



**Budapesti Műszaki és Gazdaságtudományi Egyetem**  
Villamosmérnöki és Informatikai Kar  
Méréstechnika és Információs Rendszerek Tanszék

# Biometrikus felhasználó azonosítás

DIPLOMATERV

*Készítette*  
Boér Lehel

*Konzulens*  
dr. Zainkó Csaba

2019. április 2.

# Tartalomjegyzék

<b>Kivonat</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1. Háttérismeretek</b>	<b>1</b>
1.1. Biometrikus azonosítás . . . . .	1
1.2. Beszélőfelismerés . . . . .	2
1.3. Az automatikus beszélőfelismerés története . . . . .	2
1.3.1. Jellemző kinyerés . . . . .	3
1.3.2. Jellemző normalizálás . . . . .	4
1.3.3. Beszélő modellek . . . . .	4
1.4. Korábbi eredmények . . . . .	4
<b>2. Beszélőazonosító rendszerek napjainkban</b>	<b>5</b>
2.1. Beszédatadatbázisok . . . . .	5
2.1.1. TIMIT . . . . .	5
2.1.2. CMU Arctic . . . . .	6
2.1.3. Adatok előfeldolgozása . . . . .	6
2.2. Mérési elrendezés . . . . .	6
2.3. WaveNet classifier . . . . .	6
2.3.1. WaveNet . . . . .	6
2.3.1.1. Nyújtott kauzális konvolúció . . . . .	7
2.3.1.2. SoftMax eloszlás . . . . .	8
2.3.1.3. Reziduális blokkok . . . . .	9
2.3.2. Módosított WaveNet architektúra . . . . .	9
2.3.3. Eredmények . . . . .	10
2.4. SincNet . . . . .	10
2.4.1. Eredmények . . . . .	10
<b>Köszönetnyilvánítás</b>	<b>11</b>
<b>Irodalomjegyzék</b>	<b>12</b>

## HALLGATÓI NYILATKOZAT

Alulírott *Boér Lehel*, szigorló hallgató kijelentem, hogy ezt a diplomatervet meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy autentikált felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Budapest, 2019. április 2.

---

*Boér Lehel*  
hallgató

# Kivonat

Jelen dokumentum egy diplomaterv sablon, amely formai keretet ad a BME Villamosmérnöki és Informatikai Karán végző hallgatók által elkészítendő szakdolgozatnak és diplomatervnek. A sablon használata opcionális. Ez a sablon  $\text{\LaTeX}$  alapú, a *TeXLive*  $\text{\TeX}$ -implementációval és a PDF- $\text{\LaTeX}$  fordítóval működőképes.

# Abstract

This document is a L<sup>A</sup>T<sub>E</sub>X-based skeleton for BSc/MSc theses of students at the Electrical Engineering and Informatics Faculty, Budapest University of Technology and Economics. The usage of this skeleton is optional. It has been tested with the *TeXLive* T<sub>E</sub>X implementation, and it requires the PDF-L<sup>A</sup>T<sub>E</sub>X compiler.

# 1. fejezet

## Háttérismeretek

### 1.1. Biometrikus azonosítás

A biometria az emberek fizikai jellemzőinek mérésével és elemzésével foglalkozik. Alkalmazását tekintve három területet különböztetünk meg:

- Felhasználó ellenőrzés: Az azonosító rendszer a biometrikus adatot egy, korábban vetthez hasonlítja. Ez alapján dönt, hogy a felhasználó hozzáférhet-e a kívánt erőforráshoz. Ilyen egy ujjlenyomat-olvasóval ellátott mobiltelefonon a képernyőzár feloldása. A felhasználó ellenőrzés arra ad választ, hogy az illető az-e akinek mondja magát.
- Felhasználó azonosítás: Az azonosító rendszer a biometrikus adatot több korábban vett mintához hasonlítja és arra ad választ, hogy ki a felhasználó; azaz beletartozik-e a korábban eltárolt biometrikus adatokból álló csoportba vagy nem. Ilyen lehet például egy ujjlenyomat-leolvasóval ellátott beléptetőrendszer cégek esetében.
- Duplikátum detektálás: Annak ellenőrzése, hogy egy felhasználó egynél többször szerepel-e egy adatbázisban. Csalások, például szociális támogatást többször igénylők kiszűrésére használják.

Az első biometrikus azonosítási eljárás az ujjlenyomattvételen alapuló személyiség-azonosítás volt, amely a modern kriminalisztika világában terjedt el, de manapság már megtalálható okostelefonokban, biometrikus beléptetőrendszerekben is.

A biometrikus azonosítást az ún. biometrikus azonosító rendszer végzi el. A folyamat során a biometrikus azonosító rendszer mintát vesz az azonosítandó egyén egy vagy több előre meghatározott fizikai jellemzőjéről, és ezekről digitális lenyomatokat képez. Az első, regisztrációs fázisban a biometrikus minta lenyomatát a rendszer egy adatbázisban eltárolja, majd később az azonosítás során az aktuális mintát összeveti a korábban rögzítettel és dönt az egyezésről. Ahhoz, hogy az ember egy fizikai jellemzőjét biometrikus adatként használhassuk, a következő elvárásokat támasztjuk vele szemben:

- Általánosság: A biometrikus adattal minden egyénnek rendelkeznie kell.
- Egyediség: A biometrikus adatnak egyedinek kell lennie a releváns populáción belül.
- Állandóság: A biometrikus adat nem, vagy csak keveset változzon az idő elteltével.
- Mérhetőség: Az biometrikus adat az egyén részéről legyen könnyen mérhető testi adottság.

- Teljesítmény: A biometrikus azonosító rendszerek teljesítménye: gyorsaság, pontosság, technológia.
- Elfogadottság: A releváns populáción belül a mérési eljárás mennyire elfogadott (emberi méltóság megőrzése).
- Biztonság: Mennyire nehéz utánozni, hamisítani a biometrikus adatot?

A biometrikus adat lehet fiziológiai (DNS, arc, ujjlenyomat, írisz) vagy viselkedési (hang, írás, gesztusok). Mivel ezek az adatok statisztikai jellegűek, megbízhatóságuk változó. Minél több adat van egy mintában, annál egyedibb, és minél nagyobb a releváns populáció (eltárolt minták összessége), annál valószínűbb, hogy találunk két hasonló mintát. Ennek elkerülésére manapság terjednek a multimódusú biometrikus azonosító rendszerek, amelyek több biometrikus adatot felhasználva végzik ez az ellenőrzés, azonosítás és duplikátum detektálás feladatát.

## 1.2. Beszélőfelismerés

Az emberi kommunikáció során fontos feladat a beszélő partner felismerése. A telekommunikációs technológia fejlődése miatt elterjedt a telefonon vagy interneten történő hangalapú kommunikáció; a telefonos felhasználófelismerés mint biometrikus azonosítási módszer megjelent már banki alkalmazásokban, call centerekben és az elektronikus kereskedelemben is (mobiltelefonos vásárlás). Az elektronikus kommunikáció során sokszor csak a beszélő hangjára hagyatkozhatunk, az alapján ismerhetjük fel az illetőt. A beszélőfelismerést háromféle módon végezhetjük:

- Naiv beszélőfelismerés: Az emberi, naiv beszélőfelismerés során az ismerős hangokat meglepően nagy pontossággal ismerjük fel.
- Törvényszéki beszélőfelismerés: A törvényszéki szakértői vizsgálat eredménye.
- Automatikus beszélőfelismerés: A beszélőfelismerést számítógépes rendszer végzi.

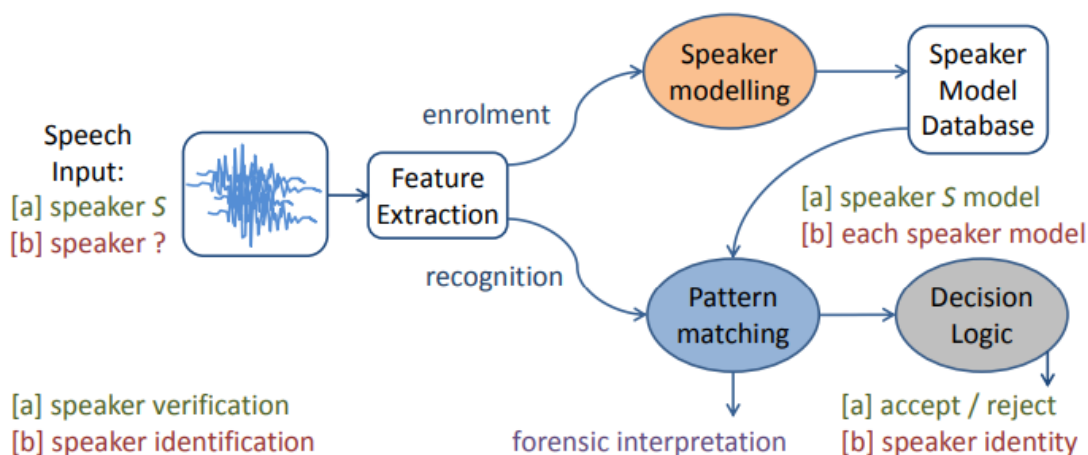
A beszélőfelismerés alatt három szűkebb fogalmat értünk. Ha a folyamat során az ismeretlen beszélőről azt ellenőrizzük, hogy az-e akinek állítja magát beszélő ellenőrzésről van szó. Beszélő szegmentáláskor a hangmintát homogén csoportokra bontjuk a beszélő személye alapján. Végül beszélő azonosításról beszélünk, ha az illető hangját rögzített hangok egy csoportjával vetjük össze és azt szeretnénk eldönteni, hogy melyikhez hasonlít a legjobban. Utóbbi felveti a kérdést, hogy mi történik ha a beszélő nem tagja a csoportnak.

Emiatt megkülönböztetjük a nyitott és zárt halmazú beszélőazonosítást. Utóbbi esetén csak olyan beszélőket ismerünk fel, akikről van hangminta az adatbázisban, míg az előbbinél ismeretlen beszélők is megjelenhetnek, így ezt is kezelni kell.

A beszélőfelismerés továbbá lehet szöveg-függő és szöveg-független attól függően, hogy a felismerő rendszer egy előre meghatározott mondatot vár, vagy bármilyen hangminta alapján működik.

## 1.3. Az automatikus beszélőfelismerés története

Az automatikus beszélőfelismerést egy számítógépes program végzi emberi beavatkozás nélkül. Az első automatikus beszéd felismerő rendszert a Texas Instruments fejlesztette és 1977-ben publikálták. A rendszer szövegfüggő beszélőellenőrzésre volt képes és az évek során a téves elutasítási és elfogadási rátája 1% alatt maradt. A hetvenes évek óta a



1.1. ábra. Automatikus beszélőfelismerő rendszer architektúrája.

beszélőazonosító és ellenőrző rendszerek rengeteget fejlődtek kezdve a vektor kvantálástól kezdve a GMM modelleken át a mély neurális hálókig.

Az automatikus beszélőfelismerő általános működését az alábbi (!) ábra szemlélteti. A beszélő-ellenőrzés és beszélőazonosítás során is az első lépés rögzített *jellemzők kinyerése*. Ezután az első, *tanító fázisban* referencia modelleket készítünk az egyes beszélőkhöz a jellemzők alapján, amelyeket eltároljuk a beszélő adatbázisban. Beszélő-ellenőrzés esetén ebben a fázisban egy küszöbérték is meghatározásra kerül. *Teszt fázisban* a rendszer kinyeri ugyanazokat a jellemzőket az aktuális hangmintából, majd *jellemző összehasonlítás* történik.

1. beszélő-ellenőrzés esetén megkeresi az ellenőrizendő személyhez tartozó modellt az adatbázisban és összehasonlítja az aktuális jellemzőkkel. Ha az eredmény a küszöbszint felett van, a rendszer egyezést mutat.
2. beszélőazonosítás esetén az aktuális jellemzőket az összes modellel összehasonlítja, majd a legjobb egyezés mellett dönt.

### 1.3.1. Jellemző kinyerés

A jellemző kinyerés célja a dimenzió csökkentése és a beszélőspecifikus információk kinyerése. Mivel a beszéd komplex jel, a beszélő azonosítása szempontjából felesleges információkat is hordoz. Ilyen például a környezet és a csatorna zaja. A kinyert jellemzőket a hangminta terjedelme alapján osztályozzuk. *Rövidtávú jellemzők* a 20-30 ms-os keretből kinyert mel-frekvenciás és lineáris prediktív kepsztrális együtthatók (MFCC és LPCC). A *prozódikus jellemzők* kinyerése 100 ms-os terjedelemben történik és a beszéd ritmusát, a hangmagasságot és a sebességet jellemzik. A *hosszútávú jellemzők* a jel akár perc hosszú kereteiből nyerjük ki. Ezek képesek reprezentálni a beszélő akcentusát illetve a szavak szemantikáját és az idiolektust.

A beszélőfelismerő rendszerek teljesítményét javította, ha a jellemzőket csak a hangminta azon részeiből nyerték ki, amikben beszéd is volt. Erre alkalmazott technika a *Voice Activation Detection (VAD)*.



### 1.3.2. Jellemző normalizálás

Jellemzőkinyerés során próbáljuk kiszűrni a beszélő szempontjából értékes részéket, ugyanakkor nincs tökéletes jellemző, amely ne változna a környezet hatására. Ezt a változást segítik minimalizálni a normalizálási módszerek.

### 1.3.3. Beszélő modellek

Kezdetben a vektor kvantálás volt az elterjed modellezési módszer, amit később a *Gaussian Mixture Model (GMM)* váltott fel. A GMM egy adathalmazt több normális eloszlás keverékeként ír le és képes nem felügyelt módon klaszterezni az adatokat. Egy beszélőhöz egy valószínűségi sűrűségfüggvényt rendel, amely különböző pontokban kiértékelve (például teszt fázisban a beszélőtől kinyert jellemzők) egy valószínűséget ad a két beszélő hasonlóságára.

A GMM megközelítés főleg beszélőazonosításra alkalmas. Beszélő-ellenőrzéshez szükség volt egy másik modellre is, ami képes leírni minden más beszélőt az ellenőrizendőn kívül. Erre adott megoldást az *Universal Background Model (UBM)*. Később jobb teljesítményt értek el, ha a teszt fázisban a beszélőkkel először UBM modelleket tanítottak és ezekből származtattak GMM-eket. Ezt nevezik GMM-UBM módszernek.

Mivel a tanító és teszt hangminták eltérő hosszúságúak lehetnek, szükség volt egy fix hosszúságú reprezentációra, ezt oldották meg a GMM szupervektorok, amelyeket az akkori megközelítés szerint szupport-vektor gépekkel vagy faktoranalízissel használtak.

Az utóbbi két módszer előnyeit kombinálva megszületett az *i – vektorok*, amelyet követve eljutunk a mai state-of-the-art módszerhez, a mély neurális hálózatokhoz (DNN).

## 1.4. Korábbi eredmények

Szerző (év)	Szervezet	Adatbázis	Módszer	Jellemzők	Hang típusa	Pontosság
Douglas A. Reynolds (1995)	Lincoln Laboratory	49	MFCC	rövid kifejezések	telefon	96.8 %
Rabah W. (2004)	King Abdulaziz University	20	SVD-alapú algoritmus	LPC/Cepstral	iroda	94 %
Yang Shao (2008)	Ohio State University	34	GFCCs	hallási jellemzők	telefon	~99.33 %
P. Krishnamoorthy (2011)	TIMIT	100	GMM-UBM	MFCC	labor	80 %
Alfredo Maesa (2012)	Voxforge.org	250	MFCC	spektrális jellemzők	beszéd-adatbázis	> 96 %
Sharada V. Chougule (2015)	Finolex Academy of Management and Technology	97	NDSF	spektrális	labor	~98-100 %

**1.1. táblázat.** Korábbi eredmények szövegfüggetlen beszélőazonosítás terén.

## 2. fejezet

# Beszélőazonosító rendszerek napjainkban

A fejezet bemutatja a tanítás során használt beszédadatbázisokat, két neurális hálózat alapú, zárt-halmazú, automatikus beszélőfelismerő rendszert; a *WaveNet* *classifiert* és a *SincNetet* illetve az ezekkel elért eredményeket.

### 2.1. Beszédadatbázisok

#### 2.1.1. TIMIT

A TIMIT beszédkorpuszt automatikus beszédfelismerő rendszerek fejlesztéséhez tervezték. 630 beszélőtől tartalmaz mintákat amerikai angol nyelven a 8 legelterjedtebb nyelvjárásban. A TIMIT archívum tartalmaz egy TRAIN és egy TEST mappát, ezek tanításhoz és teszteléshez valók. Ezeken belül további, a dialektusok sorszámaival (DR1, ..., DR8), azon belül a beszélő azonosítójával elnevezett könyvtárak találhatók. Egy beszélőhöz 10 db beszédminta tartozik 16 kHz-es NIST SPHERE fájlok formájában.

Dialektus régió (DR)	Férfi	Nő	Összesen
1	31 (63%)	18 (27%)	49 (8%)
2	71 (70%)	31 (30%)	102 (16%)
3	79 (67%)	23 (23%)	102 (16%)
4	69 (69%)	31 (31%)	100 (16%)
5	62 (63%)	36 (37%)	98 (16%)
6	30 (65%)	16 (35%)	46 (7%)
7	74 (74%)	26 (26%)	100 (16%)
8	22 (67%)	11 (33%)	33 (5%)

**2.1. táblázat.** A beszélők eloszlása dialektusok szerint.

A dialektus régiók a következők:

- DR1: New England
- DR2: Northern
- DR3: North Midland
- DR4: South Midland
- DR5: Southern
- DR6: New York City
- DR7: Western
- DR8: Army Brat (moved around)

A NIST SPHERE formátum az elején definiál egy fix hosszú fejléct, amit a hang bináris kódolása követ. Erre figyelni kell a hangfájlok beolvasásánál, Python esetében nem minden hangfeldolgozó könyvtár támogatja. Ilyen esetben kézzel el kell távolítani a fejléct a fájlok elejéről. A hangfájlokhoz tartozik egy szöveges dokumentum ami az elhangzott szöveget és annak a wav fájlbeli helyét tartalmazza. Továbbá egy WRD fájl írja le a szavakat és egy PHN kiterjesztésű a fonémákat, illetve azok időbeni elhelyezkedését a wav fájlban.

A hangfájlokhoz tartozó egyéb fájlok beszédfelismerés szempontjából fontosak. Mivel én a beszédkorpuszt beszélőfelismerésre használtam, csak a hangfájlokra volt szükségem. A TEST mappában - mivel a TIMIT-et alapvetően beszédfelismeréshez tervezték - teljesen különböző beszélők vannak a TRAIN mappához képest, ezért a TRAIN mappabeli beszélőket kell felosztani tanításhoz és teszteléshez.

### 2.1.2. CMU Arctic

A CMU Arctic beszédadatbázist beszéd-szintézis kutatásokhoz tervezték. 18 adathalmazt tartalmaz 18 különböző embertől más-más akcentussal, angol nyelven. Egy emberhez több száz beszédminta tartozik wav fájlok formájában.

### 2.1.3. Adatok előfeldolgozása

Az előfeldolgozó szkript a TIMIT adatbázis esetében a hangmintákról eltávolítja a NIST Sphere fejléct és a mondat előtti és utáni szüneteket. Ezután normalizálja a hangmintákat. A CMU Arctic esetében a hangminták alapból normalizálva vannak, ezért csak azonos méretűre kell vágni őket az egységes bemeneti dimenziók érdekében (ahogy a TIMIT esetében is).

## 2.2. Mérési elrendezés

Google Colab...

## 2.3. WaveNet classifier

A *WaveNet classifier* egy módosított WaveNet architektúra beszélőidentifikációhoz.

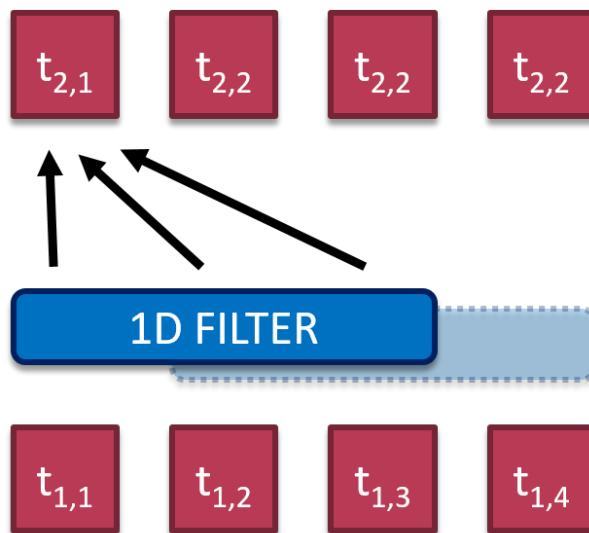
### 2.3.1. WaveNet

A WaveNet egy mély neurális hálózat audio hullámformák generálásához, amelyet a Google DeepMind publikált 2016-ban. Az ötletet az akkori felfedezések adták neurális autoregresszív generatív modellezés terén, amelyeket komplex eloszlások, például képek modellezésére használtak (van den Oord et al., 2016a;b). Ezt felhasználva audio hullámformák generálásában új eredményeket értek el.

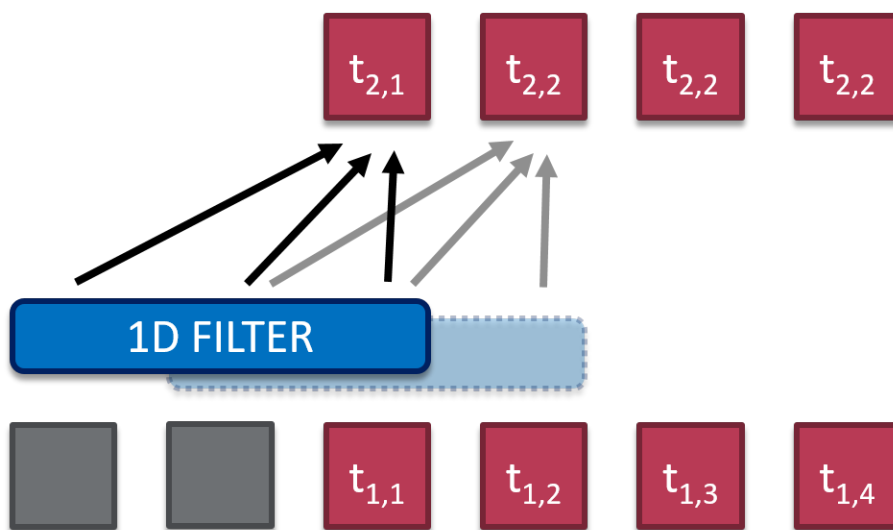
- Képes olyan természetes hangzású beszéd hullámformák generálására, amit korábban parametrikus vagy konkatenatív beszéd-szintézissel sosem értek el.
- Az audio hullámformák generáláshoz szükséges nagy receptív mezőt hatékonyan, nyújtott kauzális konvolúciókkal implementálja.
- Ha a modellt a beszélők identitásával tanítják, képes új hangok generálására.
- A zene generálás és a beszédfelismerés terén is ígéretesnek bizonyult.

### 2.3.1.1. Nyújtott kauzális konvolúció

A modell autoregresszív, vagyis a kimenete korábbi időpillanatokban felvett értékeitől függ. Ez azért fontos, mert a WaveNet egy generatív modell. A generált hullámforma  $t$ -edik időpillanatbeli értéke nem függhet jövőbeli értékektől. A kauzalitás legegyszerűbb implementációja ha legalább a  $kernel-1$  méretű paddinget adunk a konvolúcióhoz.



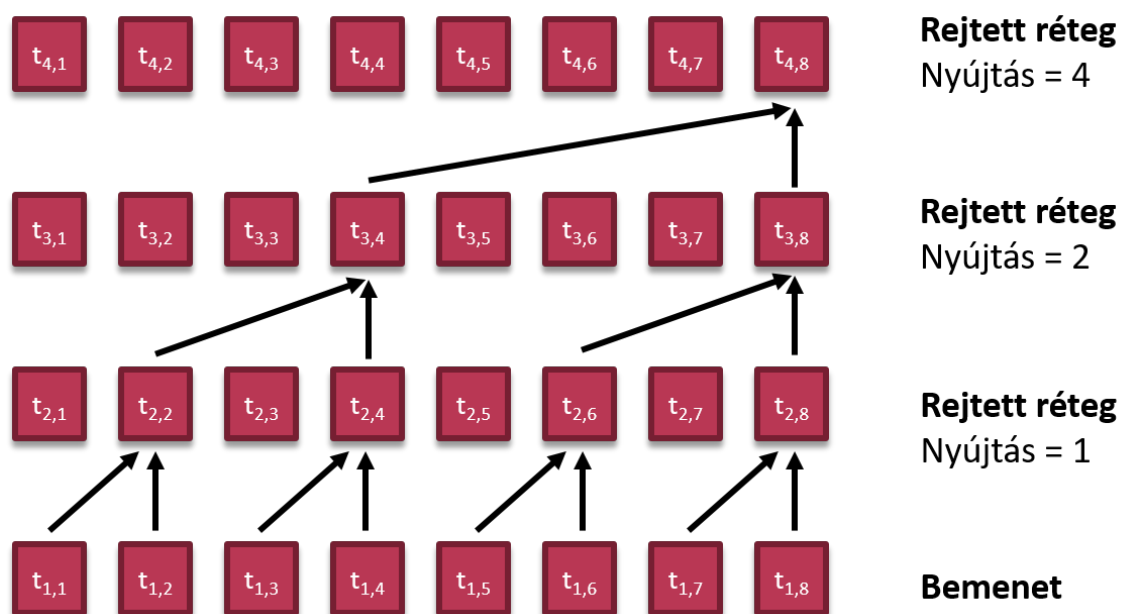
2.1. ábra. 1D nem kauzális konvolúció.



2.2. ábra. 1D kauzális konvolúció paddinggel.

A 2.1 ábra egy nem kauzális 1D konvolúciót mutat. A  $t_{i,j}$  az  $i$ -edik rétegbeli  $j$ -edik neuron. Látható, hogy a  $t_{2,1}$  jövőbeli időpillanatokból kap értékeket a szűrőn keresztül. Ennek megoldása a szűrő eltolása padding segítségével. Ezt szemlélteti a ?? ábra, ahol az egyes neuronok csak korábbi időpillanatokból kapnak értékeket.

A beszéd generálásánál a  $t$ -edik időpillanatban a hullámforma értéke a korábbi adatoktól függ. Ahhoz, hogy magas frekvenciájú, pl. 16 kHz frekvencián mintavételezett hangadattal tanítsuk a hálózatot nagy receptív mezőre van szükség. A receptív mező az a szélesség, amit a szűrő lát a bemenetből. 16 kHz esetén egy másodpercnnyi jelet 16000 szám reprezentál. Ahhoz, hogy a hálózat helyesen jósolja meg a következő generált értéket, a hosszú távú dependenciákat figyelembe kell vennie, tehát a receptív mező méretét elég nagyra kell megválasztani. A probléma ekkora mezők esetén, hogy sok konvolúciós réteget igényelnek (egy korábbi időpillanatbeli adat plusz egy konvolúciós réteget igényel), ami növeli a számítási komplexitást. A nyújtott konvolúciók erre adnak hatékony megoldást. A filter meghatározott távolságokkal kihagy valamennyi inputot, majd figyelembe vesz egyet. Egymás utáni rétegekben a nyújtási tényezőt exponenciálisan növelve a receptív mező is exponenciálisan fog nőni.



2.3. ábra. 1D nyújtott kauzális konvolúciós rétegek.

### 2.3.1.2. SoftMax eloszlás

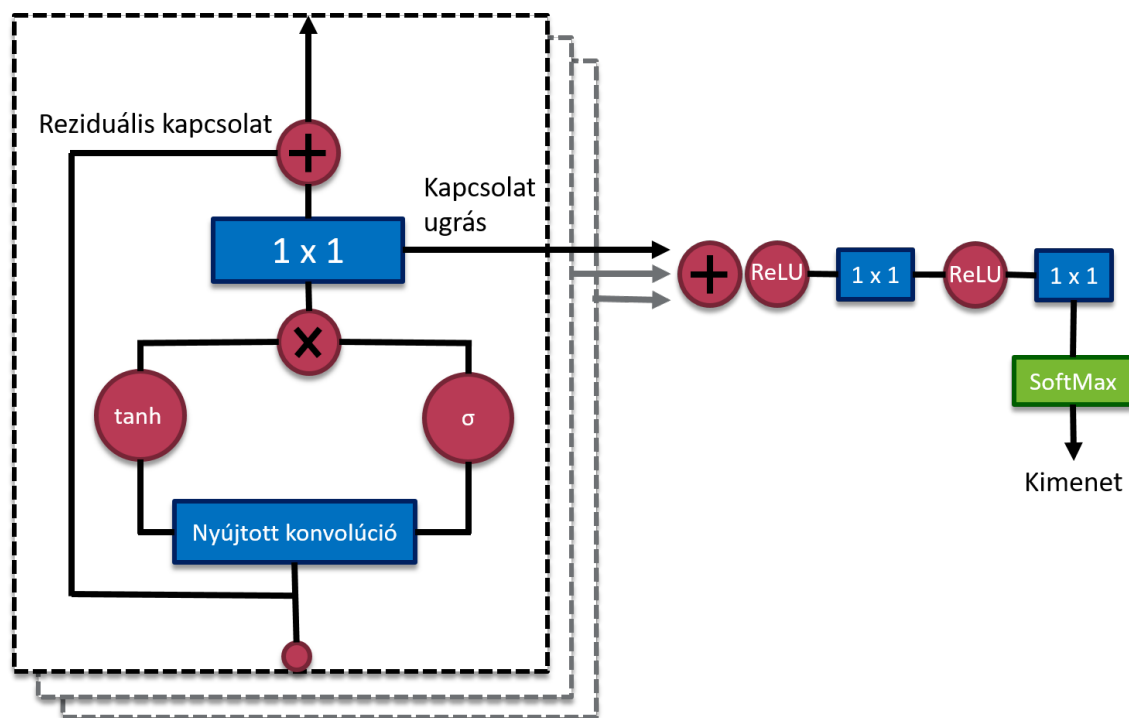
A WaveNet SoftMax réteget használ a  $p(x_t|x_1, \dots, x_t)$  feltételes valószínűségi eloszlás modellezésére. Az audio jeleket általában 16 bites egészekkel kódolják, amelyek 65536 értéket vehetnek fel. Ebben az esetben a SoftMax rétegnek 65536 valószínűséget kell kimenetként adnia, melyek összege 1. A  $\mu$ -law kvantálást alkalmazva a beszédjel 256 biten kódolható és később az inverz transzformációval jó minőségben visszaállítható.

Az emberi hallás sokkal érzékenyebb alacsony amplitúdójú hangok kvantálási zajára, mint a magasabbakéra. Erre alapozva a  $\mu$ -law kvantáló a jelet egy logaritmikus függvénnyel kvantálja úgy, hogy az alacsonyabb amplitúdójú jelek nagyobb felbontással (több bittel), a magasabbak pedig kisebbel lesznek kódolva. Ez növeli a SoftMax réteg hatékonyságát is, mert nagyobbak lesznek a valószínűségek közötti különbségek.

A tanítást a WaveNet klasszifikációs problémaként kezeli. A bemeneteket OneHot kódolással adjuk meg, a SoftMax réteg pedig az így kódolt egészekre ad valószínűségi eloszlást.

### 2.3.1.3. Reziduális blokkok

A WaveNet architektúra egymáshoz csatolt reziduális blokkokból és ún. kapcsolat-ugrásokból (skip-connection) épül föl. A reziduális hálózatok előnye, hogy orvosolják az elenyésző gradiens problémát, így sokkal mélyebb hálózat építhető.



2.4. ábra. WaveNet architektúra.

Egy reziduális blokk bemenete egy 2x1-es konvolúciós rétegen megy keresztül. Balra egy  $\tanh$ , jobbra egy szigmoid aktivációs függvényen haladnak át, majd elemenkénti szorzás és 1x1 konvolúció után egyrészt átugorja a reziduális kapcsolatot, illetve azzal együtt bemenetként szolgál a következő reziduális egységnek. Az 1x1 konvolúciós rétegek a dimenzionalitás változtatására szolgálnak. Külön 1x1 konvolúciós szűrők skálázzák a kimenetet a következő reziduális blokk bemenetére, és a kapcsolat-ugrásokhoz.

### 2.3.2. Módosított WaveNet architektúra

A módosított WaveNet architektúra segítségével a WaveNetet beszélőfelismerésre használhatjuk.

```
from WaveNetClassifier import WaveNetClassifier

wnc = WaveNetClassifier((96000,), (10,), kernel_size = 2, dilation_depth = 9,
                        n_filters = 40, task = 'classification')

wnc.fit(X_train, y_train, validation_data = (X_val, y_val), epochs = 100,
        batch_size = 32, optimizer='adam', save=True, save_dir='./')
```

```
y_pred = wnc.predict(X_test)
```

A WaveNetClassifier objektum paraméterei:

- *input\_shape*: Bemeneti dimenziók tuple formájában. Például ha a bemenet egy 6 s hosszú hullámforma 16 kHz-en mintavételezve, a bemeneti dimenziók (96000,)
- *output\_shape*: Kimeneti dimenziók tuple formájában. Például ha 100 osztály szerint klasszifikálunk, a kimeneti dimenziókból képzett tuple (100,).
- *kernel\_size*: A konvolúciós filter/kernel mérete a reziduális blokkokban.
- *dilation\_depth*: A reziduális blokkok száma.
- *n\_filters*: A konvolúciós filterek száma a reziduális blokkokban.
- *task*: Klasszifikáció vagy regresszió.
- *regression\_range*: A regresszió céltartománya lista vagy tuple formátumban.
- *load*: Előző WaveNetClassifier betöltése (bool).
- *load\_dir*: A betölteni kívánt modell könyvtára.

### 2.3.3. Eredmények

A WaveNet classifiert mindkét beszédatadtbázison teszteltem. A TIMIT beszédkorpusszal csak kevés beszélő esetén ért el jó eredményt, több mint 20 beszélő esetén a modell nem tanult. Ennek valószínűsített oka, hogy a TIMIT adatbázis beszélőnként 10 hangmintát tartalmaz.

<b>Beszélők száma</b>	18
<b>Minta/beszélő</b>	100
<b>Minta össz.</b>	1800
<b>Minta hossza</b>	4000
<b>Epochok száma</b>	43
<b>Hiba</b>	0.0013
<b>Pontosság</b>	1.0

**2.2. táblázat.** Paraméterek CMU Arctic adatbázissal.

!!! teszhalmaz mérete !!!

Tanítás után a teszt adathalmazon a hálózat 96.799 %-os pontosságot ért el.

## 2.4. SincNet

### 2.4.1. Eredmények

# Köszönetnyilvánítás

Ez nem kötelező, akár törölhető is. Ha a szerző szükségét érzi, itt lehet köszönetet nyilvánítani azoknak, akik hozzájárultak munkájukkal ahhoz, hogy a hallgató a szakdolgozatban vagy diplomamunkában leírt feladatokat sikeresen elvégezze. A konzulensnek való köszönetnyilvánítás sem kötelező, a konzulensnek hivatalosan is dolga, hogy a hallgatót konzultálja.



# Irodalomjegyzék

- [1] Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Kar: Diplomaterv portál (2011. február 26.). <http://diplomaterv.vik.bme.hu/>.
- [2] James C. Candy: Decimation for sigma delta modulation. 34. évf. (1986. 01) 1. sz., 72–76. p. DOI: **10.1109/TCOM.1986.1096432**.
- [3] Gábor Jeney: Hogyan néz ki egy igényes dokumentum? Néhány szóban az alapvető tipográfiai szabályokról, 2014. <http://www.mcl.hu/~jeneyg/kinezet.pdf>.
- [4] Peter Kiss: Adaptive digital compensation of analog circuit imperfections for cascaded delta-sigma analog-to-digital converters, 2000. 04.
- [5] Wai L. Lee–Charles G. Sodini: A topology for higher order interpolative coders. In *Proc. of the IEEE International Symposium on Circuits and Systems* (konferenciaanyag). 1987. 4-7 05., 459–462. p.
- [6] Alexey Mkrtychev: Models for the logic of proofs. In Sergei Adian–Anil Nerode (szerk.): *Logical Foundations of Computer Science*. Lecture Notes in Computer Science sorozat, 1234. köt. 1997, Springer Berlin Heidelberg, 266–275. p. ISBN 978-3-540-63045-6. URL [http://dx.doi.org/10.1007/3-540-63045-7\\_27](http://dx.doi.org/10.1007/3-540-63045-7_27).
- [7] Richard Schreier: *The Delta-Sigma Toolbox v5.2*. Oregon State University, 2000. 01. <http://www.mathworks.com/matlabcentral/fileexchange/>.
- [8] Ferenc Wettl–Gyula Mayer–Péter Szabó: *L<sup>A</sup>T<sub>E</sub>X kézikönyv*. 2004, Panem Könyvkiadó.