Beszélőazonosítás, hangbiometria rendvédelmi alkalmazása

Fejes Attila

Absztrakt

Cikkem a beszéd alapján történő személyazonosítás tudományos alapjait, módszertanát, rendvédelemben betöltött szerepét, gyakorlati alkalmazását mutatja be. Tárgyalom a szakterülethez kapcsolódó elméletet, amelynek alapos ismerete a beszélőazonosítás végrehajtásához elengedhetetlenül szükséges. Gyakorlati példákon keresztül ismertetem az akusztikai-fonetikai (félautomatikus) és a biometrikus (automatikus) azonosítási módszertant. Végül a beszélőazonosítás szélesebb körű alkalmazási lehetőségeiről írok, amelynek segítségével a nemzetbiztonsági tevékenység, a bűnüldözés hatékonysága növelhető lenne.

***Kulcsszavak**: beszélőazonosítás, hangbiometria*

Abstract

In my papers I describe the scientific basis of speaker recognition and methods, furthermore its part of the policing, and practical method. I present the paper knowledge; of which expert has to be having widely technical expertise. I introduce the acoustical-phonetical (semi-automatic), and voice biometrical (automatic) methods through practical examples. Finally I write a more widely application of speaker recognition, in which case it would be increase the efficiency of criminal investigation.

***Keywords**: speaker recognition, voice biometrics*

Bevezetés

Az elmúlt, közel másfél évtizedben - sok országhoz hasonlóan - Magyarországon is megtöbbszöröződött a különböző informatikai-kommunikációs eszközökkel lebonyolított beszélgetések, telefonhívások száma. Köszönhetően a technikai fejlődésnek, az eszközök sokrétűbbé válásának, a mindennapjaink részévé vált az a lehetőség, hogy a legváltozatosabb körülmények között készíthetünk kép- és hangfelvételeket egyre jobb minőségben, komolyabb adattárolási, és továbbítási problémák nélkül. Az információs társadalom kialakulása elhozta a bizonyos szempontból korlátlan beszédhang-rögzítés lehetőségét, a tudományos kutatások pedig egyre fejlettebb hangfeldolgozó és elemző rendszerek létrejöttét eredményezik. Mindezen folyamatok egyik következménye, hogy a nemzetbiztonsági, rendvédelmi szférában is egyre nagyobb mennyiségű (lehallgatott) beszédhang kerül rögzítésre. További kihatás, hogy a keletkezett hangfelvételek offline (vagy egyes esetekben online) elemzésének hatékonysága, így jelentősége is megnőtt, különösen a beszéd alapján történő személyazonosítás vonatkozásában. Szintén a fentiek következményeként vált lehetővé pl. a mobiltelefonhálózatok cellainformációiból felállítható kapcsolati háló elemzése, melyet a cikkemben terjedelmi, témaköri okok miatt nem tárgyalok, amely azonban bűncselekmények felderítéséhez nyújthat szintén jelentős mértékű segítséget.

Cikkemben röviden összefoglalom a beszéd alapján történő személyazonosítás tudományos alapjait, amelyhez betekintést nyújtok az emberi hangképzés és hangfeldolgozás ismereteibe, továbbá a fonetika tudományába, valamint bemutatom a beszélőazonosítás félautomatikus és automatikus (biometrikus) módszertanát és ez utóbbinak a nemzetbiztonsági, rendvédelmi szférában betöltött jelentőségét, jövőbeli felhasználási lehetőségeit.

A beszéd alapján történő személyazonosítás tudományos alapjai

Az emberi kommunikáció egyik eszköze a beszéd. Más megfogalmazásban a beszéd a nyelvhasználat hangzó formája, gondolataink megformálásának, továbbításának, valamint a gondolatok feldolgozásának eszköze[84]. Más szempontból a beszédhang humán biometriai jellemző, amely alapján lehetséges a beszélő egyedi azonosítása. Az azonosság valószínűségi fokának megállapításához a biometriát alkalmazó rendszerek matematikai-valószínűségi számítási eljárásokat

[84] *Gósy Mária: Fonetika, a beszéd tudománya, Osiris Kiadó, Budapest, 2004, pp. 11.*

alkalmaznak. (Mindezekről később, a Biometrikus beszélőazonosítás c. fejezetben részletesebben is szót ejtek.)

Ahhoz, hogy a beszélőazonosításról teljes képet kapjunk, meg kell határozni, hogy mi alapján lehetséges beszéd alapján személyt azonosítani. Legegyszerűbben fogalmazva azt mondhatjuk, hogy a beszélőazonosítás a hangképzés egyediségén, a beszédképzés sajátosságain, és a nyelvhasználat egyéni jellegzetességein alapul. Ebből a meghatározásból is látható, hogy a teljes hangszakértői módszertant alkalmazó azonosításhoz nem csak önmagában a beszédhang - mint fizikai rezgés - vizsgálata szükséges, hanem elemezni kell a beszélő által használt nyelv alkalmazásának jellemzőit is nyelvészeti szempontok figyelembe vételével.

Meg kell még említeni egy további tényezőt, ami az emberi halló- és hangfeldolgozást végző szervekhez köthető szubjektív - mérési adatokkal le nem írható -, összetevő, amelyet neurális spektrogramnak[85] nevezünk, és amely a beszédészleléssel van összefüggésben. A neurális spektrogram a beszéd akusztikai jellemzőinek agyban kódolt megfelelője, ami alapján az ember a beszélőt beazonosítja, vagy a különbözőséget detektálja. Ezt a képességünket használjuk a félautomatikus beszélőazonosításban a percepciós elemzés során – és természetesen a mindennapi életben is. Látható mindezek alapján, hogy a beszélőazonosítás módszertana tartalmaz „szubjektív" elemet is, amelynek a súlyozása, az eredmények objektív számértékekkel való összevetése nagy gyakorlatot kíván. További nehézség, hogy a beszédhang kvázi-stacionárius jellegű, amely azt jelenti, hogy az akusztikai (fizikai) paraméterek csak nagyon rövid időtávon tekinthetők állandónak, így a beszédhang – a kézíráshoz hasonlóan – nagy variabilitással rendelkezik, ezért a méréseknél mindig átlagos értékek összevetése történik. Ahhoz, hogy hangtechnikai szakértői szinten képesek legyünk személyazonosítást végezni, az eredményeket hitelesen alátámasztani, ismernünk kell a beszéd képzését, szerkezetét, feldolgozását, továbbá rendelkeznünk kell fonetikai ismeretekkel, amelyeket a következő fejezetben tárgyalok.

Beszédképzés, hangfeldolgozás, fonetika

A beszédképzésben négy fő szervünk vesz részt: tüdő, gége, toldalékcső, valamint az agy, annak is a Broca-terület elnevezésű része. Beszédlégzéskor (amely funkcióját tekintve abban különbözik az élettani légzéstől, hogy a célja nem kizárólag az oxigéncsere) a tüdő a légzőizmok működésével biztosítja a kiáramló le-

[85] *Uo. pp. 275.*

vegőt, amely a légcsőn keresztül eljut a gégéhez és a hangszalagokhoz. A kiáramló levegő megrezegteti a hangszalagokat, így létrejön a hangképzés alapja, a fonáció (zöngeképzés). Az így keletkező akusztikai energia (zönge) áthalad a toldalékcsövön (vokális traktus), majd eljut a szájnyílásig, melynek mozgásával, az ajkak és a nyelv közreműködésével, létrejön a beszéd. A vokális traktust három üreg alkotja: a garat, a szájüreg és az orrüreg, amely szervek rezonátorlánckénti működnek, és a személyenként eltérő fizikai-méretbeli kialakításuk miatt olyan egyéni sajátosságokat hoznak létre a beszédhangban, amely alapján lehetséges egy beszélőt azonosítani.

A hangfeldolgozásban részt vevő szervek anatómiailag két fő részre bonthatóak: perifériás és centrális rész. A perifériás rész a fül, a centrális rész a hallóidegből, a hallópályából és az agy azon részeiből áll, melyek az akusztikai ingerek feldolgozásában vesznek részt. A hallás során a környezet akusztikai ingerei a hallójáraton keresztül eljutnak a dobhártyára, amely mechanikai rezgéssé transzformálja a levegőmolekulák rezgését. A mechanikai rezgés a hallócsontocskák (amelyek egy négyzetcentiméteren tartalmazzák az emberi szervezet három legkisebb csontját: kalapács, üllő, kengyel) segítségével alakul át folyadékrezgéssé, amely a csigában található Corti-szerv szőrsejtjei révén elektromos impulzusok formájában idegi ingerületi alakot vesz fel. A neuron sejtekkel ez az ingerület jut el az agyba, a Wernicke-területre, ahol a beszédhang, a beszéd szemantikai tartalma feldolgozódik.

A fonetika a beszéd sajátosságaival foglalkozó interdiszciplináris tudományág, melynek célkitűzése az, hogy a beszéd objektív leírását adja. Ez azt jelenti, hogy a beszédet az artikuláció, az akusztikum és a percepció hármas egységében jellemzi. A beszéd képzését a fonetikában két részre oszthatjuk: szegmentális és szupraszegmentális hangszerkezetre, melyeket ugyanazon szervekkel hozunk létre. A szegmentális hangszerkezeten értjük a beszédhangokat, hangkapcsolatokat és hangsorokat. A szegmentális szerkezeten belül a beszédhangokat magánhangzókra (vokálisok) és mássalhangzókra (konszonánsok) oszthatjuk fel, amelyek között fonetikai szempontból az alapvető különbség az, hogy a magánhangzók képzésekor nincs akadály a szájüregben. A szupraszegmentális hangszerkezeten (prozódia) a beszéddallamot, ritmust, tempót, szünetet, hangerőt és hangszínezetet értjük. Azt is mondhatjuk, hogy a szegmentális hangszerkezetre "ráépülnek" az azzal egyidejűleg képzett szupraszegmentumok, melyek legfőbb funkciója a beszéden belül az érzelmi, értelmi tagolás, jelentésbeli megkülönböztetés. Például, szó szerint egyező mondatok esetében, ha a vesszővel jelzett szüneteket máshol tartja a beszélő, és a hangsúlyokat is eltérően alkalmazza, a jelentés is teljesen eltérő lehet, mint ahogy a neves idézet is mutatja:

„A királynét megölni nem kell, félnetek jó lesz, ha mindenki egyetért, én nem, ellenzem.

A királynét megölni nem kell félnetek, jó lesz, ha mindenki egyetért, én nem ellenzem.”

A félautomatikus beszélőazonosítási módszertan

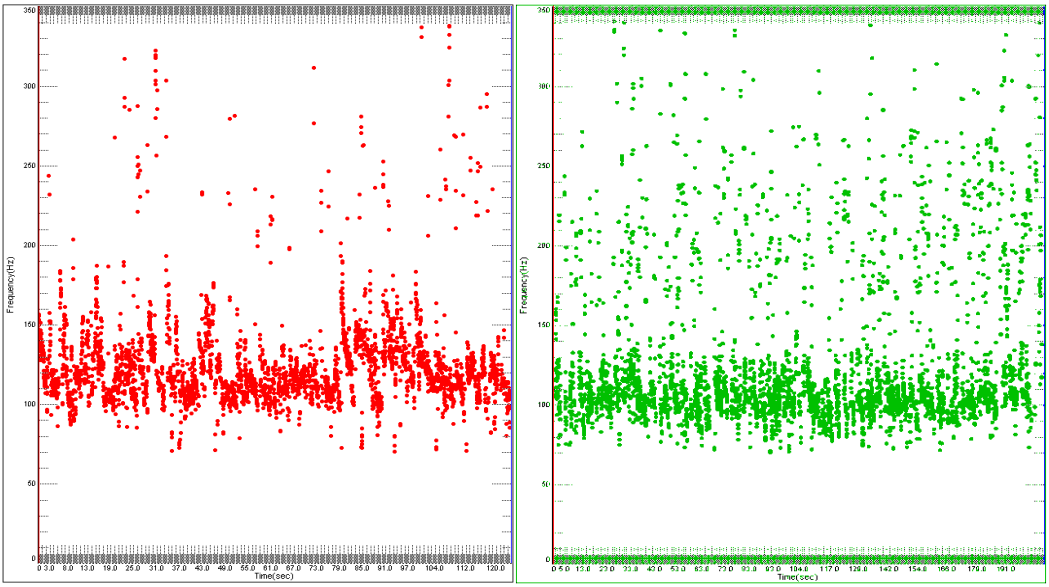
Az akusztikai-fonetikai alapú – félautomatikus – beszélőazonosítási módszertan percepciós elemzésből, és szoftveres hanganalitikai mérésekből tevődik össze. A percepciós vizsgálat a beszéd egyedi jellegzetességeinek észleléses úton történő feltárását jelenti, a számítógéppel végrehajtott elemzések pedig a hang akusztikai tulajdonságait számszerűsítik.

A hangképzés percepciós vizsgálata során legtöbbször az alábbi jellegzetességek detektálhatóak (előfordulásuk esetén): beszédhangok zártabb, nyíltabb ejtése, r, s, sz hangok képzési hibái (gyakori), diftongusok (kettőshangzók), tájnyelvi dialektus a beszédben, beszédtempó, dallamvezetés, hezitálások és azok tulajdonságai. A percepciós elemzésen belül nyelvészeti szempontú analízist is elvégzünk, melynek során leggyakrabban az alábbi tényezőket vizsgáljuk: a beszélőre jellemző tipikus szavak, szókapcsolatok, kifejezések, azok ismétlése, nyelvhelyesség a grammatikai szabályok alapján, mondatok alkotása, szóösszetétele (egyszerű-összetett), töltelékszavak alkalmazása (izé, ja, hát). Mindezen megállapítások objektív alátámasztásához nem férhet kétség, például az r hang hibás kiejtéséből adódó raccsolás (rotacismus) kimutatható a spektrografikus képen a perdület vizsgálatával, a diftongus a koartikuláció szintén spektrografikus képen történő elemzésével. A percepciós elemzésen belül megállapításokat tesz a hangszakértő a hangszínezetre, hangfekvésre, ezek alapján a beszélők hasonlóságára is. Ez utóbbi megállapítás látszólag hordoz magában szubjektivitást, hiszen mérési adatokkal ez esetben nem tudja a hallgató alátámasztani, hogy miért találja a beszélőket hasonlónak, vagy különbözőnek, azonban a neurális spektrogram létével összefüggésben kijelenthetjük, hogy az ily módon tett szakértői megállapítások is megalapozottnak tekinthetőek.

A szoftveres hanganalitikai mérések során a szegmentális és a szupraszegmentális jellemzőket egyaránt számítógépes vizsgálatoknak, méréseknek vetjük alá. Nagyon fontos tényező, hogy a beszédhang variabilitásából következően nem a megegyező, vagy eltérő mérési eredményekből, hanem a szignifikáns hasonlóságok, vagy különbözőségek alapján vonunk le következtetést. Például, mivel az alaphang-érték egy folyamatosan változó tényező, nem egy konkrét időpil-

lanatban mért adat, hanem az átlagos alaphang-érték a meghatározó. A szoftveres mérések, vizsgálatok során leggyakrabban az alábbi eljárásokat alkalmazzuk.

A beszéd átlagos alaphang-értékének meghatározása.: Jó minőségű felvételeken megbízható módon mérhető adat, amely egészséges embereknél a korral együtt változik, azonban tudományos kutatások alapján elmondható, hogy felnőtteknél kb. 10 éves időtávon belül megközelítőleg állandó. Vizsgálata során figyelembe kell venni, hogy egyes kommunikációs szituációkban, más-más érzelmi állapotban nagymértékben változik. A következő ábrákon egy beszélőnek két időpontban rögzített hangja alapján elkészített Pitch contour[86] ábrázolása látható, amely az alaphang értékeket mutatja a beszéd egy hosszabb szakaszán.



1. ábra: Pitch contour ábrázolás azonos beszélők hangja alapján

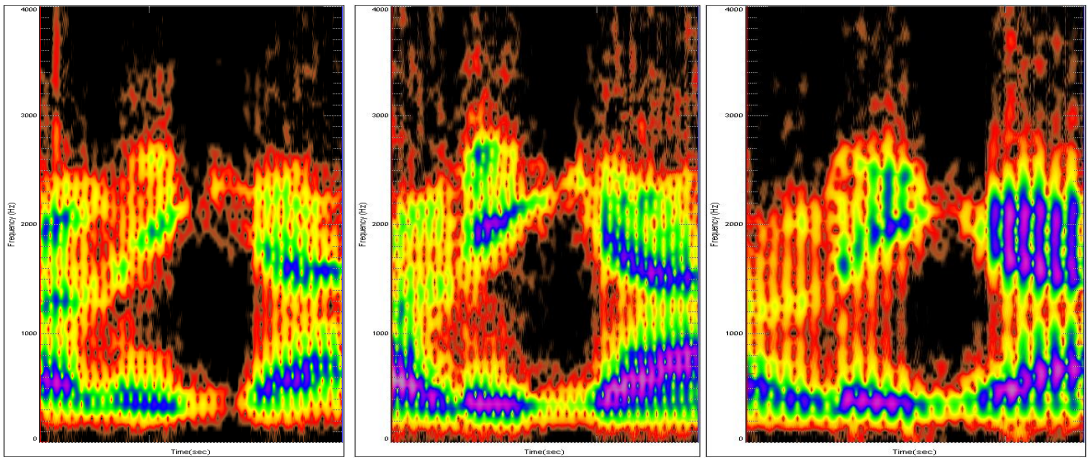
A két ábrán látható, hogy legnagyobb számban kb. 110 Hz körül szóródnak mindkét képen a különböző időpontokban mért alaphang-értékek, és az is egyértelműen detektálható, hogy a szórás értéke is nagyon közel van egymáshoz.

Időszerkezeti elemzések (a magán- és mássalhangzók hossza, néma fázis, zárfelpattanási idő, zöngekezdési idő): A magán-és mássalhangzók tiszta fázisa hosz-

[86] *Pitch Contour (alaphang-értékek görbéje)*

szának mérése csak jó minőségű hanganyagon célravezető, mert különben nem hiteles eredmények születnek. Szükséges még, hogy a beszélő jó hangképzési, artikulációs készséggel rendelkezzen, valamint, hogy megfelelő mennyiségű (több percnyi) hanganyag álljon rendelkezésre.

Spektrogram (hangszínkép)-elemzés: A spektrografikus kép (SPG) a beszédhangot frekvencia - idő tartományba konvertálva mutatja a hang intenzitás értékét, így lehetővé válik a hangok három dimenzióban történő vizsgálata, amely a formánsképek elemezéséhez nélkülözhetetlen feltétel. A formánsok a zöngének a rezonátorüregekben (garatüreg, szájüreg, orrüreg) felerősödött felharmonikusai, másképpen fogalmazva, különböző, az adott hangra jellemző frekvenciákon kialakuló energiakoncentrációk, melyek alakzata egyénre jellemző. Egyező hangok, hangkapcsolatok formánsai, alakjuk, elhelyezkedésük, frekvenciaértékük nagyfokú hasonlóságot mutatnak azonos beszélők esetében, míg különböző beszélő személyeknél a formánsstruktúrában szignifikáns eltérések észlelhetőek. A 2. ábrán egy beszélő különböző időpontokban rögzített telefonbeszélgetéseiből származó ugyanazon hangkapcsolat alapján elkészített SPG-képei láthatóak.



2. ábra Spektrografikus (SPG) képek azonos beszélők hangja alapján

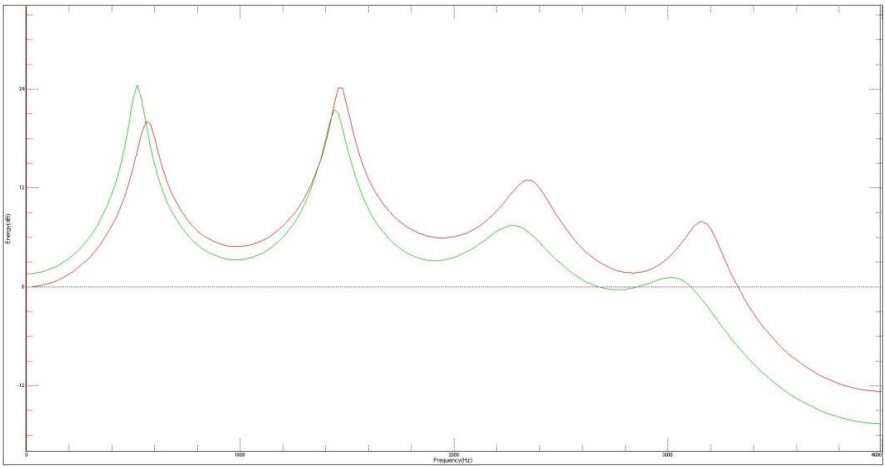
Látható, hogy a sötétkék-lila színekkel jelzett formánsstruktúra - a beszéd variabilitása következtében - nem egyezik meg teljesen, azonban az energiagócpontok alakzata, frekvencia- és időértékei nagymértékben hasonlítanak egymásra, amely tény a beszélők azonosságára utal. A félautomatikus beszélőazonosítás módszertanában a fenti vizsgálatokon kívül különböző algoritmusok al-

kalmazásával elkészített grafikonokon is elemezzük a beszédhangot. Ezek a következők lehetnek:

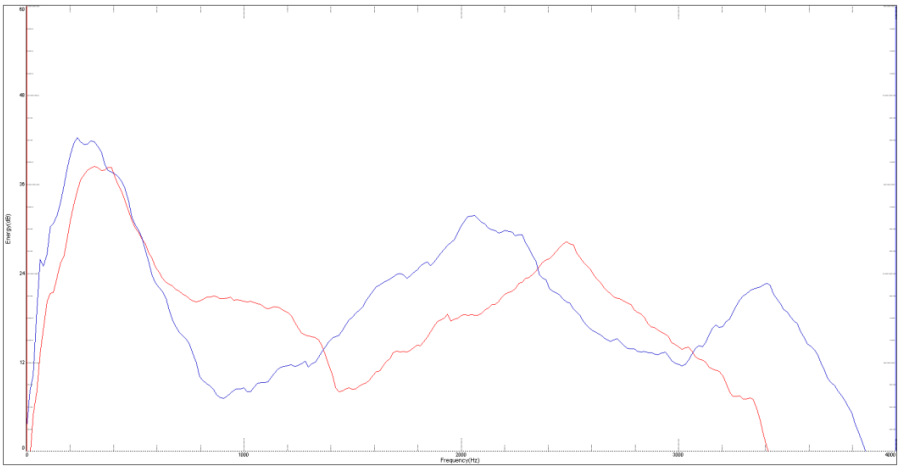
- FFT[87], LPC[88], LTA[89], Cepstrum analízis.

Az FFT analízis egy hang adott pillanatban mért hangfrekvenciás összetevőinek energiaviszonyait, az LPC analízis a szájüreg hangátviteli karakterisztikáját ábrázolja. A Cepstrum grafikon görbéje kettős FFT analízissel készül, és a fáziseltolódás miatt bekövetkező energiamaximum-pontok frekvencia-eltolódásának kiküszöbölésére használható. Az LTA analízis - hosszú idejű, átlagolt FFT analízis alkalmazásával - a beszéd hangfrekvenciás összetevőinek energiaviszonyait mutatja meg nagyobb mennyiségű hanganyagon. Előnye, hogy alkalmazásával különböző szövegtartalmú beszéd is elemezhető. A következő képeken azonos és különböző beszélők hangja alapján, többféle algoritmussal készült görbék láthatóak, melyekkel a vizsgálati módszert szemléltetem.

[87] *FFT (Fast Fourier Transformation - Gyors Fourier Transzformáció)*
[88] *LPC (Linear Prediction Coefficient - Lineáris Predikciós Együttható)*
[89] *LTA (Long Term Average - Hosszú Idejű Átlag)*



3. ábra LPC analízis azonos beszélők "a" hangjaira



4. ábra FFT analízis különböző beszélők "i" hangjaira

A 3. ábra azonos személytől más-más időpontokban rögzített hangfelvételekből kiemelt "a" hangokra készült LPC grafikont, a 4. ábra különböző személyek "i" hangjai alapján készült FFT görbéket ábrázolja. A grafikonok elemzésekor a görbék maximum- és minimumhelyeinek értékeit, a meredekségüknek, lefutásának hasonlóságát, vagy jelentős (szignifikáns) eltéréseit vizsgáljuk. Mivel ugyanazt a

hangot nem vagyunk képesek még egyszer pontosan ugyanúgy kiejteni, a fenti algoritmusokkal elkészített görbék azonos személyek esetében sem lesznek teljesen egybevágóak, viszont nagyfokú hasonlóságot fognak felvenni. A 3. ábrán látható, hogy a görbék maximumhelyei majdnem azonos helyen találhatóak, és azok lefutása, meredeksége is nagymértékben hasonlít egymásra, ami a beszélők azonosságára utal, míg a 4. ábrán a minimum- és maximumhelyek frekvencia- és intenzitás értéke, a görbék meredeksége jelentős eltérést mutat, amely a beszélők különbözőségét támasztja alá.

Fontos megjegyezni, hogy mind a spektrografikus vizsgálatoknál, mind a különböző algoritmusokkal megvalósított elemzések során nagyon sok mintán kell a méréseket elvégezni. Néhány SPG kép, grafikon alapján nem lehet a személyazonosságra megbízhatóan megállapítást tenni, ezért nagy mennyiségű hangot, hangkapcsolatot szükséges összehasonlító vizsgálat alá vonni a szakmailag teljesen megalapozott szakvélemény elkészítéséhez.

A félautomatikus beszélőazonosítás módszertan kiforrott, bevált metódusnak mondható, melyet számos európai országban alkalmaznak, azonban több hátránnyal is rendelkezik. Mivel főként egyező hangok és hangkapcsolatok összehasonlításán nyugszik, eltérő szövegtartalmú hanganyagok esetében, különösen, ha azok rövid időtartamúak, nem mindig lehet a személyazonosítást elvégezni. A teljes vizsgálati módszertan elvégzése időigényes, különösen nagy mennyiségű hanganyag esetében, és nem alkalmas idegen nyelven beszélők azonosítására – hacsak nem ismeri a hangszakértő a nyelvet anyanyelvi szinten. Látható a fentiek alapján az is, hogy bizonyos fokú szubjektivitást a vizsgálati módszerek és a metódus összessége is hordoz magában, mivel támaszkodik az azonosítást végző szakértő elméleti ismereteire, értékítéletére, gyakorlati tapasztalatára. A következő fejezetben tárgyalt hangbiometria ezen problémák döntő részét kiküszöböli, továbbá lehetővé teszi a beszédhang-feldolgozás automatizálását.

Biometrikus (automatikus) beszélőazonosítás

A biometrikus személyazonosítás az ember biológiai, vagy viselkedésbeli tulajdonságait használja fel, abból kiindulva, hogy az egyén biometriai adatai, jellemzői egyediek, minden más személytől különböznek. Egyedi, a biometrián belül vizsgálható viselkedési tulajdonság lehet pl. a billentyűleütés, a kézírás, aláírás, járás, stb. Fiziológiai jellemző a DNS, a retina, az írisz tulajdonságai, és nem utolsósorban a beszédhang, amely a vokális traktus egyedisége következtében alkalmas biometriai alkalmazásra.

Ahhoz, hogy biometrikus jellemző alapján azonosítható legyen az egyén, először meg kell határozni, hogy pontosan milyen adatot kívánunk felhasználni. A beszédhang esetében ez az akusztikai jellemzők értékei, mind a szegmentális, mind a szupraszegmentális szinten, melyek lehetnek pl.: átlagos alaphang-érték, formánsok sávszélessége és frekvenciaértéke, a hangsúly, a dallam és azok változásának mérhető adatai, beszédtempó, stb.. Látható, hogy az akusztikai-fonetikai beszélőazonosítás módszertanából már ismert és vizsgált jellemzőket használja a hangbiometria is, azt azonban, hogy pontosan milyen jellemzők, milyen súllyal és algoritmusok alkalmazásával vesznek részt a jellemzőkinyerés (feature extraction) folyamatában, a fejlesztők, érthető módon, nem hozzák nyilvánosságra. Az ismeretes, hogy legjobban a beszéd spektruma reprezentálja a beszélő vokális traktusát, ezért ebből számítják ki különböző eljárások alkalmazásával a beszéd jellemzővektorát. Az egyik alkalmazott eljárás, amikor a beszédhangot 20 ms-os ablakszélességgel, 10 ms-os átlapolással szegmentálják és MFCC[90] függvénnyel, 19 elemű vektor és annak deriváltjai felhasználásával számítják ki a jellemzővektorokat.

A következő lépés a vizsgált személy biometriai beszédhang-modelljének az elkészítése. Kutatások szerint ennek egyik leghatékonyabb módja a GMM[91] alkalmazása, melynek segítségével a beszédhangra jellemző Gauss-görbe (haranggörbe) paramétereinek legpontosabb becslését lehet elérni.

A tudományos kutatások eredményei és a rendszerfejlesztők tapasztalatai azt mutatják, hogy a biometrikus beszélőazonosítás eredményessége jelentősen javítható, hogy ha a beszédhang valószerűségi értékét normalizálják egy általános háttérmodellből származó értékkel[92]. Ehhez egy nagy adatbázison tanított háttérmodellt (UBM[93]) készítenek, amely a jellemzők beszélő-független eloszlását reprezentálja. A UBM elkészítéséhez például a BatVox automatikus azonosítórendszer fejlesztői 4 órányi, nemenként szétválasztott beszédhangot használtak fel.

Az automatikus azonosítás utolsó lépése a hasonlóság valószínűségi értékének meghatározása, amelyhez leggyakrabban a Bayes-analízist[94], illetve az abból

---

[90] *MFCC (Mel Frequency Cepstral Coefficients - rövid idejű spektrális burkológörbe)*

[91] *GMM (Gaussian Mixture Modell - Gauss-féle keverék modell)*

[92] *Beszédtudomány tanulmánykötet, szerkesztette: Markó Alexandra, MTA Nyelvtudományi Intézet 2012, pp. 344*

[93] *UBM (Universal Background Model - általános beszélő-háttérmodell)*

[94] *Joaquin Gonzalez-Rodriguez, Andrzej Drygajlo, Daniel Ramos-Castro, Marta Garcia-Gomar, Javier Ortega-Garcia: Robust Estimation, Interpretation and Assessment of*

származó LR[95] meghatározását hívják segítségül. Az eddigiek szerint áttekintve a beszélőazonosítás tudományos hátterét és a módszerek alkalmazásának alapelveit láthattuk, hogy jelenleg nem ismert olyan akusztikai jellemző, vagy jellemzők együttese, amely kifejezetten beszélő-specifikus lenne, továbbá számszerűsíthető formában (lsd. neurális spektrogram) rendelkezésünkre állna. Emiatt a beszélőazonosítás során az egyes jellemzőket összehasonlítjuk, és az eredmények alapján meghatározzuk a hasonlóság valószínűségét. Erre a műveletre kínál objektív matematikai-valószínűségi eljárást a Bayes-tétel, amelynek egyszerűsített alakja így néz ki:

posteriori esély = LR x priori esély

A posteriori esély a vádlott bűnösségének esélye a bizonyítékok és az ismérvek alapján, a priori esély a vádlott bűnösségének esélye a bizonyítékok megvizsgálása előtt, továbbá kijelenthető, hogy mindkét esély megvizsgálása a bíróság felelősségi köre. Egy büntetőeljárás bírói szakában a Bayes-tétel megközelítésében az az eldöntendő kérdés, hogy mekkora a valószínűsége a vádlott bűnösségének, illetve hogy a bizonyíték (hangfelvétel) milyen mértékben növeli, vagy csökkenti annak a valószínűségét, hogy a vádlott bűnös.

Mivel tudjuk, hogy a Bayes-tétel egy feltételes valószínűség és annak fordítottja között állít fel kapcsolatot, a fenti egyszerűsített képletbe behelyettesíthetjük a beszélőazonosítással kapcsolatos elemeket, így kapjuk az alábbi függvényt.

$$\frac{p(H_0/E)}{p(H_1/E)} = \frac{p(E/H_0)}{p(E/H_1)} \times \frac{p(H_0)}{p(H_1)}$$

posteriori esély — Likelihood Ratio — priori esély

Az egyenletben a p (probability) valószínűséget jelöl, az E (evidence) a hangfelvétel, mint bizonyíték, H betűvel pedig kétféle hipotézis van jelölve:

Likelihood Ratios in Forensic Speaker Recognition, Computer Speech and Language, 26 July 2005. pp. 5-14.

[95] *LR (Likelihood Ratio - Valószínűségi Arány)*

H0: annak valószínűsége, hogy a kérdéses felvételen az ismert személy hangja hallható

H1: annak valószínűsége, hogy a kérdéses felvételen nem az ismert személy hangja hallható

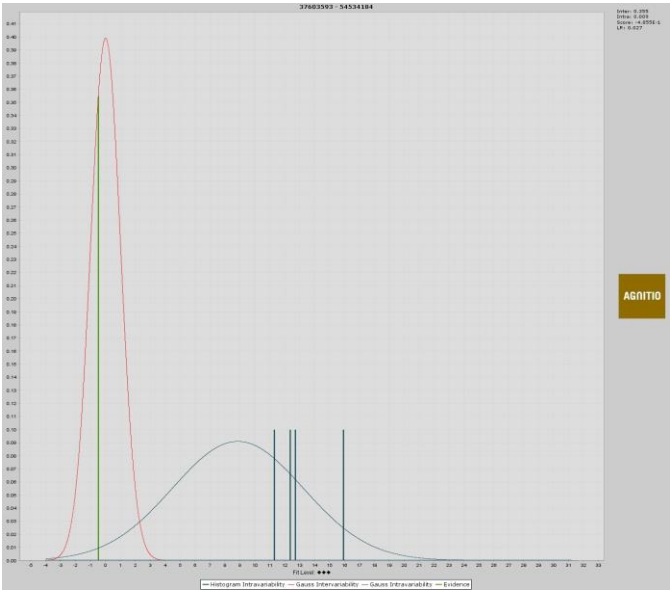
A fentiek szerint a Likelihood Ratio két valószínűség, az Intravariabilitás és az Intervariabilitás hányadosa, amelyet a biometrikus rendszer számít ki, és amelynek értékelése a szakértő feladata. Ahhoz, hogy a módszertanról teljes képet kapjunk, ki kell térnünk a biometrikus azonosítás bemeneti feltételeire is. Egy, Európában széleskörűen alkalmazott biometrikus rendszer, a BatVox esetében a hangfájl technikai paraméter-követelményei egy közepes mintavételezésű és bitfelbontású felvételnek felelnek meg (8 kHz, 16 bit, PCM wav), melyeket manapság már könnyű biztosítani. Az ismeretlen személytől legalább 7 másodperc, az ismerttől minimum 40 másodperc nettó (szünetek nélküli) hanganyaggal kell rendelkezni ahhoz, hogy az azonosítási eljárás elvégezhető legyen. A gyakorlati tapasztalatok azt mutatják, hogy legpontosabb eredményeket a 2-4 perc hosszúságú felvételek esetében lehet elérni, továbbá, hogy 4-5 percnél hosszabb felvételeket nem érdemes vizsgálat alá vonni, mert nem javul a beszélőazonosítás hatékonysága, hibaaránya.

A következőkben a már említett BatVox rendszer segítségével szemléltetem a biometrikus beszélőazonosítás eredményeinek kiértékelését. Az 5. ábrán öt, ismeretlen személyazonosságú (52893168, 62583165, 54534184, 54574746, 33204179), és egy ismert beszélő (37603593) hangja összevetésének számítási eredményei láthatóak. A 37603593 sz., ismert személyazonosságú beszélő hangja önmagával is össze van hasonlítva annak érdekében, hogy lássuk, mi a legmagasabb LR hasonlósági érték, amelyet a rendszer meghatároz. Esetünkben ez 1E10, ami exponenciális formában 1010-nek felel meg.

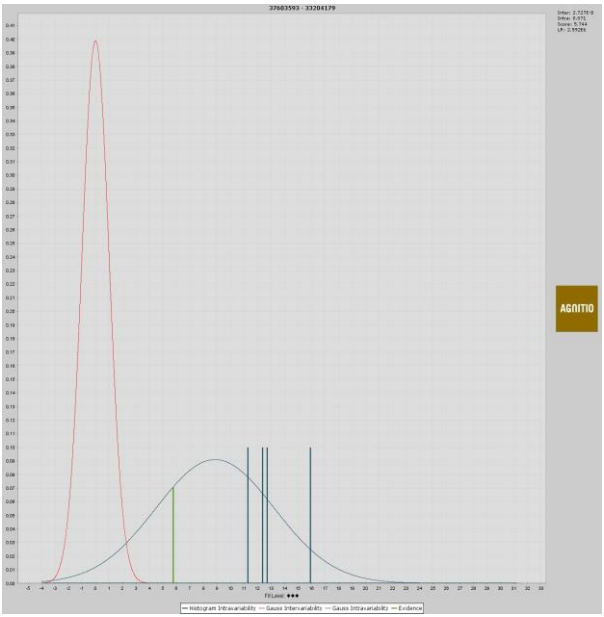
	Model Order 1			
	Name		Score	Fit Level
37603593	37603593		1E10	Ok
52893168	37603593		0.02743	Ok
62583165	37603593		0.27472	Ok
54534184	37603593		0.02657	Ok
54574746	37603593		35.08425	Ok
33204179	37603593		2.59242E6	Ok

5. ábra LR eredmények táblázata

A kiértékelésnél a következőket kell figyelembe venni, ahhoz, hogy objektív szakvélemény születhessen. A táblázat utolsó sorának LR értéke 2.59242E6, amely azt jelenti, hogy 2.59242x106 - szor valószínűbb az, hogy a 33204179 és a 37603593 számú beszélő azonos, mint hogy különböző személyekről van szó. Ennek az interpretációnak megfelelően a kék színnel jelölt, nullánál kisebb LR értékek, és az utolsó előtti sorban szereplő, nagyon alacsony 35.08425 szám is egyértelmű különbözőséget jelöl. Ezek a számok grafikusan is ábrázolhatóak, és így segítenek a szakértőnek abban, hogy az eredményeket kiértékelve a hiteles megállapításokat megtegye. A következő ábrákon különböző és azonos beszélők LR eredményeinek grafikus ábrázolása egyaránt látható.



6. ábra LR eredmény grafikus ábrázolása különböző beszélők esetén



7. ábra LR eredmény grafikus ábrázolása azonos beszélők esetén

A grafikonokon kék egyenesekkel a 37603593 sz. (ismert) beszélő hangja alapján képzett hisztogram van jelölve, piros színű a populációs adatbázis alapján számított Gauss-görbe (intervariabilitás), és kék színű szintén a 37603593 sz. személy beszédhangja alapján számított Gauss-görbe (intravariabilitás). Amennyiben a zöld színű marker egyenes (amely az ismeretlen személy beszédhangját, tehát a kérdéses hangfelvételt jelöli) a piros színű Gauss-görbén belül helyezkedik el, ez arra utal, hogy a két vizsgált beszédhang nem ugyanazon személytől származik, míg ha a zöld színű egyenes a kék színű Gauss-görbén belül helyezkedik el, ez arra utal, hogy a két beszédhang azonos személytől származik. Az, hogy a két lehetőségnek milyen a valószínűségi foka, attól függ, hogy a zöld egyenes milyen Intervariabilitás, Intravariabilitás, Score és LR értékek mellett helyezkedik el a grafikon X tengelyén. Megtévesztő lehet, hogy az 5. ábrán látható LR táblázatban és a grafikonon is Score érték található. A feloldás, hogy az 5. ábra táblázatában szereplő Score érték valójában az LR arányt mutatja (ahogy a grafikon jobb felső sorában is látható), míg a grafikus ábrázolás Score értéke egy, a beszélők biometriai modelljének hasonlóságát kifejező viszonyszámot mutat meg.

Mind az LR, mind a Score eredményekre érvényes, hogy értékelni kell azokat, mégpedig különböző szempontrendszerek (sávhatárok, küszöbszintek) alapján. Ennek elvégzéséhez a szakértőnek az elmélet alapos ismeretével, az adott rendszerrel működésével kapcsolatosan nagy gyakorlattal kell rendelkeznie ahhoz, hogy szakértői megállapításai a személyazonosságra vonatkozóan hitelesek, és alátámaszthatóak legyenek.

A hangbiometria szélesebb körű alkalmazási lehetőségei

Az eddigiekben a beszélőazonosítás módszertana 1:1 és 1:N (egy személyt egy személyhez, egy személyt több személyhez) azonosítási lehetőségeken került elemzésre. Megállapítható, hogy a félautomatikus metódus leginkább az 1:1 típusú azonosításra alkalmas, azzal a kiegészítéssel, hogy elvégezhető vele az 1:N típusú eljárás is, amennyiben csak kis (N) számú személy beszéde vesz részt az vizsgálatban. Ez tipikus feladat felderítési szakban, bűncselekmény gyanúsítottjának beazonosítása céljából, vagy ha bizonyítani kell az ügy bírói szakában, hogy a terhelt személy beszél az inkriminált hangfelvételen - amennyiben azt nem ismeri el. A hangbiometrián belül léteznek minden azonosítási metódust (1:1, 1:N, N:N) megvalósítani képes rendszerek, ezek közül a cikkemben említett BatVox különösen az 1:1 és 1:N módszertanra alkalmas, azzal, hogy itt már jóval nagyobb számú, több tíz (esetleg százas nagyságrendű) személlyel is elvégezhető a

feladat. Léteznek azonban olyan rendszerek, amelyek több tíz-, vagy százezer beszélőn is képesek azonosítás lefuttatására oly módon, hogy feldolgozható számú személyekre képesek leszűkíteni a valószínűleg azonos személyek körét. Ez egy, az azonosság valószínűségének eredménye alapján létrehozott találati listát jelent, amelyből a feldolgozó személy, esetleg további vizsgálatok során a hangszakértő, már képes behatárolni, hogy ki beszél a hangfelvételen.

A kriminalisztikában régóta alkalmazott eljárás, hogy egy bűncselekményben érintett személytől DNS-, ujjlenyomat, stb. mintát rögzítenek a bűnügyi nyilvántartásba vétel során annak érdekében, hogy egy ismeretlen minta esetén megállapítható legyen a személyazonosság, hogy volt-e már a mintát hagyó személy valamilyen eljárásban terheltként érintett. Európa számos országában alkalmazzák azt a műveletet, hogy a nyilvántartásban vétel során nemcsak a fenti jellemzőket rögzítik, hanem a gyanúsítottól, hangszakértő által kidolgozott eljárás alapján hangmintát is rögzítenek. Ezzel lehetővé válik egy ismeretlen személytől származó hanganyag esetén a beazonosítás, amennyiben a bűnügyi nyilvántartás rendszerében már rögzítve van a biometrikus beszédhang-modell, így jelentős segítséget kaphatnak a felderítést végző hatóságok.

Végül ki kell térni a személyes adatok védelmére a (hang)biometriai adatok kezelésével összefüggésben. Az adatvédelmi törvény szerint a személyes adat bármely meghatározott (azonosított vagy azonosítható) természetes személlyel (a továbbiakban: érintett) kapcsolatba hozható adat, az adatból levonható, az érintettre vonatkozó következtetés. A személyes adat az adatkezelés során mindaddig megőrzi e minőségét, amíg kapcsolata az érintettel helyreállítható. A személy különösen akkor tekinthető azonosíthatónak, ha őt – közvetlenül vagy közvetve – név, azonosító jel, illetőleg egy vagy több, fizikai, fiziológiai, mentális, gazdasági, kulturális vagy szociális azonosságára jellemző tényező alapján azonosítani lehet. Mindezek alapján könnyen belátható, hogy a hangbiometriai beszédhang modell egyértelműen a személyes adatok körébe tartozik, mivel az alapján a beszélő személy azonosítható. Amennyiben a beszélőhöz kötött biometriai adatokat - akár bűnügyi nyilvántartás keretén belül, akár telefonbeszélgetések offline, vagy online elemzésekor - széleskörűen kívánjuk felhasználni, ez csak a vonatkozó törvények, alapelvek (célhoz kötöttség, adattárolás biztonsága, személyes adatok hozzáférésének védelme, stb.) szigorú betartásával történhet jogszerűen, így talán elkerülhetővé válik a tudományos kutatások eredményeinek rossz cél érdekében történő felhasználása.

A fentiekre tekintettel a majdan létrejövő biometriai beszédhang-modellekből létrejövő adatbázisokat szigorúan védeni kell. Ez jelent adatbiztonsági védelmet, tehát, hogy illetéktelenek ne férhessenek hozzá az abban tárolt

adatokhoz, ami az információbiztonság körébe tartozó, hardverekkel, szoftverekkel viszonylag egyszerűen megvalósítható feladat. Nehezebb dió, ha arról beszélünk, hogy az adatbázisokban szereplő modelleket a hatóság, a különböző szervezetek csak a célhoz kötöttség elve alapján használhatják fel, mert az igény, hogy egy érintett személyről mindent tudni akar egy szervezet, általában szélesebb körű szokott lenni, mint ami az érintett jogszabályok szellemével összhangban van. Ezért a törvényekbe be kell majd építeni azokat a „fékeket", korlátozásokat, amelyek az egyén számára biztosítják, hogy a hangbiometriai adatait kizárólag jól behatárolható – bűncselekmények felderítéséhez kapcsolódó – cél érdekében használhatja fel az állam, elkerülve azt, hogy bárki, bármilyen okból, még egy, személyhez köthető jellemző felhasználásával totálisan ellenőrizhető legyen.

Felhasznált irodalom

1. GÓSY Mária: Fonetika, a beszéd tudománya, Osiris Kiadó, Budapest, 2004, pp. 11., 275.
2. Beszédtudomány tanulmánykötet, szerkesztette: Markó Alexandra, MTA Nyelvtudományi Intézet 2012, pp. 340., 344.
3. Joaquin Gonzalez-RODRIGUEZ, Andrzej Drygajlo, Daniel Ramos-Castro, Marta Garcia-Gomar, Javier Ortega-Garcia: Robust Estimation, Interpretation and Assessment of Likeli hood Ratios in Forensic Speaker Recognition, Computer Speech and Language, 26 July 2005. pp. 5-14.
4. TREMMEL Flórián, FENYVESI Csaba, HERKE Csongor: Kriminalisztika, Dialóg Campus Kiadó 2009, 2012 pp. 290-295.
5. CRAIG Adam: Essential Mathematics and Statistics for Forensic Science, Wiley-Blackwell 2010