

A BESZÉD AKUSZTIKAI FONETIKAI LEÍRÁSA

Vicsi Klára

1. Bevezetés

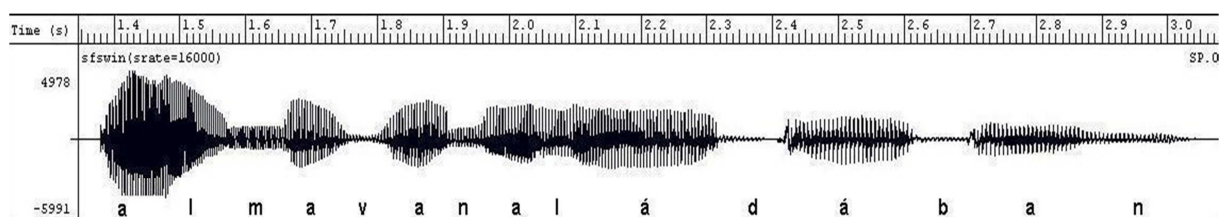
A beszélt nyelv az emberek egymás közötti kommunikációjának az eszköze. A beszélő agyában megszülető nyelvi formába öntött üzenet a beszédprodukciónál a levegő közvetítésével továbbítódik a hallgató füléig, ahol az üzenet beszédérzékelési és feldolgozási folyamat során a hallgató agyában újraformálódik. Tehát embertől emberig a kommunikációban a közvetítő közeg a levegő, amelyben a nyelvi üzenet a hanghullámok formájában továbbítódik. Ezekhez a hanghullámokhoz adódik a környezetben keletkező zaj, ezeket a hanghullámokat torzítják el a falakról visszaverődő hullámok, például egy erősen visszhangos teremben. A pszicho-akusztika a beszédpercepció fonetika és a pszicholingvisztika vizsgálja, hogy mi a kapcsolat a produkcióval létrejött beszédhanghullámok akusztikai tulajdonságai és a hallgatóban keletkező beszédérzet és megértés között. Tehát, a teljes kommunikációs folyamat megismeréséhez a beszédprodukciónál létrehozott beszéd fizikai tulajdonságaival, vagyis a beszéd akusztikai leírásával, is meg kell ismerkednünk. A tárgyalást néhány akusztikai alapfogalom ismertetésével kezdjük, mint például a rezgés frekvenciája, valamint a hangnyomásszint fogalmának bemutatása, majd rátérünk a különböző frekvenciaelemzési módszerek rövid vázlatos leírására. Rövid áttekintést adunk a beszéd akusztikai fonetikai fogalmairól, külön tárgyalva a beszéd szegmentális, és szupraszegmentális jellemzőit. A leírás rövid és összefoglaló jellegű. Azt tartottuk szem előtt, hogy olyan fogalmakat, módszereket ismertessünk röviden, amelyek elengedhetetlenek a kézikönyv fejezeteinek a megértéséhez.

1.1. A beszélt és írott nyelv közötti különbség

A beszélt és írott nyelv feldolgozása lényegesen különbözik egymástól. Annak ellenére, hogy a gyermeki nyelvfejlődés során a beszélt nyelv feldolgozása alakul ki először, a beszélt nyelv feldolgozásáról lényegesen kevesebbet tudunk, mint az írott nyelvről (Ferreira, F., Anes, M., 1994).

Az egyik nagy különbség a beszélt és írott nyelv feldolgozásában az, hogy az írott nyelvnél az olvasás sebességét az olvasó a feldolgozási sebességéhez illesztheti, lelassíthatja, felgyorsíthatja az olvasást, visszamehet, megállhat, míg a beszélt nyelv esetében a hallgatónak a beszélő tempójához kell alkalmazkodnia.

A másik nagy különbség az, hogy a beszéd képzésekor a folyamatos artikuláció következtében létrehozott akusztikai hullámforma folyamatos. A beszédhangok, amelyek a beszéd folyamat minimális egységei, az illető nyelv hangrendszerében elfoglalt helyétől függetlenül, nincsenek elkülönülve, az egyik beszédhangból az átmenet a másik beszédhangba folytonos, mint ahogy az artikuláció is az. A szavak nincsenek szünetekkel elválasztva, mint az írásban. Egy mondat hullámformájának (a hangnyomás időbeli változásának) alakulására példát az 1. ábra mutat.



1. ábra A beszédhangnyomás időbeli változása az „A l m a v a n a l á d á b a n,” mondat kiejtésekor

A szavak között általában nem tartunk szünetet, egyes fonológiai szabályok a szavak határain túl is érvényesek. Szünetet, ami rendszerint levegővétellel jár, főleg egy-egy frázis vagy mondat végén tartunk. Kontextuális információ kell ahhoz, hogy a szavak határait meg tudjuk határozni. Beszédünk úgy alakult ki, hogy a kontextuális információ mellett a prozódiai információk, a hangsúly, a hanglejtés, a beszédtempó és a hangszín segítik az emberi feldolgozó rendszert a szegmentálásban.

A harmadik nagy különbség a beszéd redundanciája. A nyelvi tartalom mellett a beszéd számos egyéb információt hordoz. Például a beszélő nemét tükrözi, kifejezi érzelmi állapotát, fizikai, egészségi kondícióit stb.

Nagy különbség továbbá az, hogy egy-egy nyelvi egység - mint például a beszédhang, szótag, szó stb. -, hossza időben igen erősen változik beszélőtől függően és egy beszélőnél is, és ez a változás nem egyforma mértékű az egységet felépítő elemeknél. Például egy szóban egyes beszédhangok, beszédhangrészek hossza különböző mértékben változik a beszédtempó függvényében.

A beszéd bonyolultsága, a beszéddel történő kísérletezés nehézségei, továbbá a beszédkísérletek költséges volta mind-mind hozzájárulhat a beszélt nyelvfeldolgozás elmaradásához az írott nyelvhez képest.

1.2. Nemzetközi fonetikai jelölésrendszer

A hangjelölő írás kialakulását követően még közelítőleg 300 év kellett ahhoz, hogy általánossá váljék használata. Szinte lehetetlennek látszott a hangzó beszéd sokszínűségének ábrázolása alig több mint két tucatra tehető írásjelből álló jelkészlettel. Továbbá, a beszéd zenei elemeinek (prozódiai jellemzők) jelölésére a írásjelek nem, illetve csak nagyon részlegesen adnak lehetőséget. Mindenesetre, a hangjelölő írás kialakulásakor a írásjelek megfeleltek a hang-típusoknak. Évszázadok során az írás keveset módosult, viszont a beszélt nyelv dinamikusan változott. Az írás és a kiejtés egyre jobban eltávolodott egymástól (Kassai I., 1998).

Szükségessé vált a kiejtés lejegyzése nemzetközi jelölőrendszerrel, ami tükrözi egy-egy beszédhang kiejtését. Így alakult ki a nemzetközi jelölésrendszer. 1889-ben a Nemzetközi Phonetikai Társaság létrehozta az IPA International Phonetic Alphabet szimbólumkészletet.

A beszéd nyelvi leírásánál azokat a beszédhangokat, amelyeknek a szavakban jelentésmeghatározó szerepük van, fonémáknak nevezik. A magyar nyelvben például a rövid 'ö' hang és a hosszú 'ő' hang két különböző fonéma, hiszen pl. az 'öt' és az 'öt' szavak más jelentenek. Hogy egy-egy beszédhang fonéma vagy nem, az mindig az adott nyelvtől függ, vagyis a fonémák egy nyelv rendszerén belül definiálhatók.

A különböző fonetikai átírásoknál a fonémák lejegyzésére általában az IPA szimbólumrendszert szokás használni (Lass, N. J. 1996). A magyar nyelvben 64 fonéma, 14 magánhangzó és 50 mássalhangzó van.

Sajnos ez az IPA jelölésrendszer nem illeszkedik a számítógép billentyűzetéhez. Különböző segédprogramok készültek az IPA jelölésrendszer gépi használatához, de a megjelenített karakterek formái erősen rendszerfüggőek voltak és sok bosszúságot okoztak a kutatóknak. Ezért nemzetközi szinten bevezetésre került egy újfajta, úgynevezett SAMPA jelölésrendszer, amely alkalmazkodik a számítógéppel kezelhető karakterkészlethez (Wells, J.C., 1997). Ezzel a számítógépes gépelés és továbbítás egyszerűen megoldható, ellentétben a hagyományos IPA jelölésrendszerrel. A magyar fonémák SAMPA jelölésének rendszerét az 1. táblázat tartalmazza (Vicsi, K. 1996, Vicsi, K. 1998)

1. táblázat: A magyar nyelv fonémáinak és néhány tipikus allofónjának (kiejtésvariánsnak) SAMPA jelölésrendszere

Betűk	Példák	IPA	SAMPA
a	hat	ɔ	O
á	hát	a:	A:
e	vet	ɛ	E
é	vét	e:	e:
i	hűt	i	i
í	szít	i:	i:
o	sok	o	o
ó	sók	o:	o:
ö	köt	ø	2
ő	söt	ø:	2:
u	fut	u	u
ú	kút	u:	u:
ü	süt	y	y
ű	füt	y:	y:

Betűk	Példák	IPA	SAMPA
p	pad	p	p
b	bab	b	b
t	tél	t	t
d	dél	d	d
k	kép	k	k
g	gép	g	g
c	cél	tʃ	tʃ
dz	bodza	dʒ	dʒ
cs	cső	tʃ	tʃ
dzs	dzsem	dʒ	dʒ
ty	tyúk	c	t'
gy	gyár	j	d'
f	fél	f	f
v	vér	v	v
sz	szép	s	s
z	zaj	z	z
s	só	ʃ	S
zs	zsír	ʒ	Z
h	hét	h	h
r	réz	r	r
l	lép	l	l
j	jön, lyuk	j	j
m	méz	m	m
n	néz	n	n
ny	nyom	ɲ	J

A /h/, /j/, /n/ és /m/ fonémák

megkülönböztetett ejtés-

változatai (allofónjai)

]	ɣsɔ]	č	x)
---	------	---	----

m	kánfőz, hamvas horn-ágy, horváltás	m̥	F
n	ing török	n̥	N

h	doh	x	x
h̥	ihlet	x̥	x
h	lehet	fi	h̥

A hosszan ejtett hang jele minden esetben a megfelelő SAMPA szimbólum után tett kettőspont.

Évszázadok alatt a beszéd dinamikus változása miatt az írás - beszéd viszonya bonyolult szabályrendszerre vált. A jelenség nyelvfüggő. Magyar nyelvre az írás - beszéd átírás, vagyis az írott szöveg kiejtés szerinti lejegyzése, szabályrendszerrel megoldható. Ilyen szabály például a zöngés hasonulás (napban /nOb:On/) vagy a frikatív összeolvadás (metszet /mEts:Et/) szabályai. Angol nyelvre ez az átírás lényegesen bonyolultabb. Az írott szöveg kiejtés szerinti lejegyzési (fonetikai átírási) szabályaira a beszédtechnológiában például a beszédészítésnél (automatikus szövegfelolvasó rendszerek) van szükség, a kiejtés írott szöveg átalakítására pedig a gépi beszédfelismerésnél (automatikus beszéd lejegyző rendszerek).

2. Akusztikai alapfogalmak

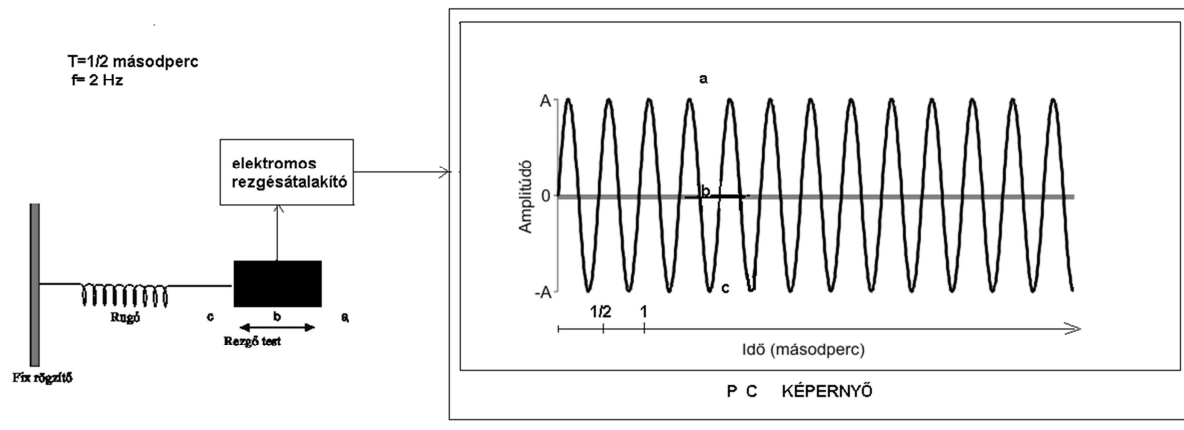
2.1. A rezgő mozgás, a hang keletkezése

A hang rezgés révén jön létre. A rezgő húr, cintányér, hangvilla rezgését átadja a környező levegő molekuláinak, úgy, hogy a környező térben a molekulák sűrűsödése és ritkulása jön létre. Ezeknek a sűrűsödéseknek és ritkulásoknak hatására a környező levegőben folytonos nyomásingadozások alakulnak ki, amelyek a levegő molekuláinak a segítségével, a molekulák egymás közötti rezgési energiájuk átadásával, hanghullámok formájában a levegőben tovább terjednek és a dobhártyát rezgésbe hozzák. Ahhoz, hogy a hangként megjelenő rezgés jellemzőit megérthessük, szükséges a rezgés fogalmát meghatároznunk. Azokat a fizikai folyamatokat nevezzük rezgéseknek, amelyek meghatározott időközönként újra meg újra

ugyanazt az állapotot éri el, vagy ugyanazon állapoton haladnak át (Tarnóczy T., 1984.). Ez az oszcillálás lehet periodikus vagy rendezetlen, véletlenszerű.

2.1.1. Harmonikus rezgőmozgás

A legegyszerűbb rezgés a harmonikus rezgő mozgás, amelynek grafikonját a számítógép képernyőjén kirajzolhatunk. Vegyünk egy rugót. Egyik végét rögzítsük, a másik végére helyezünk egy m tömeggel bíró testet (egy elméletileg súrlódásmentes felületre), a 2. ábra szerinti elrendezésben.



2. ábra Egy harmonikus rezgő test kitérésí grafikonja. Az ábra a periódusidőt, a kitérésí amplitúdót, és a frekvenciát szemlélteti.

Alaphelyzetben a test nyugalomban van, ez az ábrán a b pont. Mozdítsuk ki nyugalmi állapotából, úgy hogy például széthúzva rugót, a rugó erejének ellenében F erőkifejtéssel a testet az a pontba mozdítjuk, és magára hagyjuk. A test ellenkező irányban a b egyensúlyi helyzeten áthaladva a c pontig kitér, majd visszafordul és kitér az a pontig. Ez a mozgás *elméleti energiavesztés mentes esetben* periodikusan ismétlődik, tehát egyenlő időközönként kerül a test ugyanabba az állapotba. A maximális rezgési kitérés, vagyis a rezgés amplitúdója (A) nem változik. A kitérés időbeli változása szinuszos görbét ír le, ahol a kitérés $s = A \sin$

ωt , ahol a körfrekvencia $\omega = \frac{2\pi}{T}$, és $2\pi = 360^\circ$ valamint T egy teljes periódus ideje. Egy

periódus idő (T) az időtartam, amíg a rezgő test először kerül újra ugyanabba az állapotba, amelyben a periódus elején volt. Az egységnyi idő alatti ismétlődések, periódusok száma a *frekvencia* (f), mértékegysége a hertz [Hz], amelyet Heinrich *Hertz* német fizikusról neveztek el, és ami az 1[sec] alatti rezgések számát jelenti:

$$f = \frac{1}{T} \text{ [Hz]}, \text{ vagy } T = \frac{1}{f} \text{ [sec]}.$$

A Hz-nél nagyobb egység a kilohertz (1 kHz = 1000 Hz). Ha két teljes periódus lezajlik 1 sec alatt, vagyis a periódusidő $\frac{1}{2}$ sec, amint azt a 2. ábra szemlélteti, akkor a rezgés frekvenciája 2 Hz, ha 20 teljes periódus zajlik le 1 sec alatt, vagyis a periódusidő $\frac{1}{20}$ sec, akkor a rezgés frekvenciája 20 Hz. Tehát minél nagyobb a rezgések száma másodpercenként, annál kisebb a rezgési idő.

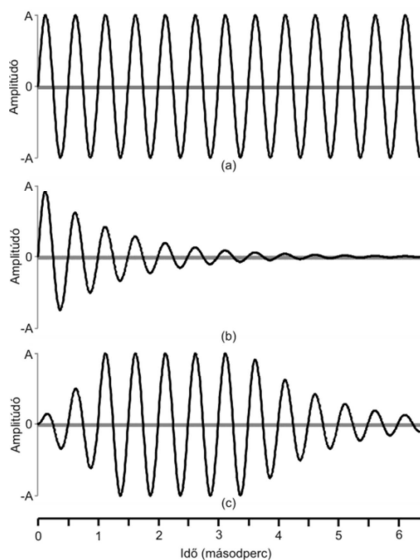
Az ábrán látható test, vagy részecske hangforrásként szerepelhet, amennyiben rezgését átadja a környező levegő molekuláinak úgy, hogy a környező térben a molekulák sűrűsödése,

ritkulása, vagyis nyomásváltozás jön létre. Elektromos átalakítóval, például mikrofonnal ezt a nyomásváltozást vesszük fel.

A szinuszos görbével leírható harmonikus rezgések, a hallgató fülében az ún. *tisztahang* érzetét keltik, csak amplitúdójukban és frekvenciájukban különböznek egymástól. Az emberi hallástartomány 20Hz-től 20 000Hz-ig terjed, tehát a mechanikai rezgésekből az emberi fül csak ezt a tartományt képes érzékelni. Idős korra a felső frekvenciahatár jelentősen lecsökkenhet.

A szinuszos formájú tisztahang ritka jelenség a hangforrások világában.

A fenti példánál energiaveszteség mentes rezgést feltételeztünk, vagyis a rezgés amplitúdója konstans. Ez a *csillapítatlan rezgés*, amely a 3. a. ábrán látható, ahol a periódusidő $T=1/2$ sec. Valójában a környezetünkben az egyensúlyi helyzetéből kimozdított és magukra hagyott rugalmas testek csökkenő amplitúdójú csillapodó amplitúdójú rezgést végeznek. A rezgő test energiájának egy része a súrlódás révén hővé alakul át, a másik része pedig hangjelenség formájában kisugárzódik és a rezgés csökkenő amplitúdójú, változatlan frekvenciájú szinuszos rezgés lesz, amelyre példát a 3.b. ábra mutat. Az egyensúlyi helyzetéből kimozdított és magukra hagyott rugalmas testek (például egy megpendített húr, egy megkoccintott üvegpohár stb.) ilyen csökkenő amplitúdójú, ún. *szabad rezgést* végeznek, és a rezgés frekvenciája, a testre jellemzően mindig ugyanaz. Ezt nevezik a test *természetes* vagy *sajátfrekvenciájának*.



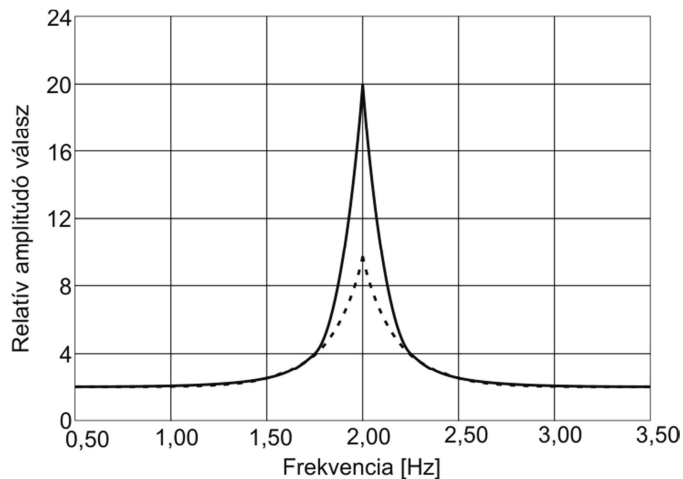
3.ábra Csillapítatlan (a) és csillapodó (b) rezgésforma, (c) berezgés, lecsengés folyamata

2.1.1. Kényszerrezgés, rezonancia

A rezgés folyamatossá tétele külső energia bevitelével lehetséges, azért, hogy az energiaveszteséget pótoljuk. Ha az egyszerű rugó-test példánkon a 2. ábrán a rugó baloldalán a fixen rögzített pontot külső erővel a rugó tengelyében előre hátra mozgatjuk, akkor a testre kényszerítő erőt fejtünk ki és a test kényszerrezgést végez. A *kényszerrezgés* frekvenciáját a kényszerítő erő (gerjesztő erő) frekvenciája szabja meg. Az amplitúdó a kényszerítő erő amplitúdójától és a kényszerítő erő frekvenciájától függ. Egy adott amplitúdójú gerjesztő erő esetében a kényszerrezgést végző test amplitúdója akkor lesz a legnagyobb, ha a kényszerítő (gerjesztő) frekvencia megegyezik a kényszerített (gerjesztett) rendszer sajátfrekvenciájával. Ezt nevezik a *rezonancia* jelenségének, tehát amikor is maximálisan együtt rezeg a

kényszerítő rendszer a kényszerítettel. Azt a frekvenciát, amin ez bekövetkezik, *rezonanciafrekvenciának* nevezik. A frekvencia függvényében felvett rezgés amplitúdó görbét rezonanciagörbének nevezük. A kényszerrezgés amplitúdója a kényszerítő rezgés amplitúdójának sokszorososa lehet.

A *rezonanciagörbe* alakja függ a csillapítástól és a súrlódástól.



4. ábra A 2 Hz sajátfrekvenciájú test kényszerrezgésekor kialakuló rezonanciagörbe a kényszerítő frekvencia függvényében. A szaggatott görbe nagyobb csillapítás esetében alakul ki

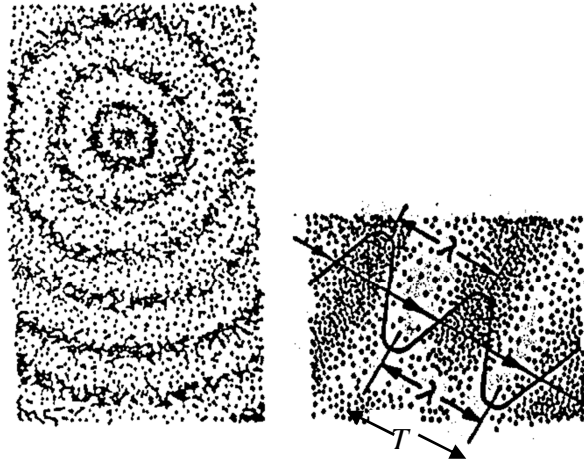
Ha egy rezgésre képes testet rezgésbe hozunk, azt tapasztaljuk, hogy a rezgés amplitúdója csak bizonyos idő múlva éri el a maximumot. Ez a *berezgés* jelensége, amely a gerjesztés kezdeti pillanatától az *állandósult állapot* elérésig tartó átmeneti állapot. Az *állandósult állapotot* a *lecsengés* követi, amely a gerjesztés megszűnésének pillanatától a nyugalmi állapot eléréséig vagy másik kényszerítő belépéséig tartó átmeneti állapot. A berezgést és a lecsengést, amelyet a 3. ábra c képe szemléltet, *átmeneti rezgési formának* vagy *tranzienst állapotnak* nevezük.

2. 2. Hang terjedése levegőben

A levegő elemi részecskéi nyugalmi állapotban állandó, rendezetlen mozgásban vannak, de úgy, hogy minden részecskének van egy átlagos „stabil” mozgási állapota, meghatározott távolsága a többi részecskétől, amelyben szeretnek tartózkodni. Tengerszinten nyugalmi állapotban $p_0=1[\text{atm}]$ nyomás mérhető. Ha valami a részecskéket ebből az állapotból kimozdítja, olyan erők keletkeznek, amelyek igyekeznek a részecskéket az egyensúlyi helyzetükbe vissza helyezni. Amikor egy test (a hangforrás) rezeg, a szomszédos levegő részecskéit a nyugalmi állapotból kimozdítja, és velük együtt rezeg, kimozdítva némi késéssel a távolabbi szomszédos részecskéket is azok nyugalmi helyzetéből. Vagyis a zavar hatása terjed tova, a részecskék csak az egyensúlyi helyzetük körül rezegnek, átadva a zavarás hatását a szomszéd részecskéknek. A hanghullámterjedés tehát a zavar mozgásának a terjedése a hangot közvetítő közegben, például levegőben, úgy hogy maguk a részecskék nem haladnak együtt a hullámmozgással. A hanghullámok a levegőben úgy terjednek, hogy a részecskék a hullám terjedési irányában rezegnek. Ezek az úgynevezett *longitudinális hullámok*, amelyeket az 5. ábra mutatja. A víz felszínén terjedő hullámoknál a részecskék le-

föl mozognak merőlegesen a terjedés irányára. Ezek a *transzverzális hullámok*. Itt is a vízrészecskék csak le-föl mozognak, és nem utaznak a hullámmal együtt.

A vivő közeg, amely valamilyen mechanikai rezgés hatását közvetíti, lehet légnemű, cseppfolyós, szilárd.



5. ábra A levegőrészecskék sűrűsödése és ritkulása. Két egyenlő sűrűségű hely távolsága a nyomáshullámhossz (λ)

A hang terjedési sebessége(c):

$c = \frac{\lambda}{T} = \lambda \cdot f$ [m/sec]. A továbbiakban csak a levegőben terjedő hanghullámokkal

foglalkozunk ahol a hang terjedési sebessége $c = 331,5$ [m/sec]. $0\text{ }^{\circ}\text{C}$ -on és 1 [atm] ($100\ 000$ [Pa]) nyomáson, $c = 343,5$ [m/sec], $20\text{ }^{\circ}\text{C}$ -on és 1 [atm]nyomáson.

A *hang hullámhossza* (λ), a hanghullám (T) periódusidő alatt megtett útja. A hang hullámhossza és a hang frekvenciája (f) fordított arányban állnak egymással. Egy 20Hz -es hang hullámhossza $16,6\text{m}$, egy $20\ 000\text{Hz}$ -es hang hullámhossza $1,66\text{cm}$.

A hanghullámok terjedésénél, mint minden hullámformánál, általában előfordulnak visszaverődések és elhajlások.

Hangnyomás ($p_{hang}(t)$):A hang terjedésekor a részecskék sűrűsödése és ritkulása egy adott pontban $p_{hang}(t)$ nyomásváltozást eredményez. Ez a nyomásingadozás igen kicsi és a légköri (sztatikus) nyomás értékére szuperponálódik, vagyis annak hangfrekvenciás ingadoztatásában nyilvánul meg. A nyomás időbeli változása tehát

$$p_{légköri} + p_{hang}(t)$$

alakban jelentkeznek. Maga a $p_{hang}(t)$ függvény tartalmazhat periódusos és statisztikusan ingadozó elemeket, de az ún. alaplajtól eltekintve véges ideig tart és rendszerint rezgési és lecsengési elemekkel is rendelkezik. Tehát matematikailag rendkívül bonyolult függvény. Ezért a megismerés formája rendszerint nem az időbeli lefolyás rögzítése, hanem valamilyen időbeli átlag, leggyakrabban a négyzetes középérték, az úgynevezett effektív érték megállapítása:

$$p_{eff} = \sqrt{p^2(t)} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} p^2(t) dt}$$

A négyzetes középérték mérése nemcsak fizikai ok, hanem egyben biológiai tapasztalat is. A fül ugyanis az ún. effektív értéket érzékeli. Tárgyalásaink során *hangnyomáson mindig effektív hangnyomást értünk* [N/m²], és (*p*)-vel jelöljük

Néhány nyomásérték összehasonlításként:

sztatikus nyomás	→ 1 atm ~ 10 ⁵ Pa ~ 10 ⁵ N/m ²
beszéd nyomásingadozás	→ 10 ⁻² - 10 ⁻¹ Pa
hallásküszöbnél a nyomásingadozás	→ 2 · 10 ⁻⁵ Pa
fájdalomküszöbnél a nyomásingadozás	→ 20 Pa

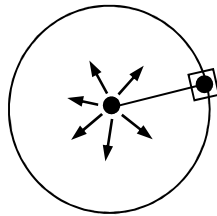
A *hangforrás elsődleges adata a hangteljesítménye (P)*: a hangforrás körüli képzett gömbfelületen időegység alatt átáramlott összes energiamentiség [Watt].

$$P = \frac{P_{eff}^2}{\rho_0 \cdot c} \cdot S$$

ahol S a felület [m²].

A hangteljesítmény tehát független attól, hogy hol végezzük a mérést, de meghatározáshoz sok mérést kell végezni. A hangteljesítmény „mennyiségi”, tehát összegező adat: az elemi értékek, a felületegységre eső teljesítménynek, vagyis az intenzitásnak a sugárzó körüli teljes gömbfelületen vett integrálja.

A *hang intenzitása (I)*: egységnyi felületen merőlegesen (ld. 6. ábra) időegység alatt átáramlott energia [W/m²],



6. ábra A *hang intenzitása az egységnyi felületen merőlegesen, időegység alatt átáramlott energia*

$$I = \frac{P_{eff}^2}{\rho \cdot c}$$

ahol $\rho \cdot c$ a közegre jellemző *akusztikai keménység*, amit az adott közegben a hang sebessége c [m/s], és a közeg sűrűsége ρ [kg/m³] szab meg.

hallásküszöb:	$I_0 = 10^{-12} \text{ W / m}^2$	$p_0 = 2 \cdot 10^{-5} \text{ Pa}$
hangos beszéd	$I = 10^{-6} \text{ W / m}^2$	$p = 2 \cdot 10^{-2} \text{ Pa}$
Fájdalomküszöb	$I = 1 \text{ W / m}^2$	$p = 2 \cdot 10^1 \text{ Pa}$

2.3. Impedanciák

A Hangtér két legfontosabb térmennyisége a nyomás és a részecskesebesség. E két térmennyiség hányadosa impedancia jellegű komplex mennyiség. Hogy a különböző

impedancia fogalmak közül melyiket használjuk az azon múlik, hogy milyen jelenséget kell leírunk és azt melyik impedancia fogalommal tudjuk egyszerűbben leírni. A fajlagos akusztikai impedancia (hullámimpedancia) a legáltalánosabb, a hangnyomás és az általa okozott hangtéri részecske sebesség hányadosa. Ezzel az impedanciával szabad térben számolunk.

| A fajlagos akusztikai impedancia (Z):

$$| Z = \frac{p}{v} = \frac{F}{A \cdot v}, \left[\frac{kg}{s \cdot m^2} \right]$$

komplex menyiség:

$$| Z = \frac{\rho \cdot c k^2 r^2}{(1+k^2 r^2)} + \frac{j \rho \cdot c k r}{(1+k^2 r^2)},$$

ahol r = forrástól való távolság,

| és $k = \frac{2\pi}{\lambda}$, ahol λ a hullámhossz [m].

A $\rho \cdot c$ annak a közegnek az akusztikai keménysége, amelyben a hang terjed.

A fajlagos akusztikai impedanciát akkor használjuk, ha a hullámoknak egyik közegből a másikba haladását vizsgáljuk. Az impedancia ugyanis részben a közeg akusztikai sajátosságait, részben a közegben terjedő hullám természetét tartalmazza. A közeg sajátosságát az akusztikai keménység mutatja, mind a valós, mind a képzetes tagban. A hullámtermeztet a kr képviseli, vagyis azt mutatja, hogy a hangforrástól milyen távolságban vagyunk.

A fajlagos akusztikai impedancia valós tagja az ellenállás, képzetes tagját pedig reaktanciának nevezzük.

Igen nagy kr értékeknél, tehát távol a hangforrástól, szabad hangtérben az ellenállás $Z = \rho c$

| $\left[\frac{kg}{s \cdot m^2} \right]$ - hez tart. A reaktancia 0-hoz tart. Ha $kr = 1$, akkor mind a két tag értéke $z = \frac{\rho c}{2}$. ezek a helyek közegenként és a hangforrás frekvenciájától függően más-más r sugarú gömb felszínén fekszenek.

| Akusztikai impedancia (Z_a):

A gerjesztett hangtér valamely hullámfelületén kijelölt felületdarabra ható hangnyomás és térfogatsebesség (részecskesebesség .felület) komplex hányadosa.

$$| Z_a = \frac{Z}{A} = \frac{F}{v A^2} = \frac{p}{v \cdot A}, \left[\frac{kg}{s \cdot m^4} \right]$$

Az artikulációt modellező különböző keresztmetszetű csőmodell matematikai leírásánál használjuk az akusztikai impedanciát. Amennyiben egy csőben azt feltételezzük, hogy nincs visszaverődés, és a hullám egy irányban terjed, az akusztikai impedancia:

$$| Z_a = \frac{\rho c}{A}, \left[\frac{kg}{s \cdot m^4} \right]$$

ahol ρ a levegő sűrűsége, c a hang terjedési sebessége az adott csőben, A pedig a cső keresztmetszeti felülete. Tehát különböző keresztmetszetű csövek akusztikai ellenállása különböző.

Ha össze akarjuk hasonlítani a hangtéri fajlagos akusztikai impedanciát az akusztikai impedanciával, akkor azt látjuk, hogy a csövekben számolt akusztikai impedancia egyenlő a szabad hangtéri fajlagos impedancia és a csőkeresztmetszeti terület hányadosával.

| *A mechanikai impedancia vagy sugárzási impedancia (Z_m):*

Erőnek és az erő támadáspontjában kialakuló erő irányú sebesség hányadosa

$$z_m = ZA = \frac{F}{v}, \left[\frac{kg}{s} \right]$$

A mechanikai, vagy sugárzási impedancia akkor használandó, ha a hanghullámok és a gerjesztő forrás csatlakozását vizsgáljuk. A rezgő rendszer mechanikai sajátosságait és a hangsugárzás tulajdonságait tartalmazza. Gerjesztő források terének és a források sugárzási jelenségeinek számításában ezért ez az impedanciafogalom használatos.

Például beszédhangképzésnél a hangtérbe jutó hang energiáját és színekét a szájnyílás kisugárzásánál a sugárzási impedancia befolyásolja. Az akusztikus energia kicsatolása, vagyis a sugárzási impedancia frekvenciafüggő. Magasabb frekvenciákon a sugárzási impedancia kisebb, mint az alacsony frekvenciákon. Ez a színek felső frekvenciatartományának a megemelését eredményezi átlagosan 6 dB/oktávval.

| *Admittance, vagyis az akusztikai vezetőképesség $\left[\frac{sm^4}{kgae} \right]$:*

az impedancia reciproka értéke, szintén komplex mennyiség, a beengedés mértéke az elektromos analógiából akusztikai vezetőképességnek is lehet nevezni, amelynek valós része a konduktancia és képzetes része a susceptancia (compliance). Azt az árammennyiséget mérik, amennyi az elnyelt energia pótlására szükséges.

2.4. Szintérték – dB fogalma

Az a legkisebb hangintenzitás-érték amelyet még épp meghallunk, vagyis az úgynevezett hallásküszöb hangintenzitás-értéke $I_0 = 10^{-12} W/m^2$ azaz $0,000\ 000\ 000\ 001\ W/m^2$, egy nagyteljesítményű repülőgép zaja 10 m távolságban kb. $I = 1\ W/m^2$, ami az emberi fájdalomküszöbhez közeli érték. Ez annyit jelent, hogy a hangintenzitás értéke az emberi hallható tartományban 12 nagyságrendet fog át, vagyis a repülőgép zajának a hangintenzitása az épp meghallható hang intenzitásának 1 000 000 000 000 szorososa.

Olyan esetekben, amikor a kezelt mennyiségek mértéke több nagyságrendet átfog, célszerű szintértékként logaritmikus viszonyszámot használni. A gyakorlatban ilyen viszonyszám a deciBel[dB], ami az adott mennyiségek arányának logaritmus, 10-vel szorozva :

$$\text{decibel[dB]: } X_{dB} = 10 \log \frac{X}{X_0} = 10 \log X - 10 \log X_0$$

Az akusztikában és vele kapcsolódó tudományágakban, mint a foniátria, fonetika, pszicholingvisztika, digitális beszédfeldolgozás stb. a hangintenzitás és a hangnyomás kezelésére dB szintértéket használunk, és a viszonyítási alap a hallásküszöb-intenzitás ill. hangnyomás értéke.

A hangintenzitás szintértéke:

$$L_{dB} = 10 \lg \frac{I}{I_0} = 10 \lg I - 10 \lg I_0 \text{ [dB]},$$

ahol a viszonyítási alap a hallásküszöb intenzitásértéke vagyis $I_0 = 10^{-12} \text{ W / m}^2$.

A hangintenzitás szint tehát a hallásküszöbnél: $L_{dB} = 10 \lg 1 = 0 \text{ dB}$,

a hangos beszédnél: $L_{dB} = 10 \log \frac{10^{-6}}{10^{-12}} = 10 \log 10^6 = 10 \cdot 6 = 60 \text{ dB}$,

egy nagyteljesítményű repülőgép zaja esetén $L_{dB} = 10 \log \frac{10^0}{10^{-12}} = 10 \log 10^{12} = 10 \cdot 12 = 120 \text{ dB}$.

A hangintenzitás a hangnyomás négyzetével arányos. Szintben kifejezve:

$$L_{dB} = 10 \log \frac{I}{I_0} = 10 \log \frac{p^2}{p_0^2} = 10 \cdot 2 \cdot \lg \frac{p}{p_0} = 20 \lg \frac{p}{p_0}.$$

A hangnyomás szintértéke:

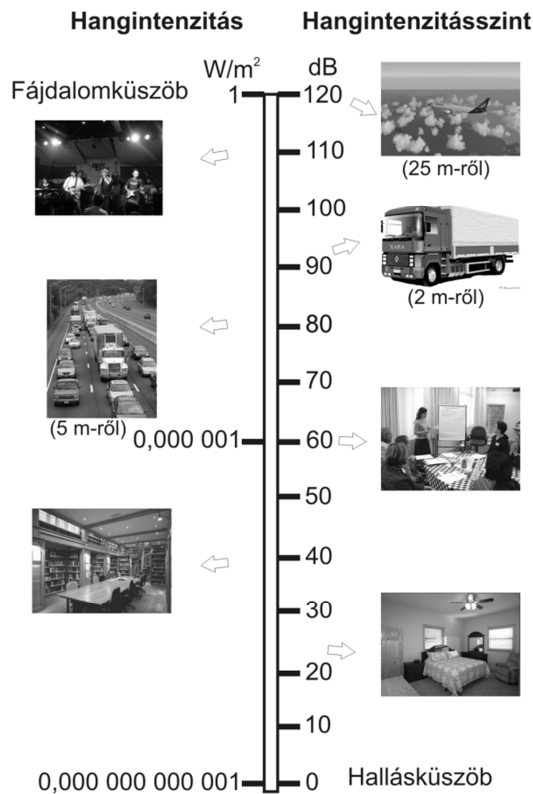
$$L_{dB} = 20 \lg \frac{p}{p_0} = 20 \lg p - 20 \lg p_0 \text{ [dB]},$$

ahol a viszonyítási alap a hallásküszöb hangnyomásértéke, vagyis $p_0 = 2 \cdot 10^{-5} \text{ Pa}$.

Ha tehát hangnyomás arányokkal számolunk dB-ben, akkor a hangnyomásarányok logaritmusának húszszorosát kell vennünk

Igy akár hangnyomás, akár intenzitásszint értékekkel számolhatunk, a szintértékek nagysága egyenlő.

A teljes hallástartomány dinamikában a 7. ábrán látható.



7. ábra A teljes hallástartomány dinamikában

Általában az akusztikai mérés technikában a sima dB-ben kifejezett L szintértékek mellett gyakran használják még az A frekvenciasúlyozású - hallási érzékenység frekvenciafüggését figyelembe vevő - L [dBA] szintértékeket. Továbbá, a zajakusztikában elterjedt az idő-egyenértékű zajszint használata az Leq szintérték, amely olyan szintértéket jelent, amely egyenlő egy hosszúidejű (pl. 8, vagy 24 órás) időben változó szintű hangesemény átlagos zajadagjával. Általában tévedések elkerülése végett zárójelben megjegyzik, hány órás zajadagról van szó (pl. Leq (8), Leq(24)).

2.4. Összetett rezgés

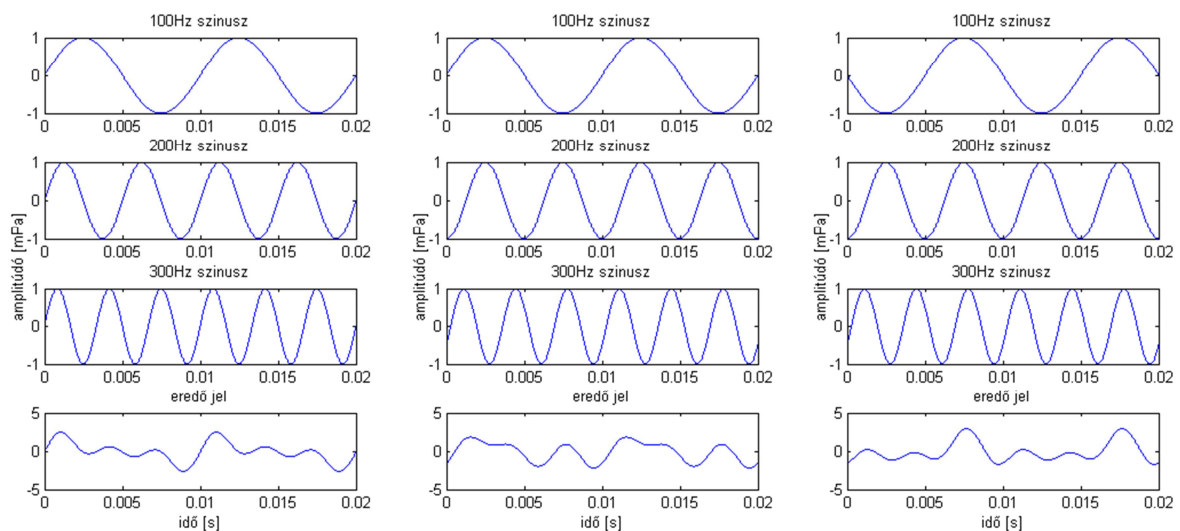
A szinuszos formájú harmonikus rezgés által keltett tisztahang ritka jelenség a hangforrások világában. A hangvilla rezgése szolgáltat tiszta hangot, vagy füttyüléssel lehet közel tiszta hangot létrehozni. A természetben előforduló rezgések összetett rezgések. Több, egymástól különböző rezgőmozgást egy anyagi részecske nem végezhet egy időben. Érvényesül a *lineáris szuperpozíció* elve, vagyis egyazon pontra ható rezgések egyszerűen összeadódnak. Például zárt térben egy hangforrás keltette hanghullám rezgéséhez a falról visszaverődő hanghullámok rezgései hozzáadódnak.

A tiszta szinuszos rezgések párhuzamos összetételéből egyszerű és összetett, azaz nem tiszta szinuszos periódusos rezgések vagy ezeknek különleges esetei származhatnak. A sokféle kombinációból mi példaként olyan rezgések összetételét vizsgáljuk meg részletesebben, ahol az összetevő rezgések amplitúdója egyenlő, és ahol az összetevő rezgések frekvenciájának aránya egész szám a legalacsonyabb frekvenciájú rezgéshez viszonyítva. Például 100 Hz 200Hz és 300Hz frekvenciájú rezgések esetén a rezgések frekvenciájának hányadosa 200/100, 300/100, vagyis 2 és 3. Az összetétel ismét periodikus rezgést eredményez, melynek frekvenciája megegyezik az összetételben szereplő legkisebb frekvenciával, alakja azonban nem szinuszos, hanem erősen függ az összetevődő rezgések kezdőfázisától. A 8. ábrán

háromféle fázisbeállításban (egymáshoz képest három különböző időeltolásban) a. b. és c. esetben mutatjuk be ugyanazokat a részrezgéseket az oszlopok első három sorában. A negyedikben, vagyis az utolsóban pedig a keletkezett összetett rezgést (Tarnóczy T., 1982). Látható, hogy az egyes részhangok fázisbeállításától mennyire függ a keletkező rezgés alakja. Az eredő rezgésalak helyességét ellenőrizhetjük, ha adott időpillanatokban az összetevő rezgések amplitúdóit egyszerűen, grafikusán összeadjuk.

Ami nagyon lényeges az az, hogy az ábra utolsó sorában mutatott három különböző hullámformájú eredő összetett periódikus rezgés ugyan azt a hangbenyomást kelti a fülben. Vagyis a fáziskülönbség nem befolyásolja az összetett hang érzeti hangzását (annak ellenére, hogy más-más a hullámforma alakja), hiszen a frekvenciakomponensek nem változnak. Fülünknek ez a tulajdonsága teszi lehetővé, hogy különböző akusztikájú terekben produkált ugyanazon beszédet, ugyanannak értjük, annak ellenére, hogy a beszédhangok hullámformáját a különböző akusztikai terek megváltoztatják.

Ha a részrezgések rezgésszámának hányadosa nem egész szám, hanem tört, az összetett rezgés periódusa meghosszabbodik, míg ha a hányados alakja igen bonyolult, a periodicitás teljesen elveszhet. Ha azonban a rezgésszámok hányadosa igen nagy, az alaprezgés periódusa ismét dominál, és a nagy szaporaságú összetevő alig észrevehetően módosítja a periódusát.



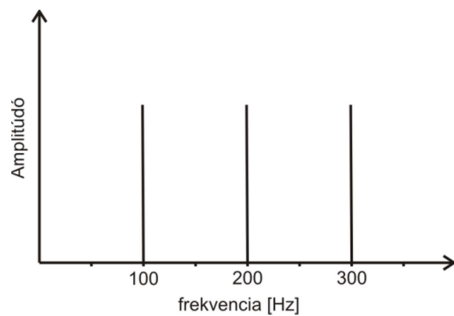
8 ábra Részrezgések összegzése. Az első három sorban háromféle fázisbeállításban (egymáshoz képest három különböző időeltolásban) a. b. és c. esetben láthatók a részrezgések. A negyedik sorban, a keletkezett összetett rezgés látható. Ugyanazon komponensekből, de más fázisértékekkel összeadott eredő jel rezgésformája különböző

2.5. Összetett rezgések frekvenciaelemzése

Mint ahogy azt már említettük a természetben előforduló rezgések összetett rezgések. Összetett rezgéseket érzékelünk a fülünkkel, és összetett rezgéseket veszünk fel mikrofonokkal hangfelvételkor.

Igen gyakran szeretnénk tudni, hogy ezek az összetett rezgések milyen harmonikus komponensekből állnak. Azt a folyamatot, amikor egy összetett rezgést (akár periodikus, akár nem) frekvenciakomponenseire bontjuk, frekvenciaelemzésnek nevezzük.

Fourier, francia matematikus a 19. század elején kimutatta, hogy lineáris rendszerekben bármely összetett rezgés, amely egyértelműen leírható időfüggvényével, felbontható különböző frekvenciájú, amplitúdójú és fázisú harmonikus komponenseire. Frekvencia komponensekre bontáskor az adott összetett rezgés hangnyomás időfüggvényét, frekvenciafüggvénnyé alakítjuk át. Ezt a frekvenciafüggvényt (frekvencia, amplitúdó és fázis adatok összességét) nevezzük *spektrumnak* vagy *színképnek*. A gyakorlatban hang spektruma lehet nyomás amplitúdó-, teljesítmény-, vagy energiaspektrum, attól függően, hogy az adott időpontban a frekvencia összetevők nyomás amplitúdó, teljesítmény, vagy energia eloszlását adja meg.



9.ábra. A 8. ábra utolsó sorában lévő összetett periodikus rezgések amplitúdó spektruma

A *periodikus rezgéseknél* az összetett rezgésnek van egy alap ismétlési periódusa, amely az összetett hangot felépítő összes összetevő közül a legmélyebb frekvencia összetevő, és amely meghatározza a komplex hang frekvencia komponenseit. Ezt a legmélyebb hangot *alaphangnak* (f_0) nevezzük. A komplex hang többi összetevőjét *felhangoknak* (f_1 f_2 ... f_n) nevezzük. A felhangok a legalacsonyabb frekvenciájú *alaphang* (f_0) egész számú többszörösei. Az alaphang és a felhangok együttesét pedig *részhangoknak* nevezzük.

A 8. ábra utolsó sorában lévő összetett periodikus rezgések például, amelyek 100Hz alapfrekvenciájúak, de különböző a hullámformájuk, felbonthatók 100, 200 és 300 Hz-es tiszta szinuszos, egyenlő amplitúdójú frekvencia-összetevőkre. A szokásos amplitúdó spektrum ábrázolásukat a 9. ábra mutatja, ahol az alaphang $f_0 = 100$ Hz, a felhangok $f_1 = 200$ Hz és $f_2 = 300$ Hz.

Egy hangszeren például, amikor lejátszunk egy dallamot, az alaphangot (f_0) változtatjuk. A felhangokat a gerjesztés módja (pengetés, ütés), a felhangok amplitúdóját a hangszer rezonanciatulajdonságai szabják meg.

A nem periodikus rezgések esetén, mint például a fehérzaj, vagy impulzus, az összetevő komponenseinek frekvencia aránya nem egész szám: folyamatos spektrumot adnak. Az energia szétszóródik egy frekvenciatartományban és nem meghatározott frekvenciáknál koncentrálódik.

Férfiak beszédében az alaphang az f_0 jellemzően 100 Hz körüli, nőknél ennek közel kétszerese. A beszédben az alaphang adja meg a beszélő személy úgynevezett hangfekvését (mély hangú, magas hangú beszélő). A hangfekvés jellemző a beszélő személyre. A beszédben a felhangok a legalacsonyabb frekvenciájú alaphang (f_0) egész számú többszörösei (például a magánhangzóknak). A periodikus rezgések tehát úgynevezett vonalas spektrummal rendelkeznek (csak a felhangok frekvenciáin vannak spektrum összetevők). Fontos megjegyezni, hogy a felharmonikusok egymás közötti távolsága lineáris a frekvenciatengelyen, és hogy két felharmonikus között a távolság megegyezik az

alaphfrekvencia értékével (lásd a 9. ábrát). A hallórendszerünk képes arra, hogy egy összetett periodikus hangból a különbségi hangokat kihallja. Ezt a hallási tulajdonságot használják ki az orgonatervezők, amikor nem készítik el a mély alaphangnak megfelelő nagy méretű sípokot.

Ebből a hallási tulajdonságból adódik az is, hogy egy beszélő személy alaphangját akkor is meg tudjuk ítélni, ha a tényleges f_0 komponens nincsen benne a beszédjelben (például a telefonon hallott hangban a férfi alaphang nincs benne az átviteli rendszer szűrése miatt). A vonalas spektrum elméleti képe a gyakorlatban azonban kissé más. A vonalak környezetében is vannak frekvenciakomponensek, mert a vizsgált jel általában nem stacionárius, ezért a jelet csak korlátozott időtartományon belül (ablakolással) tudjuk vizsgálni, és az ablakolás torzítja jelet. Továbbá a jel még az ablakon belül is változhat, ezzel további összetevőket hoz be a szinképbe. Például a beszéd esetében a hangjakak nyitódási-záródási bizonytalanságából adódó alaphfrekvencia ingadozás (kváziperiodikus jelleg).

A nem periodikus rezgés esetén, mint például olyan zörej, amelynek frekvencia összetevői minden frekvencián egyenlő intenzitással megtalálhatók (fehérzaj) vagy az impulzus jellegű, gyors lefolyású hangok (zárfelpattanás), az összetevő frekvenciakomponensek között nincs olyan szabályosság, mint ami a periodikus hangoknál volt. A nem periodikus rezgések végtelen sok frekvenciájú szinuszos összetevőből állnak, és ezek a frekvenciatartomány bármely pontján lehetnek (vagyis az összetevők nem meghatározott frekvenciáknál koncentrálnak). A nem periodikus jelek ezért folytonos spektrummal rendelkeznek.

2.6. A beszédjel frekvenciaelemzése

A beszédjel szintén összetett rezgés, amely időben folyamatosan változó, különböző rezgésmódok kombinációja.

A beszédjel elemzése nem egy egyszerű feladat, különösen két szempontból:

1. A beszéd egy biológiai produktum, ahol a beszédjel időfüggvényének egyes megvalósulásai, pl. még ugyanazon személy kitartott 'á' hangja időfüggvénye is esetről esetre más és más (nem determinisztikus). Ilyen típusú, de időben állandó (stacionárius) jeleknél viszont a hosszabb időre vett átlaguk, azaz a hosszú idejű átlagos spektrumuk (LTAS: Long Time Average Spectrum) hasonló, így egyetlen realizáció időátlagából vonjuk le következtetéseket. Ez a következtetés azután más realizációk időbeli átlagára is jó közelítéssel érvényes lesz.

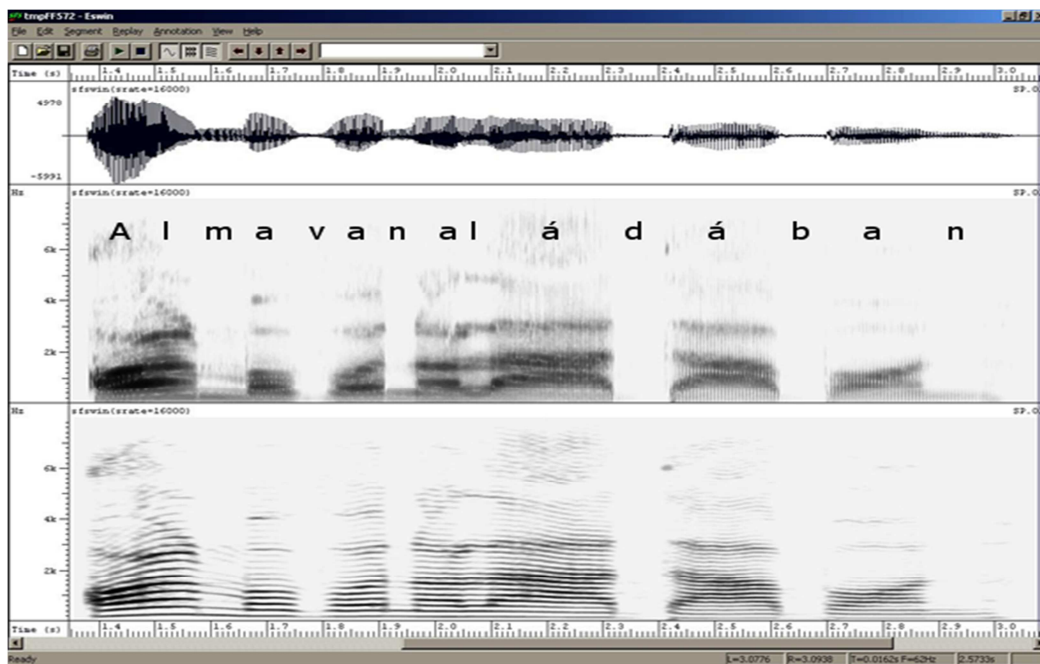
Így a leggyakrabban teljesítményszint vagy intenzitásszint sűrűség spektrumot (Fourier-transzformáltjának négyzete) szokás számolni, vagyis egy meghatározott sáv szélességre eső teljesítmény vagy intenzitásszintet [dB/Hz] (Douglas O'Shaughnessy 1987). Valójában meghatározott sáv szélességben szűrjük a jelet, és a meghatározott sáv szélességbe eső teljesítményt vagy intenzitást számoljuk. Gyakran a jellemzőt sok mérés utáni átlagszámításból adjuk meg. A *teljesítményszint spektrum ill. intenzitásszint spektrum*: a jel meghatározott időintervallumában a frekvencia-összetevők teljesítményszint ill. intenzitásszint eloszlását adja meg.

2. A beszéd-előállítási folyamat nem a fenn leírt időben állandó folyamat. A hangképzés folytonos és időben változó jelet állít elő, amelyben tranziens, közelállandó, és impulzusszerű jelek váltakozva követik egymást. Az ilyen nem stacionárius jelek matematikai kezelése nagyon bonyolult. Többségük azonban, korlátozott időtartományban közel stacionáriusnak vehető, és így egy adott időpontban, hozzá tartozó megfelelő időablakban, a beszéd részlet közel állandónak vehető, és az elemzés elvégezhető. Az ilyen jeleket kvázistacionernek hívjuk. A beszéd is ilyen kvázistacioner részek sorozatának tekinthető, és spektrum elemzés a kvázistacioner részek elvégezhető úgy, hogy az egymás után kijelölt pontokban (pl. 10-20ms-ként) mindig egy meghatározott időtartományban (pl. 25ms-ban) meghatározzuk a

teljesítményspektrumot. Tehát amikor a változást követni szeretnénk egy teljes beszédszakaszon, akkor az elemzést mindig egy meghatározott időintervallumban, idő ablakban(Δt) kell elvégeznünk, és ezt az időablakot végig kell görgetni a vizsgált beszédszakaszon. Így kapjuk meg a *gördülő teljesítményspektrumot*, vagyis a *teljesítményspektrumot* (amit a fonetikában szonogramnak neveznek). Tehát a spektrogram a teljesítményspektrum időbeli változása, vagyis mutatja, a frekvencia-összetevők, teljesítményszint eloszlásának időbeli változását.

A 10. ábrán az „Alma van a ládában.” mondat időfüggvénye alatt, a mondat spektrum változásának időbeli folyamatát, vagyis a spektrogramot mutatunk be, fenn rövid, és lenn hosszú elemzési időablakkal. Az ábrán a vízszintes tengelyen az időt mutatjuk szekundumban, a függőleges tengelyen a frekvenciát tüntettük fel 8 kHz-ig beállított elemzési frekvenciatartományban. Az adott időponthoz tartozó intenzitásszint nagysága arányos a feketedés mértékével.

Frekvenciaelemzésnél a meghatározott időintervallum, vagyis az elemzési ablak szélessége (Δt) meghatározza az elemzés frekvenciafelbontását (Δf), vagyis azt, hogy milyen részletességgel kapjuk meg a spektrum összetevőket. A $\Delta t \cdot \Delta f = \text{állandó érték}$. Finom frekvenciafelbontáshoz hosszú időablakra van szükség, a pontosabb időbeli követéshez viszont az ablak szélességet rövidre kell választanunk. A beszédelemzési technikában a spektrumelemzéshez szokásos időablak 5ms és 50ms közötti (Olaszy G.,1989).



10. ábra „Alma van a ládában” mondat amplitúdó időfüggvénye (fenn), spektrogramja rövid, 5ms elemzési ablakkal (középen), és spektrogramja hosszú 30ms elemzési időablakkal (lenn). Az intenzitásszint nagysága arányos a feketedés mértékével.

A rövid időablakú elemzéssel kapott felső spektrogramon az artikuláció folyamán bekövetkező változásokat, zárfelpattanásokat hűen tudjuk követni, de a frekvenciafelbontás elég rossz. A felhangtartalom összemosódik. A hosszabb időablakú elemzésnél a változások nem jól követhetőek, de a frekvenciafelbontás sokkal jobb, mint a felső spektrogramnál. Itt a vízszintes csíkok az egyes felhangok erősségének időbeli változását szeparáltan mutatják.

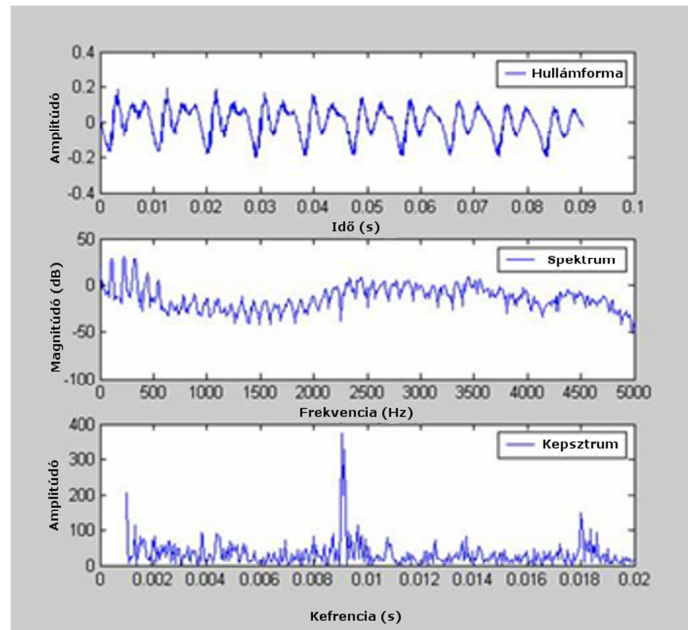
Régebben, a frekvenciaelemzéseknél 10. ábrán mutatott spektrogramokhoz hasonlóan, minden frekvenciatartományban azonos volt az elemzési sáv szélessége (frekvencia felbontása), azonban a fülünk a különböző frekvenciatartományokban más-más sáv szélességgel dolgozik, 500 Hz alatt állandó, kb. 100 Hz-es sáv szélességgel, 500 Hz. fölött az elemzési sáv szélesség a frekvenciával növekszik, 5000 Hz környékén a sáv szélesség már több mint 1000 Hz. Az emberi hallásfolyamatban működő változó sáv szélességű elemzési sávokat *kritikus sávoknak* nevezzük, és sorrendjüket [*Bark*]-ban adjuk meg (Zwicker, E. 1982). A kritikus sávoknak megfelelő frekvenciaelemzéssel kapott gördülő színeképet cochleogrammoknak is szokták nevezni.

A szűrősáv, vagy FFT vel végzett frekvenciaelemzésen túl még számos elemzési módszer létezik. Ilyen például a lineáris predikció LP (Linear Prediction) alapú, vagy a kepsztrum (Cepstrum) elemzés.

A *lineáris predikció (LP)* egy olyan matematikai módszer, amellyel egy jelenség adott időpontbeli állapotát meg lehet becsülni a jelenségre vonatkozó a priori adatok és az adott időpontot megelőző állapotok birtokában. Ezt a matematikai eljárást számos tudomány területen használják. A digitális beszédfeldolgozásban azért terjedt el, mivel alkalmas a beszédhullámforma leírására, az elhangzott időben változó beszédmintákból származtatott paraméterek egy készlete alapján. A lineáris predikció, túl az előrejelzésen, egy jelenség tömör, lényegkiemelő leírására is szolgálhat, és sokszor ezt a tulajdonságát aknázzák ki.

A *kepsztrum (Cepstrum)* a beszéd teljesítményspektrumát logaritmizálja, majd ezt a függvényt mint időfüggvényt kezeli és Fourier-transzformálja. Ezt nevezzük kepsztrum transformációnak. *Kepsztrum transzformációval* a gerjesztő jel, és a hangképző csatorna hatása jól szétválasztható. Ezen tulajdonsága miatt alkalmazzák szívesen a beszédtechnológiában. A kepsztrumelemzés jól használható az alaphang meghatározására.

A teljesítményspektrumot ekkor úgy tekinthetjük, mint sok kisebb szinusz hullámkomponens összege. Minden szinusz hullámnak van egy amplitúdója és egy „frekvenciája”. Annak érdekében, hogy elkerüljük a „frekvencia” szó zavaró elnevezését, a teljesítményspektrum hullámkomponenseit ezentúl kefrenciának hívjuk. A kefrencia a teljesítményspektrum frekvenciáinak frekvenciája (a mértékegysége a ciklus/frekvencia, azaz a másodperc). A teljesítményspektrum minden egyes hullámkomponensének amplitúdóját a kefrencia függvényében ábrázolva kapjuk a kepsztrumot. (Lásd az X1 ábrát). A kapott Fourier-transzformáció tehát a teljesítményspektrumban található információt (frekvencia tartomány) idő (kefrencia) tartományba transzformálta. A legkiemelkedőbb csúcs a kepsztrumban az alapperiódus. Az alapperiódus a komplex hullámnak tekintett spektrum domináns szinusz hullámának kefrenciája, ugyanúgy, mint ahogy az alapprofrendencia a komplex hullámnak tekintett beszédjel domináns szinusz hullámának frekvenciája. A kepsztrum kisebb amplitúdójú csúcsai a harmonikusok. Egy magas periodicitású jelnek a spektrumban lesz egy nagy kiemelkedésű csúcs az alapprofrendenciánál és több kiemelkedése ennek többszöröseinél. Ezek a csúcsok szabályos intervallumonként helyezkednek el. Ez az intervallum felel meg a kepsztrum alapperiódusának. Így egy nagy kiemelkedésű csúcs lesz látható az alapperiódusnál, amint azt az X1 ábra is mutatja. Egy alig periodikus jel spektrumában sok kiemelkedő csúcsú lesz sok frekvenciánál, amelyek nem követik egymást meghatározott szabályossággal. Így egy gyengén periodikus jel alacsony amplitúdójú sok kepsztrális csúcsot eredményez, különböző kefrenciákon. A kepsztrumba berajzolt lineáris regressziós görbével, amely reprezentálja az átlagos hangenergiát, kiküszöbölhetjük a hangkeltés amplitúdójának eltérését különböző szituációk és személyek között. A kepsztrális csúcs (a legnagyobb amplitúdóval rendelkező csúcs a kepsztrumban) és a lineáris regressziós görbe közti különbséget kepsztrális csúcs kiemelkedésnek (cepstral peak prominence, CPP) nevezzük. Gyengén periodikus jelnek ennek értéke kicsi, erősen periodikusnál pedig nagy.



X1. ábra: Kepsztrum egy periodikus hang esetén.

A különböző frekvenciaelemzési módszerekről részletes leírás a Magyar Beszéd című könyvben (Németh.....2010) található.

Napjainkban sok, a hangelemzés egyszerű elvégzésére alkalmas, szabadon használható program van, amelyből néhányat felsorolunk:

Cooledit - általános akusztikai elemző

<http://www.softpedia.com/get/Multimedia/Audio/Audio-Editors-Recorders/Cool-Edit-Pro.shtml>

Wavesurfer - beszédelemző, szegmentáló

<http://www.speech.kth.se/wavesurfer/>

Wasp- beszédelemző, szegmentáló

<http://www.phon.ucl.ac.uk/resource/sfs/wasp.htm>

Praat - fonetikai elemző, szerkesztő program

<http://www.fon.hum.uva.nl/praat/>

A fenn felsorolt programok digitális jelfeldolgozással működnek. Ezért fontos tudnunk a következőket:

A teljes hallástartományban történő hangfelvételeknél, tehát amikor a jelet 20kHz-ig át akarjuk vinni a felvételi rendszeren, a mintavételi frekvenciát minimum 40 kHz-re, az amplitúdó felbontást minimum 16 bitre kell beállítani. Beszédjel esetében max. 8 kHz frekvencia átvitele elegendő, mivel e felett a frekvencia felett már a beszédben nincsenek jelentős frekvenciakomponensek. Ekkor 16 kHz-es mintavételi frekvenciát állítunk be, így a frekvenciaelemzést 8kHz-ig végezzük, és az 512 pontos FFT-elemzés elegendő. Az elemzési időablakot célszerű Hamming-ablakra választani (Gordos-Takács, 1983; F.J. Owens, 1993.)

3. A beszéd folyamat akusztikai-fonetikai jellemzői

A beszéd nyelvi jelentéssel bíró akusztikai produktum, számos nem nyelvi jelentést hordozó információval. A beszédhangrezgéseket bonyolult, összetett és koordinált fiziológiai működéssel hozzuk létre. De nemcsak létrehozását tekintve összetett folyamat a beszéd, hanem megjelenési formája, a beszédhangrezgés is mutatja az összetettséget, tehát

akusztikailag is összetett jelsorozat, amely időben, hangerőben és frekvenciában dinamikusan változik.

Ha a beszéd nem lenne egyéb, mint az írásjelek hangos megjelenési formája, akkor nem éreznék kifejezőbbnek, elevenebbnek a nyomtatott vagy írott írásjelnél. A beszéd által hordozott mondanivaló minőségileg különbözik az írott szóalakokkal kifejezett fogalmi jellegű tartalmaktól. A beszédben kifejezésre jutó, nem nyelvi információ, mint például a beszélő neme, fizikai állapota vagy a beszélő által kifejezett érzelmek (harag, izgalom, öröm, panasz, csodálkozás, gyengédség, ijedtség stb.) létrehozását a komplex természetű akusztikai jel teszi lehetővé. Erre a komplex akusztikai jelre jellemző, hogy a beszéd nem nyelvi és nyelvi jellemzői együttesen vannak benne jelen, és szétválasztásuk igen nehéz.

A hangképző szervek működésével létrehozott beszédhangrezgés objektíven műszerekkel vizsgálható. Minthogy a beszélők beszélőszervei kisebb-nagyobb mértékben különböznek, és az artikuláció ugyanazon a nyelven belül sem egyforma, a létrehozott beszéd akusztikailag különböző. A beszéd akusztikai szerkezete a beszélőtől és a beszédhelyzettől (átviteli körülmények) függően változik, de még egy beszélő esetében is, ugyanazon beszédhangot, beszédhangsorozatot képezve az előállított beszéd akusztikai szerkezete, ha kisebb mértékben is, mint különböző személyek esetén, de eltér egymástól. Az emberi beszédfeldolgozás folyamatai azonban biztosítják, hogy az akusztikai különbségek ellenére a fonológiai döntések állandóak maradjanak (nyelvspecifikus beszédészlelés) (Gósy M., 2004).

Az akusztikai alapfogalmak című 2. fejezet alapján elmondható, hogy lényegében 3 alapparaméterrel jellemezhető a beszédhangrezgés. Ez a három alapparaméter a *hangnyomás*

$p[N/m^2]$, a *frekvencia* $f = \frac{1}{T}$ [Hz] és az *időtartam* [s]. A beszédparaméterekre általában

jellemző, hogy mint a legtöbb biológiai produktum paraméterei, nem determinisztikusak. A jellemző paraméterek legtöbbször hosszú idei átlagból, vagy sok mérés utáni átlagszámításból adhatók meg.

a. A *beszédhang erősségét jellemző fizikai alapparaméter a beszédhangnyomás*, származtatott jellemző a beszédenergia, beszédintenzitás, beszédteljesítmény. (A beszédhang energiája, intenzitása és teljesítménye a beszédhangnyomás négyzetével arányos.) A beszédhangnyomás, -energia, -intenzitás, -teljesítmény mindegyikének kezelésére [dB] szintértéket használunk. Ez a szintérték a beszéd folyamatban állandóan hullámzik, meg kell különböztetnünk az átlagos szintet a szintcsúcsoktól. A beszédhangnyomás- és beszédintenzitás-szintet a folyamatos beszéd valamely időközre vonatkoztatott effektív értékével jellemezzük, amint azt a 2.2. a „Hang terjedése levegőben” c. alfejezetben tárgyaltuk. Az időköz mértéke szerint megkülönböztetünk pillanatnyi és hosszabb időközre vonatkoztatott átlagos értéket.

A beszédproduktumban az energia nagyobb része a magánhangzókhoz (lásd később) kapcsolódik.

Szavakról felvett hangnyomásgörbéken a magánhangzók mindig nagyobb teljesítményértéket mutatnak.

b. A *hangmagasságra jellemző fizikai mennyiség a beszéd alapfrekvenciája* f_0 [Hz], amely a folyamatos beszéd alaphangját jelenti, férfiaknál 100-200 Hz, nőknél 150-300 Hz, gyermekeknél 250-600 Hz.

A beszéd hangmagasságának változásai a mondatok dallamformáit, intonációját alakítják ki.

c. A *hang színezetére jellemző, származtatott mennyiség a beszéd intenzitásszint sűrűség spektrum* [dB/Hz], amelyet a 2.5 fejezetben részleteztünk. A beszéd akusztikumának fontos jellemzője, amely a frekvencia-összetevők intenzitásszint értékeit adja meg. A beszédhangok megkülönböztetésében van lényeges szerepe. Azonban a megváltozott

színkép tükrözője lehet a beszélő állapotának, jellemzi a beszélőt is; például meg tudjuk a spektrumból ítélni, hogy férfi, nő, vagy gyermek beszél-e.

d. A beszédhangok, a beszédhangátmenetek időtartama[s], a beszédhangok időtartam arányai mind, fontos jellemzői a beszédnek. A beszédsebességre, beszédritmusra jellemző az időegység alatt elmondott beszédhangok száma. A tempó és szünetek összefüggésben vannak a beszélő egyén karakterével vagy érzelmi állapotával, de a tempóváltozásnak van logikai funkciója is, például kiemeli a nyomatékos mondanivalót.

A beszédben az akusztikai összetevők a legváltozatosabb módon kombinálódnak, s lehetővé teszik, hogy a beszéd mint komplex akusztikai jelzés az árnyalt és differenciált közlés eszköze legyen. Például, ha a beszélő megemeli a hangját, hangosabban beszél, általában a frekvenciaösszetétel is megváltozik. Nagyobb hangerő esetén a nagyobb frekvenciájú felhangok intenzitása nagyobb mértékben erősödik fel, mint a kis frekvenciájúaké.

A beszédre alapvetően jellemző továbbá, hogy az információt több szinten hordozza. Ezek a szintek a következők: az akusztikai-fonetikai szint, fonológiai, szintaktikai, szemantikai, pragmatikai szint (Ainsworth, W.A., 1976).

A beszéd leírásánál tárgyalhatjuk a beszédhangok vagy azoknál kisebb egységek akusztikai, fonetikai leírását, de vizsgálhatjuk több beszédhang átfogó együttes viselkedését is. A *beszéd szegmentális szerkezetének* vizsgálatakor a beszédhang vagy a beszédhangnál kisebb egységek akusztikai leírására kerül sor, míg a *szupraszegmentális szerkezet* tárgyalásakor a beszéd több fonémán átnyúló akusztikai jellemzőinek leírása történik. A továbbiakban a szegmentális és a szupraszegmentális szerkezet különálló tárgyalására kerül sor.

3.1. A beszéd szegmentális leírása

Hangképzés során a különböző akusztikai gerjesztésekkel, a beszédképzőszervek hangolásával olyan akusztikailag különböző hangrezgéssorozatokat állítunk elő, amely a nyelvi tartalom segítségével beszédhangok sorozatává áll össze agyunkban. Különböző gerjesztési típusokat használunk a beszédhangok képzésénél, melyek eredményeként különböző beszédhangtípushoz tartozó beszédhangokat hozunk létre. Például a hangajkak (régőbbi szakirodalmakban hangszalagok) rezgésével létrehozuk a zöngét mint akusztikai produktumot, és ha csak tiszta gerjesztés történik, akkor magánhangzók jönnek létre, vagy diftongusok, vagyis a kettős magánhangzók, pl. angolban az „ai”, de a magyar nyelvben diftongusok nincsenek. Ha a hangajkak rezgése mellett más, zörej típusú gerjesztés is létrejön, akkor a zöngés mássalhangzók különböző típusaihoz tartozó beszédhangokat képezzük. A zöngétlen mássalhangzók esetében zörej típusú gerjesztés jön létre a hangajkak rezgése nélkül.

A beszédhangok gerjesztésekor a gerjesztő jel, másnéven kényszerítő jel (a zöngé, a súrlódási zörej, a zárfelpattanási zörej) meghatározott színeképp. Ezt a gerjesztő jelet befolyásolja a hangképző csatorna üregeinek rezonanciája. Ahogy artikuláció közben változtatjuk e csatorna méretét vagy a csatornában a képzés helyét, a hangképző csatorna változó rezonanciái állandóan módosítják az eredeti gerjesztő jel színeképet (ld. részletezve a Magánhangzók c. alfejezetet.).

A hangképző csatorna valójában egy levegővel telt cső, és mint a levegővel töltött csövek, rezonátorként működik. Ez azt jelenti, hogy a hangképző csatornának van bizonyos természetes rezgési frekvenciája, és sokkal könnyebben reagál egy olyan hanghullámra, amelynek frekvenciája hasonló ennek rezonáns frekvenciájához, mint egy más frekvenciájú hanghullámra.

Tehát egy beszédhang akusztikai tulajdonságait a gerjesztés módja, típusa és a képzés helye (a hangképző csatorna állapota) együttesen határozzák meg.

A beszédhangok időtartama 50 ms – 150 ms között változik, normál beszédtempó mellett (12 beszédhang/mp). Ritmus változásakor lényegében főleg a beszédhangok kvázistacioner szakaszának ideje változik, nő vagy rövidül meg. Az átmenet, vagyis a hangok kialakulási szakaszának időtartama, valamint a tranziens komponensek, mint például a zárfelpattanás közelítőleg változatlanok.

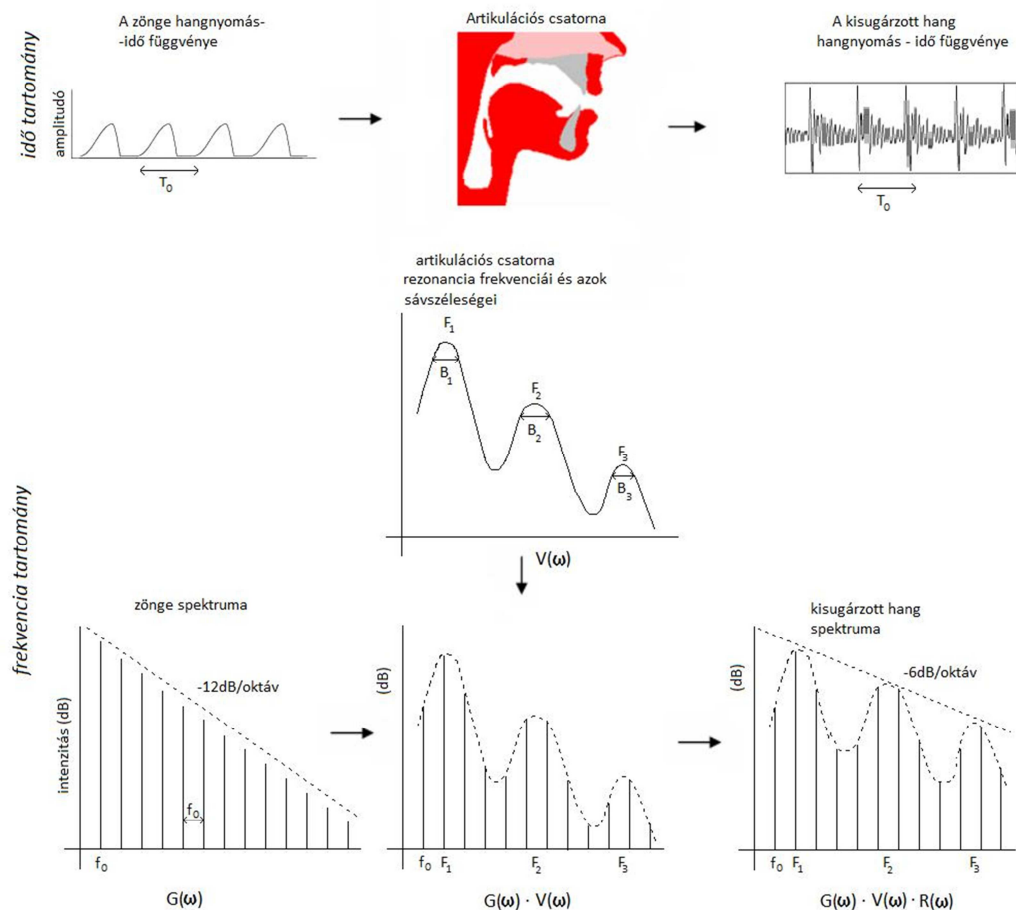
3.1.1. A beszédképzés gerjesztett szűrőmodellje

Amint azt az előző fejezetben kifejtettük, a létrehozott beszéd akusztikai tulajdonságait lényegében három tényező határozza meg: a hang előállításának a módja vagyis a hangforrás, a hozzá kapcsolódó hangképző üregrendszer rezonanciátulajdonságai és a hangtérbe való sugárzás milyensége (11. ábra). Így a beszéd alapvetően e három komponens által képzett függvény szorzataként áll össze a

következő egyenlet szerint (Fant 1960, Kent–Read 1992):

$$S(w) = G(w)V(w)R(w). ()$$

ahol $S(w)$ jelenti a beszéd eredő színeképet (Speech), $G(w)$ a hangszalag rezgésével létrehozott hangforrás színeképet (Glottal), $V(w)$ a hangképző üregekből álló hangképző csatorna átviteli függvényét (Vocal tract), $R(w)$ pedig az ajak kisugárzásánál a sugárzási impedancia átviteli függvényét (lip Radiation). A 11. ábrán látható a három összetevő hatása a beszédjelre.



11. ábra A beszéd létrehozása fiziológiai és fizikai szinten a zöngés beszédhangokra a gerjesztett szűrő modellje szerint. A beszédjel alapvetően három komponens által képzett átviteli függvény szorzataként áll össze. A formánsok: F_1, F_2, F_3, F_4 , a gerjesztett szűrő rezonanciafrekvenciáinál alakulnak ki.

Magánhangzók és zöngés mássalhangzók képzésénél, a hangjakak működésből keletkező zöngé hangnyomás-időfüggvénye közel fűrészfog jellegű (11. ábra balra fenn), T_0 alapperiódussal. Ennek szinképi összetevőit képviseli $G(\omega)$ függvény, amely leírja az alaphang (f_0) és a felhangok együttesét. A felhangok intenzitása az ilyen fűrészfog típusú időfüggvények esetén átlagosan 12 dB/oktáv meredekséggel csökken a frekvencia növekedésével. Ezt a szinképet befolyásolja a változó méretű artikulációs csatorna, amely egy üregrendszer több rezonanciafrekvenciával (Kent–Read 1992), ez a $V(\omega)$ függvény. A rezonanciafrekvenciákat a szakirodalom formánsoknak nevezi, amelyeket sorrendben $F_1, F_2, F_3, \dots, F_n$ betűkkel, a rezonancia szélességet (a rezonanciagörbe maximuma – 3 dB intenzitásértéknél mért frekvenciaszélességet) pedig $B_1, B_2, B_3, \dots, B_n$ betűkkel jelöljük. A rezonanciafrekvenciákon és azok környezetében a felhangok intenzitása megnő, más helyeken elnyomódik. Így alakulnak ki a zöngés hangok formánsstruktúrái. A végső, a hangtérbe jutó hang szinképet még a szájnyílás kisugárzásánál a sugárzási impedancia (ld. 2.3. alfejezet) $R(\omega)$ átviteli függvénye befolyásolja. Ez azt jelenti, hogy az akusztikus energia kicsatolása frekvenciafüggő. Magasabb frekvenciákon a sugárzási impedancia kisebb, mint az alacsony frekvenciákon. Ez a szinkép felső frekvenciatartományának a megemelését eredményezi átlagosan 6 dB/oktávval.

A zöngétlen hangok esetében a $G(w)$ gerjesztő függvény lehet súrlódási zörej, mint például a réshangok képzésénél, ahol a súrlódási zörejt a résen kiáramló levegő örvényleszakadása okozza, vagy lehet a hangképző csatornában keletkezett zár felpattanásakor keletkező zárfelpattanási zörej. Mindkét esetben a zörej jellegű $G(w)$ szinképet a gerjesztett üreg $V(w)$ rezonanciája befolyásolja, szinképi súlypontokat alakítva ki az eredő szinképben. Ezeknél a hangoknál az üregeknek nincs olyan éles rezonanciája, mint a magánhangzók esetében, de a szinképi súlypontok a zár vagy a rés képzési helyének függvényében változnak.

A beszédhangok képzésekor beszélhetünk tiszta (egyfajta gerjesztésű) hangokról, ahol a $G(w)$ gerjesztés vagy zöngé (például magánhangzók), vagy súrlódási zörej (réshangok), vagy pedig lökéshullámszerű zárfelpattanási zörej (zárhangok). Gyakoriak a vegyes gerjesztésű mássalhangzók is, ahol több egyidejű gerjesztés fordul elő egyetlen hang képzésén belül. A zöngés zár-, illetve réshangok képzésekor a zárfelpattanási-, illetve súrlódási zörej mellett zöngés gerjesztési hang is része lehet a hangképzésnek. A gerjesztés típusait tekintve, a legösszetettebb hangok a zöngés zár-rés hangok, amelyekben mindhárom képzési forma szerepet játszik (zöngé, súrlódási zörej és lökéshullámszerű zárfelpattanási zörej).

A magyar hangkészlet képzés szerinti típusait a 2. táblázat szemlélteti.

2. táblázat A magyar hangkészlet képzés szerinti típusai

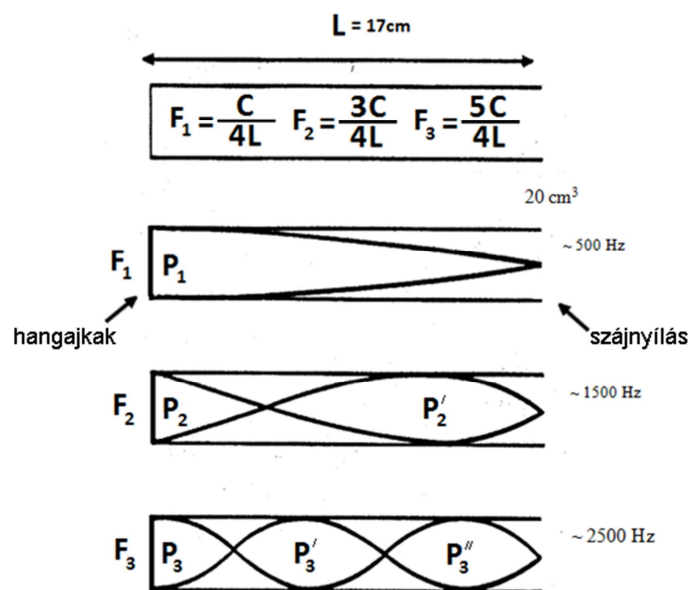
	gerjesztés típusa	akusztikai produktum	Beszédhangtípus és írásjelszimbólumai
tiszta gerjesztés	hangajkak rezegnek	zöngé	magánhangzók i, í, ü, ú, u, ú, e, é, ö, ó, o, ó, a, á
	résen kiáramló levegő turbulens áramlása	súrlódási zörej	zöngétlen réshangok f, sz, s, h
vegyes gerjesztés	zár-felpattanás	beszédenergia mentes rész + zárfelpattanási zörej	zöngétlen zárhangok p, t, k, ty
	hangajkak rezegnek+ résen kiáramló levegő turbulens áramlása	zöngé + súrlódási zörej	zöngés réshangok v, z, zs likvidák l, r, j
	hangajkak rezegnek+ zár-felpattanás	zöngé + zárfelpattanási zörej	zöngés zárhangok b, d, g, gy nazálisok m, n, ny
	hangajkak rezegnek+ zár-felpattanás + résen kiáramló levegő turbulens áramlása	zöngé + zárfelpattanási zörej + súrlódási zörej	zöngés affrikáták (zár-réshangok) dz, dzs
	hangajkak nem rezegnek + zár-felpattanás + résen kiáramló levegő turbulens áramlása	beszédenergia mentes rész + zárfelpattanási zörej. + súrlódási zörej	zöngétlen affrikáták (zárréshangok) C, cs

3.1.2. A beszédképzés gerjesztett cső modellje

Arra a kérdésre, hogy a $V(w)$ a hangképző üregekből álló hangképző csatorna átviteli függvénye hogy alakul, ma a gerjesztett csőmodell a legelfogadottabb modell. Ez az egyik legegyszerűbb modell azon a közelítő feltevésen alapszik, hogy a hangképző csatornában – mint egy keskeny csőben – csak tengelyirányban terjednek a hullámok, mivel a keresztmetszeti méretek a hullámhosszhoz képest kicsik. Ilyenkor a cső alakja egy egydimenziós keresztmetszeti függvénnyel ($A(x,t)$) leírható, ahol az x hossz a t időben folyamatosan változik, és feltételezik, hogy a hanghullám visszaverődése a csőfalról veszteségmentes, veszteség csak a száj- és orrnyíláson keresztül a térbe sugárzásból ered, a toldalékcső csatolásmentes, és a csőfalak merevek. A csőben amely a hangajkaknál zártnak, a szájnyílásnál nyitottnak tekinthető, állóhullámú légoszlop alakul ki. Az $x=0$ a hangajkaknál, és $x=L$ a szájnyílásnál. Az L értéke személyenként változik, nőknél kb. 13 cm, gyermekeknél még kisebb, férfiaknál nagyobb, átlag kb. 17 cm. Az egyik végén zárt, a másik végén pedig nyitott cső, egy orgonasípra emlékeztet, ahol a hanghullámok fel-alá haladnak a csőben, és amelyek visszaverődnek és találkoznak, erősítve vagy gyengítve a csőben lévő levegőmolekulák rezgését. Ebben az állóhullámú légoszlopban a zárt vég kimozdulási és sebességi minimumhelyet, viszont nyomásban és sűrűsödésben maximumot jelent. A nyitott vég fordított szerepet játszik. Az eredő állóhullámú rezgés rezonanciafrekvenciái normál viszonyok mellett csak a cső hosszától függenek, az alábbi képlet szerinti helyeken mutatva rezonancia frekvenciákat.

$$f_n = c/4L(2n-1); \quad n = 1, 2, \dots$$

Az első rezonancia $1/4$ hullámhossznál, és az $1/4$ hullámhossz páratlan számú többszöröseinél, alakul ki,

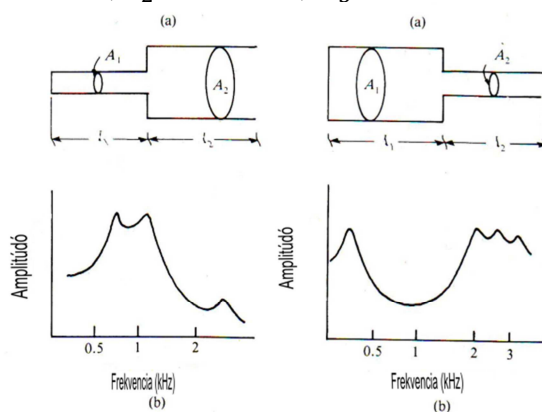


12. ábra Az egyik végén zárt (a hangajkaknál), a másik végén nyitott (a szájüregnél, orrüregnél) cső állóhullámú rezonanciahelyei: F_1, F_2, F_3 . A p_1, p_2, p_3 jelzi a maximális hangnyomású helyeket (Kent-Read, 1992)

amint azt a 12. ábra szemlélteti. Ha a cső hossza $L = 0,17\text{m}$ (átlagos férfi artikulációs csatornája), és hang terjedési sebessége $c = 340 \text{ m/s}$, 1 atmoszféra nyomáson és 20°C hőmérsékleten, akkor az első, második és harmadik állóhullámú rezonancia $F_1 = 500 \text{ Hz}$, F_2

= 1500 Hz, $F_3 = 2500$ Hz. Az egyenletes keresztmetszetű csőhöz hasonlítható a beszédképzésben a semleges magánhangzóhoz tartozó artikulációs csatorna alakja, az úgynevezett schwa magánhangzó. Ennek a beszédhangnak az első három formánsa rendre az 500, 1500, 2500 Hz körüli értékeknél látható a hang spektrogramokon. A gyakorlatban persze a hangjakak a hangképzés zárt fázisai során záródik le csak igazán. Az ajakkerekítés, vagy zárás gyakran keskenyebbé teszi az akusztikus csövet az ajkánál, ingadozást okozva az elméleti formánsfrekvenciákban. Ahogy a különböző hangokat képezzük, a hangképző csatorna keresztmetszete folytonosan változik, ami az állóhullámú rezonanciát befolyásolja és ingadozást mutat a semleges schwa magánhangzó formánsértékeihez képest. Normál esetben az eltérések néhány (vagy több) száz Hz-es tartományon belül vannak. Így, ha a hangképző csatorna 17 cm-es, F_1 általában a 300-800 Hz-es tartományba esik; F_1 ennél alacsonyabb lehet, ha például a hangképző csatorna erősen leszűkül (pl. a legtöbb mássalhangzó, és olyan magánhangzók esetében, amelyeknél a nyelv felemelkedik – magas magánhangzók). Magasabb értéket vesz fel mély magánhangzók esetében. F_2 általában a 700-2200 Hz közötti tartományba esik; F_2 nagyobb értékeket vesz fel, amikor a nyelv relatíve elől áll, és kisebb, amikor a nyelv inkább hátul van. Az F_2 tartománya általánosságban nagyobb, mint más formánsoké. F_3 általában az 1800-2800 Hz-es tartományba esik; F_3 nagyobb értékű, amikor a nyelv relatíve elől és magasán áll, és kisebbet, amikor a nyelv hátrahúzódik (mint pl. az /r/ esetében). A magasabb formánsok intenzitása általában fokozatosan gyengül, és kevésbé relevánsak a legtöbb esetben.

A formánsok alakulásának modellezése oly módon lehetséges, hogy a csatorna felosztható közel állandó keresztmetszetű szakaszokra, melyekre a hullámegyenletek pontosan leírhatók és a rezonanciaértékek kiszámolhatók (Gordos–Takács 1983). A pontos leírást persze bonyolítja, hogy a keresztmetszet változásoknál a hullám impedancia megváltozik, ezért ezeknél a helyeknél visszaverődések lépnek fel. Azonban magánhangzók esetében már két különböző A_1 , A_2 keresztmetszetű cső együtteséből álló modell átviteli függvénye is egész jó közelítést ad. Az [a:] és [i] hangokra jellemző artikulációs csatorna keresztmetszeteket modellező csőformációkat mutatunk be a 13. ábrán. Az átviteli függvények maximum helyei jól közelítik a fenti két hangra jellemző formánsokat. Az [a:] esetében a formánsok a következő értékeket közelítik meg: $F_1 = 900$ Hz, $F_2 = 1100$ Hz, $F_3 = 2900$ Hz. Az [i] esetében a formánsok a következő értékek körül ingadoznak: $F_1 = 100-280$ Hz, $F_2 = 1900$ Hz, $F_3 = 2100$ Hz.



13. ábra Az A_1 , A_2 keresztmetszetű cső együtteséből álló modell (a) és annak átviteli függvénye (b), balra az á, jobbra az i hang esetében (Stevens, 2000)

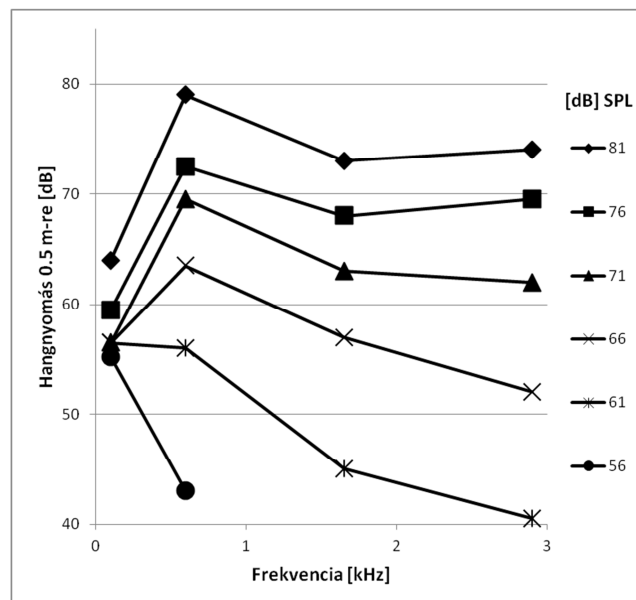
A mássalhangzók artikulációjának ilyen típusú modellezése már jóval bonyolultabb, de Stevens (1972) megmutatta, hogy a cső mentén, a csőben lévő szűkület helyének változtatásával leírható az energia maximumok kialakulása a képzési hely függvényében.

A nazális mássalhangzók háromcsöves hangképző csatorna modellel való elemzést kívánják, minthogy a nazális hangoknál az orrüreg is szerepet játszik, amikor a lágyszájpad leereszkedik az ilyen mássalhangzók kimondásakor. Az egész rendszer modellezhető három csőként (garat, orr, száj), amelyek egy ponton találkoznak (a lágyszájpadnál).

A garat+orrüreg állandó hosszúságú. E hosszak megfelelően a nazális hangok színekében 800 Hz-ként rezonanciák jelennek meg, függetlenül attól, hogy melyik nazális hangról van szó. Ugyanakkor a szájüreg hossz mérete változik a különböző nazális hangok képzésekor, ennek megfelelően a színek is változik. A nazális mássalhangzók így jellemezhetőek: (a) formánsok minden 800 Hz-nél, (b) a színekben a nazális mássalhangzók relatíve stabil formánsokkal rendelkeznek, amelyek gyengébbek, mint a magánhangzók formánsai, és kevésbé koartikuláltak. (c) szélesebbek a formáns sáv szélességek, és (d) zérók a spektrumban.

3.1.3. Hangképzés hangerejének befolyása a színekre

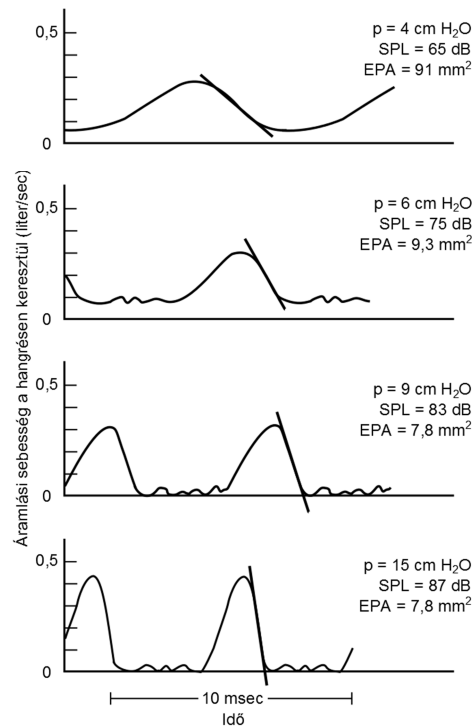
Csaknem minden hangszerre jellemző, hogy a hangerőszint emelése jobban megnöveli a magasabb harmonikusok szintjét, mint az alacsonyabbakét. Ez igaz a beszéd hangerejének növelésekor, de még inkább szembeeső éneklésnél, amint azt a 14. ábra mutatja. Hangos éneklésnél a teljes hangenergia lényegesen nagyobb része jelenik meg a magasabb harmonikusokban, mint a lágyszájpad éneklés esetében.



14. ábra. Az első négy harmonikus intenzitás szintjének (SPL) alakulása különböző hangerőszinteken. Lágyszájpad hangképzés esetében az alacsonyfrekvenciák dominálnak, de a hangerő növekedésével a magasabb harmonikusok intenzitás szintje emelkedik (Sundberg nyomán, 1987.)

Ennek a magasabb harmonikusoknál jelentkező energianyereségnek az oka látható a hangrésonánsszerűen áthaladó légáramlási sebesség hullámformáinak összehasonlításával

(„glottogramok”) a 15. ábrán. Ahogy a hangképzés hangossága növekszik, a hangrés bezáródásának sebessége (a hullámformák záró élei mentén meghúzott vastag vonalak lejtőivel jelölve) megnő. Ez az idő függvényében mutatkozó bezáródási sebesség megnövekedése (nagy emelkedési, vagy esési sebességű hullámformák) a frekvenciatartományban a magasabb harmonikusokban gazdag színeképet eredményez.



15. ábra. A hangrésen áthaladó légáramlási hullámforma négy növekvő hangerősségszint esetében. A hangrés záródásának sebessége a hangerősségszint növekedésével együtt nő. (Sundberg nyomán, 1987).

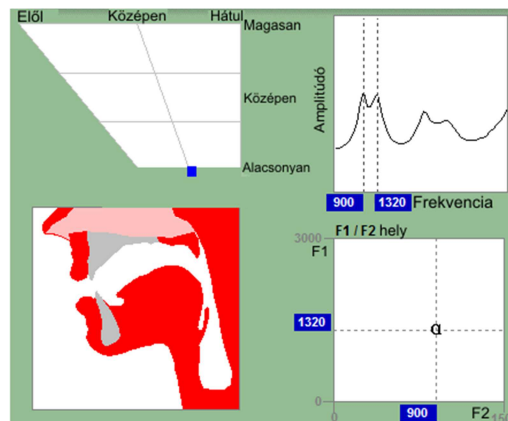
3.1.4. Magánhangzók

Mint ahogy az előző fejezetben tárgyaltuk a hangképzéskor a magánhangzók, valamint a zöngés mássalhangzók zöngéből eredő színeképek kialakulását a 11. ábra szemlélteti. A hangajkak rezgésekor a keletkező zöngé hangnyomás időfüggvénye egy közel fűrészfog jellegű függvény (T_0) alapperiódussal, és a zöngé színeképi összetevői az alaphang (f_0) és a felhangok együttese, melyek az alaphang egész számú többszörösei (amint ezt már részleteztük az előző fejezetben). Ezt a színeképet befolyásolja a változó méretű hangképző csatorna, amely egy üregrendszer, több rezonanciafrekvenciával (Kent Ray D., 1992.)

A rezonanciafrekvenciákon és azok környezetében a részhangok intenzitása megnő, más helyeken elnyomódik. Így alakulnak ki a magánhangzók, nazális mássalhangzók és a likvidák. A fonetikai szakirodalomban a rezonanciacsúcsokat formánsoknak, a rezonanciafrekvenciákat formánsfrekvenciáknak (F) nevezzük. Az egyes rezonanciacsúcsok szélessége (B) szintén jellemző az adott üregre. A magánhangzó típusát főleg az első (F_1) és második (F_2) formáns szabja meg. (Chistovich, L.,1980). A magasabb formánsok a színezetre jellemzőek.

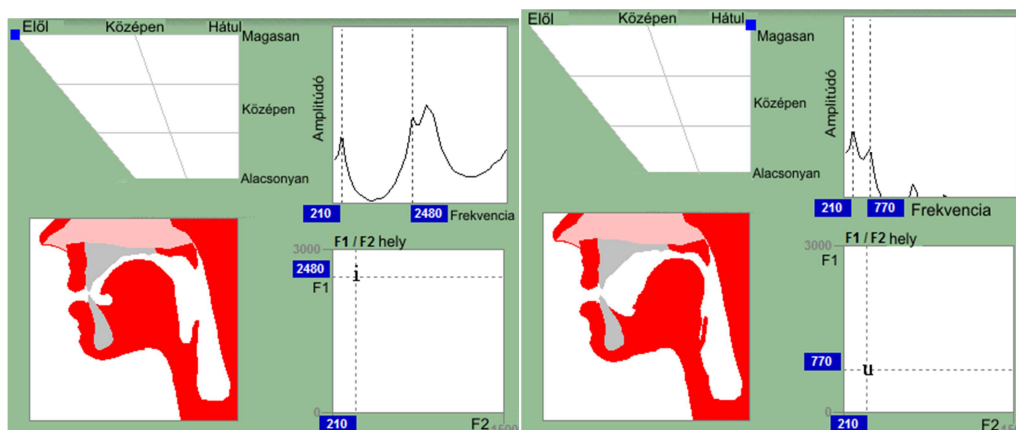
A képzési hely függvényében a hangképző üregek mérete változik, és így változik a színekép is. Erre példát a 12. és 13. ábrákon mutatunk be. Az ábrák bal felső részében az úgynevezett fonetikai négyszög látható, a jellemző artikulációs pozíciók egy kétdimenziós síkbeli

ábrázolása, amelyben vízszintesen a nyelv vízszintes mozgási pozíciója (elől, közepén, hátul), függőlegesen az állkapocs, és ezzel együtt a nyelv le-föl mozgása (magasan, közepén, alacsonyan) van feltüntetve. Az ábrák bal alsó részén a hangképző csatorna aktuális keresztmetszete látható. Jobb felső részen az adott képzési pozícióban kialakult színképet tüntettük fel, bejelölve az aktuális első és második formánsfrekvenciát. A jobb alsó ábra az aktuális magánhangzó elhelyezkedését mutatja az első az F_1 és a második F_2 formánsfrekvenciák által meghatározott síkban. A magyar „á” hangot közepén képzett alsó nyelvállású, igen nyílt állkapcsú, nyitott ajakműködésű, legtágabb hangképzéssel állítjuk elő, amint azt a 16. ábrán mutatjuk. Spektrumában az F_1 és F_2 formánsértékek egymáshoz közel, általában 800-900, ill. 1200-1300 Hz környezetében fordulnak elő férfiak ejtésében.



16. ábra Az „á” beszédhang képzési pozíciója, spektruma, F_1 F_2 formánsfrekvenciái (Sensimetrics, 1997)

Az „i” hangot pedig elől képzett, felső nyelvállású, zárt állkapcsú ajakréssel képezzük, amint azt a 13. ábra bal oldala mutatja. Formánsfrekvenciái egymástól távol, 230-300, ill. 2300-2500 Hz környezetében vannak (Magdics K., 1965).



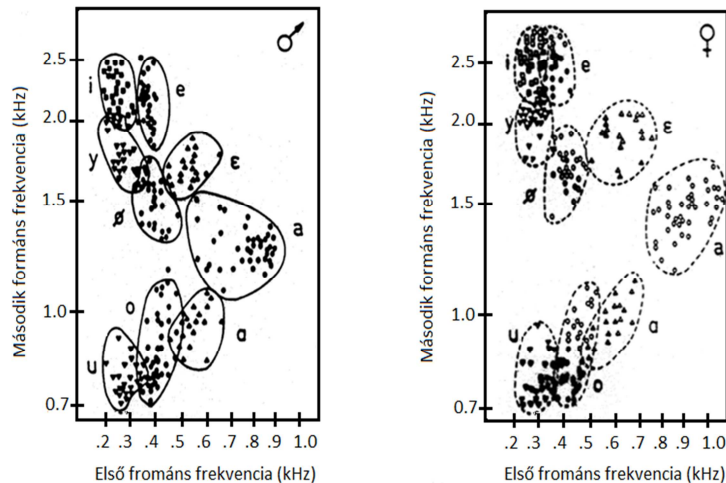
17. ábra Balra az „i” és jobbra az „u” beszédhang képzési pozíciója, spektruma, F_1 F_2 formánsfrekvenciái (Sensimetrics, 1997)

Általánosságban, nagy vonalakban elmondható, hogy amennyiben a nyelv hátramozdul, és a száj kerekedik, az F_2 csökken, az F_1 közelítőleg állandó. Ez történik, amikor „i”

beszédhangból „u” hangot képezünk. A nyelv mozgását és a hozzá tartozó formáns frekvencia értékeket a 17. ábra bal és jobb oldalán szemléltetjük.

Az állkapocsmozgás a nyelv függőleges mozgásával közel szinkronban változik, ha az állkapocs nyílik, az F_1 nő, az F_2 megközelítőleg állandó.

A magyar magánhangzók férfi és női ejtésben az F_1 és az F_2 formánsfrekvenciák által meghatározott síkban a 18. ábrán tanulmányozható. Az ábrák a mért F_1 és F_2



18. ábra Magyar magánhangzók F_2 értékei az F_1 függvényében férfi és női ejtésben, IPhA jelölésszerrel (Tarnóczy, 1974)

formánsfrekvenciák szórásterületeit mutatják 4 férfi és 4 női ejtésben. Az ábra jól szemlélteti, hogy mivel általában a női hangképző üregek kisebb méretűek, mint a férfiak hangképző üregei, ezért a formánsfrekvenciák női ejtésben valamivel magasabbak, mint a férfiaknál, és gyermekeknél még magasabb, természetesen egy adott határon belül. Nemcsak az alaphang és ezzel a felharmonikusok egymás közötti távolsága változik a hangjak mérete függvényében, hanem a spektrumban a felharmonikusok nagysága, és egymáshoz viszonyított aránya, az F_1 és F_2 értékei is.

A folyamatos artikuláció következtében a hangképző üregek méretei is folyamatosan változnak. Normál tempójú beszédben a magánhangzónak van egy kialakulási, egy kvázistacionárius (célkonfigurációt tükröző) és egy befejező szakasza. A színek, ezzel együtt a formánsfrekvenciák, a kvázistacioner szakaszban tekinthetők közel állandónak. A kialakulási és befejező szakaszban a színek a formánsokkal együtt erősen változhat a szomszédos beszédhangok képzési helye függvényében. Erre példát a 19. ábrán mutatunk be, ahol a magánhangzók színekváltozása jól követhető az ábrázolt mondattöredékben. Gyorsabb beszédtempónál a beszédhangok kvázistacioner szakasza rövidül meg, gyakran el is tűnik.

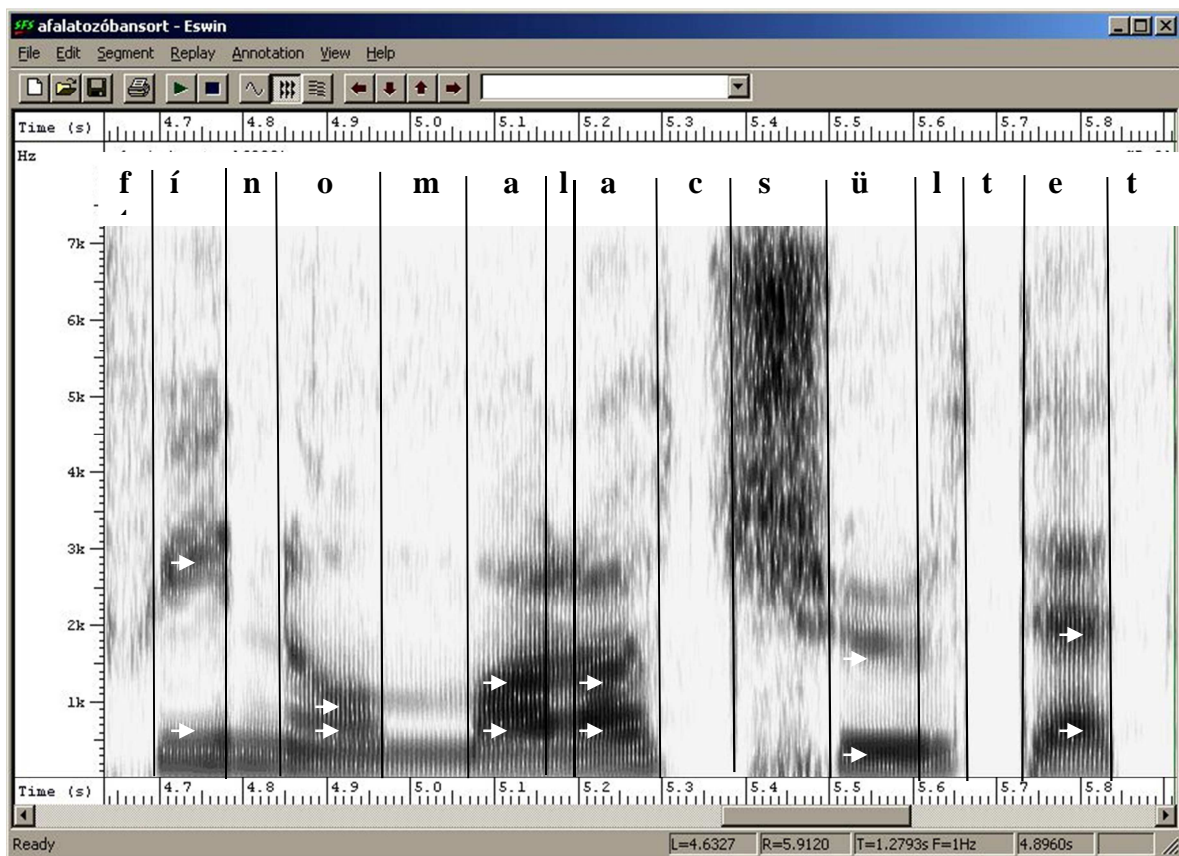
Ma a gépi beszédfeldolgozási eljárásokban nemcsak a formánsfrekvenciákkal dolgoznak, hanem figyelembe veszik a teljes spektrumot (Lass, N.,J 1996. Morgan, G. 2000. Tetschner, W. 1993.)

3.1.5. Mássalhangzók

Képzésük során a szájüregben akadály képződik, melyen a levegő átáramolva, vagy az akadályt megszüntetve kisebb, nagyobb erősségű zörej keletkezik. Vannak rezonáns jellegű mássalhangzók és zörej jellegű mássalhangzók, amelyek lehetnek zöngések és zöngétlenek.

A *likvidák* rezonáns jellegű mássalhangzók. Formánsokkal és antiformánsokkal rendelkeznek, energiájuk jellemzően kisebb, mint a magánhangzóké. Likvidákra példa az „l” hang a 19.. ábrán.

A *nazális mássalhangzók* szintén rezonáns jellegű mássalhangzók, de itt az orrüreg, egy állandó méretű rezonátorüreg is szerepet kap a hangképzésben. Ez adja meg a nazális beszédhangok jellegzetes színezetét. Amikor levegő áramlik az orrüregen és a szájüregen keresztül – nazooralis magánhangzókat, ha csak az orrüregen keresztül – nazális mássalhangzókat képezünk. A nazális formáns frekvenciája 250 – 300 Hz, a többi formáns hely a képzés helyétől függ. Az antiformáns helyek (alacsony (750 – 1250 Hz), középső (1450 – 2200 Hz) és magas (3000 Hz fölött)) szintén változnak a képzési helytel. Energiájuk jellemzően kisebb, mint a magánhangzóké. Nazális hangokra példa az „n” és „m” hang a 19.. ábrán.



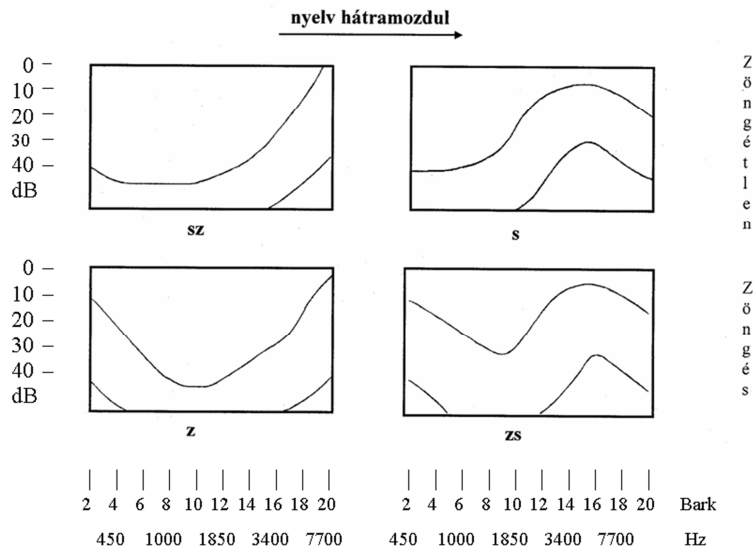
19.. ábra A „finom malacsültet” mondattöredék kézzel szegmentált és címkézett spektrogramja

A magánhangzók közepén az F_1 , F_2 formánsfrekvenciák környezetét a fehér nyilak mutatják

Réshangok képzésekor a hangképző csatornában kiáramló levegő szűk résen halad át. Turbulens áramlás keletkezik, amely széles spektrumú súrlódási zörejt eredményez. A képzési hely határozza meg, hogy a keletkező zörejben hova esnek a színeképi súlypontok. Réshangokra példa a 19.. ábrán az „f” és „s” hang.

Amennyiben a képzésnél például a rés a fogmedertől, a zöngétlen „sz” és a zöngés „z” képzési helyétől, hátrafelé tolódik a velum (lágú szájpad) felé, a zöngétlen „s” és a zöngés „zs” képzési helye felé, akkor a színekpi energiasúlypont lefelé, a kisebb frekvenciák irányába tolódik. Erre példa a 20. ábra hallási spektrumain látható a zöngétlen „sz” és a zöngés „z”, valamint a zöngétlen „s” és a zöngés „zs” esetében. Az ábrán a kritikus frekvenciasávokban mért energiaszintek szórásstartománya látható 72 beszélő ejtésében (Vicsi K., 2003).

A réshangokra általánosan jellemző, hogy személyfüggésük nincs.



20. ábra Két zöngés és két zöngétlen réshang kritikus frekvenciasávokban mért energiaszintek szórásstartománya.. A színekpi energiasúlypont változódik a képzési hely függvényében

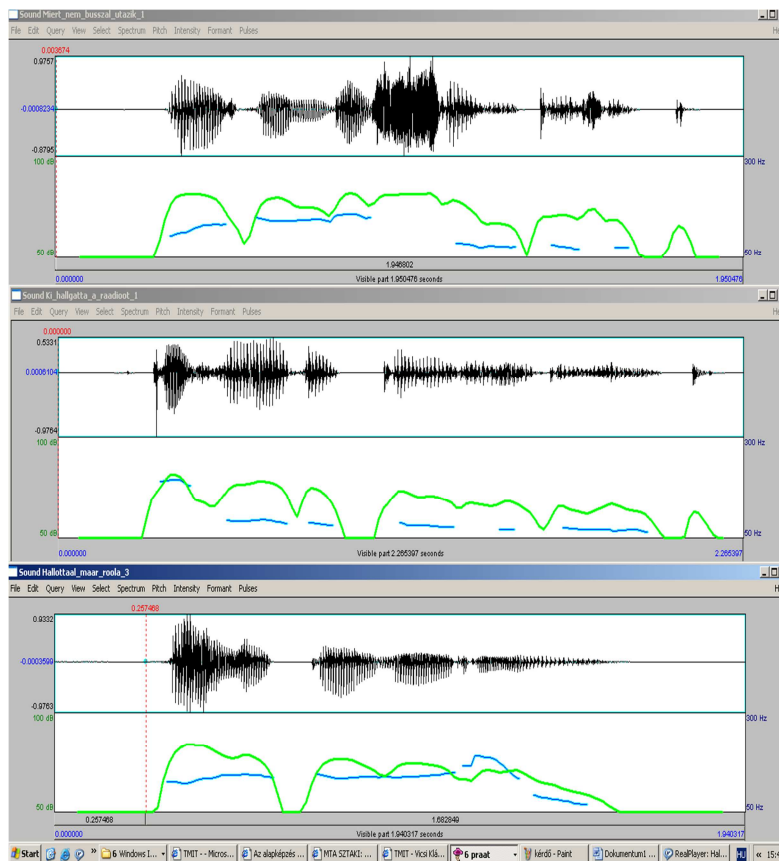
A rövid réshang időtartama 100-200 ms, hosszúé 200 - 300 ms, a beszédstílustól függően.

Zárhangok képzésekor a gerjesztés a zár felpattanásával történik. A létrejövő zöreje a zárfelpattanási zöreje, amelynek színekpi súlypontja a képzés helyétől függ, ugyanúgy, mint a réshangok esetén. Belső időszerkezettel jellemezhető, akusztikailag összetett beszédhang, zárképzési tranziens után következik a zár (vagyis zöngétlen esetben az energiamentes rész, zöngés esetben a fojtott zöngé), majd a zárfelpattanási zöreje. Igen jellemzőek a zár képzési helyére a zárhangot követő magánhangzók F_1 és F_2 formánsfrekvenciáinak átmenetei a kvázistacioner állapot elérése előtt. Zárhangokra példa a 19. ábrán a „t” hang.

Affrikáták, vagyis a zár-rés hangok képzésekor zár és rés képzése történik azonnali egymásutánban. Belső időszerkezettel jellemezhető összetett beszédhang, akusztikai komponensei: zárképzési tranziens, zár (vagyis zöngétlen esetben az energiamentes rész, zöngés esetben a fojtott zöngé), zárfelpattanási zöreje, spiráns zöreje. Affrikátára példa a 19. ábrán a „c” hang.

3.2. A beszéd szupraszegmentális leírása

A több beszédhang együttes viselkedését leíró jellemzőket nevezzük szupraszegmentális vagy prozódiai jellemzőknek. Ezek a *hangsúly*, a *hanglejtés (intonáció)*, a *beszédtempó*, *beszédritmus*, *hangerő* és a *hangszínezet*. A szupraszegmentális leíráshoz használt jellemző fizikai paraméterek az intenzitás [dB], az alaphang [Hz], időtartam [sec], a spektrum [dB/Hz]. Néhány mondat mért alaphangjának és intenzitásának időbeli változását a 20. ábrán mutatjuk be.



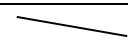

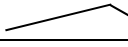
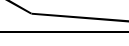
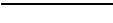

20. ábra. „Miért nem busszal utazik?” „Ki hallgatta a rádiót?” „Hallottál már róla?” kérdő mondatok hullámformája, intenzitás (zöld) és alaphang (kék) időbeli változása

Nyelvenként ezeknek a fizikai paramétereknek más-más arányú keverékei jelenhetnek meg ugyanannak a prozódiai jellemzőnek a képviselőjében. Pl. az orosz hangsúly előidézésében nagyobb mértékű az időtartam részesezése, mint a másik két fizikai paraméteré. Ugyanakkor a francia hangsúly észlelésében a hangmagasság a döntő tényező, így nem ritka, hogy egy francia mondatban a hangsúlyos szótag kevésbé intenzív, mint a hangsúlytalan, viszont hangmagasságban kiemelkedik. Továbbá az oroszban és az angolban például valamely paraméter „pozitív” értéke mellett „negatív” módon is kifejezésre jut a hangsúly: a nem hangsúlyos szótagok magánhangzóinak minőségi redukciójában (Kassai I., 1998).

3.2.1. Hanglejtés

A beszélő legalacsonyabb és legmagasabb alaphang-értékének a különbsége adja a beszédhang terjedelmét. Az alaphangterjedelmen belül figyelembe vesszük az alaphang-változás irányát, a hangmenetet, vagyis a hanglejtést. A hangmagasság-változás három lehetséges iránya, az ereszkedés, a szinttartás és az emelkedés, illetőleg ezek gyorsabb időbeli lefutású változatai, a szökés és az esés egymással kombinálódva különböző dallamsémákat hoznak létre, amelyek lehetnek emelkedő-esők, emelkedő-ereszkedők, szinttartó-ereszkedők, ereszkedő-emelkedők, stb. A magyar nyelvben, erős leegyszerűsítéssel a leggyakoribb 6 hanglejtéstípust a 3. táblázatban mutatjuk be, a részletekre nem térve ki.

3. táblázat A magyar nyelv leggyakoribb hanglejtéstípusai

Hanglejtés típusa	példák
ereszkedő 	A fán mókus volt.
gyors eső 	Melyik?
emelkedő – eső 	Balázs hol van?
gyors eső – ereszkedő 	Ki zenél?
lebegő 	Nem, már hazament.
szökő 	Áll?

Ha a dallamminták egy szón belül jellemzőek, tehát minden szónak megvan a maga önálló dallamképlete, amelyhez önálló jelentés társul, akkor *tonális* vagy *polifon* nyelvvel állunk szemben. Ha viszont a szó hanglejtése valamely magasabb rendű nyelvi egységnek (szószerkezetnek vagy mondatnak) van alárendelve, akkor monoton nyelvről beszélünk. Tonális vagy polifon nyelvre példa az Európában beszélt nyelvek közül a szerbhorvát, a litván, a svéd és a norvég. A norvég nyelvben például a kokken fonémasor attól függően jelenti azt, hogy „szakács”, vagy azt, hogy „főzni”, hogy végig emelkedő dallammal, avagy ereszkedő-emelkedő dallammal mondjuk-e ki (Kassai I., 1998).

A monoton nyelvekben, amilyen a magyar is, a hanglejtésnek a mondat szintjén teljeseedik ki a funkciója, ennek megfelelően egy szónak a hanglejtése attól függően ereszkedő, szinttartó, vagy emelkedő, hogy milyen típusú mondatban, annak mely pontján és milyen mondattani szerepben áll

3.2.2. Hangsúly

A beszélő leggyengébb és legerősebb hangja által képviselt dinamikai tartományon belül tekintetbe vesszük a hangerő szintjét (erős, közepes, gyenge) és a hangerőváltozás irányát (gyengülő, szinttartó, erősödő), valamint ezek kombinációit (gyengülő-erősödő, szinttartó-gyengülő, erősödő-gyengülő stb.). A változás időbeli lefolyása szerint beszélünk hirtelen és enyhe változásról, a mértéket a dB/s hányados fejezi ki.

Amikor a közlés valamely szótagját vagy szótagjait a többihez képest kiemelő hangsúlyúnak tekintjük, ennek a nyelvi tényezőnek a fizikai megfelelőjét keressük, még nehezebb a dolgunk, mint amikor a hanglejtés fizikai hátterét igyekeztünk tisztázni, mivel a hangsúly előidézésében az intenzitás még annyira sem tekinthető alapvetően meghatározónak, mint az intonáció létrejöttében az alaphangváltozása. Az intenzitás változásának járulékos mozzanata az alaphang változása. Szubjektív lehallgatási tesztek szerint ahhoz, hogy egy hangsúlyt észleljünk a megnyilatkozásban, a szótagok között 10 dB feletti hangerőkülönbségnek és kis tercet meghaladó hangközkülönbségnek kell lennie. Ha ennél kisebbek a fizikai különbségek, a döntés elveszti egyértelmű jellegét, és akadnak hallgatók, akik több hangsúlyt észlelnek, illetőleg tartózkodnak a döntéstől.

Amennyiben a hangsornak mindig ugyanazt a sorszámú szótagját emeljük ki, akkor kötött hangsúlyról beszélünk. De változhat a hangsúly helye szavanként is, ekkor szabad hangsúlyról beszélünk. A magyar kötött hangsúlyú nyelv, mert normális közlési körülmények között a szónak mindig az első szótagja viseli a hangsúlyt. Ide tartozik még például a francia nyelv, amelyben a hangsúly hagyományosan az utolsó szótagon van.

3.2.3. Beszédtempó

A *beszédtempó* (beszédsebesség), az időegységre jutó nyelvi jelek száma (Gósy M., 2004), függetlenül attól, hogy a közlésben volt-e szünet vagy más megakadásjelenség. A nyelvi jelek lehetnek beszédhangok, szótagok, ritkábban szavak. Egy nyelvközösségen belül a

beszédtempó viszonylag állandó jellemző. Vannak népek, amelyek gyorsabban beszélnek, mint mások. A magyar beszéd tempóját szokásosan beszédhang/másodpercben adjuk meg. Az átlagos köznapi beszéd tempóindexe a mai magyar köznyelvben 12 beszédhang/mp, amely a közléstartamba beleérti a szünetet, hezitációt, tehát minden, nem a beszédképzésre fordított időt is. Az *artikulációs tempó* az artikuláció tiszta idejére eső nyelvi jelek számát jelenti, vagyis ekkor nem vesszük figyelembe a szüneteket és megakadásjelenségeket.

Az észlelés oldaláról a beszédtempónak általában három fokozatát különböztetjük meg: a **lassú**, a **közepes** és a **gyors** tempót. A tempó időbeli változása tekintetében pedig beszélünk gyorsuló és lassuló tempóról. A beszéd sebességének érzete azonban nemcsak a beszédképzés időzítésétől függ, hanem a hangfolyamatot megszakító szünetek számától és időtartamától is.

3.2.4. Beszédritmus

Perceptuális jelenség. A beszéd egy vagy több jellemzője (időtartam, hangsúly) rendszeres ismétlődésének az eredménye. A változások rendszerint szótag egységekben valósulnak meg.

3.2.5. Hangerő

A beszéd átlagos hangerejét fejezi ki, amelyet leginkább a légzési mechanizmusunkal tudjuk szabályozni. Halkabb környezetben halkabban, zajos környezetben hangosabban beszélünk. A hangerő beállításánál is nagy szerepe van a hallórendszeren keresztüli visszacsatolásnak.

3.2.6. Hangszínezet

A spektrum vagy hangszín a beszédhangok színeképi jellemzésére használatos, míg a hangszínezet kifejezést a beszélő személy beszédének jellemzésére használja a szakirodalom (Gósy M., 2004). Az egyénre jellemző hangszínezet számos paraméter együttes és sajátos elrendeződésének a következménye. A beszélő egyéni beszéd-sajátosságai, az egyéni hangképző szervek mérete, izomzata közötti variáltság következménye. Soros kapcsolatban van a többi szupraszegmentális jellemzővel. Az érzelmek kifejezésében is fontos szerepe van (Tóth sz. L., Vicsi K., 2007).

3.3. A beszéd variáltsága

A beszéd jellemző fizikai paraméterei számos hatás következtében megváltoznak, variáltságuknak számos forrása van, amelyek az elemzést, feldolgozást megnehezítik. A beszéd fizikai megvalósulásakor a hangképző szervek méretei, a vezérlő izmok feszítettsége, az üregek falának rugalmassága stb. mind-mind befolyásolják egy-egy beszédhang fizikai jellemzőit.

A fizikai paraméterek variáltságát okozhatják a környezeti, akusztikai körülmények. Ilyenek pl. a zajos, zajtalan környezet, visszhangok, termek, telefonbeszéd stb..

A folyamatos beszédben a kapcsolódó beszédhangok megváltoztatják, elcsúsztatják az adott hang képzési helyét, ezzel a keletkezett hang fizikai paraméterei is megváltoznak.

E variáltság ellenére mégis meg lehet határozni, hogy milyen közös jegyek alapján ítélik meg anyanyelvi beszélő a nyelvi tudása alapján két hangot azonos fonetikai osztályhoz tartozónak. A variáltságot okozó tényezők számos módon csoportosíthatók, mégis talán az egy beszélőnél és a több beszélő közötti variáltság szerinti csoportosítás a legmegfelelőbb (Vicsi K., 2002).

Variáltság egy beszélőnél:

A hangképzés folyamatosan változó mozgások összessége. A folyamatos hangképzőszervi mozgások miatt az egyik hang fizikai tulajdonságai befolyásolják az azt megelőző és követő hangok fizikai tulajdonságait. Ezt nevezik koartikulációs hatásnak. Egy beszélő különböző hangkörnyezetben ejtett ugyanazon beszédhangjainak fizikai jellemzői, például formánsfrekvenciái eltérnek egymástól, természetesen adott határokon belül. Egy beszélőnél a ritmus, hangerő, hangmagasság, hanglejtés, nyomatékbeli különbségek szintén a fizikai paraméterek eltéréseihez vezetnek.

A megfázás igen nagymértékben megváltoztatja a hangok akusztikai paramétereit, hiszen a megvastagodott nyálkahártyájú, esetleg folyadékkal terhelt üregek rezonanciafrekvenciái eltolódnak. A környezeti hatások, izgalom, meglepetés stb. szintén hatással vannak a létrehozott beszéd akusztikai tulajdonságaira.

Több beszélő közötti variáltság:

Biológiai tényezők többek között a beszédképző szervek méretkülönbsége, ami az akusztikai paraméterek jelentős variáltságát okozza női, férfi, gyermekhangok esetében, de egy-egy csoporton belül is.

Nyelvi különbségekhez tartoznak az egy nyelvközösséghez tartozó embercsoportok ejtésbeli különbségei.

Környezeti hatások okozta variáltság:

A statikus (teremakusztikai hatások, utózengei idő, rögzítő berendezések stb.) és dinamikus (zaj, mikrofonpozíció stb.) hatások szintén erősen befolyásolják a beszéd akusztikai paramétereit.

4. Ajánlás

Egy átfogó kézikönyv terjedelme nem teszi lehetővé sem a beszédakusztika, sem a fonetikai alapok bővebb tárgyalását. Azonban részletekben menő, főleg kutatók számára javasolt Kenneth N. Stevens *Acoustic Phonetics* című könyve, amely mind akusztikailag, mind fonetikailag korrekt részletes ismeretet ad (Kenneth N. Stevens 2000). Átfogó, főleg nyelvész hallgatóknak ajánlott Peter B. Denes and Elliot N. Pinson: *The Speech Chain* című könyve (Peter B. Denes and Elliot N. Pinson 1993), amelyben a beszéd-kommunikáció teljes körfolyamatát végig tárgyalja, kitérve a digitális beszédfeldolgozásra is. Műszaki vagy matematikai érdeklődésű olvasóknak ajánlott Gordos Géza – Takács György *Digitális beszédfeldolgozás* című könyve (Gordos G. – Takács Gy. 1983), A magyar beszéd című könv (Németh, Olaszy, 2010) valamint Douglas O'Shaughnessy: *Speech Communications: Human and Machine* című könyve (O'Shaughnessy, D. 2000). Átfogó tudományos mű Gósy Mária könyve a *Fonetika, a beszéd tudománya* (Gósy 2004), amelyben a magyar beszéd részletes fonetikai leírása megtalálható.

Irodalmi hivatkozások:

Ainsworth, W.A.: *Mechanisms of Speech Recognition*, Pergamon Press, Oxford, 1976.

A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Szerk. : Németh Géza, Olaszi Gábor. Akadémiai Kiadó 2010.

Chistovich, L. A.: *Auditory processing of speech*, *Language & Speech* 23, 1980.

Denes, P. B. and Pinson, E. N.: *The speech Chain*, W.H. Freeman and Company, New York, 1993.

Ferreira, F., Anes, M.: 'Why Study Spoken Language?', In Gernsbacher, M. A.: *Handbook of Psycholinguistics*, Academic Press, San Diego, New York, Boston, 1994.

Ghitza, O.: *Auditory nerve representation as a basis for speech processing*. In Furui, S. and Sondhi, M. M. (Eds.), *Advances in speech signal processing*, New York: Marcel Dekker (pp. 453-485) 1992.

Gordos, G., Takács, Gy.: *Digitális beszédfeldolgozás*. Műszaki Könyvkiadó, Budapest, 1983. Műszaki Könyvkiadó, Budapest, 1989.

Gósy, M.: *Fonetika, a beszéd tudománya*, Osiris Kiadó, Budapest, 2004.

Kassai, I.: *Fonetika*, Nemzeti Tankönyvkiadó, Budapest, 1998.

Kent, R. D, Charles, R.: *The Acoustic Analysis of Speech*, Singular Publishing Group, Inc. San Diego, California, 1992.

Lass, N., J.: *Experimental Phonetics*, Mosby-Year Book, Inc. St. Louis, 1996.

Magdics, K., *A magyar beszédhangok akusztikai szerkezete*. Nyelvtudományi Értekezések

- 49, Budapest, Akadémiai Kiadó, 1965.
- Moore, B. C.J: An Introduction to the Psychology of Hearing, Academic Press, London 1982
- Morgan, Gold: Speech and Audio Signal Processing, John Wiley&Sons, Inc., New York, 2000.
- O'Shaughnessy, D: Speech Communications: Human and Machine IEEE Press, 2000.
- Owens, F. J.: Signal Processing of Speech, The Macmillan Press Ltd., London, 1993.
- Pap, J.: Hang-Ember-Hang Tudomány-Egyetem sorozat, Vince Kiadó, Budapest, 2002.
- Sensimetrics, Speech Production and Perception, User Guide, Sensimetrics Corporation, 1997
- Subosits, I.: Hangtan, Tas-11 Kft., Budapest, 2004.
- Stevens, K. N.: Acoustic Phonetics, MIT Press. Cambridge 2000.
- Tarnóczy, T.: A magánhangzók akusztikai vizsgálatának problémái, (Általános nyelvészeti tanulmányok X., A nyelv hangdomíniuma, Budapest, 1974)
- Tarnóczy, T.: Zeneakusztika, Zeneműkiadó, Budapest, 1982.
- Tetschner, W.: Voice Processing, Second Edition, Artech House, Boston, London 1993.
- Tóth, Sz. L., Sztahó, D., Vicsi, K.: Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction, In Esposito, A. : Emotion Perception by Human and Machine, Springer 2008.
- Vicsi, K.: SAMPA computer readable phonetic alphabet, Hungarian <http://www.phon.ucl.ac.uk/home/sampa/hungaria.htm>, 1996.
- Vicsi, K.: 'Beszédatbázisok', Olasz G.: 'Magyar nyelvi beszédtechnológiai alapismeretek'. http://fonetika.nytud.hu/oktat_hu.htm. NIKOL Kft. 2002.
- Vicsi, K., Vig A.: 'Az első magyar nyelvű beszédatbázis', Beszédkutatás'98, Tanulmányok az elméleti és alkalmazott fonetika köréből. MTA Nyelvtudományi Intézet, Budapest, pp. 163-178, 1998.
- Vicsi K.: Beszédkommunikáció, <http://alpha.tmit.bme.hu/speech/docs/education/beszedkomm.pdf>, 2003.
- Wells, J. C.: 'SAMPa computer readable phonetic alphabet'. In Gibbon, D., Moore, R. and Winski, R. (eds.), 1997. Handbook of Standards and Resources for Spoken Language Systems. Berlin and New York: Mouton de Gruyter. Part IV, section B. 1997.
- Zwicker, E.: Psychoakustik, Springer-Verlag, Berlin 1982

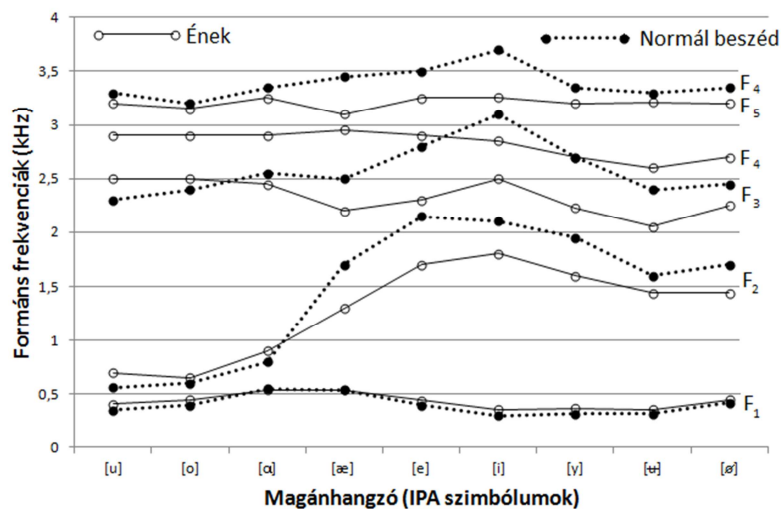
Adatok az éneklés akusztikájához

A hangjakak rezgése során keletkező akusztikus jel csaknem háromszögletű hullámforma, amely felharmonikusokban gazdag, és kb. oktávonként 12 dB-es eséssel csökken, mint ahogy azt az előző Gerjesztett szűrőmodell c alfejezetben tárgyaltuk. A hangképző csatorna, szűrő-rezonátorként szolgál e zümmögés zenei hanggá való átalakítására a magánhangzók képzésekor, és ugyancsak artikulálja a mássalhangzókat. A hangmagasság és a formánsfrekvenciák virtuálisan függetlenek egymástól a beszédben, ám képzett énekesek (különösen a szopránok) néha úgy hangolják magánhangzó formánsaikat, hogy azok az éneklési hangmagasság egy, vagy több felharmonikusához illeszkedjenek. A hangmagasság növelésével a felharmónikusok ritkábbak, mivel a felharmonikusok az alaphang egész számú többszörösei (ld. Gerjesztett szűrőmodell alfejezet 12. ábrája). Például, míg egy beszélt 100 Hz alaphangmagasságú férfihang felharmónikusai 100 Hz frekvenciaközökben jelennek meg, addig egy 440 Hz hangmagassággal éneklő énekes felharmónikusai 440Hz-enként jelennek meg. Sokszor az énekelt magánhangzó rezonanciához közeli felharmónikus frekvenciája nem esik egybe a rezonanciafrekvenciával. Ilyenkor nem erősödik fel az adott felharmónikus

annyira, amennyire a rezonancia miatt felerősödhetne. A képzett énekes a hangképző csatorna formálásával az üregrezonanciát úgy változtatja, hogy a rezonanciafrekvencia egybeessen a legközelebbi felharmónikus frekvenciájával, így érve el a maximális felhangerősödést. (esetleg ide lehet egy ábra, ha nem világos a leírás)

Az énekesformáns kialakulása

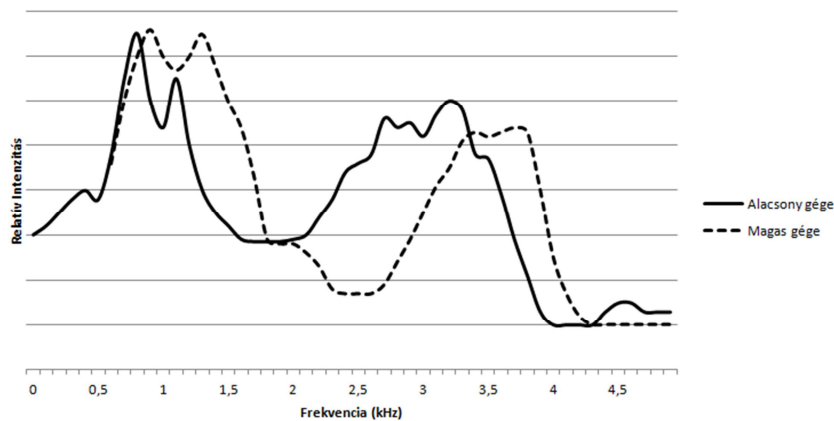
A kiénekelte magánhangzók és formánsaik eltérnek a beszélt magánhangzóktól, a legfontosabb eltérések egyike az *énekesformáns* megjelenése 2500 - 3000 Hz körül. Az énektanulás egyik fontos eredménye a gége leeresztésének elsajátítása, és a garat kinyitása ezen extra formáns létrehozására. Az énekes formáns szemléltetésére kilenc különböző kiénekelte és beszélt magánhangzó formánsfrekvenciáit a 2.ábra mutatja.



2.ábra. Hosszú svéd magánhangzók formánsfrekvenciái egy normál férfibeszédben (szaggatott vonal) és egy hivatásos férfi énekes énekében (vastag vonal). (Sundberg nyomán, 1974.)

Képzett énekesek, különösen a férfi operaénekesek erős formánst mutatnak valahol 2500-3000 Hz között. Ez az „énekesformáns”, amely többé-kevésbé függetlennek tűnik bizonyos magánhangzóktól és a hangmagasságtól, általában a harmadik és a negyedik formáns között helyezkedik el, ragyogást és erőt ad a férfi énekhanghoz.

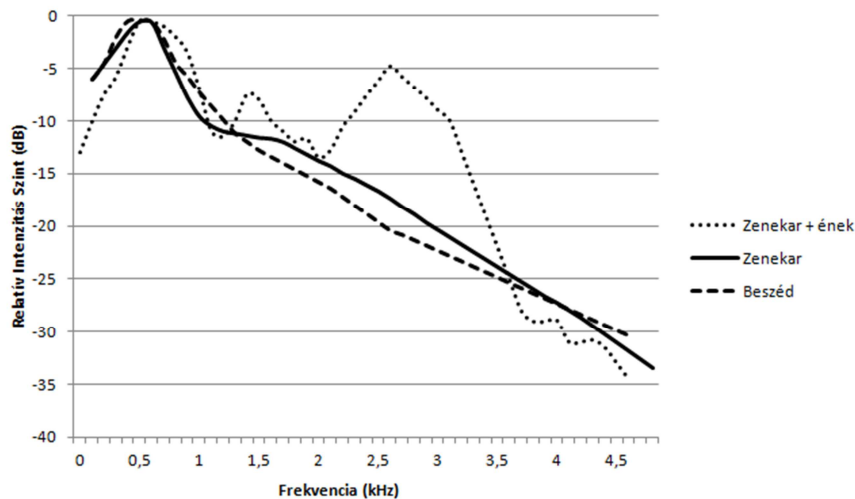
Sundberg (1974) az énekesformánst a leereszkedő gégeének tulajdonítja, amely a kiszélesedő garattal további (kb. 2 cm hosszú) rezonáns üreget képez 2500-3000 Hz-es frekvenciatartományban. A gége, amely több mint 30 mm-rel leereszkedik nyeléskor, 15 mm-rel ereszkedik le énekléskor (Shipp, 1977). Képzetlen énekesek viszont hajlamosak felemelni a gégájukat a hangmagasság emelésekor.



3. ábra. Magánhangzóspektrum az /a/ hangra, magasan álló, és leereszkedett gégével. (Rossing nyomán, 1990)

A 3. ábra mutatja az /a/ („ah”) magánhangzó spektrumát mind magasan álló, mint leereszkedett gégével, egy hivatásos bassz-bariton énekes által kiénekelve. A leereszkedett gége spektrumában 2500-tól 3000 Hz-ig terjedő széles rezonancia a harmadik magánhangzó formáns és az énekesformáns keveréke.

Mínthogy az énekesformáns kiszélesedett gégét igényel („nyílt torok”), a jó éneklés jellemzője a mellkas regiszter (lásd 2. ábra). A hivatásos alténekesnőknél gyakran megtalálható ez a formáns, ám a szopránoknál, akik főként fejregiszterben énekelnek, nem. A fejhangon éneklő férfiénekesek esetében általában ugyancsak nem. A ábra mutatja, Jussi Björling tenorénekest hogyan segítette az énekesformáns „áttörni” egy nagy zenekaron.



4. ábra. A normál beszéd és zenekari zene idealizált átlagspektruma. A pontozott görbe mutatja Jussi Björling átlagspektrumát, miközben egy hangos zenekar által kísérvé énekelt. (Sundbergnyomán, 1977a.)

Különböző énekesek (és beszélők) formánsfrekvenciái meglehetősen széles körben változhatnak, mégis érhető magánhangzókat eredményeznek. Továbbá, az éneklés bizonyos

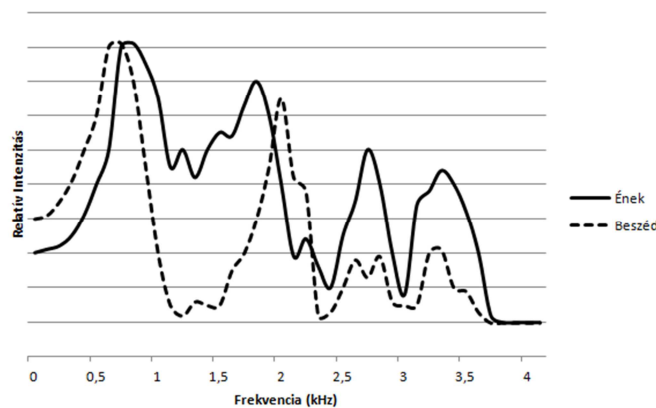
tartományában a magánhangzó formánsok lényegesen megváltoztatják normál frekvenciájukat.

Magánhangzók kiéneklésének nehézségei a hangmagasság növelésével

A hangszalagok kontrollálják a hangmagasságot, a hangképző szerv pedig a magánhangzók formánsaikon keresztül való hangzását, és ugyancsak artikulálja a mássalhangzókat. A hangmagasság és a formánsfrekvenciák virtuálisan függetlenek egymástól a beszédben, ám képzett énekesek (különösen a szopránok) néha úgy hangolják magánhangzó formánsaikat, hogy azok az éneklési hangmagasság egy, vagy több harmóniájához illőek legyenek.

Egy basszus, vagy bariton énekes esetében az alapfrekvenciát ritkán erősíti meg a formánsrezonancia. A legtöbb esetben a formánsok az alapfrekvencia magasabb harmonikusait erősítik; például, ha egy basszus az „a” hangot énekl: G_2 ($f= 98$ Hz) hangmagassággal, az első formáns a hetedik harmonikust erősíti a legjobban, a második formáns a tizenegyedik körüli harmonikusokat erősíti, a harmadik formáns pedig, bár kisebb mértékben (de fontosat amiatt a frekvenciatartomány miatt, amelyben található) a huszonnegyedik, és a huszonötödik harmonikust, és szomszédjait. A hangmagasság természetesen G_2 marad, mivel a felhangok ennek az alapfrekvenciának a harmonikusai.

Az egyéni énekelt magánhangzó formánsok elemzése a basszus és bariton hangok esetében is olyan változásokat mutat, amelyek lényegesek lehetnek. A 5.ábra mutatja az |ae| (kiejtve: „aa”) spektrumát egy hivatásos bassz-bariton énekes által énekelve és kimondva. Jegyezzük meg, hogy az első formáns virtuálisan változatlan, viszont a második formáns alacsonyabb frekvenciájú az énekelt magánhangzóban. A harmadik és negyedik formáns nagyjából ugyanazon hangmagasságban maradt, ám jelentősen erősebbek az énekelt magánhangzó esetében. Az alacsony hangok esetében az artikulációs csatorna különböző formánsai hangsúlyozzák a hangjakakból származó forráshangok különféle harmonikusait.

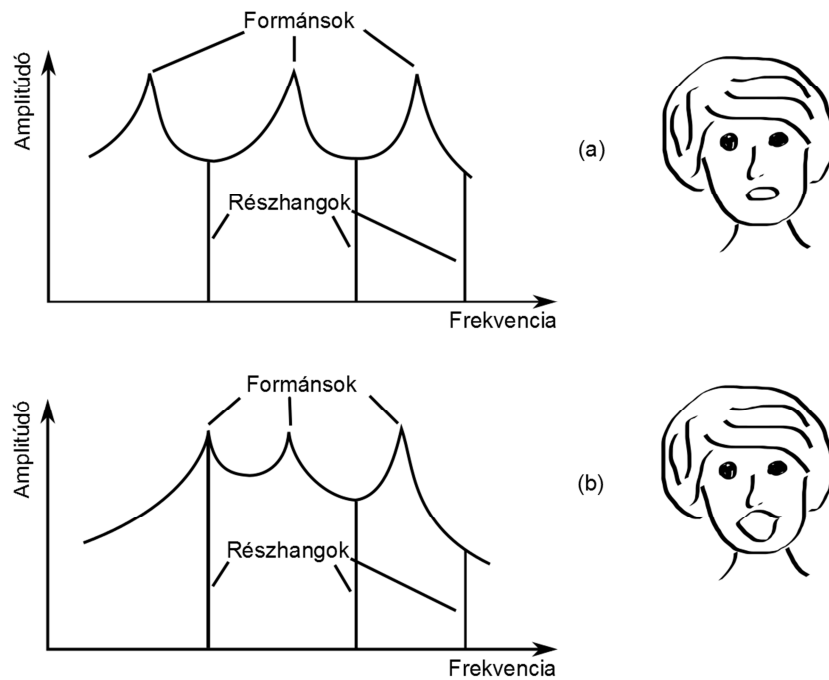


5.ábra. Az |ae| magánhangzó spektrumai egy hivatásos énekes által énekelve és kimondva (Rossing nyomán, 1990)

A szopránok azonban sokszor énekelnek abban a tartományban, amelyben a hangmagasság meghaladja az első formáns frekvenciáját. Tapasztalt szopránok megtanulták, hogyan „hangolják” formánsaikat egy reális frekvenciatartomány felett annak érdekében, hogy a formáns egybeessen az alapfrekvenciával, vagy a kiéneklt hangok felhangjainak egyikével.

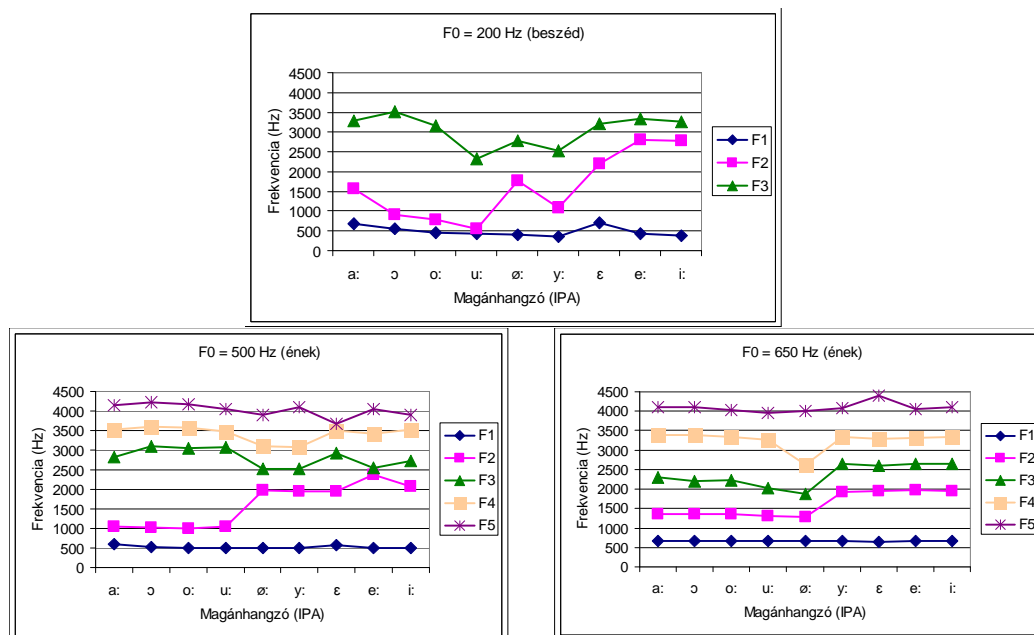
Például, ha egy szoprán $|i|$ -t („ee”) énekel F_5 (698 Hz) hangmagasságban, akkor normál első formánsát 310 Hz-nél találhatjuk, egy oktávval a hangmagasság alatt. Azonban, ha valamelyest szélesebbre nyitja az ajkait a normál pozíciónál, amikor $/i/$ -t mond., a formáns felnyomódhat a hangmagasság környékére. Ha pedig $/A/$ -t („ah”) énekel A_4 -es hangmagasságban (440 Hz), akkor normál első formánsát 700 Hz körül találja, a kiénekel hang alaphangja (440 Hz) és második harmonikusa (880 Hz) között. Az énekesnő valószínűleg sokkal kényelmesebbnek találja ebben az esetben a formánsnak a második harmonikus környékére való felemelését, hogy biztosíthassa a szükséges lendületet. A 6. ábra mutatja, hogyan vihető véghez a formánshangolás az állkapocs nagyobbra nyitásával, a artikulációs csatorna alakjának megváltoztatásával.

Formánshangolásakor esetlegesen számítani lehet a magánhangzók nemkívánatos torzulásaira, de ez nem minden esetben történik így. Hozzászoktunk ahhoz, hogy különböző hangmagasságú magánhangzókat ismerünk fel férfiak, nők és gyermekek beszédében, különböző hosszúságú artikulációs csatornákkal. Ha a hangmagasság nagy, relatíve magas formánsfrekvenciát kapcsolunk hozzá.



6.ábra. Formánshangolás szélesebbre nyitott állkapoccsal: (a) normál első formáns a hangmagasság alatt helyezkedik el; (b) első formáns felemelték, hogy egybeessen a hangmagassággal (Sundberg nyomán, 1977a).

Magyarországon Deme Andrea végzett kutatásokat énekhanggal (Deme A., 2011). Egyik kísérletében egy koloratúr szoprán hangkategóriájú női szolista énekelt magyar magánhangzókat, aki a hangját a klasszikus nyugati énektechnikának megfelelően képezte. Formánsainak alakulására kapott formánsfrekvencia értékek beszédre, valamint két különböző hangmagasságon a 6. a. b. c. ábráján láthatóak.



6.ábra. Koloratúr szoprán női szólista magánhangzóinak formánsfrekvenciái: fenn: beszéd esetén, balra lenn: 500Hz alapfrekvencián énekelve, jobbra lenn 650 Hz alapfrekvencián énekelve (Deme A., 2011)

Deme kísérlete jól szemlélteti a magánhangzók formánsfrekvenciáinak hangolását a hangmagasság növelésével. A beszélt magánhangzókra jellemző variáltság egyre jobban beszűkül. Magas frekvenciákon a magánhangzók formánsfrekvenciái alig különböznek egymástól.

A szoprán tartomány legfelső hangjához közel, ahol a formánshangolás különösen hangsúlyozott, nehéz megkülönböztetni egy magánhangzót a többitől. (Próbáljunk meg hallgatni egy szopránt, amint nagy hangmagasságon énekel különböző magánhangzókat, és meglátjuk, hogy meg tudjuk-e különböztetni őket.) A zeneszerzők meglehetősen óvakodnak ettől a magánhangzó-felismerési nehézségtől, és általánosságban nem jelenítenek meg fontos szövegeket a szoprán tartomány tetején (ha nem tehetnek mást, akkor megismétlik a szöveget alacsonyabb hangmagasságban).

Referenciák

- Appelman, D. R. (1967). *The Science of Vocal Pedagogy*. Bloomington, Indiana: Indiana University Press.
- Bartholomew, W. T. (1940). „The Paradoy of Voice Teaching”, *J. Acoust. Soc. Am.* **11**:446.
- Benade: A. H. (1976). *Fundamentals of Musical Acoustics*. New York: Oxford.
- Bloothoof, G., and R. Plomp: (1984), 1985, 1986) „Spectral Analysis of Sung Vowels. I, II and II.” *J. Acoust. Soc. Am.* **75**:1259; **77**:1580; **79**:852.
- Bjorklund, A. (1961). „Analyses of Soprano Voices”, *J. Acoust. Soc. Am.* **33**:575.
- Bouhuys, A., J. Mead, D. F. Proctor and K. N. Stevens (1968). „Pressure–Flow Events During Singing,” *Annals N. Y. Acad. Sci.* **155**:165.
- Deme Andrea 2010. Az énekelt magánhangzók fonetikai elemzése. In: Parapatics A. (szerk.) 2011. *Félúton 6. A 6. félúton konferencia (2010) kiadványa*. Url: <http://linguistics.elte.hu/studies/fuk/fuk10/>

- Large, J. (1972). „Towards an Integrated Physiologic-Acoustic Theory of Vocal Registers,” *Nat. Assn. Teachers of Singing, Bull.* **28**:18, 30.
- Leanderson, R., J. Sundberg, and C. von Euler (1987). „Role of Diaphragmatic Activity During Singing: A Study of Transdiaphragmatic Pressures,” *J. Appl. Physiol.* **65**:259.
- Mackworth-Young, G. (1953). *What Happens in Singing*. London: Neame.
- Marchesi, M. (1970). *Bel Canto: A Theoretical and Practical Vocal Method*. (Dover reproduction of original undated publication by Enoch & Sons, London.)
- Peterson, G. E. and H. L. Barney (1952). „Control Methods Used in a Study of the Vowels,” *J. Acoust. Soc Am.* **24**:104.
- Rossing, T. D., J. Sundberg, and S. Ternström (1986). „Acoustic Comparison of Voice Use in Solo and Choir Singing,” *J. Acoust. Soc Am.* **82**:830.
- Rossing, T. D., J. Sundberg and S. Ternström (1987). „Acoustic Comparison of Soprano Solo and Choir Singing,” *J. Acoust. Soc. Am.* **82**:830.
- Seymour, J. (1972). „Acoustic Analysis of Singing Voices, Parts I, II, III,” *Acustica* **27**:203, 209, 218.
- Shipp, T. (1977). „Vertical Laryngeal Position in Singing,” *J. Research in Singing* 1.16 (abstract in „*J. Acoust. Soc Am.* **58**:S95).
- Strong, W. J. and G. R. Plitnik (1977). *Music, Speech and High Fidelity*, Provo, Utah: Brigham Young University Press.
- Sundberg, J. (1974). „Articulatory Interpretation of the ‘Singing Formant’,” *J. Acoust. Soc Am.* **55**:838.
- Sundberg, J. (1975). „Formant Technique in a Professional Female Singer,” *Acustica* **32**:8.
- Sundberg, J. (1977a). „The Acoustics of the Singing Voice,” *Sci. Am.* **236**(3)
- Sundberg, J. (1977b). „Singing and Timbre”, in *Music Room and Acoustics*. Stockholm: Royal Academy of Music.
- Sundberg, J. (1978). „Waveform and Spectrum of the Glottal Voice Source,” Report STL-QPSR 2-3, 35. Stockholm: Speech Transmission Lab., Royal Institute of Technology.
- Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.
- Ternström, S., and J. Sundberg (1988). „Intonation Precision of Choir Singers,” *J. Acoust. Soc Am.* **84**:59.
- Titze, I. R. (1973). „The Human Vocal Cords: A Mathematical Model, Part I,” *Phonetica* **28**:129.
- Van den Berg, Jw. (1968). „Register Problems,” *Ann. New York Academy of Sciences* **155**:129.
- Vennard, W. (1967). *Singing: The Mechanism and the Technic*. New York: Carl Fischer.