

Debreceni Egyetem
Informatikai Kar

Akusztikai jelek elemzése

Témavezető:
Dr. Papp Zoltán
egyetemi adjunktus

Készítette:
Balázs Tamás Zsolt
programozó matematikus

Debrecen
2010

Bevezetés	1
Hangtani alapfogalmak.....	1
A hangmagasság és hangosság pszichofizikája	6
Hangmagasság	6
Kritikus sáv szint és gerjesztési szint	9
Komplex hangok hangmagassága	10
Hangosság	12
Modellek, algoritmusok.....	16
Hangmagasság	16
Időtartomány alapú módszerek	17
Frekvenciatartomány alapú módszerek	18
Statisztikai módszerek	19
Autokorrelációs módszerek a hangmagasság meghatározására	20
Algoritmus	23
Hangossági modellek implementálása Matlab-ban.....	25
Modell implementációk	27
A hangosságrepresentáció direkt implementációja.....	27
A Zwicker-modellen alapuló implementációk.....	28
A DIN45631 / ISO532B hangosságmodell implementációja	28
.....	30
Zwicker hangossági modellje, ahogyan a PSQM-ben használják.....	31
Zwicker hangossági modellje, ahogyan a PEAQ-ban használják	31
Moore és Glasberg hangossági modellje	32
A kimenet kalibrálása és tesztelés	33
Nyelvészeti fogalmak.....	34
Intonáció	34
Intonáció PC-vel.....	36
Alkalmazásfejlesztés	38
A fejlesztői környezet: Matlab 6.5 Release 13	38
Intonációs és hangossági görbék megjelenítése.....	40
Felhasznált irodalom	42

Bevezetés

A digitális jelfeldolgozás egy kiemelten fontos területe a hangfeldolgozás. Ezen belül egy szintén fontos részterület a hangjelek analízise, azok a vizsgálatok, amelyek az érzékeléssel kapcsolatosak. Ezeknek a tulajdonságoknak a mérése során meg kell vizsgálni, hogy hogyan függ az érzet nagysága a fizikai mennyiségtől. Sok alkalmazásban játszik központi szerepet a hangmagasság és a hangosság. Az irodalom számtalan modellt közöl, egyik jobb, másik kevésbé, egyik a beszédjelek esetén vált be, másik a zenei felvételek esetén. Némelyik modellt szabványként fogadták el, de a szubjektív komponens jelenlétéből adódóan állandó felülvizsgálatra szorul. A dolgozat eleje a fizikai hangtani bevezető után a két érzetre vonatkozó modelleket ismerteti bemutatva azt a sokszínűséget, amely e témát jellemzi. Az egyes modellek alapján algoritmusok is születtek. A dolgozat ezekről is ad némi betekintést. Mindkét érzet esetén kiválasztottunk egy-egy algoritmust, amelyeket részletesen ismertetünk. A hangmagasságot meghatározó algoritmus az autokorreláción alapul, a hangosságot a szabványként is elfogadott Zwicker-féle modellen alapuló algoritmus számítja ki.

Érdekes és hasznos alkalmazások középpontjában szerepel a hangmagasság és a hangosság. A dolgozat csupán egyet mutat be: hogyan lehet az idegen nyelv tanulását, ezen belül is a helyes intonáció és hangsúlyozás elsajátítását számítógéppel támogatni. Készült egy ilyen alkalmazás, amelynek ismertetése a dolgozat végén szerepel.

Hangtani alapfogalmak

Rezgés: valamely fizikai mennyiség időben periodikus ingadozása, az idő periodikus függvénye. A szinuszos vagy harmonikus rezgés a legegyszerűbb rezgés. Vannak nem harmonikus, ún. összetett rezgések is. Ha egy rezgő rendszert magára hagyunk, a rezgés csillapodik a rendszer mechanikai energiáját csökkentő súrlódás miatt. Ha csillapítatlan rezgést szeretnénk, akkor valamilyen külső energiaforrásból pótolnunk kell a felemészített energiát.

Fourier-tétel: általában bármely T periódusú $f(t)$ periodikus függvény ($T > 0$, $f(t+T) = f(t)$, T a legkisebb ilyen időköz) egyértelműen előállítható megfelelő amplitúdókkal és fázisállandókkal bíró szinuszos összetevők összegeként, melyek ún. körfrekvenciái a rezgés körfrekvenciájának ($\omega = 2\pi/T$) egész számú többszörösei: $f(t) = c_0 + c_1 \sin(\omega t + \gamma_1) + \dots + c_n \sin(n\omega t + \gamma_n)$. Ez a Fourier-sor. Ha $f(t)$ Fourier-sorában a c_n amplitúdóit a fázisállandók

figyelman kívül hagyásával ábrázoljuk az n függvényében, akkor megkapjuk az $f(t)$ periodikus függvény vonalas színeképét, spektrumát.

Hang: többféle értelemben, illetve több fogalom gyűjtőneveként használatos. Fizikai fogalomként az észlelő tudatától függetlenül meglévő hangjelenséget, élettani (fiziológiai) és lélektani (pszichológiai) fogalomként hangérzetet, hangélményt fejez ki. Hangérzetet általában megfelelő amplitúdójú mechanikai rezgések váltanak ki, frekvenciájuk 20 Hz és 16 kHz közötti. Alkalmas rezgő rendszerből, hangforrásból indulnak ki és rugalmas közegben, például levegőben terjednek.

Tágabb értelemben a hangtan több területet, illetve határterületet felölel, élet- és lélektani, zenei és gyakorlati kapcsolatai is vannak. A hangtannak tehát mindezekkel kapcsolatos területei is vannak: fizikai, fiziológiai hangtan, zenei, építészeti hangtan (a teremakusztika ennek része), elektroakusztika.

Hangérzet: szubjektív hangbenyomás, vannak zenei hangok, zörejek és dőrejek. A zeneiek periodikusak, a zörejek nem, a dőrejek pedig lökésszerű hanghullámok. A Fourier-tétel alapján a zenei hang spektruma valamely v frekvenciájú alaphangból és $2v$, $3v$, ... frekvenciájú felhangokból áll. A zörejek spektruma általában sok rezgésből áll, ezek frekvenciáinak arányai nem racionális számok. A tisztahang egy zenei hang, harmonikus, azaz szinuszos rezgés felel meg neki. Színekpe egyetlen vonal. A zenei hangok érzet szerint különbözhetnek hangosság, magasság és színezet tekintetében. A hangerő nagyrészt az amplitúdó függvénye. A tér valamely pontján a hangintenzitás a hanghullám intenzitásával egyezik meg.

Hangmagasság: az alaprezgés frekvenciájától függ, magasabb hangnak nagyobb frekvencia felel meg. Fülünk hangmagasság-megkülönböztető képessége bizonyos tartományban még két 0,3%-os frekvenciakülönbségű hangot is képes külön felismerni, azaz a fül spektrális felbontóképessége 300. A hangmagasság érzetének kialakításához mindössze 3 periódusnyi idő is elég, ez pl. 1 kHz esetében 0,003 s időtartamú hang.

A hangmagasság különbségek érzékelésének kifejezésére két fogalmat használunk: harmonikus, illetve melodikus hallás. A harmonikus hallás esetén két hang távolságát a v_2/v_1 frekvenciahányadosukkal fejezzük ki, a 2:1 arányú hangköz az oktáv, ilyenkor azt mondjuk, hogy a magasabb a mélyebbnek az oktávja. Egymástól oktávnyira levő hangokat nagyon hasonlóknak érzünk. Az átlagos hallástartomány (20 Hz-től 16 kHz-ig) közel 10 oktávnak

felel meg. A beszédhang általánosan 1 oktáv terjedelmű, a zenei hangoké nagyjából 7 oktávnyi.

Hangszín: különböző hangforrások azonos alaphangfrekvenciájú hangjai másképpen szólnak, ez a felharmonikusok frekvenciájától és viszonylagos erősségétől, az ún. spektrumtól függ. Tehát a Fourier sor $c_1, c_2, \dots$ amplitúdóinak aránya befolyásolja, a γ_i fázisállandók általában nem. A hangszínt meghatározó spektrum a hangforrás természetén kívül függ a hangkeltés módjától is, egy húr pengetve illetve vonóval más és más hangot ad. Ha a mélyebb felhangok uralkodnak, akkor a hangot tompának találjuk, ha a magasabbak, akkor élesnek. A szinuszrezgésnek megfelelő, felhangok nélküli tisztahang üresnek, színtelennek hat.

Hangtér: a térnek hanghullámokkal betöltött része, amelyben anyagi részecskék rezgő mozgást végeznek egyensúlyi helyzetük környezetében, közben nyomás-, sűrűség- és hőmérsékletingadozások lépnek fel és terjednek tovább hullámszerűen.

Hangenergia-sűrűség: a térfogategységben foglalt hangenergia időbeli középértéke.

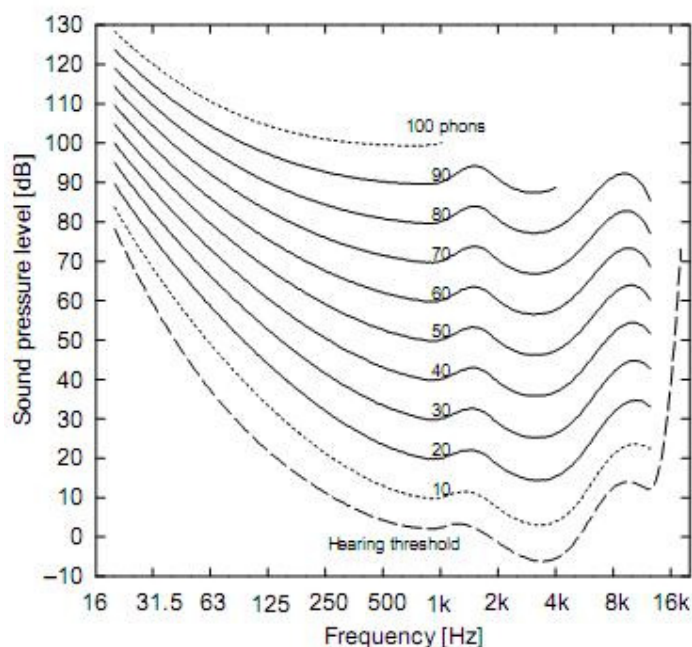
Hangintenzitás: nevezhetjük objektív hangerőnek is ingeroldalról vizsgálva. A hangenergia-áram sűrűsége, egységnyi felületen át, merőleges irányban időegység alatt átviramló hangenergia időbeli középértéke. Dimenziója: teljesítmény/felület, mértékegysége W/m^2 . 1000 Hz-es tiszta hang éppen meghallható intenzitása az ún. ingerküszöb: $I_0 = 10^{-12} W/m^2$. Az érték megállapítása óta az emberek érzékenysége romlott, ez az érték 5 dB-lel emelkedett.

Hangteljesítmény: $P = I \cdot q$, azaz egy kisméretű q felületen időegység alatt átviramló hangenergia. Hangforrás hangteljesítménye a hangforrás által időegység alatt minden irányban kisugárzott hangenergia. Mivel a különböző források elég széles energia tartományban sugároznak, ezért két teljesítményt vagy intenzitást úgy hasonlítottunk össze, hogy hányadosuk tízes alapú logaritmusát vesszük. Azt mondjuk P_1 és P_2 teljesítményszintjeinek vagy intenzitás szintjeinek különbsége: $n = 10 \cdot \lg \frac{P_2}{P_1}$ dB.

Hangosság: érzetoldalról vizsgálunk, nevezhetjük szubjektív hangerőnek is. Ez a megállapítás általánosságban is igaz. A pszichofizika alaptörvénye az inger és érzet közti kapcsolatot írja le. Ez logaritmikus, ha az érzet az inger logaritmusának lineáris függvénye. 150 évig volt uralkodó elv a Weber-Fechner-féle pszichofizikai törvény, amely szerint a hangintenzitás mint inger és hangosság mint érzet között logaritmikus összefüggés áll fenn. Az inger és érzet közötti összefüggés ún. hatványösszefüggés, ha az érzet logaritmus az inger

logaritmusának lineáris függvénye. A tudomány mai állása szerint a hangintenzitás és a hangosság között ilyen jellegű az összefüggés. A hangosság megállapításához a hang intenzitásszintjéből ($10\lg(I/I_0)$, egysége a dB) indulnak ki. Ha egy tetszőleges frekvenciájú hallható hang hallás alapján egy I intenzitású 1 kHz-es tisztahanggal egyező hangosság érzetet kelt, akkor a kérdéses hang phon-hangossága: $H=10\lg\frac{I}{I_0}$ phon. Eszerint az ingerküszöbnek megfelelő intenzitású hang magasságától függetlenül 0 phon hangosságú. Az ingerküszöb ellentéte a fájdalomküszöb, amelyhez tartozó hangosság 130 phon.

A hangerősség-hangosság összefüggésben a frekvencia szerepét és a dB és phon-skála közti különbséget jól jelzik az egyenlő hangossági görbék. Az abszcissza a frekvencia, az ordináta a decibelben kifejezett intenzitásszint. Jól látható, hogy az 1 kHz-es hangok dB intenzitásszintje és phon-hangossága azonos, más frekvenciákra azonban ez nem áll fenn. Az ábra szerint 100 Hz-nél 0 phon hallásküszöbnek 0 dB intenzitásszint helyett 40 dB felel meg. Itt tehát a fül kisebb érzékenysége miatt a hangérzet keletkezéséhez I_0 helyett $10^4 \cdot I_0$ intenzitás szükséges. Hasonlóan, egy 100 Hz-es hang akkor 40 phon hangosságú (akkor hallatszik ugyanolyan hangosnak, mint egy 40 dB intenzitásszintű 1 kHz-es tisztahang), ha intenzitásszintje 60 dB. Az egyenlő hangossági görbéket az évek során sok kritika érte, amelynek következtében apróbb módosításokra került sor. A mellékelt ábra egy már letisztult állapotot tükröz.



A phon-skála a logaritmikus összefüggés skálája. Ez nem egyenletes skála, nem arányos a phon érték az érzettel. Egyenletes skálához a hatványösszefüggésre való áttéréssel jutunk.

Ez az ún. son-hangosság skála:

phon	0	10	20	30	40	50	60	70	80	90	100	110	120	130
son	0,001	0,05	0,25	0,5	1	2	4	8	16	32	64	128	256	512

Elfedés: két hang egyidejű hangzásánál fordulhat elő, a nagyobb hangosságú megnehezíti vagy teljesen ellehetetleníti a kisebb hangosságúnak az érzékelését.

A hangmagasság és hangosság pszichofizikája

Hangmagasság

Tiszta hangok magassága különböző eljárásokkal mérhető. Egyik lehetőség: egy f frekvenciájú tiszta hang mellé az $\frac{1}{2}f$ frekvenciájú hangot úgy kell beállítania, hogy az általa hallott hang magasságát feleakkorának érezze. Ha egy tiszta 440 Hz-es hangot hall elsőként, illetve egy változó frekvenciájú tiszta hangot másodikként, és ezeket felváltva hallgatja, akkor a másodikat 220 Hz környékére állítja be. Azaz alacsony frekvenciákon egyenes arányban változik a frekvencia és a hangmagasság. Magasabb frekvenciákon váratlan hatások jelennek meg. $f_1=8$ kHz esetén f_2 -t nem 4 kHz környékére állítják az alanyok, hanem 1,3 kHz környékére. Bár nagy egyéni eltérések tapasztalhatóak, sok kísérlet igazolja az 1,3 kHz környéki beállítást. Egyéb 1 kHz feletti mérések is ezt a tendenciát igazolták: fele hangmagassághoz 2:1-nél nagyobb frekvenciaarány szükséges. Ugyanilyen módszerrel meghatározható a kétszeres hangmagasság érzete is. Mindkét esetben állandó ingert használnak leggyakrabban. Nagy frekvenciák esetén keskenysávú zörejt is használatos tiszta hang helyett. Az érzetek felezéséből, illetve megduplázásából kikísérletezhetőek az arányok, de ezek nem abszolút értékek. Ehhez referenciapontot kell definiálnunk az „arány-hangmagasság” érzetére a frekvencia függvényében: 1000 Hz = 1000 mel. Ez a skála kapta a mel nevet.

Tiszta hangok magassága mástól is függ, például a hangnyomásszinttől. Egy f_L frekvenciájú és L intenzitású tiszta hangot mérünk össze egy 40 dB-es és $f_{40\text{dB}}$ frekvenciájú hanggal. Egy 200 Hz-es hang 80 dB-en kisebb hangmagasságúnak tűnik, mint 40 dB-en. Ugyanezt ugyanígy 6 kHz-re végrehajtva éppen fordított eredményt kapunk: 80 dB-en nagyobb hangmagasságú, mint 40 dB-en. Tehát a hangmagasság megállapításához nem elég a frekvencia, kisebb mértékben függ a hangnyomás-szinttől is.
$$V = \frac{f_{40\text{dB}} - f_L}{f_L}$$

sokszor és sok alannal megismételték, ezek átlagadatok, egyedi esetek ettől erősen eltérhetnek. Az eredmények azt mutatják, hogy 40 dB nyomásváltozás kb. 3%-kal tolja el a hangmagasságot. Ez a legtöbb esetben elhanyagolható. Ha mégis pontos meghatározás szükséges, mind a hangnyomás-szintet, mind a frekvenciát meg kell adnunk.

A kritikus sávok fogalmát Fletcher vezette be. Azt feltételezte, hogy a tiszta hangot elfedő zaj annak spektrumából való. Ahhoz, hogy relatív értékek helyett abszolútakat kapjunk, további feltételezések szükségesek: elfedés akkor lép fel, ha az elfedő (maszkoló) és a maszkolt hang azonos erősségű. A zaj azon része, amely a teszthang spektrumától távol esik, nem vesz részt a maszkolásban.

A fehér zaj maszk küszöböt hoz létre, amely frekvenciafüggő, noha annak sűrűség szintje független a frekvenciától. Mindez 500 Hz-ig igaz, 1 kHz felett a görbe tízedenként 10 dB-es növekedésbe kezd. Hallórendszerünk relatív frekvenciaszelektivitásáról elmondható, hogy relatíve szűk frekvenciasávokban dolgozza fel a hangokat. Ha feltételezzük, hogy hallórendszerünk maszk küszöböt hoz létre függetlenül a frekvenciától, akkor az általunk keresett frekvenciasávoknak 500 Hz alatt frekvenciafüggetleneknek kell lenniük, akárcsak a fehér zaj sűrűségének. Ennélfogva a kritikus sávnak konstans szélességűnek kell lennie. Magasabb frekvenciákon a maszk küszöb 10 dB/tized értékkel növekszik, azaz az intenzitás sávon belül frekvenciaarányosan erősödik. Vagyis a frekvencia tízszeres növekedésére a sáv szélességeknél is tízszeresére kell módosulnia. Feltételezve – ahogyan Fletcher is tette – hogy egy hang zajban akkor hallható, ha hangerejük megegyezik (a zaj az elfedett hang középpontú kritikus sávban helyezkedik el), akkor a keresett sáv szélesség így becsülhető: 500 Hz alatt a maszkolási küszöb 17 dB-lel magasabb a hangot elfedő fehér zaj sűrűségénél. E sávon belül azonos hangerőt feltételezve mind a hang mind a küszöb esetén, sáv szélességként $10^{\frac{17}{10}}$ értéket kapunk, azaz nagyjából az 1 Hz ötvenszeresét. Ez alacsony frekvenciákon 50 Hz sáv szélességhez vezet. Azonban hallórendszerünk maszk küszöb létrehozó tulajdonságának az elfedni kívánt hang frekvenciájától való függetlenségéről vallott elképzelésünk hibás. A hang maszk küszöbnél vett ereje csak fele/negyede a sávba eső zaj erejének. Ezt a sáv szélességekről szóló kiegészítő információt felhasználva a kritikus sávok elég pontosan meghatározhatóak: 500 Hz alatt 100 Hz szélességűek, e fölött pedig a középfrekvencia 20%-ával egyeznek meg, azaz frekvenciafüggően növekednek. A fenti feltételezések mellett számos direkt módszer is létezik a kritikus sávok és azok sáv szélességeinek meghatározására.

Kritikus Sáv Arány Skála: sok alanytól gyűjtöttek adatokat, mérési eredményeket a kritikus sávok szélességének elfogadható mértékű becsléséhez. Beigazolódott, hogy alacsony frekvenciákon a közeg frekvenciaválasztát figyelembe kell vennünk, ha értelmes adatokat szeretnénk kapni. Úgy tűnik 200 Hz alatt a legmegbízhatóbb az FM–AM váltáskor

érzékelhető fáziseffektusokat használó módszer. Bár a hallható tartományban a legalacsonyabb kritikus sáv nagyon közel esik 80 Hz-hez, csábítónak tűnik a 0..20 Hz nem hallható tartomány hozzáadása a kritikus sávhoz és annak feltételezése, hogy a legalsó kritikus sáv 0..100 Hz közé esik. 70 dB szint felett kis változások fordulnak elő, ennek ellenére a kritikus sávok szélessége a frekvencia függvényében jó közelítéssel határozható meg. 500 Hz-ig 100 Hz szélességű marad. 500 Hz és 3 kHz között kicsit lassabban növekszik, mint a frekvencia függvényében várható lenne, 3 KHz felett pedig valamivel gyorsabban az elvárhatónál. Hasznos lehet konstans 100 Hz-et feltételezni 500 Hz-ig, aztán 20%-ot az e fölötti középfrekvenciákra. A következőkben megadott táblázat precízebb, a kritikus sávok alsó és felső határát adja meg, ha azok úgy helyezkednek el egymás mellett, hogy egy alacsonyabb kritikus sáv felső vágási frekvenciája megegyezik a következő sáv alsó vágási frekvenciájával. A kritikus sávok fontosak a hallás érzetének magyarázatában. Annyi hipotézisben és modellben szerepelnek, hogy mértékegység alapjául is szolgálnak, ez vezetett a kritikus sáv arány skála fogalmának létrejöttéhez. A skála azon a tényen alapszik, hogy hallórendszerünk széles spektrumot bont fel kritikus sávokra. Ha azok az előbb leírt módon kapcsolódnak egymáshoz, akkor egy bizonyos frekvencia felel meg minden metszéspontnak. Ezeket pontokkal ábrázoljuk. Észrevehető hogy a hallható frekvenciatartomány 24 kritikus sávra osztható.

Ez 0-tól 24-ig növekszik és Bark a mértékegysége (a nevét egy Barkhausen nevű kutatóról kapta, aki

z Bark	f_l, f_h Hz	f_c Hz	z Bark	Δf_G Hz
0	0			
		50	0,5	100
1	100			
		150	1,5	100
2	200			
		250	2,5	100
3	300			
		350	3,5	100
4	400			
		450	4,5	110
5	510			
		570	5,5	120
6	630			
		700	6,5	140
7	770			
		840	7,5	150
8	920			
		1000	8,5	160
9	1080			
		1170	9,5	190
10	1270			
		1370	10,5	210
11	1480			
		1600	11,5	240
12	1720			
		1850	12,5	280
13	2000			
		2150	13,5	320
14	2320			
		2500	14,5	380
15	2700			
		2900	15,5	450
16	3150			
		3400	16,5	550
17	3700			
		4000	17,5	700
18	4400			
		4800	18,5	900
19	5300			
		5800	19,5	1100
20	6400			
		7000	20,5	1300
21	7700			
		8500	21,5	1800
22	9500			
		10500	22,5	2500
23	12000			
		13500	23,5	3500
24	15500			

bevezette a phon-t, azt a hangosság szintet leíró értéket, amelyben a kritikus sáv fontos szerepet játszik). z , a kritikus sáv mértéke illetve f , a frekvencia függésének vizsgálata fontos az emberi fül számos karakterisztikájának megértésében.

Kritikus sáv szint és gerjesztési szint

Hallórendszerünk frekvenciaszelektivitása alapján a hangintenzitást olyan részekre bontjuk, amelyek különböző kritikus sávokba esnek. Ez a megközelítés vezet a kritikus sáv intenzitásának fogalmához. Ezek meghatározásához kritikus sáv szűrőket használunk, amelyek kimeneteként átlagos intenzitást kapunk, mint ingert. Eme értékek leginkább logaritmikusan (és nem lineárisan) értelmezettek. Kritikus sáv intenzitása:

$$I_G(f) = \int_{f-0,5\Delta_{G(f)}}^{f+0,5\Delta_{G(f)}} \frac{dI}{df} df, \text{ vagy (mivel } z \text{ és } f \text{ függ egymástól): } I_G(f) = \int_{z-0,5\text{Bark}}^{z+0,5\text{Bark}} \frac{dI}{dz} dz.$$

Logaritmussal és $I_0=10^{-12} \text{ W/m}^2$ referenciaértékkel kifejezve a kritikus sáv szintje:

$$L_G = 10 * \log \frac{I_G}{I_0} \text{ dB. Kritikus sáv intenzitása tekinthető a teljes, súlyozatlan hangintenzitás}$$

részeként is (ez az intenzitás egy kritikus sáv szélességével egyező szélességű frekvenciaablakba esik). Frekvencia kritikus sáv mértékké alakítása a frekvenciafüggő ablakot 1 Bark szélességűvé teszi, a kritikus sáv mértékétől függetlenül. Következésképpen egy kritikus sávnyi szélességű keskenysávú zaj kritikus sáv intenzitása a sáv mértéktől függ és háromszög alakú, melynek alapja 2 Bark szélességű. Szinuszos hang képe négyszögletes, 1 Bark szélességű.

A középértékek – például az inger vagy az ingerszint – jobban közelítik hallószervünk frekvenciaszelektivitását. Az alsó és felső csúcsokat küszöbként használják az ingerszint kontra kritikus sáv alapú mérték meghatározására. Jelen telítettség a kritikus sáv szint maximumának felel meg. A csúcstelítettség a szubjektíven mért maszkolási küszöbök csúcsának felel meg. Legtöbb esetben $L_E = 10 * \log \frac{E}{E_0} \text{ dB}$ használatos. Jelen telítettség

környezetében meghatározzuk a kritikus sáv mértéket, ezzel egyezik meg a telítettségi szint. Ha az intenzitás sűrűség durván változik a kritikus sáv mértékének függvényében (azaz alul áteresztő szűrőn átengedett zajról vagy szinuszos hangról van szó), akkor a kritikus sáv szint maximumértéke felel meg az ingerszintnek. Ebből a pontból vagy ennek a tartománynak a közepéről indulva adódnak hozzá az ingerszint csúcsoi. Ezek úgy állnak elő, hogy a maszk küszöbszint csúcsát felfelé toljuk mindaddig, amíg az ingerszint megegyezik a pillanatnyi

ingerszinttel, amely már eredetileg is elérhető volt. Vagyis a maszk küszöbértékek felfelé tolódnak a maszkolási index értékével (vagyis a kritikus sáv szint és a maszk küszöbérték közötti különbséggel) a pillanatnyi inger környezetében.

A hallásküszöb is értelmezhető belső zaj által keltett maszk küszöbként. Ez a belső zaj közép- és magas frekvenciákon frekvenciafüggetlen, alacsony frekvenciákon viszont erősen nő, és ott felelős a hallásküszöb emelkedéséért. A szívdobogás és a belső spontán izommozgások tipikus zajforrások, amelyek akusztikailag ingerlik hallórendszerünket, főleg nagyon alacsony frekvenciákon. A zajok hatása fokozható a külső hallójárat elzárásával, ilyenkor próbamikrofonnal nagyobb hangnyomásszint mérhető, mint nyitott hallójáratnál.

Komplex hangok hangmagassága

A komplex hangok tekinthetők több tiszta hang összegeként. Ha a tiszta hangok frekvenciái egész számú többszörösei egy közös alapfrekvenciának, akkor a kapott komplex hangot harmonikus komplex hangnak nevezzük. Annak ellenére, hogy a komplex hangok számos tiszta hangot foglalnak magukba, a mindennapi életben sokkal gyakrabban vannak jelen, mint a tiszta hangok. A magánhangzók vagy számos hangszer hangja harmonikus komplex hang. Komplex hangok magassága becsülhető a tiszta hangok magasságával való egyezésekből. Noha számos tiszta hangot tartalmaznak, a komplex hangok általában nem sokféle hangmagasságúak, inkább egy egyedüli vagy kiemelkedő hangmagasságot mutatnak. Alapjában véve a komplex hangok magassága közel áll egy frekvenciához, amely nincs távol az összetevők közötti frekvenciakülönbségtől, azaz az alapfrekvenciától. Közelebbi vizsgálat kismértékű, ám rendszeres eltérést mutat ettől a szabálytól. Egy 60 dB-es tiszta hang és egy 50 dB-es harmonikus komplex hang alapfrekvenciája közötti eltérés az alapfrekvencia függvényében a következő tulajdonságú: Alapfrekvenciától indulva a komplex hang az összes azonos amplitúdójú felharmonikust tartalmazza 500 Hz-ig, e fölött -3 dB/oktáv spektrális súlyozással. Az 1 kHz alatti alapfrekvencia esetén annak folyamatos csökkentése fokozatosan lassul. Tehát a -3-as súlyozás szerint egy 58,2 Hz-es tiszta hang magassága megegyezik egy 60 Hz alapfrekvenciájú harmonikus komplex hang hangmagasságával. Illetve egy 396 Hz-es tiszta hang és egy 400 Hz alapfrekvenciájú harmonikus komplex hang magassága is azonos. 1 kHz felett pedig semmi eltérés nincs.

Harmonikus komplex hangok hangmagassága függ a szinttől. Egy 200 Hz alapfrekvenciájú komplex hang magasságának eltolása a szint függvényében a következő: A

komplex hang szintjének növekedésével az eltolás negatív irányban növekszik. Ugyanez a helyzet áll fenn alacsony frekvenciákon tiszta hangok esetén. Vagyis egy komplex hang magassága az összetevőinek spektrális hangmagasságán alapszik. Ha egy tiszta és egy komplex hang magassága megegyezik, akkor a tiszta hang frekvenciája alacsonyabb a komplex alapfrekvenciájánál. Egy lehetséges magyarázat szerint a komplex hang alapfrekvenciáját a második felharmonikus tolja az alacsonyabb frekvenciák felé. Ha egy komplex hangból eltávolítjuk az alsó felharmonikusokat, a hangmagasság alig változik. Vagyis a nem teljes, alapfrekvencia nélküli harmonikus hang magassága nagyon közel áll az alapfrekvencia magasságához. Ahol ez a jelenség áll fenn, ott beszélünk virtuális, alacsony, megmaradó hangmagasságról. Nem minden harmonikusoktól megfosztott komplex hangnak van virtuális hangmagassága. A legmélyebb komponens frekvenciája és az alapfrekvencia speciális kombinációjának fennállása esetén jön létre virtuális hangmagasság. Azok a komplex hangok, melyek legalacsonyabb frekvenciakomponense is 5 kHz feletti, nem rendelkeznek virtuális hangmagassággal, bármekkora is legyen az alapfrekvenciájuk. Alacsony alapfrekvenciák esetén a határ még lejjebb tolódik. Egy 50 Hz alapfrekvenciájú komplex hangnak csak akkor van virtuális hangmagassága, ha legalacsonyabb komponensének frekvenciája nem haladja meg az 1 kHz-et. Vagyis alulról felfelé a huszadik felharmonikusig rendelkezhetnek virtuális hangmagassággal, vagy kelthetnek ilyen érzetet. A csonka harmonikus hangok magassága a teljes harmonikus hangokéhoz hasonló. A teljes esethez hasonlóan negatív relatív frekvenciakülönbséget mutatnak, igaz nagyobb, nagyjából kétszereset. A csonka harmonikus komplex hangok magasságának szintfüggősége levezethető a legmélyebb spektrális komponens szintfüggőségéből is. Ennek következtében, ha a legalacsonyabb frekvenciájú komponens 3 kHz körüli, akkor a hangmagasság eltolása növekszik a szinttel.

Egy harmonikus komplex hang bizonyos spektrális összetevőinek feluláteresztő szűréssel való eltávolítása helyett a spektrális komponensek módosíthatóak nem hallható módon is egy meredeken emelkedő elégséges szintű aluláteresztett zajjal. Ez esetben a komplex hang alsó frekvenciakomponensei teljesen maszkoltak, a csonka komplex hang legalsó frekvenciája pedig az aluláteresztő szűrő vágófrekvenciája mellett kezdődik.

Hangosság

Hangosság: az érzetek kategóriájához tartozik. Az inger-érzet kapcsolat nem építhető fel az alig észlelhető intenzitásváltozásokból direkt módon, más típusú mérések eredményeiből kell azt megszerezniünk, például magnitúdóbecslésből. Ráadásul a hangossággal kapcsolatban a hangosság szint is fontos. A maszkolás mindig jelen van, teljes csend nem létezik, illúzió csupán. A nagyon kis zörejeket, például izommozgások zaját elfedi a szervezetünk. Tehát beszélnünk kell részlegesen maszkolt hang hangosságáról, amikor a vizsgált hang mellett jelen van egy maszkoló hang is. A megmaradó hangosság értéke nulla (=maszkolási küszöb) és a részlegesen maszkolt hang magassága között változhat, általában sokkal halkabb, mint egy nem maszkolt hang esetében. A maszkolás bekövetkezhet azonos időben megszólaló vagy időben eltolva, késleltetett maszkoló hanggal. Eszerint a részlegesen maszkolt hangosság effektusai nem csupán spektrálisak, hanem időbeliek is.

Hangosság szint: hangossági összevetések, összehasonlítások sokkal pontosabb eredményekre vezethetnek, mint a magnitúdóbecslés. Emiatt jött létre a hangok hangosságérzetének jellemzésére a hangosság szint. Barkhausen vezette be az 1920-as években, a kutató, akinek a neve rövidüléséből született a kritikus sáv arány mértékegysége: a Bark. Minden hang hangossága mérhető, de a legismertebbek azok a görbék, amelyek tiszta hangok különböző frekvenciáinak hangosság szintjeit mutatják be. Ezek a görbék a hangtér azonos hangosságú pontjait ábrázolják, egyenlő hangossági görbéknek nevezzük őket. A görbék rajzolásához szükséges méréseket számos laborban elvégezték, a hangokat legalább 500 ms-ig játszva. 1 kHz-nél minden görbe áthalad azon a ponton, melynek dB-értéke megegyezik a görbe ugyanazon a ponton jelentkező phon-értékével. Vagyis a 40 phon-hoz tartozó görbének át kell haladnia a 40 dB-es ponton 1 kHz-nél. Az emberi hangosságérzékelés alsó határa, a hallásküszöb is egyenlő hangossági görbe. 1 kHz-nél a hallásküszöb nem 0 hanem 3 dB, ezért a grafikonon is 3 phon-nál kell megjelennie. Az alacsony hangosság szinthez, mondjuk 20 phon-hoz tartozó görbe alakja majdnem párhuzamos a hallásküszöb görbéjével. Ez még inkább érvényesül nagyjából 200 Hz feletti frekvenciák esetén, és abban a tartományban fennáll magasabb hangosság szintekre is. Ennél alacsonyabb frekvenciákra és magasabb hangosság szinteken a görbe ellaposodik. Egy 50 Hz-es 50 dB hangnyomás értékű hang hangosság szintje 20 phon, viszont egy 50 Hz-es 110 dB-es hang hangosság szintje már 100 phon. A phon és dB értékek különbsége nagyjából 30 az alacsony szinteken, a magasabbakon viszont csak 10. A hallás legérzékenyebb területe minden görbén

a 2 és 5 kHz közötti tartományba esik, és egy mélyedésnek felel meg. Nagy szintekre a süllyedés még intenzívebb, vagyis az abba a frekvenciatartományba eső hangok hangosabbak, mint azt párhuzamos eltolásnál várnánk.

Ezek a hangok frontálisan akadálymentesen terjednek ideálisnak feltételezett közegben, a görbéket ilyen mérések alapján rajzolják. A legtöbb esetben azonban a hangtér nem ilyen, inkább hasonlít a diffúz hangtérre, amelyben a hang minden irányból érkezik. Hallórendszerünk nem egyformán érzékeny a különböző irányokból érkező hangokra, ráadásul az irányérzékenység frekvenciafüggő is. Emiatt egy egyetlen irányból érkező hang hangtere és egy diffúz hangtér egyenlő hangosságú görbéi különböznek. A különbség a csillapodással magyarázható, amely azonos hangosságot idéz elő normál és diffúz hangtérben is. Alacsony frekvenciákon ez a csillapodás elhanyagolható, mert hallórendszerünk ilyenkor körkörös gyűjtőként működik. 1kHz-nél -3 dB a csillapodás, tehát diffúz térben az 1 kHz-es hang hangnyomásának 3 dB-lel alacsonyabbnak kell lennie, mint ugyanannak a hangnak a normál térben mért nyomása ahhoz, hogy azonos hangosságúnak érezzük őket. A csillapodásos méréseket nem szinuszos hangokkal, hanem keskenysávú zörejekkel végezték. Nagyobb frekvenciákra a csillapodás is növekszik és 2,5 kHz-nél 2 dB-t ér el. Ennél is magasabb frekvenciákra viszont már csökken. Az egyenlő hangosságú görbéket és a csillapodást felhasználva már megalkotható a diffúz hangtér egyenlő hangosságú görbéinek serege. A görbék a hangosság frekvenciafüggőségét jelzik. Azonban sok más tényezőtől is függ: sáv szélességtől, tartalmazott frekvenciáktól, időtartamtól. Azaz túl egyszerű lenne egyetlen súlyozással közelíteni a hangosság szintet, mint például A-súlyozott hangosság szint. Az ilyen megközelítés frekvenciafüggése csak alacsony frekvenciákon egyezik meg az egyenlő hangosságú görbékkel, azaz csak szinuszos hangok vagy keskenysávú zörejek alacsony szintjén képes helyes meghatározásra. Az így mért zajok vagy a komplex hangok dB(A) értékei, illetve ezek kombinációi csalókák.

A hangosság sáv szélességtől való függésének egyik lehetséges tanulmányozási módja két azonos szintű hang hangosságának mérése Δf frekvenciakülönbségük függvényében. Ilyen mérések eredményét mutatja a következő ábra 1 kHz középfrekvenciával. Minden hang hangnyomásszintje 60 dB volt. Egy 1 kHz-es hang hangosság szintje látható, amelyet az alany azonos hangosságúnak tart egy két hangból álló komplex hanggal. Kis különbségekre, egészen 10 Hz-ig, a fül képes érzékelni a két hang fázisát, ami olyan hangosságérzethez vezet, amely fázison belül elért csúcspontnak felel meg. Ez a csúcs az egykomponensű hang

hangnyomásának kétszerese és 66 phon hangosságshoz vezet. Azaz kis frekvenciaszeperációk esetén a hangosságshintet a perióduson belül a hangnyomás csúcsértéke adja. Egy komplex hang hangosságshintjét a két hang hangnyomás értékének összegzésével kapjuk meg. Nagyobb frekvenciátávolságokra, amikor Δf már 20 kHz és 160 kHz között található, az azonos hangosságúnak ítélt hang nyomása állandó, 63 dB. Ez a hangintenzitások összegződésének felel meg. Másként megfogalmazva: abban a frekvenciartományban, amelyben a fül már nem veszi számításba a hangnyomás burkoló időszakos változásait, ott már az intenzitás a hangosságot leginkább meghatározó tényező. $\Delta f > 160$ Hz esetben az azonos hangosságúnak ítélt 1 kHz-es hang hangosságshintje erős emelkedésbe kezd. Ez a tendencia folyamatos $\Delta f = 2000$ Hz-ig (ez az érték, 1 kHz az f_1 -nek 400 Hz-en és f_2 -nek 2400 Hz-en a mértani közepe), ahol eléri a 10 dB-t az egyetlen hangból álló 1 kHz-es hanghoz viszonyítva. Ez a 10 dB erősödés a hangosság duplázódásának felel meg, feltételezzük tehát, hogy nagy Δf esetén a komplex hang hangossága az öt alkotó hangok hangosságának összegzése.

A frekvenciaszeperáció intenzitásösszegzésből kellően nagy Δf esetén hangosságösszegződésbe való átmenete a kritikus sávokkal hozható összefüggésbe. Más középfrekvenciákra is hasonló összefüggéseket mértek. Vagyis a kritikus sávok fontos szerepet játszanak nemcsak a különböző sávokba eső hangok hangosságában, hanem a komplex hangok hangosságában is, Δf függvényeként.

Elfedés: ez a jelenség a mindennapi életben fontos szerepet játszik. Egy csendes utca járdáján folytatott beszélgetésben alacsony hangerő is elégséges ahhoz, hogy a beszélők értsék egymást. Ha azonban elhalad egy hangos teherautó, a beszélgetést ez erősen zavarja. Konstans hangerő mellett partnerünk már nem érti meg, amit mondani akarunk. Az ilyen elfedés kétféleképpen orvosolható. Megvárhatjuk, amíg a teherautó elhalad, és folytathatjuk a beszélgetést, vagy nagyobb hangerővel beszélünk. Mindkét esetben partnerünk újra hallani fog bennünket. Hasonló hatások érvényesülnek dalokban, dalrészletekben is. Egy hangosabb hangszer elfedhet egy halkabbat. Amint a hangosabb szünetet tart vagy elhallgat, a halkabb ismét hallhatóvá válik. Ezek a szimultán maszkolás tipikus példái. Hogy az elfedést kvantitatívan mérni tudjuk, gyakran meghatározzuk a maszkolási küszöböt. Ez tulajdonképpen egy (általában szinuszos) teszhang hangnyomáshintjének olyan értéke, amely még éppen hallható egy elfedő hang jelenlétében. Néhány nagyon ritka esetet leszámítva ez az érték a hallásküszöb felett van. Azzal csak akkor egyezik meg, ha a két hang

frekvenciája nagyon eltérő. Az elfedő hang erősítésével folyamatos átmenet figyelhető meg: az elfedett nem tűnik el azonnal, csak fokozatosan csökken a hangossága. Azaz kétféle elfedés van: részleges és teljes. Ilyesmi gyakran megfigyelhető párbeszédalkalmával.

Maszkolás akkor is történhet, ha a két hang nem egyidejűleg szólal meg. Ez esetben az elfedettnek rövidnek, kitörő impulzusúnak kell lennie, és az elfedő előtt szólal meg. Ez az elfedés az előelfedés, vagy visszirányú elfedés. Ez a hatás nem túl erőteljes, más a helyzet azonban fordított esetben. Ilyenkor előreirányuló vagy utómaszkolásról beszélünk.

Pszichoakusztikai beállítási görbék: egyes hangok más hangok által történő elfedése többféleképpen is ábrázolható. Négy változó van: a teszthang frekvenciája és szintje, valamint az elfedőé. Alapesetben a teszthang küszöbszintjét szokták ábrázolni adott frekvenciájú és szintű elfedő hang mellett az elfedett frekvenciájának függvényében. A pszichoakusztikai beállítási görbék más mintát követnek: annak a maszkoló hangnak a szintjét ábrázolják a frekvenciája függvényében, amely elégséges egy adott mély frekvenciájú és alacsony szintű teszthang elfedéséhez. Élhetünk azzal a módszerrel is, hogy az alany a teszthangot hallgatva az elfedő szintjét változtatja előbb a teszthang hallhatóvá tételéért, aztán annak elhallgattatása érdekében. Így áll elő a beállítási görbe. Az alacsonyabb frekvenciák felé a görbe laposabb, mint a magasabb frekvenciák felé, a minimumát kicsivel a teszthang frekvenciája feletti maszkoló frekvencián éri el. A paramétert felcserélve az ordinátával megváltoznak a maszkolási görbék jellemzői is. Ezek az egyetlen hangból álló maszkoló által elfedett teszthang még éppen hallható szintjét mutatják a frekvenciájának függvényében, paraméterként az elfedő hang szintjét használjuk. Ezek a görbék függenek a maszkoló frekvenciájától. Ez a függés nem áll fenn, ha az abszcisszán a kritikus sáv arányt ábrázoljuk a logaritmikus frekvenciaskála helyett. A kritikus sáv arány a klasszikus maszkolási minták normalizálásakor hasznos. Az abszcisszán a teszthang frekvenciája helyett ábrázolható a kritikus sáv arány különbség, Δz (Δz a z_m és z_t különbségeként áll elő, z_m a maszkoló, z_t a teszthang frekvenciájának kritikus sáv aránya). Feltételezzük azt is, hogy a teszthang hallásküszöbét 0 dB-re állítottuk be.

Vagyis úgy tűnik, a maszkolás alacsonyabb frekvenciákon szembetűnő, magasabb frekvenciákon kevésbé érezhető a hatása. A klasszikus maszkolási görbék és a pszichoakusztikai beállítási görbék azonos módon tárnak fel jellemzőket, függőséget, könnyen oda-vissza alakíthatóak a paraméter és az ordináta felcserélésével.

Modellek, algoritmusok

Hangmagasság

A hangmagasság megállapítás népszerű téma számos kutatási területen a személyi számítógépek elterjedése óta. Nagyjából 50 évnyi kutatás után sem elég pontosak és hatékonyak azonban a jelenlegi technikák. Mindegyikük boldogul tisztá, egyetlen hangból álló jellel, többszólamú vagy zajos hanggal azonban már gondok adódnak. A hangmagasság meghatározás vagy f_0 becslése általában a legalacsonyabb frekvenciakomponens meghatározását jelenti. Egy periodikus hullámformában a legtöbb komponens harmonikus arányban áll, azaz kis értékű egész számú többszöröse az alapharmonikusnak. A napjainkban használt felismerőmotorok gyakran ki is hagyják a felismerésből a magasabb komponensekre vonatkozó információkat, de azok nem teljesen haszontalanok. Sok szemantikai információt hordoznak fonetikai és lexikális szempontból. Hangsúlyozó nyelvekben a kiejtés közbeni magasságmódosulások lexikálisan többetinformációt közölnek, ilyenkor hasznos hozzávennünk ezeket az információkat is a becsléshez, a kétértelműség elkerülése érdekében.

A valóságban a hangoknak sok felharmonikusa van, és ezek befolyásolják a hangmagasság észlelését. Ezt erősíti meg a gazdagabb spektrum, a hangzást teltebbé, összhangzatosabbá teszi. Egy hullám minél inkább hasonló egy szinusz hullámhoz, annál határozottabban állapítható meg a frekvenciája, érzékelhető a hangmagassága. Ez az érzet a felhangok kapcsolatától is függ: minél szorosabb a harmonikus függésük, annál jobban meghatározható a hangmagasság is. Ez a meghatározás függ még többek között a hullámforma intenzitásától és időtartamától is. Az alapfrekvencia becslése bonyolult téma, számos próbálkozás volt, mindegyik kielégítően működik a saját specifikus környezetében, más területen viszont kevésbé. Ezért léteznek becslő algoritmusok külön a zenéhez és külön a beszédhez. Az eredménykiértékelés történhet sok hangminta és sok alany segítségével. Ezek az emberek, szakértők és laikusok vegyesen, megítélik a kapott eredményeket, ez a módszer csak így lehet hiteles. Más lehetőség: veszünk n darab becslési módszert és egy közös mintát. Ha a magasságot $k \approx n$ számú jól becsli meg, akkor a maradék $n-k$ tévedett.

A hangmagasság egy érzetoldali fogalom, egy periodikus vagy kváziperiodikus hullámforma f_0 értékéhez kötődik, elegendő a rezgés alapperiódusát meghatároznunk, ennek reciproka a frekvencia. Ha azonban a hullámforma egynél több szinusz hullámból áll és elég sok a felharmonikus, vagy csak néhány van, de ezek f_0 -nál nagyobb energiájúak, akkor a

becslés bonyolultabb. Az időtartomány alapú módszerek számításai általában egyszerűbbek, a frekvenciatartomány alapúaké bonyolultabbak, a statisztikai módszerek pedig valószínűségelmélettel igyekeznek segíteni a döntéshozást.

Időtartomány alapú módszerek

Idő-esemény arányérzékelés: az ilyen módszerek a jelben a rövidebb események ismétlődését keresik.

Zérusátmenet arány (ZCR): annak mérőszáma, hogy adott időegység alatt hány zérusátmenet van a jelben. A feltételezés szerint a ZCR a spektrális tartalomról nyújt információt. Ha egy hullámforma tartalmaz nagyfrekvenciájú spektrális komponenseket, akkor egy ciklus alatt akár kétszer is átmehet a nullán, azaz a módszer használata előtt szűrünk kell, ezzel kizárjuk a magas frekvenciákat, de csak óvatosan, nehogy az f_0 frekvenciakomponens is eltűnjön.

Maximumarány: a másodpercenkénti pozitív csúcsok száma. Elméletileg a jelnek minden ciklusban van maximuma és minimuma. Elég a maximumokat számolnunk, ebből a frekvencia meghatározható. Gyakorlatilag vennünk kell egy lokális csúcsok számát becslő algoritmust, és amíg egyetlen lokális maximum van, addig a másodpercenkénti csúcsszámok adják a frekvenciát.

Emelkedés-lejtés arány: ha egy jel periodikus, akkor a hullámforma emelkedései vagy lejtői is azok, tehát a maximumok vagy a nullák a ZCR-hez hasonlóan számolhatóak.

Autokorreláció: két jel közötti korreláció a hasonlóságuk mértéke. A jeleket különböző időszeleteken hasonlítjuk össze, mindegyikben a hasonlóságukat vizsgálva. A kapott eredmény egy hasonlóságsmérték a két jel kezdete eltolódásának függvényében. Az autokorreláció egy jel önmagával történő korrelálása, ami teljes egyezést jelent nulla értékű késleltetésnél, annak növekedésével pedig egyre nagyobb eltérést. A periodikus jelek autokorrelációs függvénye is periodikus. A késleltetés növekedésével és félperiódusnyi értékhez való közeledésével minimalizálódik a korreláció, mivel fáziskioltás következik be. A teljes periódusnak megfelelő értékhez való közeledéskor maximalizálódik az autokorreláció, mert a jelek fázisban vannak. Az autokorreláció első maximumértéke jelzi a periódust. Gond a felharmonikusokban bővelkedő kváziperiodikus jelekkel van.

YIN algoritmus (Alain de Cheveigné és Hideki Kawahana): a YIN-YANG filozofikus egyensúlyelméletéről kapta a nevét, és egyensúlyt keres az autokorrelációk és a

semlegesítődések között. A szorzat maximálása helyett a jel és eltolt másolata közötti különbséget próbálja minimalizálni.

Fázistér: egy jel rövidtávú történetéből próbál következtetni az ismétlődő ciklusokra. A jel értékét és meredekségét azonos időpillanatban ábrázoljuk. Periodikus jel ilyen ábrázolásban is periodikus. A kvázi-fázisteres módszer egyszerűbb, a jelet időeltolt önmagával veszi.

Fázistér és frekvencia: a fázistérben minden periodikus jel zárt ciklusú, egy ciklus útja pedig függ a jel harmonikus összetételétől. Nincs más dolgunk, mint a fázistér új értékeit vizsgálni, figyelni, hogy mikor keresztezik és mikor követik a régi nyomot. Legegyszerűbb megoldás a fázistér pontjainak távolságát figyelni. Ha ez nullává válik vagy egy minimumértéket ér el, akkor a jel periodikus, ismétlődik. Kváziperiodikus jeleknél a fázistér pontjainak útja sosem ismétlődik, bár a vonala hasonló.

Frekvenciatartomány alapú módszerek

Ebben a tartományban sokkal több az információ, amely f_0 -lal kapcsolatban állhat.

Komponensek frekvenciáinak aránya: 1979-ben Martin Piszczalski egy teljesen automatikus zeneátíró rendszeren dolgozott, melynek első lépése a hangmagasság-érzékelés. Az ő eljárása spektrális transzformációval indított és maximumkereséssel azonosította a jel komponenseit. Ezekből párokat képzett és ezek közül azt a legkisebb értékű párt kereste, amelyik a harmonikus sorba illik. Ilyen párokból hipotézisek állnak fel a jel alapfrekvenciájára. Ha mindegyikből képződött már egy hipotézis, akkor ezek közül a legerősebbnek tűnő adja majd az alapfrekvenciát. A párok súlyozottak, bizonyosak jobban beszámítanak, ezek a magasabb amplitúdó-értékűek. A módszer nem igényli az alapfrekvencia jelenlétét és nem harmonikus vagy hiányzó felhangokkal is boldogul.

Szűrőt használó módszerek: több, különböző középfrekvenciájú, keskenysávú szűrőt használunk, és a kimenetüket vizsgáljuk. Amikor egy spektrális maximumérték egy szűrő áteresztő sávjának tartományába kerül, akkor a kimeneten levő érték magasabb, mint amikor nem.

Optimum fészülő módszer: hatékony, de számításigényes. Az összetett szűrőnek számos, egymástól egyenlő távolságra levő áteresztő sávja van, a továbbiak helyét az első sáv helye határozza meg. A bemenő jelet sok különböző frekvencián szűrjük, és ha egész sor

rendezetten előálló harmonikus jelenik meg, illetve ezek egybeesnek a fészű fogaival, akkor a szűrő magasabb kimeneti értéket szolgáltat.

Állítható IIR szűrő: a felhasználó által szabályozható keskenysávú szűrő, amely átfogja a teljes frekvenciaspektrumot. Ha a szűrő és egy jól kivethető frekvenciájú frekvenciakomponens bekerül ebbe a tartományba, akkor a középfrekvenciáról leolvasható az f_0 .

Kepsztrum analízis: módosított spektrum, amint az elnevezése is utal rá. Egyfajta spektrális elemzés, ahol a kimenet a bemenet magnitúdóspektruma logaritmusának Fourier-transzformáltja. Ezt az eljárást a nemlineáris rendszerek lineárisítási tételére fejlesztették ki. Egy frekvenciaspektrumban természetes módon feltűnő komponensek súlyai gyakran eltérnek kissé egymástól, a módszer ezt próbálja orvosolni. Elmélete: a jel egyenletes távolságonként tartalmaz csúcsokat, ezek képezik a harmonikus spektrumot. A csúcsok kevésbé eltérőek, ha a magnitúdó logaritmusát vesszük, az amplitúdó kezelhetőbb skálára kerül. Az eredmény egy periodikus jel a frekvenciatartományban, amelynek periódusa – a csúcsok közti távolság – az eredeti jel alaphfrekvenciájával áll kapcsolatban. E módosított jel maximuma az eredeti jel periódusánál van. Előfeltétele: a jel komponenseinek egyenletes közönként kell előfordulniuk.

Multirezolúciós módszerek: az eredeti módszerhez képest növeljük vagy csökkentjük a felbontást, ezzel együtt az ablak méretét, figyelve a hatékonyságot. Az alaperiódus annak az ablaknak a mérete lesz, amelyikben a módszer a leghatékonyabb.

Statisztikai módszerek

Neurális hálók: a kapcsolatmodellek csoportjába tartozik, ezek önszervező mintaillesztők, amelyek osztályozó kimenettel rendelkeznek a homályos és rendezetlen bemenethez. Logikai szempontból csomópont-együttesek, közöttük súlyozott élek vezetnek. Valamely csomópontba bejövő élek jelei összegződnek, és ha ez az érték elér egy küszöböt, akkor a kimeneti éleken a megfelelő csomópontoknak impulzust küldünk.

Maximum likelihood becslők: különböző értékű hipotéziseket hasonlítanak össze és az előző értékeikből becslik meg annak valószínűségét, hogy helyesek-e.

A bemenet előfeldolgozásával, tartományának szűkítésével, az idő- vagy frekvenciatartomány felbontását módosítva a hatékonyság növelhető. Szintén hasznos lehet az emberi hallórendszer karakterisztikájának algoritmushoz fűzése, például egy konstans-Q spektrális transzformáció használata az alap Fourier-transzformáció helyett. Más hasznos

technika lehet a követés: az előző időkerethez tartozó hasonló hipotéziseket keressük. Megmelíthetők még a Rejtett Markov Modell-t (HMM) használó összehasonlító eljárások.

Vannak használatbavételre kész algoritmusok is: kettő ezek közül a Matlabhoz készített beszédelemző Colea csomag is tartalmaz, de a YIN algoritmus is ilyen.

Minden f_0 becslő 3 előfeltételezésen alapszik:

1. a bemenet egyetlen összetevőből áll, ez keretekre bontott, a vizsgált időtartamban homogén;
2. a bemenet specifikus audio tartományra szűkített, arra, amelyre az algoritmus ki lett fejlesztve;
3. az f_0 becslés a hangmagasság érzékelésével azonos.

Remélhetőleg a jövőben nem kell majd előfeltételeket szabnunk.

Autokorrelációs módszerek a hangmagasság meghatározására

Definíció szerint az akusztikai hangmagasság periódusának legjobb jelöltje a hang autokorrelációs függvénye maximumának helyéből kiindulva található meg. A meghatározás pontatlan a mintavételezés és az ablakolás miatt. Ez pontatlan időtartomány-alapú és kepsztrum-metódusokhoz vezetett a hangmagasság-meghatározásban.

Autokorreláció és periodicitás: egy $x(t)$ állandó (konstans statisztikákkal rendelkező) időjel autokorrelációs függvénye $(r_x(\tau))$ a τ késleltetés függvényében: $r_x(\tau) \equiv \int x(t)x(t+\tau)dt$. A függvény globális maximuma $\tau=0$ -ra adódik. Ha vannak még máshol is globális maximumok, akkor a jelet periodikusnak nevezzük és van egy T_0 késleltetés, más néven periódus. Vagyis minden maximum az nT_0 értékeknél van, ahol n egész szám úgy, hogy $r_x(nT_0) = r_x(0)$. Egy ilyen periodikus jel F_0 alaphérfvenciája: $F_0 = 1/T_0$. Ha 0-tól eltérő helyen nincsenek globális maximumok, lokálisak még lehetnek. Ha ezek közül a legkiemelkedőbb $\tau_{\max}$ periódusnál van és elég nagy $r_x(\tau_{\max})$ értékű, akkor azt mondjuk, hogy a jelnek van periodikus része. Annak R_0 harmonikus intenzitása pedig 0 és 1 közötti szám, amely a normalizált autokorreláció

$(r'_x(\tau_{\max}))$ lokális maximuma: $r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)}$. Ilyen $x(t)$ jel egy $H(t)$ periodikus, T_0 periódusú

jelből állítható elő $N(t)$ zaj hozzáadásával. Az első, integrálos $r_x(\tau)$ -s egyenletből következtethetünk arra, hogy ha ezek a részek korrelálatlanok, akkor a jel autokorrelációja a részek autokorrelációjának összege lesz. Nullperiódusra $r_x(0) = r_H(0) + r_N(0)$. Ha a zajrész fehér zaj, azaz nem korrelál önmagával, akkor $\tau_{\max} = T_0$ -nál $r_x(\tau_{\max}) = r_H(T_0) = r_H(0)$ értékű lokális

maximuma van. Mivel nullperiódusnál az autokorreláció értéke a jel energiájának felel meg, így $\tau_{\max}$ -nál a normalizált autokorreláció a periodikus (másnéven harmonikus) komponens relatív energiájával azonos. A komplementere pedig a zajkomponens relatív energiáját jelenti:

$$r'_x(\tau_{\max}) = \frac{r_H(0)}{r_x(0)}; 1 - r'_x(\tau_{\max}) = \frac{r_N(0)}{r_x(0)}.$$

Nem állandó (azaz dinamikusan változó) jelek esetén a rövidtávú autokorrelációt becsüljük egy t időpontban, a jel egy rövid szegmensét nézzük t középpontú ablakban. Ez $F_0(t)$ becsléseket ad a helyi alaphfrekvenciára és $R_0(t)$ becsléseket a lokális harmonikus intenzitásra. Ha azt szeretnénk, hogy egyáltalán értelme legyen eme becsléseknek, akkor a lehető legközelebb kell lenniük az integrálos egyenletből levezetett mennyiségek értékeihez, ha egy állandó jel esetén rövidtávú elemzést végzünk. Előbb azt mutatjuk be, hogy hogyan birkózzunk meg a fellépő ablakolási és mintavételezési problémákkal. Aztán az algoritmus következik. Végül annak hatékonyságát taglaljuk háromféle állandó jel esetében: perturbációk nélküli periodikus jel, additív zajjal terhelt jel és gyorsan változó jel esetén.

Ablakolás és a késleltetési tartomány: egy folytonos $x(t)$ jel alaphfrekvenciájának jelöltjei t_{mid} időpillanatban a t_{mid} középpontú rövid szegmens autokorrelációs függvénye lokális maximumának segítségével határozhatóak meg. 1. lépés: $x(t)$ -ből egy T időtartamú (T az ablakszélesség), t_{mid} középpontú szegmenst veszünk. Ebből kivonjuk a szegmens μ_x átlagát és az eredményt megszorozzuk az ablakfüggvénnyel, $w(t)$ -vel, hogy megkapjuk az ablakolt jelet:

$$a(t) = (x(t_{\text{mid}} - \frac{1}{2}T + t) - \mu_x)w(t). \text{ A } w(t) \text{ ablakfüggvény } t = \frac{1}{2}T \text{ körül szimmetrikus és } [0, T]$$

intervallumon kívül mindenütt 0. Mi a szinusznégyszeges vagy Hanning-ablakot választjuk:

$$w(t) = \frac{1}{2} - \frac{1}{2} \cos \frac{2\pi t}{T}; \text{ 2. lépés: az ablakolt jel normalizált autokorrelációs függvénye tulajdonképpen szimmetrikus függvénye a } \tau \text{ késleltetésnek (ettől fogva az alaphangokat}$$

$$\text{elnyomjuk): } r_a(\tau) = r_a(-\tau) = \frac{\int_0^{T-\tau} a(t)a(t+\tau)dt}{\int_0^T a^2(t)dt}. \text{ E maximumok legnagyobbika olyan}$$

késleltetésnél található, amely az első formánsnak felel meg, amíg mi azt szeretnénk, ha F_0 -nak megfelelő késleltetésnél lenne. Emiatt Hess egy 1992-es tanulmányában az erős formánsokra elég érzékenynek tartja az autokorrelációs módszert. Ráadásul az autokorrelációs függvény aszimmetrikussága a csúcs késleltetésének becslését túl alacsonyra teszi, így a

hangmagasság értékének becslése túl magas lesz (szinuszhullám 3 periódusának Hanning-ablakban történő vizsgálata esetén 6% a hibahatár). Az első probléma megoldására gyakran alkalmazzák az összes 900 Hz feletti frekvencia kiszűrésének módszerét (Rabiner, 1977), amely a legelső kivételével az összes formánst elnyomja, és a hangmagasságot a második maximumból becsli. Ez nem túl hatékony módszer, mivel gyakran találunk magasabb formánsokat 900 Hz alatt és alapprofrekvenciákat 900 Hz felett. Más formánst elnyomó módszerek középkihágást vagy spektrális simítást és hasonlókat alkalmaznak. Ilyen ad-hoc becslések beszéd- és beszélőfüggővé teszik az eljárást. Ezek mind sürgősszerűek, itt egy ellenszert: 3. lépés: kiszámítjuk a Hanning-ablak normalizált autokorrelációs függvényét:

$r_w(\tau) = (1 - \frac{|\tau|}{T})(\frac{2}{3} + \frac{1}{3}\cos\frac{2\pi\tau}{T}) + \frac{1}{2\pi}\sin\frac{2\pi|\tau|}{T}$. Az eredeti jelszegmens $r_x(\tau)$ autokorrelációjának becsléséhez az ablakolt jel autokorrelációs függvényét, $r_a(\tau)$ -t elosztjuk az ablak autokorrelációs függvényével, $r_w(\tau)$ -val: $r_x(\tau) \approx \frac{r_a(\tau)}{r_w(\tau)}$. Könnyen belátható, hogy konstans

$x(t)=1$ jelre pontos a becslés, periodikus jelekre 1-hez nagyon közeli autokorrelációs csúcsokat ad.

Az ablaktípusok közül a Hanning keskenyebb a többinél, így a gyorsan változó hangok kevésbé zavarják meg. Az ablakolt jel és az ablak autokorrelációs függvénye FFT-vel számolható. Ez azért lehetséges, mert az autokorreláció meghatározásához előbb az ablakolt jel Fourier-transzformáltját határozzuk meg, ez átvezet a frekvenciatartományba: $\tilde{a}(\omega) = \int a(t)e^{-i\omega t} dt$, majd az energiasűrűség $(|a(\omega)|^2)$ inverz Fourier-transzformáltját vesszük, amely a késleltetés tartományába vált: $r_a(\tau) = \int |a(\omega)|^2 e^{i\omega\tau} \frac{d\omega}{2\pi}$.

Mintavételezés és a késleltetési tartomány: vegyünk egy időben folytonos jelet, $x(t)$ -t, amely $f_{\max}$ felett nem tartalmaz frekvenciákat. Ezt a jelet mintavételezhetjük $<1/(2f_{\max})$ rendszeres időközönként, így t_n időpillanatokban x_n értékeket kapunk: $x_n = x(t_n)$; $t_n = t_0 + n\Delta t$. Ezzel a mintavételezéssel nem veszünk adatot, mivel az eredeti jel ebben a formában visszaállítható: $x(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin\pi(t - t_n)/\Delta t}{\pi(t - t_n)/\Delta t}$. A mintavételezett jelből számított autokorreláció szintén mintavételezett függvény: $r_n = r(n\Delta t)$. Az autokorrelációban $(m-1)\Delta t$ és $(m+1)\Delta t$ között lokális maximum van, ha $r_m > r_{m-1}$ és $r_m > r_{m+1}$. A periódus első, durva becslése lehetne $\tau_{\max} \approx m\Delta t$, de ez pontatlan és oktávtevesztéshez is vezet az alapperiódus meghatározásánál.

Valamivel jobb az $m\Delta\tau$ körüli parabolikus interpoláció, de még hatásosabb egy „sin x/x ” interpoláció. Mivel nem tudunk végtelen összeget számolni, jobbra és balra véges számú (N darab) minta felett interpolálunk, majd nullára szűkítjük a széleken az interpolációt egy Hanning-ablak újbóli alkalmazásával.

Algoritmus

1. Előfeldolgozás: a Hanning-ablak Fourier-transzformáltja nem kívánt túlnyúlásainak (Nyquist-frekvencia körüli komponensek) eltávolításához lágy felülmintavételezést végzünk: egy FFT-t hajtunk végre a teljes jelen, szűrünk egy szorzással, amely a Nyquist-frekvencia 95%-ától annak 100%-áig lineárisan 0 felé tart, aztán az előző FFT-nél eggyel magasabb rendű inverz FFT-t hajtunk végre;
2. Kiszámítjuk a jel globális abszolút csúcsát;
3. Módszerünk rövididejű elemzés, ezt aránylag kevés szegmensen hajtjuk végre. Minden keretben keresünk néhány késleltetés-magasság párost, amelyek jó periodicitásjelöltek lehetnek, majd minden keretre végrehajtjuk az alábbiakat (ezek a párosok biztosan tartalmazzák a mindig jelen levő zöngétlen komponenst):
 - 3.1. Vesszük a jel egy szegmensét, ennek hosszát (azaz az ablak szélességét) a MinimumPitch paraméter határozza meg, amely a keresett legalacsonyabb alaphfrekvenciát segít becsülni. Jó becsléshez az ablaknak a jel legalább három periódusát tartalmaznia kell;
 - 3.2. Kivonjuk a lokális átlagot;
 - 3.3. Az első jelölt az állandóan jelen levő zöngétlen, melynek intenzitása két küszöbparaméterrel határolható be: a SilenceThreshold és a VoicingThreshold értékekkel;
 - 3.4. Megszorozzuk az ablakfüggvénnyel;
 - 3.5. Interpolációhoz fél ablaknyi autokorrelációs érték kell, ezért hozzáfűzünk fél ablakszélességnyi nullát;
 - 3.6. További nullákat fűzünk hozzá mindaddig, amíg a minták száma kettő hatványává válik;
 - 3.7. FFT-t hajtunk végre realft (1989) algoritmussal;

- 3.8. Négyzetre emeljük a mintákat a frekvenciatartományban;
- 3.9. FFT-t hajtunk végre, amely $r_a(\tau)$ -nak a mintavételezett verzióját adja;
- 3.10. Osztunk az ablak autokorrelációjával, amelyet egyetlen alkalommal számítottunk ki a 3.5.–3.9. lépésekben. Ez $r_x(\tau)$ mintavételezett verzióját szolgáltatja;
- 3.11. $r_x(\tau)$ folytonos verziójának maximumhelyeit és értékeit határozzuk meg Brent (1989) algoritmussal. Számunkra csak azok megfelelőek, amelyek MinimumPitch és MaximumPitch közé esnek. MaximumPitch értékének MinimumPitch és a Nyquist-frekvencia közé esőnek kell lennie. Csak azokat a zöngétleneket tartjuk meg, amelyek lokális impulzusa: $R \equiv \text{VoicingThreshold} + \max(0, 2 - \frac{\text{lokális abszolút csúcs/globális abszolút csúcs}}{\text{SilenceThreshold}/(1 + \text{VoicingThreshold})})$, és a legnagyobb lokális impulzusértékkel rendelkező zöngéseket ($\text{MaximumNumberOfCandidatesPerFrame} - 1$): $R \equiv r(\tau_{\max}) - \text{OctaveCost} * \log_2(\text{MinimumPitch} * \tau_{\max})$. Az OctaveCost paraméter a magasabb alaphfrekvenciákra hatásosabb. Használatának egyik oka, hogy egy kváziperiodikus jelben minden csúcs egyenlő magasságú, ezek közül a legkisebb késleltetésűt kell választanunk. Másik oka, hogy el akarjuk kerülni az additív zaj okozta lefelé történő oktávcsúszásokat. Harmadrészt az észlelt hangmagasság és az akusztikai alaphfrekvencia eltér egymástól. A 2. pontot minden keretre végrehajtjuk, így (F_{ni}, R_{ni}) frekvencia-impulzus párok alakulnak ki, ahol n 1 és a keretek száma között változik, i pedig 1 és egy kereten belüli jelöltek száma között fut. Lokálisan legjobb jelölt minden keretben a legnagyobb R -rel rendelkező. Ám mivel közel hasonló R értékű jelöltek több keretben is előfordulhatnak, elindíthatjuk a globális útkeresőt, melynek célja, hogy minimalizálja a nagy frekvenciaugrásokat és a zöngés-zöngétlen döntés esetleges tévesztéseit;
4. Minden n keretre p_n egy szám 1 és az adott keret jelöltjeinek száma között. A $\{p_n \mid 1 \leq n \leq \text{keretek száma}\}$. Minden útnak költsége van:

$$\text{cost}(\{p_n\}) = \sum_{n=2}^{\text{NumberOfFrames}} \text{transitionCost}(F_{n-1}, P_{n-1}, F_n, p_n) - \sum_{n=1}^{\text{NumberOfFrames}} R_n p_n, \quad \text{ahol } a$$

transitionCost függvényt a következőképpen definiáljuk ($F=0$ zöngétlent jelent):

VoicedUnvoiced és OctaveJumpCost paramétereket 0,2 értékűre tanácsos megválasztanunk. Globálisan a legjobb út a legkisebb költségű. Ez tartalmazhat olyan jelölteket, amelyek lokálisan csak másodjelöltek. A legolcsóbb utak a dinamikus programozás segítségével találhatók meg, pl. Viterbi-algoritmus Rejtett Markov-modellhez (Van Alper és Van Bergem, 1989).

Hangosság modellek implementálása Matlab-ban

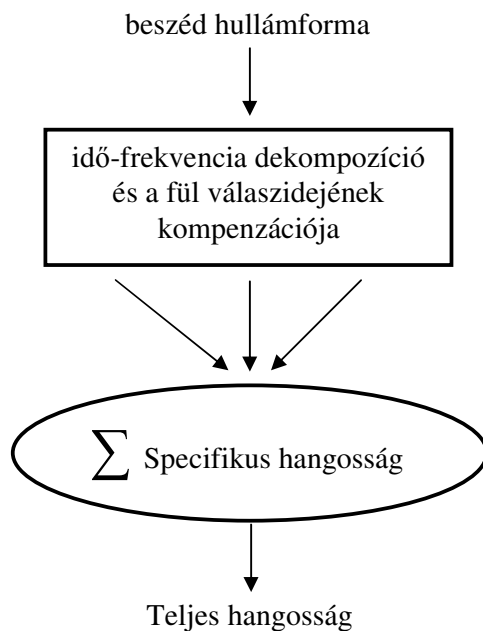
Pszichoakusztikai elemzés során az a cél, hogy olyan transzformációt hozzunk létre, amely egy időbeli hanghullámot képez le egy tartományba, amely a legjobban fejezi ki a hang érzetet. Ilyen transzformációk kulcseleme a dB-ben kifejezett intenzitás és az érzékelt hangosság közötti leképezés. E leképezésre számos modell létezik. Alább Matlab környezetben adunk meg egy szabványként is elfogadott modell implementációját.

A hangtanban az idő-frekvencia elemzés elsődleges eszköze a spektrogram. Népszerű, mivel számítógéppel gyorsan előállítható, kimenete pedig jól értelmezhető. A 90-es évektől kezdődően sok munkát fektettek jobb hangelemző eszközök kifejlesztésébe, amelyek nagyobb figyelmet fordítanak a pszichoakusztikai tulajdonságokra. Rendelkezésre állt a technológia a korábbi évtizedek pszichoakusztikai kutatási eredményeinek teljes implementálására. Napjainkban tehát számos beszéd- és hangfeldolgozó algoritmus hivatkozik pszichoakusztikai transzformációkra. E vonatkozásban egy fontos megkötést teszünk a spektrogramra nézve: a jelintenzitás ábrázolása általában dB-ben kifejezett hangnyomáásszinttel (SPL) történik. Ez lehetővé teszi az intenzitás objektív mérését, de nem fedti teljesen a hang által a hallgatóban keltett hangosság szubjektív érzetét. Ennek eléréséhez figyelembe kell vennünk a fül érzékenységet a hang frekvenciakomponenseinek különböző hangosságszintjeire. A szükséges információkat az egyenlő hangossági görbék tartalmazzák. Ezekről leolvasható, hogy a fül az alacsonyabb frekvenciájú hangokra kevésbé érzékeny és 3-4 kHz között a legérzékenyebb. E görbék alapján a dB értékek átválthatók phon értékekre, ahol azonos phon-értékű komponensek azonos hangosságúak lesznek, de különböző SPL értékűek. A phon-skála egyik hátránya, hogy az érzékelt hangossággal nem egyenesen arányos, például a phon-érték duplázása nem jelenti a hangosság megkétszerezését. A son-skála bevezetése oldja meg a problémát, ez egy lineáris skála. A son-phon skálák közötti összefüggés a következő módon írható le: $L(i) = \left(\frac{D(i)}{40}\right)^{2,642}$ a $D(i) < 40$ esetben, illetve $L(i) = 2^{0,1(D(i)-40)}$, ha $D(i) \geq 40$, ahol $L(i)$ az

érezelt hangosság son-ban, $D(i)$ pedig phon-ban kifejezve. Az alábbi ábra azt szemlélteti, hogy hogyan határozható meg a hangosság son-ban:

Az ábrán feltüntetett fokozatok implementálására különböző megközelítések léteznek. Az alapeljárás szerint első lépésként a jelet idő-frekvencia tartományba kell képeznünk. Az előírt frekvenciaelemzési pontok kapcsolatban állnak majd a fül kritikus sávokra bontó képességének érzékenységeivel. Szintén figyelniük kell az időbeni és a frekvencia szerinti elfedésre, illetve kompenzálnunk kell a hallásküszöb alatti összetevőket. Ezt az állapotot követi az idő-frekvencia szeletek intenzitásszintjeinek átváltása a frekvenciasávok specifikus hangosságába. A hangosságértékeket aztán összegezzük ahhoz, hogy megkapjuk minden egyes időszelet teljes hangosságát. Jelen esetben háromféle implementációs stratégiát elemzünk:

- egy direkt implementációt, amely egy idő-frekvencia dekompozíció alapszik, egy dB SPL→phon konverzió, amelyet az $L(i)$, $D(i)$ képlet direkt implementálása követ;
- a Zwicker-modell háromféle implementációját;
- a Moore és Glasberg hangossági modellt.



A tárgyalni kívánt implementációk némelyikének forrásai beszédminőség-mérési stratégiák. Az idő-frekvencia dekompozíciókat és a hangossági konverziókat is átvehetjük az EMBSD, a PSQM, illetve a PEAQ mérésekből.

Modell implementációk

Kalibrálás: minden implementáció első és talán legfontosabb, döntő fázisa, eleme a bemenő jel kalibrálása. Matlab-ban a wavból beolvasott hangok normalizáláson esnek át, hogy amplitúdójuk -1 és 1 közé kerüljön. Ez természetesen nem lesz befolyással a felvétel, illetve lejátszás közbeni amplitúdóra. A hang beskálázható annak érdekében, hogy elérje a kívánt dB SPL értéket. Amikor dB SPL értékeket használunk, választanunk kell egy referencia értéket. Ezt a szintet levegőben, szabad térben 20 μ Pa-ra szokták megválasztani. Ha a felvételekor alkalmazott dB SPL értéket nem ismerjük, akkor beszéd esetén 65-70 dB SPL közötti értéket érdemes feltételeznünk. Ennélfogva egy y jelvektor 70 dB SPL-re skálázása a következő:

```
SPLMeas = 70;
Pref = 20e-6;
y_refscaled = (y./Pref);
RMS = sqrt(mean(y_refscaled.^2));
SPLmat = 20*log10(RMS); % dB SPL Matlab-ban
c = 10^((SPLMeas-SPLmat)/20);
yca1 = c*(y_refscaled);
```

Ha a HUTear eszköztár telepítve van, akkor erre a célra használható a következő függvény is: `ret = pascalize(y,70);`

A hangosságrepresentáció direkt implementációja

Eme implementációhoz szükséges algoritmust az EMBSD hangminőség-mérőből vettük. A jelet keretekre bontjuk, mindegyiket ablakoljuk a Hanning-függvénnyel, és megkapjuk a spektrális teljesítménysűrűséget. Minden teljesítményspektrumot 1 Bark széles kritikus sávokra bontunk 3,4 kHz felső frekvenciakorláttal. Az EMBSD-ben Schroeder szórási modelljét alkalmazzuk a frekvenciamaszkolás kritikus sávok körüli hatásainak implementációba foglalásához. Minden kritikus sáv hangosság szintjét a szakirodalomból vett egyenlő hangossági görbék segítségével határozzuk meg. A görbék közé eső dB intenzitásértékek pedig interpolálásra kerülnek a helyes hangosság szint megállapítása végett. Majd mértékegységet váltunk, áttérünk son-ra az $L(i)$, $D(i)$ képlet segítségével.

```
NFFT = 1024; NOVERLAP = 0;
Bf = 1:18;
[Yxx,f] = psd(yca1, NFFT, fs, NFFT, 0);
Yxx_scale = (2.*Yxx)./NFFT;
[B_XX, bark] = bk_freq02(Bf, f, Yxx_scale);
C_XX = spread_new(Bf, B_XX);
P_XX = dbtophon(C_XX);
S_XX = phtosn(P_XX);
```

```
N_mbsd(1) = sum(S_XX);
```

A Matlab-kódok ugyanazok, mint Wonho EMBSD-disszertációjában. Mégis helyesbíteniünk kell, módosítunk a `dbtophon.m`-en. A kódban egy `equal.mat` állomány is meghívásra kerül, mivel ez tartalmazza az átírt egyenlő hangossági görbéket.

A Zwicker-modellen alapuló implementációk

Talán a legismertebb és a legnépszerűbb hangosságmodell a Zwicker által javasolt. Része lett az 1984-es nemzetközi harmadoktávós szabványnak is és számos ITU szabvány használta beszéd- és hangminőség-vizsgálatában. Némi eltérés azért tapasztalható ezek között.

A DIN45631 / ISO532B hangosságmodell implementációja

Ez a Matlab program a <http://widget.ecn.purdue.edu/~hastinga/Research.htm> oldalon található kód átültetése volt. Ez az implementáció egy harmadoktávós szűrőkből álló szűrőcsaládot használ a jel spektrális dekompozíciójához. Egyik hátránya, hogy csak durva megközelítést adja a szűrők formájának, illetve középfrekvenciáik helyének. H. Quast adta az N' son/Bark-os dB SPL-ben megadott L_G hangosság szint egyetlen harmadoktávós sávban mért specifikus hangosságának egyenletét: $N' = 0,0064 \cdot 10^{0,025 \cdot L_{ETQ}} \cdot [(11 + \frac{1}{4} 10^{(L_G - a_0 - L_{ETQ})}) - 1]$.

a_0 -val a csillapodást írjuk le, azaz a szabad térben terjedő hangnak a hallórendszerünkig való eljutása közben bekövetkezett gyengülést. Ezt a fejünk és a külső fül okozza. Az L_{ETQ} a hallásküszöbi gerjesztés. E két paraméter értékeit E. Zwicker, H. Fastl és C. Dallmayr adja 1984-ben (Acoustica #55, PP. 63-67, „BASIC program for calculating the loudness of sounds from their 1/3 oct band spectra according to ISO532B”). Az implementáció lefuttatásához Matlabban a következő parancsokat kell kiadnunk:

```
[Yxx, f] = PowSpec(Pref.*yca1, fs, df);
[YdB, err] = Convert2dB(Yxx,1);
```

Eme implementációban az Y_{xx} jel energiájának spektrális sűrűsége a $Pref$ alkotóelem figyelembevétele nélkül kerül meghatározásra. Továbbá csak decibelbe alakítás után használjuk, vagyis a `Convert2dB.m`-be be kell szúrunk a következő sort:

```
YdB = 10*log10((ca1^2)*Yxx/(Pref^2));
```

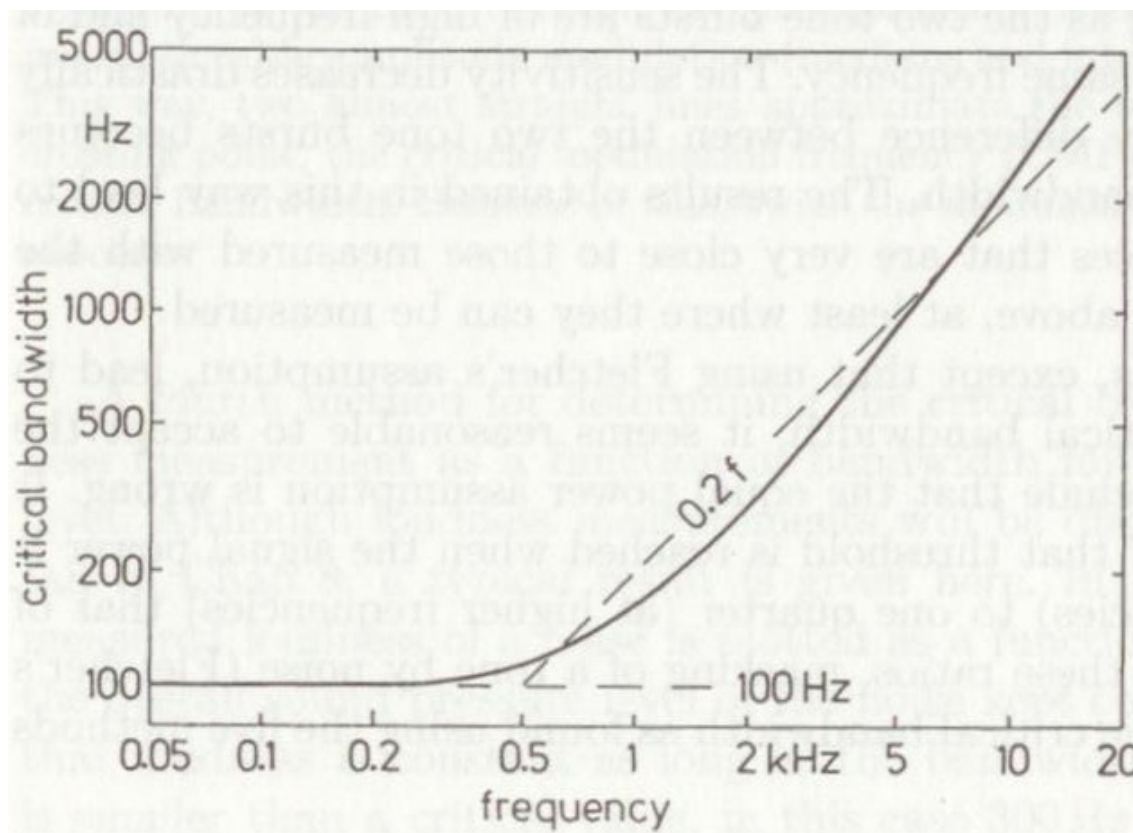
A `PowSpec` két másik bemeneti változója az fs mintavételi frekvencia és a df frekvenciarezolúció. Az alábbi kódok generálják a harmadoktávós szűrőket, a szűrő kivitelezése pedig egy MATHWORKS-ból vett programból származik., a neve `Oct2dsngn.m`. `PowSpec` kimenetét f adja. Kivitelezzük a szűrő válaszait is annak

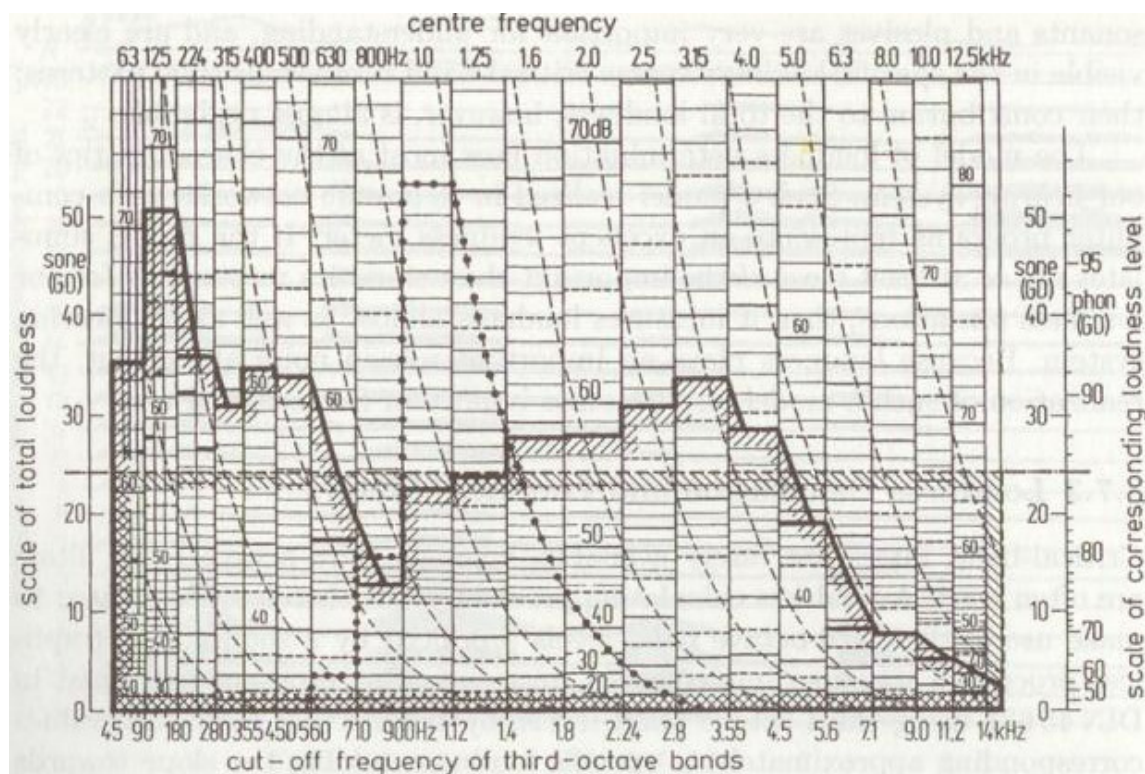
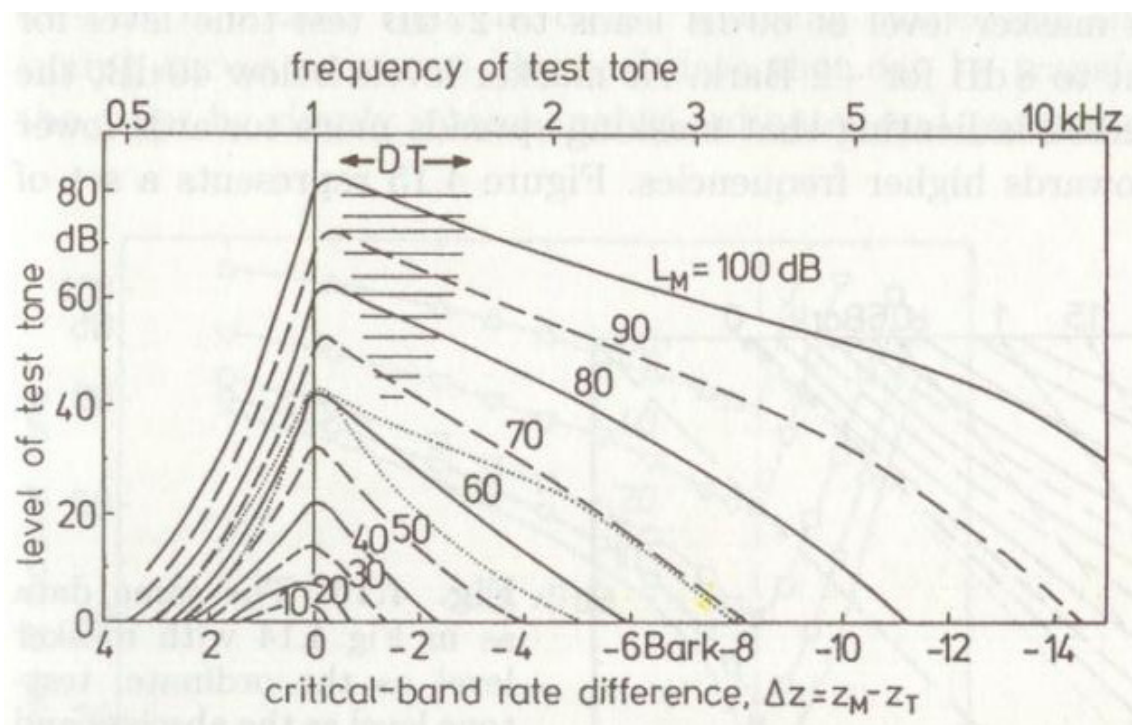
érdekében, hogy meggyőződhessünk róla: a jelbe nem került energiagyarapodás külső elemekből, forrásból. L_t adja a jel energiaspektrumának szűrt sávjait:

```
[H, err] = GenerateFilters_16000(f);
H = 0.94833723551160*H;
for ink = 1:24
    Lt(ink) = 10*log10(sum((10.^(YdB/10)).*(abs(H(ink, :)).^2))));
end.
```

Az utolsó függvényhívás adja vissza a teljes hangosságot, azaz N -t és a specifikus hangosságok vektorát, N_s -t. A bemeneti MS a hangteret adja meg és alapértelmezetten szabad térre állítják, azaz $MS = 'f'$.

```
[N, Ns, err] = DIN45631_16000(Lt, MS);
```





Zwicker hangosság modellje, ahogyan a PSQM-ben használják

A PSQM algoritmust az ITU vette át a beszédminőség megbecsléséhez, manapság váltotta fel a PESQ (ITU P.862 ajánlás: "Perceptual Evaluation of Speech Quality (PESQ), the New ITU Standard for End-to-end Speech Quality Assessment"). Az implementációban használt egyenlet eltér az előzőekben tárgyalt N' -től. Az $LX(f)_n$ specifikus hangosság

egyenlete: $LX(f)_n = S_1 * \frac{P_0(f)^\gamma}{0,5} * [(0,5 + 0,5 * \frac{PPX(f)_n}{P_0(f)})^\gamma - 1]$, ahol S_1 a skálázási faktor, $P_0(f)$

az f frekvenciára vonatkozó abszolút hallásküszöb, $PPX(f)_n$ az n . keret f frekvenciájára vonatkozó hangenergia sűrűsége, γ pedig egy konstans. Ezek a sűrűségek tulajdonképpen egyetlen keret energiaspektrumát jelentik, amelyet Bark-skálára fordítunk 0,321 Bark felbontással, a Hertzben mért sávszélességhez közeli skálázással. A PSQM dokumentum alapján $S_1 = 1/240,5$, $\gamma = 0,001$:

```
[Yxx, f] = psd(ycal, NFFT, fs, NFFT, NOVERLAP);
deltaz = 0,312;
for i = 2:57
    deltaf = Ffreqs(i)-Ffreqs(i-1);
    scal = (deltaf./deltaz);
    indice = (Ffreqs(i-1) <= f & f < Ffreqs(i));
    index = find(indice > 0);
    index_first = index(1);
    index_last = index(end);
    PPX(i-1) = (scal./(index_last-
    index_first+1)).*sum(Yxx(Ffreqs(i-1)
    <= f & f < Ffreqs(i)));
end
Lx = S1.*(P0).^gamma.*(((1-0,5+0,5.*PPX./P0).^gamma)-1);
Lx(find(Lx < 0)) = 0;
N_pseq = sum(Lx); % teljes hangosság, összhangosság
```

Zwicker hangosság modellje, ahogyan a PEAQ-ban használják

Ez a legkifinomultabb a pszichoakusztikai dekompozíciók közül (T. Thiede, „Perceptual Audio Quality Assessment using a Non-Linear Filter Bank”, Ph. D. thesis, TU Berlin, 1999). Minden keret energiaspektrumát súlyozzuk a külső- és középfül frekvenciaválaszával, amelyet egy modellből származtatunk. A spektrális energiateljesítményeket aztán egymástól 0,25 Bark-nyi távolságra levő kritikus sávokba csoportosítjuk. A sávok energiáihoz egy eltolásértéket adunk a fülben keletkező zajok ellensúlyozása érdekében. A frekvenciatartománybeli terjedés implementálására egy dB-ben mért Δ terjedési függvényt használ. A terjedési gerjesztési mintát az $\tilde{E}_{SR}$ adja. A PSQM algoritmussal ellentétben az

ingerküszöb értékeit PEAQ-ban a zárójelben fentebb említett modellből számították, ahol a c konstans, értéke: $LX(f)_n = c * (\frac{E_i(f)}{s(f)E_0})^{0,23} * [(1 - s(f) + \frac{s(f)E_{SR}(f)_n}{E_i(f)})^{0,23} - 1]$. A dB-ben mért

küszöbindex: $s_{dB}(f) = -2 - 2,05 * \tan^{-1}(\frac{f}{4000}) - 0,75 * \tan^{-1}((\frac{f}{1600})^2)$, az ingerküszöb:

$E_{idB}(f) = 3,64 * (\frac{f}{1000})^{-0,8}$. Eme függvény teljes implementációját P. Kabal dokumentuma

tartalmazza: „An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality”, TSP Lab Technical Report, Dept. ECE, McGill Univ., May 2002. X_2 a keret energiaspektruma, E_b a Bark-ba alakított spektrum, LX a specifikus, N_{tot} az összhangosság. Az alábbi kód függvénynevei megegyeznek a dokumentumban szereplőkkel, további bemeneti változók pedig: a jel hossza, len és a mintavételi frekvencia, FS :

```
X2 = PQDFTFrame(yca1, len);
Eb = PQgroupCB(X2, 'Basic', len, Fs);
E = PQspreadCB(Eb, 'Basic', Fs);
[Ntot] = PQLoud(E, 'Basic', 'FFT', Fs);
N_tot(1) = Ntot;
```

Moore és Glasberg hangossági modellje

(B. Moore, B. Glasberg, T. Baer, „A model for the prediction of thresholds, loudness and partial loudness”, J. Audio Eng. Soc., vol. 45, pp. 224-240, 1997)

Ez a modell is eltér a Zwicker-féleltől: a használt hallási frekvenciaskála az egyenértékű négyszögletes sáv szélesség (Equivalent Rectangular Bandwidth, ERB), az egy szűrési sávra eső specifikus hangosság egyenlete pedig: $N' = C * (E_{sig}^\alpha - E_{ThQ}^\alpha)$. E az egyes frekvenciasávokon belüli ingerminta, E_{ThQ} a hallásküszöbön vonatkozó ingerminta, C és α pedig két állandó. Javasolható egy Matlab-beli implementáció a <http://www.auditory.org/postings/2002/565.html> oldalról, amely a **HUTear** eszköztár segítségével számítja ki az ingermintákat. Egy 128 tagú gammaton szűrőcsoporthoz feltételezve az ajánlott modell bemeneti változói a következők:

```
model.fs = Fs;
[f, b, CentFreq] = Make_cgtbank(128, Fs, 200, 6);
save gt128_test f b CentFreq;
model.cochlea.fb.file = 'gt128_test.mat';
model.cochlea.asymmcomp = 1;
model.haircell.r = 'half';
model.haircell.c = 0,7;
model.haircell.f = '1kHz';
model.neural.function = 'mean';
```

Minden keret esetén a jel ingermintáját (excitation pattern) a HUTear eszköztár AudMod függvényének segítségével generáltuk: `Esig = AudMod (ycal, model);` Hasonlóképpen jártunk el a hallásküszöbnél generált ingermintánál is:

```
[CrctLinPwr, frq Npts, CrctdB] = OutMidCrct2('MAF', 128, Fs);
MAF = interp1(frqNpts, CrctdB, CentFreq, 'linear', 'extrap');
for i = 1:length(CentFreq)
    tones(i,:) = pascalize(sin(2.*pi.*(0:frame_len-1).*CentFreqs(i)./Fs, MAF(i)));
end;
Ethq = AudMod(sum(tones), model);
```

Ha az ingerminták ismertek, akkor már kiszámítható a specifikus és az összhangosság is. Ahhoz, hogy a specifikusból az összhangosság megkapható legyen, összegzés előtt skáláznunk kell. Ez a folyamat az ERB szűrők sávszélességén alapszik:

```
N = C.*(Esig.^alpha-Ethq.^alpha);
N(find(N<0)) = 0;
EarQ = 1/0,107939;
minBW = 24,7;
order = 1;
b = 1.019;
ERBwidth = ((CentFreq/EarQ).^order+minBW^order.^(1/order));
totalLoudness = sum((N.*ERBwidth)');
```

A kimenet kalibrálása és tesztelés

Az MBSD, a PQSD és a Moore-Glasberg hangossági modellek esetében szükségessé vált a kalibrálás. Az optimalizációs eszköztár `lsqcurve-fit.m` függvényét használtuk erre a célra. Az MBSD modellt használva példaként, a kódot a következő formában írhatjuk:

```
function[l] = mbsd_cal(coef, S_XX)
l = diag(sqrt(coef.*S_XX)*sqrt(coef.*S_XX)');
```

, ahol `coef` a kalibrációs paraméter, `S_XX` a specifikus hangosság kiszámításához használt inger, `l` pedig az összhangosság.

1000 Hz-es, 512 pont hosszúságú, 16 kHz-en mintavételezett és {40, 50, 60, 70, 80} dB SPL-re kalibrált szinuszos hangokat használtunk. Ezek összhangossága pedig {1, 2, 4, 8, 16} volt. MBSD modell esetén `S_XX`-et skáláznunk kell 0,2567 együtthatóval. PSQM modellben $S_1 = 7,63 \cdot 10^{-4}$ és $\gamma = 0,2941$, Moore és Glasberg modelljében pedig $C = 0,0002$ és $\alpha = 0,8885$. Ezen szinuszgörbék különböző modellekben előállt összhangosságai (son-ban) az alábbi táblázatban szerepelnek. Jól látszik belőle, hogy egyetlen mérés sem találja el az összhangosságot pontosan, de mindegyik nagyon közel jár az elvárt értékhez.

Modell	40 dB	50 dB	60 dB	70 dB	80 dB
MBSD	0,701	1,709	3,835	7,97	16,102
DIN45631	0,8030	1,974	4,513	9,43	18,76
PSQM	0,975	1,985	4,001	8,007	15,98
PEAQ	1,312	2,551	4,701	8,363	14,486
MOOREGLAS	0,9364	2,06	3,978	7,23	13,1

Bemeneti szinuszoid SPL értékek és a modellek kimenetei son-ban.

Következtetések: jelen elemzésben néhány hangosságmodell Matlab implementációját mutattuk be. Ahol szükség volt rá, ott a modell kalibrálása is megtörtént. Végül az eredmények is közlésre kerültek a modellek kimenetének értelmezése céljából különböző dB SPL hangosságú szinuszhullámokra.

Nyelvészeti fogalmak

Intonáció

A hangsúlyon belül fontos különbséget tenni a szóhangsúly és a mondathangsúly között: a szóhangsúly a szavaknak az a szótagja, amelyet a legnagyobb erővel (és rendszerint hangosabban is) ejtünk – ezt főhangsúlynak is nevezzük. A hangsúlyos szótagok a magyarban nem választhatóak meg szabadon: egy adott szónak egy adott szótagján van a főhangsúly, és adott szótagokon vannak a mellékhangsúlyok. Egy mondat minden szavának megvan tehát a hangsúlyos szótagja, melyet az egyes szavakon belül a legerősebben ejtünk, de ez az erősség nem azonos minden szóban: a mondat a közlés szempontjából legfontosabb szavát (vagy csak annak hangsúlyos szótagját) még erősebben ejtjük, mint a többi szót. Ez a mondathangsúly. A hangsúly nem összetévesztendő a hanglejtéssel: míg az előbbi egy-egy szótag kiejtésének *erősségére* vonatkozik, az utóbbi teljes szavak vagy mondatok szótagjainak egymáshoz viszonyított *hangmagasság-változásaira*. Tehát amit a mondat végén le szoktunk vinni, az szigorúan véve nem a hangsúly, még ha általában azt is mondjuk, hiszen nem a hangerőt csökkentjük, hanem a hangmagasságot. A mondat hanglejtése (intonációja), ha nem is változtatja meg, de mindenképpen árnyalja a mondat jelentését, viszont ez nyelvenként eltérő lehet. A magyar nyelvben ugyanaz a mondat hordozhat több különböző jelentést, ha eltérő az intonáció.

Nyelvünkben a szóhangsúly mindig a szavak első szótagjára esik, hacsak nem valamelyik másik szótag (morféma, azaz önálló jelentéssel bíró rész) jelentőségét szeretnénk kiemelni. A szótagok hangsúlyozottsága hozzátartozik a szóhoz: ha egy másik szótagra tesszük a hangsúlyt, megváltozik a szó hangzása. A pusztán "esztétikai" hibán túl egy rossz helyre tett hangsúly megnehezítheti a szó felismerését is. Ha az alapszóhoz képzőt illesztünk, annak kiejtése alapján véve nem változik meg (a *kéz-kezes* típusú rövidüléstől eltekintve).

Az intonáció tulajdonképpen a mondat dallamát jelenti, akár le is lehetne kottázni. A kijelentő mondatban és a kiegészítendő kérdésben a fontos szótól ereszkedővé válik a hanglejtés. Az eldöntendő kérdésben az utolsó előtti szótagnál felvisszük a hangot, az utolsón pedig leejtjük. Visszakérdezéskor a mondat szerkezete rendszerint kijelentő, de emelkedő dallammal, hiszen ilyenkor csak az intonáció különbözteti meg a tényleges kijelentéstől. A felszólító mondat is ereszkedő hanglejtésű, legfeljebb kicsit erőteljesebben, nagyobb hangerővel mondjuk, kiáltjuk.

Sok nyelvben a hanglejtésnek szemantikai funkciója van: ezzel lehet például állító mondatot kérdővé tenni, csodálkozást, gúnyt kifejezni. Ilyen nyelv a magyar is. Más nyelvekben a hanglejtés a szó szintjén általános jelentésmegkülönböztetésre képes – a hanglejtésnek ezt a típusát nevezik tónusnak. A tónus (szóhanglejtés, zenei hangsúly) a hangmagasság nyelvi alkalmazása szavak megkülönböztetésére. Az intonáció számos nyelvben kifejezhet hangsúlyt, nyomatékot, szembeállítást, kérdést, felkiáltást, érzelmi viszonyulást. A tonális nyelvek szótári szavakat különböztetnek meg velük: eltérő lesz egy szó jelentése, ha közepes, mély, magas, ereszkedő vagy emelkedő hangfekvésben ejtik ki. Az ilyen esetekben a tónust, mivel jelentést különböztet meg (akárcsak a fonéma), tonémának is nevezik.

A hanglejtésükben eltérő szavakat a tonális nyelvek beszélőinek agya különböző szavakként azonosítja, és e nyelvek beszélői között elterjedt az abszolút hallás. A világ nyelveinek jelentős része tonális, például a kínai, a vietnami, a thai, valamint számos más ázsiai és afrikai nyelv. A magyar és az indoeurópai nyelvek többsége azonban nem ilyen (a pandzsábitól eltekintve); némelyekben pedig csak korlátozottan érvényesülnek ilyen vonások (pl. litván, szlovén, szerbhorvát, limburgi, svéd, norvég nyelv). A tónus lehet állandó (magas, közepes vagy mély) vagy változó (emelkedő, ereszkedő, emelkedő-ereszkedő, ereszkedő-

emelkedő stb.). A lehetséges tónusok (tonémák) száma az egyes nyelvekben kettő és hat között változik. A tónusok, bár nyelvi szerepük idegen a magyar anyanyelvűek számára, könnyen képezhetőek, ha olyan mondatokat ejtünk különböző hanglejtéssel, amelyek csak egyetlen szóból állnak, illetve ha a beszéd folyamatból kiragadunk egyes szavakat, amelyek a mondat hanglejtését hordozzák.

A szavak megfelelő hangsúlyozására és a mondatok hanglejtésére tehát nem árt odafigyelnünk, ha tisztán, érthetően és egyértelműen szeretnénk beszélni.

Intonáció PC-vel

A személyi számítógépek intonációtanulásban való használatának és hatékonyságának kutatása az 1960-as évek környékén kezdődött, azonban az ezekkel történő jelelemzés és hangmagasság meghatározás technológiája csak az 1990-es években vált tömegek számára elérhetővé. Igaz konkrét implementáció kevés volt akkoriban, azonban a szükséges programok és hardverelemek rendelkezésre álltak a fejlesztés buzdítására. A hangfelismerés, azaz az emberi beszéd felismerése és „megértése” gyorsan fejlődő területe a számítástechnikának. A kutatások kereskedelmi felfedezésekhez és termeléshez vezettek, amelynek a kiejtés tanítására is hatása volt. A hagyományos kiejtésjavító módszerek elég szubjektív értékeléstől függenek, amikor a tanár javítja ki a külföldi diákot, vagy saját magának kell hibáját felfedeznie a helyes kiejtést szalagról hallgatva. Így sok hiba feltáratlan marad, szemben egy számítógép monitorával, ahol objektíven látható a hiba helye, típusa, jelentősége és a kijavításában mutatott fejlődés. Egy 1979-es kutatás eredményéből kiderült, hogy ha a diákok utánozzák a hallott mintát, vagy ha grafikus ábrázolásban látják is a helyes kiejtési módot, az eredményeik akkor sem javulnak számottevően. Ha azonban a sajátjukat egy grafikonon látják a követendővel, akkor a másik két csoporthoz viszonyítva sokkal jobb kiejtésre tesznek szert. Egy, a 80-as évek elején folytatott kísérletből derült ki, hogy azok, akiknél volt audiovizuális visszacsatolás, sokkal hamarabb érték el egy mondat pontos és helyes kiejtését, mint akiknél csak hallási visszacsatolás volt. Közben sokkal motiváltabbak voltak, többet gyakoroltak, ismételtek, több hibát javítottak ki.

Kezdetben a tanításhoz szükséges audiovizuális anyag elkészítése is szinte megfizethetetlen összegekbe került. A berendezések jórészt költséges oszcillografikusak és spektrografikusak voltak, a 90-es években viszont a hardver, főleg a beszéd digitalizálók és a hangmagasság meghatározó algoritmusok elérhetőekké váltak. Akkor érett meg az idő az

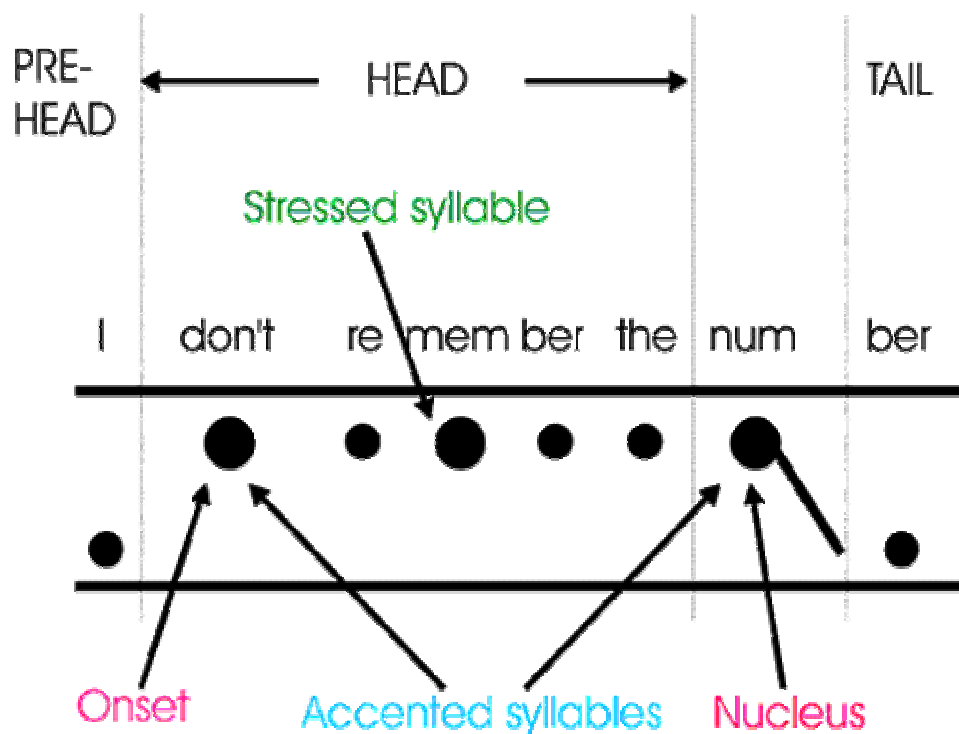
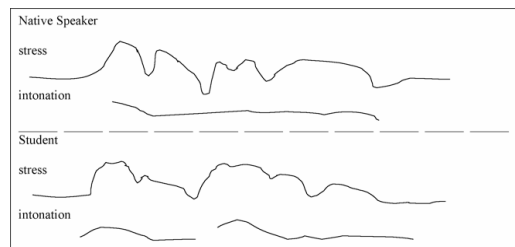
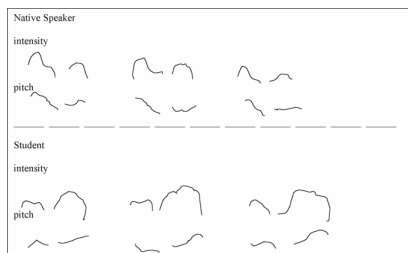
ilyen segédanyagok elkészítésére. A Kay Elemetrics Corporation New Jersey-ben egyedi, hordozható, önálló eszközt fejlesztett ki a helyes intonáció oktatásához, Visi-Pitch model 6087PC néven, beépített beszédelemzővel. Számos területen használják (logopédia, idegen nyelvek oktatása, beszéd kutatás). A Visi-Pitch-et inkább pedagógiai célra fejlesztették, így nem volt elég rugalmas, nem fedte a lingvisztikai kutatások széles spektrumát. IBM-re számos más alkalmazást fejlesztettek: Micro Speech Lab, Interactive Laboratory System, Canadian Speech Research Environment, CSpeech.

Ha az egyén meg is tanul hangokat, szavakat helyesen kiejteni az adott nyelven, akkor is idegen akcentussal beszél majd, ha natív intonációs mintáit viszi át a tanulandó nyelvbe. A nyelvtanításban fontos szerepet játszik a kiejtés. Az ezt vizsgáló kutatások bebizonyították, hogy fel kell váltanunk a kiejtés hagyományos, fonéma-alapú kezelését egy tágabb, beszéd-alapú szemlélettel. Ez átteszi a hangsúlyt a pedagógia eme részében a nyelv alaki és strukturális feléről annak kommunikatív területére. Ez a bővített megközelítés nem csupán a nyelv hangjait veszi figyelembe, hanem a prozódia elemeit (intonáció, hangsúly, ritmus) és a hangminőséget is (suttogás, magas hangú beszéd, fejhangon történő beszéd).

Abberton és Fourcin (1975) adta meg azokat a követelményeket, amelyeknek egy vizuális visszacsatolós rendszernek meg kell felelnie: 1) a kép, ábra legyen tiszta és értelmezhető; 2) a visszacsatolás történjen valós időben (minimális késleltetés még elfogadható a beszéd rögzítése és az ábrázolás között, < 1 sec); 3) a minták méretének egyeznie kell a különböző hangfekvésű alanyok esetében; 4) a közölt információ ne legyen túl részletező; 5) a rendszer legyen képes két görbe ábrázolására azonos időben ugyanazon a kijelzőn: egy helyes F_0 kontúr rajzolására és az alany próbálkozásának megjelenítésére; 6) a berendezés lehetőleg olcsó, megbízható és könnyen kezelhető legyen.

A korábbi tanulmányok is elég hatékonynak mutatták a hangmagasság vizualizáló berendezéseket, azonban ezek a 90-es években kezdtek igazán széles körben elérhetővé válni és terjedni, a technológia fejlődésével. A PC-k nagy előnye az, hogy azonnali és végtelen lejátszási kapacitással rendelkeznek (a kazettákhoz viszonyítva, amelyeket minden lejátszás után vissza kell tekernünk, illetve a sorozatos használat romlik a minőségük).

Néhány kép az irodalomból:



Alkalmazásfejlesztés

A fejlesztői környezet: Matlab 6.5 Release 13

A Matlab nagy hatékonyságú, interaktív, tudományos és műszaki számításokra és azok grafikus ábrázolására kifejezetten alkalmas rendszer. Eszközei alkalmasak a numerikus analízis, a mátrix algebra, a jelfeldolgozás és a grafikus megjelenítés feladatainak megoldására. Alapvető adattípusa a mátrix, nem kell deklarálni. Összetett numerikus feladatok oldhatóak meg a beépített vektor- és mátrixműveletekkel, nem kell foglalkoznunk a

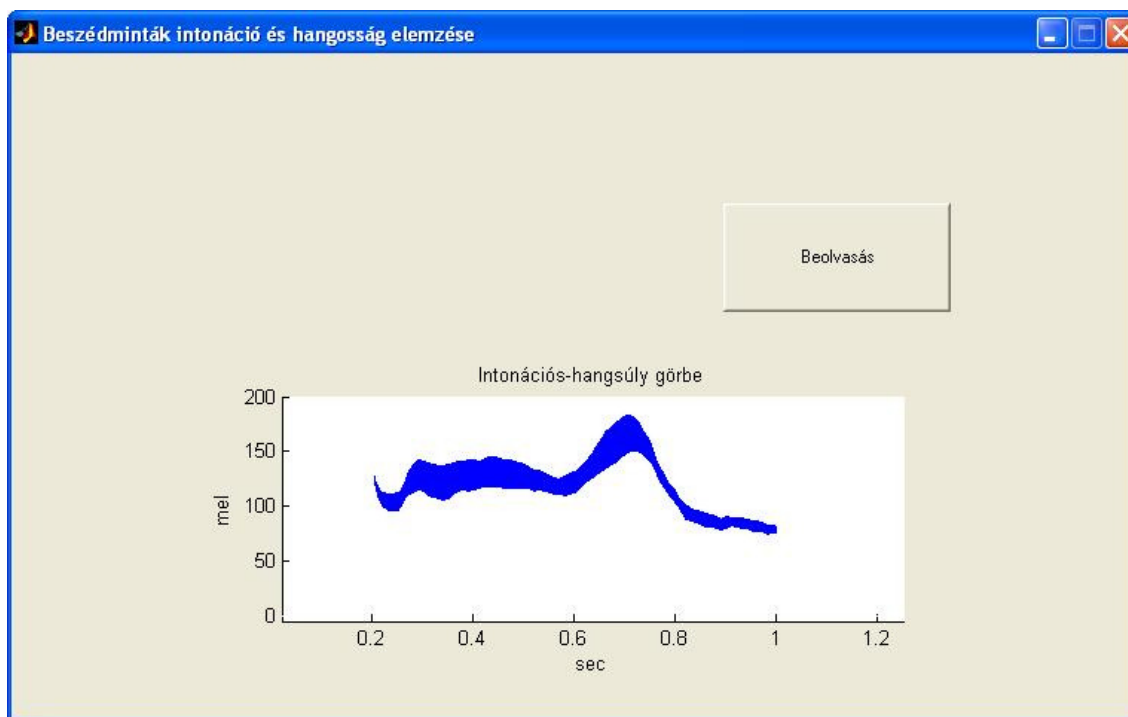
hagyományos értelemben vett programozással, nem kell időt töltenünk ilyen eljárások megírásával. A program neve a MATrix LABoratory név rövidüléséből keletkezett, a rendszer létrejöttében számos kutató több évig segédkezett. Mára világszerte elterjedt, egyetemeken is oktatják segédeszközként bizonyos tantárgyakhoz, az iparban pedig kutatásra, gyakorlati mérnöki és matematikai feladatok megoldására használják. A Matlab rendszer eszköztárakkal bővített, amelyek speciális feladatosztályok kezelésére megírt Matlab-függvények gyűjteményei, tufományterületeket igyekeznek lefedni. Ugyanakkor bárki saját eszköztárát is fejleszthet, a rendszer tehát rugalmas, könnyen kezelhető. Futtatható parancssorból is a megfelelő futtatható állományt indítva, de helyes telepítés után a Windows Start Menüjéből is elérhető. Alap adattípusa a mátrix, az index kezdőértéke 1. Vannak struktúrák, cella tömbök és objektumok is.

Nagy erőssége a hatékony és rugalmas grafikai rendszere. Létrehozhatunk két- és háromdimenziós grafikákat, forgathatjuk őket, tengelyeket címkézhethetünk, szintvonalakat rajzoltathatunk be. Ha megadták az összetartozó koordinátákat, akkor is könnyen kirajzolhatjuk a grafikont. Több grafikon is megjeleníthető ugyanazon az ábrán, ekkor a rendszer különböző színnel rajzolja őket, ha felül nem bíráljuk. Az ábrák valamely részét vagy egészét nagyíthatjuk-kicsinyíthetjük is. A grafikus rész objektum orientált, hierarchikus, gyökere a *root* (azaz a képernyő), szülője minden grafikus objektumnak, neki viszont nincs szülője. Egyetlen ilyen lehet, nem kell definiálnunk, magától létrejön indításkor. Minden grafika grafikus objektumokból áll, vannak interfészek is. A gyermekek öröklik a szülő beállított tulajdonságait és a saját beállításaiikkal együtt továbbadják saját gyerekeiknek, hacsak másképpen nem rendelkezünk, expliciten nem adunk meg mást. A *get* és *set* utasításokkal kérdezhetőek le, illetve módosíthatóak különböző tulajdonságok, megjelenítési módok. A *figure* egy grafikus ablak a képernyőn. A *patch* objektum egy kiszínezett poligon által definiált terület. A grafikus ablak rajzolási tartományát az *axes* definiálja, szülője a grafikus ablak, a *figure*, gyerekei lehetnek pl. a *line*, *image*, *patch*, *surface*, *text*.

Intonációs és hangossági görbék megjelenítése

Mint ahogy az előző fejezetekben már említettük a hangmagassági és hangossági görbék megjelenítése nagyban segíti az idegen nyelv tanulását, azon belül is a helyes intonáció és hangsúlyozás elsajátítását. Erre készült egy alkalmazás Matlab 6.5-ös környezetben. A hangosság kiszámításához szükséges az 1 kHz-es hang hallásküszöbének a meghatározása. Ezt – fizikai mérőberendezések nélkül – csak közelítőleg tudjuk meghatározni, mégpedig a felhasználó közreműködésével, annak ún. másodfajú ítéletei alapján: A felhasználó véletlenszerűen váltakozva erősödő és gyengülő hangot hall. Jelezni kell, hogy mikor tűnik el, ill. jelenik meg, válik hallhatóvá a hang. Ezt követően az alkalmazás meghatározza a hangmagassági és hangossági függvényeket, majd ábrázolja azt. Kerültük azt az ábrázolási módot, amely megosztja, vagy valamiféle konverzióra kényszeríti a figyelmet. Mivel a hangosság egyfajta mennyiségi érzet, a hangmagasság ábrázolása már a zenében is függvényábrázolás-szerű, ezért a hangmagasságot ábrázoljuk az egyenletes mel-skálában, de úgy, hogy a függvénygörbe vastagsága nem egyenletes, az egyenesen arányos a hangossággal, annak is az egyenletes son-skála szerinti nagyságával.

Az alkalmazás a GUI eszköztárának felhasználásával készült. A nyitó panelről a kalibrációs



ablakba jutunk ahol a fenti kalibrációs eljárás zajlik. Sikeres kalibráció után visszajutunk a fő ablakba, ahol beolvasás és megjelenítés váltja egymást.

Az alkalmazásnak számtalan továbbfejlesztési lehetősége van. Egy ilyen: natív és tanulói kiejtések összehasonlítása: alkalmas metrika definiálása, idővetemítés, hangfekvések egymáshoz illesztése stb.

Felhasznált irodalom

Budó Ágoston: *Kísérleti fizika I (mechanika, hangtan, hőtan)*, 1989.

E. Zwicker, H. Fastl: *Psychoacoustics, facts and models*, 1999.

Tarnóczy Tamás: *Zenei akusztika*, 1982.

Internetes források

David Gerhard: *Pitch Extraction and Fundamental Frequency: History and Current Techniques*, 2003

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.834&rep=rep1&type=pdf>

http://idegennyelvor.blog.hu/2009/09/13/hangsulyozas_es_hanglejtes_az_angolban

Joseph Timoney, Thomas Lysaght, Marc Schoenwiesner, Lorcán Mac Manus: *Implementing loudness models in Matlab*, 2004.

http://dafx04.na.infn.it/WebProc/Proc/P_177.pdf

Paul Boersma: *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*, 1993.

http://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf

Ezúton szeretnék köszönetet mondani Dr. Papp Zoltán adjunktusnak, témavezetőmnek, aki az írás során nagy segítségemre volt, felmerülő kérdéseimre választ adott, áldozatos és türelmes munkával, észrevételeivel és tanácsaival segítette a dolgozat létrejöttét.