

**A hallás természetének és a hang különböző átalakításainak
vizsgálata**

Készítette: Bánhalmi András

V. matematika-fizika

Témavezetők:

Dr. Papp Katalin

Dr. Gingl Zoltán

Szegedi Tudományegyetem, SZTE TTK Kísérleti Fizikai Tanszék

Szeged, 2002

Tartalomjegyzék

1. Bevezetés	2
2. Motiváció	4
3. Bevezető fogalmak	5
3.1. Fogalmak a számítástechnikából	5
3.1.1. Digitális hangrögzítés	5
3.1.2. Hanghullám fájl formátuma	8
3.2. Fogalmak a fizikából	9
3.2.1. Harmonikus rezgőmozgás és gerjesztés	9
3.2.2. Csillapított harmonikus rezgőmozgás	11
3.2.3. Csillapított harmonikus rezgőmozgás és gerjesztés; a rezonátor	14
3.3. A rezonátor differenciálegyenletének numerikus megoldásáról	18
3.4. A rezonátorok kapcsolata a digitális szűrőkkel	24
4. Spektrogram felvétele rezonátorok segítségével	
<i>Hiba! A könyvjelző nem létezik.</i>	
4.1. Spektrogram	25
4.2. Spektrogram előállítás rezonátorral	25
4.3. Hangolás	28
4.4. Vonal kiszélesedés	29
4.5. Időbeli megnyúlás	30
4.6. Kapcsolat a hallással	30
5. Spektrogram feldolgozása visszaalakításhoz	32
5.1. Vonal kiszélesedés megszüntetése	32
5.2. Előkészítés frekvenciakövetéshez: vonal egyértelműsítés	34
6. Spektrogram visszaalakítás frekvenciakövetéses simítással	36
6.1. Az algoritmus és magyarázata	37
7. Felhasználás	41
7.1. Hangnyújtás	41
7.2. Hangtömörítés	42
7.3. Különböző effektusok	43
7.4. Beszéd előállítás, felolvasó program	44
8. Irodalom	48
Melléklet:	49

1. Bevezetés

A dolgozatban egyrészt a *hallás természetével* kapcsolatos fogalmak (mint például a hangmagasság, a hangszín, a hangerő) megértését, elsajátítását szeretnénk elősegíteni. Ezen cél érdekében egy egyszerű fizikai jelenségből indulunk ki, ami nem más, mint a *rezonancia*. Fizika órán legtöbbször előkerül a fogalom, de negatív értelemben: leomló hidakkal és épületekkel, esetleg a földrengésekkel hozható kapcsolatba. A rezonanciajelenséget leíró differenciálegyenletet a dolgozatban *építő jelleggel* használjuk fel. A fényt prizma segítségével könnyen felbonthatjuk különböző frekvenciájú harmonikus összetevőire, de a hanggal más a helyzet. Ebben siet segítségünkre a rezonancia jelenség és a számítógép. Segítségükkel a *hangot* is harmonikus összetevőkre bonthatjuk és *vizualizálhatjuk*, amit *spektrogramnak* fogunk nevezni (és mindezt anélkül tesszük, hogy a Fourier-transzformációt vagy a waveleteket túl sokszor szóba kellene hoznunk). Mivel számítógéppel dolgozunk, elemeznünk kell a differenciálegyenlet numerikus megoldásának a tulajdonságait is. A spektrogram hanggá visszaalakítása is könnyen értelmezhető módon történik: kicsit leegyszerűsítve arról van szó, hogy a szinuszfüggvények összeadásával kapjuk meg a hanghullám állományt (és ezzel érintjük a *szuperpozíció* fogalmát is). Az igazán izgalmas dolog az, amit a spektrogramon végezhetünk. Megnézzük, hogy hogyan lehet hangot megnyújtani időben anélkül, hogy a hangmagasság megváltozna. Megvizsgálhatjuk a fordítottját is: hogyan változtassuk meg a hangmagasságot, hogy a sebesség ne változzon. Hogyan adhatunk a hanghoz visszhangot, mit hallunk, ha a frekvenciatartományból elhagyunk, vagy azt valahogy transzformáljuk. Megvizsgálunk egy egyszerű hangtömörítési lehetőséget. Legvégül egy komolyabb felhasználási lehetőségre is kitérünk, a gépi beszéd előállításra. Miközben ezeken végigmegyünk például egy fizika fakultáció keretében, a

tanulóban egy letisztult kép fog kialakulni a harmonikus komponensek szerepéről, a hallás leegyszerűsített működéséről, és azt is megtapasztalja, hogy az egyes tudományterületek, mint a fizika, a matematika, a számítástechnika, a nyelvészet (a pszichológia, biológia, stb.) szorosan összefonódnak egymással. Ezen kívül a dolgozat a további, felsőfokú tanulmányokhoz is segítséget nyújt azáltal, hogy kiemeli a harmonikus komponensekre való felbontás hasznosságát és segíti sok fontos fogalom megértését.

A dolgozat felépítése a következő: A bevezető és a motiváció utáni fejezetben áttekintést adunk a fogalmakról, amiket később felhasználunk. Először a számítástechnika ide vágó alapfogalmait tekintjük át, majd a szükséges fizikai ismereteket foglaljuk össze, miközben eljutunk a rezonátor fogalmáig, és megvizsgáljuk annak tulajdonságait. Szintén ebben a fejezetben kidolgozunk egy kiválóan alkalmazható és gyors numerikus megoldást a rezonátor differenciálegyenletének megoldásához. Ezen numerikus megoldás stabilitásának a feltételeit, és egyéb fontos jellemzőit is bizonyítani fogjuk. A harmadik és a negyedik fejezetben bemutatjuk azt a – rezonátorokat használó – algoritmust, melyet a spektrogram felvételére használunk fel, és megvizsgáljuk annak tulajdonságait. Az utolsó előtti fejezetben bemutatjuk azt az algoritmust, amit a visszaalakításhoz készítettünk el. Ezután több példán keresztül bemutatjuk a felhasználási lehetőségeket, amiket a dolgozathoz mellékel programokkal és hanganyagokkal fogunk demonstrálni.

2. Motiváció

A fizika órákon a középiskolás tanulók figyelmének felkeltésére ma már informatikai eszközök is rendelkezésre állnak. Habár nem pótolhatatlan, de sokszor kézenfekvő lehetőségek rejlenek a számítástechnikában. Például modellezzük gáZRészecskéket biliárdgolyóval, és kihasználhatjuk a számítógép nagy számítási kapacitását eloszlásfüggvények kirajzolásához, miközben csak az ütközésekre vonatkozó fizikát használjuk ki. Modellezzük bolygómozgást, csak Newton mozgástörvényeit alkalmazva, és megvizsgálhatjuk a pályákat. Mindezekhez hasonlít a mi dolgozatunk témája is. A csillapított és gerjesztett egy dimenziós rezgőmozgást modellezzük, de *nem önmagáért*, hanem mert ezáltal a hanghullámot felbonthatjuk szinuszos hullámokra. Mindezzel azt a feltételezésünket próbáljuk alátámasztani, hogy a fül is ilyen módon működik. Minderről úgy bizonyosodhatunk meg, hogy ha a megkeresett szinuszos komponenseket újra összeadjuk, akkor az eredeti hangot halljuk, illetve bizonyos transzformációk a várt eredményt adják. Motivációként érdemes bemutatni azokat a szabadon felhasználható programokat, amelyeket a TDK dolgozathoz írtunk, és a dolgozat Mellékletében megtalálható. Az egyik programmal felgyorsíthatunk illetve lelassíthatunk egy hanganyagot, megnövelhetjük, lecsökkenthetjük a hangmagasságot, változtathatunk a hangszínen, visszhangot keverhetünk a hanghoz, mindezeket valós időben. A másik program segítségével szöveget olvastathatunk fel, azaz ez egy beszéd előállító program. Harmadik programunk segítségével vizualizálhatjuk a hangot, azaz spektrogramot hozunk létre. Ha a programokkal felkeltettük a tanulók érdeklődését, akkor talán erősebben fognak koncentrálni a kicsit „száraz”, inkább elméleti jellegű (de nem túl bonyolult) részeknél, míg a végén esetleg már saját ötleteket valósítanak meg, érdekes hangátalakító eljárásokat találnak ki.

3. Bevezető fogalmak

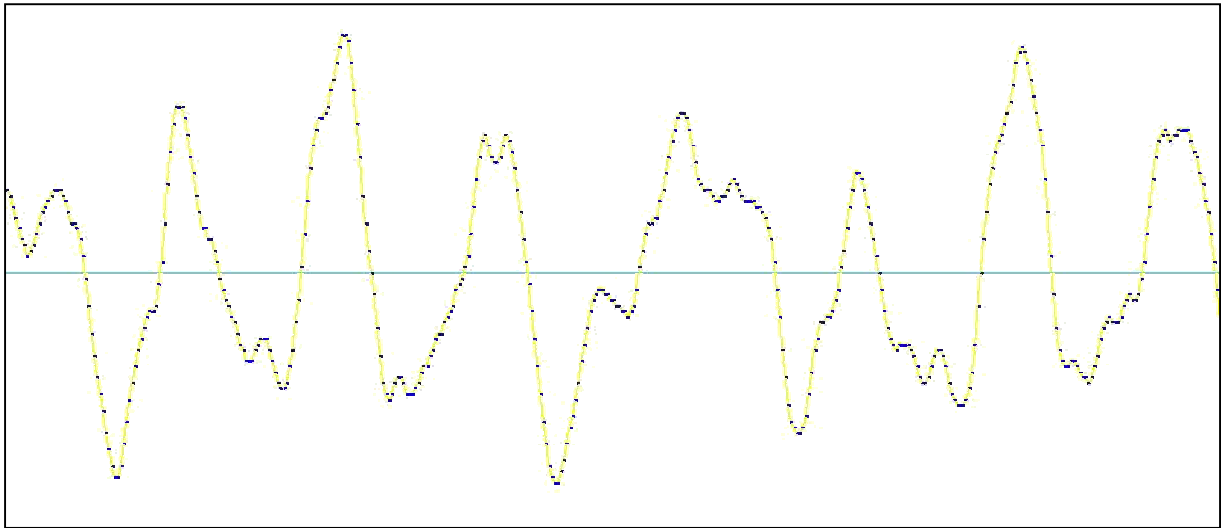
3.1. Fogalmak a számítástechnikából

Ahhoz, hogy megértsük, hogy a hang és a számítógép hogyan viszonyulnak egymáshoz, az alább leírt fogalmakkal mindenképpen tisztában kell lenni.

3.1.1. Digitális hangrögzítés

Először is foglalkozunk röviden a digitális hangrögzítéssel. Mikrofont helyezünk el a tér egy pontjába. Tegyük fel, hogy ez egy membrán mikrofon. A membrán a hang, azaz a légnomás változás hatására önmaga is rezgésbe jön. A membrán egy tekercsben van elhelyezve, így a tekercs vezetékében áramot indukál. Így létrejön egy *áramerősség ingadozás*, ami nagyjából a *nyomás ingadozásával* arányos. Ezt az áramerősség függvényt, – ami egy folytonos függvény – *analóg jelnek* nevezzük. Digitális számítógépen ezeket a folytonos jeleket nem tudjuk közvetlenül eltárolni, át kell alakítani véges sok szám sorozatára. Ezt végzi el az *analóg/digitális átalakító*. Tulajdonképpen arról van szó, a jelet bizonyos időpontokban (például másodpercenként 44100-szor) megmérjük. Ezt nevezzük *mintavételezésnek*. Ekkor már véges sok számunk van, de még ezek értéke végtelen sok féle lehet. Ekkor legegyszerűbb esetben azt tehetjük, hogy kijelölünk egy maximális értéket, és beskálázzuk a pozitív és negatív tartományt a maximális értékig egyenletesen. A digitális számítógépek felépítése miatt az osztópontok szokásos száma 256 (8 bit) vagy 65536 (16 bit). Az előbb leírt eljárást *kvantálásnak* nevezzük. További jellemzője a digitális hangmintáknak a *csatornák száma*. Egy csatorna esetén mono, két csatorna esetén stereo, és elterjedtek már a több csatornás hangminták is.

A következő ábra egy digitális hangmintát mutat (vízszintes tengely az idő, a függőleges az amplitúdó):



1. ábra: digitális hanghullám

Még két jelenséget szokás a mintavételezéssel és a kvantálással kapcsolatban megemlíteni: Shannon tételét (mintavételezési tétel), és a kvantálási zajt.

A *mintavételezési tétel* szerint ha egy jelsorozat *mintavételezési frekvenciája* $2f$ (azaz egy másodperc alatt $2f$ mintát veszünk), akkor a jelsorozat által egyértelműen ábrázolható *maximális* frekvencia f . Fordítva: ha a figyelt frekvenciatartomány felső határa f , akkor a mintavételezési frekvencia legalább $2f$ legyen. A mintavételezési tételre a leghétköznapibb példa a kocsikerék látszólagos mozgása. Órán szemléltethetjük is a tételt a táblán. Vegyünk például egy szélsőséges esetet: a mintavételezési frekvencia 10 Hz , és egy 10 Hz -es szinuszt digitalizálunk. A táblára felrajzolt pontok amplitúdója meg fog egyezni, tehát a 10 Hz digitálisan a 0 Hz -nek fog megfelelni (általában is elmondható, hogy tükrözni kell f -re: 8 Hz -nek 2 Hz fog megfelelni). Mindezek miatt mintavételezés előtt az analóg jelet át kell engedni egy aluláteresztő szűrőn.

A *kvantálási zaj* a digitális értékeknek a valódi értékektől való eltéréséből ered. Ez a különbség egy véletlenszerű szám, ami fehér zajt eredményez. Minél kisebb a kvantálás értéke, annál intenzívebb a zaj. Ezzel ellentétesen befolyásoló tényező az, hogy ha a kvantálás értéke több, akkor a hanganyag tárolásához több tárterület szükséges.

Mivel a hallható frekvenciatartomány az embereknél kb. 20 kHz-ig terjed, a szokásos mintavételezési frekvencia 44100 Hz. A CD minőséghez tartozó kvantálás 16 bites. A 44100 érték pontosabb okairól és még sok másról [9]-ben olvashatunk.

A digitális hangminta visszaalakításakor a digitális értékeket feszültség amplitúdóvá alakítjuk és a szintek közötti ugrásokat kisimítjuk egy digitális/analóg átalakító segítségével. Az előbb említett f -nél nagyobb frekvenciákat aluláteresztő szűrővel tüntetjük el, majd erősítőkön keresztül a most már analóg jelet a hangszórókba vezetjük.

A valós idejű hangátalakítás fogalmáról még nem szóltunk. Ez azt jelenti, hogy ahogyan beérkezik egy minta a mintavételezés során, akkor minimális késéssel készül el a kimenő jel. Nem valós idejűek azok az algoritmusok, amelyeknek szükségük van időben későbbi hangmintára a jel feldolgozásához. A dolgozat algoritmusai valós idejűek.

Két egyszerű hangátalakítási lehetőséget említünk még meg: a túlmintavételezést és az alulmintavételezést. Alulmintavételezés esetén csak annyit teszünk, hogy például elhagyunk minden második mintát a hangállományból, a mintavételi frekvencia értékét viszont nem változtatjuk meg. Ez az eljárás természetesen azt eredményezi, hogy megnő a hangmagasság, miközben a lejátszás felgyorsul. Olyan ez, mint amikor nagy bakelitlemezt játszunk le a kislemez nagyobb fordulatszámán a régi lemezjátszón. A túlmintavételezéskor pedig hozzá rakunk a hangfájlhoz hangmintákat úgy, hogy például két hangminta közé berakjuk az átlagukat (persze jobb eredményt érhetünk el, ha nem lineáris interpolációval dolgozunk). Ha

a mintavételi frekvencia értékét nem változtatjuk meg (tehát ugyanannyi minta/sec sebességgel játszuk le a hangot, mint az eredetit), akkor mélyebb, de lassabb hangot kapunk.

3.1.2. Hanghullám fájl formátuma

Csak a PCM-WAV formátummal foglalkozunk. Az ilyen fájl két részből áll: egy un. fejrészből és egy ezt követő adatrészből. A fejrész néhány bájtos, a mintavételezési frekvenciát, a kvantálás bitmélységét és a csatornák számát tartalmazza. A pontos formájára nem térünk ki, szükség esetén az Interneten bármikor megkereshető. Az adatrész tartalmazza az egyes mintákat (egy vagy több bájton ábrázolva), időbeli sorrendben. Ha több csatorna van, akkor az egyes csatornák egy időponthoz tartozó mintái egymás után vannak.

3.2. Fogalmak a fizikából

Ebben az alfejezetben bevezetjük a rezonátor fogalmát, ami az összes további vizsgálódásunk alapja lesz. A rezonátor tulajdonképpen egy fizikai rezgőrendszer, aminek megvannak mind a mechanikában, mind az elektronikában a megfelelő példái. Ezeket a rezgőrendszereket differenciálegyenletek segítségével tudjuk leírni, és ezen differenciálegyenletek kezdeti érték feladatának megoldása leírja a rendszer állapotát az idő függvényében. Egy olyan rendszerig jutunk el ebben a fejezetben, amit ha egy adott függvénnyel „gerjesztünk” (ez a függvény a hangminta lesz), akkor ennek a gerjesztő függvénynek egy előre adott frekvenciához tartozó komponenséről szerezhethetünk információkat: jelen van-e és milyen energiával van jelen egy adott frekvenciájú harmonikus rezgés a jelben.

3.2.1. Harmonikus rezgőmozgás és gerjesztés

A harmonikus rezgőmozgás példajaként a rugóra akasztott test esetét szokták a mechanikában felhozni, abban az ideális esetben, amikor nincsen közegellenállás. Ha a testet kitérítjük az egyensúlyi helyzetéből, akkor harmonikus rezgőmozgást végez. Az elektronikában pedig egy ideális tekercsből és kondenzátorból álló, ellenállás nélküli áramkör erre a példa, ez az ún. rezgőkör. A példákat közösen harmonikus oszcillátornak nevezzük. Az ilyen rendszereket a következő differenciálegyenlettel írhatjuk le [1]:

$$\ddot{x} + \omega^2 x = 0,$$

ahol $x = x(t)$ az időtől függő valamilyen mennyiség (például a rugóra akasztott test kitérése az idő függvényében), az ω pedig egy konstans mennyiség (a körfrekvencia).

Közismert, hogy ennek a homogén állandó együtthatós lineáris differenciálegyenletnek az általános megoldása

$$x(t) = A \cos(\omega t) + B \sin(\omega t)$$

alakú, ahol A és B együtthatók a kezdeti értékekből számíthatók ki [2]. Az ω mennyiséget a rendszer sajátfrekvenciájának nevezzük (a megoldás egy ω körfrekvenciájú harmonikus rezgés).

Kicsit maradva még az ideális rugós példánál, képzeljük el a rugóra akasztott test mozgását, ha a rugó eddig rögzített végét a kezünkbe vesszük, és a rugó kezünkbe tartott végét szinuszosan mozgatójuk fel és le, azaz azon egyenes mentén, amelyen a rugóra akasztott test mozog. Azt tapasztalhatjuk, hogy a test mozgása felveszi a kezünk mozgásának a frekvenciáját, és a rezgésének az amplitúdója erősen függ attól, hogy a kezünk milyen frekvenciával rezeg. Ha a kezünket azzal a frekvenciával mozgatójuk, ami megegyezik a rendszer sajátfrekvenciájával (ω), akkor a test mozgásának az amplitúdója nagyon nagyvá válik. Ezt a berezgetést gerjesztésnek, a jelenséget rezonanciának, a kényszererő hatása alatt rezgő rendszer mozgását kényszerrezgésnek nevezzük. A rendszert leíró differenciálegyenlet a következő [1]:

$$\ddot{x} + \omega^2 x = f(t)$$

ami azt jelenti, hogy a testre nem csak a kitérés -1 -szeresével arányos erő hat, hanem egy $f(t)$ gerjesztő vagy kényszererő is. Ezt a differenciálegyenletet később, általánosabban, csillapítás figyelembe vételével fogjuk megvizsgálni.

3.2.2. Csillapított harmonikus rezgőmozgás

Az ideális, csillapítás nélküli eset után rátérünk a reálisabb esetre, mikor is a rendszer energiája csökken, mert például a rugóra akasztott testnél, közegellenállás, súrlódás lép fel. Ez az erő fékezi, „csillapítja” a rezgést, ami ezért egy idő után nagyon kicsivé válik. Az, hogy milyen gyorsan csökken a rendszer energiája, az a súrlódás mértékétől függ. Például egy rugóra akasztott test energiája levegőben sokkal lassabban csökken, mint ha a rendszert vízbe rakjuk. A csillapítást úgy fejezzük ki, hogy azt mondjuk: legyen a csillapodást kifejező (súrlódási) erő a sebességgel arányos, és ellentétes irányú (fékező). Ekkor a rendszer differenciálegyenlete a következő [2]:

$$\ddot{x} + 2k\dot{x} + \omega_0^2 x = 0,$$

ahol ω_0 neve ez esetben is sajátfrekvencia, a k neve csillapítási tényező.

Ennek az állandó együtthatójú, homogén, lineáris differenciálegyenletnek a megoldásait keressük a szokásos $x(t) = e^{\lambda t}$ alakban. Behelyettesítés után azt kapjuk, hogy

$$\lambda^2 + 2k\lambda + \omega_0^2 = 0$$

$$\lambda_{1,2} = -k \pm \sqrt{k^2 - \omega_0^2}.$$

A differenciálegyenlet általános megoldása:

$$x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}.$$

Három eset van: valós, komplex, és a kettős gyökök esete.

a) $k < \omega_0$

Ekkor legyen $\omega = \sqrt{\omega_0^2 - k^2}$, ezzel

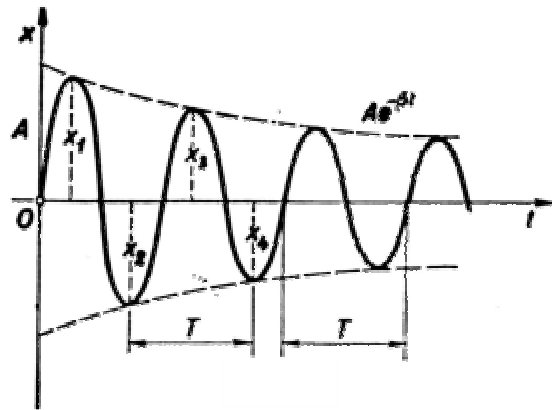
$$x(t) = e^{-kt} (c_1 e^{i\omega t} + c_2 e^{-i\omega t}).$$

Nekünk csak a valós megoldásokra van szükségünk. Akkor és csak akkor valós az $x(t)$ megoldás, hogyha a c_1 és c_2 komplex konstansok a következő alakban írhatók fel [4]:

$$c_1 = \frac{a}{2} e^{i\alpha}, \quad c_2 = \frac{a}{2} e^{-i\alpha}, \quad \text{ahol } a \text{ valós konstans, így}$$

$$x(t) = ae^{-kt} \cos(\omega t + \alpha),$$

illetve felírhatjuk a megoldást a szokásosabb szinuszos alakban is (az a és α' konstansok a kezdeti feltételekből határozhatók meg):



2. ábra: az $x(t)$ megoldás kinézete

$$x(t) = ae^{-kt} \sin(\omega t + \alpha').$$

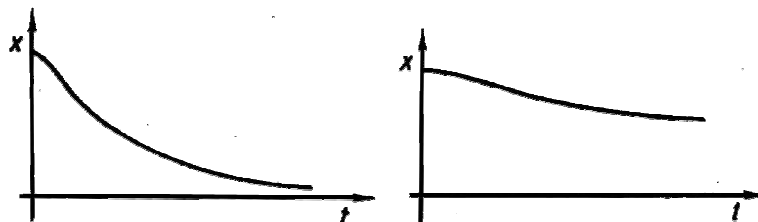
Tehát az általános megoldás egy szinuszgörbe, megszorozva egy exponenciálisan lecsengő burkolóval. A csillapítás időbeli mértékét egyértelműen a k csillapítási tényező befolyásolja. A lecsillapodó rezgés ω frekvenciája is függ az előzőekben leírt módon a k -tól.

b) $k = \omega_0$

Ezt az esetet aperiodikus határesetnek nevezzük. Ekkor az egyik partikuláris megoldása a differenciálegyenletnek e^{-kt} . Kell még egy másik, ez lehet a te^{-kt} . Ekkor az általános megoldás [2]:

$$x(t) = e^{-kt} (c_1 + c_2 t)$$

A következő ábra bal oldali részén látható a megoldás.



3. ábra: anharmonikus megoldások

c) $k > \omega_0$

Az általános megoldás [2]: $x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$, ahol $c_1, c_2, \lambda_1, \lambda_2$ valósak, és λ_1, λ_2 negatívak. A mozgást az előző ábra jobb oldali része szemlélteti.

Fizika óra keretében elvégezhetünk néhány, a csillapított rezgőmozgást szemléltető kísérletet. Ez alapján felállíthatunk egy hipotézist a $\ddot{x} + 2k\dot{x} + \omega_0^2 x = 0$ differenciálegyenlet megoldására: $x(t) = ae^{-kt} \cos(\omega t + \alpha)$, majd behelyettesítéssel meggyőződünk arról, hogy ez valóban megoldás, és arról, is, hogy milyen feltételek mellett. Ehhez fölteszük, hogy a tanulók hallottak már a derivált függvény fogalmáról, a szorzat deriváltjának kiszámításáról, a szinusz és az exponenciális függvény differenciálásáról. A számítás:

$$x(t) = ae^{-kt} \cos(\omega t + \alpha).$$

$$\dot{x}(t) = -ake^{-kt} \cos(\omega t + \alpha) - a\omega e^{-kt} \sin(\omega t + \alpha)$$

$$\ddot{x}(t) = a(k^2 - \omega^2)e^{-kt} \cos(\omega t + \alpha) + 2ak\omega e^{-kt} \sin(\omega t + \alpha)$$

Behelyettesítés után azt kapjuk, hogy:

$$a(\omega_0^2 - \omega^2 - k^2)e^{-kt} \cos(\omega t + \alpha) = 0.$$

Ez csak akkor lehet igaz, ha $(\omega_0^2 - \omega^2 - k^2) = 0$. Tehát $\omega = \sqrt{\omega_0^2 - k^2}$, és csak $k < \omega_0$ esetében létezik ez a megoldás.

3.2.3. Csillapított harmonikus rezgőmozgás és gerjesztés; a rezonátor

Egy gerjesztett csillapított harmonikus rezgőmozgást úgy valósíthatunk meg, hogy egy rugóra akasztott test mozgását figyeljük meg, miközben a rugó másik végét fogjuk és periodikusan mozgatjuk, és ezt az egész rendszert mondjuk vízbe helyezzük. A víz képviseli a csillapítást, a rugó biztosítja a kvázielasztikus erőt, a kezünk jelenti a gerjesztést, a test mozgását kényszerrezgésnek nevezzük, a rendszert pedig nevezzük rezonátornak. Még jobban demonstrálhatók a kényszerrezgések az un. Pohl-féle készülékkel [1]. A rezonátor tulajdonképpen bizonyos információt (pontosabban, egy frekvenciakomponens energiáját, lásd később) fog megmondani nekünk a gerjesztésről, a gerjesztő jelről. Ennek megértéséhez először meg kell vizsgálnunk a rezonátor működését. Egy adott $f(t)$ gerjesztő jel esetén a rezonátort leíró differenciálegyenlet a következő [4]:

$$\ddot{x} + 2k\dot{x} + \omega_0^2 x = f(t)$$

A homogén megfelelőjének (tehát amikor nincsen gerjesztés) az általános megoldását az előző pontban láttuk, most keressük meg az inhomogén differenciálegyenlet egy (partikuláris) megoldását. Ennek megkereséséhez először ne ezt a differenciálegyenletet tekintjük, hanem:

$$\ddot{x} + 2k\dot{x} + \omega_0^2 x = a \cos(\omega t). \quad (3.2.3.1)$$

Keressük a megoldást $x(t) = A \cos(\omega t + \varphi)$ alakban.

Ekkor $\dot{x}(t) = -A\omega \sin(\omega t + \varphi)$, és $\ddot{x}(t) = -A\omega^2 \cos(\omega t + \varphi)$. Ha behelyettesítjük az egyenletbe, azt hogy:

$$A(\omega_0^2 - \omega^2) \cos(\omega t + \varphi) - 2Ak\omega \sin(\omega t + \varphi) = a \cos(\omega t), \text{ ebből:}$$

$\frac{A}{a}[(\omega_0^2 - \omega^2) \cos(\omega t + \varphi) - 2k\omega \sin(\omega t + \varphi)] = \cos(\omega t)$. Bővítjük a bal oldali törtet:

$$\frac{A\sqrt{(\omega_0^2 - \omega^2)^2 + 4k^2\omega^2}}{a} \left[\frac{(\omega_0^2 - \omega^2)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4k^2\omega^2}} \cos(\omega t + \varphi) - \frac{2k\omega}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4k^2\omega^2}} \sin(\omega t + \varphi) \right] = \cos(\omega t)$$

A kapcsos zárójelen belüli tagnak meg kell egyeznie $\cos(\omega t)$ -vel, az együtthatónak pedig egynek kell lennie. Ebből következnek a következő képletek:

$$\operatorname{tg}(\varphi) = \frac{2k\omega}{\omega^2 - \omega_0^2}$$

$$A = \frac{a}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4k^2\omega^2}}$$

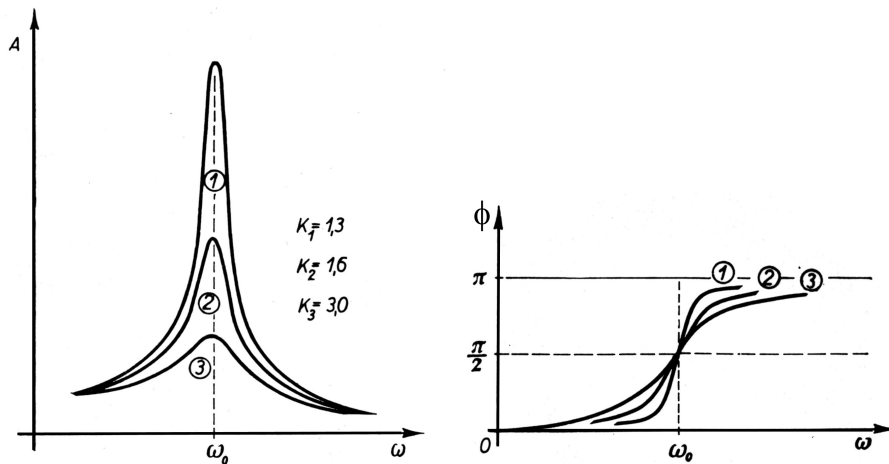
Tehát, a megoldás az előző képletekkel megadott konstansok felhasználásával: $A \cos(\omega t + \varphi)$.

A kényszerregés differenciálegyenletének általános megoldása pedig (figyelembe véve az előző pontban a homogén egyenletre kapott megoldást) [1]:

$$x(t) = be^{-kt} \sin(\sqrt{\omega_0^2 - k^2}t + \alpha) + A \cos(\omega t + \varphi)$$

A homogén egyenletből örökölt első tagra az ott leírtak vonatkoznak, a benne lévő állandók a kezdeti feltételekből kiszámíthatók. Ez a tag úgy interpretálható, hogy a rezonátor gerjesztő jel előtti állapota exponenciálisan lecseng, mialatt a rendszer rááll a gerjesztő frekvenciára (a megadott amplitúdóval).

Ami még a mi szempontunkból nagyon érdekes, az az ún. rezonanciagörbe, az $A(\omega)$, amire a rá vonatkozó összefüggést már levezettük.



4. ábra: rezonanciagörbe és a fáziseltérés

Ennek a görbének a maximumhelyét könnyen megkaphatjuk [1]:

$$\frac{d}{d\omega} \left((\omega^2 - \omega_0^2)^2 + 4k^2\omega^2 \right) = 0, \text{ innen}$$

$$\omega_r = \sqrt{\omega_0^2 - 2k^2} \quad (3.2.3.2), \text{ a neve rezonancia-frekvencia.}$$

Az is látható, hogy a rezonancia-frekvencián felvett amplitúdó annál nagyobb, minél kisebb a csillapítás. Az is könnyen bebizonyítható, hogy a rezonanciagörbe ún. félértékszélessége körülbelül $2k$ -val egyenlő. Tehát ha a csillapítást csökkentjük, akkor a rezonanciagörbe élesebbé és magasabbá válik, határesetben (ha $k = 0$) az amplitúdó a rezonanciafrekvencián végtelen nagyra válik. Viszont a csillapítás csökkentésének van egy hátránya is: az exponenciálisan lecsengő tag lassabban cseng le. Másrészt nem biztos, hogy az lesz a célunk a rezonátorral, hogy minél kisebb frekvenciatartományról gyűjtsünk információt, lehet, hogy éppen jól jön a kicsit nagyobb félértékszélesség.

Most nézzük meg az eredeti differenciálegyenletünk megoldását. Emlékeztetőül:

$$\ddot{x} + 2k\dot{x} + \omega_0^2 x = f(t)$$

Az általánosság megszorítása nélkül feltehetjük, hogy a hangot megkaphatjuk több harmonikus függvény összegeként [4].

Tehát legyen $f(t) = a_1 \cos(\omega_1 t + \varphi_1) + \dots + a_n \cos(\omega_n t + \varphi_n)$, n darab harmonikus összetevő összege. Legyen $x_1(t), \dots, x_n(t)$ a következő differenciálegyenletek megoldása:

$$\begin{aligned} \ddot{x}_1 + 2k\dot{x}_1 + \omega_0^2 x_1 &= a_1 \cos(\omega_1 t + \varphi_1) \\ &\vdots \\ \ddot{x}_n + 2k\dot{x}_n + \omega_0^2 x_n &= a_n \cos(\omega_n t + \varphi_n) \end{aligned}$$

Adjuk össze az egyenleteket, azt kapjuk, hogy:

$$\ddot{x} + 2k\dot{x} + \omega_0^2 x = f(t)$$

ahol $x(t) = x_1(t) + \dots + x_n(t)$. Tehát azt mondhatjuk, hogy $f(t)$ hatására létrejövő $x(t)$ rezgés

felírható $x(t) = \sum A_n \cos(\omega_n t + \phi_n)$ alakban (a lecsengő tagtól most eltekintve), ahol

$$A_n = \frac{a_n}{\sqrt{(\omega_0^2 - \omega_n^2)^2 + 4k^2 \omega_n^2}}$$

és ϕ_n -re is a megfelelő formula érvényes [4]:

$$\phi_n = \arctan \frac{2k\omega}{\omega^2 - \omega_0^2} + \varphi_n.$$

Tehát azt kaptuk, hogy ha a gerjesztő jel több harmonikus összetevő összege, akkor a létrejövő $x(t)$ rezgés is ugyanezen harmonikus összetevők összege, viszont ezen összetevők A_n / a_n amplitúdó erősítése a rezonancia frekvencián a legnagyobb (vö. 3.2.3.2). A rezonátor felhasználásáról a következő fejezetben lesz szó.

3.3. A rezonátor differenciálegyenletének numerikus megoldásáról

Nyilvánvaló, hogy a differenciálegyenlet egzakt megoldását numerikusan csak közelítőleg (vagy még úgy sem) tudjuk meghatározni. A differenciálegyenletek numerikus integrálásának elméletében ([5], [6]) elég sokféle módszert találhatunk a differenciálegyenlet közelítő megoldására. Ilyenek például az egy lépéses, többlépéses, Runge-Kutta, prediktor-korrektor és a magasabb rendű deriváltakon alapuló módszerek. A cél az, hogy minél gyorsabban futó algoritmus kapjunk (ezért célszerű valamilyen numerikus egy lépéses módszert választani a differenciálegyenlet folyamatos, közelítő megoldására).

Először is vizsgáljuk meg a differenciálegyenlet egzakt megoldását. A differenciálegyenlet konstans együtthatós, így teljesülnek a Picard-Lindelöf tétel feltételei [7], emiatt a differenciálegyenletnek pontosan egy megoldása létezik adott kezdőértékek mellett. Az egyetlen megoldást az előző alfejezetben meg is kaptuk.

Most egy lépéses numerikus módszereket nézünk meg abból a szempontból, hogy jól alkalmazhatók-e. Írjuk fel a másodrendű differenciálegyenletünket elsőrendű differenciálegyenlet rendszer alakban:

$$\begin{aligned}\dot{x} &= v \\ \dot{v} &= -2kv - \omega_0^2 x + f(t)\end{aligned}$$

Numerikus megoldási módszerek közül az állandó lépésközűekkel foglalkozunk. Ekkor élhetünk a $\Delta t = 1$ választással a számítások egyszerűbbé tétele és gyorsítása érdekében, ami együtt jár a k és az ω_0 paraméterek bizonyos transzformációjával. Intuitíven azt mondhatjuk: ha $f(t)$ egy 22050 Hz mintavételezésű hang, akkor a Δt legyen 1/22050 sec, mivel ennyi idő elteltével van új jel. Az ω_0 frekvenciát pedig behangoljuk mondjuk 100 Hz-re. Ezzel viszont ekvivalens az a rendszer, amiben $f(t)$ -t egy 1 Hz mintavételezésű hangként

fogjuk fel, azaz a következő jel $\Delta t = 1$ idő múlva következnek, de akkor az ω_0 frekvenciát is a képletben át kell írni $\omega_0 / 22050$ -re. Egzakt módon kifejezve az előbb leírtakat, az idő tengelyt transzformáljuk: $t' = 22050t$. Ezt behelyettesítve a $\ddot{x} + 2k\dot{x} + \omega_0^2 x = f(t)$ egyenletbe, azt kapjuk, hogy $22050^2 \ddot{x}(t') + 22050 \cdot 2k\dot{x}(t') + \omega_0^2 x(t') = f(t'/22050)$. Most ha az eredetiek helyett vesszük a következő, új paramétereket:

$$\begin{aligned} k &= k / 22050 \\ \omega_0 &= \omega_0 / 22050 \\ f(t) &= f(t / 22050) / 22050^2, \end{aligned}$$

akkor ezzel a helyettesítéssel éppen az eredeti differenciálegyenletet kapjuk. Tehát $\Delta t = 1$ választással élhetünk, de ekkor a paramétereket következetesen kell megadni.

(a) Nézzük meg, hogy mi történik, ha ezt a differenciálegyenlet rendszert a következő triviális egylépéses numerikus módszerrel oldjuk meg ($\Delta t = 1$):

$$\begin{aligned} x_{n+1} &= x_n + v_n \\ v_{n+1} &= v_n - 2kv_n - \omega_0^2 x_n + f(n). \end{aligned}$$

Vezessük be az $\begin{pmatrix} x_n \\ v_n \end{pmatrix}$ vektort, ezzel felírva az előző egyenletrendszert kapjuk, hogy

$$\begin{pmatrix} x_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -\omega_0^2 & 1 - 2k \end{pmatrix} \begin{pmatrix} x_n \\ v_n \end{pmatrix} + \begin{pmatrix} 0 \\ f(n) \end{pmatrix} = A \begin{pmatrix} x_n \\ v_n \end{pmatrix} + B_n \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ ahol}$$

$$B_n = \begin{pmatrix} f(n) & 0 \\ 0 & f(n) \end{pmatrix}.$$

Írjuk fel $\begin{pmatrix} x_n \\ v_n \end{pmatrix}$ vektort rekurzívan, $\begin{pmatrix} x_0 \\ v_0 \end{pmatrix}$ vektorig visszamenve. Azt kapjuk, hogy

$$\begin{pmatrix} x_n \\ v_n \end{pmatrix} = A \left(A \begin{pmatrix} x_{n-2} \\ v_{n-2} \end{pmatrix} + B_{n-2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) + B_{n-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \dots = A^n \begin{pmatrix} x_0 \\ v_0 \end{pmatrix} + (A^{n-1}B_0 + \dots + AB_{n-2} + EB_{n-1}) \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Ez azt mutatja, hogy az $\begin{pmatrix} x_n \\ v_n \end{pmatrix}$ vektor két tagból áll. Az első tagról ki kellene derülni, hogy az a lecsengő tag, a másodiktól pedig az, hogy ő a szinuszos. Illetve kiderülhet az is, hogy ez a numerikus megoldás nem stabil. Az utóbbi kérdésre az A mátrix determinánsa adhatja meg a választ. Ahhoz, hogy az első tag lecsengő legyen, szükséges, hogy $\det(A) < 1$ teljesüljön, mert különben a determinánsokra vonatkozó szorzástétel ($|AB| = |A| \cdot |B|$) alapján $|A^n|$ tartana a végtelenbe.

$$\begin{vmatrix} 1 & 1 \\ -\omega_0^2 & 1-2k \end{vmatrix} < 1, \text{ innen}$$

$$1 - 2k + \omega_0^2 < 1,$$

$$\frac{\omega_0^2}{2} < k.$$

Ez az egyenlőtlenség, azaz a csillapítás alulról korlátozása viszont számunkra nem kedvező, mert ekkor nem tud a rezonátor a közép, illetve magas frekvenciatartományban a nagy csillapítás miatt megfelelően működni, csak alacsony frekvenciatartományban használható.

(b) Most egy, az előzőnél kedvezőbb A mátrixot adunk meg. A számítás a következő:

$$\begin{aligned} v_{n+1} &= v_n - 2kv_n - \omega_0^2 x_n + f(n) \\ x_{n+1} &= x_n + v_{n+1} \end{aligned} \tag{3.3.1}$$

Ehhez a közelítéshez is felírjuk a nemrekurzív formulát, mint az (a) pontban is tettük:

$$\begin{pmatrix} x_n \\ v_n \end{pmatrix} = A^n \begin{pmatrix} x_0 \\ v_0 \end{pmatrix} + (A^{n-1}B_0 + \dots + AB_{n-2} + EB_{n-1}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3.3.2)$$

ahol $A = \begin{pmatrix} 1 - \omega_0^2 & 1 - 2k \\ -\omega_0^2 & 1 - 2k \end{pmatrix}$ és $B_n = \begin{pmatrix} f(n) & 0 \\ 0 & f(n) \end{pmatrix}$.

Vizsgáljuk meg ismét az A mátrix determinánsát! A következőt kapjuk:

$$\begin{vmatrix} 1 - \omega_0^2 & 1 - 2k \\ -\omega_0^2 & 1 - 2k \end{vmatrix} < 1, \text{ innen}$$

$$1 - 2k < 1, \text{ azaz } 0 < k.$$

Ez semmiképpen nem jelent túl nagy megkötést a k értékére nézve.

A további numerikus vizsgálatok apró betűvel szerepelnek, ezeknek elég csak az eredményeit közölni a tanulókkal. Számítsuk ki az A mátrix sajátértékeit:

$$\lambda_{1,2} = \frac{-(\omega_0^2 + 2k - 2) \pm \sqrt{(\omega_0^2 + 2k - 2)^2 - 4(1 - 2k)}}{2}.$$

A sajátértékeket fel akarjuk írni $r \cdot e^{i\omega}$ alakban. Könnyen kiszámolható, hogy a következő képletek az érvényesek:

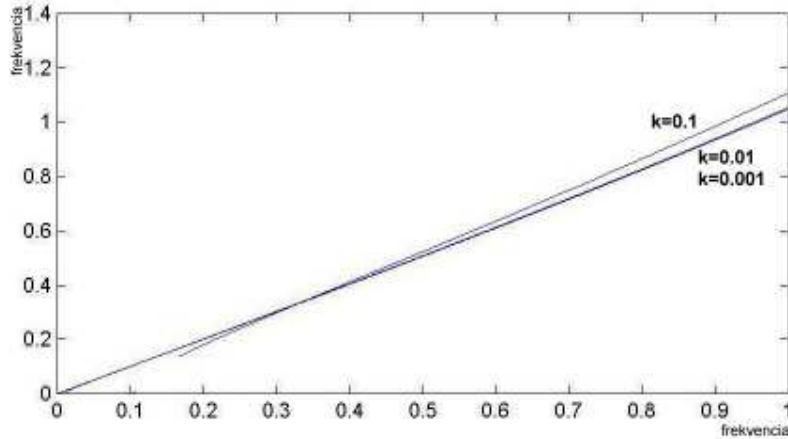
$$r = \sqrt{1 - 2k}$$

$$\omega = \pm \text{arc ctg} \frac{\frac{\omega_0^2 + 2k - 2}{\sqrt{4 - 8k}}}{\sqrt{1 - \left(\frac{\omega_0^2 + 2k - 2}{\sqrt{4 - 8k}}\right)^2}}$$

Látható, hogy a sajátértékek abszolút értéke mindig kevesebb egynél, ha a csillapítás nem 0. Azt a feltételt is kaptuk a csillapításra, hogy:

$$k < \frac{1}{2}$$

viszont ez nem jelent a gyakorlatban lényeges megkötést. Az ω -ra kapott formula szintén megfelelőnek mondható, nem túl nagy k esetén lehet azt mondani, hogy $\omega \approx \omega_0$. Ezt szemlélteti a következő ábra, amin az $\omega(\omega_0)$ függvény van ábrázolva bizonyos k értékekre:



5. ábra: ω és ω_0 eltérése

Tehát megkaptuk az A mátrix sajátértékére a megfelelő összefüggéseket, amely szerint

$$\lambda_{1,2} \approx \sqrt{1-2k} e^{\pm i\omega_0}.$$

Ezek szerint az A^n mátrix sajátértékei: $\lambda_{1,2} \approx (\sqrt{1-2k})^n e^{\pm in\omega_0}$, ami miatt az $A^n \begin{pmatrix} x_0 \\ v_0 \end{pmatrix}$ tag egy exponenciálisan lecsengő szinuszos függvény.

Ha össze akarjuk hasonlítani a k értékét az egzakt megoldásbeli (3.3 alfejezet) k értékkel, definiálhatjuk

$\kappa = -\ln \sqrt{1-2k}$. Ekkor $\lambda_{1,2} = e^{-\kappa} e^{\pm i\omega_0}$, másrészt kis k esetén igaz a következő összefüggés, ha $\ln(1-x)$ -et az ő Taylor sorának az első tagjával közelítjük: $\kappa = -\ln \sqrt{1-2k} = -\frac{1}{2} \ln(1-2k) \approx k$.

Most fordítsuk a figyelmünket a második tagra: $(A^{n-1}B_0 + \dots + AB_{n-2} + EB_{n-1}) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. A zárójelben lévő mátrix (jelöljük C_n -nel) sajátértékeire vagyunk kíváncsiak. Mivel C_n tulajdonképpen A polinomja, az egyik sajátértékét (jel. $\lambda_1^{(n)}$) a következőképpen írhatjuk fel:

$$\lambda_1^{(n)} = \sum_{j=0}^{n-1} r^j e^{ij\omega_0} f(n-1-j).$$

Tegyük fel, hogy $f(n) = \sin(n\omega + \varphi)$, ahol ω tetszőleges, és írjuk fel $\lambda_1^{(n+1)}$ -et $\lambda_1^{(n)}$ segítségével:

$$\lambda_1^{(n+1)} = \sin(n\omega) + r \cdot e^{i\omega_0} \lambda_1^{(n)}.$$

Ha felhasználjuk, hogy $2\sin(n\omega) = e^{i\omega} - e^{-i\omega}$, akkor $\lambda_1^{(n)} = c_1 e^{ion} + c_2 e^{-ion}$ alakú megoldáshoz jutunk.

A sajátértékek ilyen változása miatt a (3.3.2) összefüggés második tagja egy ω körfrekvenciájú szinusz. Tehát azt kaptuk, amit szerettünk volna: (3.3.2) első tagja egy exponenciálisan csillapodó szinusz, a második tagja pedig a gerjesztő jel frekvenciájával azonos frekvenciájú szinusz. Így az egzakt megoldással valamilyen értelemben egyenértékű numerikus megoldáshoz jutottunk el.

A következő fontos kérdés vetődik még fel: mi az összefüggés az $x_n = Ae^{i(\omega n + \varphi)}$ A amplitúdója és a gerjesztő jel amplitúdója között? Egy olyan összefüggést szeretnénk kapni, ami hasonlít a 3.2 fejezetbeli

$$A = \frac{a}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4k^2\omega^2}} \text{ formulára.}$$

Ehhez írjuk át a (3.3.1) formulát a v_n -ek kiküszöbölésével a következő alakúra:

$$x_{n+1} = x_n(2 - 2k - \omega_0^2) - x_{n-1}(1 - 2k) + f(n)$$

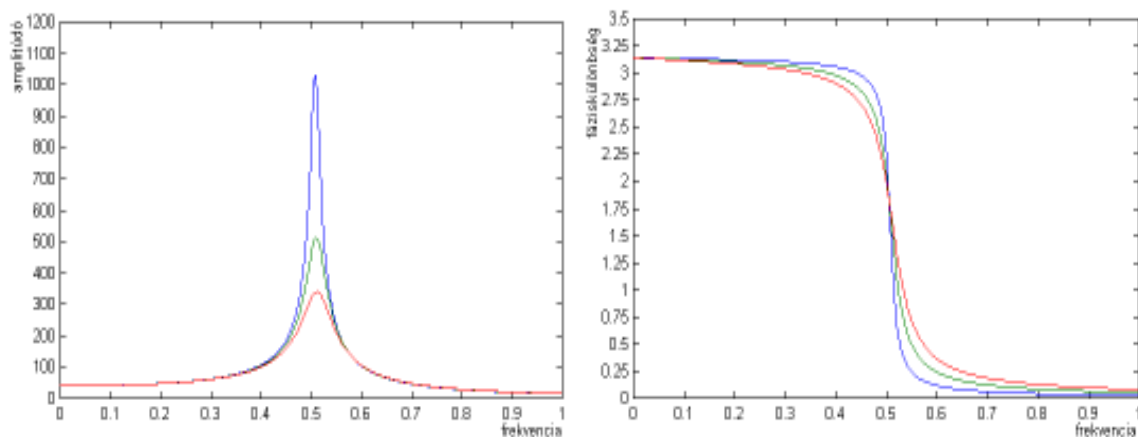
Az előző bekezdésekből tudjuk, hogy a lecsengő tagtól eltekintve $x_n = \text{Re}(Ae^{i(\omega n + \varphi)})$ alakú, ha a gerjesztő jel $f(n) = \text{Re}(Ae^{i\omega n})$ alakú. Helyettesítsük be az előbbi rekurziós formulába a két komplex sorozatot. Átrendezés után:

$$Ae^{i\varphi}(e^{i\omega} + (\omega_0^2 - 2 + 2k) + (1 - 2k)e^{-i\omega}) = a$$

A két oldal abszolút értékének meg kell egyeznie, ebből azt kapjuk, hogy:

$$A = \frac{a}{\sqrt{((2 - 2k)(\cos \omega - 1) + \omega_0^2)^2 + (2k \sin \omega)^2}}$$

Ez a formula kis ω esetén (ha a szinusz függvényt a Taylor polinomja első, a koszinusz függvényt az első kettő tagjával helyettesítjük) azonos a (3.3)-ban megadott formulával. A következő ábra az $A(\omega)$ függvényt szemlélteti $\omega_0 = 0.5$ esetén, különböző csillapításokra ($k=0.01, k=0.02, k=0.03$):



6. ábra: A numerikus megoldás rezonanciagörbéje és fáziseltérése

3.4. A rezonátorok kapcsolata a digitális szűrőkkel

A digitális szűrők tulajdonképpen az analóg szűrők digitális megfelelői. Arra jók, hogy a jel bizonyos frekvenciájú szinuszos összetevőinek amplitúdóját, fázisát megváltoztassák.

Mint láttuk a rezonátor differenciálegyenlet rendszerének (b) megoldásánál, az x_n megadható a következő módon:

$$x_{n+1} = x_n(2 - 2k - \omega_0^2) - x_{n-1}(1 - 2k) + f(n) \quad (3.4.1)$$

IIR szűrőnek ([3], [9]) nevezünk egy olyan rendszert, ami felírható a következő formában:

$$y[n] = \sum_{k=0}^M b_k x[n-k] + \sum_{k=1}^M a_k y[n-k],$$

ahol $y[n]$ a rendszer kimenete, $x[n]$ a rendszer bemenete, és legalább egy a_k nem nulla (tehát a kimenet nem kapható meg a bemenet előző pár értékének lineáris kombinációjaként). Ennek alapján, ha a (3.4.1) formulában $f(n)$ helyett $x[n]$ -t írunk, x_n helyett pedig $y[n]$ -t, akkor látjuk, hogy egy IIR szűrő alakjához jutunk el.

Annak, hogy nem szűrőként tárgyaltuk a dolgot, habár megtehettük volna, annak az az oka, hogy a szerző számára így egyszerűbbnek tűnt, másrészt nem szűrt jel, hanem amplitúdó spektrum kinyerése a végső célja a dolgozat eddigi részének.

4. Spektrogram felvétele rezonátorok segítségével

4.1. Spektrogram

Az előző fejezetben bemutatott rezonátor segítségével szeretnénk spektrogramot készíteni. A spektrogram a hangot vizualizálja olyan módon, hogy megmutatja minden időpontban azt, hogy a hang milyen frekvenciájú szinuszos komponensekből tevődik össze, és ezeknek mekkora az amplitúdója. Három mennyiségünk van tehát: az idő, a frekvencia, és az amplitúdó. *Az idő a vízszintes tengely, a frekvencia a függőleges tengely, és színnel ábrázoljuk az amplitúdót.* Mielőtt fizika órán rátérnénk a következő fejezetekre, azt ajánljuk, hogy mutassuk be különböző hangok spektrogramját. Ha lineáris frekvenciaskálát használunk, akkor be tudjuk mutatni a zenei hangok és a beszélt zöngés hangok tanulmányozásával azt, hogy a jelenlevő frekvenciaértékek egy alapfrekvencia egész szám-szorosai (ezeket nevezzük felharmonikusoknak).

4.2. Spektrogram előállítás rezonátorral

A spektrogram elkészítése a következő módon történik: Veszünk n darab rezonátort. A rezonátorokat behangoljuk: megadjuk az őket leíró differenciálegyenlet ω_0 és k paramétereit úgy, hogy a rendszerek rezonanciafrekvenciája a kívánt frekvencia legyen ($\omega_r = \sqrt{\omega_0^2 - 2k^2}$). Ezután numerikusan megoldjuk az n darab differenciálegyenletet párhuzamosan, egy megadott hangfájl (ez szolgáltatja az $f(t)$ függvényt) esetére. Közben adott időközönként kiszámoljuk az egyes rezonátorok energiáját, és tulajdonképpen ebből származik a spektrogram.

Elevenítsük fel a rezonanciagörbe fogalmát:

$$A = \frac{a}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4k^2 \omega^2}}. \quad (4.1.1)$$

Ez az összefüggés az eredeti jel maximális kitérése (azaz az amplitúdója (a)), és a kényszerrezgés amplitúdója (A) között. A kényszerrezgés amplitúdóját egy adott időpontban pedig meg tudjuk mondani, még ha a rendszer nincs is a maximális kitéréshez tartozó állapotban, mégpedig azáltal, hogy a rendszer energiáját kétféleképpen írhatjuk fel:

$$E = \frac{1}{2} \omega^2 A^2$$

$$E = \frac{1}{2} v^2 + \frac{1}{2} \omega^2 x^2.$$

Minden egyes rezonátornak csak azon a frekvenciaértéken kell pontosnak lennie, ahová behangoltuk, ezt az elvárást támasztjuk. Ebben az esetben $\omega = \omega_0$. Ez alapján a második képlet alapján az energiát minden iterációs lépésben ki tudjuk számítani, és ebből az energiaértékből megkaphatjuk az amplitúdót az első egyenlet segítségével. Ezek után tegyük fel, hogy a hang ($f(t)$) tartalmaz az egyik rezonátor rezonanciafrekvenciájához közeli harmonikus rezgést. Ekkor, ha k csillapítás nem túl nagy (ekkor igaz, hogy $\omega_0 \approx \omega_r \approx \omega$), a (4.1.1) egyenletekből látszik a következő becslés:

$$a = A \cdot 2k\omega_0 = 2k\sqrt{2}\sqrt{E} = 2k\sqrt{2}\sqrt{\frac{1}{2}\dot{x}^2 + \frac{1}{2}\omega_0^2 x^2}.$$

Azaz ha a jelben egy rezonátor rezonanciafrekvenciájához közeli harmonikus összetevő jelen van, akkor az eredeti jelösszetevő amplitúdóját az előző képlet segítségével tudjuk becsülni. De mit ad ez a számítás más a gerjesztő jel frekvenciájához képest távoli frekvenciára hangolt

rezonátor esetén? Mivel A nagyon kicsi akkor, ha ω nincs ω_r közelében, ezért a is kicsi lesz.

(Hogy mennyire kicsi, azt a (4.1.1) formula, a rezonanciagörbe határozza meg.)

Összefoglalva a fentieket, most leírjuk az algoritmust (n darab rezonátor, 50 iterációnként (egy iteráció=egy minta feldolgozása) számítjuk ki az n db. komponensű amplitúdó vektort):

Spektrogram számítás

```
Rezonátorhangolás(n, minfrek, maxfrek, ...);  
While (van még jel a fájlban)  
    J=Következő jel a fájlból();  
    Iteráció minden rezonátorra(J);  
    If (iterációszám=0 mod 50) spektrum vektor számítás();
```

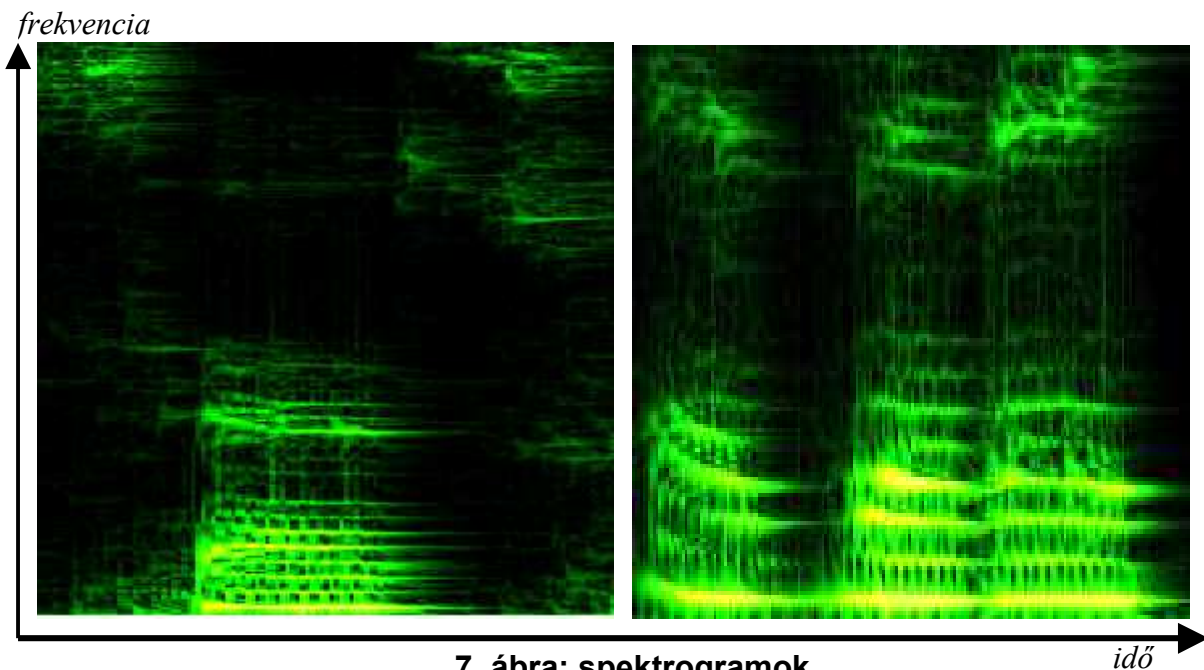
Iteráció minden rezonátorra(jel)

```
For i=1 to RezonátorSzám  
    sebesseg[i]+=(-w[i]*hely[i]-k[i]*sebesseg[i]+jel);  
    hely[i]+=(sebesseg[i]);
```

amplitúdó vektor számítás

```
For i=1 to RezonátorSzám  
    amp[i]= k[i]*sqrt(  
        (sebesseg[i]*sebesseg[i]+w[i]*w[i]*hely[i]*hely[i]) );
```

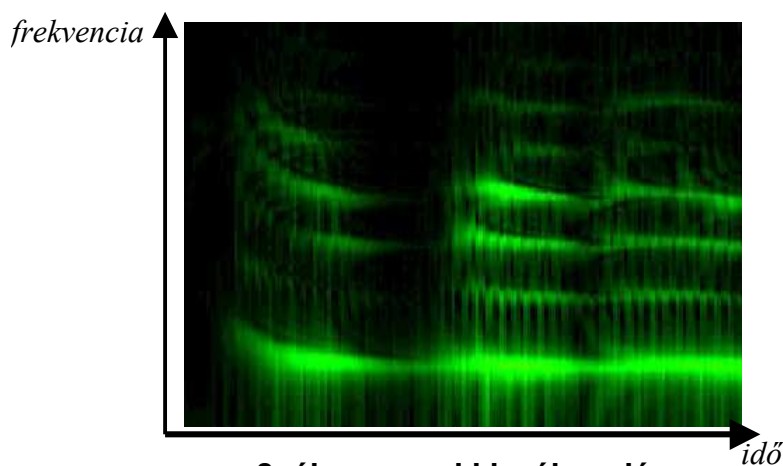
Ilyen spektrogramokat mutatnak a következő ábrák:



7. ábra: spektrogramok

4.3. Hangolás

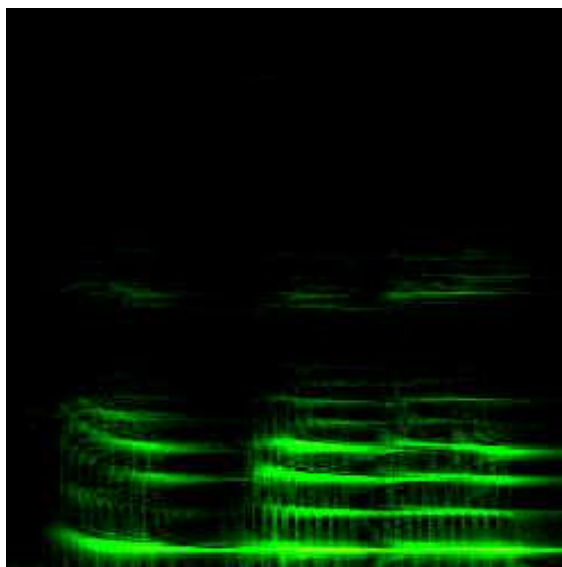
Az elérendő cél határozza meg azt, hogy a rezonátorok milyen rendszerét választjuk. Ha a cél az, hogy megmondjuk, hogy a jelben jelen van-e egy adott frekvenciájú komponens, és ennek a komponensnek mekkora az amplitúdója, akkor elegendő egyetlen egy rezonátort behangolnunk, kis csillapítással. Ha a cél a hangmagasság meghatározása, pl. beszéd esetén, akkor az alacsony frekvenciatartományra kell sok rezonátort beállítani, mint például az ábrán. Ha sok rezonátor van besűrítve szűk frekvenciatartományba, akkor egy problémába ütközünk: több rezonátor is viszonylag nagy energiájú gerjesztést jelezhet, tehát a spektrumvonalak kiszélesednek. Erre még visszatérünk. Olyan alkalmazás is elképzelhető, amikor csak egyes frekvenciasávok közötti becsült energiaeloszlásra vagyunk kíváncsiak. Ekkor berakhatunk az egyes sávok közepére egy-egy erősen csillapított rezonátort. Azt kell csak figyelembe venni k beállításakor, hogy a rezonanciagörbe félértékszélessége kb. $2k$, másrészt a differenciálegyenlet numerikus megoldásából adódó feltételek is jó, ha teljesülnek.



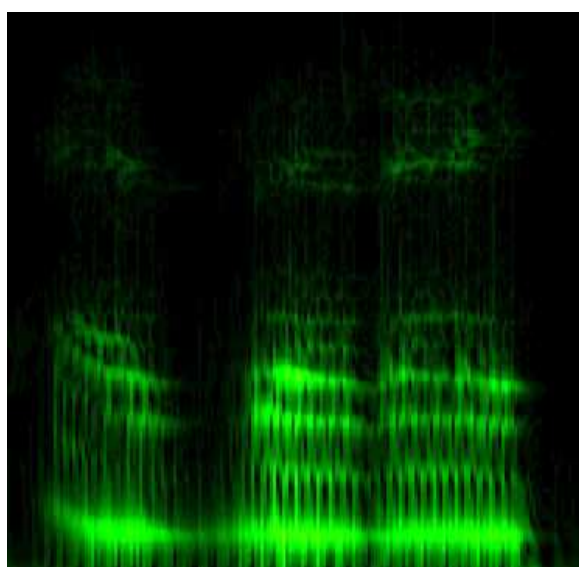
8. ábra: vonal kiszélesedés

4.4. Vonal kiszélesedés

Az előző bekezdésben említett vonal kiszélesedés a csillapítási tényezőtől és a rezonátorok frekvenciatartománybeli sűrűségétől függ, vagyis az egyes rezonátorok rezonanciagörbéinek összeéréséből, fedéséből következik. Ekkor azt lehet mondani, hogy a maximális amplitúdót szolgáltató rezonátor közelíti legjobban a jel ottani frekvenciájú komponensének amplitúdóját, vagy azt is megtehetjük, hogy megnézzük még a szomszédos rezonátorokat is, és azok amplitúdó értékei alapján valamilyen közelítéssel pontosítjuk a hangmagasság és amplitúdó becslését. Itt megjegyezzük azonban, hogy ekkora fokú pontosításra általában nincs szükség. A vonal kiszélesedés hatékony megszüntetésével a következő fejezet foglalkozik.



9. ábra: kis csillapítás



10. ábra: nagy csillapítás

A fenti képeket kis illetve nagy csillapítás esetén kaptuk, és mutatják a vonal kiszélesedést illetve az időbeli magnyúlást.

4.5. Időbeli megnyúlás

Az időbeli megnyúlás szintén a csillapítási tényezőhöz kapcsolódik, itt arról van szó, hogy ha a jelben az adott frekvenciájú komponens megszűnik, attól még a rezonátor továbbra is nem 0 amplitúdó értéket szolgáltat, ez az amplitúdó érték exponenciálisan cseng le. Pontosabban $\exp(-kt)$ burkológörbe szerint (illetve lásd a numerikus megoldásnál elmondottakat). Ha a csillapítás nagyobb, akkor időben kisebb tehetetlenségű, pontosabb, gyorsabban lecsengő függvényt kapunk. Egy másik megoldási lehetőség az, ha egyszerűen kivonjuk a megelőző spektrum vektor $\exp(-kt)$ -szeresét az újból, így a megszűnő hang lecsengése csökkenthető.

4.6. Kapcsolat a hallással

Az előbbieken leírt rezonátor rendszer a következő módon kapcsolódik az emberi halláshoz: az emberi hallás is tulajdonképpen egy rezonancia jelenség (habár nem lineáris, hanem több dimenziós), ami hosszú idő alatt a tökéletességig fejlődött az evolúció által. Igazából a halláson gondolkozva sokféle elméleti modellt készíthetünk. Amiről a dolgozat szól, az egy egyszerű példája ennek. Az emberi hallást ha összehasonlítjuk a rezonátor rendszerrel, akkor azt mondhatjuk, hogy a hallás, mint analóg rendszer nyilvánvalóan sokkal jobb. Érdekeség, hogy a hallásnál tapasztalható különböző jelenségeknek meg van a megfelelője a rezonátoroknál, és fordítva. Az ún. elfedés azt jelenti, hogy egy nagyobb intenzitású szinuszos komponens jelenlétekor az ő frekvenciájának kis környezetében lévő frekvenciájú, de kisebb intenzitású hangokat nem érzékeljük. Az elfedésnek a megfelelője a rezonátoroknál a vonal kiszélesedés. A hallást tanulmányozó kutatók, pszichológusok

felfigyeltek bizonyos időbeli gátlásokra a hallásban: például egymás után gyorsan megszólaltatott azonos frekvenciájú hangok nem választhatók szét. Az ehhez hasonló időbeli gátlásoknak, elfedésnek illetve „tehetetlenségnek” pedig megfeleltethető lehet a digitális rezonátoroknál az időbeli megnyúlás, a differenciálegyenlet lecsengő tagja.

Ha tovább részletezzük fizika órán ezt a kérdést, akkor még mélyebben tárgyalhatjuk a fül működését. A hallás során az ideghártyán lévő érzékelő neuronok nagyjából az általunk előállított spektrogramnak megfelelően stimulálódnak, amikor a hang megrezegetti azokat. Az agy neurális hálózata ki kell, hogy élesítse a képet, el kell hagyni a nem releváns információkat. Az élesítésből származó információvesztésből következik lényegében az elfedés jelensége. Számítógépen is léteznek már több évtizede neurális hálózatok (neten: előrecsatolt neurális hálózat, feed-forward neural network), és különböző beszédfelismerő rendszerek azon alapulnak, hogy spektrum vektoroknak valamilyen transzformált változatából álló adatbázison tanítják a gépi neuronhálókat.

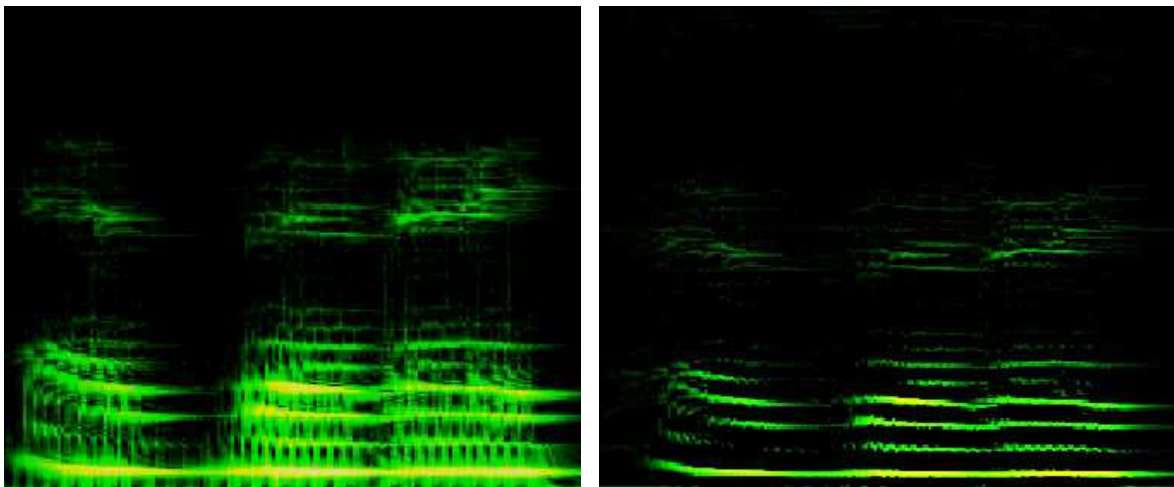
5. Spektrogram feldolgozása visszaalakításhoz

Ahhoz, hogy a spektrogramot a következőkben leírt módon visszaalakíthassuk, különböző változtatásokat kell rajta elvégezni.

5.1. Vonalkiszéledés megszüntetése

A vonalkiszéledés megszüntetésére alkalmazhatunk dekonvolúciót illetve szűrést ([11], [10]). A már megkapott spektrum vektort FFT-vel átalakítjuk, a kapott vektor komponenseit megszorozzuk egy dekonvolúciós vektor megfelelő komponenseinek reciprokával, vagy a szűrő vektorral, majd az inverz FFT segítségével megkapjuk a kiélesített spektrum vektort (az esetleg keletkező negatív komponenseket 0-val kell helyettesíteni).

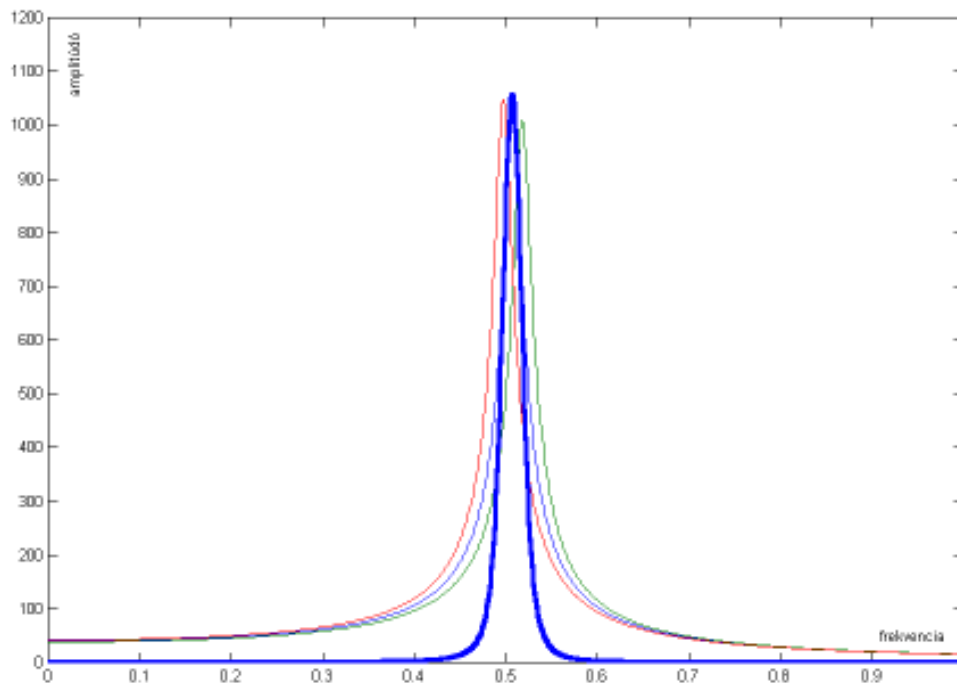
A következő ábra a vonalkiszéledést csökkentő dekonvolúció alkalmazását szemlélteti (dekonvolúció előtt és után):



11. ábra: spektrogram a transzformáció előtt és után

Az általunk javasolt kiélesítési mód, amit fizika órán is esetleg bemutathatunk, az a rezonátorok rezgéséből, a rezonanciagörbékből és a fáziseltérés görbékből indul ki. Ha ránézünk két közel azonos frekvenciára hangolt rezonátor előbb említett görbéire, akkor azt látjuk, hogy a rezonanciafrekvenciától bizonyos távolságra lévő frekvenciákon a két görbe

már magyarából összesímul. Ebből könnyen adódik a következtetés: vonjuk ki a szomszédos rezonátorok mozgásállapotát egymásból, és akkor egy élesebb rezonanciagörbéhez jutunk. Ezt szemlélteti a következő ábra, de itt már mindkét szomszéd mozgásállapotának a felét vonjuk ki a középső rezonátor mozgásállapotából. A következő ábrán a három vékonyabb vonallal ábrázolt görbe az $\omega_0^{(i)} = 0.5$ és $k^{(i)} = 0.01$ paraméterekkel rendelkező rezonátor, az $\omega_0^{(i-1)} = 0.49$, $k^{(i-1)} = 0.01$ és az $\omega_0^{(i+1)} = 0.51$, $k^{(i+1)} = 0.01$ paraméterű rezonátorok rezonanciagörbéjét mutatja. A vastag vonallal a különbség-rezgéshez tartozó rezonanciagörbét ábrázoltuk.



12. ábra: az új rezonanciagörbe

A mozgásállapotból most is számíthatunk energiát, ebből pedig visszakövetkeztethetünk az eredeti amplitúdóra, mint ahogyan tettük ezt az egyszerű rezonátorok esetében is, csak kicsit összetettebb formulával kell szorozni.

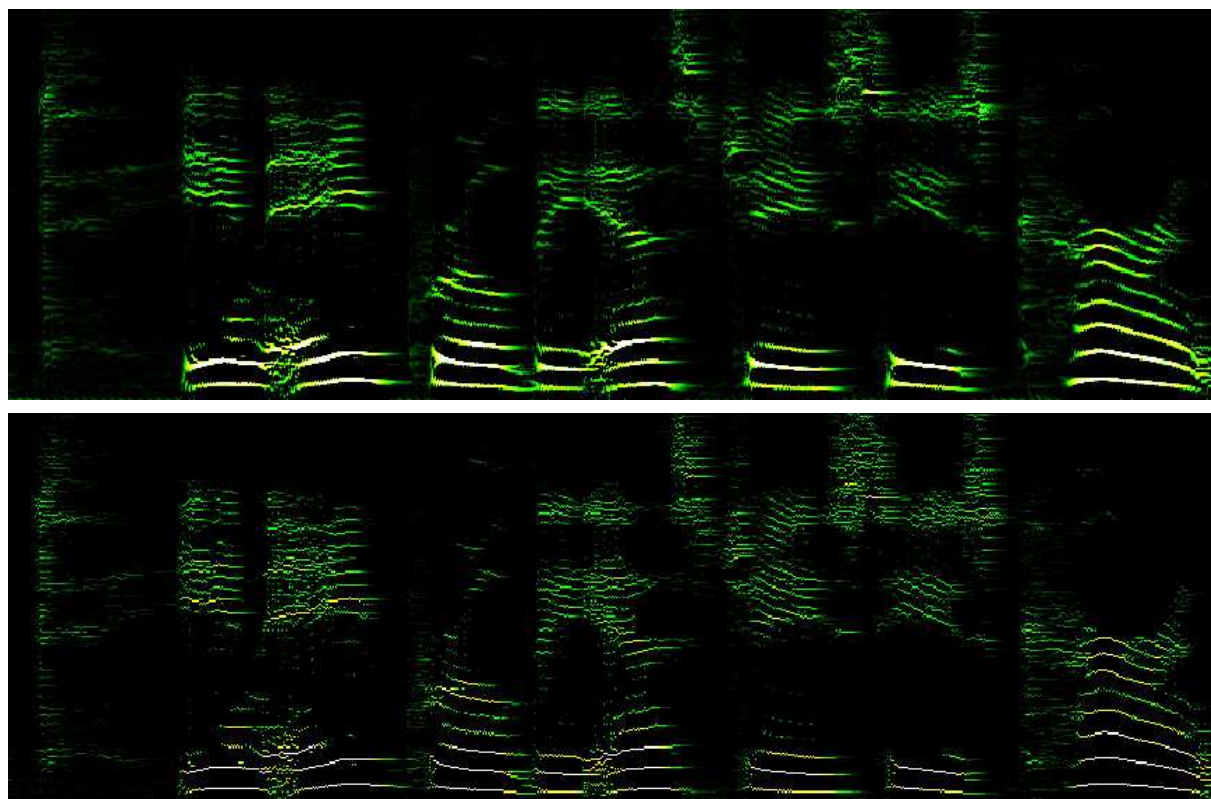
5.2. Előkészítés frekvenciakövetéshez: vonal egyértelműsítés

Az említett élesítési technikák után is fennmaradhat bizonyos vonal kiszélesedés, például abban az esetben, amikor a jelben jelenlévő harmonikus komponens frekvenciája két egymás melletti frekvenciára behangolt rezonátor frekvenciái közé esik, és a rezonátorok elég sűrűn vannak hangolva. Ezt célszerű megszüntetni, kiválasztani a nagyobbikat és a másikat kinullázni. Ezzel egyértelműsítjük azt, hogy a jelben található eredeti összetevőt milyen frekvenciájú jelnek vesszük visszaalakításkor; ha ezt nem tesszük meg, akkor visszaalakításkor a másik hamis komponens miatt lebegés és zaj lép fel, tehát tulajdonképpen olyan komponenset hagyunk el, ami valószínűleg nem is volt jelen az eredeti hangban (ha jelen is lett volna, az emberi hallás elfedési jelensége miatt hagyható el). Másrészt az elhagyásra a majdani szuperpozíciós visszaalakító eljárás miatt is szükségünk van. Tehát az itt javasolt algoritmus az, hogy az eredeti spektrum vektor egymás melletti amplitúdó értékei közül az algoritmus kimenő vektorában csak legfeljebb az egyiket (a nagyobbat) tartjuk meg. A pontos algoritmust C nyelv-szerűen adjuk meg, mert minden informatikával foglalkozó ezt ismeri:

```
for(i=0; i<rezonatorSzam-1; i++)
    if(amp[i] < amp[i+1] )
        amp[i]=0;
for(i=rezonatorSzam-1; i>0; i--)
    if(amp[i] < amp[i-1] )
        amp[i]=0;
```

Ez az algoritmus egyrészt egyértelműsíti az értékeket, élesíti a spektrumot, másrészt elő is készíti a spektrogramot a következő részben leírt visszaalakításra. A következő ábrák

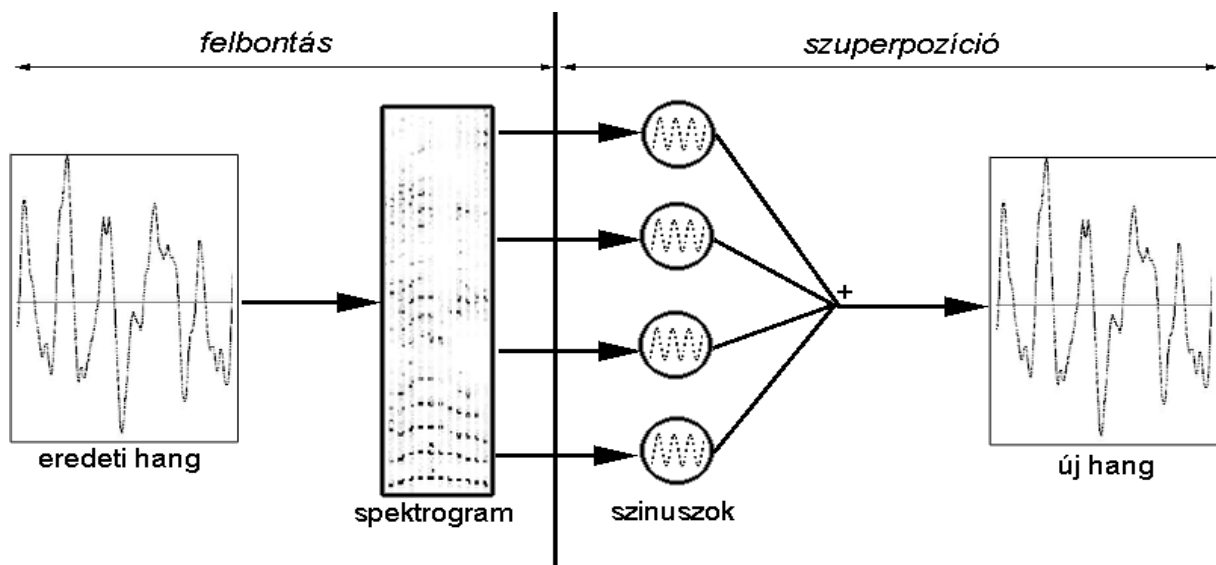
szemléltetik a spektrogramon végrehajtott változást, az első ábra a transzformáció előtti, a második pedig a végrehajtás utáni spektrogramot mutatja.



13. ábra: az eredeti spektrogram, és az egyértelműsített

6. Spektrogram visszaalakítás frekvenciakövetéses simítással

Az eljárás alapvetően a szuperpozíció elvén alapszik: a spektrogram alapján feltárt harmonikus komponenseket (azaz a szinusz függvényeket) összeadjuk. Ezt szemlélteti a következő ábra jobb oldali része, mellyel össze is szeretnénk foglalni az eddigieket:

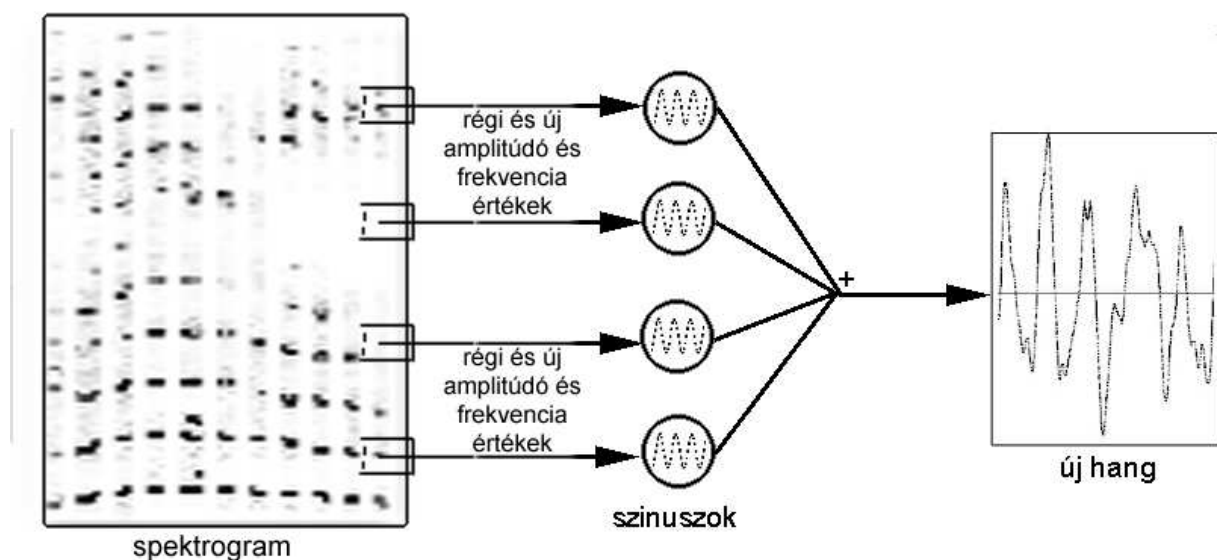


14. ábra: hangátalakítás

A valóság mégis kicsit bonyolultabb: általában a hangok magassága (frekvenciája) és erőssége (amplitúdója) is változik, de ezt folytonosan teszi, azaz követhetően. Itt nyer értelmet az un. vonal egyértelműsítés: követni tudjuk a hangmagasság változását. Egy spektrum vektor visszaalakításakor tudjuk azt is, hogy a megelőző spektrum vektorban milyen magas volt ugyanez a hang, nyomon tudjuk tehát követni. Emiatt úgy tudjuk előállítani a szinuszokat, amiket összeadjunk, hogy azok követik a spektrogramban jelenlévő hangmagasság és amplitúdó változásokat.

Tehát ha egy adott időpontban egy adott frekvenciájú komponens jelen van, a következő időpontban pedig ismerjük ennek a komponensnek a folytatását, akkor a folytatás

fázisa legyen hozzáigazítva az előző fázishoz. Például ha egy hang magassága emelkedik, akkor a két egymást követő spektrum vektorban a megfelelő i -edik illetve $i+1$ -edik érték lesz nem 0 (a szomszédosokat kinulláztuk, így ez követhető). Ekkor azzal a feltételezéssel élünk, hogy az időben későbbi vektor $i+1$ -edik (nem nulla) komponensének értéke az előző vektor i -edik értékének folytatása, és ezért a fázis (és az amplitúdó) folytonosságát biztosítjuk (pl. valamilyen simítással). Ezt szemlélteti a következő ábra: a megelőző vektor bizonyos komponenseit is figyelembe vesszük.



15. ábra: visszaalakítás

6.1. Az algoritmus és magyarázata

Főciklus:

```

1.   Inicializálás();
2.   While (van még spektrumvektor) {
3.       Spektrum=KövetkezőSpektrumVektor();
4.       Visszaalakítás(Spektrum);
5.   }

```

A *Visszaalakítás(Spektrum)* függvény algoritmus C-ben:

A változók jelentése:

amp[i]: az i-edik frekvenciájú szinusz aktuális amplitúdója, statikus (azaz az értékét az eljárás visszatérése után is megőrző) tömb. Ez tartalmazza a kezdőértékeket az amplitúdó simításához (a végérték az új vektor, spektrum[i]).

előző_amp[i]: az előző iteráció utáni amplitúdó vektor eltárolt másolata.

arg[i]: az i-edik frekvenciájú szinusz aktuális fázisa, argumentuma.

arg_növ_akt[i]: az i-edik frekvenciájú szinusz aktuális numerikus szögsebessége, azaz megadja az arg[i] (fázis) aktuális növekedését. Ez a másik mennyiség, amit simítunk (az amplitúdó mellett).

arg_növ[i]: az i-edik frekvenciájú szinusznak a szögsebessége, rögzített érték, nem változik. A szögsebesség vektor „célvektora” a simításkor. Inicializáláskor beállítandók a tömb elemei a megfelelő körfrekvenciákra (amelyek megegyeznek a spektrogram felvételekor használt körfrekvenciákkal, ha nem szándékozunk torzítani a hangot).

out[j]: a kimenő jel, hangminta.

spektrum[i]: az aktuális spektrumvektor. Az amplitúdó simítás „célvektora”.

szorzó[i]: Az egyes frekvenciaértékekhez tartozó amplitúdó szorzókat tároló tömb, használható frekvencia kiemelésre, szűrésre. Inicializáláskor beállítandó.

szinusztömb[k]: A szinusz függvény értékeit tárolja, szinusztömb[0],..., szinusztömb[40000] felbontással. Inicializáláskor kiszámítandó.

frekvenciaSzám: megadja a spektrogram felvételekor használt rezonátorok számát.

Visszaalakítás(spektrum):

```
1.   for (i=0;i<frekvenciaSzám;i+=1){
2.       if(spektrum[i]>0){
3.           spektrum[i] = spektrum[i]*szorzó[i];//szűréshez
4.
5.           if( (i>0) és (amp[i-1]>0)){
6.               arg[i]=arg[i-1];
7.               amp[i-1]=0;
8.               amp[i]=előző_amp[i-1];
9.               arg_növ_akt[i]=arg_növ_akt[i-1];
10.          }
11.          else if((i<N-1) és (amp[i+1]>0)){
```

```

11.         arg[i]=arg[i+1];
12.         amp[i+1]=0;
13.         amp[i]=előző_amp[i+1];
14.         arg_növ_akt[i]=arg_növ_akt[i+1];
15.     }
16. }
17.     if (amp[i]<=0.5)
18.         amp[i]=0;
19. }

20. while(j<hossz){
21.     jel=0;
22.     for (i=0; i<frekvenciaSzám;i+=1)
23.         if(spektrum[i]>0 || amp[i]>0){
24.             amp[i]=0.9*amp[i]+0.1*spektrum[i];
25.             arg_növ_akt[i]=0.9*arg_növ_akt[i]+0.1*arg_növ[i];
26.             arg[i]+=arg_növ_akt[i];
27.             if (arg[i]>40000.0) {arg[i]-=40000.0;}
28.             D=(szinusztömb[(int)(arg[i])]*amp[i]);
29.             jel=jel+D;
30.         }
31.         out[j]=jel;
32.         j=j+1;
33.     }
34.     Átmásol(amp,előző_amp);
35.     return;

```

Magyarázat:

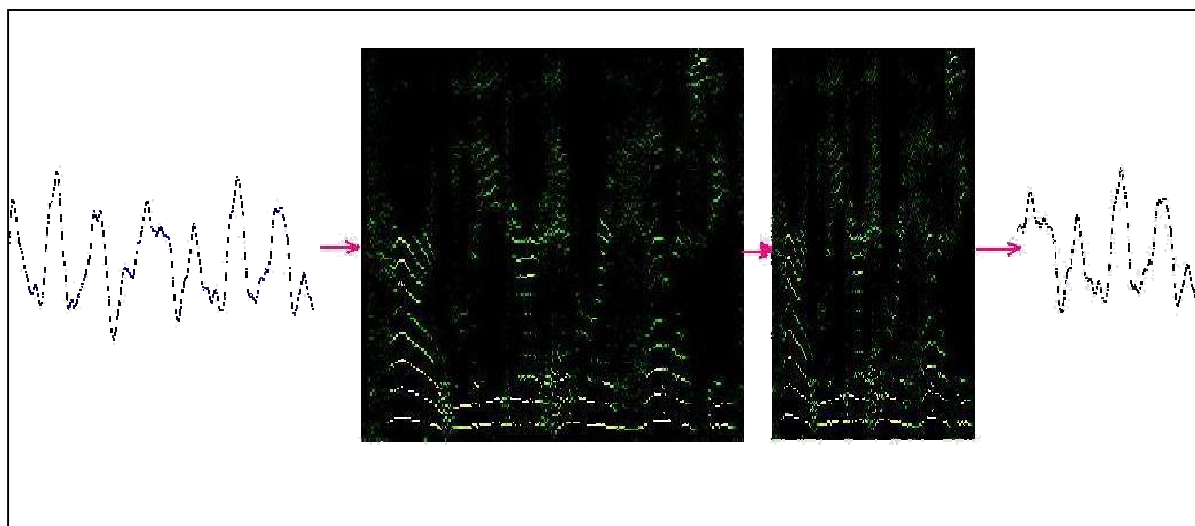
Az 1-19 sorokban a követés és simítás kezdeti értékeit állítjuk be, az előző iteráció befejező értékei alapján. Végigmegeyünk a spektrumvektor összes nem nulla komponensén (legyen az aktuális komponens az i -edik), és ha az előző iterációból megmaradt amplitúdó vektor (`amp[]`) szomszédos($i-1,i+1,i$) komponense nem 0 (4. és 10. sor), akkor ehhez a komponenshez mint kezdő értékhez fogjuk az amplitúdót (7. 13.) és a szögsebességet(8. 14.) hozzásimítani az új iterációban. Hogy csak egy komponenshez simíthassunk, és hogy lebegés ne legyen (egyszerre ne legyen jelen két szomszédos komponens jelen), arról a kinullázások (6. 12.) gondoskodnak. A 17. és a 18. sorokban a kis amplitúdókat kinullázzuk, az algoritmus gyorsítása érdekében. Ezt nyilvánvalóan megtehetjük, mert a kimenő kvantált jelben nem jelentkezik nagy változás.

A 20.-33. sorokban elvégezzük a simítást. A szögsebesség és az amplitúdó simítását a 24. és 25. sorok végzik el. A 26. és 27. sorokban megváltoztatjuk a fázist a szögsebességgel, majd ellenőrizzük, hogy a fázis 40000 értéket túllépte-e (mivel a szinusz tömbünk 40000 elemből áll). A *szinusztomb[0-40000]* tömbben vannak előre kiszámítva a szinusz függvény értékei, ezeket használjuk fel (28. sor) a jel összeállításához (29. sor). Tehát a megfelelő amplitúdóval megszorzott szinuszokat összeadjuk és az output tömbben az összeget eltároljuk (31. sor). A *hossz* változó a két egymást követő spektrumvektor közötti időt adja meg. Ha a *hossz* változó értéke megegyezik azzal az értékkel, ami megadta a spektrogrammá alakításakor azt, hogy hány iterációnként számítsunk spektrum vektort, akkor a kapott hang ugyanolyan sebességű, mint az eredeti. A 35. sorban eltároljuk a simítás végeztével keletkező amplitúdó vektort, hogy a következő iterációban felhasználhassuk (ez lesz a következő simítás kezdőértéke).

7. Felhasználás

7.1. Hangnyújtás

A hang időbeli nyújtása/zsugorítása nagyon egyszerűen oldható meg: Egy megfelelően behangolt rezonátor sokaság szolgáltatson K_1 iterációnként (tehát minden K_1 -edik minta feldolgozása után) spektrumvektort. Ha a visszaalakító algoritmusban minden új spektrumvektor feldolgozásakor K_2 db mintát állítunk elő, és $K_1=K_2$, akkor a kimenő hang ugyanolyan gyors, mint a bemenő. Ha $K_1>K_2$, akkor felgyorsul, ha $K_1<K_2$, akkor pedig lelassul.



16. ábra: hangnyújtás

7.2. Hangtömörítés

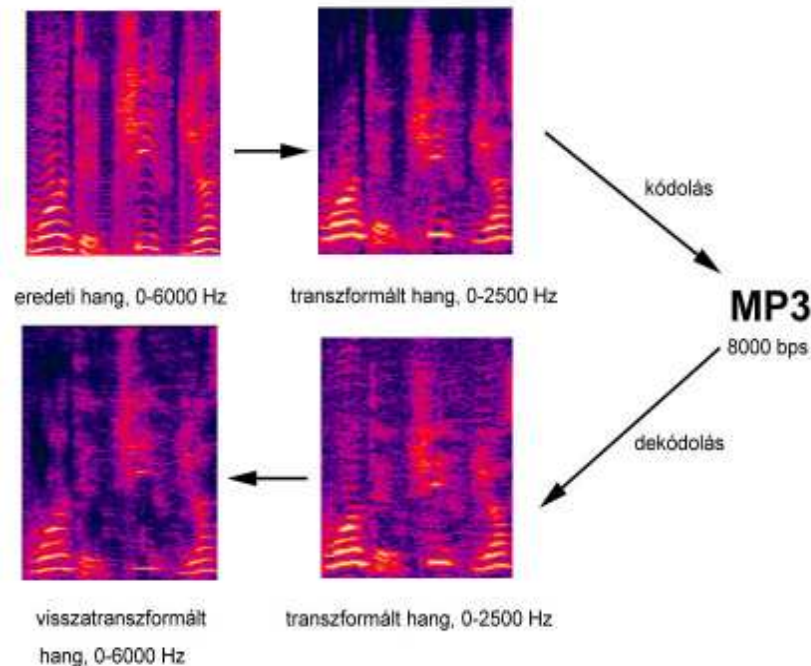
Most csak veszteséges hangtömörítéssel foglalkozunk (veszteségmentesnek akkor nevezünk egy tömörítést, ha az eredeti hang minden egyes mintája pontosan visszaállítható). A tömörítéseknek van egy lényeges paramétere, amely a tömörítés sávszélességét adja meg. Egy 8 kbps-os (8 kilobit per sec.) tömörítés azt jelenti, hogy 8000 bit, azaz 1000 bájt tárterület szükséges 1 másodperces hanganyag tárolására. A legegyszerűbb tömörítés az alulmintavételezés és a kvantálás értékkészletének a csökkentése.

A dolgozat eljárásainak felhasználása hangtömörítéshez több módon is lehetséges:

Egyik megoldás az, hogy a spektrumvektort megfelelően sok iterációnként (150-300) veszünk, és ezen vektorokból is csak az első 20-30 legnagyobb értékű komponenset tartjuk meg, amit még tovább tömörítünk. Ekkor kb. 12kbps-os beszéd-tömörítést érhetünk el, megfelelő minőségben.

A másik megoldás az, ami az előzőnél lassabb, de erősebb, hogy egy hangállomány spektrogramját leképezzük nemlineárisan egy alacsonyabb tartományra. Például a 0-6000Hz tartományt leképezzük nemlineárisan a 0-2000Hz tartományra. Azért képezzük le nemlineárisan, hogy az alsó frekvenciatartomány, ami fontos minden szempontból csak kevésbé transzformálódjon, torzuljon, a távolságok itt csak kicsit változzanak, mert akkor kitömörítésnél pontosabb eredményt kapunk. (Legyen kicsi a relatív hangmagasság hiba.) Tehát a spektrogram leképezése után egy 0-2000 Hz frekvenciatartománybeli komponenseket tartalmazó hanghullámhoz jutunk, amit pl. csak alulmintavételezéssel az eredeti (tfh. 22050 mintavételezésű) hang kb. ötöd részére tömöríthetünk. Ez viszont még csak egy hanghullám fájl, amit tovább lehet tömöríteni más módszerekkel. Visszaalakításkor a rezonátorokat most a 0-2000 Hz tartomány megfelelő frekvenciaértékeire beállítva, egy újabb nemlineáris

transzformációval (az előző inverzével) a komponensek frekvenciáit visszaállítjuk. Például használhatjuk az mp3 tömörítést. Ekkor a transzformáció utáni mp3 tömörítéssel 8 kbps-os (a 8kbps-os mp3 kb. 2500 Hz-nél levágja a spektrumot) tömörítés érhető el, elfogadható minőségben. (Talán mp3 helyett más tömörítő eljárást használva többet is el tudunk érni.)



17. ábra: hangtömörítés frekvenciatartomány transzformációval

7.3. Különböző effektusok

Mindenféle olyan effektusokat lehet létrehozni, amiket a spektrogramon mint képen végrehajthatunk, azaz mindenféle nemlineáris leképezés, frekvenciatartomány kivágás, eltolás, nyújtás, szűrés, stb. A sebesség változtatását már említettük. Érdekes effektust lehet létrehozni a csillapítás kicsire állításával, ami a beszédet szinte zenévé alakítja. Ha nem szinuszos függvényeket adunk össze visszaalakításkor, az a hangszín megváltozását

eredményezi. Megoldható az egy hangból többet előállító kórus effektus is, visszhang hozzáadása, illetve a hangmagasság csökkentése vagy növelése. A hangmagasságot úgy változtathatjuk meg, ha a spektrogramot a frekvenciatengely irányában nyújtjuk. Van egy egyszerűbb módja is: a visszaalakításkor összeadandó szinusz függvények szögsebességei legyenek a spektrogram felvételekor használt körfrekvencia értékek konstansszorosai. Láthatjuk tehát, hogy nem csak a spektrogramon ügyeskedhetünk, hanem a visszaalakító algoritmus paramétereit is okosan változtatgathatjuk, akár menet közben. A visszaalakító algoritmust kell kiegészíteni pár sorral ahhoz is, hogy visszhang effektust kapjunk.

7.4. Beszéd előállítás, felolvasó program

Végül, egy újszerű felolvasó programot is bemutatunk az elmülethez kapcsolódó gyakorlati lehetőségek demonstrációjaként.

A rendszer a következő részekből áll: egy hangadatbázisból, amit bemondott szavakból vágunk ki, ez csak részben egyezik meg a magyar fonémakészlettel. Ez körülbelül a következő elemeket tartalmazza:

- magánhangzók (a, á, ...),
- magánhangzó párok(aá ,..., eo, ...),
- egyes mássalhangzók önmagukban(s, sz, z, cs, c, t,...),
- zöngés (pl. „d” előtt) és zöngétlen (pl. „t” előtt) szünet(jele: +, -),
- j, l, m, n hangok difónban (kettes hangkapcsolatban) magánhangzókkal (mivel itt a hangok egymásra hatása nem elhanyagolható, azaz a száj mozgataa folytonos és nem gyors): (aj, ja, ..., en, ..., mu), stb.

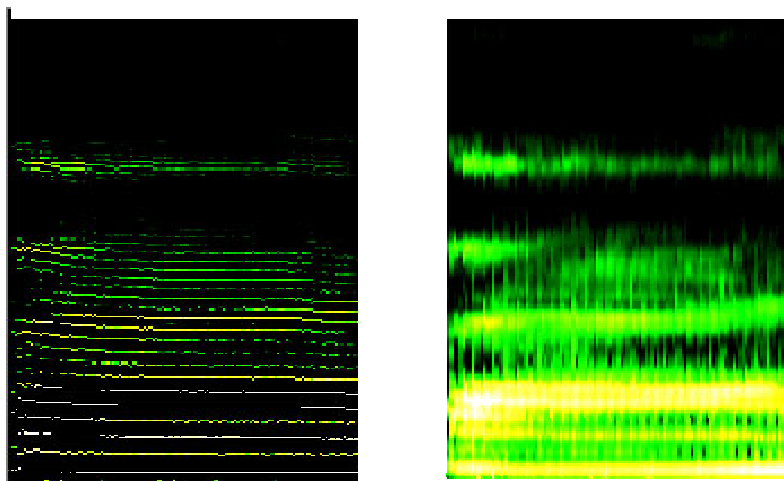
A rendszer következő része a hangmagasság meghatározó program, ami a zöngés hangok magasságát becsli, és kiírja egy fájlba. Ezt úgy teszi meg, hogy egy rezonátorrendszer van beállítva az alacsony frekvenciatartományban (40-200 Hz); a spektrum legnagyobb komponenséhez tartozó frekvencia adja az alaphang magasságát. Egy másik módszerrel is lehet becsülni a hangmagasságot, ekkor a frekvenciatartomány szélesebb, és nem csak az első maximumot vesszük figyelembe, hanem mondjuk az első hármát (felharmonikusokat). Ez azért tesz pontosabb hangmagasságbecslést lehetővé, mert a relatív hibát csökkenti, és elkerülhető egy esetleges al hangmagasság becslés (egy erős alacsony frekvenciájú hang esetén).

A harmadik egység a zöngés hangokat azonos hangmagasságra konvertáló program. Ezt egyszerű túl- vagy alulmintavételezéssel is megoldhatnánk, de ez a hangot torzítja. Ezért csak az a megoldás jó, amelyik a formánsok frekvenciatartományát ugyanott tartja. Ezért először egy adott magánhangzó éles spektrogramja mellett felvesszük az abból konvolúcióval erősen elkent képét is. Ezt a formánsokat megmutató spektrogramot nevezzük most el karakterisztikának. (Formánsoknak a karakterisztika nagyobb léptékű maximumhelyeit nevezzük. A magánhangzók felismerése jórészt a formánsok számával, helyével, szélességével és intenzitásával áll kapcsolatban. A karakterisztika tulajdonképpen azt tükrözi, hogy a szájüregben lezajló erősítések és gyengítések hatására a szájból kijövő hangban milyen mértékben erősödnek fel a különböző frekvenciájú komponensek, amit a nyelv helyzete, a száj nyitottsága, stb. határoz meg.) A hangmagasságra korrigált spektrumot úgy kapjuk, hogy az eredetit leosztjuk a karakterisztikával, a kapott spektrogramot a hangmagasság változtatásának megfelelően transzformáljuk, majd a karakterisztikával visszaszorozzuk. Tehát ekkor a végeredmény egy-egy spektrogram minden zöngésre, azonos hangmagasságon. A spektrogramokat zöngétlenekre is előállítjuk.

Ekkor kész van a felolvasó programhoz szükséges hang adatbázis.

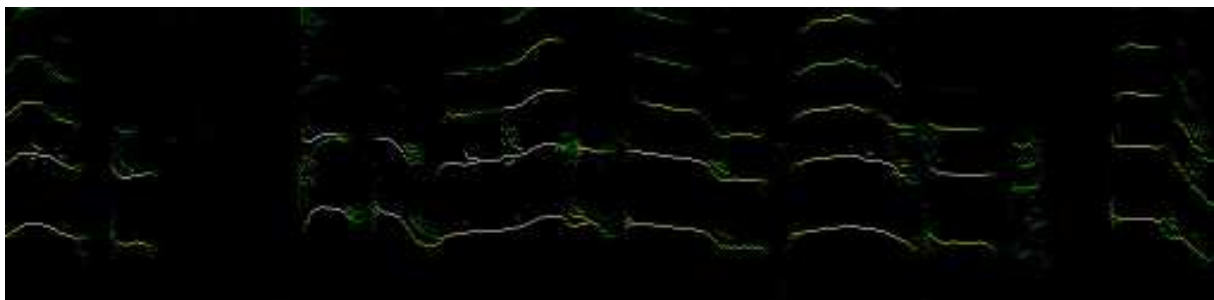
A felolvasó program maga az előző részben leírt követéses simítás szerint működik. A szövegfájlt könnyen a megfelelő egységekre, hangokra bontja, és ezután egymás után végigmegy a hangoknak megfelelő spektrogramokon, persze a különálló egységek között is simítva, így az „összeácsoltság” nem érzékelhető. A hangmagasság változtatásra menet közben a már előre kiszámított karakterisztikák segítségével nyílik lehetőség. Természetesen a sebesség változtatása nem okoz gondot.

A következő ábra egy szóból kivágott „á” magánhangzó spektrogramját és karakterisztikáját mutatja:



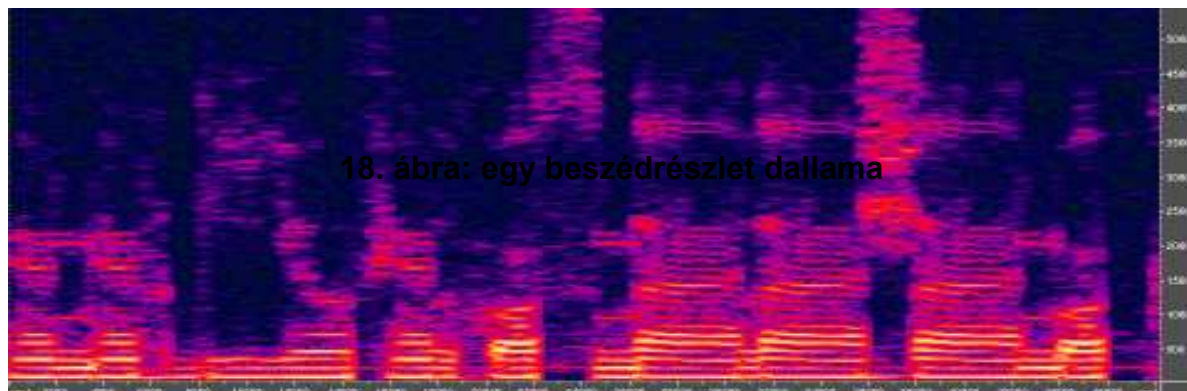
18. ábra: zöngés hangból nyert karakterisztika

A következő ábra pedig egy beszédrészlet dallamát szemlélteti:



19. ábra: dallam

A felolvasó program egy kimenetének spektrogramját mutatja a következő ábra (a szó: „energiafelhasználásának”):



20.-ábra: a beszéd előállító program egy kimenetének spektrogramja

8. Irodalom

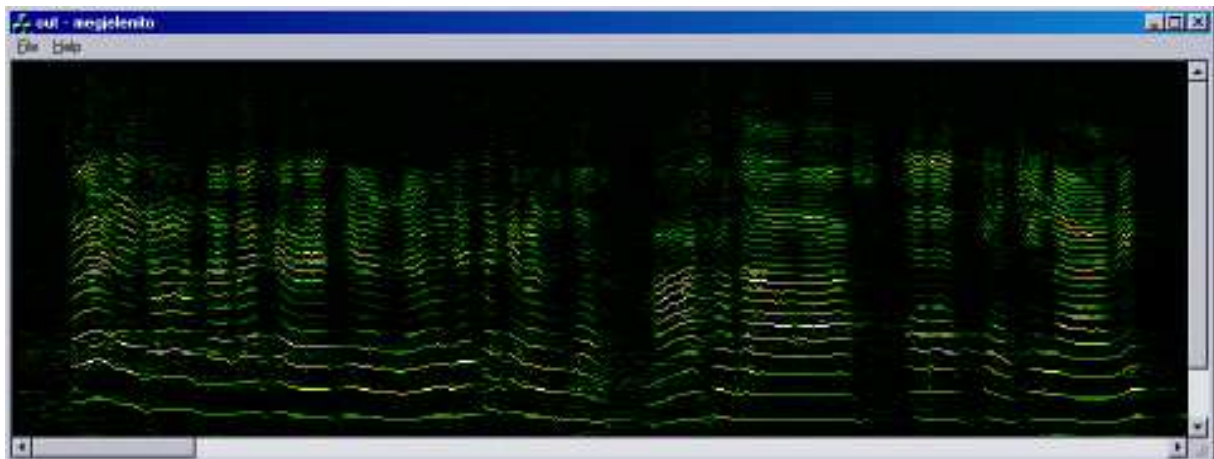
1. Budó Ágoston: Kísérleti fizika I., Nemzeti tankönyvkiadó, Budapest, 1979.
2. Budó Ágoston: Mechanika, Nemzeti tankönyvkiadó, Budapest, 1979.
3. A. V. Oppenheim- R. W. Schaffner: Discrete-time Signal Processing, Prentice Hall, Englewood Cliffs, New Jersey 1989.
4. Angster Judit, Arató Éva: Akusztikai példatár, Akadémiai kiadó, Budapest, 1986.
5. N. Sz. Bahvalov: A gépi matematika numerikus módszerei, Műszaki könyvkiadó, Budapest, 1977.
6. Anthony Ralston: Bevezetés a numerikus analízisbe, Műszaki könyvkiadó, Budapest, 1969.
7. Terjéki József: Differenciálegyenletek, POLYGON, Szeged, 1997.
8. Móricz Ferenc: Differenciálegyenletek numerikus módszerei, POLYGON, Szeged, 1998.
9. Bajusz-Bors-Csibra-Horváth: A PC-k hangja, COM-WARE, 1995
10. Proakis–Manolakis: Introduction to Digital Signal Processing, MacMillan, New York, 1989.
11. Székely Vladimír: Képkorrekció, hanganalízis, térszámítás PC-n, ComputerBooks, Budapest, 1994.

Melléklet:

Egy CD lemezen mellékeljük a demonstrációs hangmintákat és a példaprogramokat. A hangminták meghallgatásához az index.html lapról induljunk el. A hangok elérhetőek a következő címen is: <http://oasis.inf.u-szeged.hu/Banhalmi.Andras>

Három példaprogramot mellékelünk: egy felolvasó programot, egy hangtranszformáló programot és egy spektrogram megjelenítő programot.

Megjelenito.exe: megjeleníti a .spd kiterjesztésű saját spektrogram fájlokat. Ilyeneket például az rthang.exe program segítségével hozhatunk létre. Egy kép róla:



Betuz.exe: felolvasó program, egyszerűbb demonstrációs célra. Egy sima szövegfájl szavait egyesével felolvassa. Kicsi és rossz minőségű adatbázisa miatt a hangminőség nem az igazi.

Rthang.exe: hangátalakító program. A funkcióit tekintsük meg a saját helpjében. Nemlineáris hangspektrum transzformációt az input és output frekvenciaskála megfelelő beállításával érhetünk el. A programról egy kép, amelyen a főbb funkciók is láthatók:

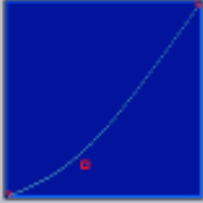
RTSound

Állások


alacsony 1.000000
 magas 1.000000
 alacsony 1.000000
 szűk 1.000000
 tozható 0.000000
 kábel magas 1.000000

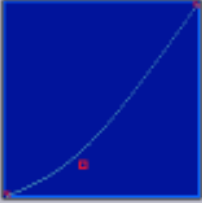
Viszhang

bal 15v. 50.00000
 bal1. energi. 0.000000
 bal2. 15v. 50.00000
 bal2. energi. 0.000000
 Kézi
 Buffer méret 12000
 Belső mintavétel 80



bemeneti frekvencia leképezés





kimeneti frekvencia leképezés

Spektrogram fájl kimenet:

PCM wav fájl kimenet: