

# EXPRESSIVENESS IN MUSICAL PERFORMANCE: COMPARISON OF NOVICE AND EXPERT-LISTENER PERSPECTIVES

*Emery Schubert*

*Dorottya Fabian*

School of Music and Music Education, University of New South Wales, Sydney, Australia

## ABSTRACT

**Background:** As part of our ongoing research into the perception of baroque performance practice, we investigated the relationship between expert and novice baroque music listeners.

**Aims:** To test the hypothesis that expert baroque listeners are more consistent and accurate in their responses to baroque music than novice listeners and to compare the aesthetic and performance feature judgements of novices with those of experts.

**Methods:** 30 experts and 31 novices in baroque music listening made aesthetic and performance feature assessments to four excerpts from J. S. Bach's instrumental compositions, with at least two interpretations of each piece. Aesthetic parameters of preference, stylishness, baroque expressiveness and romantic expressiveness were rated, in addition to performance parameters of articulation, flexibility, intensity and legato.

**Results & Conclusions:** Experts and novices tend to agree on the performance parameter ratings more so than they do on aesthetic parameter ratings. Novices tend to respond with greater reliability than experts, however, expert responses are more valid and complex than novices.

## 1. BACKGROUND

Our ongoing research has focussed on better understanding the musicological question of what devices determine expressiveness in baroque music performance (Fabian Somorjay & Schubert, 2000; Fabian & Schubert, 2002a, b, c, d; Schubert & Fabian Somorjay 2000; Schubert & Fabian 2001, 2002). We have been using empirical methodology which involves asking various participants to make responses. By examining multiple responses, a more complete picture of baroque music practice, from a perceptual point of view, may be determined. We argue that this approach can provide a broader understanding than the more traditional approach of the lone expert's contemplations. However, multiple response collection is also fraught with problems. For example, from whom does one gather the responses? Individual differences and their influence upon musical responses have received some attention (eg. Gromko, 1993; Kemp, 1997; Smith, et al., 1994), but little is known about how experience in baroque music influences the kinds of judgements which interest us. The role of the lone expert is precarious, because at best they can only hope to represent the views of individuals sharing the same 'preparatory set' (Meyer, 1956; Bartel, 1988) – that is, a highly specialised group of individuals. And perhaps even within this group there will still be disagreements because their knowledge and experiences may be so refined and individualistic through their home-conservatory

traditions, detailed reading of music literature and specialised knowledge of music scores that their reactions may become apparently inconsistent and possibly dogmatic.

The present concern is of individual differences along this dimension of musical expertise. Regarding the question of who to test, there are three possibilities: (1) Experts with considerable experience in listening and performing: In the case of baroque music specialisation, this may include knowledge of period instruments and of historically informed performance practice. (2) Experienced musicians, though not necessarily experts (those commonly used in music-psychology based studies). (3) Novices who have no specialised knowledge of music, but may be interested in it.

Ideally, each of these groups will respond to music in the same way. The experts learn and disseminate their playing styles and knowledge which are then absorbed by listeners whether musicians or novices. Past research shows that judgement of musical/acoustic features can be done fairly accurately by experienced and less experienced musicians, provided the task is presented clearly (Smith et al, 1997). However, there appear to be some differences in aesthetic responses between these groups. Gromko (1993) found that experts tend to be more analytic and less experienced people more emotional in their aesthetic responses to music. We will, therefore, address the question of the relationship between responses to performance versus aesthetic parameters.

If there is a difference between these groups, a consequent question is which group will provide the most valid, reliable and efficient results? Experts should be the most valid if we assume that they are the holders of musical knowledge within a given culture. If this were the case, knowledge about their responses should be compiled and disseminated to musicians and novices wishing to improve their musical sophistication. However, as mentioned, experts may have such refined knowledge that their views may fluctuate greatly across individuals. This problem can be redeemed by assuming multiplicity of solutions to musical, performance and aesthetic problems. By the same token, variance in responses might be interpreted as demonstrating a lack of reliability. That is, how do we know that experts respond consistently? This is an important question, because inconsistencies within individuals suggest a kind of musical, performance and aesthetic anarchy, making expert knowledge very difficult to use for educating less experienced people about performance practices. Further, what relevance does the debate about the aesthetic value of historically informed performance options have to non-specialists? Are they sensitive to the subtleties which experts claim to detect? Answering these questions is the aim of the present paper. Greater understanding

of these issues will help experimental design for future investigations of the perception of expression in music. The focus on baroque music specialists allows us to focus on a highly specialised group of experts.

## 2. AIMS

The aim of this paper is to test the hypothesis that expert baroque listeners provide reliable and valid assessments about baroque music and its expressive means. More specifically, we investigated whether this reliability and validity applies to aesthetic parameters, performance parameters, both or neither.

## 3. METHODS

### 3.1. Participants

30 experts and 31 novices in baroque music took part. The experts were recruited through professional baroque ensembles and personal contacts. They each had significant experience in baroque performance or listening. It included individuals who were experienced listeners though not necessarily performers: for example, broadcasters involved in preparing and/or presenting programs on baroque music. We argue that novices recruited must be interested in music but lacking in expertise. If they had no interest, we could not expect them to engage in the task in a serious manner. Therefore, we recruited students from a university undergraduate music fundamentals course with no music prerequisites. These individuals were tested within the first week of the course. The pool was embellished with other individuals who fulfilled the criteria of being novice.

### 3.2. Stimuli

A variety of performers (eg. historically informed and traditional), and a variety of musical moods (slow and fast movements) were selected for rating. Pieces were restricted to instrumental excerpts by J. S. Bach. In two cases, the stimulus was tested twice, and in another a stimulus was modified by filtration (see Fabian & Schubert, 2002a). For each piece at least two versions (i.e. different performances) were rated. The 14 stimuli used are described in the Appendix. Within each piece, different performances were presented in random order except that repeated pieces were never presented in immediate succession.

### 3.3. Procedure

Participants sat at a computer and completed a questionnaire about their baroque musical experience and their own definitions of terms pertinent in baroque music performance such as stylishness and expressiveness.

Examples were then presented. For each example, responses were made along 4 performance feature scales (with extreme range of a 9 point scales in parentheses, from 9 down to 1): **Legato** (*Very legato* [9] to *Not at all Legato* [1]); **Articulation** (*Very articulated* to *Not at all articulated*); **Intensity of Playing** (*Very intense playing* to *Not at all intense playing*); **Flexibility**

**of Playing** (*Very flexible playing* to *Not at all flexible playing*) and 5 aesthetic feature scales: **Baroque Expressiveness** (*Very baroque-like expressive* to *Not at all baroque-like expressive*); **Romantic Expressiveness** (*Very romantically expressive* to *Not at all romantically expressive*); **Stylishness** (*Very stylish performance* to *Not at all stylish performance*); **Preference** (*I like this performance very much* to *I don't like this performance at all*); **Historically Informed** (*Very historically informed* to *Not at all historically informed*).

## 4. RESULTS

One way ANOVA for each piece was conducted in addition to correlations across the 14 pieces for each mean response dimension. Table 1 summarises the general findings of the analysis. The high percentages in the second column indicate that the two groups provide distinctly different scores for each dimension. However, values approaching 1 in the third column provide evidence for agreement in the direction of scores across the 14 pieces between experts and novices. Thus, there is agreement for all performance dimensions, but not for the aesthetic dimensions. For example, *preference* ratings share the same region on the scale between the two groups (non-significance for [100-29=] 71% of the 14 pieces), but they do not agree with each other across pieces or performances (i.e. non-significant correlation [ $r = 0.142$ ] at  $p = 0.05$ ) (See also Fig. 1 for a specific example). This disagreement on matters of aesthetics supports similar findings by Gromko (1993). However, how to explain these differences is a matter which will be discussed below. *Legato* has the strongest agreement between novices and experts. With a correlation of 0.920, it seems that novices are just as 'good' at judging legato as are experts.

Response Scale	% Sig Diffs (/14) at $p=0.05$	$r$ * $p=0.05$
<i>Baroque Expressiveness</i>	64	0.495
<i>Historically Informed</i>	71	0.294
<i>Preference</i>	29	0.142
<i>Romantic Expressiveness</i>	42	0.473
<i>Stylishness</i>	57	0.166
<u>Articulation</u>	42	<b>0.555*</b>
<u>Flexibility of playing</u>	57	<b>0.605*</b>
<u>Intensity of playing</u>	14	<b>0.606*</b>
<u>Legato</u>	43	<b>0.920*</b>

**Table 1:** Experts versus Novices by response collapsed across the 14 performances judged. Sig Diff is the tally of significant differences between experts and novices based on One-way ANOVA for each dimension.  $r$  is the correlation coefficient of the 14 mean scores between expert and novice. *Italics denotes aesthetic dimensions*, Underline denotes performance dimensions. In nearly all cases, mean scores were further away from the centre (5) of the scale for experts than for novices.

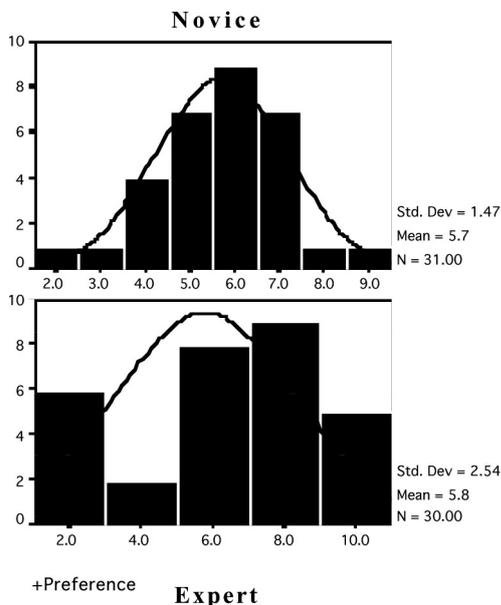
A further, descriptive analysis was conducted to investigate why deviation scores appeared to be larger for experts than for novices (Sum of squares of standard error of mean across pieces by dimensions for experts = 256, for novices = 119), particularly for aesthetic dimensions (134 for experts, 64 for novices). Histograms showed frequent occurrences of multimodal distributions in aesthetic responses by the expert group. A selection of these are shown in Figures 1 and 2.

## 5. DISCUSSION

The main aim of the present analysis was to investigate relationships between novice and expert responses. We are still analysing the data, and space prevents us from providing more details, but several pertinent findings can be discerned from the analysis.

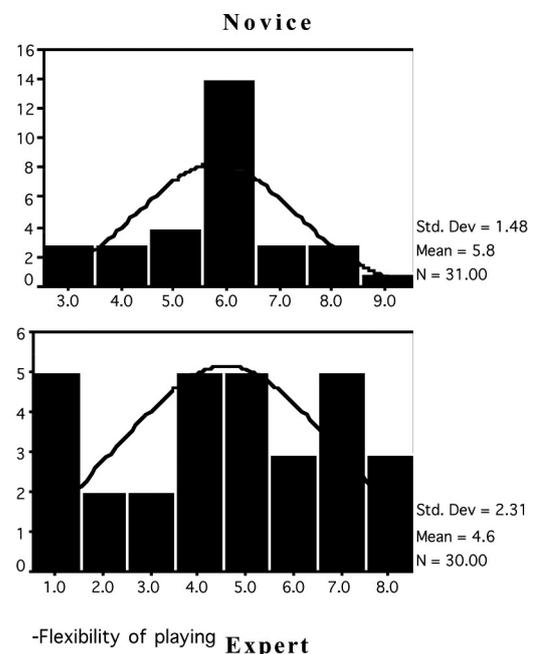
(1) Novice responses are more reliable than experts because, in general, their deviation scores tend to be smaller for a single performance on any single scale. For example, compare expert and novice histograms and deviation scores in Figs. 1 and 2.

(2) Expert responses, although more varied, are more valid than novice responses because they make clearly distinct responses across pieces, and they demonstrate the richness and complexity of aesthetic problem solving – namely that aesthetic problems may have multiple solutions. Examples of this are demonstrated by the multimodal nature of the expert histograms in Figs. 1 and 2, compared to the unimodal histograms which are typical of the novice responses. While the wider fluctuation in expert responses do not logically confirm validity, it is reasonable (if not essential) to assume that the responses of the experts are valid because of the nature of their relationship of the task for which they were recruited.



**Figure 1.** Preference rating histogram of the first solo of Harnoncourt's Allegro movement from Brandenburg Concerto No 4. Normal distribution curve is superimposed.

On the other hand, we cannot be certain that novices really meant to rate closer to the middle of the scale, or whether they did not understand the task. While this matter is discussed in finding 3 below, some of our data suggest that novices did understand the task but were not as informed as the experts. For example, novices seem to link *stylishness* to *romantic expressiveness*, whereas experts seem to relate it to *baroque expressiveness*. Perhaps the novices have a more old-fashioned view of *stylishness*, and new knowledge about baroque performance practice has not filtered through to them. In contrast, novices also seem to rate recordings that sound 'new' as being more stylish. For example, the filtered Luca recording (digitally manipulated for this study—see Appendix and Fabian & Schubert, 2002a) was rated as being less stylish by novices than the original, modern recording. The possibility that novices are unsure of how to respond, and therefore tend to 'play it safe' by responding near the middle of the scale indicates a third finding:



**Figure 2.** Flexibility rating histogram of Grumiaux's opening 30 seconds of Sarabande for solo violin. Normal distribution curve is superimposed.

(3) Novices appear to lack the language required for describing various (though not all) elements of music. For example, their judgements of *legato* were in almost complete agreement with the experts, however, their judgements of *articulation*—which should have been negatively correlated with *legato*—were more typically rated toward the centre of the scale. That is, novices probably didn't understand what was meant by the term 'articulation' in the musical context of the present study when compared to their understanding of 'legato' (consistent with the findings of Smith et al. 1994). There is, however, a further complication to this finding. Menuhin's performance of the Sarabande for solo violin (see Appendix), for example, received low scores on *historically informed* (Mean 2.87) and *baroque expressive* scales (Mean 2.87) from experts, however, novices rated these much higher (6.03 and

5.90 respectively) suggesting that novices used either different criteria for assessing these dimensions, compared with experts, or that they did not understand the meaning of these terms. To help resolve the issue, we examined definitions of terms provided by participants. A typical expert described *baroque expressiveness* in great detail such as the following response:

*“Clarity of detail within a musical line. Observance of emotional content of single sounds, intervals, rhythmic figures, and harmonic features within an overall rhetorical approach. Variety and energy, that makes each performance different, as it has to be, as soon as one parameter changes, as this has to result in reactions to that change. Generally, even in the case of a larger group, the impression would be more like that of a chamber music experience.”*

A definition by a novice with a slightly higher, though still very low musical-experience score defined it as:

*“Would have extra ornaments that are in a baroque style, but fuller and more expressive.”*

And a novice definition from a very low experience score as:

*“Probably that it is loud and exciting.”*

Other definitions further articulate the distinction between the novice and the expert. Experts tend to disagree on matters of detail and substance, whereas novices disagree on matters of understanding and basic concepts. This evidence, coupled with the low correlations with experts points to a distinction between the groups which is in need of further investigation. Rather than their criteria being different from experts, the self reported definitions typical of the novice point to a lack of knowledge and hence lack of understanding of the task, rather than their simply possessing a different construct or definition of these aesthetic terms.

(4) Our study suggests that in judging baroque music, novices can adequately determine the performance features of different performances, but (5) Novices and Experts disagree on ratings of aesthetic dimensions such as *preference* and *stylishness*.

## 6. CONCLUSIONS

Rather than providing a simple reduction to the problem of whether to test novices or experts, or whether it doesn't matter, our study presents a more complex picture. However, we will speculate about the implications of some of our findings. First, novices might be a simple, cheap and reliable substitute for making judgements about performance features, provided that ambiguities about terminology are minimised (eg. use terms such as 'legato' in favour of 'articulation'). Second, if the view is taken that the expert musician and scholar is the holder of musical knowledge in our culture, we need greater access to their opinions to help identify and disseminate the critical aspects of aesthetic parameters, such as *stylishness*, *historical informedness* and *expressiveness* for several reasons. (a) Statistically their responses are not as reliable as that of novices, but this is because conventional inferential statistical analysis fails to take into account the possibility of multiple solutions to aesthetic consideration. Such a conclusion explains larger variance in

expert scores. (b) Experts use greater extremes of scales in which aesthetic judgements are made, and these are often significantly different to the responses of novices. It therefore stands to reason that knowledge about aesthetics from experts may need to be filtered down to novices with an interest in enhancing their culturally specific, aesthetic understanding of, in the present case, baroque music. It is important that researchers in expression be cognisant of the outcome that aesthetic responses may not be reducible to a single, 'correct' value, however, they may well be reducible to a couple of values, as reflected by the multimodal responses which typified the expert respondents.

As analysis continues on this project, future aims include identifying what aspects of individual difference among experts distinguishes the multimodal responses. We conjecture that it will be related to variables such as knowledge about performance practice issues, music-educational background/traditions and possibly even age. We also plan to embellish our data with a larger range of musical backgrounds, and to examine how results are varied if responses are made to non-Bach and non-baroque music.

## 7. REFERENCES

1. Bartel, L. R. (1988). *A study of the cognitive affective response to music*. Unpublished doctoral dissertation, University of Illinois at Urbana-champaign.
2. Fabian Somorjay, D. & Schubert E. (2000). The Multidimensional Perception Space of Baroque Music. *Proceedings of the 6th International Conference of Music Perception and Cognition* Keele, England (5-10 August).
3. Fabian, D. & Schubert, E. (2002a). Is there only one way of being expressive in musical performance? - Lessons from listeners' reactions to performances of J. S. Bach's music. *Proceedings of the 7th International Conference of Music Perception and Cognition*. Sydney, Australia (17-21 July).
4. Fabian, D. & Schubert, E. (2002b). Listening to Bach: musical performance features and the perception of expressiveness in Baroque Music. *Proceedings of the 17th Congress of the International Association of Empirical Aesthetics*. Tarakazuka, Japan (4-8 August)
5. Fabian, D. & Schubert, E. (2002c). Bach and overdotting - how empirical investigations can inform performance practice. *Proceedings of the 10th Anniversary Conference 'Musical Creativity'*. European Society for the Cognitive Sciences of Music, University of Liege, Belgium (5-8 April 2002).
6. Fabian, D. and Schubert, E. (2002d). Dotting in Bach: An empirical investigation of performance practice. *SPRMME (Society for Research in Psychology of Music and Music Education conference) 'Investigating Music performance'*, London, Royal College of Music, Centre for the Study of Music performance (CSMP), England (12-13 April).

7. Gromko, J. E. (1993). Perceptual differences between expert and novice music listeners: A multidimensional scaling analysis. *Psychology of Music*. 21, 34-47.
8. Kemp, A. E. (1997). Individual differences in musical behaviour. In D. J. Hargreave & A. C. North (Eds.), *The Social Psychology of Music* (pp. 25-45). Oxford: Oxford University Press.
9. Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
10. Schubert E. & Fabian Somorjay, D. (2000). The Perception of Dotted Rhythms and the Kerning Illusion. *Proceedings of the 6th International Conference of Music Perception and Cognition* Keele, England (5-10 August).
11. Schubert, E. & Fabian, D. (2001). Perception and preference of dotted 6/8 patterns by experienced and less experienced Baroque music listeners. *Journal of Music Perception and Cognition*. 7 (2), 113-132.
12. Schubert, E. & Fabian, D. (2002). Dotting your ears: Some implications of the kerning illusion in the study of performance practice. *Proceedings of the 7th International Conference on Music Perception and Cognition*. Sydney, Australia (17-21 July).
13. Smith, J. D., Nelson, D. G. K., Grohshkopf, L. A., Appleton, T. (1994). What Child Is This - What Interval Was That: Familiar Tunes and Music Perception in Novice Listeners. *Cognition*. 52, 23-54.

## 8. APPENDIX

### Stimuli used in the analysis reported:

#### Adagio, Brandenburg Concerto No. 1 BWV 1046

1. Akademie für Alte Musik Berlin. Harmonia Mundi HMX 2908074.77, recorded: 1999
2. Academy of Ancient Music / Christopher Hogwood. L'Oiseau Lyre 455 700-2, recorded: 1985
- 3 & 4. Academy of St Martin in the Fields / Neville Marriner. Philips 426 088-2, recorded: 1972 (repeated)

#### Allegro, Brandenburg Concerto No. 4 BWV 1049

- 5 & 6. Marlboro Festival Orchestra / Pablo Casals. Sony SRCR 2021, rec: 1965 (repeated)
7. Concentus Music Wien / Nikolaus Harnoncourt. Teldec Das Alte Werk 8.42823 242 925-2, recorded: 1981
8. English Concert / Trevor Pinnock. Archiv Produktion 410 501-2, recorded: 1982

#### Sarabande, D minor Partita for Solo Violin BWV 1004

9. Arthur Grumiaux. Philips 438 736-2, recorded: 1961
10. Sergio Luca. Electrola Nonesuch 7559 73030-2, recorded: 1977
11. Filtered S. Luca (same as 10, but filtered so as to emulate a 1930s recording)
12. Menuhin. EMI Reference CHS 7 63035 2, recorded: 1934

#### Adagio, Violin Concerto in E Major BWV 1042

13. Elizabeth Wallfisch / Orchestra of the Age of Enlightenment. Virgin Classics 72435 61558 2 4, recorded: 1999
14. Frank Peter Zimmermann / English Chamber Orchestra. EMI 7243 5 72559 2 9, recorded: 1998