

A szűrők sávszélességének hatása az SNN-alapú kulcsszó-felismerésre: háromszög és keskenysávú szűrő összehasonlítása neuromorf MEMS-szenzor tervezéshez

Zeffer Tamás^{1,2}, Csikósné Pap Andrea², Volk János¹

¹HUN-REN Energiatudományi Kutatóközpont, 1121 Budapest, Konkoly-Thege Miklós út 29–33., 29. épület

²Óbudai Egyetem, Anyagtudományok és Technológiák Doktori Iskola, 1081 Budapest, Népszínház utca 8.

Absztrakt: Ebben a cikkben javaslatot teszünk az impulzus alapú neurális hálózatokkal (SNN) megvalósított neuromorf beszéd felismerésre, amelynek célja a hagyományos mélytanulási megközelítésekhez képest az energiahatékony megvalósítás. Spirál alakú konzolos megközelítésünk tervezése során azonban fontos tervezési szempont annak megértése, hogyan befolyásolja az osztályozási teljesítményt a (magas jósági tényezők által korlátozott) keskeny szűrő sávszélesség. Jelen cikk a szűrő sávszélességének hatását vizsgálja az SNN-alapú beszédosztályozásra, összehasonlítva a standard háromszög (változó sávszélességű) szűrőket a MEMS gyártási korlátjaihoz optimalizált keskeny sávú (58,6 Hz állandó sávszélességű) szűrőkkel. A 35 szóosztályt tartalmazó Google Speech Commands adatbázis segítségével értékeltük az osztályozási teljesítményt, a precizitást (Precision), az érzékenységet, az F1-pontszámot és a pontosságot (Accuracy) mutatói alapján. Eredményeink azt mutatják, hogy a háromszög szűrők 62,02% pontosságot érnek el, míg a keskeny sávszélességű szűrők 50,18% pontosságot, ami 11,84% teljesítménykülönbséget jelent. Bár ez a különbség nem elhanyagolható, a keskeny sávszélességű konfiguráció elegendő teljesítményt mutat a neuromorf szenzor megvalósításához. Ezek az eredmények gyakorlati tervezési útmutatást nyújtanak a frekvenciaszelektív konzolos rendszereket fejlesztő MEMS-mérnökök számára, és arra mutatnak rá, hogy a 60 Hz sávszélességű 10 csatornás rendszer megvalósítható első generációs neuromorf beszédérzékelő képvisel.

Kulcsszavak: impulzus alapú neurális hálózat, MEMS konzol, szűrő sávszélesség, neuromorf kódolás, beszédosztályozás, hardvertervezés

1. Bevezető

A neuromorf számítástechnika jelentős potenciált kínál az energiahatékony peremhálózati számítások (edge computing) terén a Dolgok Internete (IoT) alkalmazásokban [2]. A folyamatos értékű jeleket feldolgozó hagyományos mesterséges neurális hálózatokkal ellentétben, az impulzus alapú neurális hálózatok (SNN-ek) diszkrét eseményekkel (impulzusokkal) kommunikálnak, lehetővé téve az eseményvezérelt feldolgozást, amely drasztikusan csökkentheti az energiafogyasztást [3]. Auditív alkalmazásokhoz olyan neuromorf megközelítést javasunk, amely frekvenciaszelektív MEMS konzolos szerkezetet kombinál SNN hálózattal. Ebben az architektúrában mechanikai rezonátorok végzik a frekvenciafelbontást az analóg tartományban, ami – teljes hardveres megvalósítás esetén – kiválthatja a hagyományos beszédfelismerő rendszerekben általában szükséges, nagy számításigényű gyors Fourier-transzformáció (FFT) alkalmazását.

Hardverfejlesztésünk során a szűrőszélességet az SNN osztályozási teljesítményét befolyásoló kulcsparaméterként azonosítottuk. A hagyományos hangfeldolgozás jellemzően változó sávszélességű Mel-skálás szűrőket [4] alkalmaz, amelyek alacsony frekvenciákon finom, míg magas frekvenciákon durvább frekvenciafelbontást biztosítanak. Azonban a szilíciumra hordozókra gyártott, jellemző, magas jósaági tényezővel (Q-faktoral) rendelkező MEMS konzolok természetüknél fogva keskeny, állandó sávszélességgel rendelkeznek, amelyet mechanikai tulajdonságaik határoznak meg. Ez felvet egy alapvető kérdést: a keskeny sávú MEMS szűrők képesek-e az algoritmikus háromszög szűrőkhöz hasonló osztályozási teljesítményt elérni?

Ez a tanulmány két szűrőkonfiguráció szisztematikus összehasonlításával foglalkozik egy 10 csatornás SNN-alapú beszédosztályozóban. A 10 csatornára összpontosítunk, mert laboratóriumunk korábban már gyártott egy 16 karos spirális alakú konzoltömböt [1], amelyből 12 konzol mutatott működőképes rezonanciát a 281–672 Hz-es tartományban. Míg a gyártott hardverünk az alacsonyabb hangfrekvenciás tartományban működik, a tanulmányban szereplő 10 csatornás szoftvermodell 300–8000 Hz-es sávra terjed ki, hogy a teljesítményt a teljes beszéd spektrumon értékelhessük. E szélesebb frekvenciatartomány hardveres eléréséhez a spirálgeometria újratervzésére lenne szükség az egyes konzolok magasabb rezonanciafrekvenciákra hangolása érdekében, ami a konzolméretek és a szeizmikus tömeg módosításával megvalósítható. A meglévő hardver ugyanakkor reális alapot biztosít egy 10 csatornás neuromorf szenzorterv értékeléséhez. A két összehasonlított szűrőtípus a következő:

- háromszögszűrők: változó sávszélességű Mel-skála szűrők (amelyek a hangfeldolgozásban szabványosnak számítanak);

- keskenysávú szűrők: a MEMS gyártás által korlátozott szűrők, amelyeknél a konzolok magas jósági tényezője (Q-faktora) keskeny frekvenciaszelektivitást eredményez.

Jelen munka hozzájárulásai a következők:

- a háromszög és a keskenysávú szűrők közvetlen teljesítmény-összehasonlítása 10 csatorna esetén;
- a pontosság és sávszélesség közötti kompromisszum számszerűsítése a MEMS szenzortervezés számára;
- annak bemutatása, hogy a keskenysávú szűrők közelítőleg 50% pontosságot érnek el a 35 osztályt tartalmazó Google Speech Commands feladaton [5], igazolva ezzel a hardveres megvalósíthatóságot;
- gyakorlati tervezési javaslatok kidolgozása a neuromorf MEMS konzolos rendszerekhez

A cikk további felépítése a következő: a 2. fejezet áttekinti a kapcsolódó szakirodalmat, a 3. fejezet ismerteti kísérleti módszereinket, a 4. fejezet bemutatja az eredményeket, az 5. fejezet tárgyalja a hardvertervezési vonatkozásokat, a 6. fejezet pedig az összegzést tartalmazza.

2. A kutatás háttere

A neuromorf hallószenzorok célja a biológiai csiga frekvenciaszelektivitásának reprodukálása. Jelen munkában hardveres alapként a MEMS konzoltömbökre összpontosítunk, amelyek előnyöket kínálnak a miniatürizálás és a gyártási megbízhatóság terén az alacsony fogyasztású hallóeszközök számára.

Korábbi tanulmányok, például Jang et al. [7], téglalap alakú, AIN bevonatú konzolos szerkezeteket mutattak be mesterséges baziláris membránokhoz, amelyek 2,9–12,6 kHz tartományban értek el frekvenciaszelektivitást. Az ilyen megközelítések kiterjesztése az alacsonyabb frekvenciákra (300–1000 Hz), amelyek a teljes hangfelismeréshez szükségesek, kihívást jelent a konzolos szerkezetek hosszának növekedése miatt. Az MFA Nanoszenzor Laboratóriumában legyártott spirális konzolkialakítás ezt arkhimédészi

spirálgeometria és szeizmikus tömeg alkalmazásával oldja meg, 300–700 Hz-es tartományban érve el rezonanciákat egy kompakt, 2 mm × 2 mm-es alapterületen.

A konzolos rendszerek természetüknél fogva keskeny rezonancia-csúcsokat hoznak létre, amelyek sáv szélességét a Q-tényező határozza meg. Ez magas frekvenciaszelektivitást biztosít, de korlátozza a csatornánként rögzített információt, így kompromisszumot teremt a szelektivitás és az információs sáv szélesség között. Míg a szoftveralapú szűrők tetszőleges sáv szélességeket tudnak megvalósítani, ezeket a hardver által előírt korlátozásokat még nem vizsgálták szisztematikusan a neuromorf hallószensorok összefüggésében.

3. Alkalmazott módszerek

3.1 Hardveremuláció és rendszeráttekintés

Megközelítésünk egy MEMS konzolos érzékelőkből, VO₂ memrisztor alapú impulzusgenerátorokból és egy SNN osztályozóból álló hardverrendszert emulál. A konzolos érzékelők frekvenciaszelektív mechanikus rezonátorokként működnek, a memrisztorok a mechanikus rezgéseket impulzussorozatokká alakítják, az SNN pedig az osztályozást végzi. Bár a teljes fizikai rendszer még fejlesztés alatt áll, az elgondolást szoftveres emulációval igazoljuk, a legyártott eszközeinken mért hardverparaméterek felhasználásával.

A MEMS konzolos érzékelők spirál alakú piezoelektromos szerkezetek, amelyeket alumínium-nitrid (AlN) piezoelektromos anyaggal ellátott szilícium-szigetelő (SOI) hordozókra gyártott a laboratóriumunk. A gyártott MEMS konzolos szerkezeteken mért Q-tényezők 117-254 között vannak, ami keskeny rezonancia sáv szélességnek felel meg. Az itt értékelt 10 csatornás szoftvermodell 58,6 Hz-es állandó sáv szélességet használ a hardver viselkedésének közelítésére.

3.2. Szűrőkonfigurációk

A Mel-spektrogram előállításánál során alkalmazott két szűrőbank-kialakítást hasonlítjuk össze:

- Háromszögű szűrő (Standard): Ez a konfiguráció 10 változó sáv szélességű háromszögű szűrőt használ, amelyek logaritmikus eloszlásúak a Mel-skálán a 300–8000 Hz tartományban. Minden szűrő átfedésben van a szomszédos szűrőkkel, követve a szabványos hangfeldolgozási konvenciókat. A sáv szélesség a középfrekvenciával együtt nő, finom felbontást biztosítva az alacsony frekvenciákon.

- Keskenysávú szűrő (Hardver által korlátozott): Ez a konfiguráció 10 háromszög szűrőt használ állandó 58,6 Hz-es sávzélességgel (a háromszög alapjának teljes szélességében mérve). A középfrekvenciák eloszlása megegyezik a háromszögű szűrő esetével, de minden szűrő rögzített, keskeny sávzélességgel rendelkezik, amely a MEMS konzol viselkedését közelíti.

Egy tudományos cikkben a számításokat és levezetéseket érdemes kiemelni a folyó szövegből, hogy áttekinthetőbbek legyenek. A Wordben ezt általában képletszerkesztővel szokás csinálni, de szövegesen is megformázható úgy, hogy profin nézzen ki.

Az 58,6 Hz-es sávzélesség meghatározása a 31,25 Hz-es frekvenciafelbontáson alapul, amelyet az alábbiak szerint számítottunk:

$$\Delta f = \frac{f_s}{N_{FFT}} = \frac{16000 \text{ Hz}}{512} = 31,25 \text{ Hz}$$

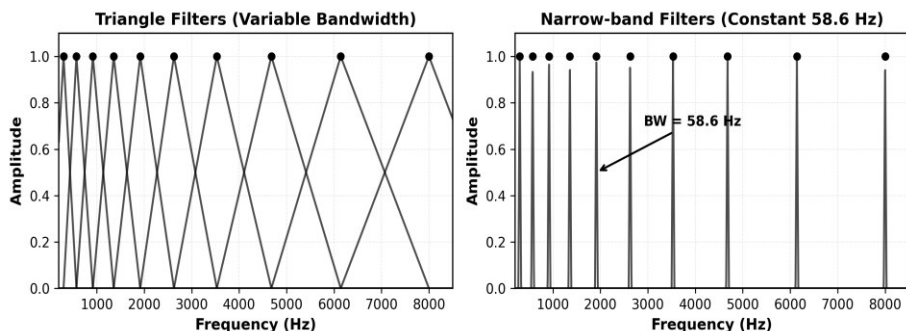
A háromszög szűrő minimális sávzélességére vonatkozó követelmény a -3 dB-es pontoknál:

$$B_{-3dB} = 2\sqrt{2} \cdot \Delta f \approx 2,83 \cdot 31,25 = 88,4 \text{ Hz}$$

Ez az érték a háromszög alapjánál mérve – feltételezve, hogy a -3 dB pont a csúcspont 70,7%-ánál helyezkedik el – körülbelül 58,6 Hz-nek felel meg. Az 1. ábra mind a standard háromszög (változó sávzélességű), mind a keskeny sávú (állandó 58,6 Hz) szűrőkonfigurációkat szemlélteti, bemutatva a standard Mel-skála szűrés és a hardverkorlátozott megközelítés összehasonlítását.

1. ábra

Szűrők frekvenciaválaszának összehasonlítása 10 csatornás konfiguráció esetén. Baloldali képen látható, hogy a változó sávzélességű háromszög szűrők szélesebb spektrális lefedettséget biztosítanak magasabb frekvenciákon. Az állandó 58,6 Hz sávzélességű keskeny sávú szűrők megközelítik a MEMS konzolos viselkedést, ami a jobb oldali ábrán látható. A fekete pontok a középfrekvenciákat jelölik.

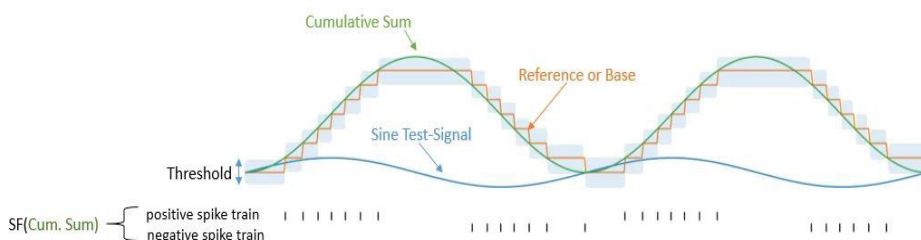


3.3 Neuromorf impulzuskódolás

A Mel-spektrogram kimenetét kumulatív küszöbérték-alapú megközelítéssel alakítjuk impulzusokká (2. ábra). Minden egyes frekvenciacsatornánál az $A(t)$ amplitúdó időbeli összegzése történik, és impulzus generálódik minden alkalommal, amikor a kumulatív összeg meghalad egy beállítható küszöbértéket. A szinuszhullámmal szemléltetett folyamat során az impulzusok akkor keletkeznek, amikor a kumulatív összeg átlépi a referencia-alapvonalat. Ez a VO_2 memrisztor oszcillátor áramköreink viselkedését emulálja, amelyek a bemeneti amplitúdóval arányos impulzussűrűséget generálnak. Hardveres méréseink lineáris kódolást mutatnak 100–800 Hz közötti impulzussűrűséggel, 25–60 nm-es gerjesztési amplitúdók esetén.

2. ábra

Neuromorf impulzuskódolási eljárás kumulatív küszöbérték-alapú megközelítéssel. A bemeneti jel kumulatív összegét a rendszer egy referencia-alapvonalhoz (narancssárga) hasonlítja. Amikor a kumulatív összeg átlépi a küszöbértéket, pozitív (felfelé mutató) vagy negatív (lefelé mutató) impulzusok keletkeznek. A megjelenített szinuszhullám szemléltetési célt szolgáló tesztszel; a gyakorlatban ezt a kódolást az egyes frekvenciacsatornák Mel-spektrogram amplitúdóértékeire alkalmazzuk. Ez a módszer a VO_2 memrisztor oszcillátor áramkörök viselkedését emulálja, ahol az impulzussűrűség arányos a bemeneti amplitúdóval.



3.4. Impulzus alapú neurális hálózat (SNN) architektúra

Az SNN négy, teljesen összekapcsolt rétegből áll, amelyek mindegyike 256 szivárgó integráló-tüzelő (LIF) neuront tartalmaz. A bemeneti réteg 10 impulzussorozatot fogad (frekvenciacsatornánként egyet), a kimeneti réteg pedig 35 kimenetet állít elő, amelyek a Google Speech Commands szókészletének felelnek meg. A hálózati architektúrát és a tanítási paramétereiket a 1. táblázat foglalja össze.

1. táblázat

Az SNN architektúra paramétereinek és a tanítási konfigurációnak összefoglalása a 10 csatornás beszédosztályozáshoz.

Kategória	Paraméter	Érték/Beállítás
Hálózati Struktúra	Architektúra	4 teljesen összekapcsolt réteg
	Rétegenként a neuron szám	256
	Bemeneti csatornák	10 (impulzussorozat)
	Kimeneti csatornák	35 (GSC szókészlet)
Neuron Model	Típus	Leaky Integrate-and-Fire (LIF)
	Membránon lecsengés (β)	0.9
	Tüzelési küszöb	0.5
Tanítás	Veszteségfüggvény	Impulzussűrűségek átlagos négyzetes hibája (MSE)
	Optimalizáló	Adam
	Keretrendszer	snnTorch [6]

Ez a kompakt architektúra a hardveres megvalósíthatóságot helyezi előtérbe, miközben fenntartja az elegendő megkülönböztető képességet, vagyis a reprezentációs kapacitástl a 35 osztályos osztályozási feladathoz. A viszonylag kis hálózati méretet (4 réteg $\times$ 256 neuron) szándékosan választottuk, hogy lehetővé tegyük a jövőbeli megvalósítást neuromorf hardverplatformokon.

¹ A reprezentációs képesség (vagy más néven kifejezőerő – expressive power) a gépi tanulásban azt jelenti: a neurális hálózat azon képessége, hogy mennyire összetett összefüggéseket, függvényeket vagy mintázatokat tud megtanulni és leképezni.

3.5 Adatkészlet és képzés

A Google Speech Commands (GSC) adatkészletet [5] használjuk, amely körülbelül 105 000 egy másodperces hangmintát tartalmaz 35 gyakori szóból. Az eredeti kulcsszó-felismerési feladattal (12 osztály) ellentétben mi mind a 35 osztályt betanítjuk, hogy értékeljük a rendszer teljes szókincs-felismerési képességét. Az audiojeleket 16 kHz-es mintavételi frekvenciával rögzítjük. A Mel-spektrogramot 512 mintás Hamming-ablakkal és 20 mintás ugrási hosszúsággal (1,25 ms) generáljuk, ami megfelelő időfelbontást biztosít a beszédjellemzőkhez.

4. Eredmények

4.1. Teljesítmény-összehasonlítás 10 csatorna esetén

A 2. táblázat mutatja be a háromszög és a keskenysávú szűrők közötti teljesítmény-összehasonlítást a 10 csatornás konfigurációra. Minden metrikát a GSC tesztadatbázison értékeltünk.

2. táblázat

A háromszög és a keskenysávú szűrők teljesítményének összehasonlítása 10 csatornás SNN osztályozás esetén a Google Speech Commands adatbázison [5] (35 osztály).

Szűrőtípus	Precizitás (%)	Érzékenység (%)	F1-Score (%)	Tesztpontosság (%)
Háromszög (Változó sáv szélesség)	61.32	57.61	56.75	62.02
Keskenysávú (58.6 Hz)	51.77	46.63	46.00	50.18
Teljesítménykülönbség	9.55	10.98	10.75	11.84

A háromszög-szűrő konfiguráció 62,02%-os tesztpontosságot ér el, 11,84 százalékponttal felülmúlva a keskenysávú szűrőt (50,18%). Ez a teljesítménykülönbség minden tanulási mutatónál következetesen jelentkezik; a precizitás, az érzékenység és az F1-érték hasonló, körülbelül 10–11% eltérést mutat.

A teljesítménykülönbség ellenére a keskenysávú szűrő 50,18% pontosságot mutató értéket ér el. Összehasonlításképpen, a véletlenszerű osztályozás egy 35 osztályos besorolás esetén mindössze 2,86% pontosságot eredményezne (35-ből

1), ami azt bizonyítja, hogy még a sávszélességben korlátozott szűrők is lehetővé teszik az érdemi osztályozást. Ez a teljesítmény alacsonyabb, mint az ezen az adatbázison a hagyományos mesterséges neurális hálózatok (ANN) esetében közölt eredmények [5], viszont igazolja a kis teljesítményű, hardveres megvalósítás kivitelezhetőségét.

5. Értelmezés és összefüggések

5.1. Kompromisszumok: teljesítmény és gyárthatóság

MEMS tervezési megközelítésünk egyik központi kompromisszuma az osztályozási teljesítmény és a hardveres gyárthatóság között húzódik. A változó sávszélességű szűrők hardveres megvalósítása vagy (1) változó jósági tényezőjű (Q-faktorú) konzolokat igényelne, ami tömeggyártásban nehezen szabályozható, vagy (2) aktív hangolóáramköröket, amelyek növelik az energiafogyasztást és a bonyolultságot. Az állandó sávszélességű megközelítés számos előnnyel jár:

- Egyszerűbb gyártás: Minden konzol azonos szerkezeti kialakítást használhat, csupán hosszban eltérve.
- Kiszámítható viselkedés: Az állandó jósági tényező egyszerűsíti a modellezést és a kalibrációt.
- Kompakt alapterület: 2 mm × 2 mm-es kialakításunk illeszkedik a középfül méretkorlátaikhoz.

A 11,84% teljesítménykülönbség elfogadható lehet olyan alkalmazásokban, ahol az energiahatékonyság és a kompaktság fontosabb az abszolút pontosságnál.

5.2. Összehasonlítás a kapcsolódó munkákkal

A háromszög és a keskenysávú szűrőkkel elért eredményeinket összevethetjük a korábbi SNN-alapú kulcsszófelismerési munkákkal. Számos korábbi tanulmány a 10–12 osztályos benchmarkkal [5] foglalkozik, vagy mélyebb hálózati architektúrákat alkalmaz [8].

A szabványos benchmark 10 cél-kulcsszót használ, kiegészítve az „ismeretlen” és „csend” osztályokkal a nem célzott beszéd kezelésére.

Munkánk újdonsága nem a maximális pontosság elérésében rejlik, hanem abban, hogy bebizonyítottuk, hogy a hardverre jellemző szűrő sávszélességek képesek támogatni a jelentősegteljes osztályozást. Tudásunk szerint a mi tanulmányunk az első olyan szisztematikus kutatás, amely összehasonlítja az algoritmikus és a hardveresen korlátozott szűrőszélességeket SNN kontextusban.

5.3. Út a nagyobb teljesítmény felé

Míg a 10 csatornás konfiguráció igazolja a megvalósíthatóságot, a 20–40 csatornára való skálázás csökkentheti a teljesítménykülönbséget a keskenysávú és a háromszög szűrők között. Bővebb (előkészületben lévő) tanulmányunk azt mutatja, hogy a 40 csatornás keskenysávú szűrők jobban megközelítik a háromszög szűrők teljesítményét. Ugyanakkor 40 konzol jelentősen növeli a chipfelületet és a gyártási bonyolultságot. Az optimális csatornaszám az adott alkalmazás követelményeitől függ.

5.4. Korlátok

Jelen tanulmánynak számos korlátja van. Elsőként, szoftveres emulációt használunk a fizikai MEMS hardverrel történő tesztelés helyett. Míg emulációnk mért konzolparamétereken alapul, a valós tényezők, mint például a gyártási szórás és a mechanikai áthallás befolyásolhatják a teljesítményt. Másodszor, elemzésünk egyetlen sávszélesség-értékre (58,6 Hz) összpontosít; a sávszélességek szélesebb tartományának vizsgálata feltárhatná az optimális kompromisszumos pontokat.

6. Következtetések

Jelen cikk a szűrők sávszélességének SNN-alapú kulcsszó-felismerő alkalmazásra gyakorolt hatását vizsgálta, összehasonlítva a standard háromszög szűrőket a MEMS gyártásra optimalizált keskenysávú szűrőkkel. A 10 csatornás konfigurációt alkalmazva megállapítottuk, hogy a keskenysávú (58,6 Hz) szűrők körülbelül 50%-os pontosságot érnek el a 35 osztályos Google Speech Commands adatbázison, szemben a háromszög szűrők 62%-ával. Bár ez a kb. 12%-os teljesítménykülönbség nem elhanyagolható, a keskenysávú konfiguráció elegendő pontosságot mutat a koncepcióigazoló (proof-of-concept) neuromorf szenzorok számára. Legjobb tudásunk szerint ez az első szisztematikus összehasonlítása az algoritmikus és a hardver által meghatározott szűrőmegoldásoknak a neuromorf hangfeldolgozás terén.

Eredményeink gyakorlati útmutatást nyújtanak a MEMS-mérnökök számára, igazolva, hogy egy 10 csatornás, csatornánként közel 60 Hz sávszélességű

rendszer megfelelő alapja lehet egy első generációs neuromorf beszédszenzornak. A spirális konzolgeometria lehetővé teszi a kompakt megvalósítást ($2\text{ mm} \times 2\text{ mm}$ chipenként), alkalmassá téve a rendszert méretkorlátos alkalmazásokra.

A jövőbeli munka a 10 csatornás MEMS tömb fizikai megvalósítására és tesztelésére, a VO_2 memrisztor impulzusgenerátorokkal való integrációra, valamint a sávszélesség-korlátozott bemenetekre optimalizált SNN architektúrák vizsgálatára összpontosít. Továbbá, az adaptív sávszélesség-kiosztás vizsgálata – ahol a kritikus frekvenciatartományok keskenyebb, szelektívebb szűrőket kapnak – javíthatja a teljesítmény és a bonyolultság közötti kompromisszumot.

References

- [1] P. Udvardi et al., "Spiral-Shaped Piezoelectric MEMS Cantilever Array for Fully Implantable Hearing Systems," *Micromachines*, vol. 8, no. 10, p. 311, Oct. 2017.
- [2] C. D. Schuman et al., "Opportunities for neuromorphic computing algorithms and applications," *Nat Comput Sci*, vol. 2, no. 1, pp. 10–19, Jan. 2022.
- [3] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.
- [4] Y.-Y. Yang et al., "TorchAudio: Building Blocks for Audio and Speech Processing," arXiv:2110.15018, 2022.
- [5] P. Warden, "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition," arXiv:1804.03209, 2018.
- [6] J. K. Eshraghian et al., "Training Spiking Neural Networks Using Lessons From Deep Learning," arXiv:2109.12894, 2023.
- [7] J. Jang et al., "A microelectromechanical system artificial basilar membrane based on a piezoelectric cantilever array and its characterization using an animal model," *Sci. Rep.*, vol. 5, p. 12447, 2015.
- [8] Q. Yang, Q. Liu, and H. Li, "Deep residual spiking neural network for keyword spotting in low-resource settings," in *Proc. Interspeech*, 2022, pp. 3023–3027.