

Hangelemzés és -szintézis

A mérési leírást kidolgozta: Rucz Péter

Mérési leírás
Akusztika és hangtechnika laboratórium

1. Bevezetés

Jelen mérés célja a gyakorlatban széles körben alkalmazott hangelemző módszerek bemutatása. A mérés során az elemző eljárásokat elsősorban zenei jelekre fogjuk alkalmazni, de a módszerek használhatóak pl. zaj-, rezgés-, vagy akár beszédjelek analízisére is. Mivel praktikusan csak digitalizált (mintavételezett és kvantált) hangjeleket tárolunk, így a 2. szakasz jelentős része is a diszkrétizált hangjelek spektrális vizsgálatát tárgyalja.

A mérés 3. szakaszban tárgyalta, hangszintézissel kapcsolatos részében a jel-alapú additív szintézis alapvető módszereit mutatjuk be. A laboratóriumi foglalkozás végén egy hangszer hangját fogjuk elemezni a frekvencia- és időtartományban, majd hangszintézissel modellezzük a hangszer megszólalását.

2. Hangelemzés

2.1. Spektrális elemzés

A spektrális elemzés során a hangjelet a frekvenciatartományban vizsgáljuk. Zenei jelek esetében ez azért is kézenfekvő, mert a hangjel spektrumából következtethetünk az érzett hangmagasságra, hangszínre. A diszkrét spektrum rövid bemutatása után az ablakozás és a spektrális átlagolás hatásait tárgyaljuk, majd a spektrális burkoló alapján kiértékelhető jellemzőket vizsgáljuk. Végül az idő-frekvenciaanalízishez a spektrogram fogalmát vezetjük be.

2.1.1. Folytonos és diszkrét spektrum

Folytonos jelek spektrumát a Fourier-transzformációval számíthatjuk, melyet az inverz művelettel szokás definiálni:

$$p(t) = \mathcal{F}^{-1} \{ \hat{p}(\omega) \} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{p}(\omega) e^{j\omega t} d\omega, \quad (1)$$

ahol t az idő, ω a körfrekvencia, $j = \sqrt{-1}$ a képzetes egység, $p(t)$ és $\hat{p}(\omega)$ pedig rendre az időtartománybeli jel és annak spektruma a frekvenciatartományban. A Fourier-transzformáció operátorát \mathcal{F} módon jelöljük. Amennyiben a $p(t)$ jel négyzetesen integrálható, úgy a $\hat{p}(\omega)$ spektrum a

$$\hat{p}(\omega) = \mathcal{F} \{ p(t) \} = \int_{-\infty}^{+\infty} p(t) e^{-j\omega t} dt \quad (2)$$

Fourier-transzformációval számítható.

A (2) transzformáció kiértékeléséhez a $p(t)$ jelet a teljes t időtengelyen ismernünk kellene. Hangminták esetén azonban időben véges hosszúságú jelek spektrumára vagyunk kíváncsiak, a jelnek pedig legtöbbször diszkrét időpontokban vett mintáit ismerjük. Ezért a továbbiakban, amikor a jel spektrumáról beszélünk, nem a (2) definícióval megadott (folytonos) spektrumra, hanem a diszkrét spektrumra, vagyis a jel diszkrét Fourier-transzformáltjára (DFT) gondolunk. A diszkrét spektrum számítását kétféleképp vezethetjük be: 1) vagy a (2) képletben szereplő integrál téglányösszeg-közelítést írjuk fel az integrálási határokat $[0, T]$ -nek választva, vagy pedig 2) a $[0, T]$ időintervallumon értelmezett függvény időbeli periodikus kiterjesztésének Fourier-sorát vizsgáljuk. A két megközelítés konstans szorzó erejéig azonos eredményre vezet, ezt az A. függelék mutatja be.

A jel T hosszúságú időszelvényének mintáit ablaknak, regisztrátumnak vagy blokknak szokás nevezni. Ha a mintavételi frekvenciánk f_s , a mintavételi időköz pedig $T_s = 1/f_s$, akkor a regisztrátum $N = T/T_s$ mintát tartalmaz. Ekkor, bevezetve a $p_n = p[n] = p(nT_s)$ jelölést a regisztrátum n -edik mintájára, a DFT k -adik mintáját a

$$\hat{p}_k = \sum_{n=0}^{N-1} p_n e^{-2j\pi \frac{k}{N} n} \quad (3)$$

összefüggésből kaphatjuk meg. Az inverz művelet (inverz diszkrét Fourier-transzformáció, IDFT), mellyel a spektrumból visszaállíthatók a minták:

$$p_n = \frac{1}{N} \sum_{k=0}^{N-1} \hat{p}_k e^{2j\pi \frac{n}{N} k}. \quad (4)$$

A k index a frekvenciabin indexét jelöli, melyhez a $k f_s / N$ frekvencia tartozik. Ebből adódik a DFT frekvenciafelbontása is, ami alatt két szomszédos bin frekvenciakülönbségét értjük: $\Delta f = f_s / N = 1/T$. Érdemes megfigyelni, hogy a frekvenciafelbontást nem a mintavételi frekvencia, hanem a regisztrátum T időtartama határozza meg.

A (3) definíció alapján megfigyelhetjük, hogy a \hat{p}_k diszkrét spektrum k -ban periodikus, a periódus hossza pedig pontosan N .¹ Vagyis, a diszkrét spektrum a frekvenciatengelyen $N\Delta f = f_s$ periódussal megismétlődik. Így a diszkrét spektrumot sokszor a $[0, f_s)$ vagy a $[-f_s/2, f_s/2)$ frekvenciahatárok között ábrázoljuk.

Érdemes még megemlíteni a következőket. A (3) összefüggésben a p_n minták komplex értékűek is lehetnek. Ha

¹Ezt az állítást beláthatjuk, ha a (3) összefüggésbe k helyébe $k + N$ -t helyettesítünk. Ekkor az exponenciálisból kiemelhető egy $\exp^{-2j\pi N/N} = 1$ tag, amivel $p_k = p_{k+N}$ adódik.

a p_n minták valóságosak, akkor a diszkrét spektrumra is jellemző lesz a szimmetria, miszerint: $\hat{p}_n = \hat{p}_{-n}^*$ (a $*$ felso index a komplex konjugálást jelöli), vagyis $\Re(\hat{p}_n) = \Re(\hat{p}_{-n})$, illetve $\Im(\hat{p}_n) = -\Im(\hat{p}_{-n})$. Azaz, valós jelek spektrumának valós része (és magnitúdója) az $f = 0$ frekvenciára szimmetrikus, képzetes része (és fázisa) pedig antiszimmetrikus. Emiatt valós jel esetén a spektrumot legtöbbször a $[0, f_s/2]$ tartományon ábrázoljuk. Valós jeleknél gyakran alkalmazzuk a $\hat{p}_+(f)$ egyoldali spektrumot (*single side band spectrum*), melyet mind folytonos, mind diszkrét spektrumok esetén a

$$\hat{p}_+(f) = \begin{cases} 2\hat{p}(f) & \text{ha } f > 0, \\ 0 & \text{ha } f < 0 \end{cases} \quad (5)$$

módon definiálunk.

Ahogy a folytonos spektrumra, úgy a diszkrét spektrumra is érvényesek a Plancherel és Parseval-tételek. Utóbbi azt mondja ki, hogy a spektrum energiatartó, vagyis

$$\sum_{n=0}^{N-1} |p_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{p}_k|^2. \quad (6)$$

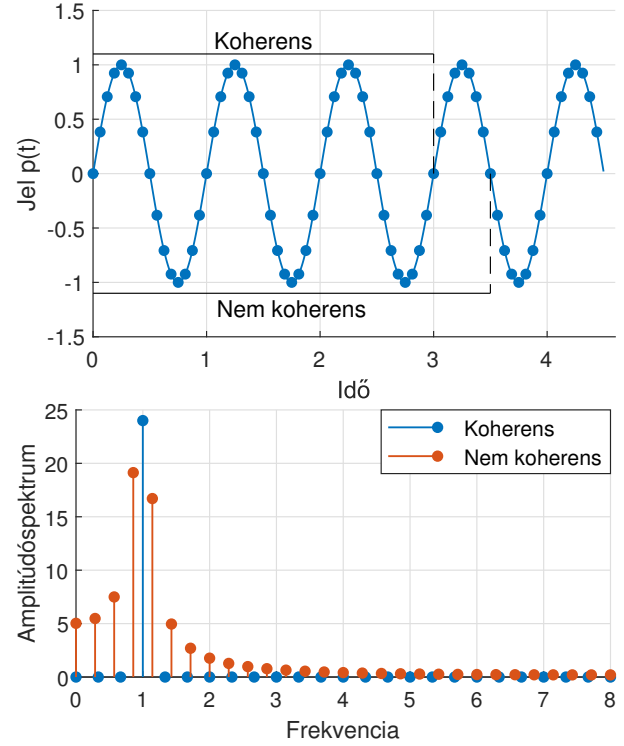
A (6) egyenlet jobb oldalán megjelenő $|\hat{p}_k|^2$ a jel diszkrét spektrumának abszolút érték négyzete, melyet a jel *periodogramjának* is szokás nevezni.

A diszkrét Fourier-transzformációt a gyakorlatban a gyors Fourier-transzformáció (*fast Fourier transform*, FFT) algoritmusával értékeljük ki, mely a \hat{p}_k ($k = 0, 1, \dots, N-1$) értékek a (3) szerinti definíciójában az $\mathcal{O}(N^2)$ művelet hatékonyan átrendezve $\mathcal{O}(N \log N)$ számú művelet elvégzésével számítja ki. Az inverz transzformációhoz az IFFT-t használjuk.

2.1.2. Koherens mintavételezés és ablakozás

Egy ω_0 körfrekvenciájú harmonikus jel folytonos spektrumra két Dirac-impulzust tartalmaz a $\pm\omega_0$ körfrekvenciáknál. A diszkrét spektrumtól is hasonló viselkedést váránk: harmonikus jelet mintavételezve a diszkrét spektrumban a pozitív és negatív frekvenciákon egy-egy különálló impulzusra számíthatunk. Ezt azonban csak akkor kapjuk meg, ha a regisztrátumunk koherensen mintavételezett, vagyis a blokk hosszúsága a jel periódusának egész számú többszöröse. Ha a mintavételezésünk nem koherens, számolnunk kell a spektrális szivárgással és a léckerítés-hatással. Ezekre mutat példát az 1. ábra, ahol ugyanannak a harmonikus jelnek a koherens és nem koherens mintavételezéssel kapott diszkrét amplitúdóspektrumát ábrázoltuk. Az ábrán az idő és a frekvencia egysége lényegtelen, ezért ezeket nem tüntettük fel. Érdekes megfigyelni, hogy bár mindkét spektrumot az $f_s/2$ Nyquist-frekvenciáig ábrázoltuk, a frekvenciabinekhez az eltérő ablakhosszak miatt más frekvenciák tartoznak a két esetben. Látható, hogy a koherens mintavételezés $N = 48$ mintát tartalmazó blokkja esetében az $f = 1$ frekvencián megkapjuk a várt $N/2$ értéket, a többi binre pedig zérus amplitúdó adódik. Ezzel szemben a nem koherens mintavételezés $N = 56$ -os ablakhosszával az $f = 1$ frekvenciára nem esik bin, aminek következményeként a teljes spektrum képe lényegesen megváltozik.

Spektrális szivárgásnak (*spectral leakage*) nevezzük azt a jelenséget, amikor a nem koherens mintavételezés hatására a regisztrátum spektruma a valós frekvenciához



1. ábra. Harmonikus jel diszkrét spektruma koherens és nem koherens mintavételezés esetén.

közeli frekvenciabinekben szétterül. Léckerítés-hatásnak (*picket fence effect*) azt a rokon jelenséget nevezzük, hogy amennyiben a jel frekvenciája nem esik pontosan frekvenciabinre, úgy a tényleges amplitúdót nem tudjuk közvetlenül leolvasni a diszkrét spektrumból, mivel az energia a környező frekvenciabinekbe szivárog. A jelenség onnan kapta a nevét, hogy a diszkrét spektrumot a folytonos spektrum² diszkrét frekvenciákon vett közelítéseként kapjuk, így pedig mintha egy léckerítés lécei közötti réseken keresztül néznénk a folytonos spektrumra.

A spektrális szivárgás hatásait ablakozással (*windowing*) csökkenthetjük. Ilyenkor a regisztrátum mintáit egy alkalmasan megválasztott ablakfüggvény (*window function*) mintáival előszorozzuk, majd az így előállt jel DFT-jét képezzük.

Az ablakfüggvény hatását a diszkrét spektrumra úgy érthetjük meg, ha meggondoljuk, hogy az idő- illetve frekvenciatartománybeli szorzás a transzformált tartományban konvolúciónak felel meg:

$$\begin{aligned} \mathcal{F}\{x(t)y(t)\} &= \mathcal{F}\{x(t)\} * \mathcal{F}\{y(t)\} = \hat{x}(\omega) * \hat{y}(\omega) \\ \mathcal{F}^{-1}\{\hat{x}(\omega)\hat{y}(\omega)\} &= \mathcal{F}^{-1}\{\hat{x}(\omega)\} * \mathcal{F}^{-1}\{\hat{y}(\omega)\} = x(t) * y(t), \end{aligned} \quad (7)$$

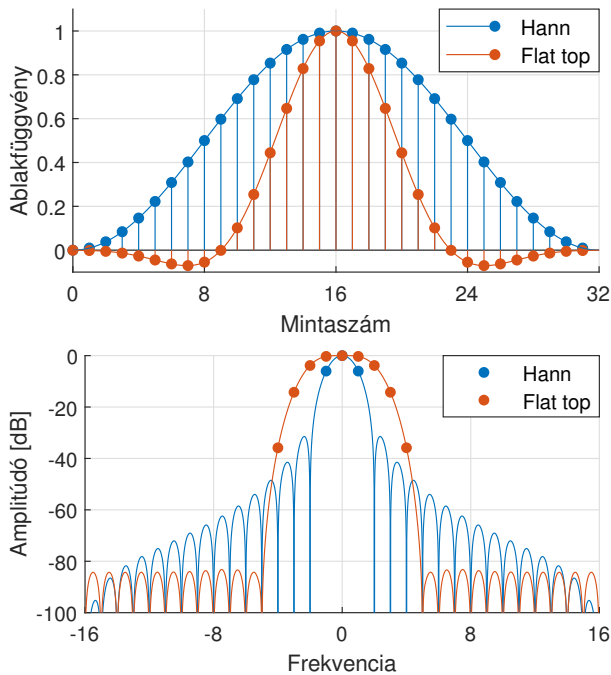
ahol a konvolúció műveletét szokásosan a $*$ operátorral jelöltük. Diszkrét spektrum esetén azonos összefüggések érvényesek, azzal a különbséggel, hogy a konvolúció helyett a *cirkuláris konvolúció* műveletét kell írunk.

$$\begin{aligned} \text{DFT}\{x[n] \cdot y[n]\} &= \text{DFT}\{x[n]\} \otimes \text{DFT}\{y[n]\} \\ \text{IDFT}\{\hat{x}[k] \cdot \hat{y}[k]\} &= \text{IDFT}\{\hat{x}[k]\} \otimes \text{IDFT}\{\hat{y}[k]\} \end{aligned} \quad (8)$$

ahol a cirkuláris konvolúció művelete

$$(x \otimes y)_k = \sum_{n=0}^{N-1} x[k]y[(k-n) \bmod N]. \quad (9)$$

²Pontosabban a diszkrét-idejű Fourier-transzformált (DTFT), lásd az A.3. függeléket.



2. ábra. Két gyakran használt ablakfüggvény (Hann és Flat top) jelalakja és spektruma. A diagramok folytonos vonalai a mintavételezés felszorzásával kapott kvázifolytonos jelalakot és spektrumot mutatják.

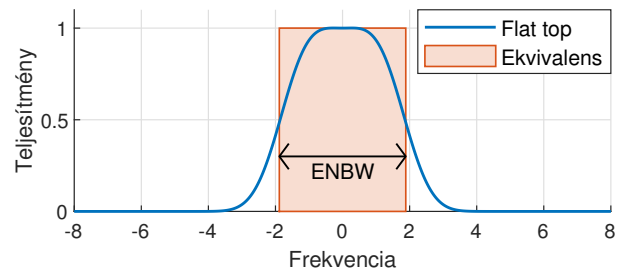
A (9) definícióban szereplő modulo művelettel a $k - n$ indexek $0 \dots N - 1$ tartományon körbefutnak, ami a jel periodikus kiterjesztését jelenti.

Tehát az ablakfüggvény mintáival előszorozva a jel mintáit, a diszkrét spektrumot az ablakfüggvény spektrumával (cirkulárisan) konvolváljuk, ami a spektrum „elkenését” jelenti.

Az adott jelfeldolgozási feladattól függően választunk ablakfüggvényt. Koherens mintavételezés esetében ablakfüggvény alkalmazása nélkül is számíthatunk spektrumot, ezt négyszögablaknak (*rectangular* vagy *boxcar window*) is szokás nevezni. A 2. ábra két gyakran alkalmazott ablakfüggvény, a Hann- illetve Flat top-ablakok viselkedését mutatja be. Az időtartományt mutató diagramon megfigyelhetjük, hogy mindkét ablak a regisztrátum szélein a zérus értékhez simítja a jelet, melynek célja, hogy a regisztrátum periodikus kiterjesztése ablakozás után folytonos legyen (azaz, ne legyen benne ugrás). A spektrumábrákon diszkrét frekvenciát a megszokottól eltérően úgy ábrázoltuk, hogy a zérus frekvencia kerüljön középre.

A Hann-ablak³ emelt koszinuszablak, mely függvény előnye hogy első két időderiváltja is folytonos, amivel a spektrumban erős melléknáláb-elnyomás (-60 dB/dekád, vagyis -18 dB/oktáv) érhető el. Érdekes megfigyelni azt is, hogy a Hann-ablak diszkrét spektrumában a zérus frekvencia mellett csak a két szomszédos binen látunk zérustól különböző amplitúdót, a többi bin pontosan az oldalnyálábok zérushelyeire esik.

A Flat top-ablakot akkor érdemes használni, ha nem koherens mintavétel mellett szeretnénk a spektrális amplitúdókat pontosan megfigyelni. A spektrumábrán jól látható, hogy a zérus frekvenciánál kapott csúcs környékén az amplitúdóspektrum lapos (innen is a „flat top”



3. ábra. Ekvivalens zaj-sávszélesség (ENBW) definíciója a Flat top-ablakfüggvény példáján.

elnevezés), ami a (8) szerinti konvolúcióban a léckerítés-hatást csökkenti. A Hann-ablakhoz hasonlóan a középső bin mellett csak a szomszédos négy-négy binben kapunk nemzérus amplitúdót, a többi bin a melléknálábok zérushelyeire esik. A melléknálábok szintje alacsony, viszont a csúcstól távolodva egy adott határ után a nyálábok szintje nem csökken, csúcértékük konstans marad.

Az ablakfüggvények jellemzésére a melléknáláb-nyomáson túl gyakran használt paraméter, az ekvivalens zaj-sávszélesség (*equivalent noise bandwidth* – ENBW), mely annak a frekvenciatartománybeli négyszögablaknak a szélessége, mely az ablakfüggvénnyel azonos teljesítményt enged át és magassága az ablakfüggvény spektrumának maximális amplitúdójával egyezik meg, ahogy a 3. ábra szemlélteti. Az ábrán az előbbi Flat top ablak teljesítményspektrumát és az ekvivalens négyszögablakot látjuk. Az ENBW definíció szerinti számítása a $w[n]$ ablakfüggvényre:

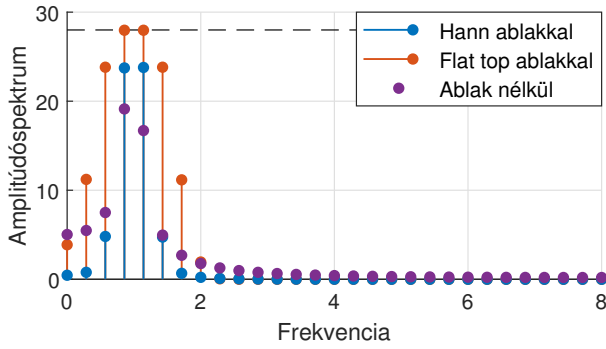
$$\text{ENBW} = N \frac{\sum_{n=0}^{N-1} w[n]^2}{\left| \sum_{n=0}^{N-1} w[n] \right|^2} = N \frac{E_w}{|\hat{w}[0]|^2}, \quad (10)$$

ahol kihasználtuk, hogy a számláló az ablak teljes E_w energiáját, a nevezőben szereplő szumma pedig az ablak zéró frekvenciához tartozó spektrumértékét adja meg. Ha az ENBW-t nem dimenziótlan, hanem fizikai frekvenciában fejezzük ki, úgy az N szorzótényezőt az f_s mintavételi frekvenciára cseréljük.

Az ablak sávszélessége azért is fontos jellemző, mert az ablakfüggvénnyel való szorzás nem csak a jelalakot, hanem természetesen a regisztrátum spektrumának teljesítmény- és amplitúdószintjeit is befolyásolja. Az eredeti amplitúdók kinyeréséhez az ablak hatását kompenzálnunk kell. Egységnyi amplitúdójú szinuszjelet koherensen mintavételezve ablakozás után nem a várt $N/2$ értékű csúcsokat kapjuk meg, hanem $|\sum_n w[n]|/2 = |\hat{w}[0]|/2$ nagyságú csúcsokat kapunk, amit az amplitúdó kiértékelésekor figyelembe kell vennünk. Az amplitúdókorrekció és a zaj-sávszélesség közti kapcsolatot alább a 2.1.3. szakaszban részletezzük.

A 4. ábra a korábbi 1. ábrán látható, nem koherens mintavételű esetre ($N = 56$ mintából álló regisztrátumra) mutatja be az ablakozás hatását, figyelembe véve az imént tárgyalt amplitúdókorrekciót. A valódi amplitúdót a vízszintes szaggatott vonal mutatja. Figyeljük meg, hogy a Flat top ablakkal nagyon jó közelítéssel megkapjuk az ideális amplitúdót. Ennek ára, hogy a spektrum „szoknyája” jóval szélesebb lesz, mint Hann-ablak használata esetén, ami viszont a jel amplitúdóját becsli alul. Az ablak nélküli esetben az amplitúdót sem kapjuk meg

³Esetenként „Hanning”-ablakként szerepel, azonban ez a megnevezés téves.



4. ábra. Nemkoherensen mintavételezett szinuszjel spektruma különböző ablakkal, amplitúdókorrekció után.

helyesen, és a spektrális szivárgás hatása is nagy. Megjegyezzük, hogy az ábrán látható eset különösen rossz (szinte *worst case*), mert a jelből három és fél periódusnyit tartalmaz a regisztrátum. A félperiódus miatt a spektrumban zéró frekvenciás (DC) szint is megjelenik, illetve a valódi spektrális csúcs pontosan két bin közé fog esni (v.ö. az 1. ábra koherens esetével).

2.1.3. Spektrális átlagolás

Bármilyen valós mérés esetén a mérésünket zaj terheli. Általában a zaj hatását a mérendő jeltől független, adott autokorrelációs függvénnyel jellemezhető sztochasztikus folyamat realizációjaként értelmezzük. Legegyszerűbb esetben azt feltételezzük, hogy a minták sorozatát az ideális jel mintáihoz egymástól független, zérus várható értékű, Gauss-eloszlású valószínűségi változók realizációit hozzáadva kapjuk. Utóbbi feltevés akkor helytálló, ha a valódi zaj teljesítménysűrűsége a Nyquist-frekvenciáig jó közelítéssel állandónak tekinthető. Ekkor

$$p[n] = p_0[n] + \sigma v[n], \quad (11)$$

ahol p_0 az ideális (zajmentes) jel, $v[n]$ egy normális eloszlású valószínűségi változó egy realizációja, σ^2 pedig a zaj teljesítménysűrűsége. A jel-zaj viszonyt (*signal to noise ratio* – SNR) az ideális jel és a zaj időablakba eső energiájának hányadosával definiáljuk:

$$\text{SNR} = \frac{\sum_{n=0}^{N-1} p_0[n]^2}{\sum_{n=0}^{N-1} \sigma^2 v[n]^2}. \quad (12)$$

Az időablak hosszát növelve a nevező várható értéke az $N\sigma^2$ értékhez tart, a számláló pedig a jel energiáját adja meg. Vagyis az SNR a jelteljesítmény és zajteljesítmény⁴ arányához fog tartani.

A mérés jel-zaj viszonyának növelése érdekében átlagolást használunk. Amennyiben koherens mintavételt tudunk biztosítani, úgy L egymást követő időablak időbeli átlagából, vagyis a

$$\langle p[n] \rangle = \frac{1}{L} \sum_{l=0}^{L-1} p[n + lN] \quad (13)$$

átlagolt jelből számíthatjuk ki a spektrumot. Az időbeli átlagolás előnye, hogy míg a hasznos jelet azonos fázisban vesszük minden regisztrátumban, a zajkomponensek

⁴Zajteljesítmény alatt itt a zaj Nyquist-frekvencia alá eső teljesítményét értjük.

fázisa véletlenszerűen változik az egyes regisztrátumokban. L regisztrátumot átlagolva a zajteljesítmény várható értéke $1/L$ részére csökken, így a zérus várható értékű zajminták a regisztrátumok L számának növelésével fokozatosan kiátlagolódnak, ahogy azt az 5(a). ábra is mutatja. Érdeemes megjegyezni, hogy mivel a DFT művelete lineáris, a (13) módon előálló időátlagból számított spektrum azonos lesz az egymást követő regisztrátumok spektrumának átlagával.

Nem koherens mintavételezésnél azonban nem használhatunk időbeli átlagolást, mivel az nemcsak a zaj, hanem a hasznos jel amplitúdóját is csökkentené az egyes időablakok eltérő fázisú mintavétele miatt. Ilyenkor spektrális (amplitúdó)átlagolást alkalmazhatunk, mellyel a hasznos jelet terhelő zaj spektrális teljesítménysűrűségének becslését javítjuk. Az átlagos amplitúdóspektrumot a

$$\langle \hat{p}_k \rangle = \sqrt{\frac{1}{L} \sum_{l=0}^{L-1} |\hat{p}_k^{[l]}|^2} \quad (14)$$

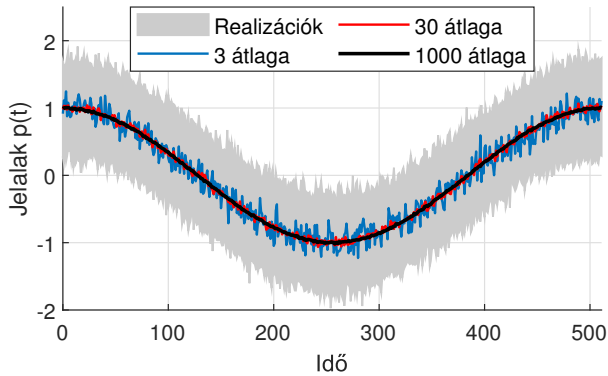
módon számítjuk, ahol $\hat{p}_k^{[l]}$, az l -edik időablakból számított diszkrét spektrum k -adik mintája. Mivel csak az amplitúdókat átlagoljuk, az átlagolás során a fázisinformáció elvész. A spektrális átlagolás során a jelet terhelő zaj teljesítményét nem tudjuk csökkenteni, ezt mutatja az 5(b). ábra is. Ahogy látható, az időtartománybeli additív zaj az egyes regisztrátumok zajplotójának ingadozásában jelenik meg, melyet a spektrális átlagolás a regisztrátumok számának növelésével fokozatosan kisimít.

Az 5(c). ábra a jel-zaj viszony alakulását mutatja a felső két ábrarészben bemutatott szimulált jelek esetén, különböző számú regisztrátum átlagolásával. A jel-zaj viszony kiértékeléséhez az ideális jelteljesítményt a zajkomponensek spektrumokból kiszámított teljesítményével hasonlítottuk össze. Láthatjuk, hogy a spektrális átlagolás a jel-zaj viszonyt valóban nem javítja. Időátlagolás esetén pedig visszakapjuk, hogy a jel-zaj viszony az átlagolt regisztrátumok L számával egyenesen arányosan nő.

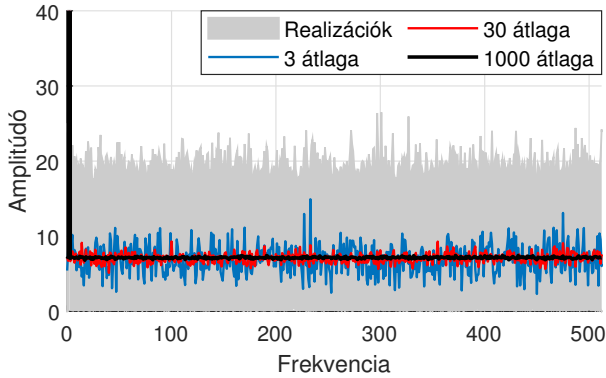
Végül vizsgáljuk meg a 2.1.2. szakaszban tárgyalt ablakozás hatását a mérést terhelő zaj spektrumára a spektrális átlagolás kihasználásával. Az egyes regisztrátumokból az ablakfüggvény alkalmazása után számítunk spektrumokat, majd ezekből spektrális átlagot, melyen elvégezzük az ablakozás hatását korrigáló amplitúdókompenzációt. E műveletek elvégzése után hasonlítsuk össze a teljesítményspektrumok szélessávú zajt tartalmazó szakaszait, ahogy azt a 6. ábra mutatja. Figyeljük meg, hogy az ablakozás és az amplitúdókompenzáció együttes hatására a szélessávú zaj szintje megemelkedik, az emelkedés mértéke pedig az ablak (frekvenciabinekből kifejezett) ekvivalens zaj-sávszélességével (ENBW paraméterével) egyezik meg, amit az ábrán dB skálán ábrázoltunk. Emiatt a dimenziótlan ekvivalens zaj-sávszélességet energiakorrekciós tényezőnek (*energy correction factor*, ECF) is szokás nevezni.

Az 1. táblázat összefoglalja a fent megismert ablakfüggvények amplitúdó- és energiakorrekciós tényezőit szorzótényező és dB alakban, az alábbi formulákkal kiszámítva:

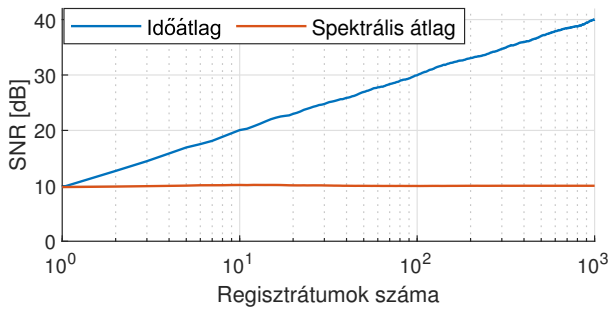
$$\text{ACF} = \frac{N}{\left| \sum_{n=0}^{N-1} w[n] \right|} \quad \text{ECF} = \frac{N \sum_{n=0}^{N-1} w[n]^2}{\left| \sum_{n=0}^{N-1} w[n] \right|^2}. \quad (15)$$



(a) Időbeli átlagolás

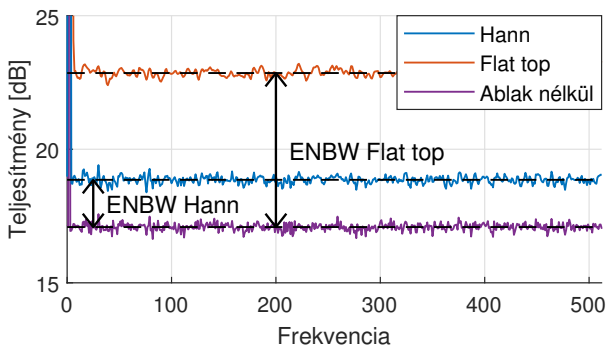


(b) Spektrális átlagolás



(c) Átlagolás hatása a jel-zaj viszonyra

5. ábra. Időbeli és spektrális átlagolás.



6. ábra. Az ablakozás hatása és az ekvivalens zaj-sávszélesség megjelenése a spektrális átlagolásban.

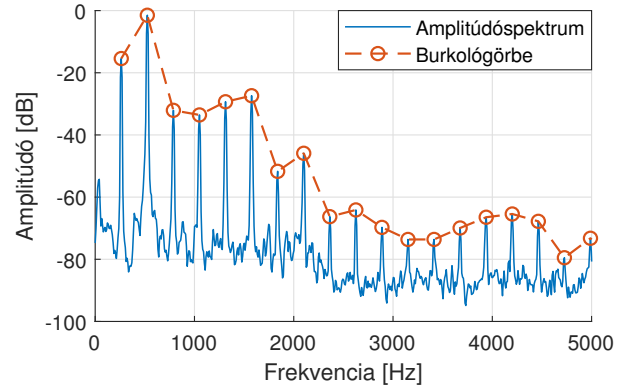
Elegendően nagy N értékek esetén az eredmények a táblázatban megadott adatokhoz tartanak.

2.1.4. Burkoló és formánsszerkezet

Hangelemzés során gyakorta vagyunk kíváncsiak pl. zenei hangok hangszínére, illetve azokra az objektív mérő-

Ablakfüggvény	ACF		ECF	
		dB		dB
Négyszög (nincs ablak)	1.000	0.000	1.000	0.000
Hann	2.000	6.021	1.500	1.761
Flat top	4.639	13.329	3.770	5.764

1. táblázat. Néhány ablakfüggvény amplitúdó- (ACF) és energiakorrekciós tényezője (ECF).



7. ábra. Fagott C4 ($f_1 \approx 263$ Hz) hangjának spektruma és burkológörbéje.

számokra, melyekből az érzékelt hangszínre következtethetünk. Ekkor az előbbi szakaszokban tárgyalt ablakozást és spektrális átlagolást arra használhatjuk, hogy a rögzített hangjelekből a lehető legtöbb információt kapjuk a jel spektrumára nézve. Periodikus vagy kváziperiodikus⁵ jelek spektruma a jellegzetes harmonikusokból álló szerkezetet mutatja, melyre a 7. ábra mutat példát. Ahogy látható, a harmonikus részhangok (*partials*) a spektrumban az f_1 alaphfrekvencia többszöröseinél jelennek meg, jóval az alapzaj (*baseline*) szintje felett. Mivel a spektrumot a harmonikusok dominálják, ezek kiértékelésére fókuszálunk.

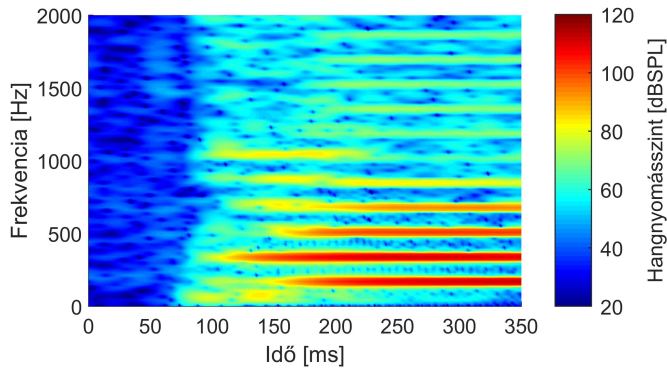
A harmonikus csúcsokat összekötő (képzeltbeli) görbét a spektrum burkológörbéjének (*spectral envelope*) nevezzük. Ahogy a 7. ábrán a burkológörbe frekvenciamentében is láthatunk jellegzetes lokális maximumhelyeket. Ezeket formánsoknak szokás nevezni, a lokális csúcsok eloszlását pedig a spektrum formánsszerkezetének (*formant structure*). A formánsok vizsgálatának például a beszédfeldolgozásban van nagy szerepe, ugyanis az emberi beszédben az egyes magánhangzók kiejtések a hangszálaink által meghatározott f_1 frekvencia közel állandó, az érzett hangmagasság- és hangszínbeli különbségeket az eltérő formánsszerkezetnek köszönhetjük, melyet a toldalékcső (*vocal tract*) alakjának változtatásával idézünk elő. A formánsfrekvenciákat legtöbbször az F_1, F_2, F_3, \dots módon jelölik.

A hangszín jellemzésére szolgáló egyszerű mérőszám a spektrális súlypont (*spectral centroid*), melyet a $|\hat{p}[k]|^2$ teljesítményspektrumból a

$$C = \frac{f_s}{N} \frac{\sum_{k=0}^{N/2} k |\hat{p}[k]|^2}{\sum_{k=0}^{N/2} |\hat{p}[k]|^2} \approx f_1 \frac{\sum_{n=1}^{N_h} n |\hat{p}(nf_1)|^2}{\sum_{n=1}^{N_h} |\hat{p}(nf_1)|^2} \quad (16)$$

módon számíthatunk ki. A jobb oldali közelítés (kvá-

⁵Itt: az ideális periodikus jeltől csak kicsiny ingadozások és zaj következtében eltérő.



8. ábra. Orgona ajaksíp hangindításának spektrogramja.

zi)periodikus jelek esetén használható, ahol csak az N_h darab harmonikus komponenset vesszük figyelembe, kihasználva, hogy ezek dominálják a spektrumot. Figyeljük meg, hogy a (16) egyenletben a számláló a frekvencia spektrum-négyzettel (teljesítménnyel) súlyozott átlaga, melyet a nevező a jel teljes energiájával normalizál.

2.1.5. Spektrogram

Spektrogram alatt a jel egymást követő időszeleteiből képzett diszkrét spektrumainak sorozatát értjük. Segítségével az idő- és frekvenciatartományban egyszerre vizsgálhatjuk a jelet. Ahogy azt a 2.1.1. szakaszban láttuk, a diszkrét spektrum frekvenciafelbontását a spektrumszámításhoz használt időablak T hossza határozza meg. Így egymást követő időablakkal az elérhető ΔT idő- és Δf frekvenciafelbontást a

$$\Delta f \Delta T \geq 1 \quad (17)$$

határozatlansági reláció korlátozza.

A spektrogram látszólagos időfelbontását átlapoló időablakok használatával növelhetjük, vagyis az egymást követő spektrumokat részben átfedő időablakokból számítjuk. Ekkor az egymást követő spektrumok nem lesznek függetlenek, hiszen részben azonos időtartománybeli jelekből képezzük azokat. A spektrogram számítása során leggyakrabban Hann-ablakot használunk a spektrális szivárgás minimalizálása és a szélessávú zaj szintjének alacsonyan tartása végett. Hangjelek spektrogramját szonogramnak is szokás nevezni.

A 8. ábra egy orgonasíp megszólalásának kezdeti szakaszát mutatja be spektrogram formájában. Jól megfigyelhető, hogy az állandósult állapot elérése előtt az egyes részhangok amplitúdója időben más-más módon változik: például a hatodik részhang kezdetben egész erősen megjelenik, az állandósult állapotban viszont majdnem teljesen eltűnik. A spektrogram szokásos ábrázolásában az idő a vízszintes, a frekvencia a függőleges tengely mentén változik, az amplitúdót pedig színskálával ábrázoljuk. Esetenként a tengelyek szerepe felcserélődhet, vagy 3D felület formájában is ábrázolhatjuk a spektrogramot. Utóbbi hátránya, hogy a szintek leolvasása nehézkes.

2.2. Frekvenciamérés és újramintavételezés

Zenei jelek feldolgozása esetén gyakori, hogy kváziperiodikus jelet dolgozunk fel. Ilyenek például a fúvós vagy

vonós hangszerek tartott hangjai, vagy az énekek hangok állandósult állapotú szakaszai, de közelítésként lassan lecsengő hangokat, pl. pengetett vagy megütött húr hangját is tekinthetjük kváziperiodikusnak. Előbbiek esetén a gerjesztés (megfúvó nyomás vagy vonóerő) kicsiny ingadozásai, míg utóbbi esetben a lecsengésből adódó lassú amplitúdócsökkenés miatt nem tökéletes a periodicitás. Természetesen más alkalmazásoknál is sokszor találkozhatunk kváziperiodikus jelekkel, ahol a tökéletes periodicitástól való eltérés változatos fizikai magyarázatra vezethető vissza.

Bár a diszkrét spektrum frekvenciafelbontását a regisztrátum T hossza korlátozza ($\Delta f = 1/T$, ahogy fent láttuk), kihasználva, hogy a jel közel periodikus, sokkal jobb becslést adhatunk a pontos frekvenciára. A B. függelék néhány, a gyakorlatban is jól alkalmazható frekvenciabecslő módszert mutat be röviden. Ezek közül alkalmazásfüggően – vagyis, annak függvényében, hogy mit feltételezünk a mért jelről, – választjuk ki a megfelelőt. Például egy pisztonfon kalibrátor mért jeléről feltételezhetjük, hogy a frekvenciája állandó, ekkor jól használható a harmonikus illesztés. Ugyanakkor, ha a frekvenciamérést nem újramintavételezéshez akarjuk használni, elegendő lehet a gyorsabb spektruminterpoláció is. Lassan lecsengő vagy felfutó jelek esetén pedig a komplex exponenciális illesztése lehet a legjobb választás.

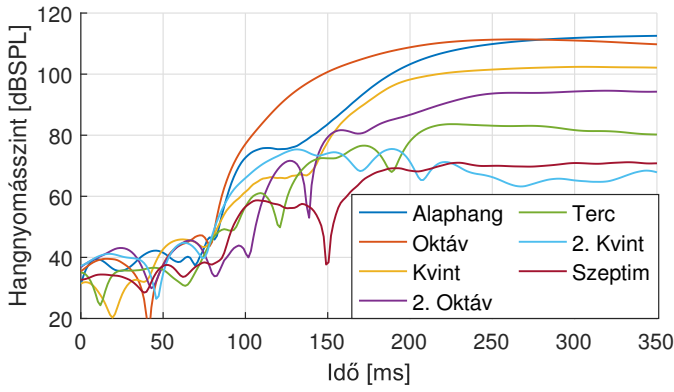
A pontos frekvenciamérés után lehetőségünk van a (kvázi)periodikus jel újramintavételezésére, melynek során az új mintavételi frekvenciát úgy választhatjuk meg, hogy az a jel alaphfrekvenciájának egész számú többszöröse legyen. Az újramintavételezés gyakorlatban használt módja az „up-fir-down” eljárás, mely a jel egész arányú interpolációját (mintavételi frekvencia növelése), FIR aluláteresztő szűrését, végül pedig egész arányú decimálást (mintavételi frekvencia csökkentése) foglalja magában. A hatékony implementáció a három műveletet nem egymás után végzi el, hanem Ezt az eljárást használja a Matlab `resample` függvénye is.

Az újramintavételezés célja, hogy garantáljuk, hogy a jel egy periódusába pontosan egész számú (tipikusan 2^m) minta essen. Így az új mintavételi frekvenciát $f'_s = 2^m f_1$ -nek választjuk, ahol f_1 a jel alaphfrekvenciája. Ezután olyan ablakméretet választunk a további feldolgozáshoz, amivel koherens mintavételezést biztosítunk. Így jelentősen csökkenthetjük, ideális esetben pedig teljesen megszüntethetjük a spektrális szivárgást és a léckerítés-effektust.

2.3. Tranziens hangelemzés

Zenei hangok vizsgálatában kiemelt szerepe van a hang kezdeti tranziens szakaszának, ugyanis a hang indítása nagyban hat az észlelt hangminőségre. A tranziens elemzés kihívása természetesen a részhangok amplitúdóinak gyors változása.

Ha a kezdeti tranziens (más néven berezgést) állandósult állapotú szakasz követi, akkor az állandósult szakaszon pontos frekvenciamérést tudunk végezni, majd az így meghatározott alaphfrekvencia alapján újramintavételezzük, az előző szakaszban tárgyalt módon. Ha az újramintavételezésnél az $f'_s = R f_1$ arányt választjuk, akkor bármely $N = RM$ (M pozitív egész) mintát tartalmazó ablakhoz kiszámított diszkrét spektrumban pontos



9. ábra. Orgonahang indításának újramintavételezés utáni, alternatív ábrázolási módja.

san az M . frekvenciabinre esik az f_1 frekvencia. A (kvázi)periodikus jel további harmonikus komponensei pedig a kM . ($k = 2, 3, \dots$) binekre fognak esni. Mivel a mintavételezés koherens, a kiválasztott harmonikushoz tartozó frekvenciabin amplitúdóértékeit az egymást követő ablakokban leolvassva jó időfelbontással kapjuk meg a harmonikus komponensek amplitúdójának időbeli változását.

Ilyen módon készült elemzésre mutat példát a 9. ábra, mely a 8. ábrán spektrogrammal mutatott tranziens hangjel újramintavételezés utáni, alternatív ábrázolási módját szemlélteti. Az $R = 64$ aránnyal újramintavételezett jel $N = 512$ mintából álló, 97%-ban átlapoló ablakokra bontottuk, majd ezek Hann-ablakokkal számított spektrumából nyertük ki finom időfelbontással az első hét részhang amplitúdóit. Az ilyen módú ábrázolás előnye, hogy könnyebben leolvashatók az amplitúdók, mint a spektrogramról. Ugyanakkor a nem harmonikus komponensekről nem kapunk információt.

3. Hangszintézis

Ebben a szakaszban zenei hangok szintézisével foglalkozunk. Két alapvető szintézismódszert különböztethetünk meg: jelalapú és fizikai szintézist. Fizikai szintézis esetén a létrejövő hangminta egy fizikai modell kimenete, amely a hangszer működését fogalmazza meg fizikai törvények alapján, több-kevesebb egyszerűsítés alkalmazásával. A jelalapú szintézis ezzel szemben a hangszer rögzített hangmintáinak elemzéséből kinyert jellemzőkből kiindulva, esetleg az egyes paramétereket a felhasználó által változtatható téve, igyekszik a hangszer hangzását reprodukálni. Mindkét módszernek megvannak a maga előnyei és hátrányai: a fizikai modellel például vizsgálható különböző geometriai paraméterek hatása a megszólaló hangra, ugyanakkor a modell számításigénye jóval nagyobb lehet az egyszerűbb, jelalapú szintézisnél. Emellett a létrejövő hang jellemzőit a fizikai modell mindig korlátok közé szorítja, tetszőleges hangszín nem lesz elérhető. Jelalapú szintézissel akár a fizikai korlátok figyelmen kívül hagyásával is változtathatók a megszólaló hang tulajdonságai. Arra viszont nem kapunk információt a jelalapú modellből, hogy a megszólaló hangot a fizikai paraméterek hogyan befolyásolják. A továbbiakban csak jelalapú szintézissel foglalkozunk, ezen belül is az additív szintézismodellel.

3.1. Additív szintézis

Az additív szintézismodellelben a jel komponensek $k = 1, 2, \dots, K$ összegeként tekintjük. Először éljünk azzal az egyszerű feltételezéssel, hogy az egyes komponensek szinuszosak, adott (a frekvenciájukhoz képest lényegesen lassabban változó) időfüggő $A_k(t)$ amplitúdóval, ekkor a $p(t)$ jel a következő formában írhatjuk fel:

$$p(t) = \sum_{k=1}^K A_k(t) \cos(2\pi f_k t + \phi_k), \quad (18)$$

ahol ϕ_k a k . komponens fázisa. Ezzel az egyszerű módszerrel is már sokféle hangszer hangját imitálhatjuk, legyen szó pengetett, vonós, idiofon vagy éppen fúvós hangszerekről. Az $A_k(t)$ időfüggvényeket például rögzített hangok analízisével határozhatjuk meg, hasonlóan az f_k frekvenciák egymáshoz mért arányaihoz. Ahhoz, hogy a szintetizált jel valamelyest életszerűbb legyen, additív zajt keverhetünk a jelhez, esetleg kicsiny véletlenszerű amplitúdómodulációt alkalmazhatunk.

3.2. Modulált jelek előállítása

A (18) módon definiált jel amplitúdóját modulálhatjuk a $p_{\text{mod}}(t)$ modulációs jellel:

$$p_{\text{AM}}(t) = [1 + d p_{\text{mod}}(t)] p(t), \quad (19)$$

ahol d az amplitúdómoduláció modulációs mélysége. Amplitúdómodulációval pl. a *tremolo* effektust modellezhetjük, mely a hangerő periodikus változtatását jelenti, modulációs jelként néhány Hz frekvenciájú harmonikus jelet használva.

Frekvenciamodulációval a jel komponenseinek frekvenciáját változtathatjuk az idő függvényében:

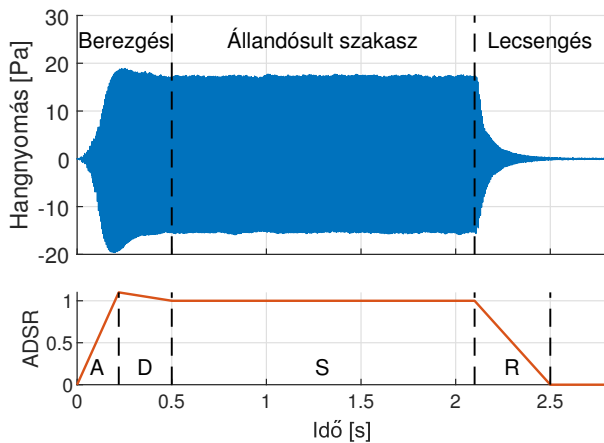
$$\begin{aligned} p_{\text{FM}}(t) &= \cos \left[2\pi \int_0^t (f_c + f_d p_{\text{mod}}(\tau)) d\tau + \phi \right] = \\ &= \cos \left[2\pi f_c t + 2\pi f_d \int_0^t p_{\text{mod}}(\tau) d\tau + \phi \right], \end{aligned} \quad (20)$$

ahol – a rádiózásban szokásos elnevezéssel – f_c a vi-vő (*carrier*) frekvencia, f_d pedig a modulációs mélység. Fontos megfigyelni, hogy a frekvenciamodulációnál a jel fázisát moduláljuk, amit a modulációs jel integrálásával érünk el.

A frekvenciamodulációval pl. a *vibrato* effektust utánozhatjuk, mely a frekvencia néhány Hz-es periodikus modulációját jelenti, tipikusan 1–2% modulációs mélységgel.

3.3. ADSR szintetizátor

Egyszerűbb digitális orgonákban gyakorta alkalmazzák az ADSR (*attack-decay-sustain-release*) szintézist, mely a jel négy szakaszra szakaszra osztja. Az *attack* szakaszban a jel amplitúdója felfut, tipikusan az állandósult értéknél valamivel nagyobb szintet érve el. A hang indításának második szakasza a *decay*, ahol az amplitúdó az előbbi túllövés után az állandósult szintre áll be. A tartott szakasz a *sustain*, végül pedig a lecsengést a *release* szakasz jelenti.



10. ábra. Orgonahang szakaszai és az ADSR burkológörbe.

A 10. ábrán egy valós orgonahang mért hangnyomás-jele látható, alatta pedig egy ADSR-burkológörbe, mely egyszerűen lineáris amplitúdóváltozást ír elő az egyes szakaszokban. Láthatjuk, hogy az egyszerű burkoló jellegre helyesen reprezentálja az orgonahang egyes szakaszait. Természetesen ennél szofisztikáltabb, pl. exponenciális felfutás és lecsengés is alkalmazható.

4. A felkészülést segítő kérdések

1. Koherensen mintavételezett, egységnyi amplitúdójú harmonikus jel diszkrét spektrumát számítjuk. Az ablak hossza $N = 2048$ minta, a jel periódusa $N_p = 256$ mintányi. Rajzold fel a DFT amplitúdóját!
2. Mely szimmetriatulajdonságok jellemzőek valós értékű jelek diszkrét spektrumára? Hogyan definiáljuk az egyoldali spektrumot?
3. Mit mond ki a Parseval-tétel? Írd le képlettel is!
4. Mit nevezünk koherens mintavételezésnek? Milyen hatásai vannak, ha egy jelet nem koherensen mintavételezünk?
5. Mit jelent a spektrális szivárgás? Mi okozza és hogyan csökkenthető a hatása?
6. Két jel szorzatának diszkrét spektruma milyen összefüggésben van az egyes jelek spektrumával?
7. Mit jelent az ekvivalens zaj-sávszélesség (ENBW)?
8. Miért van szükségünk ablakfüggvény alkalmazása esetén a spektrális amplitúdók korrekciójára?
9. Feladatunk egy ismeretlen frekvenciájú kváziperiodikus jel alaphérvenciához tartozó amplitúdójának pontos megmérése spektrális elemzés segítségével. Milyen ablakfüggvényt használnál a méréshez? Mik a számítás lépései?
10. Hogyan definiáljuk egy zajjal terhelt jel jel-zaj viszonyát?
11. Miben különbözik az időbeli és a spektrális átlagolás? Milyen esetekben használhatjuk ezeket?
12. Mit jelent a spektrum burkolója és a formánsszerkezet?

13. Hogyan definiáljuk a spektrális súlypontot?

14. Mit értünk a jel spektrogramja alatt? Mit jelent a spektrogram frekvencia- és időfelbontása közti határozatlansági reláció? Hogyan növelhető a spektrogram látszólagos időfelbontása?

15. Egy periodikus jelet $f_s = 48$ kHz frekvencián történő mintavételezése után az alaphérvenciát $f_1 = 263$ Hz-nek becsültük. Ezután a jelet újramintavételeztük $f'_s = 128f_1$ frekvenciával. Az $N = 512$ mintát tartalmazó ablakban melyik binnekhez fognak tartozni az f_1 , $2f_1$ és $3f_1$ frekvenciák? (Mely megadott adatok feleslegesek a kérdés megválaszolásához?)

16. Mi a különbség a jelalapú és a fizikai alapú hangszintézis módszerek között?

17. Mit értünk additív szintézis alatt?

18. Mit jelent a *tremolo* és *vibrato* effektus? Milyen típusú modulációval modellezhetjük ezeket?

19. Miért van szükség a (20) képletben a modulációs jel integrálására? Miért nem lesz helyes a naiv

$$p_{FM} = \cos[2\pi(f_c + f_d p_{mod}(t))t + \phi]$$

módon előállított modulált jel? (Vizsgáld meg, hogy a fenti képletben adott p_{mod} érték mekkora fázisváltozást okoz pl. $t = 1$ illetve 100 s esetén.)

20. Mit jelent az ADSR szintézis? Rajzold fel az ADSR burkológörbét!

5. Mérési feladatok

Az egyes pontokhoz tartozó feladatokat érdemes külön szkriptek formájában megoldani. Így könnyebb visszatérni az egyes elemzésekhez pl. a hangszintézis feladatoknál.

5.1. Hanganalízis

1. Ablakozás vizsgálata

Állítsd elő $f_s = 48$ kHz mintavételi frekvenciával egy egységnyi amplitúdójú, $f = 1$ kHz frekvenciájú harmonikus jel $T = 1$ s hosszúságú regisztrátumát.

Vége a jelből különböző hosszúságú szakaszokat és számítsd ki a jel spektrumát koherens illetve nem koherens mintavételezés esetén. Hasonlítsd össze a spektrumokat. Ezután végezd el a számítás Hann, majd Flat top ablakfüggvények alkalmazásával is. Mit tapasztalsz?

Keverj a jelhez Gauss eloszlású fehér zajt úgy, hogy a (várható) jel-zaj viszony 20 dB legyen. Számítsd ki a jel spektrumát időátlagolással illetve spektrális átlagolással. (A spektrális átlag számításához használhatod a mellékelt `spec_aver` függvényt.) Vizsgáld meg a különböző átlagolások hatását a hasznos jel illetve a zaj spektrális összetevőire az átlagolt regisztrátumok számának növelésével.

2. Kalibrációs jel feldolgozása.

Töltsd be a `calibration.wav` hangfájlt, mely egy pisztonfon kalibrátor felvett jelét tartalmazza. A kalibrátor 114 dB SPL szintű, 1 kHz frekvenciájú harmonikus jelet állít elő, melyet ± 5 V tartományú A/D konverterrel digitalizáltunk. (A `wav` fájlban a ± 5 V fizikai itartomány így a $[-1, +1]$ dimenziótlantartománnyá képződik le.) Határozd meg minél pontosabban a vett jel tényleges frekvenciáját, illetve a mérőmikrofon érzékenységét!

Használd az IEEE 1057 szabvány szerinti harmonikus illesztés eljárást, melyet a `harmonic_fit` függvény implementál. Az illesztés előtt érdemes a felvételtől a kisfrekvenciás zavarokat felüláteresztő szűrővel kiszűrni, majd az illesztést az állandósult állapotú szakasz rövid szegmensein elvégezni.

3. Orgonahang elemzése.

Töltsd be a `organ_pipe.wav` hangfájlt, mely egy orgona ajaksíp közelítőben felvett hangját tartalmazza. Alakítsd a mintákat hangnyomássá az előbbi feladatban a kalibrációs jel feldolgozásával meghatározott érzékenység felhasználásával. Válaszd ki az orgonahang stacionárius (állandósult állapotú) szakaszát és a komplex exponenciális illesztés eljárásával mérd meg a hang f_1 alapfrekvenciáját. Ezután mintavételezd újra a jelet az $f'_s = 64f_1$ új mintavételi frekvenciával. A további feldolgozáshoz az újramintavételezett jelet használd.

Ábrázold az állandósult szakasz spektrumát spektrális átlagolással a spektrum burkológörbéjével együtt. Ehhez gondold meg, milyen hosszú ablakokat érdemes venni a jelből. Számítsd ki az ekvivalens szintet az első néhány harmonikus részhang teljesítményének összegzésével. Számítsd ki a spektrális súlypont értékét.

Vágd ki a hang tranzienst szakaszát és átlapoló időablakok használatával számítsd ki, majd jelenítsd meg a harmonikus részhangok időfüggő amplitúdóit. Milyen ablakfüggvényt érdemes használni?

5.2. Hangszintézis

1. Egyszerű additív szintézis

Készíts additív szintézissel harmonikus komponensekből álló jelet az $f_1 = 200$ Hz alapfrekvenciával. A komponensek amplitúdói legyenek frekvencia növekedtével egyre kisebbek.

Alkalmazd az ADSR-burkológörbét a jel amplitúdója időbeli futásának beállításához.

Kis mértékben változtasd meg az egyes komponensek frekvenciáját. (Pl.: 201, 403, 605 Hz ...) Hallgasd meg az így keletkezett jelet! Mit tapasztalsz és mi a jelenség oka?

2. Modulált jelek előállítása

Készíts $f_1 = 330$ Hz frekvenciájú harmonikus alapjelet. A *tremolo* illetve *vibrato* effektusok szimulálásához készítsd el az alapjel amplitúdó- vagy frekvenciamodulált változatát. Hallgasd meg a jeleket, illetve a frekvenciamodulált jel spektrogramján ellenőrizd, hogy a várt eredményt kaptad-e.

3. Vibrafon hangszintézise

Töltsd be a `vibrafon.wav` hangfelvételt és vizsgáld meg a jelalakot. Állapítsd meg, hogy mely frekvenciák dominánsak a hangjelben! Egymást követő időablakok elemzésével számítsd ki a jel fő komponenseinek időfüggő amplitúdóit. Ezután állíts elő szintetizált hangjelet, melyben azonos frekvenciájú komponenseket adsz össze hasonló időbeli lefutással. Hasonlítsd össze az eredeti és a szintetizált jeleket!

A. Diszkrét Fourier-transzformáció

A.1. Téglányösszeg-közelítés

Ezt a szakaszt [1] alapján készítettük. Alkalmazzuk a (2) formulát a $[0, T]$ tartójú jelre, az $\omega = 2\pi f$ helyettesítéssel. Ekkor az integrált elég a $0 \leq t \leq T$ szakaszon kiértékelni, melyet téglányösszeggel, vagyis az integrandus T_s mintavételi időközönkénti kiértékelésével és összegzésével közelítünk:

$$\hat{p}(f) = \int_0^T p(t) e^{-2\pi j f t} dt \approx \sum_{n=0}^{N-1} T_s p(nT_s) e^{-2\pi j f nT_s}. \quad (21)$$

Az összegzést kiértékelhetjük tetszőleges f frekvencián, azonban a szükséges Δf frekvenciafelbontást a frekvenciatartománybeli mintavételi tétel is megadja: ha egy jel csak T időtartamon belül különbözik zérustól, akkor Fourier-transzformáltja hibátlanul helyreállítható a spektrumát $\Delta f \leq 1/T$ távolságokban mintavételezve. A $\Delta f < 1/T$ választás feleslegesen redundáns adatokat eredményezne, így célszerűen a $\Delta f = 1/T$ felbontást választjuk.

A mintavételezési fizikai időközt T_s reprezentálja, ezt a digitális jelfeldolgozásban azonban egységnyiinek szokás tekinteni, mellyel a $[0, f_s]$ frekvenciatartományt a $[0, 1]$ dimenziótlantartományra skálázzuk át. Így T_s elhagyásával és a $p[n] = p(nT_s)$, $\hat{p}[k] = \hat{p}(k\Delta f)$ index jelöléssel kapjuk:

$$\hat{p}[k] = \sum_{n=0}^{N-1} p[n] e^{-2\pi j \frac{kn}{N}} \quad k = 0, 1, \dots, N-1. \quad (22)$$

A.2. Periodikus kiterjesztés

A T periódusidejű $p(t) = p(t - mT)$, $m \in \mathbb{Z}$ periodikus jel Fourier-sorfejtését a C_k komplex együtthatókkal a

$$p(t) = \sum_{k=-\infty}^{\infty} C_k e^{2\pi j \frac{k}{T} t} \quad (23)$$

módon definiáljuk, ahol

$$C_k = \frac{1}{T} \int_0^T p(t) e^{-2\pi j \frac{k}{T} t} dt \approx \frac{1}{NT_s} \sum_{n=0}^{N-1} T_s p(nT_s) e^{-2\pi j \frac{nT_s k}{T}}. \quad (24)$$

A fentihez hasonlóan a téglányösszeg-közelítést használtuk az integrál kiértékelésére. T_s -sel egyszerűsítve azonnal láthatjuk, hogy a C_k együtthatók csak az $1/N$ szorzótényezőben különböznek a (3) összefüggés \hat{p}_k együtthatóitól. Mivel ez csak konstans szorzót jelent, az $1/N$ tényezőt a (24) helyett alkalmazhatjuk a (23)-ban is.

Kérdés még, hogy a (23) sorban hány darab együtthatót kell figyelembe vennünk. Ha feltételezzük, hogy a T_s mintavételi időközt a mintavételi tétel betartásával választottuk meg, akkor láthatjuk, hogy legfeljebb a $|k| < N/2$ együtthatók lesznek zérustól különböző értékűek.⁶

A.3. Diszkrét idejű Fourier-transzformáció

Érdemes még megemlíteni a diszkrét idejű Fourier-transzformációt (*Discrete-time Fourier transform*, DTFT), melyet a fentiek alapján legegyszerűbben a (21) jobb oldalán szereplő összeg-közelítéssel vezethetünk be. A mintavételi időközt ismét egységnyiinek véve bevezethetjük a $-\pi \leq \vartheta \leq \pi$ dimenziótlan körfrekvenciát, amivel a következő összefüggést kapjuk:

$$\hat{p}(\vartheta) = \sum_{n=0}^{N-1} p[n]e^{-j\vartheta n}. \quad (25)$$

A $\hat{p}(\vartheta)$ komplex értékű jel a $p[n]$ mintavett jel diszkrét idejű Fourier-transzformáltja. A dimenziótlan körfrekvencia kapcsolata a valós frekvenciával: $\vartheta = 2\pi f/f_s$. Hasonlóan a DFT-hez, a DTFT is periodikus a ϑ változóban, 2π periódussal.

A DTFT-ben a ϑ változót nem diszkrétizáljuk, tetszőleges értéket vehet fel, mellyel egy diszkrét jel spektruma nagyobb felbontással számítható ki. Ezt egyes szűrőtervezési algoritmusok használják ki, fent pedig a 2. ábra nagy felbontású spektrumainak számításához használtuk fel. Figyeljük meg, hogy a (25) összefüggésben a $p[n]$ minták az $e^{-j\vartheta}$ komplex értékű változó polinomegyütthatói. Emiatt a $\hat{p}(\vartheta)$ jelölés mellett a $\hat{p}(e^{j\vartheta})$ jelölés is használatos a diszkrét idejű Fourier-transzformáltra.

B. Frekvenciabecslő módszerek

B.1. A spektrum interpolációja

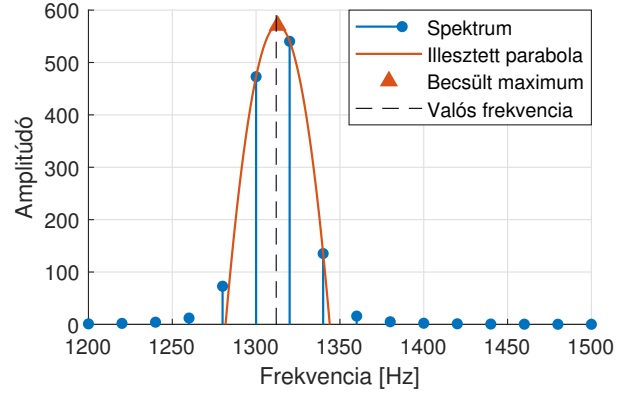
Egyszerű módszer a spektrum alapú frekvenciabecslés pontosítására az amplitúdóspektrum interpolációja. A rövid időszeléből, tipikusan Hann-ablakkal számított diszkrét amplitúdóspektrum maximumát megkeresve, a csúcshoz és a két szomszédos binhez parabolát illesztünk. Az illesztett parabola maximuma jó becslést ad a tényleges frekvenciára, a maximális érték pedig a jel amplitúdóját (az ablakozás hatását még magában foglalva) becsli. A parabolaillesztésre mutat példát a 11. ábra, ahol a módszer alkalmazásával a $\Delta f = 20$ Hz felbontású spektrum alapján kisebb, mint 1 Hz hibával becsüljük a harmonikus jel frekvenciáját. A parabolaillesztés módszerét egyszerűsége miatt más, szofisztikáltabb frekvenciabecslő módszerek is felhasználják kezdeti becsléshez [2, 3].

B.2. Harmonikus illesztés

Az IEEE 1057 [4] szabvány algoritmusokat definiál ismert és ismeretlen frekvenciájú harmonikus illesztésére mért, digitalizált jelre. A jel p_n mintáinak modellje

$$p_n = A \cos(\omega_0 t_n) + B \sin(\omega_0 t_n) + C, \quad (26)$$

⁶Itt is igaz, hogy a (24)-ben alkalmazott közelítés az együtthatókat k -ban N -periodikussá teszi.



11. ábra. Frekvenciabecslés parabolaillesztéssel. $\Delta f = 20$ Hz, $f_0 = 1312.86$ Hz, $f_0 = 1312$ Hz.

ahol $t_n = nT_s$, A , B , és C pedig meghatározandó paraméterek. Emellett a valós mintákat természetesen additív zaj is terheli.

Első lépésben feltételezzük, hogy a valódi ω_0 frekvenciát ismerjük. Ekkor a p_n ($n = 1, 2, \dots, N$) mintákra a következő mátrixalakot írhatjuk fel:

$$\underbrace{\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix}}_{\mathbf{p}} = \underbrace{\begin{bmatrix} \cos(\omega_0 t_1) & \sin(\omega_0 t_1) & 1 \\ \cos(\omega_0 t_2) & \sin(\omega_0 t_2) & 1 \\ \vdots & \vdots & \vdots \\ \cos(\omega_0 t_N) & \sin(\omega_0 t_N) & 1 \end{bmatrix}}_{\mathbf{D}_0} \underbrace{\begin{bmatrix} A \\ B \\ C \end{bmatrix}}_{\mathbf{x}}, \quad (27)$$

Az illesztendő paraméterek \mathbf{x} vektorát a (27) erősen túlhatározott egyenletrendszer legkisebb négyzetek módszerével történő megoldásával kapjuk (\mathbf{D}_0^T -tal szorozva mindkét oldalt, láthatjuk):

$$\mathbf{x} = (\mathbf{D}_0^T \mathbf{D}_0)^{-1} \mathbf{D}_0^T \mathbf{p}. \quad (28)$$

Következő lépésként tegyük fel, hogy a jel ω_0 körfrekvenciája nem ismert pontosan, viszont rendelkezésre áll egy $\tilde{\omega}_0$ becslés, amit például a jel diszkrét spektrumából kaptunk meg. Ekkor iteratív módszerrel becsülhetjük a frekvenciát, a következő megfontolásokkal [5]. Legyen az i -edik iterációban becsült körfrekvencia, koszinusz és szinuszmódok jelölése rendre $\tilde{\omega}_i$, \tilde{A}_i és \tilde{B}_i . Az i -edik iterációhoz tartozó frekvenciahibát pedig jelölje $\Delta\omega_i = \omega - \tilde{\omega}_i$. Kihasználhatjuk, hogy $\Delta\omega_i T \ll 1$, amivel a következő közelítéseket alkalmazhatjuk (elsőfokú Taylor-sorfejtés):

$$\begin{aligned} \cos \omega_0 t_n &\approx \cos \tilde{\omega}_i t_n - t_n \Delta\omega_i \sin \tilde{\omega}_i t_n \\ \sin \omega_0 t_n &\approx \sin \tilde{\omega}_i t_n + t_n \Delta\omega_i \cos \tilde{\omega}_i t_n \end{aligned} \quad (29)$$

A (29)-t a 26 jelmodellbe helyettesítve:

$$\begin{aligned} p_n &\approx A \cos(\tilde{\omega}_i t_n) + B \sin(\tilde{\omega}_i t_n) + C \\ &\quad - A t_n \Delta\omega_i \sin(\tilde{\omega}_i t_n) + B t_n \Delta\omega_i \cos(\tilde{\omega}_i t_n). \end{aligned} \quad (30)$$

A paramétervektort kibővítve a $\Delta\omega_i$ paraméterrel:

$$\mathbf{x}_i^T = [A_i, B_i, C_i, \Delta\omega_i]^T, \quad (31)$$

A nemlineáris (30) egyenletet linearizálhatjuk, kihasználva, hogy $\Delta\omega_i \approx 0$. A és B helyébe az előző iterációból kapott \tilde{A}_{i-1} és \tilde{B}_{i-1} becsült értékeket helyettesítve a (27)-hez hasonló mátrixalakot kapunk:

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\tilde{\omega}_i t_1) & \sin(\tilde{\omega}_i t_1) & 1 & -\tilde{A}_{i-1} t_1 \sin(\tilde{\omega}_i t_1) + \tilde{B}_{i-1} t_1 \cos(\tilde{\omega}_i t_1) \\ \cos(\tilde{\omega}_i t_2) & \sin(\tilde{\omega}_i t_2) & 1 & -\tilde{A}_{i-1} t_2 \sin(\tilde{\omega}_i t_2) + \tilde{B}_{i-1} t_2 \cos(\tilde{\omega}_i t_2) \\ \vdots & \vdots & \vdots & \vdots \\ \cos(\tilde{\omega}_i t_N) & \sin(\tilde{\omega}_i t_N) & 1 & -\tilde{A}_{i-1} t_N \sin(\tilde{\omega}_i t_N) + \tilde{B}_{i-1} t_N \cos(\tilde{\omega}_i t_N) \end{bmatrix}}_{D_i} \underbrace{\begin{bmatrix} A_i \\ B_i \\ C_i \\ \Delta\omega_i \end{bmatrix}}_{\mathbf{x}_i} \quad (32)$$

Az iteratív algoritmus lépései ezzel a következőképp alakulnak. Egy kezdeti $\tilde{\omega}_0$ becsült körfrekvencia alapján a (27) segítségével meghatározzuk a $\tilde{A}_0, \tilde{B}_0, \tilde{C}_0$ becslőket. Ezután a (32) megoldásával megkapjuk a következő iteráció becslőit. A becsült körfrekvenciát frissítjük: $\tilde{\omega}_{i+1} := \tilde{\omega}_i + \Delta\omega_i$. Az utóbbi két lépést ismételjük amíg az \mathbf{x}_i paramétervektor konvergenciáját el nem érjük.

Az iteráció tipikusan gyorsan konvergál, ha a mérendő frekvencia valóban konstans és az additív zajon kívül nem terheli más zavar a jelet. Ha a frekvencia valamelyest ingadozik, rövidebb regisztrátumok sorozatán érdemes futtatni az algoritmust, a zavarok elnyomásához pedig az illesztés előtt sáváteresztő szűrőt használhatunk.

B.3. Komplex exponenciális illesztés

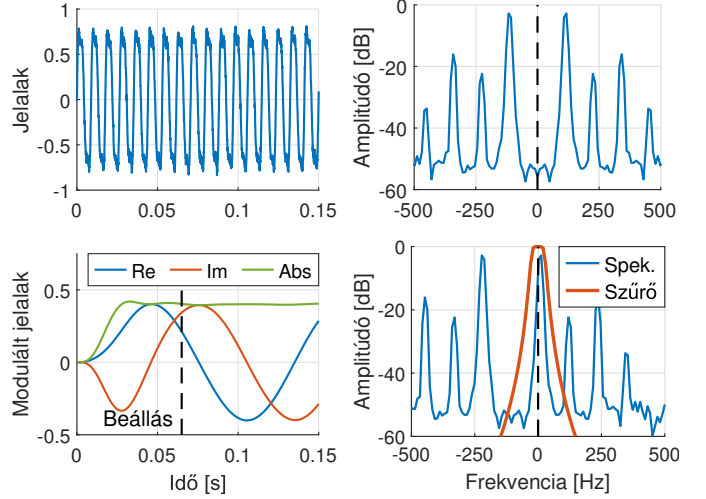
A komplex exponenciális illesztésen alapuló frekvencia-becslés a harmonikus illesztéshez hasonlóan akkor használható, ha a jel alapfrekvenciájáról rendelkezésünkre áll az \tilde{f}_1 becsült érték. A módszer működésének szemléltetéséhez a 12. ábrát használjuk fel.

A becsült \tilde{f}_1 frekvenciát használva az eredeti jelünk spektrumát egyoldali modulációval eltoljuk. A $p_m(t)$ modulált jel ennek hatására komplex értékű lesz:

$$p_m(t) = p(t)e^{-j2\pi\tilde{f}_1 t}. \quad (33)$$

A moduláció hatására, a Fourier-transzformáció eltolási tételének következményeként a 12. ábrán látható spektrum balra tolódik a becsült \tilde{f}_1 frekvenciával. Az eredeti $p(t)$ jelünk valós volt, melynek spektruma páros, ez azonban a modulált $p_m(t)$ jelre már nem teljesül. A modulált jel frekvenciatartománybeli képét a 12. ábra jobb alsó diagramja mutatja. A spektrum két maximumhelyét az $f = f_1 - \tilde{f}_1 \approx 0$ és $f = -f_1 - \tilde{f}_1 \approx -2\tilde{f}_1$ frekvenciákon találjuk. Utóbbi komponenst aluláteresztő szűrővel szűrhetjük ki a modulált jelből. Példánkban a szűrő vágási frekvenciáját $\tilde{f}_1/4$ -nek választottuk, az átviteli karakterisztikát a spektrum diagramján piros vonallal jelöltük. Érdemes megemlíteni, hogy az aluláteresztő szűréssel a jelet terhelő szélessávú zaj (pl. kvantálási zaj) teljesítménye is jelentősen csökken.

A 12. ábra bal alsó diagramján megfigyelhetjük, hogy a modulált szűrt jelünk nem csak DC komponenst tartalmaz. Bár a szűrő beállása után az amplitúdó nagyjából állandó, a valós és képzetes rész alakulásán lassú változások láthatóak. Mivel az f_1 frekvenciájú harmonikus jelet a \tilde{f}_1 frekvenciával toltuk el, a szűrés után $f_1 - \tilde{f}_1$ frekvenciával forgó fazort kapunk. Az ábrán azt is megfigyelhetjük, hogy a fázor a pozitív irányban forog, ami azt jelenti, hogy ebben az esetben $\tilde{f}_1 < f_1$.



12. ábra. Egyoldali moduláció és szűrés hatása. Felső sor: eredeti jel. Alsó sor: modulált jel. Bal oszlop: idő tartomány. Jobb oszlop: Frekvenciatartomány.

A $\Delta f = f_0 - \tilde{f}_0$ frekvenciakülönbséget időtartománybeli illesztéssel kapjuk meg. Tudjuk, hogy a p_f szűrt modulált jelünk mintái a

$$p_f[n] = p_f(nT_s) = Ae^{j(2\pi\Delta f nT_s + \varphi)}. \quad (34)$$

alakban írhatók fel, a zaj hatását figyelmen kívül hagyva. Vegyük az $n + 1$ -edik és a n -edik minta hányadosát! Figyeljük meg, hogy a hányados az n értékétől független:

$$\frac{p_f[n+1]}{p_f[n]} = e^{j2\pi\Delta f T_s} = \gamma. \quad (35)$$

Mivel a szűrt modulált jel mintáit ismerjük, γ értékét meghatározhatjuk. Ehhez egy erősen túlhatározott egyenletrendszert kell megoldanunk:

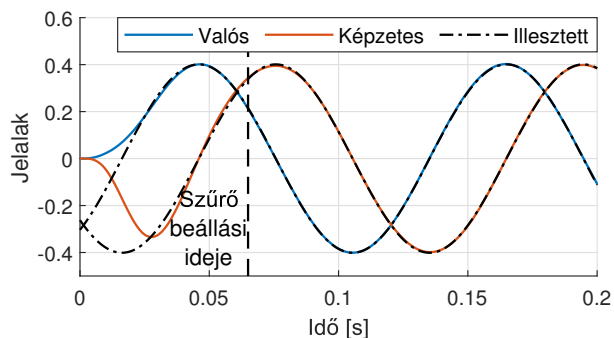
$$\begin{Bmatrix} p_f[N_0 + 1] \\ p_f[N_0 + 2] \\ \vdots \\ p_f[N] \end{Bmatrix} = \begin{Bmatrix} p_f[N_0 + 0] \\ p_f[N_0 + 1] \\ \vdots \\ p_f[N - 1] \end{Bmatrix} \gamma, \quad (36)$$

ahol N_0 az aluláteresztő szűrőnk tranziense utáni első minta sorszámát jelöli. A szűrő beállási idejét a 12. ábra bal alsó diagramján a szaggatott függőleges vonal mutatja.

Végül γ alapján kapjuk a Δf frekvenciakülönbséget

$$\Delta f = \frac{\arg\{\gamma\}}{2\pi\Delta t}, \quad (37)$$

illetve a becsült $\hat{f}_0 = \tilde{f}_0 + \Delta f$ frekvenciát. Δf meghatározása után a p_f jelre illeszthetjük a Δf frekvenciájú komplex exponenciális, újfent a legkisebb négyzetek módszerét alkalmazva. Ezt mutatja a 13. ábra.



13. ábra. Komplex exponenciális illesztés eredménye.

Hivatkozások

- [1] BME MIT Tanszéki munkaközösség. *Digitális jelfeldolgozás*. Tantárgyi jegyzet a BMEVIMM4084 (Di-

gitális jelfeldolgozás) tantárgyhoz. Műegyetemi Kiadó, 2006.

- [2] A. de Cheveigné és H. Kawahara. „YIN, a fundamental frequency estimator for speech and music”. *The Journal of the Acoustical Society of America* 111.4 (2002), 1917–1930. old. DOI: 10.1121/1.1458024.
- [3] O. Das, J. O. Smith és C. Chafe. „Improved real-time monophonic pitch tracking with the extended complex Kalman filter”. 68.1/2 (2020), 78–86. old. DOI: 10.17743/jaes.2019.0053.
- [4] *IEEE Standard for Digitizing Waveform Recorders*. IEEE Std. 1057. 1994.
- [5] P. Handel. „Properties of the IEEE-STD-1057 four-parameter sine wave fit algorithm”. *IEEE Transactions on Instrumentation and Measurement* 49.6 (2000), 1189–1193. old. DOI: 10.1109/19.893254.