Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

Hangfelvételek automatikus kottázása

Önálló laboratórium zárójegyzőkönyv
2015/16. I. félév

Nemes Marcell

V. évf, villamosmérnök szakos hallgató
BSc Beágyazott információs rendszerek szakirány

Konzulens:

dr. Bank Balázs docens
(Méréstechnika és Információs Rendszerek Tanszék)

1. Bevezető

A ma is használt zenei szimbólumokhoz hasonlatosakat már a 13. században is használtak, nyomtatott formában a 15. század végétől jelennek meg. A számítógépek megjelenésével lehetőség nyílt a zenei művek kottaszerkesztő programok segítségével való lejegyzésére.

A számítógépek számítási kapacitásának rohamos fejlődésével megnyílt az út a hangfelvételek automatikus kottázására. Mindezek ellenére a feladatnak a mai napig sem létezik olyan definitív megoldása, mely utolérhetné a képzett zenészek hallás utáni kottázási képességét.

Az automatikus kottázás fogalma már a 70-es évek végén megjelent [1], míg napjainkra már erre specializálódott szoftverek is elérhetőek a piacon, bár a pontosságuk koránt sem tökéletes.

A feladat kidolgozásával a célom az volt, hogy egy olyan algoritmust hozzak létre, mely a bemeneti hangfelvételből egy olyan, köztes leírást hozzon létre, melyből a kottázás könnyen elvégezhető. Ilyen forma például a MIDI szabvány, melyből kottázó szoftverek segítségével pontos kottakép rajzolható ki.

Az önálló laboratórium keretében megismerkedtem az automatikus kottázás fő irányvonalaival, és létrehoztam egy olyan egyszerű algoritmust, mely bizonyos korlátok között jó pontossággal felismeri a zenei hangokat. Az alábbi fejezetekben az algoritmus megvalósítását taglalom.

2. A probléma megközelítése

Az automatikus kottázásnak a jelentőségét legkönnyebben a számos felhasználási terület közül néhány felsorolásával lehet bemutatni. A teljesség igénye nélkül ezek a következők.

1. Oktatás – a zeneoktatásban kiemelkedő szerepe lehet egy hasonló programnak, hiszen a tanuló játékát önmagában ki tudja értékelni, jelezni tudja pontosan mikor és miben tévesztett. Az önálló gyakorlást nagyban elősegítheti, hiszen bármely hallott zeneszámot tanulható kottává képes alakítani.
2. Zeneszerzés – a zeneszerzők egy automatikus kottázó segítségével rengeteg munkát megspórolhatnak, hiszen a lejátszott dallamokat azonnal lekottázza, az elektronikus szerkesztő programok segítségével pedig utólag szabadon korrigálhatja azt.
3. Zenefelismerés – az automatikus kottázás segítségével a zenefelismerő programok működését lehet fejleszteni, könnyebb lehet a hasonló dallamokat összevetni

Számos megközelítés létezik a probléma megoldásához, az emberi hallás modellezésétől a hangok fizikai tulajdonságainak vizsgálatáig. [2, 3]

A különböző hangszerek eltérő hangkarakterisztikája, a harmonikus hangok átfedő spektruma mind-mind nehezítik a megoldás keresését. Az időtartománybeli vizsgálat nem alkalmas az egyszerre megszólaló hangok magasságának azonosítására, míg a frekvenciatartománybeli vizsgálat az ütemdetektálást nem teszi lehetővé.

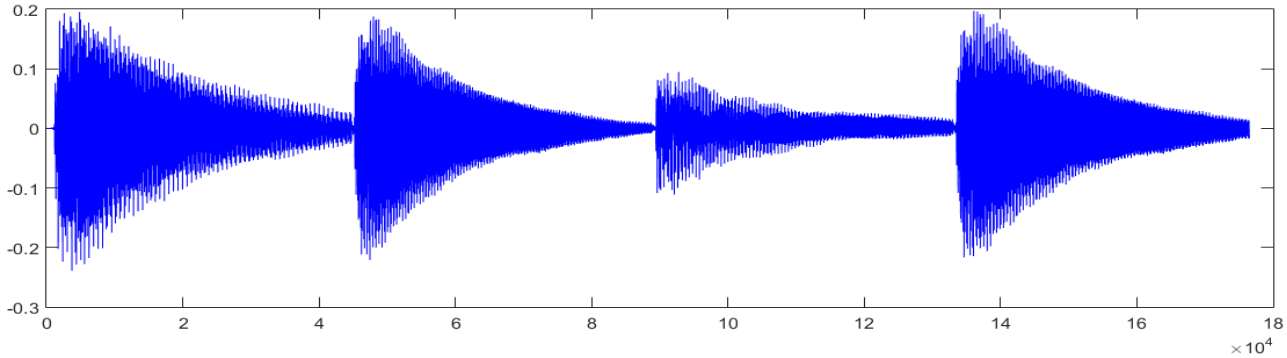
A feladat jellegéből adódóan két fő részre lehet azt osztani. Ezek a következők:

1. Onset detection (támadás detektálása), azaz az elkülönülő zenei hangok elkülönítése az időtartománybeli jelben.
2. Multi-pitch estimation (hangmagasság számítása), azaz a feldarabolt mintákban az egyes megjelenő hangok szétválasztása és azonosítása a frekvenciatartományban.

A fenti felbontást követve értelemszerűen az előbbivel kell kezdeni a feladat megvalósítását.

3. Onset detection

Első lépésként beolvasom a feldolgozni kívánt hangfelvételt. Egy példa bemenet az időfüggvénye látható az 1. ábrán, 44.1 KHz-es mintavételezés mellett.



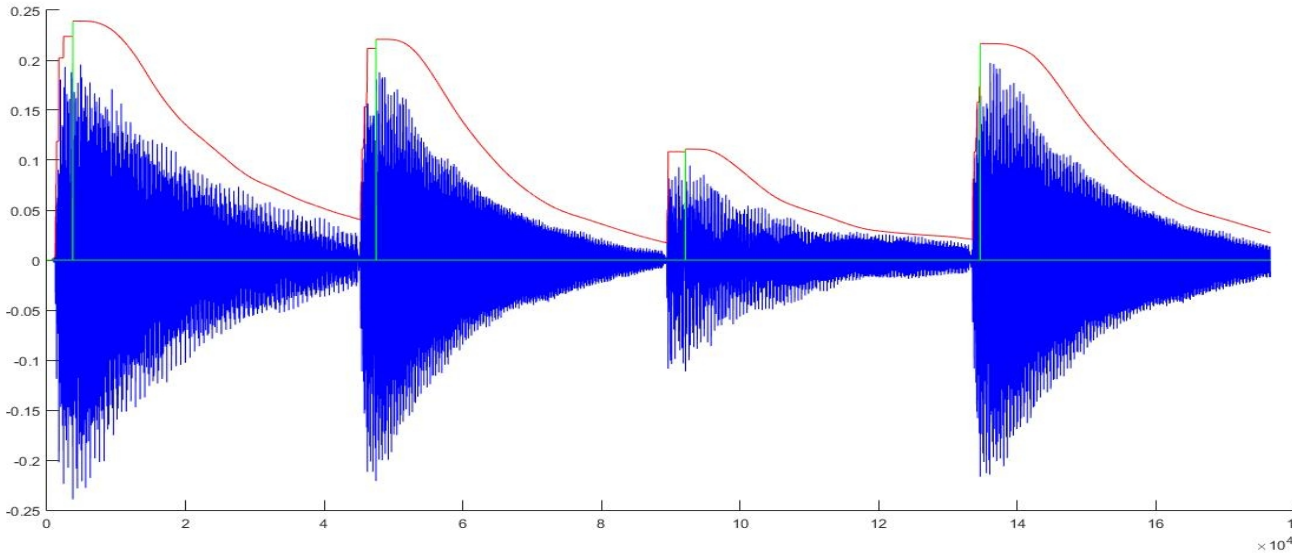
1. ábra: hangfelvétel időfüggvénye

A fenti ábrából következtethetni lehet, hogy az adott mintában hány időben elkülöníthető hang/hangegyüttes szerepel.

Ehhez a jelformához generálok egy detektáló függvényt, aminek a csúcsai fogják adni az egyes hangok/hangegyüttesek helyét. Ezt az alábbi burkoló nemlineáris függvény képezésével nyerem, ahol y jelzi a detektáló függvény diszkrét időfüggvényét, x pedig a bemenet időfüggvényét, p a meredekséget szabályzó paraméter, valamint L a bemenet hossza.

$$k = 2..L \qquad y[1] = x[1]$$
$$x[k] \geq y[k-1]: \qquad y[k] = x[k]$$
$$x[k] < y[k-1]: \qquad y[k] = (1-p) * x[k] + p * y[k-1]$$

A 2. ábrán látható a burkoló piros színnel, valamint zölddel jelölve az egyes felismert csúcsok. Ezeknek a csúcsoknak a környékén kell keresni az egyes hangokat/hangegyütteseket. Ezt a szemléltetésül szolgáló példában a csúcs előtti 50ms-tól a következő csúcs előtti 100ms-ig végzem, illetve az első és utolsó hangnál/hangegyüttesnél a minta elejétől, valamint a minta végéig.



2. ábra: ütemdetektálás az időtartományban

A burkoló 60ms-nál közelebbi csúcsai közül csak a legnagyobb amplitúdójúakat veszem számításba, mivel az emberi fül számára is megkülönböztethetetlenek az egymáshoz ilyen mértékben közeli hangok, így kiküszöbölve több téves detektálást. [4]

A fenti maximumhelyeknek megfelelő minták és a mintavételi frekvencia ismeretében kiszámítható a hang/hangegyüttes támadási időpontja az alábbi képlet segítségével, ahol t_i a keresett hang/hangegyüttes kezdeti időpontja szekundumban, x_i a megtalált maximum, *Fs pedig* a mintavételi frekvencia.

$$t_i = \frac{x_i}{Fs} - 50\,ms$$

Az így kialakult időpontok már felhasználhatóak a hangfelvétel kottázásakor.

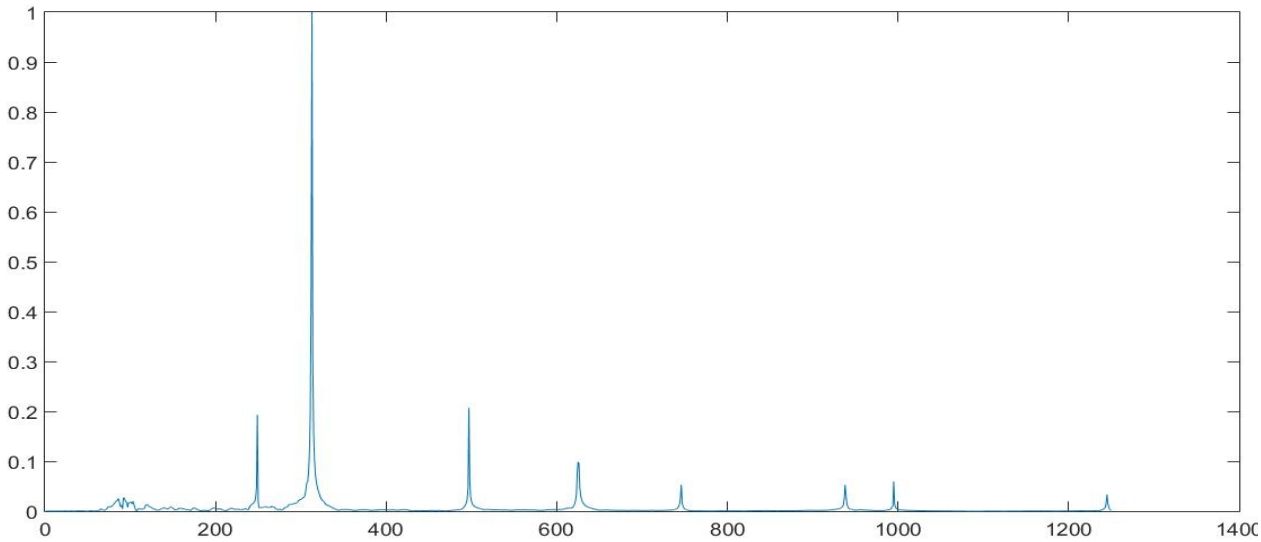
A maximumok által határolt hangmintákban található hangok magasságát a következő fejezetben ismertetett módon számítható ki.

4. Multi-pitch estimation

A feldarabolt minta egyes részeiben ismeretlen számú egyszerre szóló hang lehet jelen. Ezek elkülönítése időtartománybeli számításokkal nem lehetséges, így a minták spektrumát képzem.

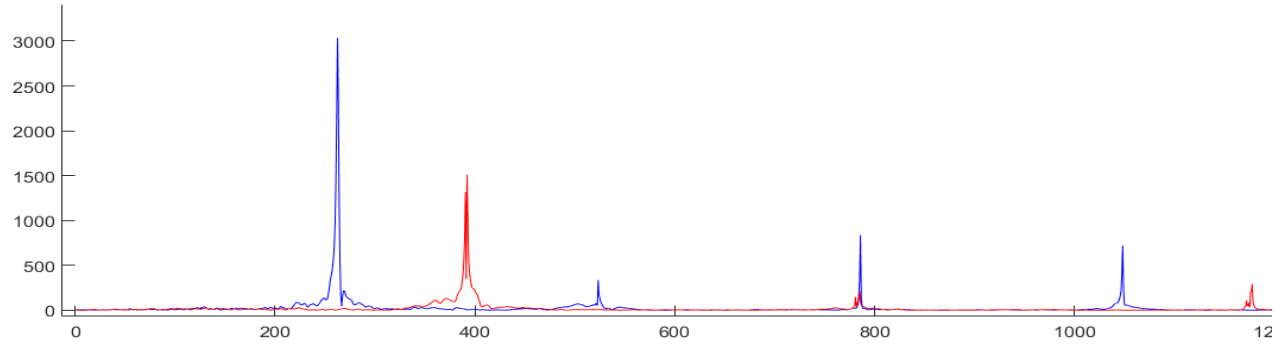
A hangmagasságok megtalálásához elegendő a spektrum pozitív valós része, így ezt képzem az *fft* függvénnyel képzett minta abszolút értékének első feléből.

Az így keletkező spektrum képéről már hasznos információk olvashatóak le. A 3. ábrán látható csúcsok egy-egy csoportja lineárisan oszlik el a frekvencia tengely mentén. Ebből arra következtethetünk, hogy a mintában harmonikus hangok vannak jelen, melyek tulajdonsága, hogy felharmonikusaik egy alapharmonikus egész számú többszörösének megfelelő frekvenciákon jelennek meg.



3. ábra: hangminta spektruma

Mivel nem biztos, hogy egy hangnak az első néhány harmonikusa vagy akár csak az alapharmonikusa is jelen van a mintában, az egyes csúcsok önmagukban nem határoznak meg zenei hangokat. Egy-egy harmonikus több hang spektrumában is szerepelhet, ahogy az a 4. ábrán látható.



4. ábra: közös harmonikus

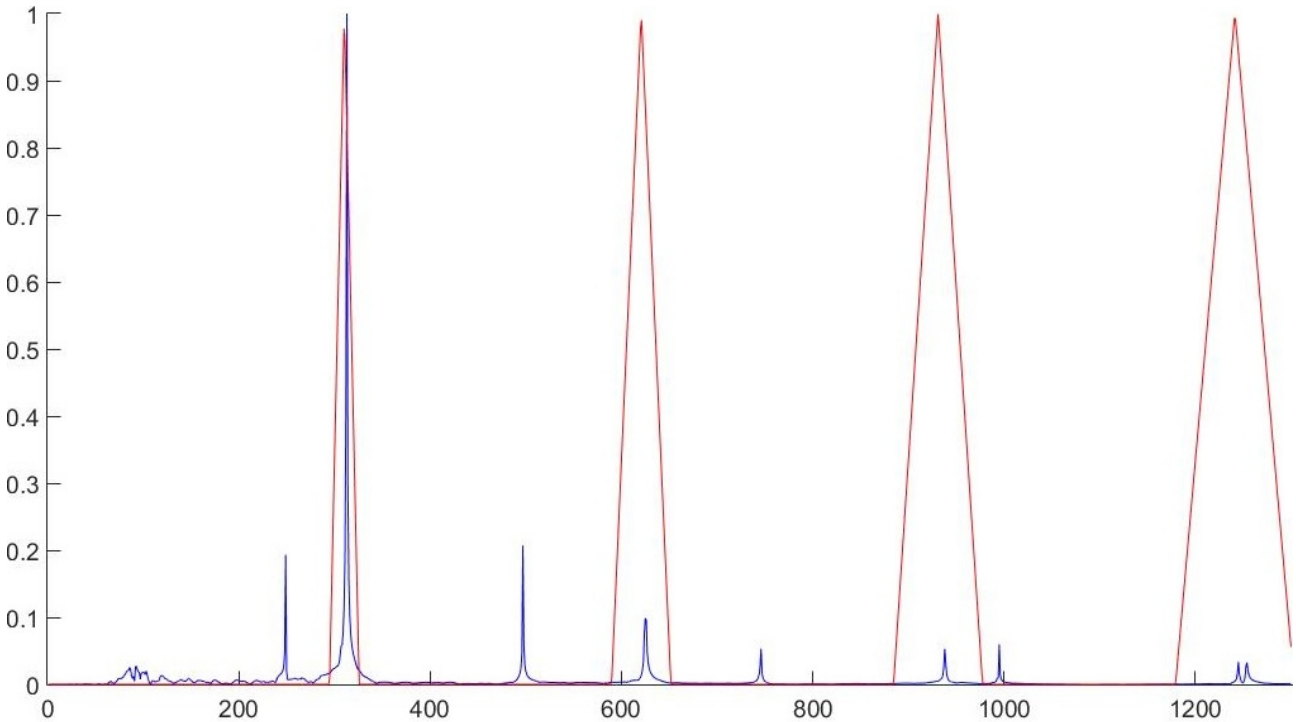
A fenti ábrán látszik, hogy a 261.63 Hz alapfrekvenciájú C hang 2. felharmonikusa (784.89 Hz) közel egybeesik a 392 Hz alapfrekvenciájú G hang 1. felharmonikusával (784 Hz).

Emiatt a hangok megtalálásához a teljes ideális spektrumukkal kell összevetni. Erre célszerű módszer az egyes zenei hangokhoz szerkesztett fésűs szűrők alkalmazása, mellyel a frekvenciatartományban lehetséges a minta ablakozása az egyes hangoknak megfelelően. [5]

Ehhez elsőként kiszámítom a zenei hangok alapfrekvenciáit C0 és B9 (angolszász jelöléssel) (azaz 16.35 Hz és 15,8 KHz) között. (Az emberi fül által hallott tartomány kb. 20 Hz és 20KHz közé esik.) A számításokat a C0-B0 oktáv ismert frekvenciái és az alábbi képlet segítségével végzem, ahol N_i az N hang i. oktávbeli frekvenciája, F_{0N} pedig az N hang ismert 0. oktávbeli frekvenciája.

$$N_i = F_{0N} * 2^i$$

Ezután a zenei hangokon végigiterálva létrehozom a hozzájuk tartozó fésűs szűrőket, ami jelen példában 4 harmonikusnak megfelelő, az ideális frekvencia alatt és felett 0.5%-kal kezdődő, illetve végződő háromszög alakú ablak, ahogy az az 5. ábrán látható.



5. ábra: a fésűs szűrő alkalmazása

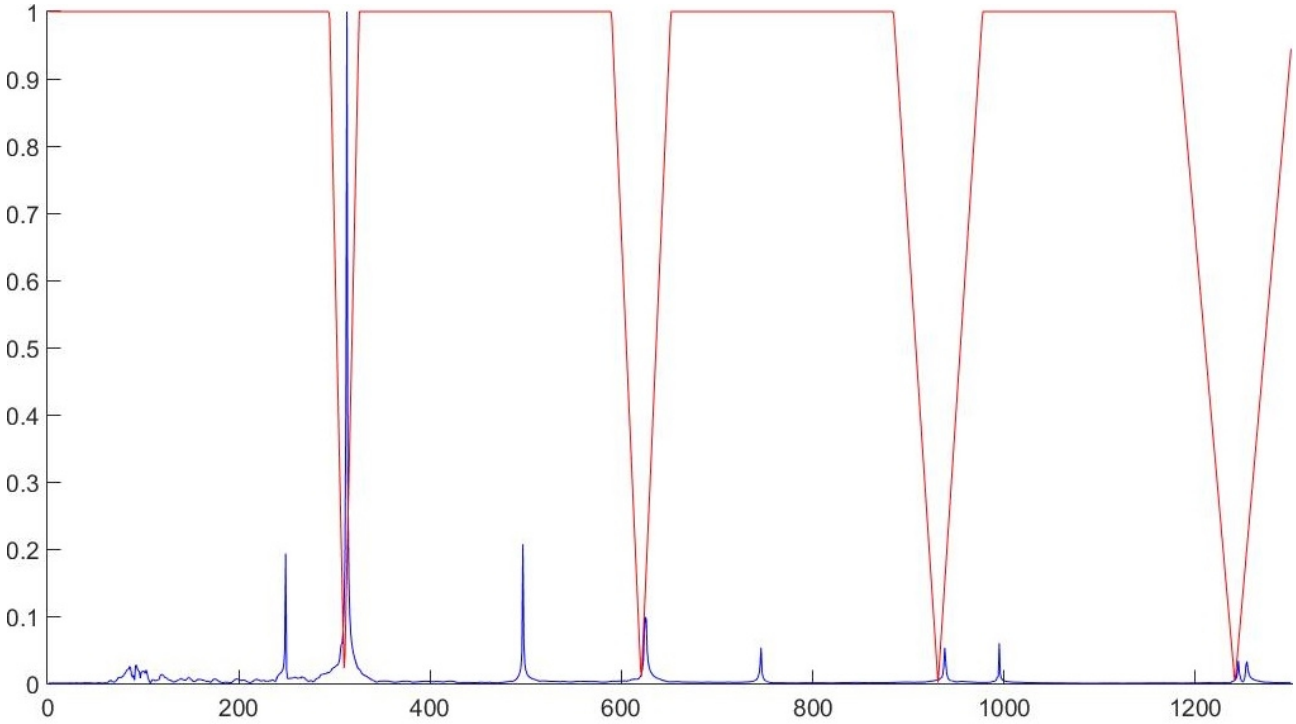
Minden egyes zenei hanghoz kapcsolódóan eltárolom a spektrum és a szűrő szorzatából számított súlyát, majd ezek közül a „legsúlyosabbat” léptetem elő 1. számú felismert hangnak.

Ez a klasszikus gitár hangtartományában (E2-C6), két konkurens hang esetén minden alkalommal jó eredményt produkált. Teszteléshez a hangtartomány összes hangjából képeztem véletlenszerűen hangpárokat, majd az ismert bemeneti párokkal hasonlítottam össze a program kimenetét.

A 2. (és minden további) hang felismerése esetén két eltérő módszert valósítottam meg a programban, összehasonlítás céljából.

1. Második (n-edik) legerősebb „súlyú” hang kinevezése 2. (n-edik) számú felismert hangnak
2. Az 1. (n-1-edik) számú felismert hang felharmonikusainak kivonása a spektrumból, majd az 1. számú felismert hanggal megegyező módon keressük a 2. (n-edik) számú felismert hangot, ahogy az a 6. ábrán látható.

Az első módszerrel, bár futásidőben kedvező algoritmusnak bizonyult, 60%-os pontosságot sikerült elérni a fentebb ismertetett teszt során.

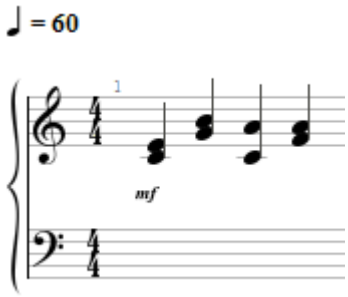


6. ábra: az ideális spektrum kivonása

A második módszerrel sikerült elérni, hogy a legnagyobb „súlyú” hang erős harmonikusai ne nyomják el a valóban jelen lévő többi hangot. A fenti teszt esetén a 2. hangot 87% pontossággal találta el, a fennmaradó hibák többsége pedig oktáv-, kisebb részben kvinttévesztésként jelent meg. Ez azzal magyarázható, hogy a hangok oktávainak harmonikusai egymásnak többszörösei vagy egybeesnek, így a mélyebb hang távolabbi harmonikusainak is tűnhetnek.

5. A teszteset kiértékelése

Az algoritmus bemutatásához használt példa egy négy hangközből álló, zongorán játszott dallam volt, melynek kottája a 7. ábrán látható.

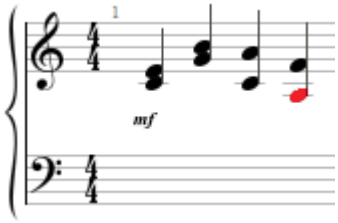


7. ábra: a teszt során használt dallam

A 4. fejezet végén kifejtett második módszert alkalmazva a program az alábbi táblázatban látható kimentet adta.

1. számú megtalált hang	**2. számú megtalált hang**	**időpont**
E4	C4	0.0347 s
G4	B4	1.0263 s
C4	A4	2.0389 s
F4	A3	3.0031 s

A kimenet alapján felrajzolható kotta a 8. ábrán látható, annak figyelembevételével, hogy a 60-as tempójelzés másodpercenként egy negyedhangnak felel meg, amit az időpontok kis eltérését az ideálistól eltekintve vissza is kapunk.



8 ábra: a teszt során reprodukált dallam

Az ábrán pirossal jelöltem az egyetlen tévesztést, ez a 4. hangköz A4 hangja helyett felismert A3 hangot jelöli, ami összhangban van a fentebb jelzett hibaaránnyal.

A tesztet értékelve elmondható, hogy a kitűzött célok megvalósultak, egy továbbfejleszthető algoritmus született, mely segítségével sikerült egy egyszerű dallamot lekottázni.

6. További lehetőségek

Az önálló laboratórium során végzett munkának koránt sincs még vége, a meglévő algoritmust sokáig lehet még pontosítani, további funkciókat hozzáadni.

A lehetséges irányok közül az alábbiakban kiemelek néhány megvalósítandó feladatot.

1. MIDI kimenet generálása – a meglévő információk (hangok időpontja és hangmagasság) ismeretében minden a rendelkezésre áll egy MIDI átalakító megalkotására, hogy egy univerzálisabb köztes leírásban lehessen az adatok további feldolgozását végezni
2. Hangmagasság felismerés pontosítása – megvalósítandó feladat a maradék fennálló hibalehetőség kiküszöbölése, a kérdéses esetekben a spektrum további vizsgálatával hatékonyabb algoritmus alkotható
3. Hangok számának meghatározása – megalkotható egy olyan logika, mely segít meghatározni, hogy az egyes hangmintákban hány hang szól párhuzamosan, általánosítva a megoldást
4. Tempódetektálás – a dallam és a ritmus ismétlődésének vizsgálatával a tempószám meghatározására alkalmas algoritmust lehet készíteni
5. Hangnemfelismerés – a már felismert hangok hangsorokra való illesztésével meghatározható az egyes dallamok hangneme
6. Precíz ritmusdetektálás – a feladat két részének korrelációjával a kitartott hangok felismerése, és az újonnan megszólaltatott hangok elkülönítése is megvalósíthatóvá válik
7. Kezelőfelület – az algoritmus paramétereinek és az import/export lehetőségek könnyebb kezelése érdekében érdemes kezelőfelület készítésén dolgozi

7. Irodalomjegyzék

[1] Martin Piszczalski és Bernard A. Galler – Computer Analysis and Transcription of Performed Music: A Project Report
Computers and the Humanities Vol. 13, No. 3 – 1979
http://www.jstor.org/stable/pdf/30207256.pdf

[2] Anssi P. Klapuri – Introduction to Music Transcription – 2006
http://www.cs.tut.fi/sgn/arg/klap/amt-intro.pdf

[3] Anssi P. Klapuri – Automatic Music Transcription as We Know it Today
Journal of New Music Research Vol. 33, No. 3 – 2004
http://www.cs.tut.fi/sgn/arg/klap/jnmr_klapuri.pdf

[4] Julien Richard – An Implementation of Multi-band Onset Detection
http://www.music-ir.org/evaluation/mirex-results/articles/all/ricard.pdf

[5] Zheng Guibun és Liu Sheng – Automatic Transcription Method for Polyphonic Music Based on Adaptive Comb Filter and Neural Network
IEEE International Conference on Mechatronics and Automation – 2007
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4303965