

Beszámoló 1. ea - 2009.09.07

Bevétel

Olaszy Gábor (\*) olaszy@tmit.bme.hu  
Némethi Géza nemethi-11-

B226 05 noba

Zle: nov. 25 17-19<sup>00</sup> 90 perc Dec 15  
Pzh: dec. 9 12-14<sup>00</sup> PPZH + U13FA!

Gyakorlat: hétfő 14<sup>00</sup> - 16<sup>00</sup> kithelenséint  
1. gyab. sept. 28.

1. beszédminőség, 2. adaló - 8
3. HMM-es beszédplism. - 8 4. a sz. nívók
5. beszédletés 6. b. inf. szed. ló.

Kötelező előadási anyag:

- 1) magyar nyelvű beszédtechnológiai alapismeretek (~600o. mm.)  
↑ letp://speechlab.tmit.bme.hu | alpha.tmit.bme.hu/beszéd
- 2) magyar hangképzésről nevezeti bemutatása  
↑ letp://fonetika.nyttud.hu/cvvc

C = consonant (msh)

V = vowel (msh)

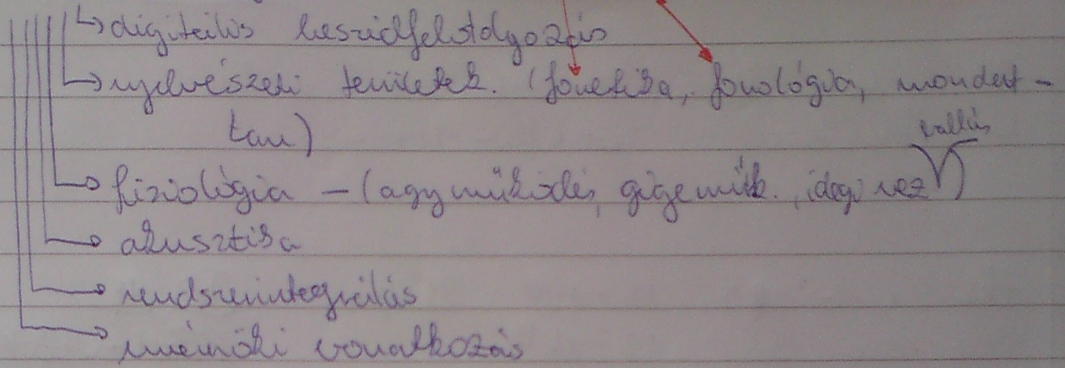
megérkező, intenzitás (t) fo.  
spektrum  
időreus.

Ajánlott irodalom: Földes Géza 1983 Digit beszédfeld.  
... Olasz: letp://beszedelall.

Multibox - növegyelohasó du

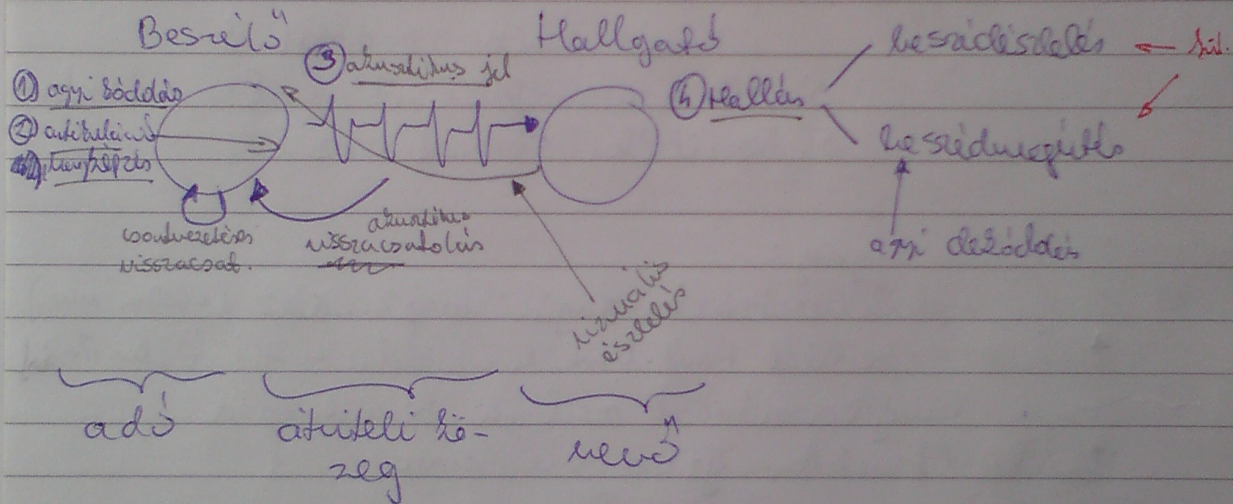
Kéllítés: neplehp. / kempelen (plunkeris. letp://ig)

Témakör: beszédtechnológia



beszédhangok, megértés, artikulációk  
nyelvi működés

Beszédlánc (1. rész)

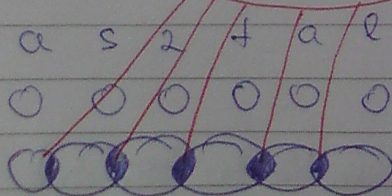


Hétföldkörüvel.

a.) Kempelen Farkas - 1791: mechanikus beszédgép (artikulációs szintézis)

egy 11 részes kóza léte a hangokat, külön szó

↳ hangátmenet felfedezése



- klyf.
- b) Alexander Graham Bell : 1876 - szabadalom AT&T
- c) Puskeis Tivadar : 1893 - ~~ka~~ kelfonokzp.
- d) Helmholtz : 1903 - rezonátor
- e) Bárány Miklós : 1916 - (későbbes nagy felde. als. és. gép)  
beszédszintézis szabadalom
- f.) PCM - 1938 Reeves / 1939: 1. elektronk. beszéd-sint. VODER (USA)
- g.) Tamóczy Tamás fizikus kísér. a V-2 rezonanciáit  
(Foucault) 1941
- h.) Békassy György 1947 Nobel-díj (?)
- i.) Gábor Péter 1950
- j.) Davis Klatt (MIT) 1980 fizikus, vezető  
DEC talb. mintakészítő - 1. kész. technik
- k.) Hungarovox 1982 - nagy. szűrés. ok.
- l.) Eurospeech világkongr.: 1999
- m.) Eurotel konferencia: Westel 1999

886-8490 : gyűjtemény

2. év. 2009. 09. 10.

(U. : nov 25)

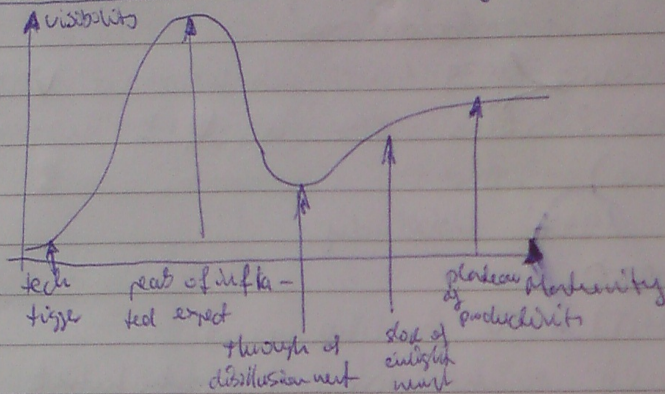
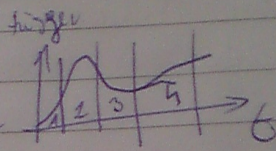
## Beszédtechnológiai alapterületek

- 1) Beszédtechnológia: km. és. lán. bármely elemi és gépi  
gép. sziválas, megvalósítása
- 2) Felismerés (hangátvitel): 'digital speech processing'
- 3) Gépi beszéd-előállítás: 'speech synthesis'  
'speech generation'
- 4) gépi beszéd-modosítás: 'speech modification' (o → ? word/dat)  
(usual → szöveg, szöveg)
- 5) gépi beszéd-felismerés: 'automatic speech recognition' ASR  
pl.: nyelv- és hang- (lang. ac.) • beszéd- és felismer.

• azonosítás (szach. szint. & id. utif.)

• beszed megertes

Gyujtelési kerdés: Gartner: huppe cycle



Veltes: pontok a nyelvtudologia: slope: beszédfejles. cell centrum

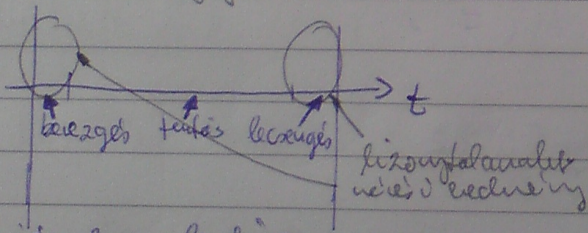
(2) Per.

A hang fizikai leirasa:

(Tamóczy Tamás) hang def.: az az a fiz. folyamat, amelynek során a levegő részecskéi a hangforrás körül az áll. ott élnek el, v. ugranak az áll. on haladnak át. Ez az osc. és lehet per. és v. rendszeren, néllenszerű.

hang. rezg. mozg. : tisztá hang

CD: 3.2-es fejezet



néllenszerű: hang leirása

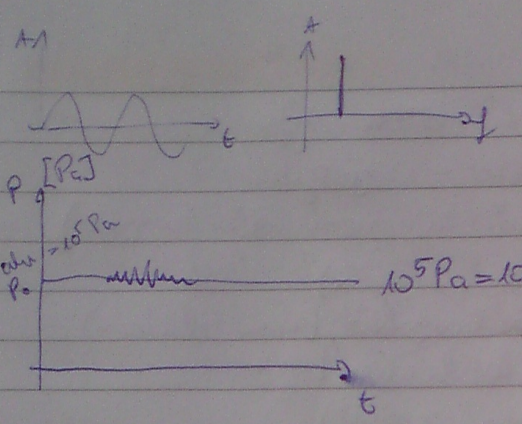
hang fontos adatai: • periódusidő =  $T$       $\frac{1}{T} = f_0$  ( $F_0$ )

alpfrekvencia

•  $\lambda$  v.  $\nu$ -összetett /  $\nu$  számszoros-e /

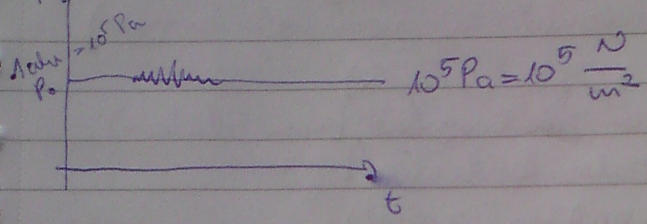
• periódikus vagy aperiodikus (széjszerű)

A1



sinuszhang mechanikus leje

hangnyomás



$$P_{\text{összes}} = p_0 + p(t)$$

$p(t) \sim 10^{-3} \text{ Pa}$  nagyságrend  
 $2 \cdot 10^{-5} \text{ Pa}$  a hallásküszöb

hangintenzitás:  $\frac{W}{m^2}$  (átvitelt energia mennyiség)  
(I)

$$2 \cdot 10^{-5} \text{ Pa} = 10^{-12} \frac{W}{m^2}$$

nagy átlagolás

repülőgéplajtómű zaj (fejdelőműködés) =  $1 \frac{W}{m^2}$

$$dBel = dB = 10 \log \frac{I}{I_0} = 10 \log \frac{I_2}{I_1} = I (dB)$$

hallásküszöb

$$P (dB) = 20 \log \frac{P}{P_0} \leftarrow \text{AKUSZTIKAI dBel}$$

60 dB + 60 dB = 63 dB

$$I + I (dB) = I + 3 dB$$

$$P + P (dB) = P + 6 dB$$

gyorsulás sebesség:

$$c = \frac{\lambda}{T} \quad \lambda \cdot h. \text{ hossz}$$

$$c_{\text{szövet}} = 340 \frac{m}{s}$$

A hallási rendszer (3) Bew

CD 3.2

Def.: a hallósz. v. hallási sz. komplex akusztikai, veles, lidván-i, elektv. jelátalakító, deprezetési és agyi szerkezet.

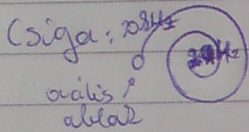
Mindezek pontosan nem ismét.

(külső fül → dobhártya)

- Fül részei:
- külső fül (fülkagyló - dobh. - ig)
  - középfül (dobhártyától - csigáig)  
(hallószalék, kalapács - üllő - lemez)
  - belső fül (csukló labirintus)  
(csigák - aqulis ablak)

mechanikus erősítő (1,3x)  
 megfelelő erős hangoktól (innen elhúzza)

Dobhártya:  $10^{-8}$  mm kinyúlás már érethet  
 $0,6 \text{ cm}^2$  felület  
 hang → mechan. energi



a hang elmozdítja a csigában lévő folyadékot  
 oszlopot

Alaphártya hossza: 35 mm

csigában érethető szél → idegsejt → inger

Frekvenciaérzékelés: 20 Hz - 20 kHz

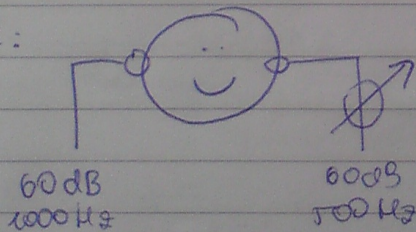
beszéd: 80 Hz - 10 kHz

hogyan érethető az egyes frekvenciákat?

$$I_1 \text{ dB} + I_2 \text{ dB} = ? \quad - \quad I + 3 \text{ dB} ?$$

F Hz ?

Példék:



hangerősség-érték = [phon]

(abba...)

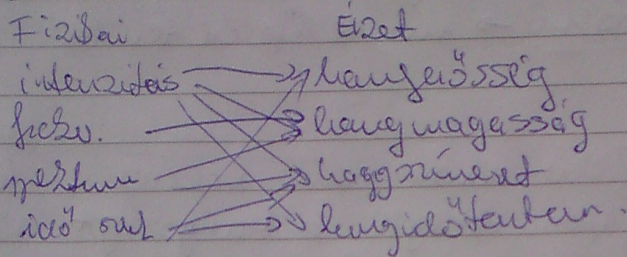
Fletcher-Munson görbe



phon def.: 1 hang erőssége annyi phon, ahány dB a vele azonos a vele azonos hangosság értéket kellet 1 kHz → szám hang hangnyomás szintje.

3. ea. 2009.09.14

Psichodikusok:



phon:  $L_N$

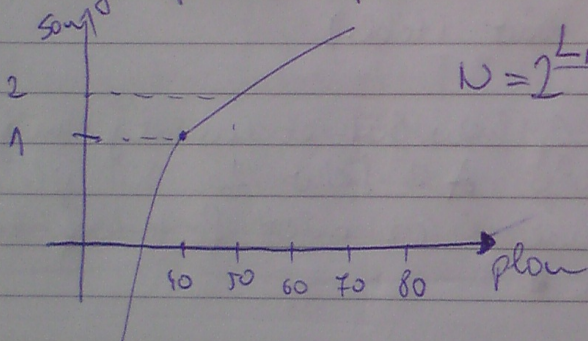
Vissanyitott hangosság

son: Barkhausen (1927)

$N$  (phon)

? : phon(son) összefüggés.

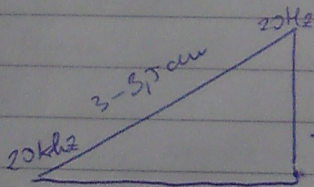
- megállapítás: 40 phon = 4 son



$$N = 2 \frac{L_N - 40(\text{phon})}{10} \quad (\text{son})$$

son skála: valóságos érzeti vélelés skálája.

def.  $T$ : a hangot által kiváltott érzet összehasonlításánál



hangelfedés a  $f$ -tartományban:

fogad mosást, és vki beszél, ha hall  
"has. frekv. hangot elnyomja a nagyobb arh. zaj."

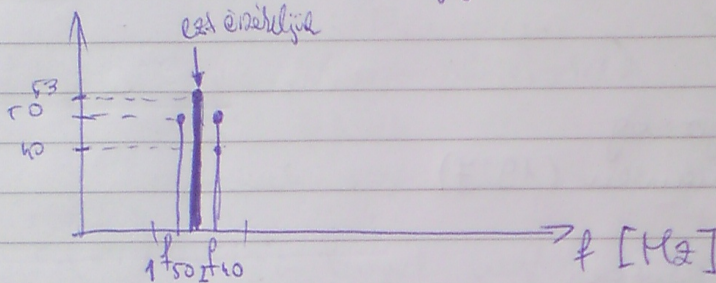
elfedési szint IdB<sub>T</sub>-ában, ha a zaj hangján nyomja el  
a hangot. Tul.: nem szimmetrikus:  $f \uparrow$  jobban  
elnyomja a hangot

elfedési görbe tulajdonságai:

- $d$  mérték:  $f \rightarrow$  jobban elnyom // felt: nagy  $f$  és  $r$  + távolság //
- ha az intenzitás nő  $\rightarrow$  nő az elfedési tényező

Kritikus sávok:

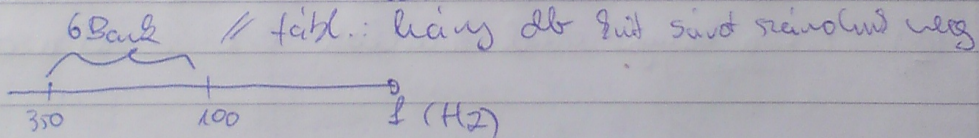
Def.: ha 1 adott frekv. sávon belül hallásunk legkevesebb gyenge intenz. alapon összegező a hangosságát, akkor a kritikus sávon belül vagyunk.



Az emberi fül 24 db kritikus sávot ~~számlál~~ 6-t.  
1 mm  $\sim$  1 db kritikus sáv (tabl.)

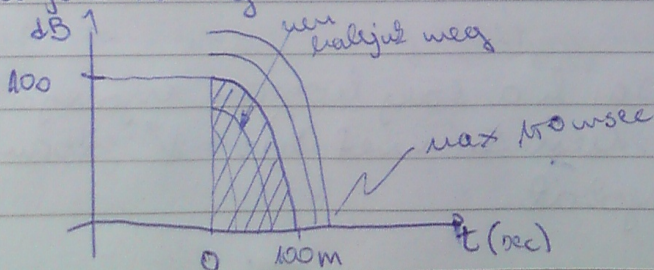
Bark/Def.: (Bark-skála) 2 frekv. sáv-ot mutatja meg pszichoakb. szempontból. A 2 sáv 2-er sáv jel-je, hogy hányx lehet felneves az adott  $f$ -tartományban a kritikus sávoké-ét. ■

(350 Hz  $\div$  1000 Hz : 6 Bark)



hangelfedés az időtartományban:

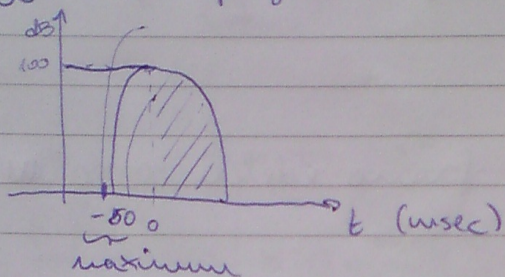
Utáelfedés: elős hangingerés után a fhangébbel nem halljuk meg.



figyelőgözei 150. kft - e.



előfeldetés: eős hanginger dőtti hangot nem halljuk meg.



• nagy sűrűs. jel gyorsabban halad, elyongja a kisebbet.

Összefoglalás:

- hangnyomás:  $p_a$   $p_0 = 10^{-5} Pa$
- hallásk.:  $2 \cdot 10^{-7} Pa$
- hangnyomásszint:  $L [dB] = 20 \lg \frac{p [Pa]}{p_0 [Pa]}$  (alk. db)  $\parallel 2 \times 20 = +60 dB //$
- hangintenzitás:  $I [\frac{W}{m^2}]$

halláskiszoró:  $10^{-12} \frac{W}{m^2}$

fejdelonul.:  $1 \frac{W}{m^2}$

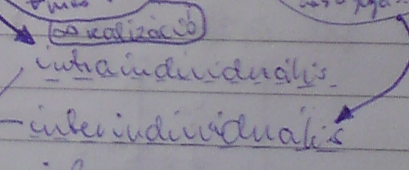
- hangintenzitátszint:  $L [dB] = 10 \lg \frac{I [\frac{W}{m^2}]}{I_0 [\frac{W}{m^2}]}$   $\parallel 2 \times 10 = +20 dB //$
- hangfrekvencia szék:  $p_{low}, 1000 Hz = dB = p_{high}$   $\parallel 1000:$
- na is f-on: a görbéről a másik le

- hangosság (normalizált):  $son$  összefügg.

(4) A beszédnyelvi szerkezetek:

beszéd, 200.06.17  
00:01 - 00:04  
Procty, 2000.06.17

szó szinten  
még a hangja  
00:01 - 00:04



Akustikai mint: folyamatos időfo. — interindividuális  
Agyi szint: mappá // diszkrét elemű jelstruktúra

Nyelv — beszédhang — szótag — szó — gondolati 1seg —  
prosódiai 1seg

(5) Hangképzés (CD 2.3), (21.0.)

ahát fuchi sell

Hangtípusok: • gongos (hátriperiodikus), V + löngis - C

(tubulens áramlás, (vibráló) - vészhangok +  
+ rezonancia.

- közelemléim - 1xi folyamat - zárhangoz.

## (6) Hangszintanalízis:

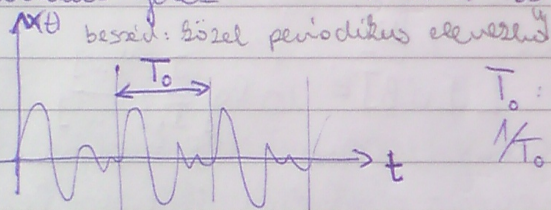
Gaelos kövök : 64 fajúna eitékú beszédhang

rejt  
/T<sub>0</sub>

4.00 - 2009.09.18

## (f) Spektális vizsgálata

a.) periodikus jel:  $\leftrightarrow$  Fourier sor



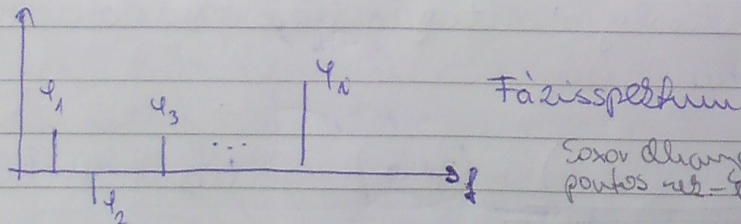
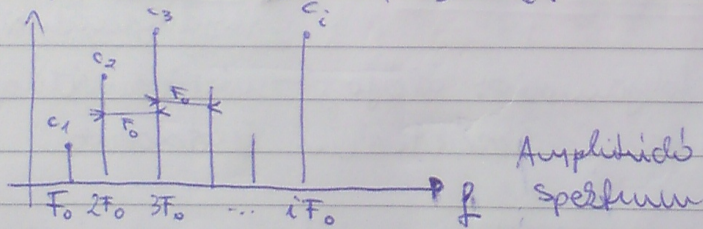
$T_0$  : periódusidő

$f_0 = \frac{1}{T_0}$  : alaphérv. **PITCH**

$$x(t) = x(t + k \cdot T_0) \quad k = \pm \infty$$

$$x(t) = c_0 + \sum_{k=1}^{\infty} c_k \cos(k\omega_0 + \varphi_k) \quad \omega_0 = \frac{2\pi}{T_0}$$

A jel spektruma:  $\{ \omega_0, c_k, \varphi_k \}$



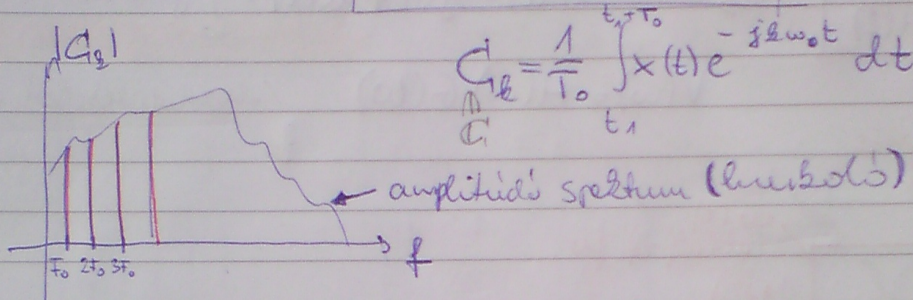
Periodikus jel spektruma: vonalas  
vonalak távolsága:  $F_0$

Telefon sáv: 300 Hz  $\pm$  3400 Hz  
 $\uparrow$   
 3,12 Hz

$\sim f_0$  fekvény: 100 Hz

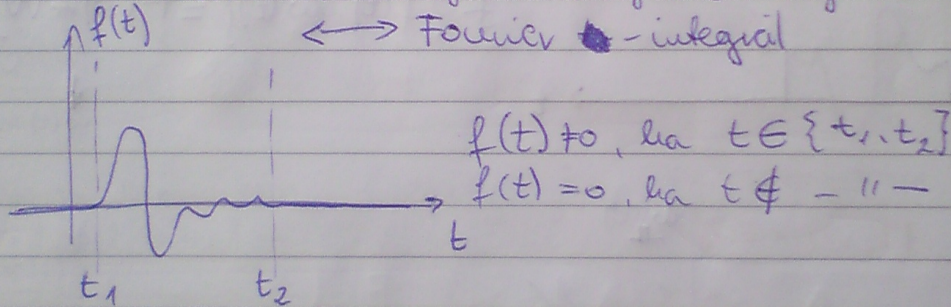
$\sim f_0$  idővel: 200 Hz

$$\cos x = \frac{e^{-jx} + e^{jx}}{2} \Rightarrow x(t) = \sum_{k=-\infty}^{\infty} C_k e^{j k \omega_0 t}$$



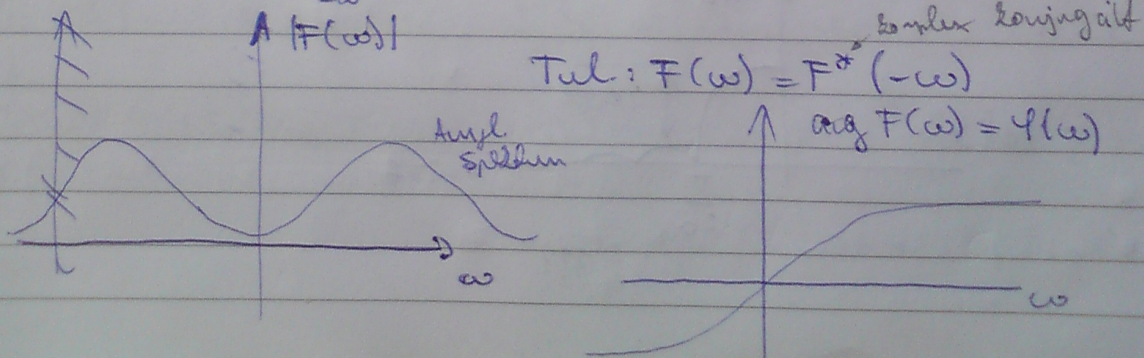
teljesítmény: Teljesítmény:  $P = \sum_{k=-\infty}^{\infty} |C_k|^2$

b.) Lokálisul (1x1 lefutású jel = single-shot)  
 $\longleftrightarrow$  Fourier  $\bullet$ -integral



$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt = \mathcal{F}\{f(t)\}$$

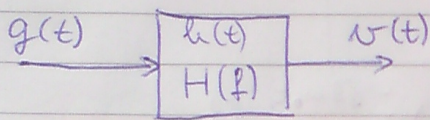
$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{+j\omega t} d\omega = \mathcal{F}^{-1}\{F(\omega)\}$$



ellenőrzés: Energia:  $E = \int_{-\infty}^{\infty} f^2(t) dt = \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega$

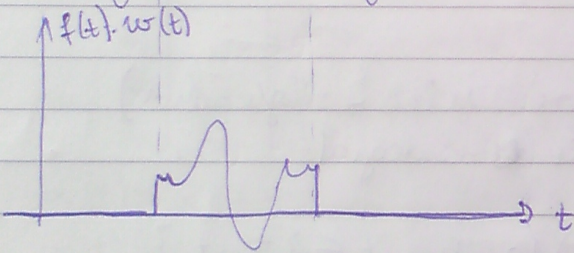
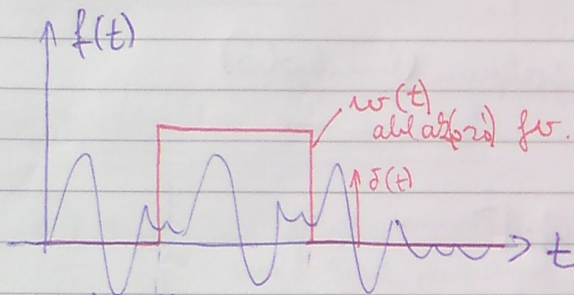
$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} f^2(t) dt$

Lineáris nss. hataisa az átított jelre



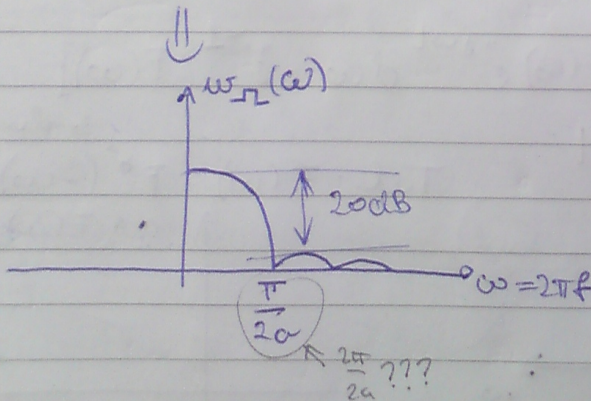
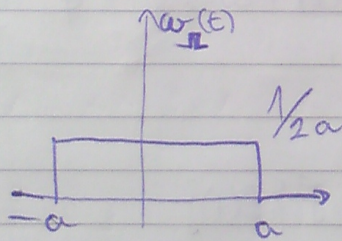
$w(t) = g(t) * h(t) = \int_{-\infty}^{\infty} g(\tau) h(t-\tau) d\tau$

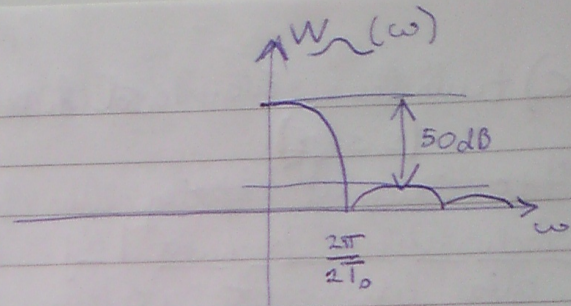
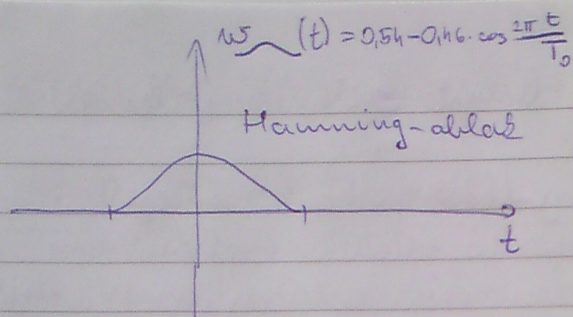
$V(\omega) = H(\omega) G(\omega)$  /abszolútértékkel kell ide!/



Spektrum ↔ gördülő spektrum

$\mathcal{F}\{f(t) \cdot w(t)\} = F(\omega) * W(\omega)$   
 $W(\omega) \mathcal{F}(\omega)$   
 $\downarrow$   
 $w(t) = 1$

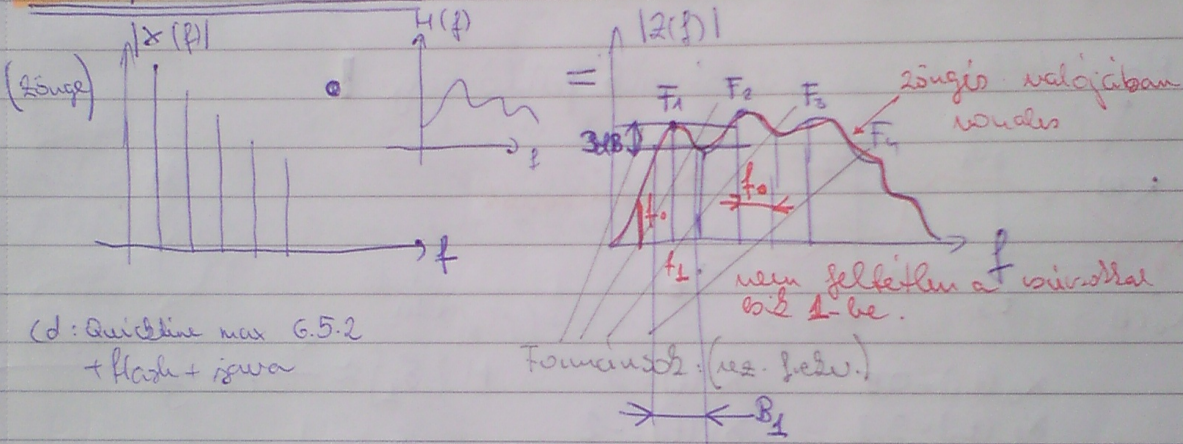




gyab-ban: Matlab, Simulink

5. eq. - 2009.09.21

Thallg valba a krombe, ah. 1x

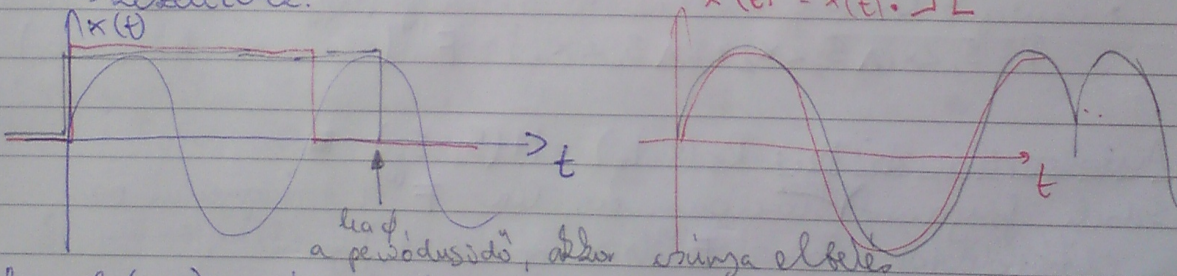


(d: Quicktime max 6.5.2 + flash + java

- niggyszög, hanning, ... ablak (+hanning)

Spektum számítási a valóságban: DFT, FFT

↳ kiegészítés!



ha szel (~10) periódust vágunk ki, az. nem nagy az elbesz

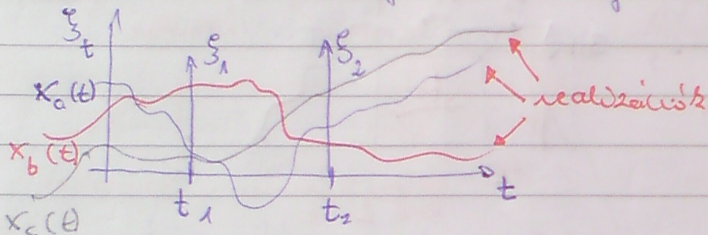
↓  
Zöngésinken, Zöngés szinuszok elbesz (pitch synchronous, asynchronous analysis)

c) turbulens áramlással képzett hangok spektrális jellemzése  
(S, sz, f)

Stacionárius stochasztikus folyamatok spektrális sűrűsége  
stochasztikus folyamat: def.: használt nélkül mechanizmus  
által létrehozott végtelen számosságú sokaság egyetlen tagja.  
(pl.: autót liften operátorral az úton)

Sokaság: stoch. folyamat  
komponens: időfü., realizáció

Modellalkotás a felhasználó felelőssége!



$x_a(t_1) = 0,5$	$x_a(t_2) = 3$
$x_b(t_1) = 3,2$	$x_b(t_2) = -2$
$x_c(t_1) = -0,4$	$x_c(t_2) = -1$

valószínűségi elvárás

$$M\{S_1\} \approx \frac{x_a(t_1) + x_b(t_1) + x_c(t_1)}{3} = 1,2$$

$$M\{S_2\} = \frac{x_a(t_2) + x_b(t_2) + x_c(t_2)}{3} = 0$$

Stochasztikus folyamatok teljes értékűen ismerjük, ha:

- ismerjük N-edrendű valószínűségi eloszlás függvényét. ■

$$P\{S_1 < x_1 \wedge S_2 < x_2 \wedge \dots \wedge S_n < x_n\} = F^{(n)}(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n)$$

Autokorrelációs fu.:  $R_S(t_1, t_2) = M(S_1; S_2)$  ■

Stoch. folyamat <sup>stacionárius</sup> stacionárius, ha  $F^{(n)}(\cdot)$  független az

időeltolástól, azaz:  $F^{(n)}(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) =$

$$= F^{(n)}(x_1, x_2, \dots, x_n, t_1 - t_0, t_2 - t_0, \dots, t_n - t_0)$$
 ■

Stoch. folyamat stacionárius, ha valószínűségi elvárás konstans,

azaz  $M(S) = \beta \in \mathbb{R}$

• az autokorr. fu-e csak  $(t_2 - t_1)$ -től függ,  $R(t_2 - t_1) = R(\tau)$ .

• az  $R_S(\tau)$  0-lal folytonos. ■

Spektrális sűrűség:  $S_g(f) = \mathcal{F}\{R_g(\tau)\} = \int_{-\infty}^{\infty} R_g(\tau) e^{-j\omega\tau} d\tau$

$R_g(\tau) = \mathcal{F}^{-1}\{S_g(f)\} = \int_{-\infty}^{\infty} S_g(f) e^{j2\pi f\tau} df$

Ha egy <sup>stationárius</sup>  $\xi_t$  stochasztikus folyamat  $\neq$  realizációja átlalad egy  $H(f)$  átviteli fű. időinv. lin. rsz.-en, akkor a kimeneten megjelenő  $\eta_t$  stochasztikus folyamat (ami a  $\xi_t$   $H(f)$ -en átlaladva realizációjából áll) mindig stationárius, s

$$S_\eta(f) = |H(f)|^2 \cdot S_g(f)$$

Ergodikus 1. folyamat, ha a várható érték átlagolás jó közelítéssel megegyezik az egy realizációból származott átlaggal

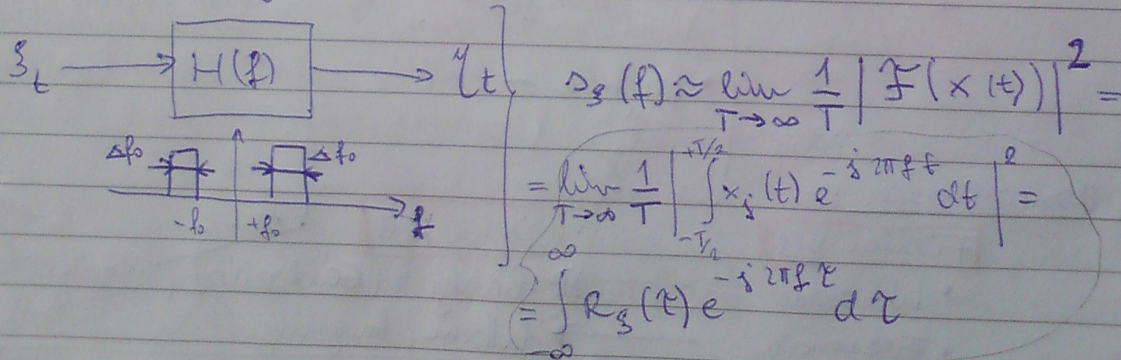
Pl.: az autokorreláció  $R_g(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{+T/2} x_j(t) x_j(t+\tau) dt$

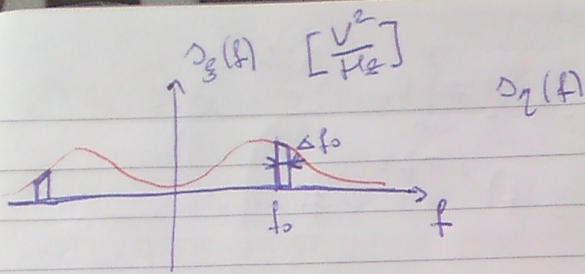
Általában: Ha  $\mathcal{F} = M\{g(\xi_1, \xi_2, \dots, \xi_n)\}$

~~$\mathcal{F} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{+T/2} g_j(t_1) g_j(t_2) \dots dt$~~

↳ utarész...  
M.M.M.

- Egy ergodikus folyamat mindig stationárius.
- Ha 1 foly. erg. az autokorrel. fű.-e, akkor



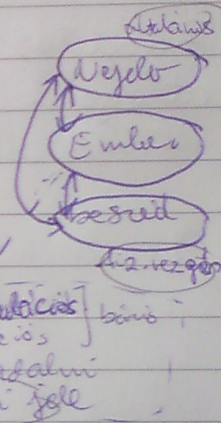


$$P_S(f_0, \Delta f_0) \approx 2 S_S(f_0) \Delta f_0$$

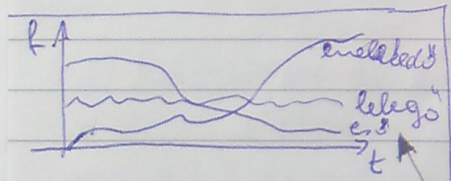
6. eo. - 2009.09.25

⑧ A magyar beszéd finom szerkezete

A beszéd abs. szerk. iróg.  
 bonyolult → más rezgésel (fűtő, szivár, stb.)  
 Miért? → beszéd: komplex rendszer eredménye



Mi van szűkösség? • Mondattan — hangjelzés



- hangsúly előjelezés
- szünettartás

- Morfológia — szótárszerkezet (toldalék, rag, oszlat)
- Fonológia — absztrakt módon foglalkozik a nyelvi vázral.

//hasznellás, hanghíres, hangfeloldás//  
 elmondta  
 fűtő → fűtőszék  
 ↳ hiánytöltés

- fonetika — konkrét realizációkkal foglalkozik

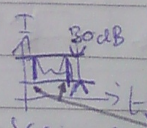
Kapcsolatközpont:

**Fizikai jel**

- Intenzitás
- Frekvencia
- Időjél (karban, les - hang, rezgés, ...)

**Nyelvi jel**

- hangzó (válaszadás), hangsúly
- dallam, hangsúly
- hangjelölés, dallamjelölés, ritmus



működési események  
 ↳ több mondatra



(CVVC)

A nyelv és a beszéd kapcsolata:

A nyelv szimbólumrendszerét alkotó építőelemek a fonémák.  
A fonéma a nyelvi sz. legkisebb eleme, mely jelentést megkülönböztető szerepet lát

PH vad → veir : a, é fonéma

soz → zob : b, h fonéma

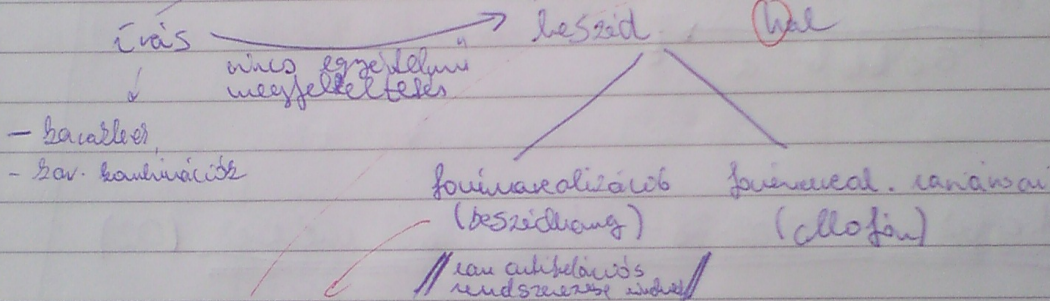
V: 14 db (alap 9 + 5 db hosszú)

C: 50 db (alap 25 + hosszú- rövid pár)

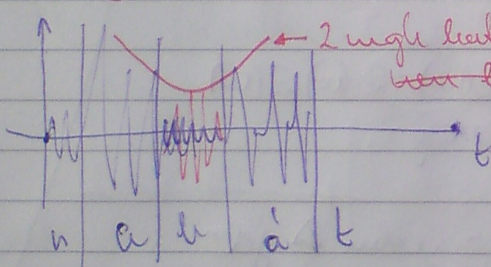
Σ 64!

bajt - 8 bit  
szépséget  
nem tisztított fonéma az a

Fonéma felépítése:



h - nek 3 variánsa van



írásig ugyanaz → más beszédhang  
flaagat vaudal

az a hang → más írásig  
j jly

esteje: gípfü ~ ✓  
esteje: karáson ~ ✓

Nemzetközi jelölésrendszer (hangok) flaagat /j/ ipa jh.

IPA - internat. phonetic association

SAMPA - szg. áll. időse

beszédszókészlet elnevezés: hangtöbbség C, V

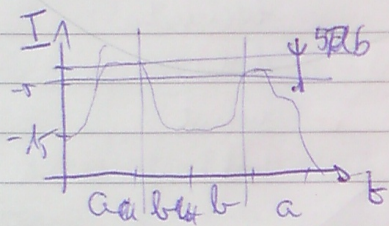
hangszalék

megnevezés szerint  
- a beszéd alap építő elemei  
(alakától függetlenül)

szópra segédlet szerint  
Hangszalék! Segédlet szerint alakul  
melyekből áll az összerakás

- hangok (64 fonéma + variációk) ○○○○○○ alakuktól függő
- hangintenzitások: ○○○○○○
- specifikus intenzitások: CP 100. oldal

(• - II - időtartamok  $\square\square$  arányúak az  
ideiglenes/leggyakoribb rendszámot tudni (C, V)  $L_{hang/sec}$  (10-20)



1. gyűjtés - 2009.09.28 - akusztikai mérés (02)

□ hangkeltésére nem (melod)

- hangszalék: zöngés gép
- az. paraméterek:  $P(\omega) = F(\omega) \cdot U(\omega) R(\omega) Z(\omega)$   
fónia rez. üveg

F: fónia · zöngés | zöngés v. zöngétlen:  $\Delta$  jel  
zöngés (s. sz) | hangszal. rezg. hossza léte  
felállítás

- U: rezonátor üveg → szűrő (kiegyenlítés)
- R: sugárzási ellenállás → hő, szűrő helye  
nagy hő, nagy elnyelés.
- Z: környezeti gáz → mindig MINDIG

megh: az anyag nagy része!

modális traktus  
Fonéma: a rez. üreg / abl. felcsúszott kéllangyadatok } !  
↑  
előtér helyei

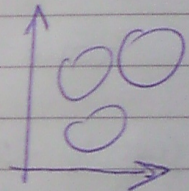
rez. üreg tel-a,  
-szegmentális és supra-segmentális

Feladat: 3 magh. i/u/a } 1. 2 fonéma felelőseket  
magas-nagy jelleg: 2. fonéma alapul

Spektrum ↑ intenzitás == kéllangy  
t ritmus = kéllangy = fonéma

Seán → cör, táv: 1. kéllangy

magh. + kéllangy első 2. fonéma kéll



2h minta feladatok

- 1) a v. sz: 5000 felett sz. } "zaj" / "sz"
- 2) magh
- 3) L felkattanó hang.

kife v. sz: 150kHz → f<sub>h</sub>

szó

1. ee. - 2009. 09. 02

ism.: szegmentális és supra-segmentális szók.

akadémikus jelleg: 10-20 szók. ciklus: 13-14

hangszíny, lecsúszás, ritmikai jellemzők

hangidőt. ad. sz. váltás minetes

PROZÓDIA

