Issues arising from a forensic analysis of a musical recording to investigate its origins

David Howard(1) and Allen Hirson(2)
(1) Signal Processing: Voice and Hearing Research Group,
Department of Electronics, York University, York YO1 5DD, UK.
(2) Speech Acoustics Laboratory,
Department of Clinical Communication Studies, City University,
Northampton Square, London EC1V 0HB, UK.

1. Abstract

A recent forensic investigation into alleged fraud involved a dispute as to the physical origin of a piece of music: whether one recording was based on a direct copy of another or whether it was a studio re-recording of the score. The development of an appropriate technique for analysis and the results obtained are described. A particular issue arises from the paucity of data available for analysis, which in this case was only a few seconds. The data was quite inadequate for any standard test for statistical significance. We examine the constraints imposed by such limited data sets for those working on forensic investigations, and the relationship between the established body of scientific knowledge in the literature on the one hand, and new methodology devised to solve particular questions of interest on the other. Important parallels in the forensic analysis of speech recordings are also discussed.

2. Background Case

A recent civil litigation case in the UK involved a dispute as to the origin of a musical recording. On the one hand there was the claim that a disputed recording (R1) was a direct copy of a previous recording (R2), whilst on the other, it was maintained that R1 consisted of music that had been played by a session musician in a studio and recorded "afresh". The question to be answered was whether R1 was a direct copy of R2 or not. A copy of each recording was made available for analysis.

2.1 The available data

Neither of the master recordings were made available for analysis: R1 was provided in vinyl format and R2 as a ferric cassette which was an unknown number of generations removed from its master tape. The studio histories of the recorded materials were not available, and requests for either direct access to or digital copies of the master tapes were denied. It was therefore not known what post-processing, if any, and copying procedures had been used in either instance.

On listening to the two recordings, it was immediately clear that they were non-identical since: (a) a drum track had been included in recording R1 which was not on R2, and (b) R1 and R2 replayed at slightly different pitches. The length of the disputed section was very short, being only 2.5s, and consisting of a fifteen note melodic line played on an acoustic guitar .

2.2 Experimental method

At the outset, direct quantitative methods in the time and frequency domains were considered: (a) a comparison of the spectral characteristics of equivalent notes in the two recordings, and (b) a comparison of the temporal relationships between the individual notes in each recording. A frequency domain analysis was discounted due to the presence of the drum track in R1, which was at both a considerably higher level and covered a wider frequency range than the guitar notes. In addition, the lack of knowledge regarding the use of EQ or other post-processing would have made it impossible to allow for any spectral variations imposed during any re-mix or tape copying. It was therefore decided to make use of a time domain analysis. In particular, comparison of the relative durations of individual notes using the ratio of each note duration to that of the complete snippet which would compensate for any overall duration differences.

The experimental methodology involved the use of a Kay Elemetrics Sonagraph Model (No. 5500) to locate and measure the onset time of each note by means of a calibrated wide-band spectrogram and time-aligned waveform plot (e.g. see [1] for a review of spectrography). The sampling rate used was 81kHz, and FFT-based wide-band spectrograms were used with an equivalent filter bandwidth of 300Hz. The Kay Sonagraph has a particular advantage in that it enables specified sections of captured sounds to be replayed repeatedly for detailed study, as well as permitting objective measurements of frequency and time in selected portions of the signal. Recording R1 contained notes that were accompanied by percussion events (drum and cymbal beats) and it was not possible to make an accurate measurement of note onset times, since these events were recorded at a considerably higher level than the guitar melody line. Therefore for the purposes of this analysis, only the onset times of those notes which had no accompanying drum beat were quantified.

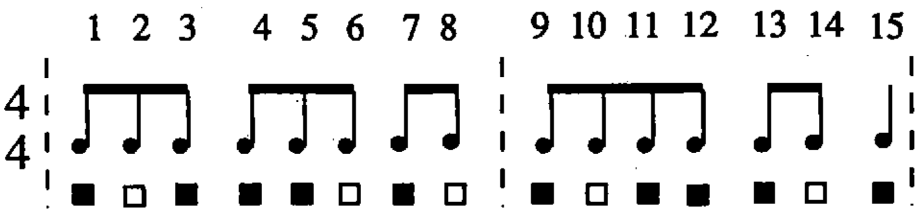

Figure 1: *The rhythmic structure of the musical phrase in question. The filled squares denote those notes which have drum beats associated with them, and the notes without drum beats, whose onset times were used for the analysis, are indicated by the open squares.*

The rhythmic nature of the melodic snippet is shown in figure 1, along with an indication of which notes had percussive events associated with them. In practice, the presence/absence of percussive events was readily identified spectrographically (e.g. see [2]). The measurement of note onset time or the spectral characteristics of the guitar note itself was not possible for those notes with percussive beats associated with them, thus the analysis was restricted to the five notes not associated with percussive events.

The first non-percussive note (shown as note 2 in figure 1) was used as the reference point in time from which the onset times of other non-percussive notes were measured. The measured times between each note onset and this reference point were expressed as a percentage of the total time between the onsets of notes 2 and 14. The difference in overall durations between R1 and R2 was calculated as being consistent with the difference in pitch between the recordings.

2.3 Results

The results of the normalised note timing analyses for the two musical snippets were directly compared by calculating their differences and expressing these as percentages (see table 1). The non-percussive notes serve a musically unstressed function in this snippet as indicated by the positions of the barlines, shown as vertical dotted lines in figure 1. Therefore they are notes which one would expect temporally to drift furthest from the main rhythmic anchor points of the music during repeated performances, and any timing differences between the performances should be most clearly shown in their note onset times.

The mean difference between relative note onsets in R1 and R2 was 1% (see table 1). In order to put these results in an appropriate context, musicological studies relating to timing precision during musical performances were investigated. A study by Shaffer et al. [3] found a variation between a musician's repeated performances of the same piece of piano music by Satie of 1.8% of the mean beat duration. We were unable to use a beat duration metric directly due to the limited data set, but after time normalisation, our data compares closely with the Shaffer et al. data. It suggests that R1 is *not* the result of another (or even the same) musician attempting to make a performance copy of R2, which we would expect to result in a % error measure significantly greater value than the 1.8% found by Shaffer et al.

It should also be noted that our figure includes the experimental error in positioning the timing cursors available on the spectrograph, ±0.003s for the time scale used, which represents a mean error of 0.85%. Thus our measured variation could, in fact, be entirely due to experimental error. While the results therefore strongly suggest that the two extracts have a common origin, one should not loose sight of the extreme paucity of the data set available in this case.

3. Forensic acoustics

The forensic comparison of two sound recordings in order to establish their cognate relations is influenced by: (a) the choice of method, (b) research which informs the interpretation of results; and (c) definitions of similarity and difference, defined scientifically and interpreted legally. These issues are discussed below.

In order to establish identity between recordings both the format and histories of the recorded materials need to be identical or the signal transformations undergone must be understood. In the current case the formats were rather different, and overall duration differences as well as the added drum beat suggested either post-processing or that the recordings had distinct histories. The onus rested with the authors to devise a method which overcame differences in phase relations, frequency spectra, bandwidth etc. Since details of post-processing and copying were unknown, methods in the frequency or phase domain were discounted, and a method was devised based on the more robust variable of temporal relations between individual notes.

The case presented here is a good example of how a methodology can be developed in response to the problems of a specific case. This search for, and development of new methods where appropriate is the stuff of science, but the restrictions imposed by the quantity of available data impose severe limits on the degree of certainty which might be attached to any conclusions reached [4].

The evidence presented to the forensic scientist often comes in small quantities e.g. hairs, fibres, flecks of paint etc. and the music fraud case is not unusual in this respect. Small quantities of material may be sufficient to derive statistically valid data if it is possible to make enough accurate measurements on the relevant scale, and if such measurements may be interpreted in terms of some relevant body of knowledge.

Close consideration of the method shows two potential weaknesses. Timing measurements were possible with an acceptable measurement error, but there were less data points than would normally be required to draw positive conclusions of identity with any statistical certainty.

Furthermore, the music research available at the time of the investigation related to classical rather than temporally simpler popular music. There is no a priori reason to suggest different levels of accuracy for different styles of music, but neither can there be complacency about over-generalising the Shaffer et al. data [3]. Further research is required.

This paper would be incomplete if it did not consider parallels in the field of forensic speech analysis. The vast majority of such analyses involve what is described as speaker identification in which a disputed speech recording is compared with a recording of the speech sample of one or more suspects, e.g. [5, 6, 7, 8]. The type of expertise required for such analyses differs from that involved for music mainly because speech sounds require auditory phonetic interpretation as well as rather complex acoustic evaluation. In addition, it would be unusual for one side in a case to claim that two tapes derive from a single recording, whereas it is common for there to be a claim that two recordings originate from a single speaker.

Rather in the way that a musical performance may vary from one performance to the next, so speech is characterised by great plasticity. Speech variability is only poorly understood, and as a consequence techniques are not currently available to make judgements with the degree of certainty which is possible in other forensic sciences. For example, fingerprints and genetic makeup do not alter in the short term and therefore have powerful forensic applications [9], whereas aspects of voice alter both in the short term, and over longer periods of time. Situational variables are also powerful forces influencing speech production. The difficulty in establishing, for example, the fundamental frequency range for a single speaker may also be compounded when vocal imitation or voice disguise is involved (e.g. [10]). In short, similarities and differences between recorded speech samples are evaluated for their forensic significance in terms of a very incomplete understanding of inter- and intra-speaker variation. Comparing musical recording in terms of intellectual property may involve similar fuzzy sets, but for the current case where direct physical copying was suspected, the comparison is more direct.

Establishing whether tapes have been copied, over-recorded or tampered with ("tape authentication") has been used for some years in the forensic analysis of tape recordings (see e.g. [11], [12]). This includes microscopic examination of the magnetic patterns on the original recording, but since original recordings were not available to us this approach was impossible. On the other hand, acoustic transients may provide tell-tale signs of such interference. In this area, music technologists may lead the field with some of the ingenious signal processing methods they use, for example, for dealing with clicks and scratches on old vinyl recordings [13]. However, such techniques have not yet been applied in forensic speech analysis.

4. Conclusions

The case work described demonstrates the requirement to work with restricted data sets, since the investigation can rarely be planned in the manner of a research experiement. The nature of the data itself can require the investigators to develop their own methodology based on experience and common sense, since there may be little supporting literature to turn to. This is a feature common to acoustic investigations as well as other branches of forensic science, and the conclusions presented in court by the expert witnesses concerned should reflect this.

Forensic scientists recognise the need for supporting research work in their field, and some are starting to carry out appropriate experiments. We have planned a computer-based experimental arrangement which will test how closely musicians can repeat/copy existing pieces of music in a variety of styles. It is hoped that this will enable data to be gathered which might in the future support conclusions such as those presented above ..This research is motivated in part from an active interest in the field, and also in response to proper critical comment from colleagues on our paper [2]. We believe that this tension between scientific doubt about earlier findings on the one hand, and more definitive legal judgements on the other is both constructive and healthy for a sound forensic science.

R1 Note No.	Rel time T1	ratio (tN-t2)/(t14-t2) r1	R2 Note No.	Rel Time T2	ratio (tN-t2)/(t14-t2) r2	Differences: Diffs (r1-r2) D	%Diff D/r1*100 P
1	0		1*	0	0	0	
2*	0.2031	0	2*	0.1906			
3	0.3687		3*	0.3562			
4	0.5719		4*	0.5187			
5	0.7437		5*	0.7000			
6*	0.9531	0.338	6*	0.8719	0.331	0.007	2.16
7	1.116		7*	1.034			
8*	1.316	0.502	8*	1.225	0.503	0.001	0.18
9	1.494		9*	1.387			
10*	1.688	0.670	10*	1.559	0.665	0.005	0.70
11	1.869		11*	1.740			
12	2.006		12*	1.897			
13	2.228		13*	2.065			
14*	2.419	1	14*	2.247	1	0	
15	2.597		15*	2.415			

Key: * = no accompanying percussive beat

Experimental error in spectrograph cursor positioning: 3ms or 0.84%
Mean of % differences: 1%

Table 1: *Relative note timing measures from the two snippets R1 and R2. An asterisk indicates notes without added percussive beats. Time measurements are in seconds.*

5. References:

[1] SR ROSEN and P HOWELL (1991) 'Signals and systems for speech and hearing', London: Academic Press.

[2] DM HOWARD, A HIRSON and GA LINDSEY (1992) 'Acoustic techniques to trace the origins of a musical recording', *Journal of the Forensic Science Society*, in press.

[3] LH SHAFFER, EF CLARKE and NP TODD (1985) 'Metre and rhythm in piano playing', *Cognition*, 20, 61-77.

[4] BH REPP (1992) 'Objective Performance analysis as a tool for the musical detective', *Journal of the Acoustical Society of America (Research briefs/Technical Notes*, in press.

[5] F NOLAN (1984) 'The Phonetics of Speaker Identification', Cambridge: Cambridge University Press.

[6] F NOLAN (1991) 'Forensic phonetics: Notes and Discussion', *Journal of Linguistics*, 27, 483-493.

[7] J BALDWIN and JP FRENCH (1990) 'Forensic Phonetics'. London: Pinter Publishers.

[8] H HOLLIEN (1991) 'The acoustics of crime. The new science of forensic phonetics'. New York: Plenum.

[9] P LOWRIE and S WELLS (1991) 'Genetic fingerprinting', *New Scientist*, 16, 1-4.

[10] A HIRSON and M DUCKWORTH (1992) 'Glottal fry and voice disguise: A case study in: forensic phonetics', *Journal of the Biological Engineering Society*, in press.

[11] BE KOENIG (1990) 'Authentication of Forensic Audio Recordings', *Journal of the Audio Engineering Society*, 38 (12), 3-33.

[12] D DEAN (1989) 'The use of Ferrofluids in the forensic examination of magnetic recordings', *Scientific Research & Development Branch. Home Office, British Government*. London: Publication 15/89.

[13] SV VASEGHI and JW RAYNER (1988) 'A new application of adaptive filters for restoration of archived gramophone recordings', *IEEE - ICASSP*, 5, A1.15, 2548-2551.