



NJSZT Informatikatörténeti Fórum –
A beszéd számítógépes feldolgozása Magyarországon

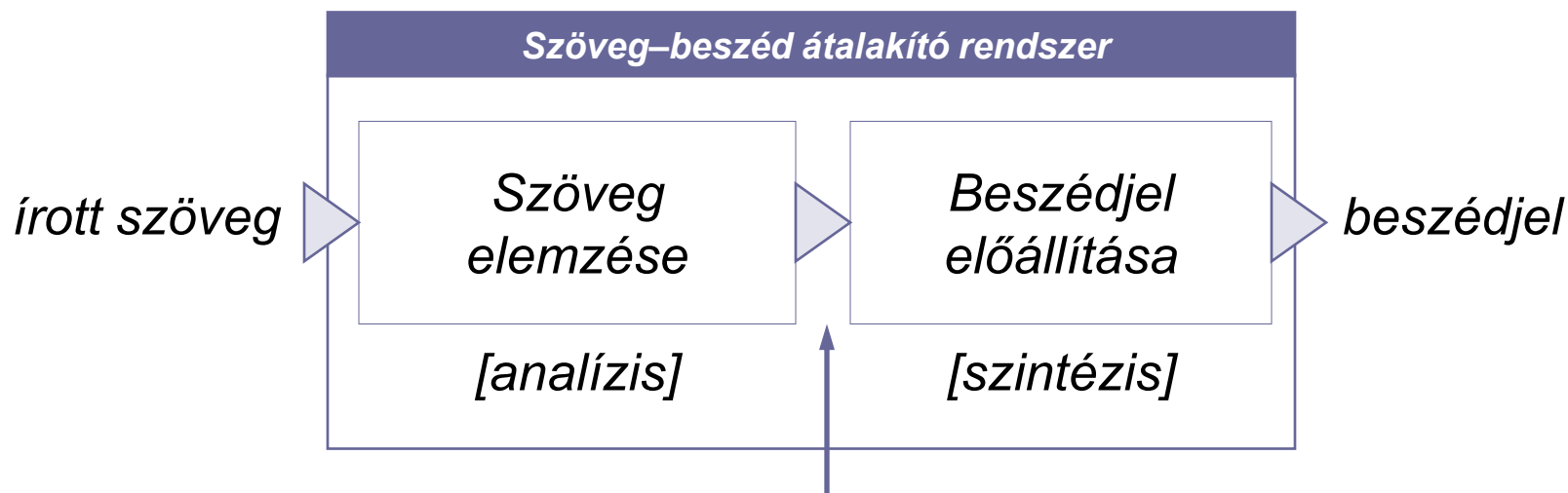
A hullámforma összefűzés bevezetése a beszéd szintézisben – Profivox TTS

Olaszi Péter, Kiss Géza, Olaszky Gábor, Németh Géza

- **Parametrikus szintézis:** A hangképző szervek működését modellezzük. A beszédjel előállítása mesterséges gerjesztéssel és szűrősorral – formánszintézis (HUNGAROVOX, MEA8000 chip)
Jellemzői: gépies jellegű hangzás; kis számítási és tárigény; tág határok között változtatható paraméterek
- **Hullámforma-kódolás:** Természetes emberi beszédből kivágott részleteket tárolunk, és ezek összefűzésével építjük fel a beszédjelet
Jellemzői: természetesebb hangzás; nagyobb számítási és tárigény; szűkebb határok között változtatható paraméterek
Példa: a *hullámforma-összefűzéses* technológia
- Megvalósítható a fenti két módszer kombinációja is

- **A 80-as évek közepéig:** lassú processzorok, kevés memória
Cél: érthető beszédjel előállítása – parametrikus szintézissel
Univoice és Hungarovox: 1982, MTA Nyelvtudományi Intézet
„1982 májusára elkészült az első automatikus, magyarul beszélő szövegszintetizáló rendszer, az Univoice alapváltozata.” 
- **A 90-es évek közepéig:** gyorsabb processzorok, több memória
Cél: többnyelvű szintézis, valamint a *prozódia*, azaz a beszéddallam, az idő- és intenzitás-szerkezet, a hangsúlyozás javítása
Multivox 3: 1986–94, Olasz Gábor – Németh Géza, BME TTT
„Das Multivox Sprachausgabe spricht jetzt mit einer Sound Blaster Karte.” 
Multivox 4: *„Most a Multivox beszéd-szintetizátor négyes változata hallható.”* 
- **90-es évek közepétől:** gyors processzorok, rengeteg memória
Cél: az emberi ejtést megközelítő hangminőség és prozódia elérése; mélyreható nyelvi elemzés; hullámforma-összefűzéses technológia
Profivox: 1999, Olasz Gábor – Németh Géza, BME TTT
„Most a Profivox professzionális szövegfelolvasó beszél.” 

A szöveg–beszéd átalakító rendszer szerkezete

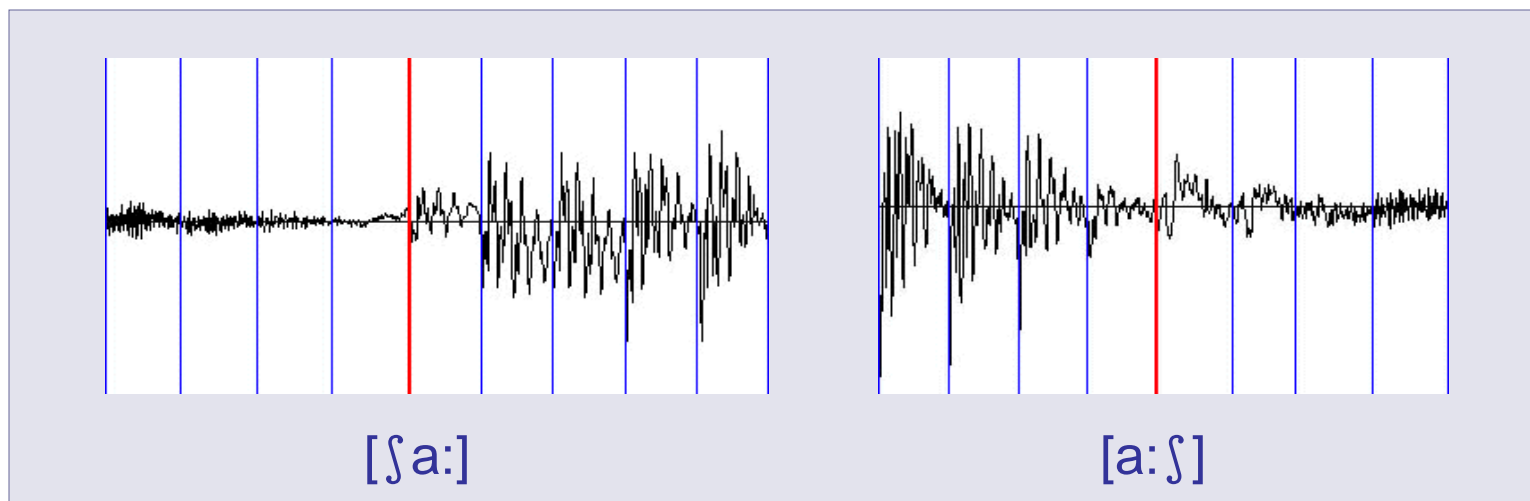


Beszédhangsor: fonetikus szöveg

Prozódiai vezérlés:

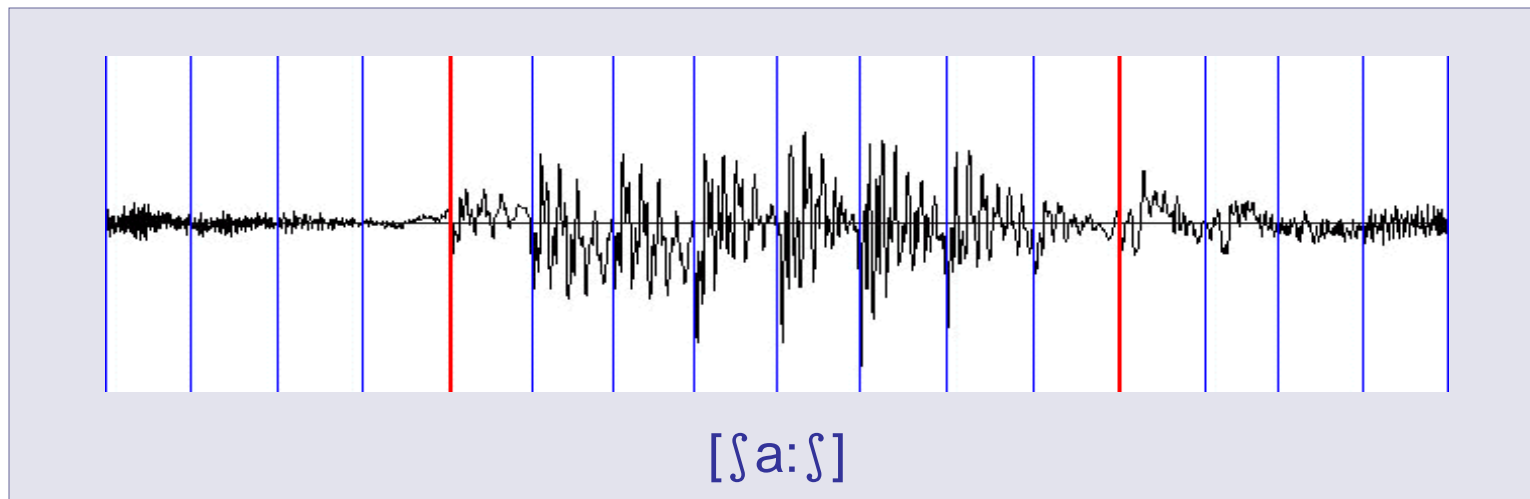
hangidőtartamok, beszéddallam,
hangsúlyozás, intenzitásmenet

A beszédhang-hangadatbázis felépítése diád-elemekkel



- Diádelem: két beszédhang kapcsolata;
a vágási pont a beszédhangok középső, stabil fázisában van
- A beszédjelet a diádelemek összefűzésével állítjuk elő
Példa: *ötven* → [_ø] [øt] [tv] [vε] [εn] [n_] (6 elem)
- Az elemek illesztésénél torzítás léphet fel

A beszédhang-hangadatbázis felépítése triád-elemekkel



- Hanghármasok: kevesebb illesztési pont
- Többnyire msh-mgh-msh kapcsolatok
Példa: *ötven* → [_øt] [tv] [vɛn] [n_] (4 elem)
- Az adatbázisban lehetnek diádok, triádok és hosszabb elemek is
Példa: [øtvɛn]

- Minden beszédhangnak minden beszédhanggal való kapcsolatát rögzíteni kell
- Diádelemek száma: ~1000
Hanghármasok száma: ~4000
- Szisztematikus módszer: hangkapcsolatok felvétele vivőszótagokba ágyazva



1.	báb	abába
2.	báp	abápa
3.	bád	abáda
4.	bát,c,cs	abáta
5.	bág	abága
6.	bák	abáka
7.	bágy	abágya
8.	báty	abátya
9.	bám	abáma
10.	bán	abánat
11.	bány	abánya
11.	báj	abája
13.	báh	abáha
14.	báv	abáva
15.	báf	abáfa
16.	báz	abáza
17.	bász	abásza
18.	bázs	abázsa
19.	bás	abása
20.	bál	abála
21.	bár	abára

1.	bab	ababa
2.	bap	abapa
3.	bad	abada
4.	bat,c,cs	abata
5.	bag	abaga
6.	bak	abaka












- Prozódia: időtartamok, hangmagasság, intenzitás változásai
- Egymás mellé fűzött adatbázis elemek, prozódiai vezérlés nélkül



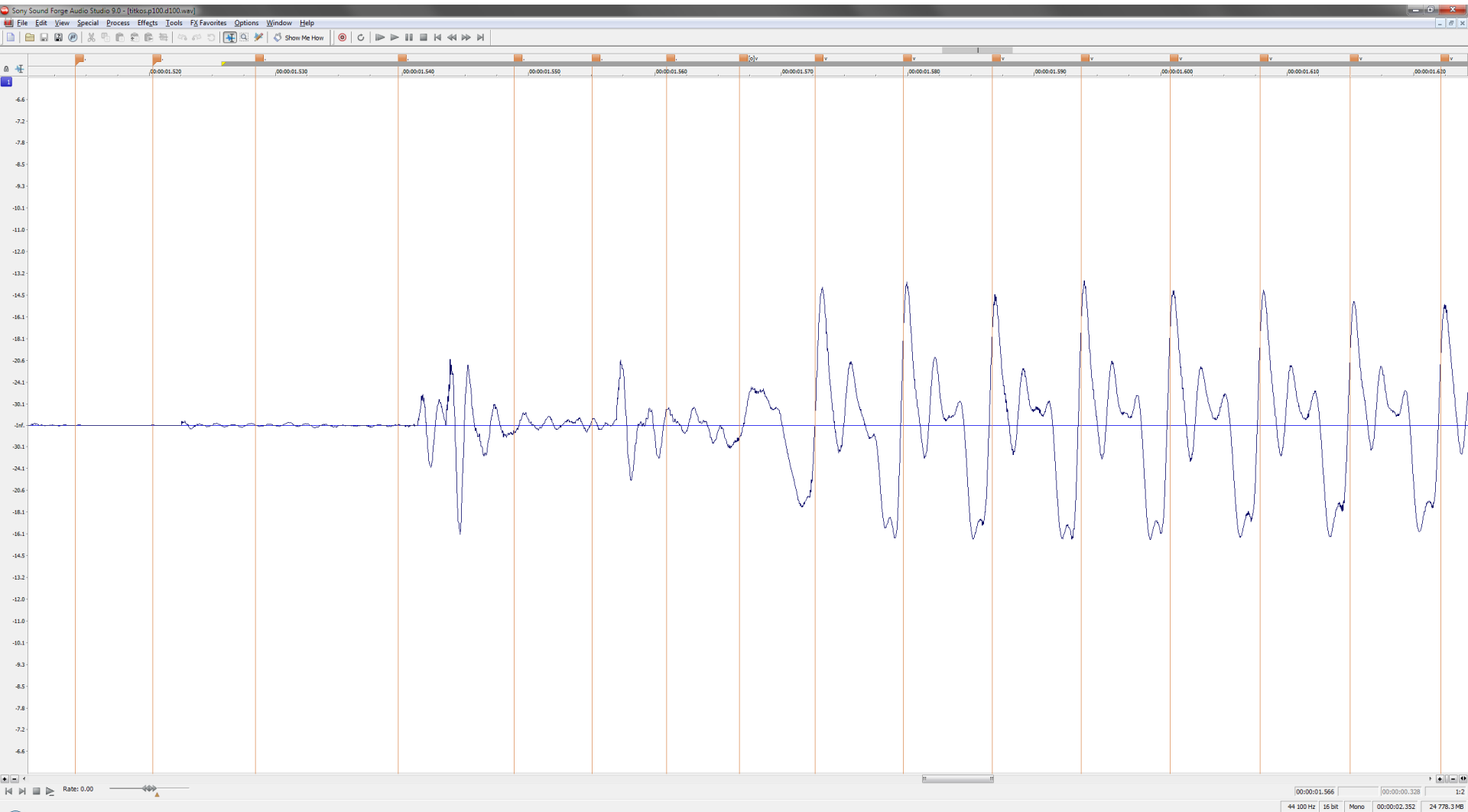
- Ugyanaz a mondat, ráültetett prozódiával



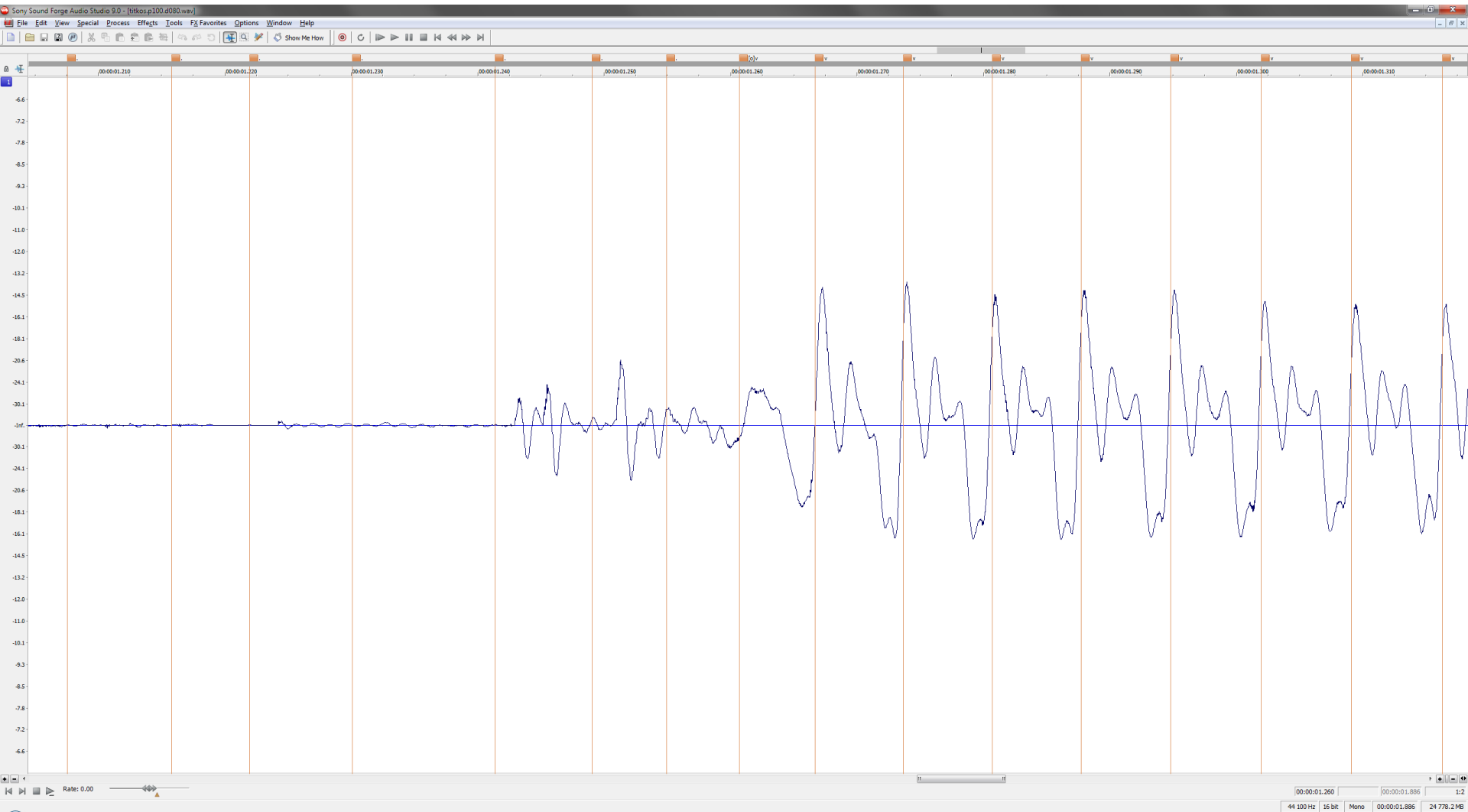
Prozódiai módosítások – összehasonlító táblázat

		Hangmagasság [%]							
		40	60	80	100	120	140	160	180
Időtartam [%]	80								
	100								
	140								
	200								

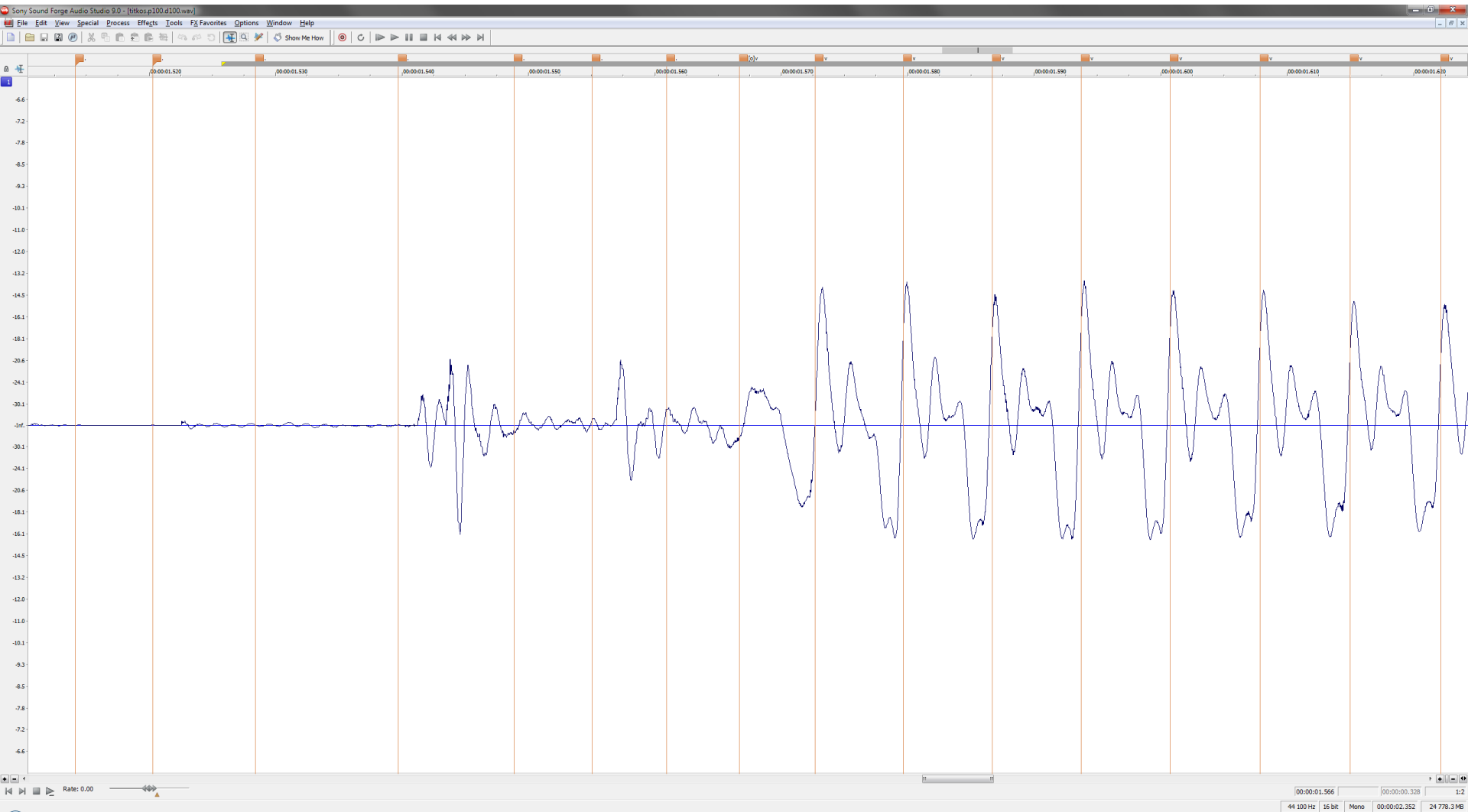
A hangidőtartam módosítása – eredeti beszédjel



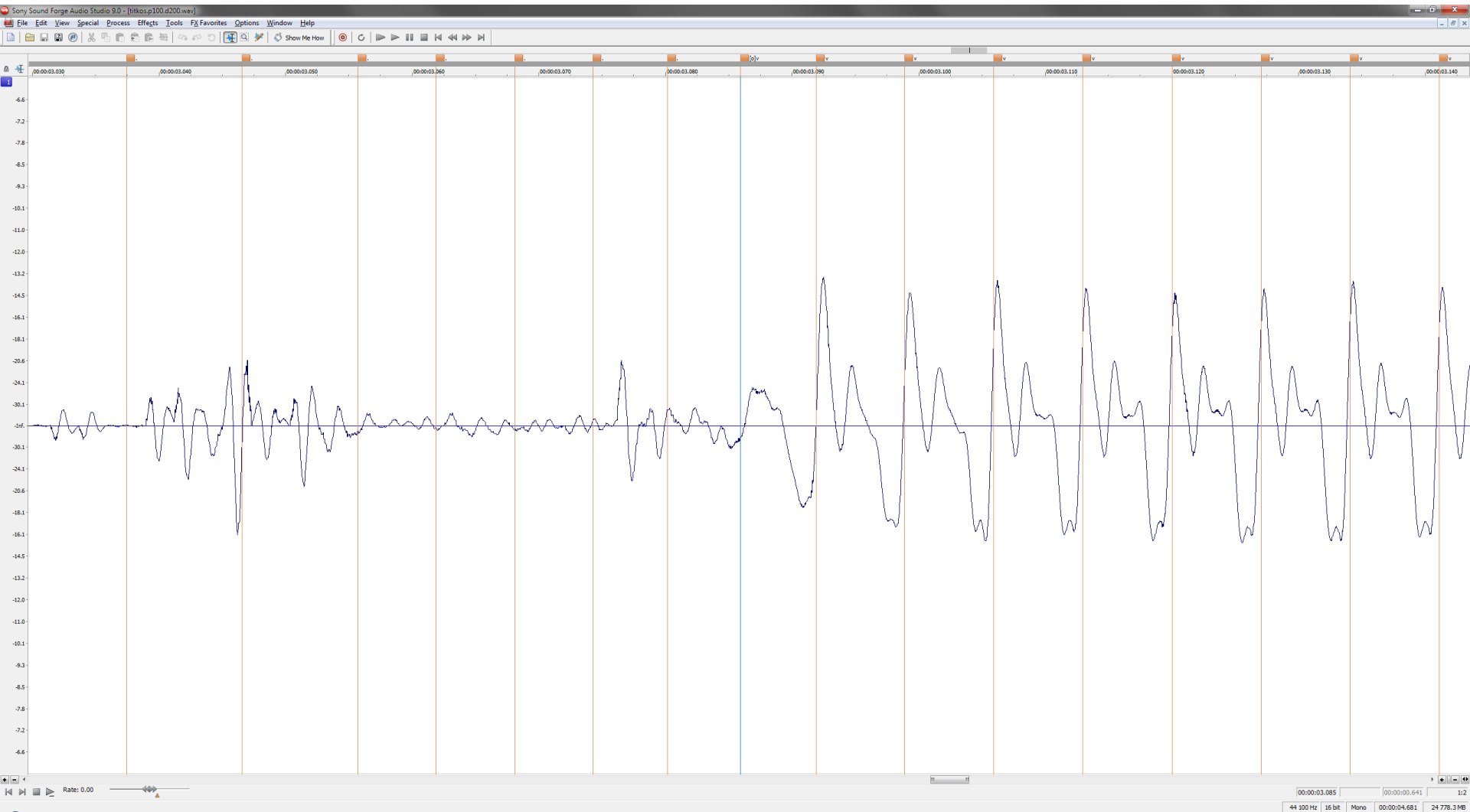
A hangidőtartam módosítása – csökkentett hangidőtartam



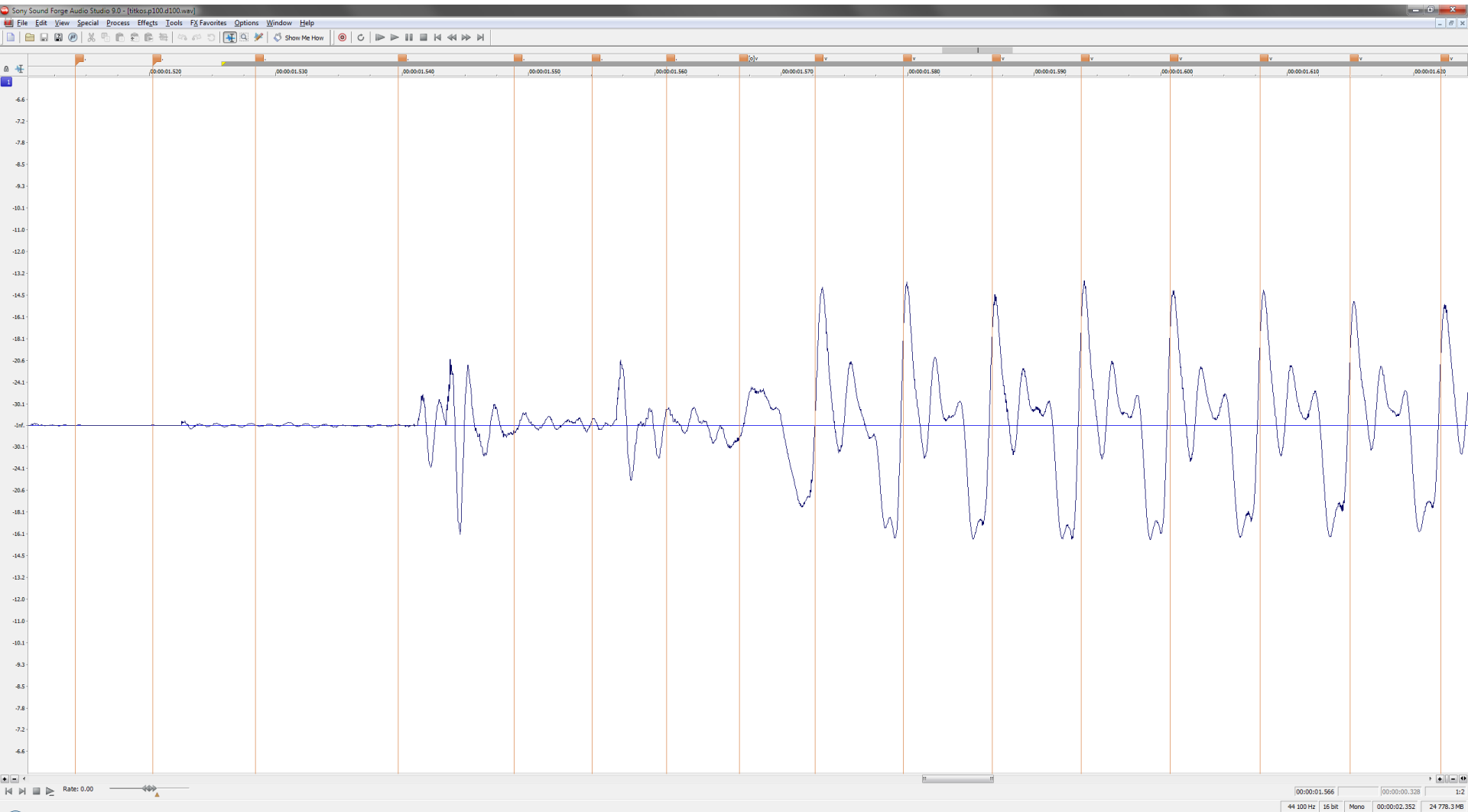
A hangidőtartam módosítása – eredeti beszédjel



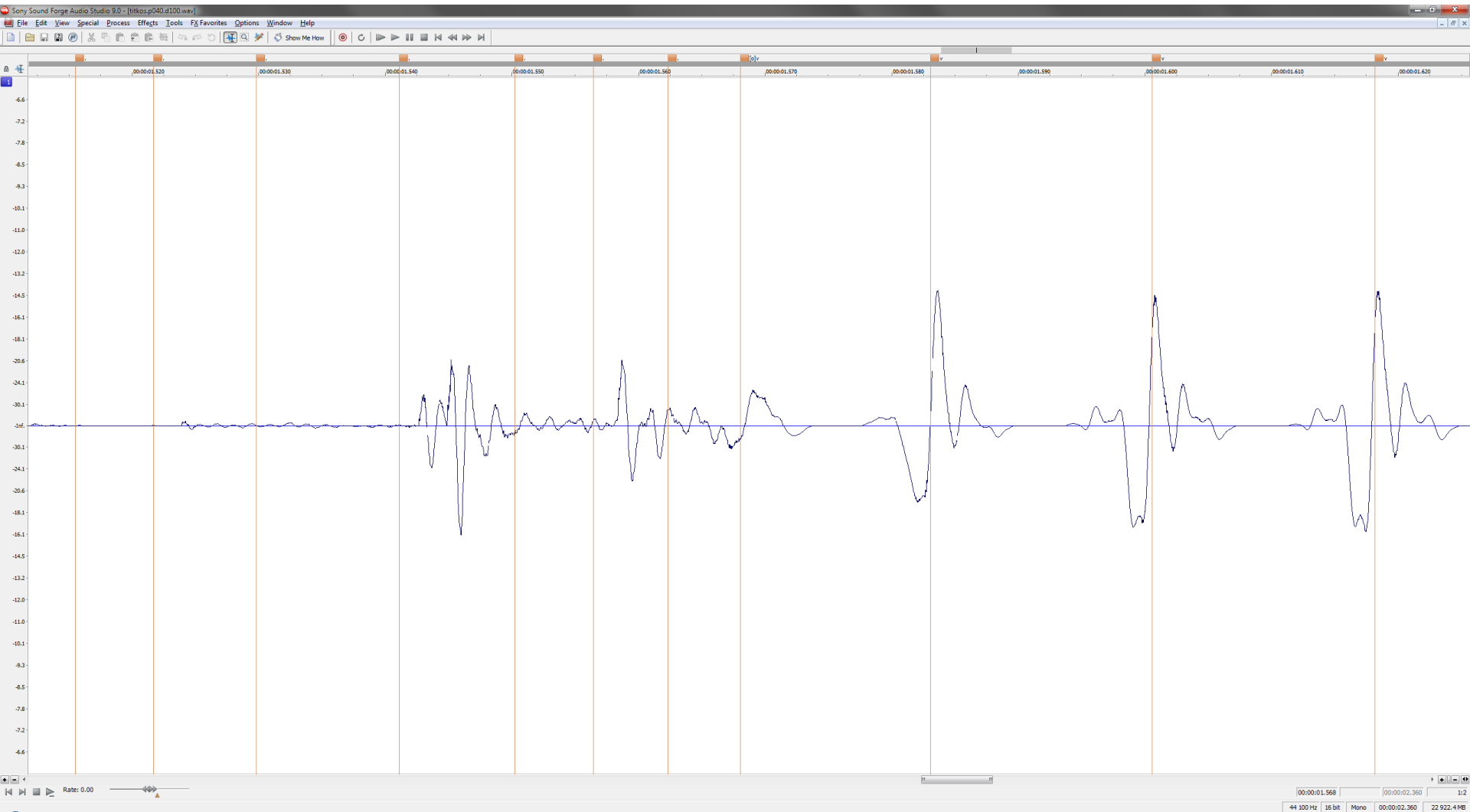
A hangidőtartam módosítása – növelt hangidőtartam



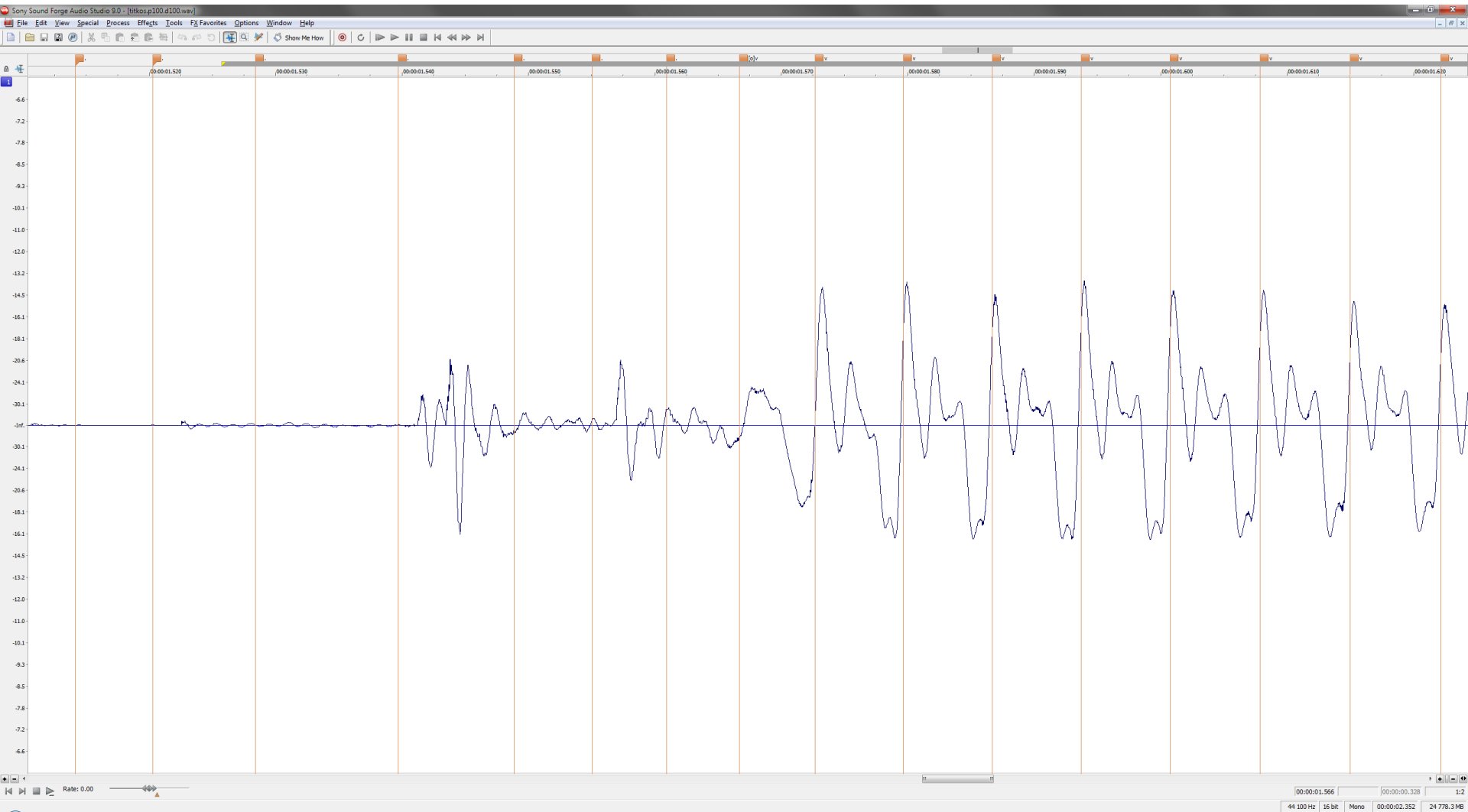
A hangidőtartam módosítása – eredeti beszédjel



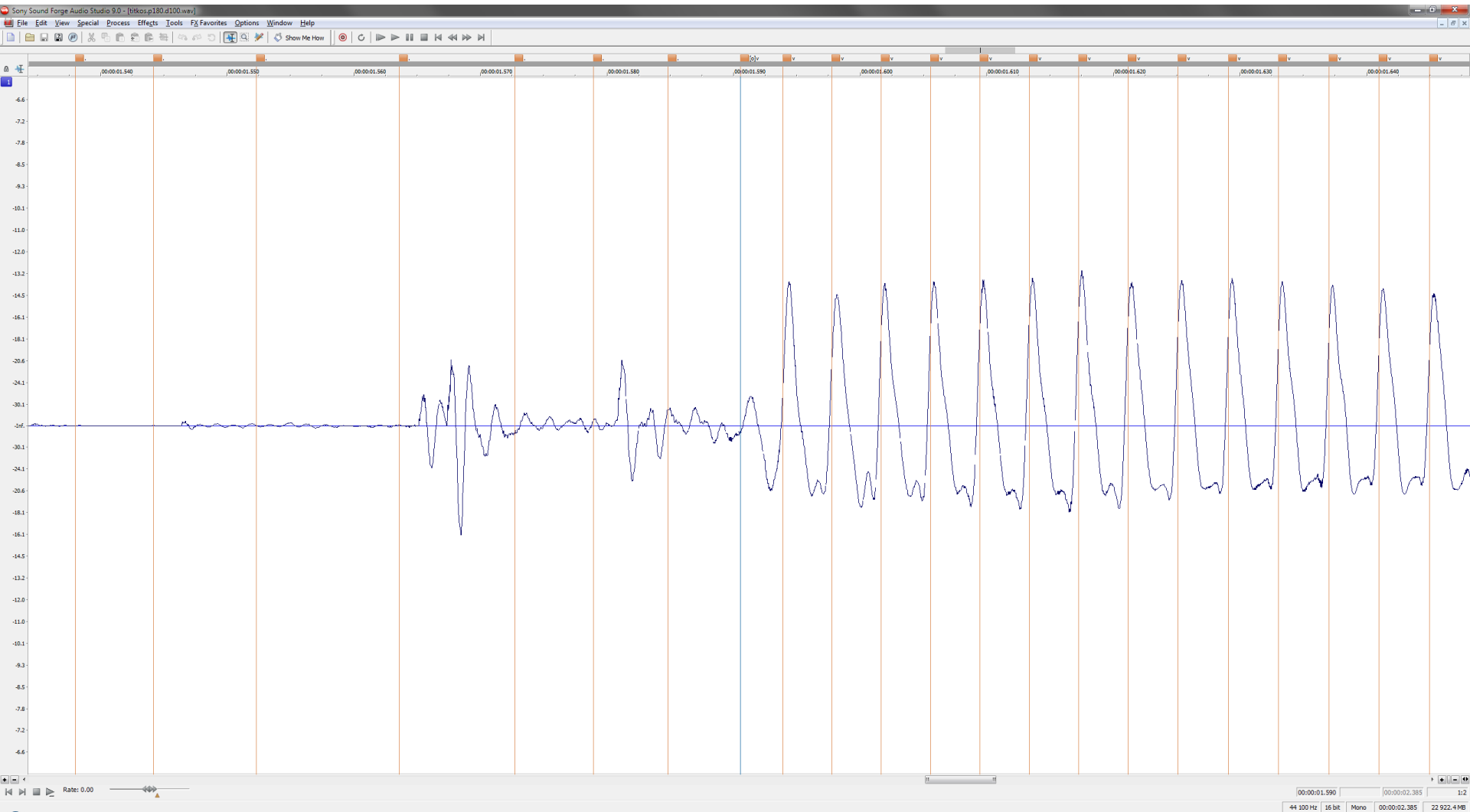
A hangidőtartam módosítása – csökkentett hangmagasság









A hangidőtartam módosítása – eredeti beszédjel



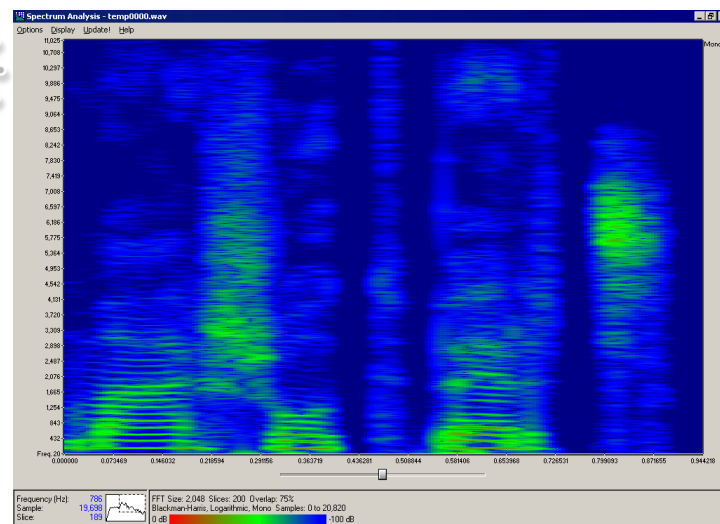
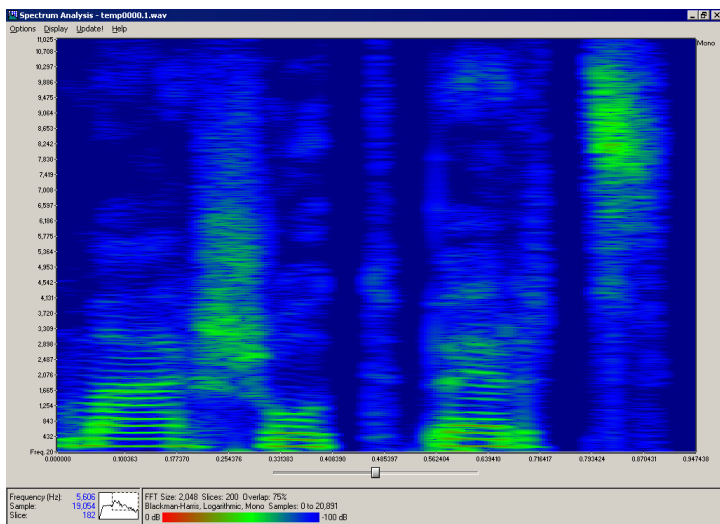
A hangidőtartam módosítása – növelt hangmagasság



- A hangidőtartamok, intonáció és intenzitás-változások átültetése egyik beszédjelről a másikra
 - csak ugyanarra a beszédhangsorra működik
- Természetes beszédjel, férfi beszélő: 
- Az eredeti mondat prozódiaja átültetve a szintetizált beszédjelre:
 - Azonos beszélő: 
 - Másik beszélő, szintén férfi: 
 - Női beszélő, időtartamok átültetve, és
 - hangmagasság változtatás nélkül: 
 - hangmagasság eltolva, és
 - eredeti hangmagasság-dinamikával 
 - női beszélőhöz igazított hangmagasság-dinamikával 

Spektrum-kompresszió telefonos beszédátvitel céljára





- A G.722 ITU-T standard szélessávú beszédkódoló felső sávkorlátja 7 kHz.
- A beszédjelben a réshangok és zár-réshangok tartalmaznak akár 10 kHz-es összetevőt is. Ezek az összetevők elvesznek a telefonos beszédátvitel során
- A spektrum-kompressziós módszer az érintett beszédhangok spektrumát összenyomja – az így módosított hang minden lényeges komponense visszaállítható a vevőoldalon.



- Jaws for Windows – képernyőolvasó szoftver vakok és gyengénlátók számára
 - Nagy mértékben gyorsítható a beszéd
 - 2018-ban is ez működik
- Robobrace fájl konverter (2018)

RoboBraille.org



- WESTEL SMS mondó 
- *Dömdödöm* kötött szótáros szövegfelolvasó 
- Banki számlainformációs rendszer 
- Általános szövegfelolvasó 



„Szóltam, meghallgattatok, értitek, ítéljetek.”

Lüsziasz
(i.e. 445–380)

