Baroque expressiveness and stylishness in three recordings of the D minor Sarabanda for solo violin (BWV 1004) by J. S. Bach

Dorottya Fabian and Emery Schubert

The University of New South Wales, Sydney, Australia

ABSTRACT: The *perception* of baroque music has until recently been somewhat neglected in music research. However, the existence of recordings documenting various styles of performing baroque music offers researchers the opportunity to investigate the efficacy of musicological claims about historically informed performance (HIP) practice from the perspective of the modern listener. The paper examines the existence of the perceptual variable of Baroque Expressiveness (BE), and its relationships to other aesthetic variables. Responses were collected from 31 musically trained participants to recordings of the D minor Sarabanda for Solo Violin by J. S. Bach made by Menuhin (1934), Grumiaux (1961) and Luca (1976), and to a filtered version of the Luca that sounded as if it was made in the 1930s. A variety of performance and aesthetic parameters were rated. Results indicated that recording quality did not influence judgements of BE. Participants were able to distinguish BE or lack thereof across the three recordings in the predicted way. Success of Performance and Stylishness were highly correlated with BE and in a regression analysis were each best predicted by perceived rhythmic flexibility. However, subtle distinctions could also be identified in the perceived performance features used across these aesthetic variables. Furthermore, partial correlations showed a stronger relationship between Success and Stylishness than BE and Stylishness. This indicates that different types of expressivenesses exist and BE is not the sole arbiter of success. Our study provides an example of the possibility of experimental examination of musicological discourse.

KEY WORDS: Historically informed performance, aesthetic response, sound recordings, music performance, perception

BACKGROUND

Although in his recent book Bruce Haynes (2007) pronounced the "end" of Early Music, investigations of what constitutes a stylish performance of baroque music have not been exhausted. In particular, the relationship between what is regarded "stylish" and what is regarded "expressive" in a "historically informed" sense has not been examined in a systematic way. Promoters of historically informed performance (HIP) claim that interpretations that utilize period instruments and playing techniques described in historical treatises project a sound that is closer to the one envisaged by the composer and thus more suitable for creating an aesthetically successful performance. Matching more closely the compositional style, the performances are, by implication, more "stylish". Although originally writing to address claims of historical "authenticity" by advocates of a particular style of HIP to be described below, critics of this view, led by Daniel Leech-Wilkinson (1984) and Richard Taruskin (1982, 1984, 1988), argue that there is precious little historically informed about such performances. Audiences consider them "stylish" because they sound crisp, clean and sleek, and thus match expectations of modern taste.

An interest in historical performing traditions can be traced to at least the end of the nineteenth-century. Its modern foundations have been laid by pioneers such as Dolmetsch (1949) and Landowska (1904/1964); it developed rapidly after the Second World War and gained momentum in the 1970s, eventually becoming the unquestioned alternative around the mid-1990s (Haskell, 1988; Butt, 2002; Fabian, 2003; Haynes, 2007). Importantly, the performance characteristics of what is considered HIP have evolved considerably with the passing of time (see next section). Therefore it is not unreasonable to view the movement's development as a reflection of changing taste (Taruskin, 1984; Leech-Wilkinson, 1984, p. 14) and an ever-recurring reaction of one generation of performing musicians to the style of their forbears (Butt, 1999). The various stages and metamorphoses of the HIP style – reflected also in the changing name of the movement from Early Music through Authenticity to Historically Informed Performance Practice – highlights the difficulties in trying to recreate lost traditions of distant generations.

Although the philosophical assumptions and practical achievements of HIP have been much debated since the late 1980s (e.g. Kenyon, 1988; Taruskin, 1995; Kivy, 1995; Fabian 2001; Butt, 2002; Walls, 2003), there has been little published research, as far as we could ascertain, that gathered empirical evidence regarding listeners' perception of different *performing traditions*; the performance features that listeners associate with various aesthetic qualities.

Of particular import is the question of expressiveness in musical performance because it is closely linked to the questions of what constitutes a stylish and a historically informed interpretation. What does it mean to be expressive in a baroque piece as opposed to a piece in another style? How does one play baroque music expressively? What are the performance features that generate a stylistically appropriate expressiveness? Are they the same across all styles and historical periods? Can the appropriateness of expressive devices be differentiated objectively on the bases of historical documents or compositional styles and aesthetics? And, crucially for the present investigation, can listeners hear these nuances in expression? The answers to these questions are embedded in the claims of HIP practitioners as well as their critics. Yet they remain untested assertions, however plausible, unless

we are able to provide some empirical evidence for or against them.

To investigate these issues a series of listening experiments was devised. In the first stage (Schubert & Fabian, 2006) a semantic differential study was conducted in which 44 listeners rated multiple performances of several pieces by Bach using 40 baroque performance descriptors to identify the underlying dimensions of the baroque music listening experience. Two factors emerged that were labelled Stylishness and Expressiveness. Stylishness was associated with items that were indicative of the success of the performance, such as high ratings of the Good Performance and Stylistically Appropriate scales and also on more specific ones, like All Baroque Instruments, Articulated, Detailed, Well Ornamented, and so on. The Expressiveness factor reflected changes from no expression through to highly or even overly-expressive. To investigate the issue further, in particular the kind of expressiveness that listeners associate with baroque style, we conducted a larger study which examined in more detail the relationship between performance characteristics and aesthetic judgements in a variety of pieces in multiple performances. Here we report the results of one of the experiments in which we tested the role of particular performance features in creating Stylish and Baroque Expressive interpretations of a sarabande by J. S. Bach. The aim was to refine our understanding of the interrelationship between aesthetic variables. We wanted to find out whether performances that utilized historically documented playing conventions were perceived as merely stylish or whether such interpretations were also judged to be expressive and/or successful. The term Baroque Expressiveness was coined and first reported by Fabian and Schubert (2002) to explore the existence of alternative style-specific expressive characteristics.

The musicological context

When discussing performance parameters that potentially discriminate between expressive interpretations, the majority of pre-1990 literature on baroque performance practice tends to emphasize the role of period instruments, ornamentation and tempo in creating a historically stylish rendering (Donington, 1989; Frotscher, 1963/1980; Neumann, 1978; Haskell, 1988; Kenyon, 1988; Elste, 2000). Authors of analytical studies of the 1950s and 1960s who discussed the stylistic features of baroque music laid emphasis on motor rhythms (e.g. Cone, 1968) and highlighted the rhythmic drive such textures generate while remaining oblivious of the metric hierarchy, so important to eighteenth-century theorists, and discussed at great length in historical treatises (Houle, 1987). All this (and possibly other cultural factors not discussed here) led to a literalistic and metronomic approach to playing baroque music, especially in England, from about the 1940s to the 1980s and was duly criticized by Leech-Wilkinson, Taruskin and others in their wake. At the beginning it was called "sportive" (The Gramophone, 1948, p. 118), "direct and forthright" (The Gramophone, 1972, p. 679), and described as having a "springy rhythmic liveliness" (The Gramophone, 1976, p. 1587). With hindsight it was labelled "sewing-machine style" (Dreyfus, 1983) for its unrelenting, clockwork-like evenness of rhythm and dynamics, and repression of expressive flexibilities: the pre-1980 "'authentic' performance is characterized by relatively uniform tempo and dynamics, a 'clean' sound and at least an attempt to avoid interpretative gestures beyond those notated or documented as part of period performance practice" (Leech-Wilkinson, 1984, p. 14).

In contrast, music making on the European Continent (especially in Holland and Austria) has developed rather differently since the mid-1960s, eventually challenging the status quo in the English speaking world and becoming the generally accepted HIP style since the mid-1990s. These performances often seem to come close to displaying techniques and producing sounds described in eighteenth-century instrumental tutors and theoretical treatises as we understand them at the beginning of the twenty-first-century. At first they sounded so novel and strange that they were scorned by critics who labelled them mannered or eccentric (see Fabian, 2003, pp. 43-46). Subsequently musicologists (e.g., Houle, 1987; Butt, 1990, 1991; Lawson & Stowell, 1999; Fabian, 2003; Haynes, 2007) codified their characteristics, identifying the eighteenth-century performance style as consisting of locally nuanced and clearly punctuated articulation, well defined metric groups and strongly projected/inflected rhythmic gestures, shallow and selectively used vibrato, and a general revelling in the characteristics of eighteenth-century instruments (e.g., the uneven bow strokes, the variety of tonguing patterns and their effect on tone qualities). These performance features create what is described as the "rhetorical delivery", that is, an interpretation that "speaks" (Harnoncourt, 1982/1988). Eighteenth-century aestheticians compared music to speech and drew analogies for performance and composition from the arts of public speaking (oration) and rhetoric (Bartel, 1997; Bonds, 1991). Period instrumental techniques developed in accordance with this prevailing aesthetic view and clearly punctuated articulation was embedded in every player's technique (Houle, 1987). This fostered rhythmic flexibility through rapidly changing timbre while highlighting metrical hierarchies.

Systematic analysis of over 100 recordings of J.S. Bach's compositions representing 30 years of mid- to late-twentieth-century playing of baroque music provided much evidence for this re-evaluation (Fabian, 2003). This examination showed that articulation stemming from period playing technique, pulse and metre, together with rhythmic flexibility and tone production, are crucial components of an interpretation that aims to approximate the performance style described in historical documents as desirable and effective. It also showed that performances utilizing these techniques were less "uniform [in] tempo and dynamics" and did not eschew "interpretative gestures beyond those notated or documented" (Leech-Wilkinson, 1984, p. 14). They were more expressive to such an extent that commentators have started discussing their style in terms of "neo-romanticism" (e.g. Dulak, 1993) even though their means of achieving the expressive quality seem to be entirely different from those associated with playing romantic music (Fabian, 2003).

Given the variety of approaches to performing baroque music, it is crucial to investigate experimentally the relationship between perceived performance features and aesthetic judgement. This would be a step towards a systematic accounting for the relative importance of certain interpretative and technical means of expression in creating performance style. As it is impossible to recreate the eighteenth-century listener, the study does not aim to judge historical verisimilitude. Rather, it aims to find out how well listeners are able to distinguish between different performing traditions and, whether they consider stylish the modern, "dispassionate" way of playing a baroque piece or the rhythmically flexible, locally articulated and phrased manner that is currently considered HIP. Although it might be unjustified to claim "more authenticity" for either of these approaches, the latter way of playing baroque music certainly recreates many more specific techniques and features described in period treatises and instrumental tutors than the "dispassionate", literalistic style

that emphasizes motor rhythms. If it is just a matter of personal taste unfettered by a dominant cultural consensus then judgement will be random or fairly equally distributed among the performing approaches exemplified. If a particular approach is consistently rated as being more stylish than another, this may indicate either enculturation or natural support for that particular performance aesthetic. Neither would mean that the approach is historically authentic, but both would at least allow us to gain an insight into whether, or to what extent, listeners uphold musicological claims about the relationship between performance features and overall aesthetic effect. As musical education is likely to have an impact on the judgement as to whether a performance is historically informed, listeners' views and beliefs about HIP must be examined for a valid interpretation of their responses.

To sum up, some scholars infer that there are specific parameters that performers manipulate to achieve particular kinds of expression. These reflect different aesthetic convictions regarding what might be a stylistically appropriate expressiveness. Musicologists have started to chart stylistic changes in twentieth-century performance history (Day, 2000; Butt, 2002; Philip, 2004; Leech-Wilkinson, 2009). In relation to baroque music at least three stages have been identified representing different approaches: an "expressive-emotional" stance most typical up until the 1930s, a "modernist-literalistic" position that followed in the 1940s-1970s and was considered "historically authentic" at the time, and the "baroque-expressive" that is currently considered HIP. Naturally, there is considerable diversity in each of these periods and categories. From a historical point of view it might be more fruitful to think of a continuum rather than categories. But for the sake of investigating the problem at hand experimentally these three well-established and broadly defined categories are useful.

The availability of sound recordings documenting these performance traditions makes it possible to test some of the musicological claims regarding expressivity, stylishness and success from the perspective of today's listener. By selecting a piece composed during the baroque period which has been performed in ways now regarded baroque appropriate (HIP), modernist-literalistic and expressive-emotional, and obtaining responses to these recordings, we are in a position to (1) examine whether listeners are able to detect differences in performing styles and expressiveness, and thus (2) to further interrogate the nature of Baroque Expressiveness, by (3) identifying salient links between perceived performance features and aesthetic qualities. Possible correlations between perceived performance features and aesthetic variables will illuminate the basis on which listeners form their judgements and whether these reflect the views contained in historical treatises or not. By using a novel term, *Baroque Expressiveness*, we invited listeners to rely, at least partially, on instinctive discrimination between performances because the term is not as loaded with cultural meaning as "success" and "stylish" are. In doing so we hoped to see whether these variables measured the same aesthetic constructs or whether they made contributions to different aspects of the perceived aesthetic experience.

METHOD

Stimuli

Based on extended familiarity with commercially available sound recordings of J.S. Bach's compositions (Fabian, 2003; Fabian, 2005) for this experiment we chose three recordings of the Sarabanda (bars 1-8) from the D minor Partita for solo violin (BWV 1004). The piece was chosen because (1) it is unaccompanied and thus performance decisions are easier to identify, (2) violin music offers the opportunity to directly contrast performances on modern and eighteenth-century violins and comparable instrumental techniques, and (3) sarabandes are slow movements that are supposed to have expressive potential. As it is a popular piece we studied over 40 recordings that covered the range of expressiveness and performance styles under examination. The chosen sound recordings represented three generations of twentieth-century violinists: Menuhin (1916-1999), Grumiaux (1921-1986) and Luca (1943-).

We assert that the interpretation of Yehudi Menuhin (recording date: 1934) is an expressive-emotional performance (click here to listen). It is quite slow (average tempo = 35 beats per minute [bpm]) and employs heavy legato bowing and intense, somewhat fast and wide vibrato (average rate = 7 cycles per second [cps]; average width = 0.5 semitone [sT]).

We claim Arthur Grumiaux's version (recording date: 1961) to be modernist-literalistic and therefore not stylish by today's [2009] HIP standards (click here to listen). When first released, it was described in The Gramophone (1962, p. 458) as "straightforward with an additional element of scholarship (or perhaps historical taste)" and it exemplifies the anti-romantic, neo-classical aesthetic that refrains from overt emotionalism, "letting the music speak for itself" (Taruskin, 1982, p. 458). It is quite fast (average tempo is 51 bpm compared to Menuhin's 35 and Luca's 44 while the average among all studied recordings is 39), with fairly homogenous dynamics, fast but not too intense vibrato (rate 7.1 cps; width 0.4sT), straight rhythm and clear, bright tone.

We posit that the rendering by Sergiù Luca (recording date: 1976) is an example of HIP and predict it to be rated the most stylish (click here to listen). It was the first commercial recording in which the performer used a baroque bow and explicitly aimed at recreating historical practices (as outlined in the liner notes). It is rhythmically flexible, articulated in small gestures, has much less vibrato (rate: 7.1 cps; width 0.15sT; but few notes played with vibrato), and a shaded tone reflecting the characteristics of the baroque bow. See Discography for full details.

One of the weaknesses of the design of such a study is that the recording quality may provide clues about the age and therefore both stylishness and expressiveness of the recording. For example, participants might have rated Menuhin's performance as less stylish and (possibly therefore) less baroque expressive than the other performances, simply because of its poorer recording quality. Our solution was to include an additional, degraded version of the most recently recorded and supposedly baroque expressive of the selected examples (click here to listen). As the long term average spectrum of the modified recording shows (Figure 1), the frequency response of the "filtered Luca" recording resembles the 1930s frequency response of the Menuhin recording more than it does the frequency response of the Grumiaux and original Luca recordings. The Luca (original) and the Grumiaux have a gradual roll-off (Ballou, 2005) toward the higher frequencies, compared to the

sudden roll-off from 5-7kHz and 3-5kHz for the Menuhin and filtered Luca respectively. At frequencies higher than the roll-off, some residual energy can be observed in the Menuhin recording at around -80dB below the peak intensity, which is lower in the Luca recording (less then -90dB below the peak intensity). However, at -80dB below the peak intensity the signal is barely audible, and was not expected to produce problems in the listening task. In fact, it is evident that the commercial Menuhin recording has also been filtered to reduce high frequency noise. The lower frequency cut-off point for the Luca was deliberate to try to minimise its resemblance to the original Luca recording. In future experiments noise (such as hiss, pops and clicks) could be added to further mimic the quality of historical recordings as was done more recently by Timmers (2007).
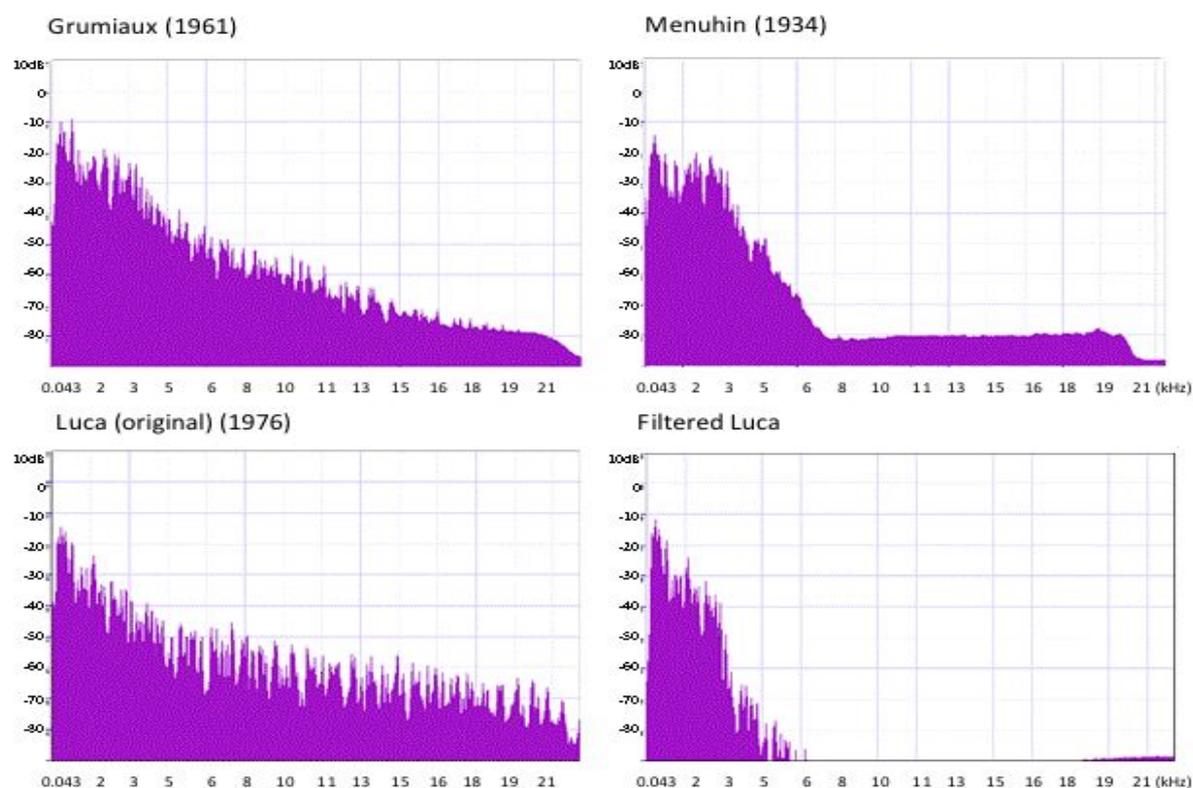

Figure 1: Long-term averaged spectra of the four recordings used.

Participants

Another problem with such an experiment is the choice of listeners. How does one control for educational and cultural influences? Music students may be indoctrinated to hear and label nuances in a particular way and to approve of HIP. Listeners who are not musically educated may be unable to answer questions as they may not know what the musical terms mean. We attempted to control for the problem by using listeners of varied musical backgrounds and interests, a mixture of technical and general descriptors of performance features, and by asking participants about their understanding of and beliefs about the two key concepts under investigation: Stylishness and Baroque Expressiveness.

31 volunteers took part in the study aged between 18 and 53 (median = 26.6) years. Twenty-four of them were Bachelor of Music students at the beginning of their second year of study at the University of New South Wales (Sydney, Australia). Some of them were jazz majors (n=6) and none of them had yet taken a course in historical performance practice. The other participants (n=9) were more experienced musicians with an interest in eighteenth-century performance practice. We created a Baroque Experience Score for each participant based on their answers to questions detailed under Procedures. Five reported never having listened to baroque music and seven reported never or rarely playing baroque music while only two claimed to listen to baroque music for more than 20 hours a week. The distribution of baroque music listening and playing experience is shown in Tables 1 and 2. There is a wide distribution of listening and playing backgrounds among the 31 participants, with the mode being 0-2 hours of listening to baroque music per week (n=14), and playing a musical instrument as an amateur (n=13).

Table 1: Participants' self report regarding their baroque music listening time per week

Baroque Music Listening hours per week	None	0 to 2 hours	2 to 4 hours	4 to 10 hours	>20 hours
Number of Participants	5	14	9	1	2

Table 2: Participants' self report regarding their level of playing an instrument

Playing level	Don't play any musical instrument of any kind	Play an instrument but never (or rarely) baroque music	Play(ed) as an amateur	Play(ed) as a free-lance or semi-professional	Play(ed) as a professional
Number of Participants	1	7	13	9	1

Rating scales

Twelve scales, grouped into three categories, were rated by the participants. The categories (and corresponding scales) were: *Phrasing* (Continuous, Legato, Articulated, Detailed), *Tone Production* (Vibrato, Intensity, Straightness, Lightness), and *Rhythm* (Grouped, Strict, Measured, Flexible). The categories and scales were chosen on the basis of literature on the characteristics of various performing conventions and included at least one general term that a non-musically educated person could also understand. For instance, Articulated and Detailed are, in this context, identical descriptors, but the latter is probably more meaningful to a non-musician. Each scale had a five-point range (very [+2], slightly [+1], neutral [0], slightly lacking [-1], completely lacking [-2]). Participants had to rate at least one scale in each category that they considered to be the most relevant for describing the particular musical example, but were free to rate as many as they wished. They also rated a further three aesthetic categories: Baroque Expressiveness, Stylishness, and Success of Performance. The scale values were the same. All scales were presented as pop-up menus on a computer interface.

Procedure

The data collection was controlled by software written by the second author. Participants filled out a questionnaire regarding their general and baroque music experience. Specifically we asked four questions: Their age, instrument, playing level, and the number of hours per week they spent listening to baroque music (cf. Tables 1 and 2). They were also asked to describe what their understanding was of the terms Stylishness and Baroque Expressiveness. Following this they started listening to the examples and rated each on specific scales as described above. They were encouraged to report their first response but were allowed to listen to the same stimulus up to four times. Once they submitted their response they could not return to that example. The piece was identified for them but not the performer, nor details of the recording. The performances were presented in random order.

RESULTS AND DISCUSSION

General definitions of stylishness and baroque expressiveness

The participants' responses to general definitions of Stylishness and Baroque Expressiveness (gathered prior to the listening tasks) were collated. The range of indicative answers for Stylishness is illustrated by the following selection of statements ordered from more to less specific:

1. Played with character, conviction and in the appropriate style for the piece, with correct articulation, dynamics etc.;
2. Within the style of the baroque period: little/no vibrato, played with the baroque performance technique in mind, not over done;
3. Stylish would indicate to me a pleasant use of ornamentation, clear tone, with confidence and competence;
4. Stylish: played in a style appropriate to the period, with details such as trills and rhythm (e.g., double dotting etc.) taken into consideration;
5. One which adheres to the rules of performance of the time e.g. some features as stated in various literature on Baroque performance practices;
6. Attention to all the markings on the music, and how they are performed in the characteristic style of that particular composer;
7. This description often means that the listener has very little understanding of the performer or the music, however if they do, it means the performance was played in a fashion which followed the performance conventions of that time considered to be correct; and
8. Stylish to me does not give a positive impression...sounds like the performer is trying to do too much to create an impression rather than performing to the intents of the music.

Most commonly participants conceptualised Stylishness as being in the correct style, i.e. as a synonym for historically informed. Some of the responses provided quite specific criteria; others were more general, at times quite unclear. Apart from a small number of responses (e.g., 3 above), the definitions include the term "style" making those parts of the definition

of limited practical value (defining a word in terms of itself). However the features and characteristics mentioned generally reflect the scholarly literature. Interestingly some of the less specific definitions demonstrate some suspicion of the use of the term (e.g. 9). While not adhering to the cultural norms, these seem to indicate support for the more arbitrary nature of the concept as argued by Taruskin (1982, 1984) and Leech-Wilkinson (1984).

Participants' understanding of the newly coined term Baroque Expressiveness appeared to be diverse and often vague (e.g., "not as expressive as romantic or twentieth-century music"). Some reflected traditional emphases (e.g., delivery of ornaments, restrained expression); others seemed to be aware of more recent views regarding eighteenth-century performing conventions (e.g. importance of articulation, using lower positions and less vibrato). Three claimed not to know what the term meant. Others offered statements such as (listed from general to more specific content order):

1. Easy to listen to;
2. I think in Baroque period … [musicians] make a big difference between melody line and accompaniment;
3. They have interpreted the music in a more passionate way, still staying in the limits of a baroque technique: still uses vibrato as an ornamentation, but the players have a very good skill ability in playing within baroque technique, so they have played it expressively, without making it sound romantic;
4. Expressing a mood or feeling using the elements such as tone, timbre and dynamics, creating an emotion in the listener;
5. The performance is expressive in a way that the piece played retains its essential Baroque characteristics and style;
6. Moderate use of dynamics and effects, restrained personal expression, attention to appropriate articulation;
7. An understanding of the "affects" and where they occur within the work and an interpretation which fully "realises" them in a historically accurate way;
8. A baroque performance that was played expressively in the baroque sense would refer to dynamics and how they were played, and to ornamentation - in particular things like trills, and whether they began on the upper or lower note;
9. A baroque expressive performance would perhaps include the minimal use of rubato and vibrato. For strings it would mean staying in lower positions as much as possible, with the use of open strings in most cases. Also perhaps that annoying swelling sound they do (soft growing loud on longer notes);
10. Convincing characterisation of each musical pattern with a tone quality to match (i.e., not uniform in evenness or vibrato) but rather choosing the inflection according to the rhythmical principles of the music; and
11. Identical to stylish, with the addition, perhaps, of factors like timing fluctuations, ornament-like use of portamento and vibrato, etc.

These answers indicate that the level of familiarity with baroque performance conventions as outlined in relevant sources was varied, at times quite limited, because of the few explicit mentions made of characteristics that we proposed to be Baroque Expressive. Furthermore, the two key terms – Stylishness and Baroque Expressiveness – appear to have some shared meaning (e.g., 11). A few of the participants reiterate doctrines now regarded mis-

leading (e.g. 2 and 6), others remain general (e.g. 3). Some seem to recall specific doctrines such as the execution of trills (8) – a preoccupation of the 1960s and 1970s now regarded as pedantic and not so important. So the concept of baroque expressiveness as represented by self-reports is far from clear cut. This may be a reflection of the invented nature of the term, and perhaps its limited value.

Comparison of recordings

Next we analysed participants' judgements of the three performances of the Sarabanda and the filtered version of the Luca recording. A repeated measures ANOVA was conducted using the four performances as the within-subjects factor for each of the response variables. As shown in Table 3, significant differences were obtained for Continuous, Articulated, Vibrato, Lightness, Baroque Expressiveness, and Stylishness ratings.

Table 3: Repeated measures ANOVA summary table

| Variable	Mauchly Sphericity Test Significance	Hypothesis df	Error df	F	p
Continuous*	0.201	3.000	33.000	3.253	0.034
Legato	0.003+	1.618	17.800	1.375	0.274
Articulated**	0.070	3.000	42.000	4.647	0.007
Detailed	0.088	3.000	30.000	2.172	0.112
Vibrato**	0.255	3.000	33.000	15.564	0.000
Intensity	0.235	3.000	45.000	2.807	0.050
Straightness	0.158	3.000	21.000	0.881	0.467
Lightness**	0.551	3.000	24.000	5.395	0.006
Grouped	0.740	3.000	18.000	0.889	0.466
Strict	0.148	3.000	30.000	0.752	0.530
Measured	0.920	3.000	24.000	1.000	0.410
Flexible	0.816	3.000	36.000	1.679	0.189
Baroque Expressiveness**	0.009+	2.343	60.928	5.469	0.004
Stylishness**	0.253	3.000	75.000	6.445	0.001
Success of Performance	0.249	3.000	33.000	1.819	0.163

*$p < 0.05$; **$p < 0.01$.*

+ Greenhouse-Geisser adjustment used when sphericity assumption violated (at significance level of 0.05).

Repeated-measures t-tests using Bonferroni adjustments (see further, Keppel & Zedeck, 1989) were used to identify which performances were significantly different from one another for each response scale. These differences are marked by asterisks in Figure 2. The figure shows the average values for each of the scales rated grouped by performance.

In general, the greatest contrast can be seen between the ratings of Menuhin's and Luca's performances, with Grumiaux's being in the middle. The Menuhin has a significantly higher rating of Continuous than any of the other recordings. The Luca and the Grumiaux are perceived as more Articulated than the Menuhin, and the Luca is perceived as being Lighter and using less Vibrato than the other two performances. Luca received the highest rating for both Stylishness and Baroque Expressiveness, while Menuhin received the lowest. Importantly, the rating of the Stylishness scale was significantly different for all three recordings (with the filtered Luca being significantly different only from Menuhin's version). At the same time there was no significant difference between the Success of the Performances.

With respect to the original Luca performance, the filtered version was rated as relatively less Articulated, less Grouped and Detailed (matching the Menuhin ratings) and having even less Vibrato (the significant difference between Luca and both the Menuhin and Grumiaux is increased compared with the significant difference between Menuhin and Grumiaux), while remaining significantly Lighter than the other two versions and more Stylish than Menuhin's. Interestingly, it was still considered the second most Baroque Expressive and Successful Performance (second only to its original version). So, according to the mean participant responses, the filtered Luca recording appears to share some characteristics with the Menuhin, but still retains HIP qualities. This suggests that the participants were *nearly* deceived into believing that the filtered Luca was an older recording, because certain performance characteristics were reported in a slightly different way (for example, Detailed and Vibrato). We take this as evidence that recording quality, at least in this case, does *not* influence judgement of whether the performance is Baroque Expressive. This finding has some parallels with Timmers (2007) who found that listeners' affective ratings were not influenced by the age of a recording or the effects of simulated age (through digital processing of the sound recording). Furthermore, while the Stylishness of the filtered Luca is still the second highest, it is nevertheless noticeably lower than its Baroque Expressiveness. Participants may have thought it was less stylish because it sounded "older" and less up to date. On the other hand while a statistically significant difference remained between the Menuhin and the filtered Luca recordings on the Stylishness scale, no such difference was found between the two on the Baroque Expressiveness scale. These results indicate that participants were, to some extent, able to make distinctions in Baroque Expressiveness independently of Stylishness.

This interpretation of the data has important implications for the view that assumes that people will judge a performance historically informed (in this case Baroque Expressive) when they find it stylish. Our analysis indicates that the story is more complex. Stylishness judgements may be more susceptible to factors such as an old sounding recording, whereas the ratings of Success and Baroque Expressiveness are more robust because they are probably tied to the detection of specific performance features, regardless of recording quality.
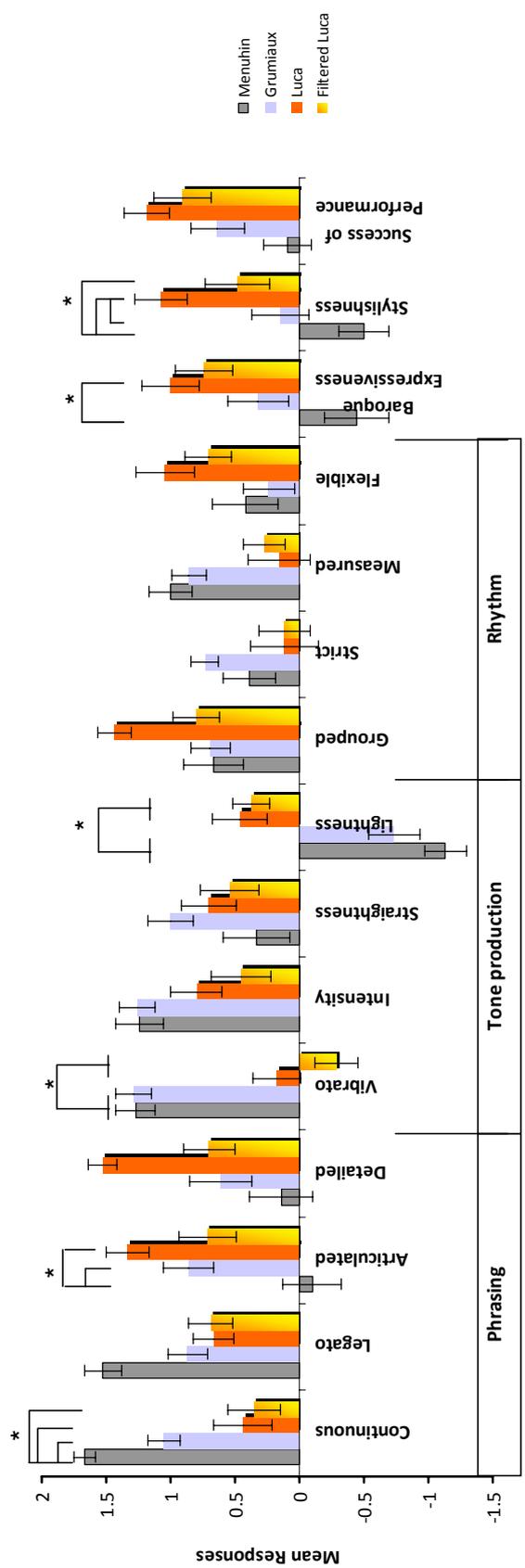
Figure 2: Mean responses and ±1 SE along each participant-rated scale (+2 [very], − 2 [completely lacking]) for the listed recordings of Bach's Sarabanda (bars 1-8) from Partita in D minor for solo violin, BWV 1004. Asterisks denote pairs of performances that are significantly different from one another according to post-hoc repeated measures t-tests with Bon-ferroni adjustment.

The nature of the aesthetic variables

A Pearson's correlation analysis among the three aesthetic scales was conducted (Table 4). All scales were highly correlated (at $p = 0.01$), with the highest coefficient being between Stylishness and Baroque Expressiveness. In itself this result would support the view that listeners consider a performance to be stylish when they believe it to be historically informed. However, the correlation with Success of Performance was slightly lower; perhaps indicating that listeners are sometimes able to make judgements of performance styles independent of their approval. For example, it may be possible to conclude that a performance of the Sarabanda was Successful even if it was not Baroque Expressive. The recordings were all made by highly successful and respected performers and it should also be kept in mind that the Success ratings were generally high for all recordings (none had a mean rating below zero), and therefore the reduced variability in the data may have contributed to the slightly lower correlation coefficient.

Table 4: Correlation and partial correlation matrix for aesthetic variables

	Baroque Expressiveness	Success of Performance
Success of Performance	.658[**] (pr .233)	
Stylishness	.753[**] (pr .448***)	.714[**] (pr .351**)

*N=69, ** p < 0.01, *** p < 0.001*

A possible criticism of our interpretation of these analyses may be that we are reading too much into small differences in correlation coefficients which are all significant and relatively high. Perhaps each of these variables is part of the same aesthetic construct? We can further test this possibility by examining (1) partial correlations and (2) whether performance features predict these aesthetic responses, and if so, to what degree.

To see whether any pairs of these aesthetic response variables were more strongly correlated with the influence of the third variable controlled, we ran partial correlations. When the common influence of Stylishness was excluded, Baroque Expressiveness and Success of Performance had a non-significant correlation (pr = 0.233, $p = 0.056$). When the common influence of Success of Performance was excluded, Baroque Expressiveness and Stylishness had a correlation of 0.448 ($p < 0.001$). When the common influence of Baroque Expressiveness was excluded, Success of Performance and Stylishness had a correlation of 0.351 ($p = 0.003$). This result suggests that Stylishness is correlated with both Baroque Expressiveness and Success. However, when Stylishness is kept constant, there is only marginal correlation between Success and Baroque Expressiveness. Thus Success is more related to Stylishness specific characteristics but not with features specific to Baroque Expressiveness. The result is illustrated in Figure 3.

Although the three aesthetic variables are all interrelated, it is noteworthy that Baroque Expressiveness seems to stand somewhat apart when the success of a performance is considered. This suggests that interpretations that utilize different kinds of expressive means may all be considered successful. Therefore, if our interpretation is correct, Baroque Expressiveness is not the sole arbiter of success. This may appear to be contradicting the re-

sults of the comparison of the individual interpretations since the Baroque Expressive performance of Luca (and also its filtered version) received higher ratings of Stylishness and Success than the other two versions (Menuhin and Grumiaux). However, the partial correlation provides a more subtle analysis because it captures responses to all pieces and all ratings, providing greater variance in responses than performance-wise analysis alone.

The correlation analysis suggests subtle differences between Baroque Expressiveness and the other two aesthetic scales. To see if these might be genuine differences we wanted to find out if they would be discriminated through different associations with the rated musical features.
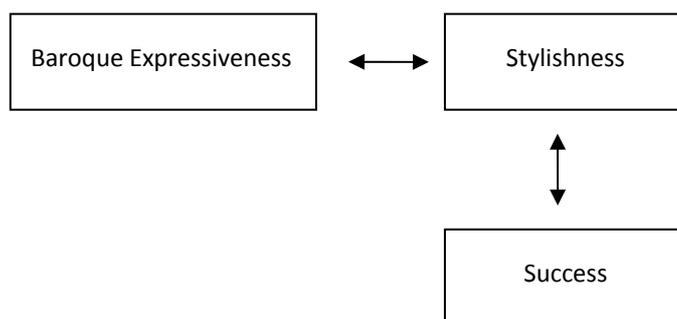

Figure 3: Interrelationship between aesthetic variables showing Baroque Expressiveness to stand apart from Success. Different kinds of expressive styles may all be judged as contributing to a relatively successful performance.

Three stepwise regression models were generated — one for each aesthetic variable — with musical feature ratings used as predictors. The results are summarised in Table 5. The Success of performance model explained the largest amount of variance (73%), as a function of two variables: Articulated [phrasing] and Flexible [rhythm], each with positive coefficients. In other words, the more flexible and articulated the performance was perceived to be, the more successful it was. The Stylishness model explained the next most variance (63.4%) as a function of three variables: Flexible [rhythm], Detailed [phrasing] and Straightness [of tone]. The Baroque Expressiveness model also explained a considerable amount of variance, though less than the other two models (44.3%). Baroque Expressiveness was predicted by two variables: Lightness [of tone] and Flexible [rhythm].

The unifying predictor across each aesthetic response was perceived rhythmic flexibility: A rhythmically flexible performance was likely to be rated as Baroque Expressive, Stylish and Successful. Therefore rhythmic flexibility should be regarded an overarching performance characteristic underlying all three aesthetic variables tested. Such a feature is foreign to the modernist literalistic approach which eschews flexibility of any kind (because of its "ambience of emotional detachment" (Taruskin, 1988, p. 189)) but is advocated by the historical sources. These and HIP theory current at the time of writing [2009] posit that rhythmic flexibility is crucial to projecting a metrical hierarchy within bars, to grouping notes into gestures, and thus to creating the highly inflected speaking quality — the ideal upheld by the many references to and analogies with rhetoric in baroque treatises. This is of particular interest given that the analyses collapsed responses together across the four performances.

So, even though Menuhin and Grumiaux were not unsuccessful, rhythmic flexibility – exhibited to a greater degree in the performance by Luca – is still seen as an important feature. The other performance variables contributing to the aesthetic models also reflect a match with HIP theory: Straightness and Lightness of tone, and Articulated and Detailed phrasing are features often mentioned in discussions of the constituents of baroque performing conventions. In sum, there is a great deal of overlap in the three-way relationship between what is regarded a Stylish, a Successful and a Baroque Expressive performance.

Table 5: Regression model summary showing how performance features could be predicted by three aesthetic variables

	Baroque Expressive	Stylishness	Success
Articulated phrasing			0.595
Detailed phrasing		0.387	
Lightness of tone production	0.451		
Straightness of tone production		0.359	
Flexible in rhythm	0.367	0.51	0.418
Adjusted R^2	0.443	0.634	0.73

Standardised coefficients for variables accepted into any of the stepwise regression models for the three aesthetic variables under investigation using F probability = 0.05 for entry and 0.10 for removal. The final row shows the adjusted R^2 for each model. The seven perceived performance feature variables not shown were left out because they did not produce significant coefficients for any of the three models.

It is important to examine in more detail the subtle distinctions among the performance features that contribute to each of the three aesthetic variables. These underscore the results of the experiment; listeners were able to perceive differences in performance features, which in turn influenced their aesthetic evaluation. They tended to rely on articulation to provide a finer grade of distinction in the Success of the performance. Lightness of tone production was used to further distinguish Baroque Expressiveness. Stylishness called in to play two further perceived performance features which did not contribute to either Success or Baroque Expressiveness: when the phrasing was perceived to be more Detailed and the tone Straighter, the performance was rated the more Stylish without it being necessarily more Successful or Baroque Expressive.

Given our earlier assertion that Detailed and Articulated are interchangeable terms, the former being easier to understand for musically less educated listeners, it is likely that the Stylishness judgement is more reflective of intuitive listening than deliberate reporting based on learnt musical knowledge. This interpretation of the data is supported by the third scale – Straightness of tone – that contributed most to the Stylishness rating. Taruskin's papers (1982, 1984, 1988) imply that it is this clarity of detail and crispness of tone that is fashionable with modern listeners.

In clarifying the subtle differences, it is worth noting that the third variable contributing to Stylishness was *not* Vibrato but Straightness (of tone). Several primary and secondary sources advocate limited and selective use of vibrato and thus, presumably, a relative straightness of tone production. Yet it is not evident that the relationship between these

two scales is complementary because eighteenth-century instruments tended to have an uneven tone quality. Vibrato and straightness could therefore refer to different aspects of tone quality, as can be seen in the contrasting ratings of Grumiaux's and Luca's recordings on these two scales: the former being rated highly on both (i.e. regarded as being played with a lot of vibrato, and simultaneously very straight), while the latter is rated low on both (i.e., regarded not at all vibrato yet not at all straight either). Furthermore, according to the ANOVA, there was a significant difference in perceived Vibrato between Luca's and the other two versions, whereas the Straightness of tone scale did not distinguish any of the three recordings from each other. One could assume that Straightness of tone reflects something more complex than the perceived amount of vibrato, probably something to do with the "directness" or "stability" of the tone. We therefore propose that listeners find the "cleanness" of the sound — whether it be in relation to the tone quality of the instrument or recording quality and so forth — as in itself stylish. If so, this aspect of Stylishness has little to do with compositional style (and therefore historical periods) and more to do with sound production quality and contemporary values, supporting Taruskin's view.

CONCLUSIONS

We have undertaken three tasks: (1) to determine whether listeners are able to detect differences in performing styles and expressiveness; and (2) to further interrogate the nature of style-specific expressiveness, by (3) identifying salient links between perceived performance features and aesthetic qualities.

The comparison of responses to the four stimuli used in the study provides evidence for listeners' ability to distinguish features discussed in relation to baroque period performance practice. The differences among Luca's HIP interpretation (Articulated, Light, Detailed, Grouped, Flexible with little Vibrato), Menuhin's expressive-emotional (Continuous, Legato, Vibrato) rendering, and Grumiaux's modernist-literalist (Vibrato, Straightness, Strict, Measured) reading were perceived and Luca's HIP recording was rated as significantly more Stylish than Grumiaux's or Menuhin's and more Baroque Expressive than Menuhin's.

Together, Baroque Expressiveness and Stylishness provide subtly different perspectives on aesthetic responses. This is consistent with the previous study (Schubert & Fabian, 2006), which identified an Expressiveness and a Stylishness dimension as the two factors underlying a baroque listening experience. The present study showed that Success is related more to Stylishness than to Baroque Expressiveness. We therefore speculate that the stylishness of a performance embraces different types of expressivenesses.

Perceived musical features can be used to differentiate between aspects of aesthetic responses (e.g. Straightness of tone predicted Stylishness while Lightness of tone predicted Baroque Expressiveness) but some performance features can contribute to several aspects of aesthetic response within a given musical style (e.g. rhythmic flexibility predicted all three tested aesthetic variables). This suggests that expressive performance has some essential characteristics as well as a certain flexibility and multiplicity of interpretative solutions.

Further investigations should examine responses from a broader range of participants (we used mainly undergraduate music students (cf. Tables 1 and2)) and different approaches to the systematic treatment of the sound recordings, such as adding noise and

compression (Timmers, 2007). Nevertheless, our data suggest that baroque expressiveness is a useful concept for aesthetic measurement of baroque performance. It assists both the understanding that expression is style specific, as well as the identification of features thought to be appropriate to baroque music performance by the modern, early twenty-first-century Western listener. While participants struggled to define the term with consistency, they nevertheless seemed to have an implicit understanding of the term, evident from their responses to a selection of performances that ranged widely in expressive content. It will be worth investigating further how people who have only crude factual and analytical knowledge can nevertheless respond quite precisely to stylistic differences manifest in musical performance.

REFERENCES

Ballou, G. (2005). *Handbook for sound engineers*. Oxford: Focal Press.

Bartel, D. (1997). *Musica poetica*. Lincoln: University of Nebraska Press.

Bonds, M. (1991). *Wordless rhetoric: Musical form and the metaphor of the oration*. Cambridge, Mass.: Harvard University Press.

Butt, J. (1990). *Bach interpretation: Articulation marks in primary sources*. Cambridge: Cambridge University Press.

Butt, J. (1991). *Bach: B minor mass*. Cambridge: Cambridge University Press.

Butt, J. (1999). Bach recordings since 1980: A mirror of historical performance. In D. Schulenberg (Ed.), *Bach perspectives 4* (pp. 181-198). Lincoln and London: University of Nebraska Press.

Butt, J. (2002). *Playing with history: The historical approach to musical performance*. Cambridge: CUP.

Cone, E. T. (1968). *Musical form and musical performance.* New York: Norton.

Day, T. (2000). *A century of recorded music: Listening to musical history*. New Have: Yale University Press.

Dolmetsch, A. (1949). *Interpretation of seventeenth- and eighteenth-century music* (Rev.ed.). London: Novello.

Donington, R. (1989). *The interpretation of early music* (Rev.ed.). London: Faber and Faber.

Dreyfus, L. (1983). Early music defended against its devotees: A theory of historical performance in the twentieth century. *The Musical Quarterly*, *69*(3), 297-322.

Dulak, M. (1993). The quiet metamorphosis of "early music". *Repercussions, 2*(2), 31–61.

Elste, M. (2000). *Meilensteine der Bach-Interpretation 1750-2000*. [Milestones of Bach interpretation 1750-2000]. Stuttgart-Weimar and Kassel: Metzler-Bärenreiter.

Fabian, D. (2001). The meaning of authenticity and the early music movement – A historical review. *International Review of the Aesthetics and Sociology of Music*, *32*(2), 153-167.

Fabian, D. (2003). *Bach performance practice 1945-1975: A review of sound recordings and literature*. Aldershot: Ashgate.

Fabian, D. (2005). Towards a performance history of Bach's sonatas and partitas for solo violin: Preliminary investigations. In L. Vikárius & V. Lampert (Eds.), *Essays in honor of László Somfai: Studies in the sources and the interpretation of music* (pp. 87-108). Lanham, Maryland: Scarecrow Press.

Fabian, D., & Schubert, E. (2002). Is there only one way of being expressive in musical performance? –– Lessons from listeners' reactions to performances of J. S. Bach's music. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition* [CD-ROM] (pp. 112-115). Adelaide: Causal Productions.

Frotscher, G. (1980). *Performance practices of early music –– A comprehensive reference work about music of past ages for musicians, interpreters and amateurs*. (K. Michaelis, Trans.). New York: Heinrichshofen. (Original work published 1963).

Gramophone, The. (1948). *Boyd Neel string orchestra (Boyd Neel). Brandenburg Concerto No. 3 in G Major, 25*, 118. [Review by W.R.A.]

Gramophone, The. (1962). *Bach sonatas and partitas for solo violin. Arthur Grumiaux (violin) Philips, 39*, 458. [Review by J.N.]

Gramophone, The. (1972). *Bach. Brandenburg Concertos Nos. 1-6, BWV 1046-51. Virtuosi of England conducted by Arthur Davison, 50*, 679. [Review by S.S.]

Gramophone, The. (1976). *Bach: Brandenburg Concertos Nos. 1-6, BWV 1046-51. English Chamber Orchestra directed from the harpsichord by Raymond Leppard, 53*, 1587-8. [Review by M.E.O.]

Harnoncourt, N. (1988). *Baroque music today: Music as speech.* (M. O'Neill, Trans.). Portland, Oregon: Amadeus Press. (Original work published 1982).

Haskell, H. (1988). *The early music revival – A history*. London: Thames and Hudson.

Haynes, B. (2007). *The end of early music: A period performer's history of music for the twenty-first century*. New York: Oxford University Press.

Houle, G. (1987). *Metre in music 1600-1800: Performance, perception and notation*. Bloomington: Indiana University Press.

Kenyon, N. (Ed.). (1988). *Authenticity and early music*. Oxford: Oxford University Press.

Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs: Analysis-of-variance and multiple regression/correlation approaches.* New York: W.H. Freeman.

Kivy, P. (1995). *Authenticities*. Ithaca: Cornell University Press.

Landowska, W. (1964). Landowska on music. (D. Restout & R. Hawkins (Eds.), Trans.). New York: Stein and Day. (Original work published 1904)

Lawson, C., & Stowell, R. (1999). *The historical performance of music: An introduction*. Cambridge: Cambridge University Press.

Leech-Wilkinson, D. (1984). What we are doing with early music is genuinely authentic to such small degree that the word loses most of its intended meaning. The limits of authenticity: A discussion. *Early Music*, *12*(1), 13-16.

Leech-Wilkinson, D. (2009). *The changing sound of music: Approaches to studying recorded musical performances*. London: Centre for the History and Analysis of Recorded Music. Retrieved June 28, 2009 from www.charm.kcl.ac.uk/studies/chapters/intro.html

Neumann, F. (1978). *Ornamentation in baroque and post-baroque music*. Princeton: Princeton University Press.

Philip, R. (2004). *Performing music in the age of recording*. Cambridge: Cambridge University Press.

Schubert, E., & Fabian, D. (2006). The dimensions of baroque music performance: A semantic differential study. *Psychology of Music*, *34*(4), 573-87.

Taruskin, R. (1982). On letting the music speak for itself: Some reflection on musicology and performance. *Journal of Musicology*, *1*(3), 101-17.

Taruskin, R. (1984). The authenticity movement can become a positivistic purgatory, literalistic and dehumanizing. The limits of authenticity: A discussion. *Early Music, 12*(1), 3-12.

Taruskin, R. (1988). The pastness of the present and the presence of the past. In N. Kenyon (Ed.), *Authenticity and early music* (pp. 137-210). Oxford: Oxford University Press.

Taruskin, R. (1995). *Text and act – essays on music and performance*. Oxford and New York: Oxford University Press.

Timmers, R. (2007). Perception of music performance on historical and modern commercial recordings. *Journal Acoustic Society of America*, *122*(5), 2872-80.

Walls, P. (2003). *History, imagination and the performance of music*. Woodbridge: Boydell: 2003.

DISCOGRAPHY

Bach, J.S., 'Sarabanda, – D minor Partita for Solo Violin BWV 1004', Arthur Grumiaux, Philips (1961), 438 736-2.

Bach, J.S., 'Sarabanda, – D minor Partita for Solo Violin BWV 1004', Sergiù Luca, Electrola Nonesuch (1977), 7559 73030-2.

Bach, J.S., 'Sarabanda, – D minor Partita for Solo Violin BWV 1004', Yehudi Menuhin, EMI (1934), Reference CHS 7 63035 2.

DOROTTYA FABIAN is an Associate Professor and lectures in musicology at The University of New South Wales. Her research focuses on changing performance styles evidenced by sound recordings. She aims to combine historical-analytical investigations with experimental examinations of listeners' perception. Currently she is working on two large projects funded by the Australian Research Council; one investigates expressiveness in baroque and romantic music performance, the other aims to identify individual artistic signatures of prominent violinists of the past century in interpretations of violin sonatas by Bach, Beethoven and Brahms. [Email: d.fabian@unsw.edu.au]

EMERY SCHUBERT is an Associate Professor at the University of New South Wales in Sydney, Australia. He works in the music and music education discipline and is a Leader of the Empirical Musicology Group. His research interests are concerned with the "scientific" study of music aesthetics, including measuring and predicting emotional responses to music continuously. Emery is currently the President of the Australian Music and Psychology Society (AMPS). [Email: e.schubert@unsw.edu.au]