Appendix B -- Comparison of Multi-F0 estimation procedures

Appendix B shows the typical example of what happen when processing a polyphonic signal when the two compared methods are used; e.g., direct cancellation on 1-D spectrum, and iterative extraction on bispectrum. For this purpose, in this Appendix, an example of Multi-F0 Estimation procedure step-by-step, carried out by the transcription system presented in the paper. The results are compared with those obtained by a transcription method performing an iterative 1-D pattern matching in the spectrum domain, and subsequent direct cancellation of the harmonic pattern of estimated notes.

The audio input source is a real signal taken from the RWC Database, analyzed in a single frame for the purpose of the example. In the processed frame, notes G_2, D_4 and B_4 are playing, corresponding to MIDI notes 43, 62 and 71, respectively. These notes present a significant partials overlapping. Actually, denoting the fundamental frequencies as f_{01}, f_{02} and f_{03}, respectively, they stay in the following ratios each other:

$$f_{02} = 3f_{01},\ f_{03} = 5f_{01},\ f_{03} = \frac{5}{3}f_{02}.$$

which are approximated, in the frequency-log scale adopted in our system (following the well-tempered scale) with distances of 19, 28 and 9 semitones.

In Figure B.1 the amplitude spectrum and bispectrum before the F0 estimation process are presented.

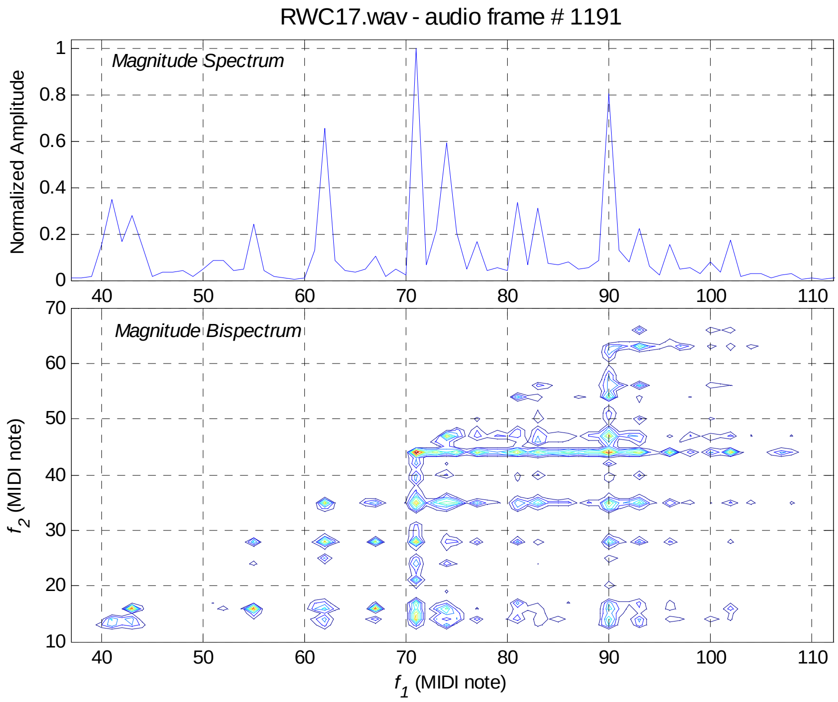

Figure B.1. Amplitude spectrum and bispectrum of audio signal before Multi-F0 estimation.

In Figures (Figures B.2 and B.3) a direct comparison between both the Multi-F0 estimation procedures is depicted, by plotting the normalized 1-D and 2-D cross-correlations for each step.

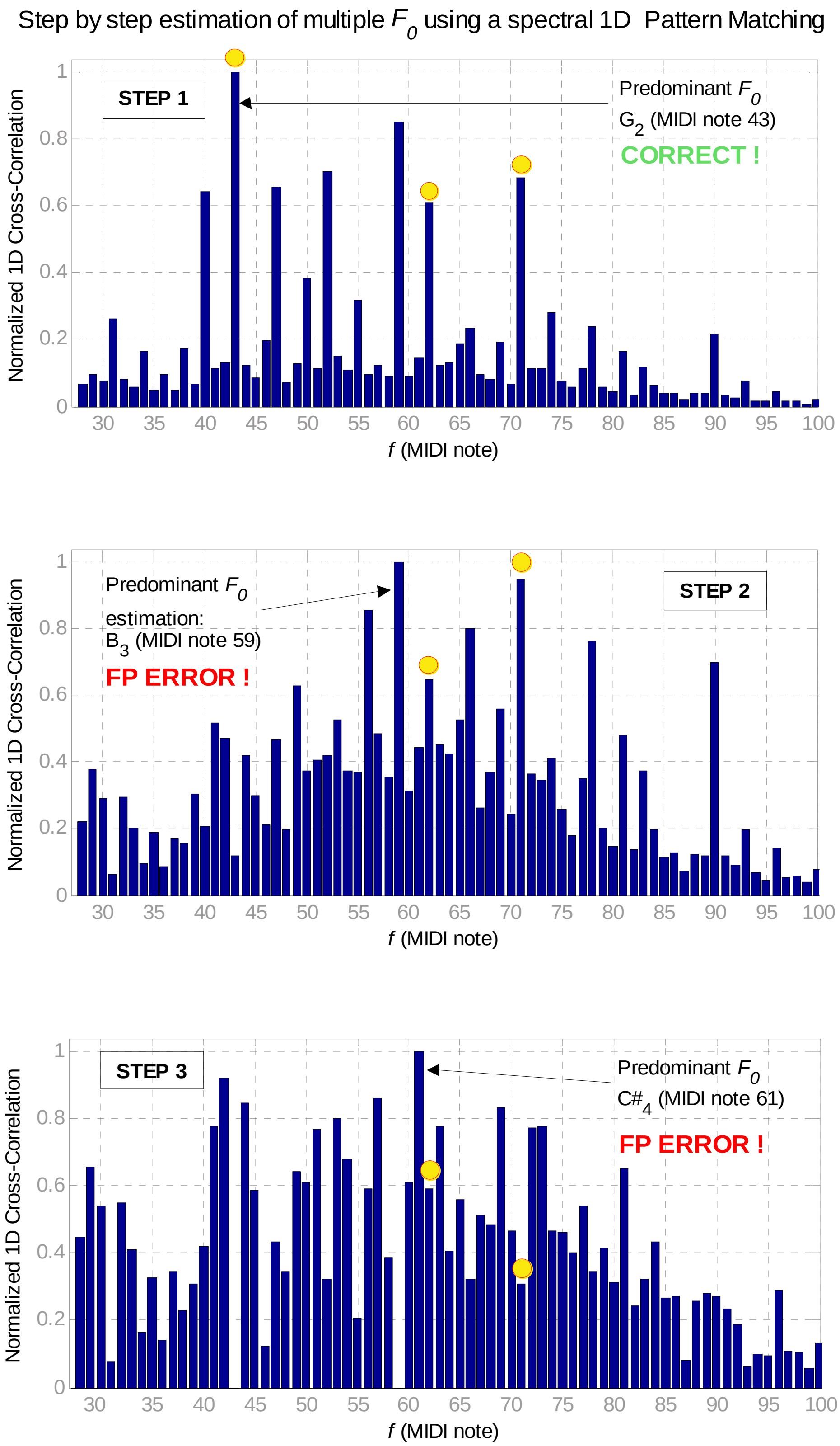

Figure B.2. Step by step Multi-F0 estimation procedure with iterative spectral 1-D pattern matching and *direct cancellation technique*. The dots identify the notes played in the audio source signal.

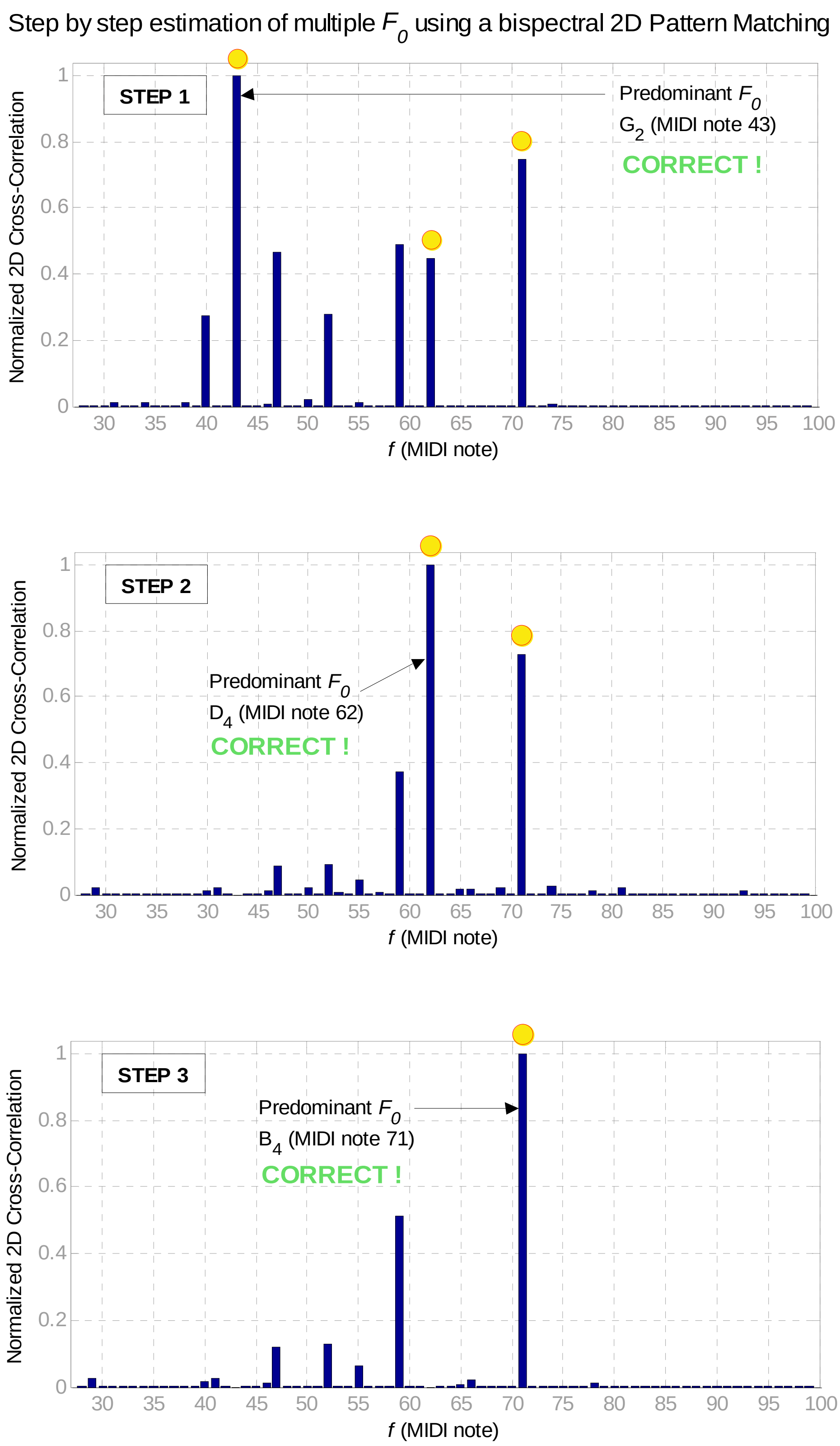

Figure B.3. Step by step Multi-F0 estimation procedure with iterative *bispectral 2-D pattern matching and pattern extraction technique*. The dots identify the notes played in the audio source signal.

It should be noted that, the 2-D bispectral correlation is much clearer than the 1-D spectral correlation. As stated in the paper (see section III.C), denoting the normalized 2-D cross-correlation as $\rho(f_1, f_2)$, if a monophonic sound has a fundamental frequency corresponding to index q in the discrete log-frequency array, then the maximum of $\rho(f_1, f_2)$ is expected to be found at (q, q). For this reason, the cross-correlation $\rho(f_1, f_2)$ is computed only for $f_1 = f_2 = q$, that is only upon the points belonging to the first quadrant bisector.

Moreover, comparing Figures B.2 and B.3, it can be observed that after *Step 1* (in which the lowest note G_2 is correctly identified by both the algorithms). On the other hand, the direct cancellation of spectral G_2 pattern (in the 1-D F0 estimation method) deletes also some coinciding partials of the two higher sounds, including the fundamental frequencies of both D_4 and B_4, as shown in Figure B.4.

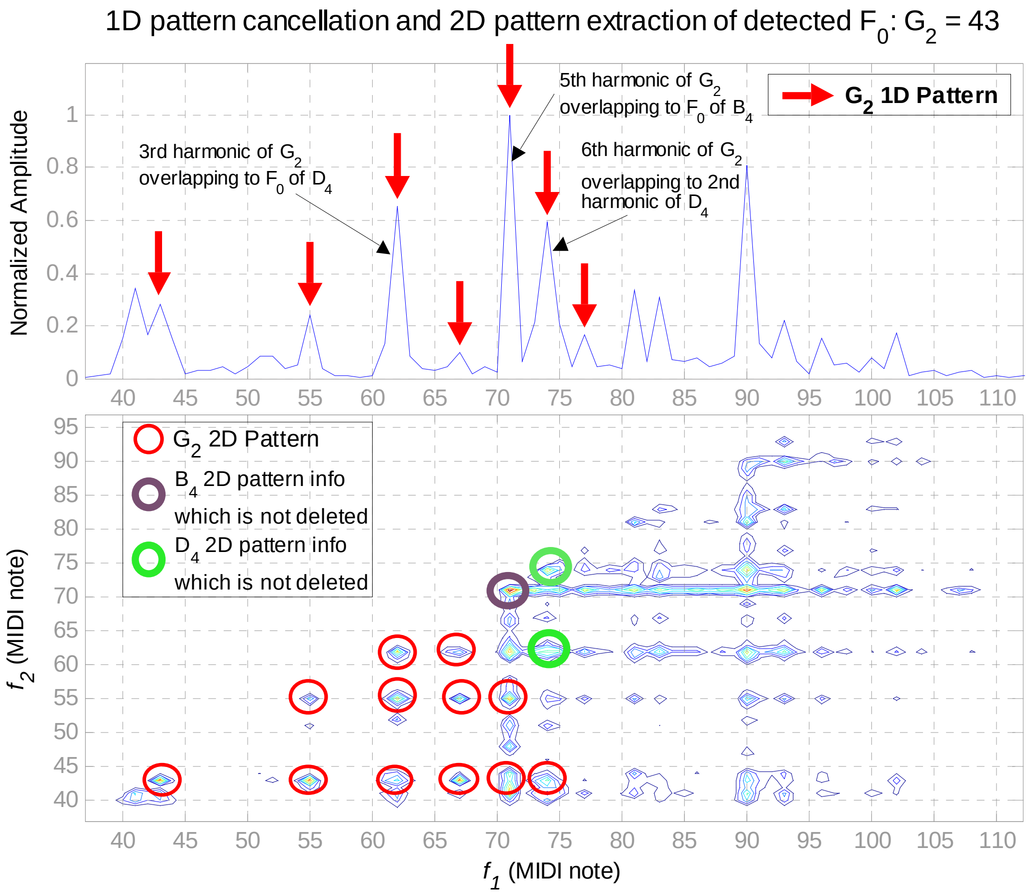

Figure B.4. Graphical comparison between direct cancellation of 1-D pattern from the spectrum (above) and extraction of 2-D pattern from the bispectrum (below).

In general, the bispectral representation cannot help to resolve the underlying components of interfering partials; while it is the mechanism of extraction of the 2-D monophonic pattern of G_2 in the proposed bispectrum-based algorithm which allows keeping critical information about the peak positions of the other sounds harmonic 2-D patterns, which are:

$$(f_{02}, 2f_{02}),\ (2f_{02}, 2f_{02}) \text{ and } (f_{03}, f_{03}).$$

In conclusion, this Appendix B has shown the typical example of what happen when processing a polyphonic signal when the two compared methods are used; e.g., direct cancellation on 1-D spectrum, and iterative extraction on bispectrum.

Thus, the system performing the iterative 2-D pattern matching and pattern extraction in the bispectrum domain successfully identifies all the three notes played in the audio source file. The system performing the iterative 1-D pattern matching and direct cancellation of the pattern in the spectrum domain identifies only the lowest note, G_2, and commits two false positive errors, due to the removal of partials of the higher sounds in the direct cancellation procedure.