinion makes the music "fall apart." This effect is made worse by Ugorski's expressive playing that, according to the critic, does not reflect a "compelling interpretive vision," but rather results in "studied eccentricity" (February 1993, p. 62). This critic's judgement echoes in expert listeners' ratings. Figure 1 shows ratings of liking, expression, and tension for both experts and students. Ugorski's excessive use of expressive inflections seems reflected in experts' expression ratings. In line with the reviewer, this use of expression does not add to the overall liking of the performance. Students, on the other hand, seemed to appreciate Ugorski's slow tempo and use of expression, so much so that this recording was their favourite among the five.

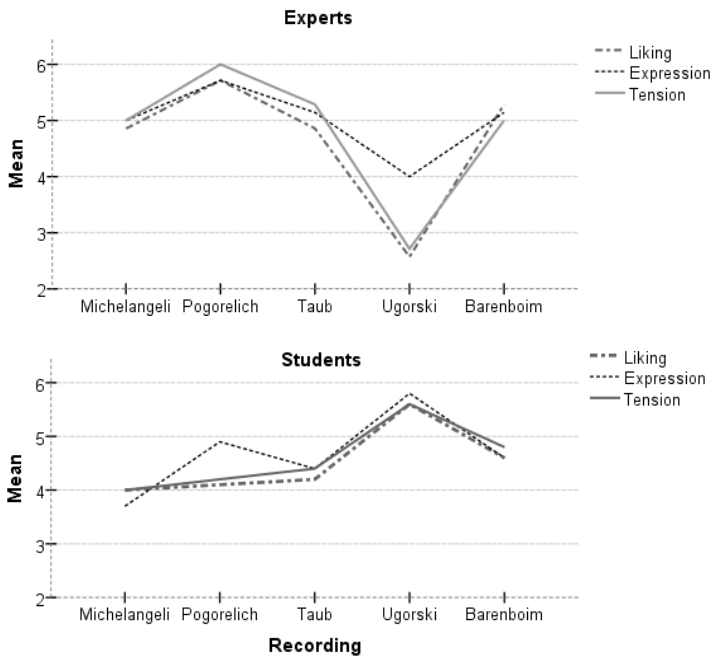


Figure 1. Experts' and students' ratings of liking, expression, and tension.

DISCUSSION

This study opens an exploration of the relationship between critics' judgments of performances and wider audience's preferences and, in a wider perspective, of the role of critical practice in the classical music market. Likings of one or the other interpretation were significantly different for the two groups of listeners; in particular, preferences expressed by more experienced listeners correlated more highly with critics' judgments, even though the level of correlation remained moderate. These results seem to support the hypothesis that critics' preferences for a given performance may be sharable only by similarly informed listeners; however, this needs to be substantiated by repeating the test with more participants and different sets of recordings. The difference between experts' and students' ratings depended mainly on students' liking of Ugorski's performance. Participants were given no information regarding the nature of the recordings; however, the quality of both the recording and performance may have led participants to think that they

were listening to commercially available professional recordings. This in turn might have reduced students' self-confidence in giving low ratings. On the other hand, this interpretation fails to explain why students rated Ugorski's recording as their favorite. An alternative explanation could be linked to the different level of familiarity with the piece and its renditions (cf. Levinson 1987). It could be that the slower tempo employed by Ugorski allows non-familiar listeners better to grasp the structure of the work, while it is perceived as dull and unexciting for listeners who know the piece well. It could also be that, with increased familiarity and knowledge of the piece, listeners develop a more precise idea of how the work should be performed, thus becoming less ready to appreciate interpretations that move away from this idea. In any case, to better understand these results follow-up studies should focus particularly on performances that are negatively reviewed by critics. Finally, if the hypothesis that critics' preferences for given performances are shared only by similarly informed listeners should be further supported, the next step would be to investigate to what extent critics' preferred performances can in fact offer the best possible aesthetic experience also to listeners who do not like them at first.

Acknowledgments

This study was partially supported by a grant from the Swiss National Science Foundation (no. 13DPD6_130 269).

Address for correspondence

Elena Alessandri, Lucerne University of Applied Sciences and Arts, Zentralstrasse 18, Lucerne 6003, Switzerland; *Email*: elena.alessandri@hslu.ch

References

- Frith S. (2009). Going critical. Writing about recordings. In N. Cook, E. F. Clarke, D. Leech-Wilkinson, and J. Rink (eds.), *The Cambridge Companion to Recorded Music* (pp. 267-282). Cambridge: Cambridge University Press.
- Levinson J. (1987). Evaluating musical performance. *Journal of Aesthetic Education*, 21, pp. 75-88.
- Pollard A. (1998). *Gramophone: The First 75 Years*. Harrow, UK: Gramophone Publications.
- Schwartz B. (2008). Can there ever be too many flowers blooming? In S. J. Tepper and W. Ivey (eds.), *Engaging Art* (pp. 239-256). New York: Routledge.