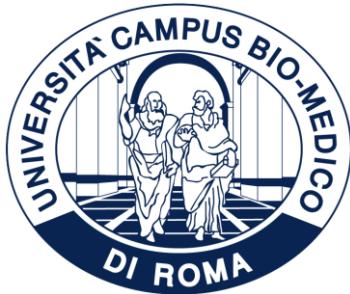


Matr. N 15905



**UNIVERSITA'
CAMPUS BIO-MEDICO DI ROMA**

**FACOLTA' DIPARTIMENTALE DI INGEGNERIA
CORSO DI LAUREA IN INGEGNERIA INDUSTRIALE**

**PROGETTAZIONE E SVILUPPO DI UN
DISPOSITIVO DI SOSTITUZIONE SENSORIALE
AUDIOVISIVA**

Relatore

Prof. Giorgio Pennazza

Correlatore

Ing. Costanza Cenerini

Laureando

Gianluca Barnaba

ANNO ACCADEMICO 2023/2024

Indice

1. INTRODUZIONE	2
1.1. Sostituzione sensoriale.....	2
1.2. Sostituzione audiovisiva	3
1.3. Obiettivi	4
2. MATERIALI E METODI	4
2.1. Hardware.....	4
2.2. Software.....	6
2.3. Sonificazione	8
3. RISULTATI E DISCUSSIONE.....	10
3.1. Flusso di lavoro del sistema	10
3.2. Test.....	12
4. CONCLUSIONI E SVILUPPI FUTURI.....	13

1. INTRODUZIONE

1.1. Sostituzione sensoriale

La sostituzione sensoriale è definita come la tecnica che "consente di sperimentare un senso attraverso un altro, per lo più utilizzando input provenienti da modalità artificiali" [1]. Questa pratica consente di colmare le lacune sensoriali derivanti dalla perdita o dal danneggiamento di un senso, sfruttando i segnali provenienti da un altro sistema sensoriale. Uno dei principali sostegni scientifici su cui si basa la sostituzione sensoriale è la neuroplasticità cross-modale, un fenomeno per cui alcune aree della corteccia cerebrale, normalmente coinvolte in un certo tipo di percezione sensoriale, si riorganizzano, sia a livello strutturale che funzionale, per poter essere occupate dai sistemi sensoriali intatti, in modo da elaborare informazioni provenienti da questi ultimi [2].

I primi studi nel campo della sostituzione sensoriale risalgono agli anni '60 e vennero portati avanti da Paul Bach-y-Rita, il pioniere dell'uso estensivo di questa tecnica [3]. Egli dimostrò che le persone cieche potevano "vedere" tramite il tatto, convertendo informazioni visive in segnali aptici: progettò un dispositivo¹ dotato di una telecamera che convertiva immagini in stimoli tattili percepiti sulla pelle, sviluppando così uno dei primi modelli basati su questa tecnologia [1].

Negli ultimi decenni, la ricerca sulla neuroplasticità cross-modale ha ulteriormente chiarito i meccanismi cerebrali alla base di questo fenomeno, permettendo ai nuovi dispositivi di sfruttare al meglio diversi canali sensoriali, in particolare quello tattile [4]. Il neuroscienziato David Eagleman, figura di spicco in questo ambito di ricerca, sì è concentrato su questa modalità sensoriale, investigando la percezione aptica anche nell'ambito dell'espansione e dell'aggiunta sensoriale [5].

Oltre a quella visiva e tattile, sono numerose le modalità sensoriali coinvolte nelle applicazioni di questa pratica, sia come strumenti di percezione che come sensi da sostituire. Infatti, le categorie di dispositivi, per la sostituzione sensoriale, sono denominate dal nome del senso sostituito e dall'aggettivo che descrive il sistema sensoriale sostituente: udito tattile, visione tattile, equilibrio tattile, consapevolezza spaziale uditiva o tattile, visione uditiva, tatto uditivo e visivo. In alcune applicazioni specifiche, invece, non è necessario che il nuovo canale sensoriale sia diverso da quello che si vuole sostituire, ma viene modificata esclusivamente la sua collocazione fisica. È il caso della ricollocazione sensoriale tattile, ossia la trasmissione di uno stimolo tattile da un dispositivo protesico o da una

¹ Il dispositivo prende il nome di "*Tactile Vision Sensory Substitution*" (TVSS)

porzione di pelle priva di sensibilità a un'altra, in modo da restituire la propriocezione di una certa parte del corpo [1].

Dall'analisi della letteratura si evince che il suono ed il tatto siano i sensi più utilizzati, nella maggior parte dei sensory substitution devices (SSD), per compensare la perdita della vista. Anche se, storicamente, i dispositivi aptici sono stati quelli maggiormente prodotti [7]. Tuttavia, negli ultimi anni sono emersi i vantaggi della sostituzione uditiva della vista: l'udito umano è in grado di gestire modelli sonori complessi e in rapida evoluzione, oltre a saper riconoscere, con un'ottima sensibilità, la frequenza e l'intensità dei segnali acustici [7]. Tali considerazioni hanno portato alla scelta di realizzare, in questo lavoro, un dispositivo di visione uditiva.

1.2. Sostituzione sensoriale audiovisiva

La sostituzione audiovisiva è un'applicazione specifica della sostituzione sensoriale che permette di trasformare informazioni visive in segnali acustici, offrendo alle persone con disabilità visive la possibilità di percepire l'ambiente circostante attraverso il suono [6].

Uno dei più noti SSD di sostituzione audiovisiva è il vOICe, che trasforma le informazioni catturate da una telecamera in segnali sonori, i quali variano in base alla posizione, forma e luminosità degli oggetti [6]. Il software di conversione codifica tre parametri di ogni pixel dell'immagine (coordinata orizzontale, coordinata verticale e luminosità) nei tre parametri dell'onda sonora (tempo, frequenza, ampiezza): il suono di uscita sarà la sovrapposizione di tutte le onde che codificano ogni pixel. Il vOICe, come la maggior parte di questi SSD, si concentrano principalmente sulla rappresentazione di forme o spazi [6], trascurando la componente cromatica: tuttavia, il colore è una parte essenziale della percezione visiva e arricchisce notevolmente la comprensione dell'ambiente per l'utente.

È interessante come alcuni utenti quotidiani del vOICe abbiano riportato la comparsa di sinestesie percettive da cui prima non erano affetti. La sinestesia percettiva è un fenomeno sensoriale in cui l'attivazione di un senso provoca automaticamente la percezione di un'altra sensazione in un senso diverso, creando una "contaminazione" tra i sensi [7]. Quella sperimentata dagli utenti, però, è una sinestesia acquisita, alquanto differente da quella innata o evolutiva presente nei soggetti vedenti o che non hanno mai utilizzato il vOICe, proprio perché basata sull'esperienza del dispositivo stesso. Questo fenomeno, che si è verificato in alcuni utenti anche quando il dispositivo non è indossato, conferma la teoria della plasticità cross-modale che è alla base della sostituzione sensoriale [6].

Oggi l'evoluzione tecnologica ha portato l'ideazione di nuovi SSD per la sostituzione audiovisiva, ancora più precisi ma sempre più complessi. Uno dei più recenti è il Colorphone 2.2, un dispositivo che cerca di tradurre la posizione e il colore dei pixel di un'immagine in tempo reale. La conversione si basa su un algoritmo simile al vOICe ma presenta diversi

parametri in più; inoltre, la corrispondenza colore-suono si basa su esperimenti sulla sonificazione e sulla corrispondenza cross-modale [9]. Il progetto, però, presenta ancora molti limiti: la scarsa illuminazione provoca un ritardo significativo nel processo di conversione e la traduzione degli stessi colori con diversi suoni, oltre al verificarsi di interferenze tra pixels vicini che a volte vengono processati come un unico pixel [9]. Sopravvive, oltretutto, un problema che persiste già dal vOICe, ovvero la complessità nella traduzione video-sonora, che implica un lungo tempo di istruzione dell’utente al dispositivo e provoca spesso uno scarso utilizzo, lasciando questi dispositivi limitati all’ambiente di ricerca e inutili alla riabilitazione visiva [3].

1.3. Obiettivi

Il fine di questo lavoro è di progettare e sviluppare un dispositivo di sostituzione sensoriale audiovisivo, tentando di risolvere alcune problematiche riscontrate nella letteratura e puntando ad aumentare il tasso di utilizzo di questi dispositivi. In particolare, il dispositivo mira a colmare, nell’ideazione della sonificazione, la mancanza di engagement affettivo, causata dall’incapacità, presente in molti SSD, di trasmettere emozioni da un senso all’altro [8]. Inoltre, il progetto cerca di ridurre la complessità e i tempi di addestramento al dispositivo, provando a rendere il suo funzionamento accessibile e intuitivo. Infine, ci si propone di sviluppare una soluzione economica e che possa essere integrata in un accessorio indossabile, tentando di superare i limiti di costo ed ingombro che affliggono molti SSD [3].

2. MATERIALI E METODI

2.1. Hardware

La progettazione di questo SSD è basata sull’utilizzo della UnitV k210 Edge Computing AI Camera (*figura 1*), un dispositivo embedded progettato per applicazioni di AI Visioning. Le sue dimensioni sono 40 x 24 x 13 mm e, al suo interno, integra una videocamera e un System-on-chip (SoC).



Figura 1: UnitV k210 Edge Computing AI Camera

L'elaborazione dell'immagine avviene real time on edge e richiede una bassa potenza computazionale, grazie all'implementazione sul SoC.

Unità di Elaborazione Il Kendryte K210 è un system-on-chip, il suo principale punto di forza è l'utilizzo di un bassissimo consumo energetico che riesce comunque a garantire un'alta prestazione di visione artificiale e udito automatico. Infatti, questo chip consuma circa un terzo rispetto ai suoi concorrenti (0,3 W rispetto a 1 W [12]), oltre a disporre di un'ampia gamma di unità periferiche, per un gran numero di scenari applicativi, e tante altre funzionalità. Questa unità è orientata allo sviluppo "a soglia zero" ("zero threshold"), ovvero punta a ridurre significativamente i tempi e la complessità del processo di sviluppo, essendo progettato per essere facilmente accessibile e integrabile in vari progetti [11]. Solitamente, nel contesto dei sistemi embedded, i SoC come il K210 vengono inseriti in un sistema hardware più ampio detto acceleratore. Infatti, nella UnitV, il Kendryte K210 è incorporato in un acceleratore KPU (Neural Processing Unit), mostrato in *figura 2*. Il datasheet del SoC è disponibile online [13].



Figura 2: Kendryte K210 integrato nel KPU

Videocamera La CameraChip OV7740 è un sensore di immagine CMOS VGA a basso consumo e alta sensibilità che fornisce funzionalità complete di una fotocamera VGA a chip singolo in un ingombro ridotto, solo 5 mm. Inoltre, l'OV7740 vanta una buona sensibilità alle basse luci di 6,0 V/lux-sec e può operare a 30 fotogrammi al secondo (fps) con risoluzione VGA e 60 fps in risoluzione QVGA, con completo controllo da parte dell'utente sulla qualità dell'immagine, la formattazione e il trasferimento dei dati in uscita. Il sensore utilizza la tecnologia proprietaria per migliorare la qualità dell'immagine riducendo le comuni fonti di illuminazione contaminante dell'immagine, come disturbi e sbarature, per produrre un'immagine a colori più pulita e stabile. A scopo di memorizzazione, l'OV7740 comprende anche una memoria programmabile (OTP) [14]. Il datasheet è disponibile online [14].



Figura 3: CameraChip OV7740

Speaker Lo speaker utilizzato è il modello 8O3W-JST-PH2.0 di CQRobot. Il connettore JST-PH2.0, oltre a poter essere collegato e scollegato agilmente, permette a questi speaker di essere facilmente integrabili con progetti IoT e, in particolare, con microcontrollori. L’altoparlante è poco ingombrante, solo 70 x 31 x 15 mm, possiede un’impedenza di 8 ohm, che lo rende compatibile con molti circuiti audio, e consuma al massimo 3 W. Maggiori informazioni sono disponibili online [17].



Figura 4: Speaker 8O3W-JST-PH2.0

2.2. Software

Il firmware del SSD è stato sviluppato in Micropython utilizzando come ambiente di sviluppo integrato (IDE) Maixpy.

IDE Maixpy è una piattaforma di programmazione, sviluppata da Sipeed, che combina hardware e software per applicazioni di intelligenza artificiale e visione artificiale. La funzione primaria del progetto Maixpy è quella di rendere Micropython il linguaggio con cui programmare sul chip Kendryte K210 [15]. L’IDE fornisce un display che mostra in tempo reale le acquisizioni della fotocamera e un terminale seriale per verificare l’output del programma, oltre ai classici tools per quanto riguarda lo script.

Linguaggio di programmazione MicroPython è una versione ottimizzata e più leggera di Python per programmare con SoC e microcontrollori. Questo linguaggio permette di programmare diversi dispositivi con risorse limitate, come i sistemi embedded, utilizzando un linguaggio di alto livello come Python. Infatti, la sua compattezza gli permette di essere eseguito in soli 256 kB di spazio per il codice e 16 kB di RAM [16]. Offre una serie di moduli specifici per l'hardware, permettendo il controllo diretto di pin GPIO, sensori e altri componenti.

Funzioni di AI Visioning L'AI Visioning è un campo della computer vision che si occupa di analisi e interpretazione delle immagini utilizzando tecniche di intelligenza artificiale, in particolare il machine learning e il deep learning [12]. L'UnitV k210 integra diverse funzioni per l'elaborazione dell'immagine acquisita dalla telecamera.

Abilità di AI Visioning	Funzioni Maixpy
Object detection	<ul style="list-style-type: none"> • <i>kpu.load()</i>: Carica un modello di rilevamento oggetti • <i>kpu.run_yolo2()</i>: Esegue l'algoritmo YOLO v2 per il rilevamento di oggetti
Face detection and recognition	<ul style="list-style-type: none"> • <i>face_detect.init()</i>: Inizializza il rilevatore di volti • <i>face_detect.run()</i>: Esegue il rilevamento dei volti su un'immagine • <i>face_recognition.load()</i>: Carica un modello di riconoscimento facciale • <i>face_recognition.recognize()</i>: Riconosce i volti in un'immagine
Pose estimation	<ul style="list-style-type: none"> • <i>kpu.load()</i>: Carica un modello di stima della posa • <i>kpu.run_yolo2()</i>: Può essere usato per rilevare punti chiave del corpo

Feature detection	<ul style="list-style-type: none"> <i>image.find_blobs()</i>: Trova blob di colore in un'immagine <i>image.find_lines()</i>: Trova linee in un'immagine <i>image.find_circles()</i>: Trova cerchi in un'immagine
Object tracking	<ul style="list-style-type: none"> <i>image.find_keypoints()</i>: Trova punti chiave per il tracciamento <i>image.match_keypoints()</i>: Confronta punti chiave tra frame
Image processing	<ul style="list-style-type: none"> <i>image.gaussian()</i>: Applica un filtro gaussiano <i>image.median()</i>: Applica un filtro mediano <i>image.dilate()</i>: Dilata un'immagine
Motion detection	<ul style="list-style-type: none"> <i>image.difference()</i>: Calcola la differenza tra due immagini
Gesture recognition	<ul style="list-style-type: none"> <i>kpu.load()</i>: Carica un modello di classificazione <i>kpu.forward()</i>: Esegue l'inferenza su un'immagine

Tabella 1: Funzioni di AI Visioning

2.3. Sonificazione

Come già anticipato, la sonificazione utilizzata è frutto dello studio sperimentale citato in bibliografia [8], quindi rispecchia le preferenze espresse dagli adulti e dai bambini che hanno partecipato all'esperimento.

La mappatura utilizza il modello di colore espresso in tonalità, saturazione e luminosità (HSB). Per identificare l'accordo musicale sono stati scelti i seguenti parametri: la tonica (RT) dell'accordo (quindi Do Do# Re ecc.), l'ottava (OC) in cui viene suonato (ad

esempio Do1 Do2 ecc.), e la modalità (MO) dell'accordo (come maggiore, minore, diminuita ecc.).

Tra le sei possibili combinazioni per l'associazione tra i parametri visivi e quelli sonori, la mappatura H-OC, S-MO, B-RT è stata quella più votata e che, sia adulti che bambini, hanno considerato più naturale.

Per la realizzazione di questa proof of concept, si è utilizzato un modello di sonificazione discreto che traduce in suono 12 range di colore. Di seguito viene riportata la tabella di associazione.

Range	Hue	Saturation	Brightness	Accordo
1	300-60°	0-50%	0-50%	Do3 Maggiore
2	300-60°	50-100%	0-50%	Do3 Minore
3	300-60°	0-50%	50-100%	Re3 Maggiore
4	300-60°	50-100%	50-100%	Re3 Minore
5	60-180°	0-50%	0-50%	Do4 Maggiore
6	60-180°	50-100%	0-50%	Do4 Minore
7	60-180°	0-50%	50-100%	Re4 Maggiore
8	60-180°	50-100%	50-100%	Re4 Minore
9	180-300°	0-50%	0-50%	Do5 Maggiore
10	180-300°	50-100%	0-50%	Do5 Minore
11	180-300°	0-50%	50-100%	Re5 Maggiore
12	180-300°	50-100%	50-100%	Re5 Minore

Tabella 2: Sonificazione

3. RISULTATI E DISCUSSIONE

In questa sezione si presenta il processo di traduzione del colore nell'accordo musicale.

3.1. Flusso di lavoro del sistema

Il processo di traduzione si divide in tre blocchi:

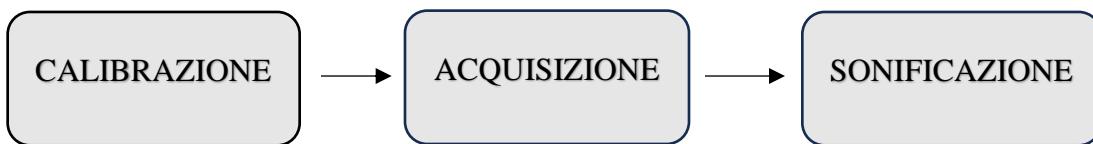


Figura 5: Flusso di lavoro del sistema

Calibrazione del colore della videocamera: la calibrazione è fondamentale affinché l'intero dispositivo funzioni in diversi ambienti con differente illuminazione. All'avvio del programma, è richiesto all'utente di puntare entro 5 secondi una cartolina dal colore noto² e di ripetere il procedimento per 3 cartoline, nel seguente ordine: rossa, verde e blu. I valori RGB delle cartoline utilizzate sono riportati in *tabella 2*. La funzione Maixpy sviluppata rileva i valori RGB delle cartoline inquadrata, che sono quindi valori relativi all'illuminazione specifica, e li confronta con i rispettivi valori canonici (noti), ricavando così i fattori di correzione (o calibrazione) da applicare alle successive acquisizioni. I valori RGB delle cartoline inquadrata sono la media dei valori al centro dell'immagine (10 x 10 pixels), quindi basta puntare le cartoline affinché siano al centro del display, senza troppa accuratezza. Mentre è essenziale che la procedura di calibrazione sia ripetuta ogni volta che si modifica l'illuminazione nell'ambiente circostante.

Cartolina	Codice Pantone	R	G	B
Rosso	199	213	0	50
Verde	383	168	173	0
Blu	18-4530	6	107	139

Tabella 2: Valori RGB delle cartoline

² Nelle prove effettuate, sono state utilizzate cartoline del marchio *Pantone*

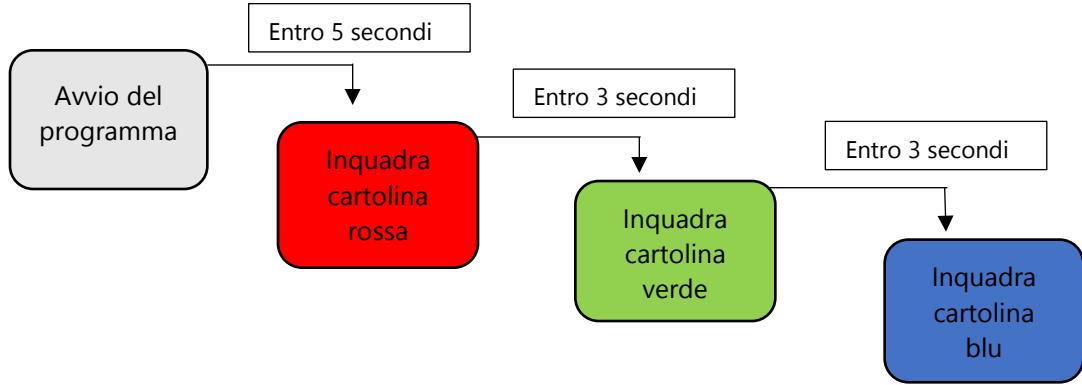


Figura 6: Procedura di calibrazione

Acquisizione del Colore: questo blocco è quello che meglio sfrutta le capacità di visione artificiale della UnitV, che non solo è in grado di rilevare i valori RGB di ogni pixel inquadrato, ma riesce a riconoscere macchie di colore (blobs) in tempo reale. Dopo la fase di calibrazione, parte automaticamente il rilevamento nell’immagine dei blobs, che vengono mostrati sul display con un rettangolo bianco che li delimita (*figura 7*): il dispositivo prima seleziona il blob più grande tra quelli dello stesso colore e ripete il procedimento per ogni colore; successivamente confronta i blobs selezionati per poi rilevare il blob di dimensione maggiore in assoluto, così da tradurre esclusivamente il colore dell’oggetto più rilevante nell’inquadratura. Nel codice MicroPython, il metodo `find_blobs` necessita, come parametri di ingresso, i valori `thresholds`³ nel modello Lab del colore dei blobs da cercare. La conversione dal modello RGB al modello Lab richiede il passaggio intermedio nel modello XYZ. Perciò, è stato necessario implementare tre funzioni, per la traduzione da RGB a Lab, e ricavarsi i fattori di calibrazione anche per il modello Lab.



Figura 7: Visualizzazione dei blobs rilevati sul display

Sonificazione: nell’ultimo step, i valori RGB del blob più esteso vengono tradotti nella rappresentazione del colore HSB, attraverso una funzione dedicata, per poi selezionare l’accordo corrispondente. Come per la calibrazione, viene eseguita una media dei valori RGB dei 100 pixel centrali del blob. Il passaggio tra modelli di colore è stato necessario

³ Esprimono il range, per ogni parametro (L, a, b), in cui si trova il valore del colore da rilevare

dato che il sensore della fotocamera non è supportato per rilevare direttamente il colore dei pixel nel modello HSB e la sonificazione è stata studiata proprio con questo modello, essendo quello che meglio si adatta alla percezione umana.

3.2. Test

In questa sezione vengo descritte le due prove che sono state effettuate per verificare il corretto funzionamento del dispositivo.

Prova di rilevamento

- **Obiettivo:** In questo test si è accertato che il dispositivo traducesse il colore del blob di dimensione maggiore, inquadrato dalla videocamera, nel corretto suono corrispondente.
- **Setup:** Dopo aver effettuato la calibrazione, sono state puntate tre cartoline con diversa dimensione (*figura 8*): una blu e due rosse con diversa saturazione, a cui corrispondono due accordi diversi. Questa procedura è stata ripetuta tre volte, modificando, ad ogni ripetizione, il ruolo della cartolina più estesa.
- **Risultati:** Ciascuna delle tre prove ha portato alla corretta sonificazione.

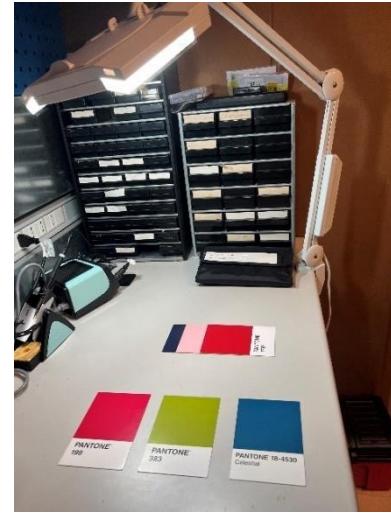


Figura 8: Setup della prova di rilevamento

Prova di sonificazione

- **Obiettivi:** In quest'altra prova è stata verificata la traduzione sonora di ogni range di colore prefissato nella sonificazione.
- **Setup:** Sono state scelte 12 cartoline Pantone in modo che ognuna appartenesse ad un range cromatico diverso e ciascuna di esse è stata esposta sia ad una fonte di luce calda che ad una fonte di luce fredda.
- **Risultati:** Quindi sono state effettuate in totale 24 misurazioni e di queste solamente due hanno portato alla selezione scorretta dell'accordo corrispondente. Le cartoline in questione sono la rossa appartenente al range 1, quando esposta a luce calda, e la verde appartenente al range 6, quando esposta a luce fredda.



Figura 9: Setup della prova di sonificazione

4. CONCLUSIONI E SVILUPPI FUTURI

In questo lavoro si è progettato, sviluppato e testato un dispositivo di sostituzione sensoriale audiovisiva. È stato realizzato un sistema embedded, dotato di intelligenza artificiale, in grado di rilevare il colore dell'oggetto più grande inquadrato e riprodurre l'accordo musicale corrispondente.

Il dispositivo ha dimostrato risultati promettenti durante i test. Ha identificato correttamente l'oggetto più grande inquadrato nel 100% dei casi, con una sonificazione accurata nel 95% delle situazioni. Gli unici due errori di sonificazione si sono verificati con cartoline aventi un valore di saturazione del 48%, molto vicino alla soglia stabilita. Tutte le altre conversioni sono state eseguite con successo.

Tuttavia, la soluzione proposta presenta alcuni limiti. L'uso di quattro modelli di colore, necessario a causa delle implicazioni del firmware, aumenta notevolmente il carico computazionale. Questo approccio genera anche una propagazione dell'errore che, in alcuni casi, può essere rilevante. Anche il setup piuttosto elaborato della calibrazione e la mediocre qualità della videocamera possono essere considerati delle limitazioni, sebbene siano oggettivamente dimensionate, considerando il contesto dell'applicazione.

Nonostante queste sfide, il progetto apre la strada a diverse prospettive. In primo luogo, c'è ampio spazio per perfezionare e ampliare la sonificazione. Dividendo in maniera più fitta il dominio cromatico, si possono tradurre un numero più consistente di colori, arricchendo notevolmente l'esperienza sensoriale degli utenti. Inoltre, il futuro sviluppo di una funzione di autocalibrazione eliminerebbe la necessità di calibrare manualmente il sistema ogni volta che cambia l'ambiente luminoso. Infine, le capacità di AI visioning del Soc permettono di poter integrare al dispositivo altre numerose funzionalità, che possono rilevarsi particolarmente preziose per i soggetti non vedenti, come il riconoscimento dei volti e degli oggetti. Tali funzionalità avrebbero il potenziale di trasformare il dispositivo da uno strumento di percezione dei colori a un assistente multifunzionale per la vita quotidiana.

Queste prospettive future non solo migliorerebbero le prestazioni e l'utilità del dispositivo, ma potrebbero anche ampliarne significativamente il campo di applicazione, offrendo ai non vedenti un supporto sempre più completo e sofisticato nella loro interazione con il mondo visivo.

Bibliografia

- [1] Schmidmaier, Matthias. "Sensory substitution systems." *Media Informatics Advanced Seminar on Multimodal Human-Computer Interaction*. 2011.
- [2] Mowad, Theresa G., et al. "Compensatory cross-modal plasticity persists after sight restoration." *Frontiers in Neuroscience* 14 (2020): 291.
- [3] Maidenbaum, Shachar, Sami Abboud, and Amir Amedi. "Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation." *Neuroscience & Biobehavioral Reviews* 41 (2014): 3-15.
- [4] Eagleman, David M., and Michael V. Perrotta. "The future of sensory substitution, addition, and expansion via haptic devices." *Frontiers in Human Neuroscience* 16 (2023): 1055546
- [5] Eagleman, D. (2015). "Can We Create New Senses for Humans?" TED
- [6] Ward, Jamie, and Peter Meijer. "Visual experiences in the blind induced by an auditory sensory substitution device." *Consciousness and cognition* 19.1 (2010): 492-500.
- [7] Auvray, Malika, Sylvain Hanneton, and J. Kevin O'Regan. "Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with 'The Voice'." *Perception* 36.3 (2007): 416-430.
- [8] Cenerini, Costanza, et al. "INVESTIGATING COLOUR-SOUND MAPPING IN CHILDREN AND ADULTS: A PILOT STUDY." (2023).
- [9] Dominik Osinski et al. «Colorophone 2.2-A Spatial, Real-Time Color-to-Sound Sensory Substitution Device». In: Authorea Preprints (2024)
- [10] "UnitV K210 Edge Computing AI Camera (OV7740)." *M5Stack Store*, <https://shop.m5stack.com/products/unity-k210-edge-computing-ai-camera-ov7740> Ultimo accesso: 15 settembre 2024
- [11] "K210 Overview." *Sipeed Maixduino*, <https://maixduino.sipeed.com/en/hardware/k210.html>. Ultimo accesso: 16 settembre 2024
- [12] "Kendryte AI." *Canaan Creative*, www.canaan-creative.com/product/kendryteai Ultimo accesso: 18 settembre 2024
- [13] "Kendryte K210 Datasheet." *GitHub*, Kendryte, <https://github.com/kendryte/kendryte-doc-datasheet/blob/master/en/001.md> Ultimo accesso: 18 settembre 2024

- [14] "OV7740." *OmniVision*, <https://www.ovt.com/products/ov7740/> Ultimo accesso: 19 settembre 2024
- [15] "MaixPy Documentation." Sipeed Wiki, <https://wiki.sipeed.com/soft/maixpy/en/> Ultimo accesso: 20 settembre 2024
- [16] "MicroPython." *MicroPython*, www.micropython.org/ Ultimo accesso: 20 settembre 2024
- [17] "*Stereo Enclosed Speaker (3 W, 8 ohm) - CQR0508.*" CQRobot Wiki, [http://www.cqrobot.wiki/index.php/Stereo_Enclosed_Speaker_\(3_W,_8_ohm\)-CQR0508](http://www.cqrobot.wiki/index.php/Stereo_Enclosed_Speaker_(3_W,_8_ohm)-CQR0508). Ultimo accesso: 28 settembre 2024.

Ringraziamenti

Innanzitutto, desidero ringraziare il mio relatore, il professor Giorgio Pennazza, per il suo supporto e i suoi consigli, nonché la mia correlatrice l'Ing. Costanza Cenerini per la sua pazienza e la sua gentilezza.

I miei più sentiti ringraziamenti vanno alla mia splendida ragazza Elisa, che mi ha sempre ascoltato e sostenuto, nonostante i miei morbosi pianti, e che mi ha reso felice tutti i giorni.

Poi devo ringraziare i miei fantastici amici che mi hanno accompagnato durante questi tre anni universitari: il robot Luca per l'esempio che mi dà con la sua determinazione e le nottate a casa sua prima degli orali, il coatto pompatto Fede per le sue sfrenate risate e le serate passate insieme, la pazza Arianna per i suoi consigli di vita e che ci ospita sempre per le intense sessioni di studio, quel simpaticone di Mario per il suo umorismo e per la sua bontà, quell'ubriacona di Marta per la sua pazienza e la sua solarità.

Vorrei dedicare un grazie a tutti i miei amici esterni all'università per le serate e le esperienze insieme, senza di voi i libri e Trigoria mi avrebbero risucchiato.

Infine, ringrazio tutta la mia famiglia che mi ha sempre spronato a dare il massimo, in particolare la mia mamma che più di tutti ha voluto indirizzarmi verso questo percorso.