Appendix A – Analysis of bispectrum for polyphonic sounds

In this Appendix, the bispectrum of polyphonic sound is theoretically treated, together with some examples. In particular, the cases regarding polyphonic signals with two or more sounds have been considered. In the case of bichords, one of the most interesting cases, being a perfect fifth interval, since it presents a strong partials overlap ratio. In this case, the analysis of residual coming from the difference of the real bispectrum of the bichord signal with respect to the linear composition of the single bispectra of concurrent sounds, has been performed. The formal analysis has demonstrated that the contributions of this residual are null or negligible for proposed multi-F0 estimation procedure. This theoretical analysis has been also confirmed by the experimental results, as shown with some examples. Moreover, the case of tri-chord with strong partial overlapping and a high number of harmonics per sound has confirmed the same results.

A.1 Bispectrum of a polyphonic signal: a bichord

In this section, the behaviour of the bispectrum for a polyphonic signal is analyzed. Let us to recall the spectrum (positive frequencies only) of a generic monophonic sound with fundamental frequency f_0:

$$X(f) = \sum_{\substack{k=1,\dots,P \\ kf_0 \in H}} \delta(f - kf_0),$$

where H is the set of harmonics of the sound, consisting of P partials (fundamental frequency included): $H = \{f_0, 2f_0, 3f_0, ..., (P-1)f_0, Pf_0\}$.

Consider now, as an example and without loss of generality, two synthesized sounds, S_1 and S_2, each one composed by five partials, so that $H_1 = \{f_{01}, 2f_{01}, 3f_{01}, 4f_{01}, 5f_{01}\}$ and $H_2 = \{f_{02}, 2f_{02}, 3f_{02}, 4f_{02}, 5f_{02}\}$. The generated spectra are denoted as $X_1(f)$ and $X_2(f)$, respectively. Accordingly with the linearity of the Fourier Transform, let $X(f) = X_1(f) + X_2(f)$ be the spectrum of the polyphonic signal S, composed by the mixture of S_1 and S_2. Under these assumptions, the bispectrum of the polyphonic signal, computed with the direct method (defined by eq. (3), Section II.B in the paper) can be expressed as the follows:

$$\begin{aligned}
B_S(f_1, f_2) &= X(f_1)X(f_2)X^*(f_1+f_2) = \\
&= \big(X_1(f_1)+X_2(f_1)\big)\big(X_1(f_2)+X_2(f_2)\big)\big(X_1(f_1+f_2)+X_2(f_1+f_2)\big)^* = \\
&= X_1(f_1)X_1(f_2)X_1^*(f_1+f_2) \\
&\quad + X_1(f_1)X_2(f_2)X_1^*(f_1+f_2) \\
&\quad + X_2(f_1)X_1(f_2)X_1^*(f_1+f_2) \\
&\quad + X_2(f_1)X_2(f_2)X_1^*(f_1+f_2) \\
&\quad + X_1(f_1)X_1(f_2)X_2^*(f_1+f_2) \\
&\quad + X_1(f_1)X_2(f_2)X_2^*(f_1+f_2) \\
&\quad + X_2(f_1)X_1(f_2)X_2^*(f_1+f_2) \\
&\quad + X_2(f_1)X_2(f_2)X_2^*(f_1+f_2).
\end{aligned} \tag{A.1}$$

A.2 Analysis of bispectrum nonlinearity

The first and the last terms of the sum in equation (A.1) are equal to $B_{S_1}(f_1, f_2)$ and $B_{S_2}(f_1, f_2)$, respectively. The bispectrum is not linear, actually $B_S(f_1, f_2) \neq B_{S_1}(f_1, f_2) + B_{S_2}(f_1, f_2)$. Let $B_{diff}(f_1, f_2)$ be the difference $B_S(f_1, f_2) - B_{S_1}(f_1, f_2) - B_{S_2}(f_1, f_2)$:

$$
\begin{aligned}
B_{diff}(f_1, f_2) = {} & X_1(f_1)X_2(f_2)X_1^*(f_1+f_2) \\
& + X_2(f_1)X_1(f_2)X_1^*(f_1+f_2) \\
& + X_2(f_1)X_2(f_2)X_1^*(f_1+f_2) \\
& + X_1(f_1)X_1(f_2)X_2^*(f_1+f_2) \\
& + X_1(f_1)X_2(f_2)X_2^*(f_1+f_2) \\
& + X_2(f_1)X_1(f_2)X_2^*(f_1+f_2).
\end{aligned}
\qquad \text{(A.2)}
$$

Let us analyze each term of the sum in equation (A.2), in order to better understand the behaviour of $B_{diff}(f_1, f_2)$.

The **first term** yields:

$$X_1(f_1)X_2(f_2)X_1^*(f_1+f_2) = \sum_{\substack{k=1,\dots,5 \\ kf_{01} \in H_1}} \delta(f_1 - kf_{01}) \sum_{\substack{l=1,\dots,5 \\ lf_{02} \in H_2}} \delta(f_2 - lf_{02}) \sum_{\substack{m=1,\dots,5 \\ mf_{01} \in H_1}} \delta(f_1 + f_2 - mf_{01}) = \prod_{\{1,2,1\}};$$

the product $\prod_{\{1,2,1\}}$ is not null only if each term of the product itself is not null. Concerning the first two terms, this happens when $f_1 = kf_{01}$ (that is, when f_1 takes the value of any of the partials belonging to H_1) and, similarly, when $f_2 = lf_{02}$. This involves that, considering the third term, the entire product is non-zero only when it exists at least an integer value m such that: $mf_{01} = kf_{01} + lf_{02}$, $k = 1,\dots,5$, $l = 1,\dots,5$ and $mf_{01} \in H_1$. To satisfy this condition, it is necessary (but not sufficient, depending on the length of H_1 and H_2) that the sounds present overlapping partials; a sufficient condition is that the two harmonic series, H_1 and H_2, share at least one frequency value.

As an example: consider two sounds, with harmonic sets H_1 and H_2, generate a perfect fifth interval (which presents a very strong partials overlap ratio); this implies $2f_{02} = 3f_{01}$. Under these conditions, the contribute of $\prod_{\{1,2,1\}}$ would be non-zero only for the following couples (f_1, f_2):

$$(f_{01}, 2f_{02}) \text{ and } (2f_{01}, 2f_{02}),$$

with $f_{01} + 2f_{02} = f_{01} + 3f_{01} = 4f_{01} \in H_1$, and also $2f_{01} + 2f_{02} = 2f_{01} + 3f_{01} = 5f_{01} \in H_1$. It is worthy to notice that these two couples are located in the upper triangular region of the plane (f_1, f_2), above the first quadrant bisector, and so they are outside the non-redundant region considered in the computation of the bispectrum (see Section II.B and Figure 2 of the paper). For this reason, the contribute of $\Pi_{\{1,2,1\}}$ to $B_{diff}(f_1,f_2)$, in this context, is zero. This analysis can be generalized for all the terms of the sum in $B_{diff}(f_1,f_2)$, as reported in the following.

Considering the **second term** of the sum in equation (A.2):

$$X_2(f_1)X_1(f_2)X_1^*(f_1+f_2)=\sum_{\substack{k=1,\dots,5\\ kf_{02}\in H_2}}\delta(f_1-kf_{02})\sum_{\substack{l=1,\dots,5\\ lf_{01}\in H_1}}\delta(f_2-lf_{01})\sum_{\substack{m=1,\dots,5\\ mf_{01}\in H_1}}\delta(f_1+f_2-mf_{01})=\Pi_{\{2,1,1\}};$$

the term $\Pi_{\{2,1,1\}}$ is non-zero only if exist at least an integer values m such that $mf_{01}=kf_{02}+lf_{01}$, $k=1,\dots,5$, $l=1,\dots,5$ and $mf_{01}\in H_1$. Following the example of the two sounds generating a perfect fifth interval, this happens only for the couples of frequencies

$$(2f_{02},f_{01}) \text{ and } (2f_{02},2f_{01}).$$

As it can be noticed, this is the symmetric case of $\Pi_{\{1,2,1\}}$, with respect to the first quadrant bisector, and in this circumstance these points are inside the non-redundant region considered for bispectrum computation. Therefore, $\Pi_{\{2,2,1\}}$ is not null in these points; however, $B_1(f_1,f_2)$ also generates nonnull values in correspondence of these two couples, in the equivalent form of $(3f_{01},f_{01})$ and $(3f_{01},2f_{01})$ (see equation (5), Section III.C of the paper). For this reason, $\Pi_{\{2,1,1\}}$ does not generate any additional peaks in the (f_1,f_2) plane; the only effect is to add an amplitude contribute to bispectral peaks generated by $B_1(f_1,f_2)$, at the same positions in the (f_1,f_2) plane. At the end of these considerations we will show that these contributes can be considered not relevant in the computation of normalized 2-D cross-correlation, within the Multi-F0 estimation procedure.

Consider now the **third term** in equation (A.2):

$$X_2(f_1)X_2(f_2)X_1^*(f_1+f_2)=\sum_{\substack{k=1,\dots,5\\ kf_{02}\in H_2}}\delta(f_1-kf_{02})\sum_{\substack{l=1,\dots,5\\ lf_{02}\in H_2}}\delta(f_2-lf_{02})\sum_{\substack{m=1,\dots,5\\ mf_{01}\in H_1}}\delta(f_1+f_2-mf_{01})=\Pi_{\{2,2,1\}};$$

$\Pi_{\{2,2,1\}}$ is non-zero only if exist at least an integer value m such that $mf_{01}=kf_{02}+lf_{02}$, $k=1,\dots,5$, $l=1,\dots,5$ and $mf_{01}\in H_1$. In our example, such a case occurs for the couple

$$(f_{02},f_{02}),$$

actually $f_{02}+f_{02}=\frac{3}{2}f_{01}+\frac{3}{2}f_{01}=3f_{01}\in H_1$. This shows that $\Pi_{\{2,2,1\}}$ only adds an amplitude contribute to a bispectral peak originated by $B_2(f_1,f_2)$ at the same position in the (f_1,f_2) plane, without generating any additional peaks.

Consider the **fourth term** in equation (A.2):

$$X_1(f_1)X_1(f_2)X_2^*(f_1+f_2) = \sum_{\substack{k=1,\dots,5 \\ kf_{01}\in H_1}} \delta(f_1 - kf_{01}) \sum_{\substack{l=1,\dots,5 \\ lf_{01}\in H_1}} \delta(f_2 - lf_{01}) \sum_{\substack{m=1,\dots,5 \\ mf_{02}\in H_2}} \delta(f_1 + f_2 - mf_{02}) = \Pi_{\{1,1,2\}};$$

$\Pi_{\{1,1,2\}}$ is non-zero only if exist at least an integer value *m* such that $mf_{02} = kf_{01} + lf_{01}$, $k = 1,\dots,5$, $l = 1,\dots,5$ and $mf_{02} \in H_2$. In our example, this happens for the following couples of frequencies:

- $(f_{01}, 2f_{01})$, actually $f_{01} + 2f_{01} = 3f_{01} = 2f_{02} \in H_2$;
- $(f_{01}, 5f_{01})$, actually $f_{01} + 5f_{01} = 6f_{01} = 4f_{02} \in H_2$;
- $(2f_{01}, 4f_{01})$, actually $2f_{01} + 4f_{01} = 6f_{01} = 4f_{02} \in H_2$.

These three couples are outside the non-redundant region considered for bispectrum computation; $\Pi_{\{1,1,2\}}$ is not null only in correspondence of the following couples, which are the symmetric ones of the three ones listed above (with respect to the first quadrant bisector):

- $(2f_{01}, f_{01})$: this adds an amplitude contribute to the bispectral peak generated by $B_1(f_1, f_2)$ at the same position in the (f_1, f_2) plane;
- $(5f_{01}, f_{01})$ and $(4f_{01}, 2f_{01})$; in correspondence of these two couples $\Pi_{\{1,1,2\}}$ gives origin (in this particular case) to two additional peaks in the bispectrum: they represent an extension to the five harmonics 2-D monophonic pattern of the sound at pitch f_{01}, (according to equation (5), Section III.C of the paper). The reason why $B_1(f_1, f_2)$ does not generate peaks in correspondence of these two couples is that the considered harmonic set H_1 is composed by five partials.

Consider the **fifth term** in equation (A.2):

$$X_1(f_1)X_2(f_2)X_2^*(f_1+f_2) = \sum_{\substack{k=1,\dots,5 \\ kf_{01}\in H_1}} \delta(f_1 - kf_{01}) \sum_{\substack{l=1,\dots,5 \\ lf_{02}\in H_2}} \delta(f_2 - lf_{02}) \sum_{\substack{m=1,\dots,5 \\ mf_{02}\in H_2}} \delta(f_1 + f_2 - mf_{02}) = \Pi_{\{1,2,2\}};$$

$\Pi_{\{1,2,2\}}$ is non-zero only if exist at least an integer value *m* such that $mf_{02} = kf_{01} + lf_{02}$, $k = 1,\dots,5$, $l = 1,\dots,5$ and $mf_{02} \in H_2$. In our example, this happens for the following couples of frequencies:

- $(3f_{01}, f_{02})$ and $(3f_{01}, 2f_{02})$, in correspondence of which add $\Pi_{\{1,2,2\}}$ adds an amplitude contribute to the bispectral peaks generated by $B_2(f_1, f_2)$ in $(2f_{02}, f_{02})$ and $(2f_{02}, 2f_{02})$;
- $(3f_{01}, 3f_{02})$, which is outside the non-redundant region considered in the computation of the bispectrum.

Consider, finally, the **sixth term** in equation (A.2):

$$X_2(f_1)X_1(f_2)X_2^*(f_1+f_2) = \sum_{\substack{k=1,\dots,5 \\ kf_{02}\in H_2}} \delta(f_1 - kf_{02}) \sum_{\substack{l=1,\dots,5 \\ lf_{01}\in H_1}} \delta(f_2 - lf_{01}) \sum_{\substack{m=1,\dots,5 \\ mf_{02}\in H_2}} \delta(f_1 + f_2 - mf_{02}) = \Pi_{\{2,1,2\}}.$$

As it can be noticed, this is the symmetric case of the previous $\Pi_{\{1,2,2\}}$, with respect to the first quadrant bisector. Therefore, $\Pi_{\{2,1,2\}}$ is non-zero only when exist at least an integer value *m* such that $mf_{02} = kf_{02} + lf_{01}$, $k = 1,\dots,5$, $l = 1,\dots,5$ and $mf_{02} \in H_2$. In our example, this happens for the following couples of frequencies:

- $(f_{02}, 3f_{01})$, which is outside the non-redundant region considered in the computation of the bispectrum;
- $(2f_{02}, 3f_{01})$ and $(3f_{02}, 3f_{01})$, in correspondence of which $\Pi_{\{2,1,2\}}$ adds an amplitude contribute to the bispectral peaks generated by $B_2(f_1, f_2)$ in $(2f_{02}, 2f_{02})$ and $(3f_{02}, 2f_{02})$.

Eventually, let us to remember that we have illustrated an example in which the two interfering sounds present a strong partials overlap ratio. For a generic synthesized bichord, the contribute of $B_{diff}(f_1, f_2)$ gains more relevance with the increasing number of partials in the harmonic sets of the sounds, and with the increasing partials overlap ratio. In the other cases, when the two sounds don't share the value of any of their partials within their harmonic sets, the value of $B_{diff}(f_1, f_2)$ is zero.

A.3 Empirical example: a synthesized bichord

A graphical example could be useful to illustrate in a clearer way this argumentation.

In Figure A.1, the contour plot of the bispectrum of a synthesized 5 harmonics bichord: $\mathrm{C}_4 - \mathrm{G}_4 \left(f_{0_{\mathrm{C}_4}} = 261.63 \text{ Hz}, f_{0_{\mathrm{G}_4}} = 392 \text{ Hz} \right)$, which forms a perfect fifth interval; then in Figure A.2, the contour plot of the sum of the bispectra of C_4 and G_4**,** is shown. In Figure A.1, the monophonic 2-D patterns of the two sounds are distinguishable, and also the two additional peaks generated by the contribute of the product $\Pi_{\{1,1,2\}}$, located at $(5f_{01}, f_{01})$ and $(4f_{01}, 2f_{01})$, which appear to have a smaller amplitude.

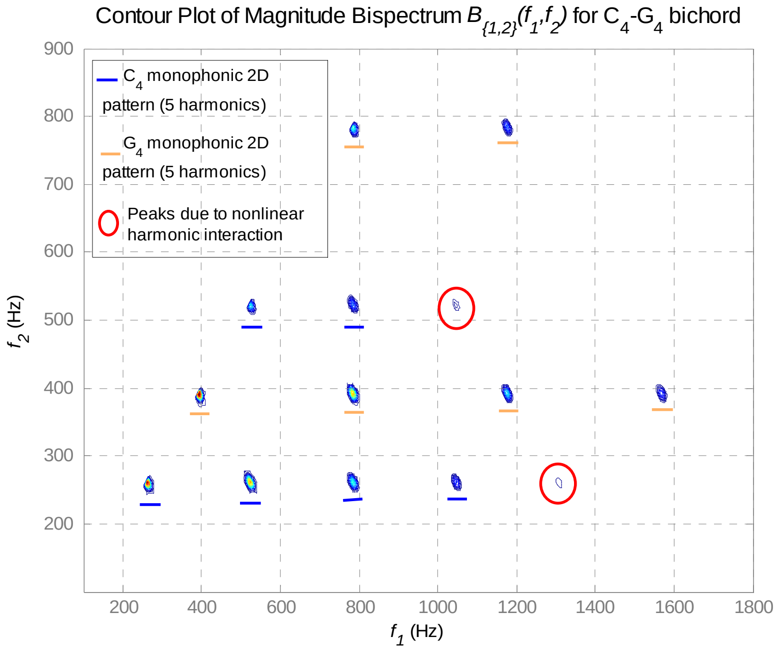

Figure A.1. Contour plot of the bispectrum of synthesized bichord C_4 - G_4.

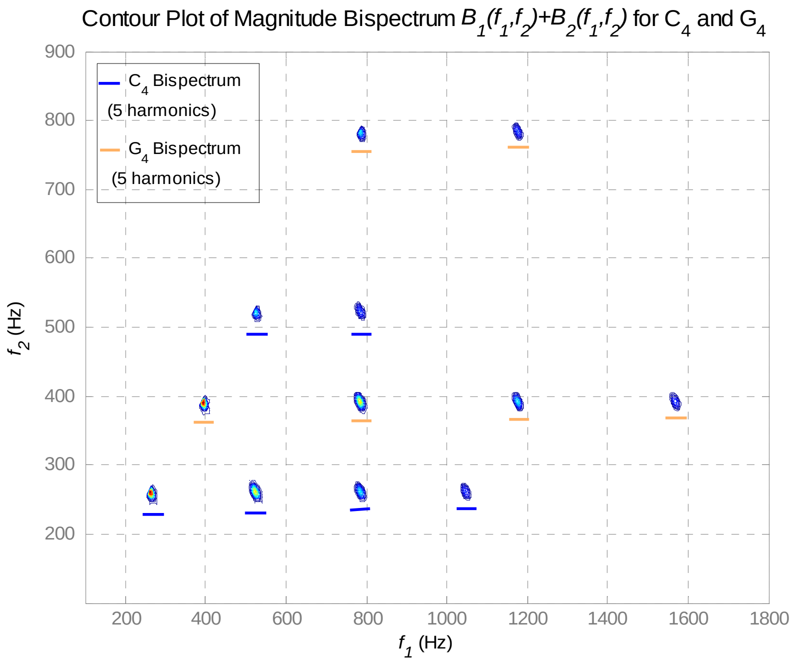

Figure A.2. Contour plot of the sum of the bispectra of two synthesized sounds: C_4 and G_4.

Dealing with real sounds, it is impossible to quantify the amplitude contribute given by each single term present in $B_{diff}(f_1,f_2)$, if the number of partials and their amplitude model is not known in advance for each concurrent sound. For this reason, it is difficult to perform a general qualitative analysis. On the other hand, it is possible to evaluate the normalized 2-D cross-correlations between both $B_{\{1,2\}}(f_1,f_2)$ and $B_1(f_1,f_2)+B_2(f_1,f_2)$ with a 2-D pattern, equivalent to the one used in the Multi-F0 estimation procedure which is the core of the system proposed in the paper. The results of the two normalized 2-D cross-correlation (denoted as $\rho_{B_{\{1,2\}}}$ and $\rho_{B_1+B_2}$) and the array obtained by subtracting $\rho_{B_{\{1,2\}}}$ and $\rho_{B_1+B_2}$, are shown in Figure A.3.

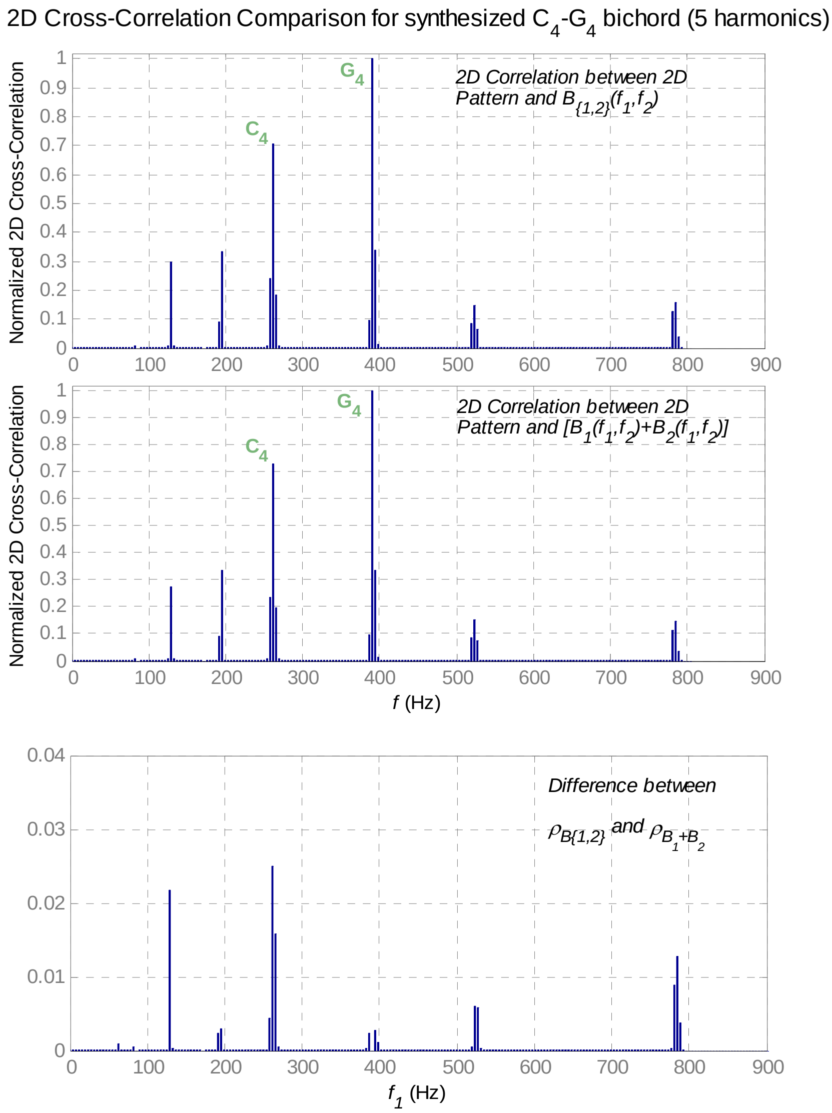

Figure A.3. Comparison of normalized 2-D cross-correlation for 5-harmonics synthesized bichord C_4 - G_4, and the difference of them (with a different scale).

It can be noted that there are no relevant differences between the two cases (in Figure A.3, bottom part reporting the difference, the *y-axis* scale has been enlarged to make difference array more readable).

Moreover, the same normalized 2-D cross-correlation for other two synthesized sounds has been calculated with the same pitch by using 10 harmonics instead of 5. This operation was made in order to show that the contribute of $B_{diff}(f_1, f_2)$ would not affect significantly the values of 2-D correlation (and, therefore, the results of Multi-F0 Estimation procedure) with increasing number of partials. The results are shown in Figure A.4.

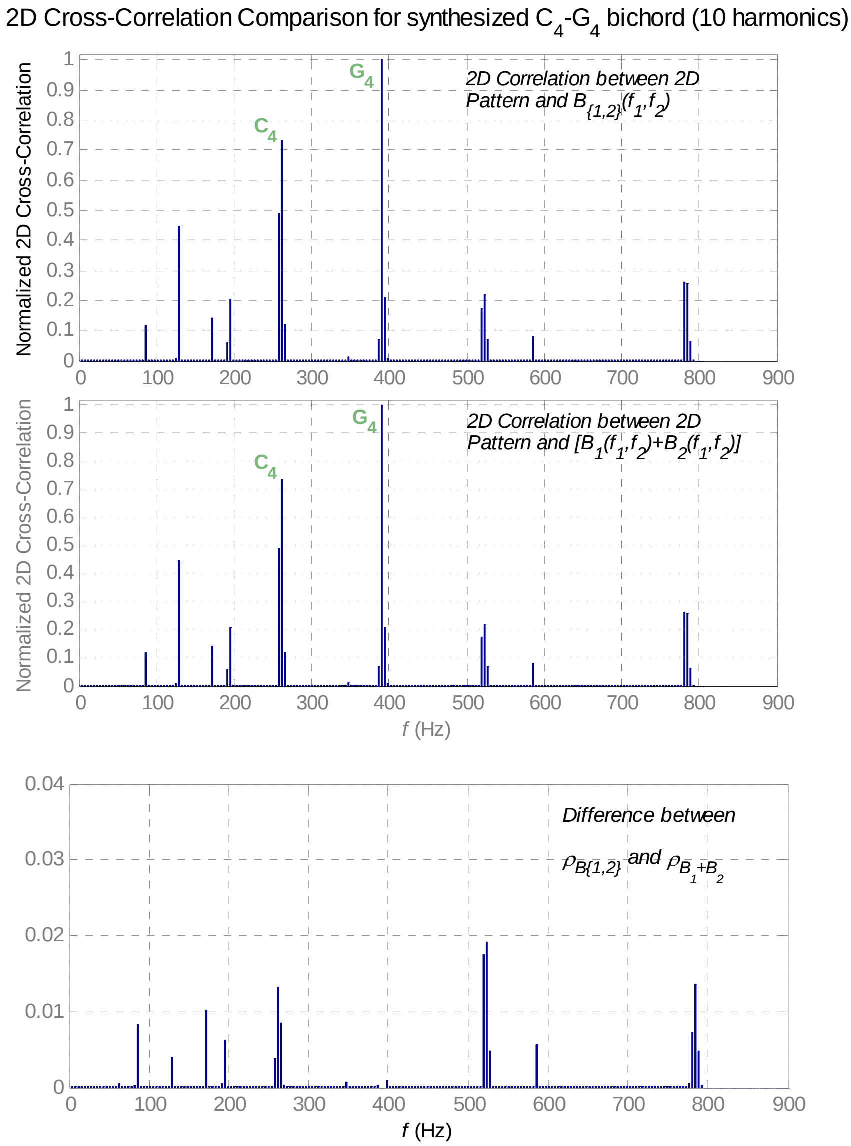

Figure A.4. Comparison of normalized 2-D cross-correlation for 10-harmonics synthesized bichord $C_4 - G_4$, and the difference of them (with a different scale).

A.4 Bispectrum of a polyphonic signal: qualitative analysis for three or more sounds

When a polyphonic audio signal is composed by more than two concurrent sounds, it can be shown, by extending the analysis performed for a bichord in section A.2, that the signal bispectrum may present additional peaks, with respect to the sum of the bispectra of the single monophonic sounds. Even in this case, they do not affect the result of normalized 2-D cross-correlation, since they are located at coordinates which do not belong to the generic pattern of a monophonic sound (see equation (5), Section III.C of the paper). This is shown in the example reported in Figure A.5 where the test signal is a trichord C_4-E_4-G_4, which presents strong partials overlapping. Therefore, in these cases, the effect of $B_{diff}(f_1, f_2)$ is null, whereas it is relevant (in the sense of computation of 2-D cross-correlation), only for those frequency couples given, in case, by the combinations of bichords harmonic sets. If the intersection between the harmonic sets of the concurrent is empty, $B_{diff}(f_1, f_2)$ is null.

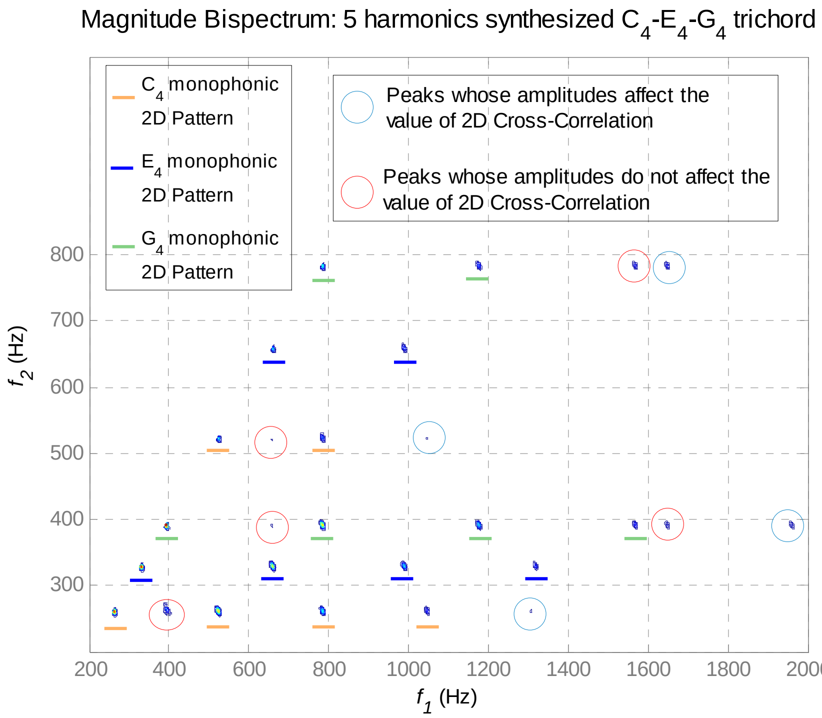

Figure A.5. Contour plot of the bispectrum of 5-harmonics synthesized trichord C_4 - E_4 - G_4, and graphical classification of the peaks of the residual $B_{res}(f_1, f_2) = B_{\{1,2,3\}}(f_1, f_2) - B_1(f_1, f_2) - B_2(f_1, f_2) - B_3(f_1, f_2)$.

Also in this example, the normalized 2-D cross-correlations $\rho_{B_{\{1,2,3\}}}$ and $\rho_{B_1+B_2+B_3}$ have been calculated for the same test signal C_4-E_4-G_4, though this time it was synthesized with 10 harmonics. The result is shown in Figure A.6. The contribute of the residual is still to be considered not relevant, although this time the difference between $\rho_{B_{\{1,2,3\}}}$ and $\rho_{B_1+B_2+B_3}$ is slightly higher, especially for note C_4 and its sub-octave. This is due to the fact that C_4, being the lowest note played in the audio signal, presents partials overlapping with both notes E_4 and G_4.

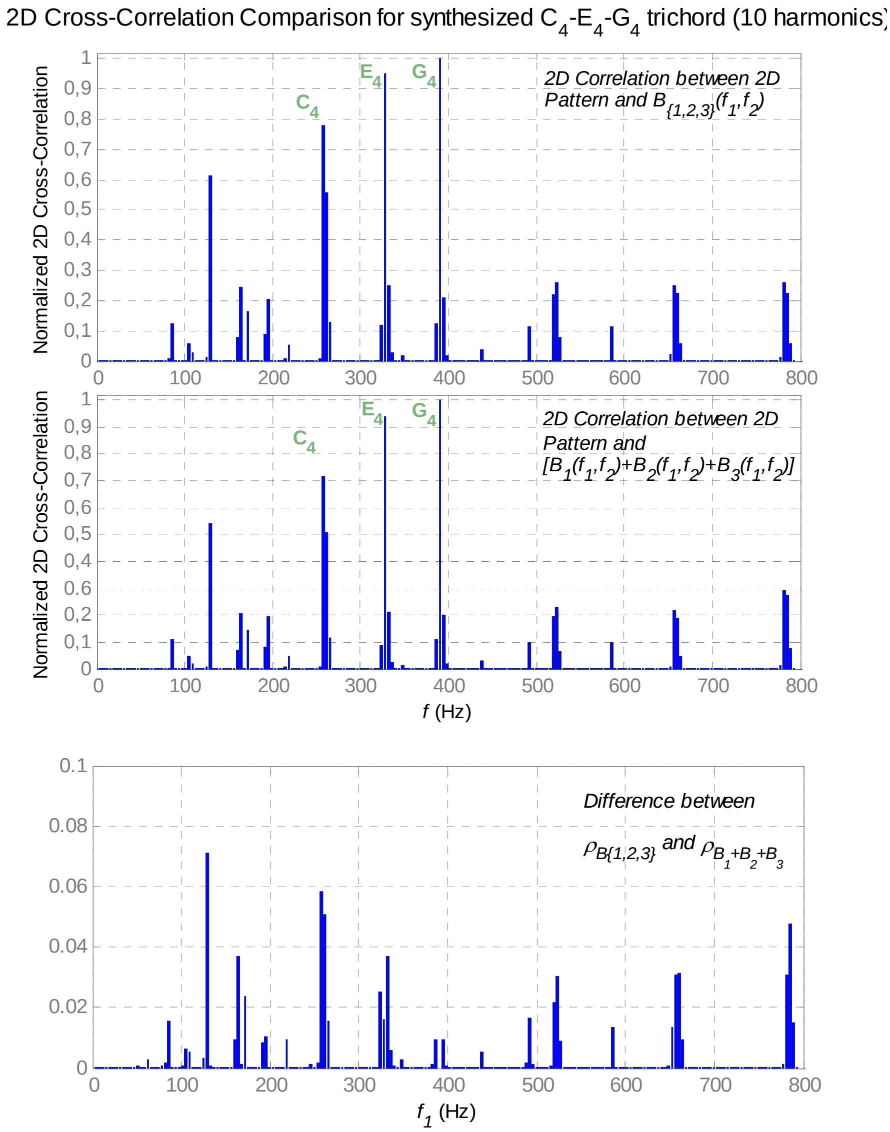

Figure A.6. Comparison of normalized 2-D cross-correlation for 10-harmonics synthesized trichord C_4 - E_4 - G_4.