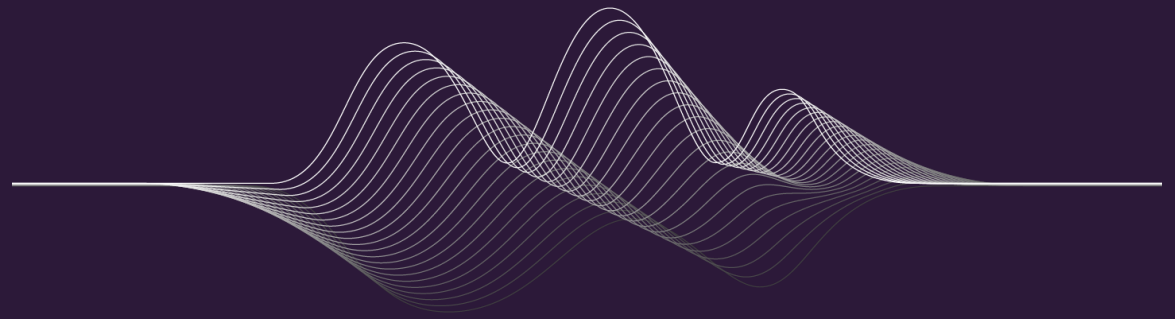


inspire

Inspiring Signal Intelligence with AI

IA al servizio di tecnologie musicali accessibili

prof. Leonardo Gabrielli – INSPIRE – DII – UNIVPM

AI@DII – 29 Ottobre 2026

Agenda

1 Introduzione

2 IA specializzata: music inpainting e BCI

3 AGI: surface reader e applicazioni industriali

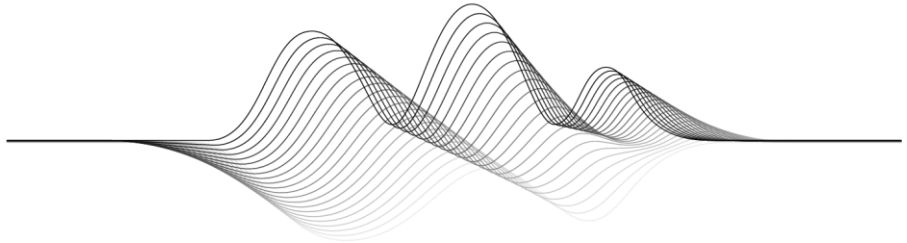
5 Conclusioni

inspire

AI: una keyword tante interpretazioni

La prospettiva INSPIRE (INtelligent Signal Processing for Innovation and Research):

- algoritmi di Signal Processing come circuiti a tempo discreto
- Neural Networks come estensione al caso nonlineare
- prime attività di ricerca del gruppo negli anni '90



inspire

Inspiring Signal Intelligence with AI

Attività in ambito musicale

Storico legame con il distretto industriale

Attività di R&D:

- collaborazione con imprese del territorio e estere
- ricerca congiunta con gruppi di ricerca nazionali e internazionali

Attività didattiche:

- corsi nella laurea elettronica e nella laurea in ing. informatica videogame
- corsi di perfezionamento dal 2015 (Computer Music Production & Sound Design, Nuove tecnologie per la performance)

inspire

Esperienze con Generative AI per la sintesi sonora

2016: Google Deep Mind lancia WaveNet

2017: UNIVPM porta il primo paper su AI per sintesi e sound design al DAFx17 – seguito da brevetto

2018: collaborazione con il compositore Økapi per Rima Glottidis, basata su AI generativa

...

Oggi abbiamo esperienze numerose e partecipazioni in lavori scientifici, progetti industriali e forniamo supporto ad artisti

inspire

Narrow AI vs. AGI

IA specializzata (narrow)

Una rete neurale, un solo compito.

Addestrata su dati mirati. Può essere piccola, veloce, prevedibile.

Esempi: ricostruire frammenti di segnale, riconoscere eventi sonori

IA generale (AGI)

Un sistema capace di qualunque compito intellettuale umano.

Oggi è un obiettivo di ricerca, non ancora pienamente realizzata

ANCHE SE... ☺

Esempi: guidare un umano in compiti difficili, comprendere informazioni tecniche

inspire

Musica d'insieme e terapia remota

Suonare insieme a distanza

Networked Music Performance: musicisti in luoghi diversi che suonano come se fossero nella stessa stanza (approccio "RIA").

Prove e concerti

Ensemble distribuiti, artisti in città diverse

Didattica

Lezioni di strumento senza spostarsi

Terapia

Musicoterapia a casa del paziente, con il terapeuta in clinica

Requisiti: bassa latenza e assenza di perdite.

Due vincoli che le videochiamate non rispettano, e che l'IA specializzata può aiutare a soddisfare.

inspire

La musica come terapia

- Tra gli interventi non farmacologici più efficaci per chi ha un declino cognitivo soggettivo
- Utile intervento per la socializzazione o il recupero di forme autistiche
- Ascoltare e fare musica stimola cognizione, linguaggio e memoria
- Aiuta l'umore e la relazione con il terapeuta

Di solito si fa in presenza.

Per chi ha difficoltà a muoversi o vive isolato, raggiungere il terapeuta è già un ostacolo.

inspire

La musica come terapia

- Tra gli interventi non farmacologici più efficaci per chi ha un declino cognitivo soggettivo
- Utile intervento per la socializzazione o il recupero di forme autistiche
- Ascoltare e fare musica stimola cognizione, linguaggio e memoria
- Aiuta l'umore e la relazione con il terapeuta

Inoltre:

- pazienti con immunodeficenza

inspire

Perché non una videochiamata?

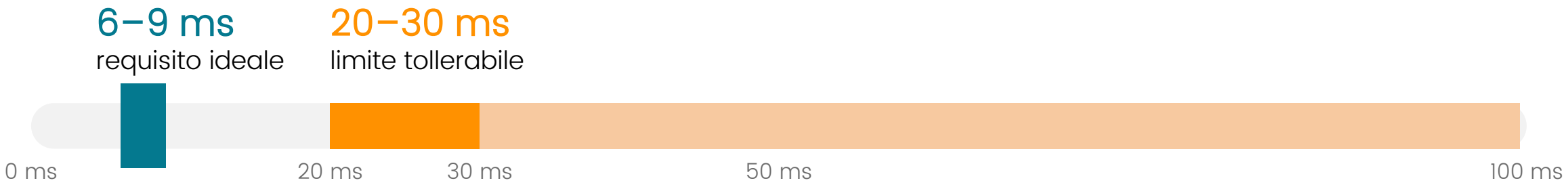
Ritardo

Le piattaforme accumulano il segnale in buffer per evitare interruzioni

Qualità

Audio compresso e «piatto»: si perde la sensazione di stare nella stessa stanza.

Per suonare insieme, il suono deve arrivare entro poche decine di millesimi di secondo.



inspire

Un esempio: il progetto VITALITY

Progetto PNRR: DII + INRCA di Ancona.

Proff. Squartini, Cecchi, Spinsante, Marcantoni

Obiettivo: sessioni terapeutiche e sociali a distanza che sembrino in presenza.

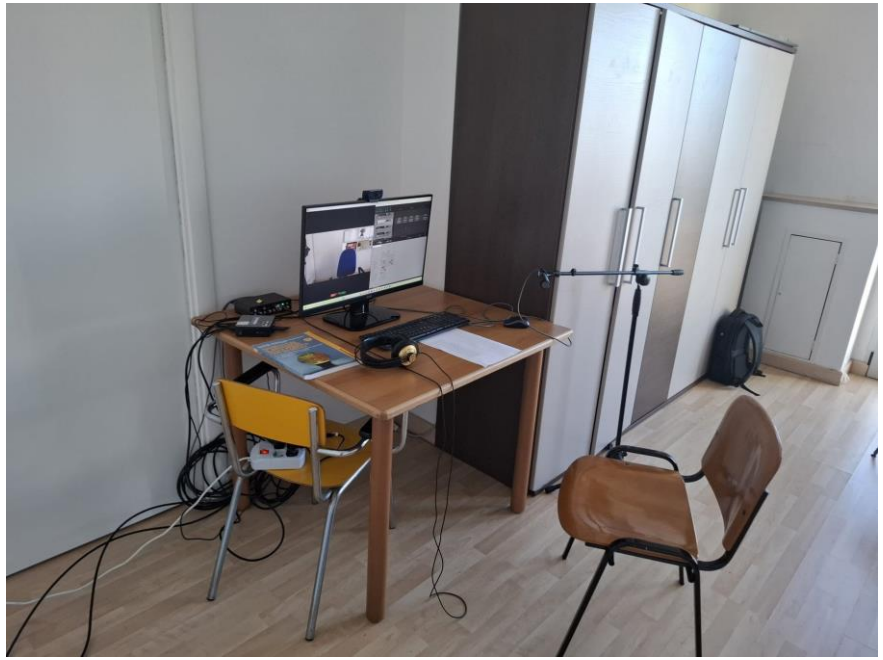
1. **Immersivo**
   il suono avvolge, come nella stessa stanza
2. **Bassa latenza**
   si può suonare insieme
3. **Zero configurazione**
   il paziente si siede e inizia

inspire

Esperimento presso INRCA



Stanza del partecipante



Stanza dell'operatore

IRCCS INRCA, Ancona · 12 persone con SCD · età media 75 anni · 2 sessioni da 40 minuti

inspire

Cosa ci hanno detto

«Sembra di essere a teatro. C'è profondità.»

«Ero isolato, non c'era rumore di fondo.»

100%

vorrebbe continuare

82%

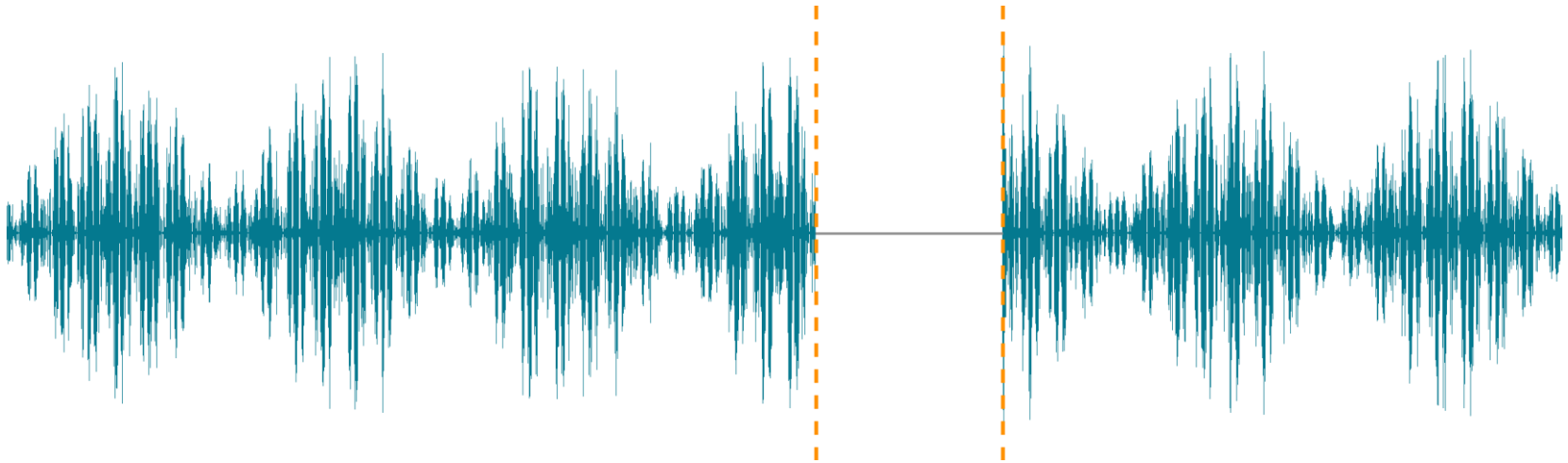
nessun problema tecnico durante le sessioni

inspire

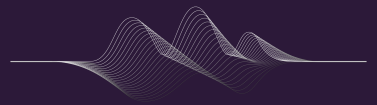
Sfide pratiche

In clinica la rete è dedicata e affidabile. A casa si passa da Internet: alcuni pacchetti audio arrivano tardi, o non arrivano.

Per tenere basso il ritardo non possiamo aspettarli: restano buchi nel suono.



inspire



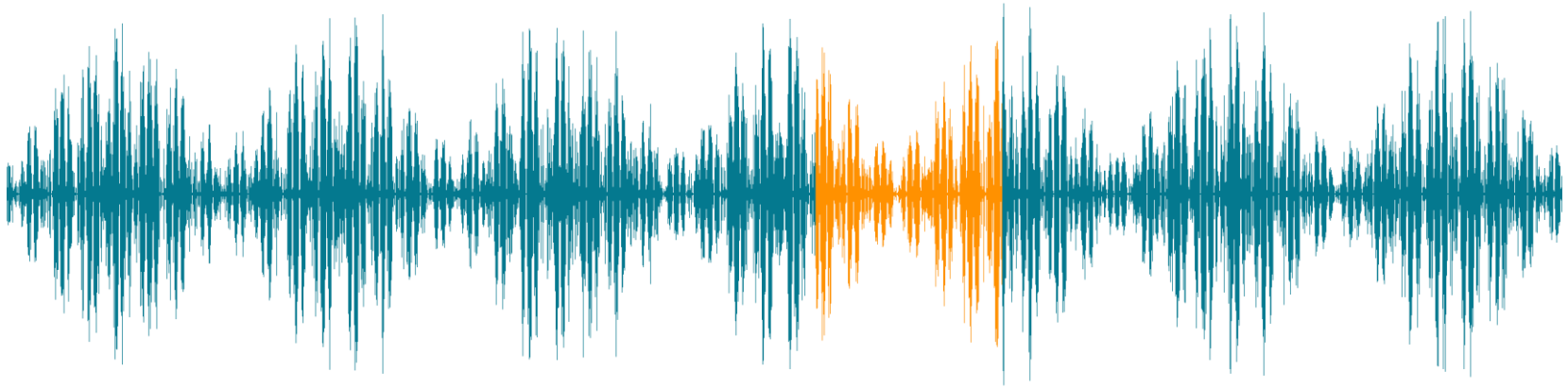
inspire

Generative AI for Music Inpainting

Packet Loss Concealment

Quando un pacchetto si perde, una rete neurale genera al suo posto un frammento plausibile.

Necessarie forti capacità predittive: quale sarà la prossima articolazione, nota, evento sonoro?



inspire

Sfida ulteriore: farlo in real-time

Offline: disponibile informazione successiva



Si conosce ciò che viene prima e dopo il "buco"
Nessuna restrizione sul tempo di calcolo

In tempo reale: processing causale



Manca conoscenza del futuro
Pochi ms di tempo per il calcolo!

Aironi, Gabrielli, Cornell, Squartini, «Complex-bin2bin: A Latency-Flexible Generative Neural Model for Audio Packet Loss Concealment», IEEE Trans. on Audio, Speech and Language Processing, 2025

inspire

Sfida ulteriore: farlo in real-time

Offline: disponibile informazione successiva



Si conosce ciò che viene prima e dopo il "buco"
Nessuna restrizione sul tempo di calcolo

In tempo reale: processing causale



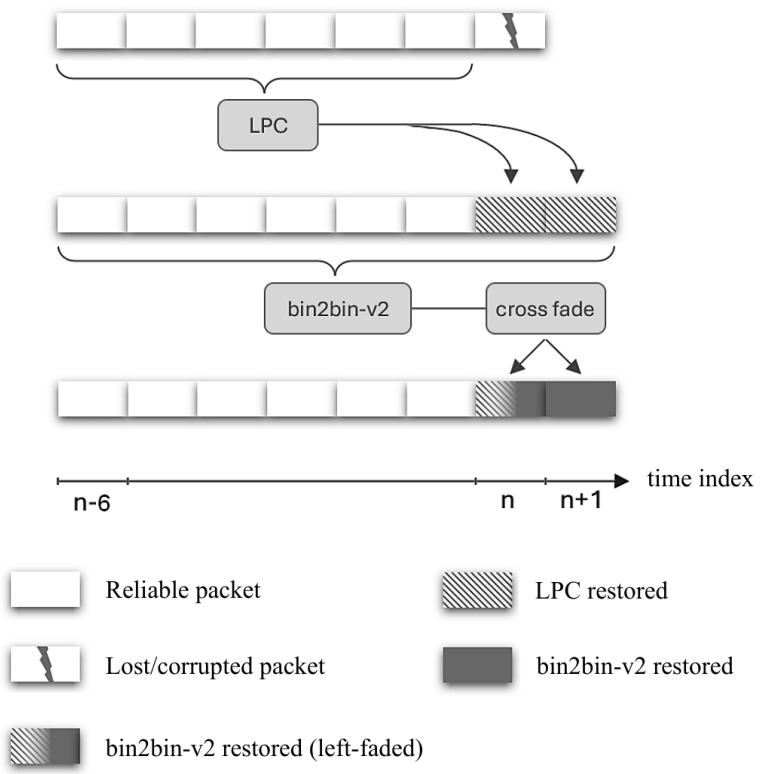
Manca conoscenza del futuro
Pochi ms di tempo per il calcolo!

I2S PLC CHALLENGE 2025 WINNER (Aironi et al.)

Il metodo che mostreremo ora è risultato vincitore della PLC Challenge 2025 organizzata all'interno di IEEE I2S

inspire

Algoritmo Proposto alla Challenge



1

Una stima veloce

La predizione lineare (LPC) «prolunga» il segnale appena ricevuto.

2

Una rete generativa la rifinisce

bin2bin-v2 è una GAN leggera "made in UNIVPM"

3

Una dissolvenza incrociata

Il pezzo ricostruito si raccorda al resto, senza click.

Tempi: 8,9 ms per pacchetto su CPU (<11,6 ms, limite temporale challenge)

Aironi, Gabrielli, Cornell, Squartini, «Restoring Music Integrity via Predictive Modeling and Spectral Inpainting», IEEE IS² Music PLC Challenge, 2025

inspire

Futuro prossimo: «parlare» in token

Oggi l'audio viaggia sempre più spesso compresso da **codec neurali**, come sequenze di simboli (token).

Allora il pacchetto perso si può prevedere come fa un LLM con la parola successiva: un Transformer causale indovina i token mancanti.

Lavoro in revisione (PLCLM), oggi sulla voce.



Il modello completa la sequenza: niente buchi da riempire nel segnale, solo simboli da indovinare.

inspire

PLCLM: i risultati

Esperimenti: voce compressa con codec neurali, packet loss: dal 10% al 60%

> +100%

di qualità percepita rispetto al segnale con i buchi

< 10%

di parole sbagliate fino al 30% di pacchetti persi

7 ms

per ricostruire un pacchetto perso

Nonostante i vincoli imposti dal processing causale: dal 30% di perdite in su il modello causale batte quello che «vede» il futuro, con metà dei parametri e un quarto della latenza (11,6 contro 46,4 ms).

inspire

inspire

Obiettivo: accessibilità

Brain-Computer Interfaces

Cosa sono: interfacce composte da elettrodi posizionati sul cuoio capelluto o all'interno del cranio (e.g.: Neuralink)

Risultati notevoli solo per i dispositivi invasivi (attualmente 80 persone in tutto il mondo hanno ricevuto un intervento di questo tipo)

Con sistemi non invasivi: segnali deboli, rumorosi. Inoltre sono diversi da persona a persona. Per alcuni utenti il sistema non funziona, e la letteratura li chiama «BCI illiterate».

Il nostro obiettivo è poter dotare persone con disabilità motorie la possibilità di interagire attraverso la musica



inspire

Case study: Motor Imagery

1

Si immagina un movimento

«Muovo la mano destra» oppure «la sinistra», senza muoversi

→

2

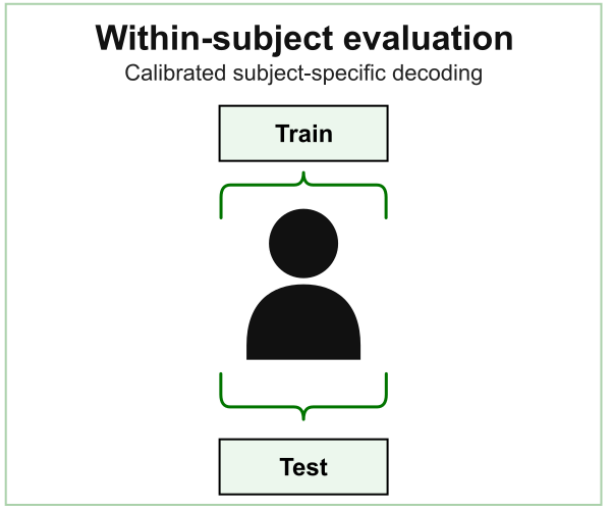
Un algoritmo lo riconosce

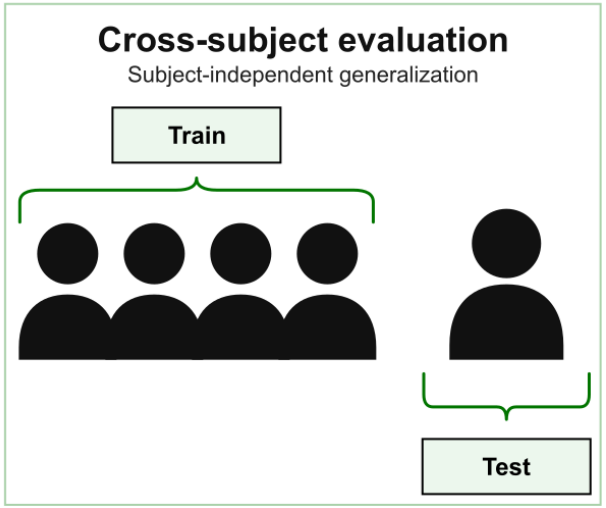
E lo trasforma in un comando, per chi non può usare le mani

Task "semplice" ma difficile riconoscere correttamente: su molti dataset la Accuracy di classificazione del movimento può essere attorno al 70% e non oltre

inspire

Training modes





Calibrato sulla persona

si allena il sistema con i dati dello stesso utente

Pronto all'uso

si allena su altre persone e si prova su una persona nuova

inspire

Scelta del metodo

Calibrato sulla persona

Miglior risultato con tecniche DSP+ML tradizionali

43 su 54 soggetti

Pronto all'uso

Miglior risultato con Deep Neural Networks (EEGnet)

35 su 54 soggetti

Molte persone che col metodo classico sembravano «BCI illiterate» ottengono risultati migliori usando Deep Neural Networks in classificazione!

Bergamino, Fioretti, Gabrielli, Squartini, «Benchmarking Machine and Deep Learning Approaches for Motor Imagery BCIs», WIRN, 2026

inspire

Next steps

Realizzare uno spettacolo multimediale con soggetti disabili

Copenhagen – giugno 2027

inspire

La promessa dell'AGI per l'accessibilità

Caso specifico: persone non vedenti o ipovedenti (BLV)

1. Il LLM ascolta la domanda
2. Il LLM consulta documenti e istruzioni, osserva la situazione
3. Il LLM fornisce indicazioni vocali al soggetto

Una promessa enorme per l'accessibilità: un assistente che vede, legge e spiega.

inspire



Senza di lui non avremmo mai cominciato...

inspire

Case study: software musicale

Con lo screen reader: l'interfaccia letta ad alta voce, o su un display braille

1

L'app descrive ogni elemento

Nome, ruolo e stato di pulsanti, menu, cursori

→

2

Lo screen reader lo annuncia

Con la sintesi vocale o in braille

→

3

L'utente naviga da tastiera

Un elemento alla volta, avanti e indietro

Esempio: *«Mute, pulsante, disattivato»*

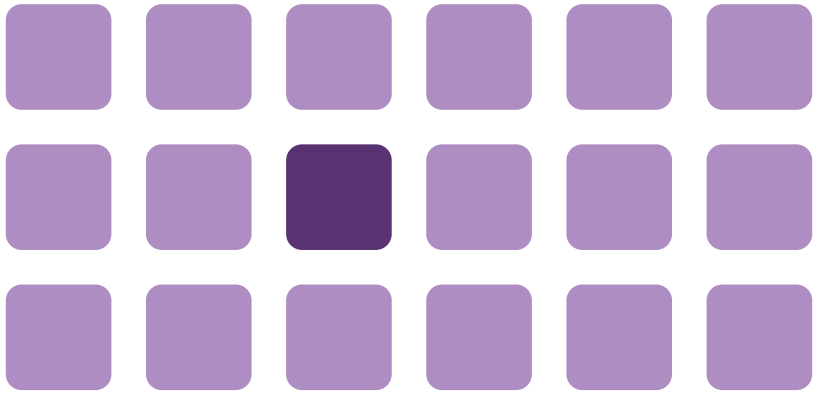
inspire

Il limite: tutto in fila

Ascoltare un'interfaccia è come farsi descrivere una stanza un oggetto alla volta.

Bisogna ricordare tutto: cosa c'è, dove si trova, cosa è cambiato. Più l'interfaccia è complessa, più cresce il carico sulla memoria.

Con la vista: tutto insieme



Con lo screen reader: uno dopo l'altro



inspire

Le interfacce musicali

Software musicale

Workstation e plugin con decine di manopole, spesso invisibili allo screen reader. Molti produttori ciechi dipendono da un assistente vedente.



inspire

Plugin che si lasciano leggere

Stiamo lavorando attivamente all'interno della community di ricercatori per sensibilizzare verso uno sviluppo progettato fin dall'inizio attorno alle esigenze BLV. L'ecosistema "JUCE" si è da poco dotato di strumenti adatti per gli sviluppatori.



inspire

Il paradosso dell'hardware

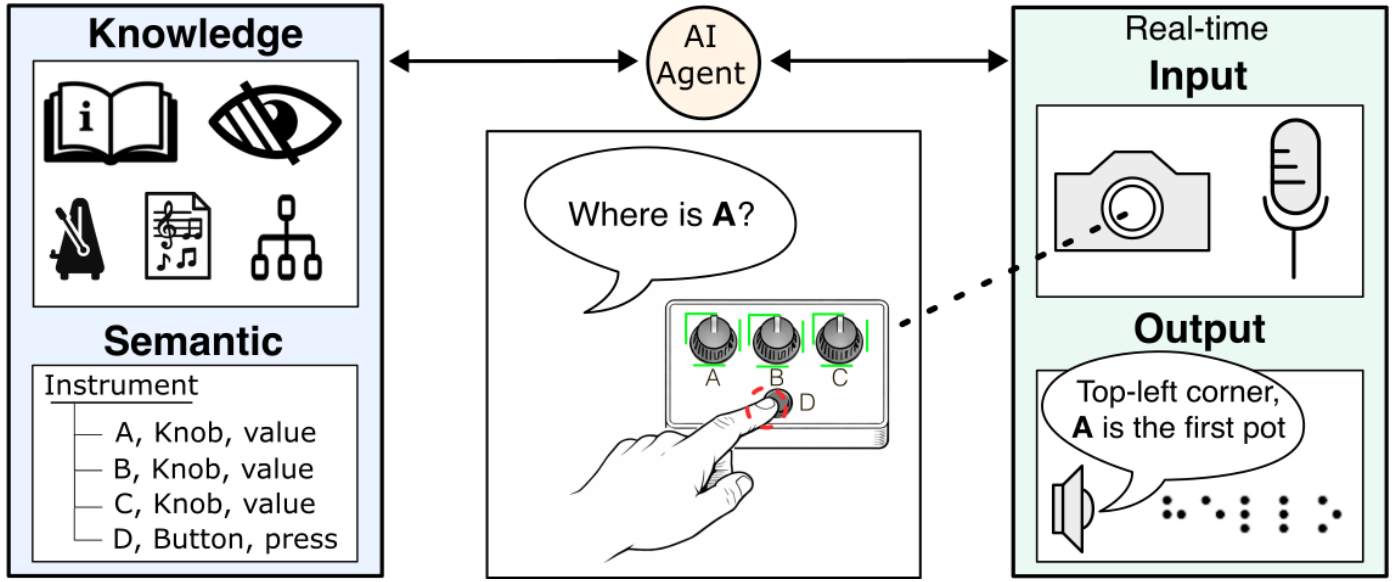
Oggi un buon software accessibile può essere più accessibile di uno strumento hardware!

L'obiettivo

Un aiuto che funzioni sugli strumenti che esistono già, senza toccarne hardware o firmware.



Un'AI che «legge» lo strumento



Il musicista chiede «Dov'è A?». L'agente guarda con la telecamera, consulta ciò che sa dello strumento e risponde a voce o in braille: «In alto a sinistra, è la prima manopola».

inspire

Flusso informazioni

Prerequisiti

- Manuale utente
- Istruzioni web accessibile (WAI-ARIA)
- Istruzioni da un esperto di comunicazione con non vedenti

Osservazione

- Dove si trova la mano
- Quali LED sono accesi, cosa dice il display
- Le precedenti azioni dell'utente

Risposta

- A voce o in braille
- Spiega di più quando si impara
- Quasi in silenzio sul palco

inspire

Un esperimento: Korg Monotribe

«Quale controllo uso per cambiare l'ottava dell'oscillatore?»



«Usa il selettore OCTAVE. È in alto a sinistra, il primo controllo della sezione dell'oscillatore.

Al tatto: è un selettore a scatti, non una manopola continua.»

Risposta dell'agente Surface Reader (tradotta e abbreviata)

inspire

Primi risultati

Un attuale LLM commerciale allo stato dell'arte compie numerosi errori:
- lessico non sempre adeguato
- non comprende alcune problematiche di localizzazione per i non vedenti
- costo in termini di token troppo elevato per uso continuativo interattivo

E ancora:
- competenze di sound design scarse o assenti
- competenze musicali scarse

inspire

Costruire competenze musicali

Lo stesso limite, in ambito industriale

Costruiamo sistemi RAG per diverse aziende del settore audio e delle tecnologie musicali.

1

Assistenza clienti

Risposte rapide e corrette su prodotti, accessori, procedure

2

Memoria aziendale

Conservare conoscenze sparse in manuali, schede e persone

3

Lavoro più veloce

Tecnici e sviluppatori trovano subito l'informazione giusta

Un LLM generalista non è in grado di dare risposte adeguate, non comprende alcune complessità

Inoltre: non è in grado di rispondere a domande che coinvolgono suoni

Come per Surface Reader, la competenza va costruita.

inspire

Un dominio da specialisti

Per rispondere a un cliente servono competenze che raramente stanno in una sola persona.

Organologia	Storia della musica	Specifiche dei prodotti	Competenze di Elettronica	Sound Design

Domande tipiche

- *«Che funzione ha il pistoncino A.P.?»*
- *«Quale uscita è compatibile con cuffie da 40 Ω?»*
- *«Come assegno un canale a un'uscita ausiliaria?»*

inspire

Cosa abbiamo imparato

I nostri prototipi possono lavorare su una workstation di costi ridotti

90%

domande complesse con risposta corretta

5 s

per risposta con modelli da pochi mld params

< 0,5 s

per cercare i documenti: il resto del tempo lo spende il modello

ricerca e riordino dei risultati

Possibilità di sfruttare reasoning per collegare documenti tra loro e costruire nuove conoscenze prima non presenti

inspire

Servono anche le orecchie

Molta assistenza tecnica si fa ascoltando: un ronzio, un tasto che «suona male», un registro stonato.

Machine listening

Algoritmi che ascoltano un difetto e lo riconoscono, come farebbe un tecnico esperto.

Classificazione audio

Dal suono registrato dal cliente a un'ipotesi di guasto, da unire alle risposte del RAG.

inspire

Vincoli reali

Dati e sicurezza

- Documenti riservati: niente invio a servizi esterni
- Requisiti di sicurezza delle informazioni, come TISAX
- Accessi diversi per clienti, tecnici, sviluppatori

Sostenibilità

- Economica: macchine di calcolo potenti ma non troppo costose
- Durata: hardware che non diventi obsoleto in poco tempo
- Ambientale: consumi proporzionati al compito

La risposta: modelli leggeri per cercare, rispondere e ragionare, eseguiti in locale.

Focus importante sulla fase di retrieval e l'assistenza all'azienda nel curare i dati ed espandere le conoscenze in mano ai dipendenti mai documentate per iscritto

inspire

Conclusioni

- Che significa AI? Tanti metodi per sfide e scopi diversi
- Metodi Deep Learning specializzati per l'elaborazione dei segnali ci aiutano a risolvere meglio problemi complessi
- Metodi basati su LLM ci permettono di immaginare nuovi scenari applicativi

- Le più disparate applicazioni musicali possono trarre beneficio da queste tecnologie...

... senza dimenticare i fondamenti teorici e le tecniche tradizionali di elaborazione

inspire

OPEN FOR COLLABORATIONS!

prof. Leonardo Gabrielli <l.gabrielli@staff.univpm.it>
DII - UNIVPM

