

Folytonos paraméterű vokóder rejtett Markov-modell alapú beszéd-szintézisben

- magyar nyelvű kísérletek 12 beszélővel

Csapó Tamás Gábor, Németh Géza
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
{csapot,nemeth}@tmit.bme.hu



1. Bevezetés

1.1 Rejtett Markov-modell alapú beszéd-szintézis (HMM-TTS)

- Egyik legkorszerűbb TTS technológia
- Statisztikai parametrikus beszéd-szintézis
- Flexibilis, kis erőforrás igényű
- Adott célbeszélőre adaptálható hang

1.2 Vokóderek HMM-TTS-ben

- Vokóder = parametrikus beszéd-kódoló
 - Analízis: beszéd → paraméterek
 - Szintézis: paraméterek → beszéd
- Vokóder típusok
 - Impulzus-zaj / kevert gerjesztés
 - Glottális forrás alapú
 - Harmonikus-zaj modell alapú
 - Maradékjel alapú
- Vokóder kihívások
 - Beszéd „zizegősségének” csökkentése
 - Alacsony számításigény
 - Beszédhangok egységes kezelése (pl. zöngés és zöngétlen)
 - Beszédminőségek kezelése (pl. glottalizáció, levegős beszéd)

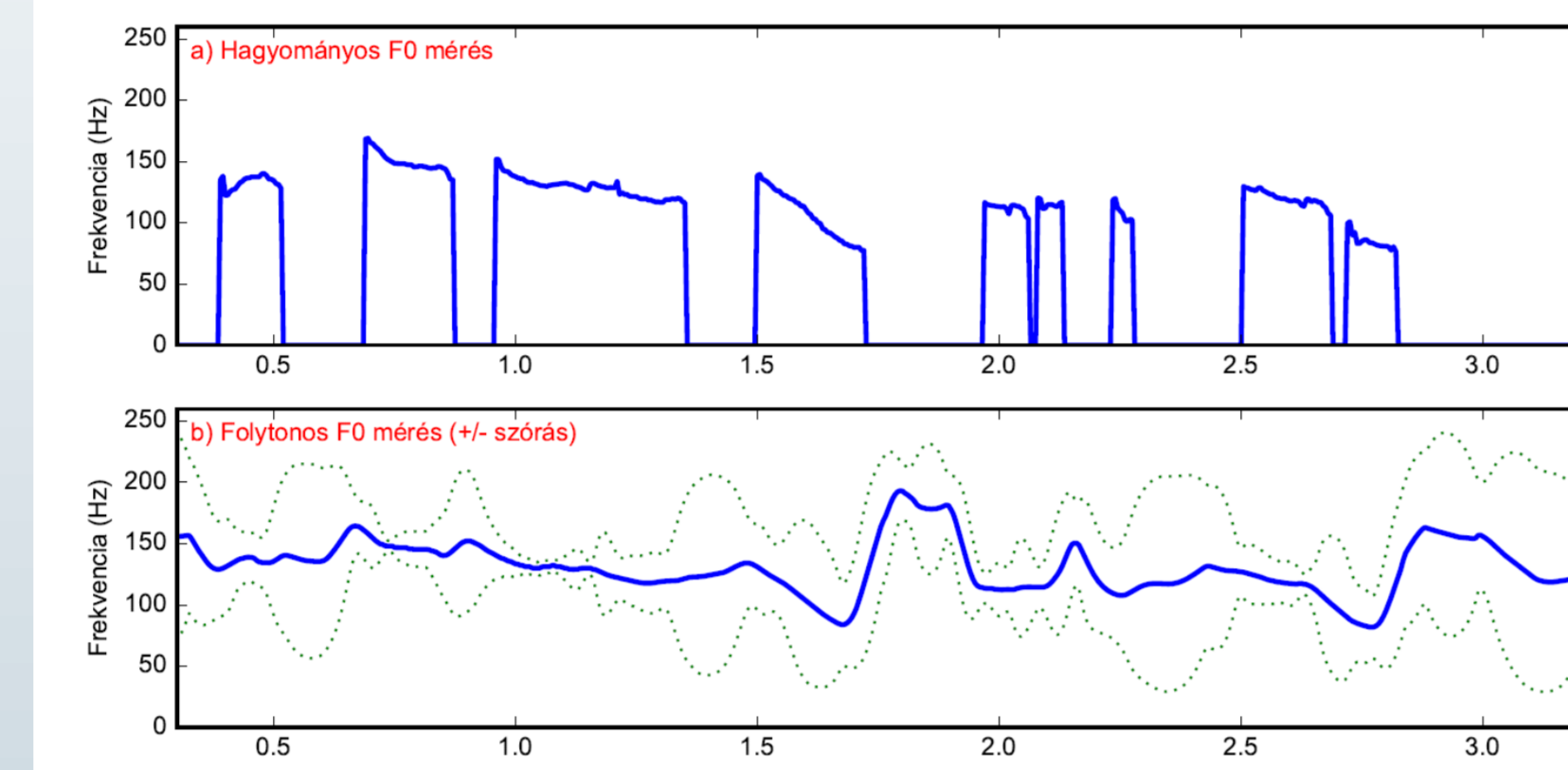
1.3 A jelen kutatás célja

- Alacsony komplexitású vokóder HMM-TTS-ben
 - Okostelefonon is gyors működés
- Folytonos paraméterek
 - Zöngés és zöngétlen beszédhangok modellezése nincs különválasztva

2. Módszerek

2.1 Analízis

- Folytonos alaphangfrekvencia detektor

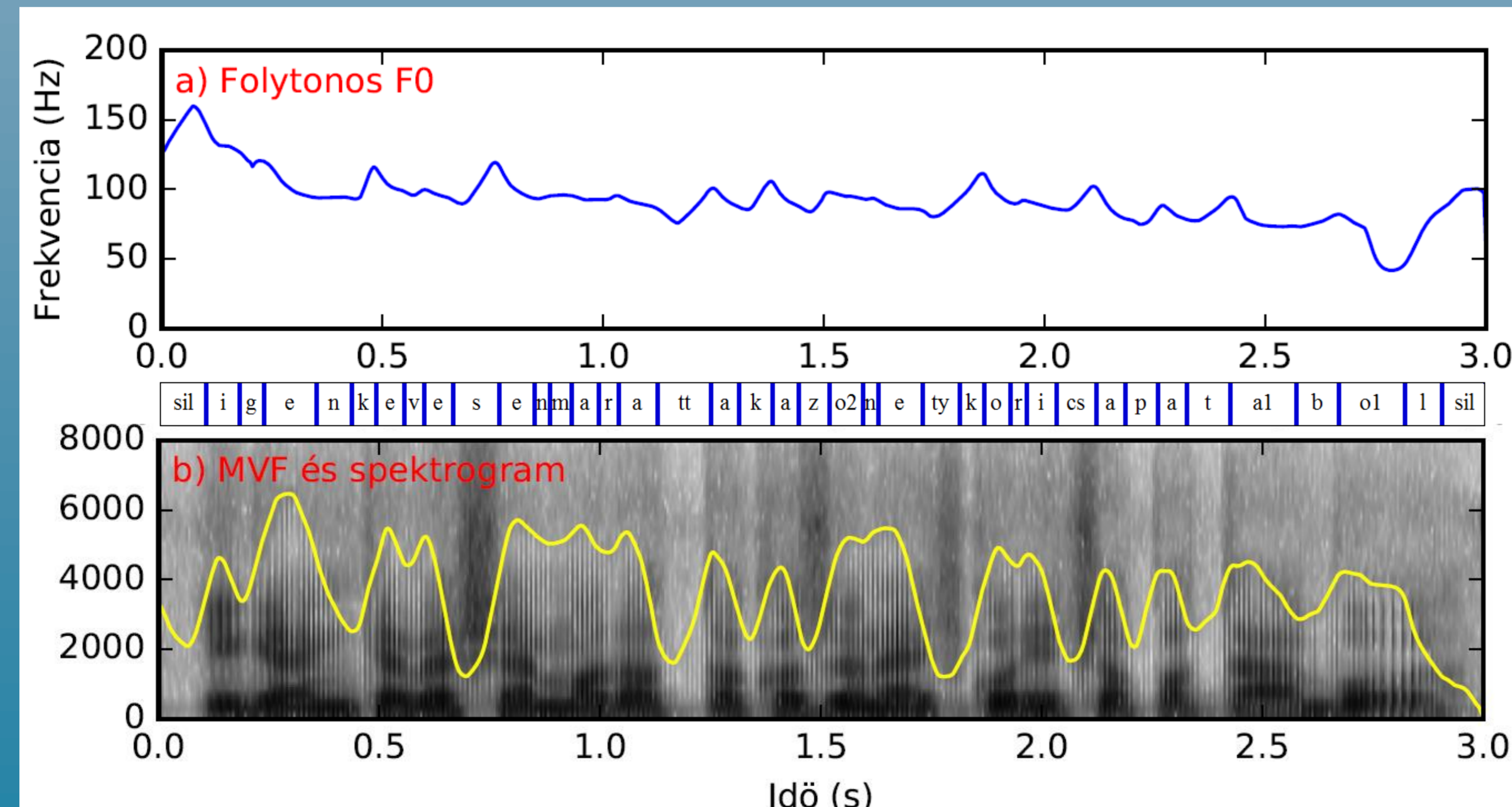


1. ábra: Az F0 mérés eredménye a) a Snack hagyományos F0 számító algoritmusmal, b) az SSP folytonos F0 számító algoritmusmal. A kék folytonos vonal az F0 kontúr, míg a zöld pontozott vonalak a +/- szórást jelölik.

- Maximális zöngességi frekvencia (MVF)
- Zöngés / zöngétlen frekvencia komponensek

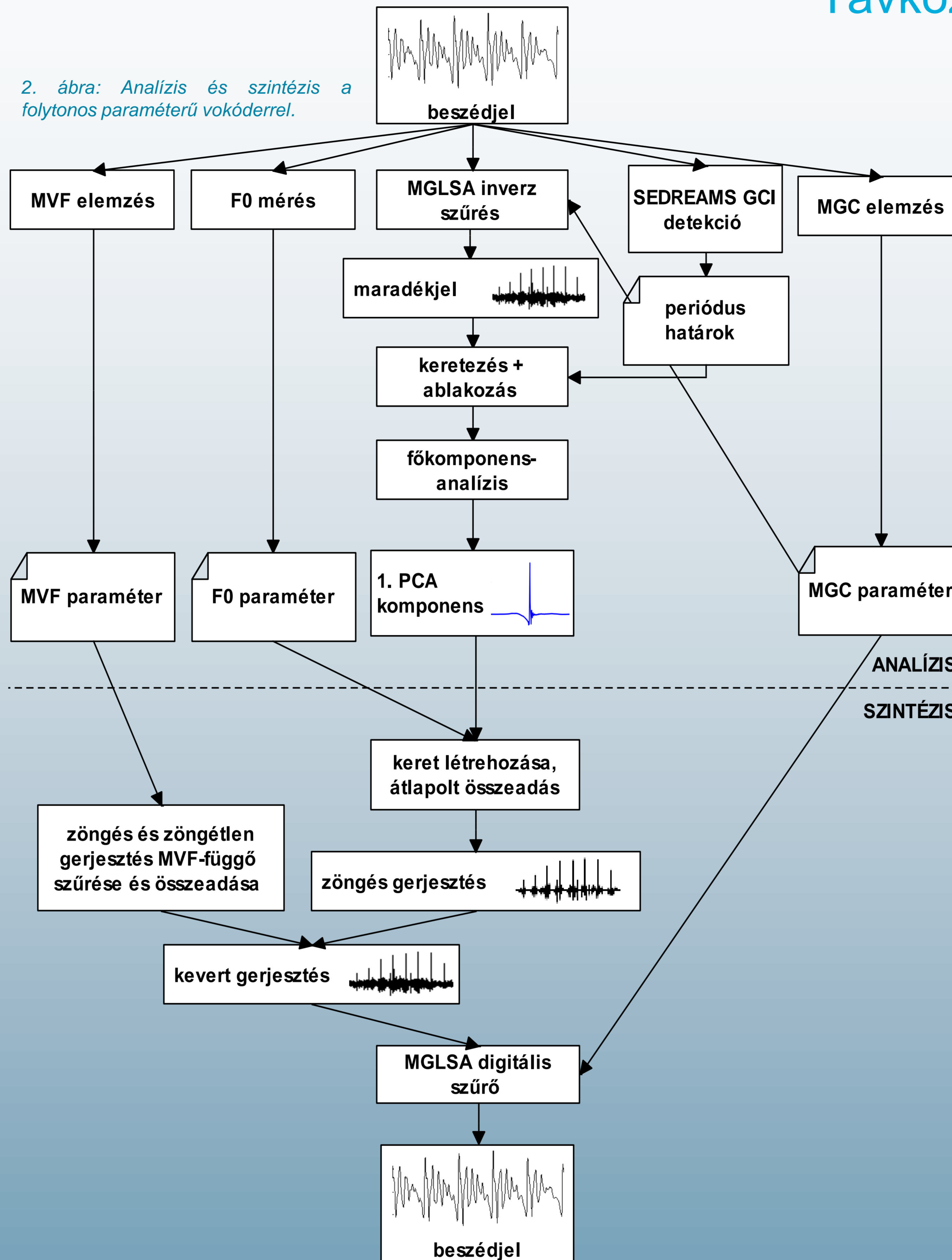
2.2 Szintézis

- Zöngétlen komponens:
 - fehérzaj, MVF feletti tartományban
- Zöngés komponens:
 - maradékjel 1. PCA komponens átlapoltan összeadva, MVF alatti tartományban



3. ábra: Szintetizált beszédminta egy férfi beszélőtől: „Igen kevesen maradtak az Ön egykori csapatából.”

2. ábra: Analízis és szintézis a folytonos paraméterű vokóderrel.



3.2 Átlaghang és beszélő adaptáció

- Átlaghang 10 beszélővel
- Átlaghang 5 férfi beszélővel
- Átlaghang 5 női beszélővel
- VoiceBank alkalmazás (Android)
 - Új beszélőtől hangminta gyűjtése
 - 50-50 mondat 5 beszélőtől
 - Beszélő adaptáció

3.3 Meghallgatásos teszt

- 12 beszélőtől és férfi / női átlaghangból 100-100 mondat szintetizálása
- Melyik férfi / női beszélő preferált?
- 5 tesztelő (beszédtechnológiai szakértők)

	Férfi beszélők						Női beszélők								
	1.	2.	3.	4.	5.	6.	7.	1.	2.	3.	4.	5.	6.	7.	
FF1	0	0	1	1	0	2	1	NŐ1	0	0	0	0	1	1	3
FF2	1	1	0	3	0	0	0	NŐ2	0	1	1	1	0	2	0
FF3	4	1	0	0	0	0	0	NŐ3	0	3	0	1	1	0	0
FF4	0	0	1	0	1	1	2	NŐ4	0	1	0	0	3	1	0
FF5	0	2	0	1	2	0	0	NŐ5	5	0	0	0	0	0	0
FF6	0	0	0	0	1	2	2	NŐ6	0	0	1	1	0	1	2
FF_átlag	0	1	3	0	1	0	0	NŐ_átlag	0	0	3	2	0	0	0

1. táblázat: A meghallgatásos teszt eredménye.

3.4 Androidos implementáció

- Közel valós időben működő HMM-TTS
- Folytonos paraméterű vokóder
- Természetesebb, mint az alaprendszer

4. Összefoglalás

- Természetesebb ember-gép kommunikáció
 - Gépi szövegfelolvasás korlátozott erőforrású eszközökön (pl. Android okostelefon)
- Beszéd-sérülteket segítő kommunikációs eszközök
 - Eredeti beszélőre emlékeztető beszédhang

3. Kísérletek és eredmények

3.1 Beszélőfüggő tanítás 12 beszélővel

- HTS-HUN rendszer kiegészítése
- PPBA adatbázis 6 férfi és 6 női beszélő
 - Beszélőnként 2 óra, kb. 2000 mondat

Főbb hivatkozások

Csapó, T.G., Németh, G., Cernak, M., Garner, P.N.: Modeling Unvoiced Sounds in Statistical Parametric Speech Synthesis with a Continuous Vocoder. In: EUSIPCO. pp. 1338–1342., Budapest, (2016).
Drugman, T., Stylianou, Y.: Maximum Voiced Frequency Estimation: Exploiting Amplitude and Phase Spectra. IEEE Sig. Process. Lett. 21, 1230–1234 (2014).

Garner, P.N., Cernak, M., Motlicek, P.: A simple continuous pitch estimation algorithm. IEEE Signal Process. Lett. 20, 102–105 (2013).
Tóth, B. P.: Rejtett Markov-modell alapú gépi beszéd-keltés. PhD disszertáció. BME TMIT (2013).